Towards Improved Breast Cancer Diagnosis using Microwave Technology and Machine Learning

Presented by:
Bárbara Luz Oliveira

to:
Electrical and Electronic Engineering,
College of Engineering and Informatics,
National University of Ireland Galway,

in fulfillment of the requirements for the degree of
Doctor of Philosophy.

Supervised by:
Dr Edward Jones

Co-supervised by:
Dr Martin Glavin
Dr Martin O'Halloran

September 2018

Contents

Declaration of Originality

I, the Candidate **Bárbara Luz Oliveira**, certify that this thesis entitled **"Towards Improved Breast Cancer Diagnosis using Microwave Technology and Machine Learning"**:

- Is all my own work;
- Has not been previously submitted for any degree or qualification at this University or any other institution;
- And where any work in this thesis was conducted in collaboration, appropriate reference to published work by my collaborators has been made and the nature and extent of my contribution has been clearly stated.

Name: _____
Bárbara Luz Oliveira

Abstract

Today, the early detection of breast cancer for asymptomatic women primarily relies on generalised screening programmes with x-ray mammography. However, the long-term value of screening mammography has been questioned due to e.g. the high false positive rate, resulting in unnecessary biopsies and overdiagnosed cancers. In this context, a need exists for new breast screening modalities with greater specificity. In this thesis, a machine learning platform using microwave technology is investigated for the purpose of diagnosing breast cancer. The proposed platform is evaluated by means of numerical and experimental phantom sets designed and developed in this research.

The proposed numerical tumour phantom set is designed to ensure tumour models are clinically-realistic; a validation procedure was undertaken with clinicians to verify the applicability of the models. Experimental tumour and breast models were also developed using tissue mimicking materials. Both the level of spiculation in the tumour models and varying levels of glandular content were included as novel elements of the phantom set.

Breast cancer diagnosis was investigated through the development of a 3-stage automated platform to analyse backscattered signals, which includes: data acquisition; data pre-processing through tumour windowing and feature extraction; and diagnosis through a random forests classifier in a two-level architecture. The results demonstrate the usefulness of creating high-similarity groups of signals before the classification and how the extraction of features can capture the characteristics of a tumour, without the need for *a priori* information. Results also show that benign tumours are more often correctly classified than malignant tumours, suggesting that microwave breast systems may be capable of achieving high specificity rates. In the context of current approaches to breast cancer care management, the results of this work support the potential value of microwave breast diagnosis systems to impact patient outcomes.

Acknowledgements

The work that I proudly, and finally, present in this thesis was developed in the DSP and Communications Laboratory within the Discipline of Electrical and Electronic Engineering, NUIG. I would like to sincerely thank my primary supervisor, Dr Edward Jones, and co-supervisors, Dr Martin Glavin and Dr Martin O'Halloran, for trusting me with the opportunity to join their research group, and for all of the support throughout the years. I thought the journey would be long, but your guidance and advice allowed me to reach a final destination that I am proud of!

Without Science Foundation Ireland, this project would not have been possible, so I thank them for backing up my work (Grant No. 12/IP/1523 and 11/SIRG/I2120). I would also like to acknowledge the support of the MiMed COST Action (TD1301), which gave me the unique opportunity to collaborate with international research groups. And thank you also to all of my GRC members for their advice, Dr Maeve Duffy and Prof Gearóid Ó Laighin, and all the staff of Electrical and Electronic Engineering for their support, in particular, Mary Costello, Myles Meehan, Martin Burke and Liam Kilmartin.

A special word of gratitude goes to Dr Raquel Conceição. Raquel, you have been such an integral part of my PhD, I could thank you forever. Thank you for your friendship, for your guidance, and for all the knowledge you were willing to share with me! I would also like to thank the IBEB institute in Lisbon, for hosting me (once again) and providing us the space to develop more work together.

To the TMD Lab and Dr Martin O'Halloran, thank you for your advice and for allowing me the opportunity to put my technical work into perspective in a clinical environment. Having this additional space to work from was invaluable! To Dr Emily Porter and Dr Adam Santorelli from the TMD Lab, thank you for sharing your expertise on breast phantoms, hardware, and overall guidance on how to produce good research.

I would also like to sincerely thank Dr Rachel Hilliard for all of the work and effort in organising Thesis Bootcamp. Being able to attend the weekend

camps gave me the space I needed to be able to process my thoughts and finally put my thesis down in words. It would have been a longer road without it, thank you!

To my friends in Ireland. You have made my life here all the better, and I am lucky for it. To everyone in Engineering, in the TMD Lab, the brunch club, and Tagtonic Plates, thank you! To name but a few, Anthony, Ashkan, Brian, Clodagh, Damien, Emma, Katie, Kevin, Martin, Niamh, Pat, Richie, Shane, Toto, Úna, thank you for all of the lunch and tea time talks, and for all of the good times we have had on the pitch. Edel, Filipe, Heidi, Kevin and Mar, your friendship is truly invaluable, thank you for always being there. And to Dave, Darragh and Declan, what a pleasure it was seeing this journey through with you!

To my friends in Portugal. To Cláudia, Luís, Rita, Xana and Rui. To my friends from Biomedical Engineering. I am so immensely grateful we have been able to keep our friendship going strong. Having you always ready to listen and help meant the world throughout my PhD.

And finally, on a more personal note, I could not thank my parents, Palmira and Victor, brother, Gonçalo and grandmothers, Etelvina and Fátima, enough. I truly could not have done it without your endless support and love. Thank you for all of the calls, thank you for helping me see past any obstacles, thank you for your encouragement, thank you for everything!

List of Figures

List of Tables

Acronyms

AC All Channels.
AUC Area Under the ROC Curve.

B-F Back-to-Front.
BI-RADS Breast Imaging Reporting and Data System.
BRIGID Breast Imaging and Diagnosis.

DAS Delay-And-Sum.
DCIS Ductal Carcinoma *In Situ*.

EA Equal Angle.
EAC Equal Angle Combined.

FDTD Finite-Difference Time-Domain.
FE Feature Extraction.
FWHM Full-Width Half-Maximum.

GRS Gaussian Random Spheres.

H-T Head-to-Toe.

KNN k-Nearest Neighbours.

L-R Left-to-Right.
LCIS Locular Carcinoma *In Situ*.
LDA Linear Discriminant Analysis.

M Mean.

MA Multiple Angle.

MARIA® Multistatic Array Processing for Radiowave Image.

MRI Magnetic Resonance Imaging.

PCA Principal Component Analysis.

PLA PolyLactic Acid.

PSD Power Spectral Density.

QDA Quadratic Component Analysis.

RIDER Reference Image Database to Evaluate Therapy Response.

ROC Receiver Operating Characteristic.

SD Standard Deviation.

SEER Surveillance, Epidemiology, and End Results.

SSIM Structural Similarity.

SVM Support Vector Machine.

TMM Tissue-Mimicking Material.

TSAR Tissue Sensing Adaptive Radar.

TW Tumour Windowing.

UWCEM University of Wisconsin Computational Electromagnetics Laboratory.

WHO World Health Organization.

CHAPTER 1

Introduction

1.1 Societal Context

Breast cancer is a high-incidence disease worldwide that primarily, but not exclusively, affects women, with an estimated 1% of breast cancer cases being detected in men [1]. Although the incidence and mortality rates due to breast cancer vary globally, breast cancer remains the most commonly diagnosed cancer and the most common cause of cancer death in women. According to the World Health Organization (WHO), in the year 2012, an estimated 1.7 million women were diagnosed with breast cancer worldwide and 520,000 died from the disease [2].

Early detection is widely established as the cornerstone of breast cancer care; breast cancers identified at an early stage allow for more effective treatment to be used, thus reducing the probability of death from the disease [3]. In fact, the five-year survival rate for early localised breast cancer exceeds 80% in countries with the appropriate resources for early detection of cancer and timely access to treatment [4]; in contrast, the survival rate may decrease to 10–40% in countries with limited health resources, where cancer is often only detected at a later stage [5]. Survival rates may also be subject to regional fluctuations depending on the different geographical profiles within a country. For example, the urban versus rural areas of a country may find that the level of access to health care differs, which, in turn, may impact early detection of breast cancer. Regional patterns may also reflect a number of other risk factors such as the age profile, ethnicity, lifestyle, among others [6]. Considering the numerous risk factors for breast cancer, risk assessment plans personalised to different populations would be beneficial in breast cancer care; they are, however, not yet a reality, further highlighting the importance of early detection of breast cancer [7].

As defined by WHO, early diagnosis and screening programmes are two strategies that work together towards the early detection of breast cancer [7]. Early diagnosis relies on improving public and professional awareness of the signs and symptoms of breast cancer and focuses on providing timely access to clinical assessment and necessary treatment. In contrast, screening programmes may rely on a number of medical examinations, for example breast imaging, to identify cancer in asymptomatic populations, ideally at a stage where it poses less of a threat to life.

Whether for the clinical assessment of symptomatic cancer or for screening programmes, the imaging of the breast has become immensely important as it provides the means to observe the interior of the breast in a non-invasive manner, potentially allowing identification of the presence and stage of development of cancer. The following sub-section provides an overview of the methods commonly used in clinical practice to image the breast.

1.2 Current Screening Practice

X-ray mammography, usually just referred to as mammography, is one of the leading breast imaging methods to be used in a hospital setting.

Mammography uses low-energy x-rays to identify abnormalities in the breast (i.e., tumours); the x-ray attenuation coefficient varies depending on the tissue, resulting in an image that shows the contrast in x-ray attenuation between healthy and cancerous breast tissues. The dark areas of a mammogram correspond to the fatty, healthy tissues of the breast, and the white areas correspond to higher density tissues of the breast, i.e. potentially cancerous masses or the dense breast tissue. Figure 1.1 shows an example mammographic image, of a heterogeneously dense breast [8].

The fact that cancerous masses and dense breast tissue attenuate x-rays similarly may result in missed cancer detections, which is especially concerning in younger women who are known to have denser breast tissue. This limitation of mammography affects the detection performance, namely in terms of the sensitivity[1]; other limitations of mammography have been identified that may impact the detection performance, such as the technology used (film-based or digital) or the inter-reader variability. For example, a study for the population of the United States between 2004 and 2009, identified that the performance of mammography in a screening context may vary with age; particularly the sensitivity varied between 73.4% and 88.5%, and

[1]Sensitivity measures the proportion of positive results (i.e., having cancer) that are correctly identified as such; sensitivity can also be referred to as true positive rate. Imaging methods with high sensitivity yield low levels of false negatives.

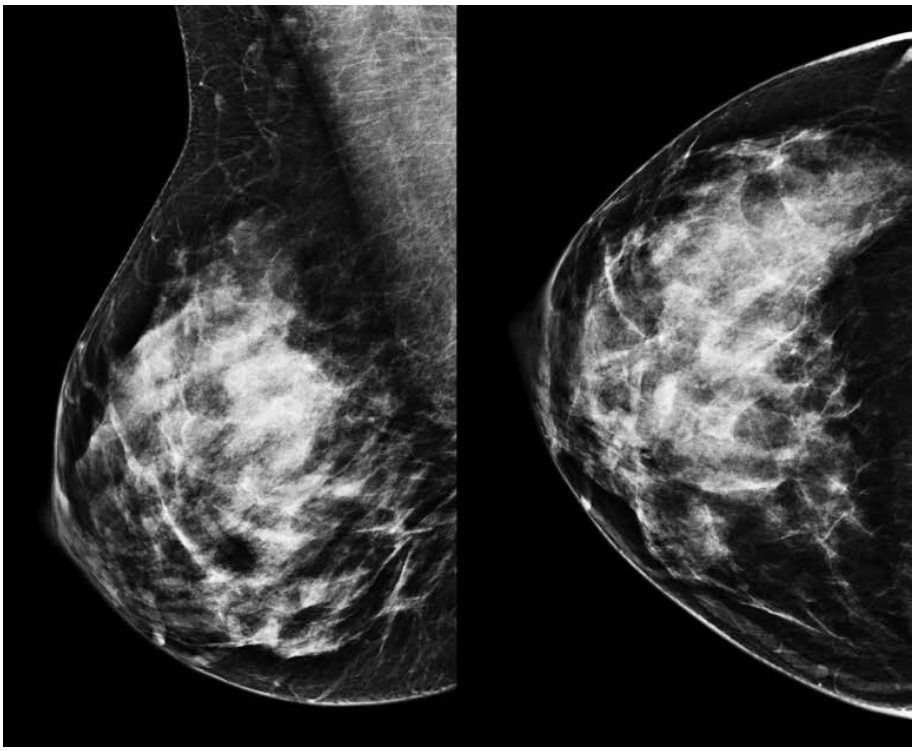


Figure 1.1: Example of a mammogram [8]: the medio-lateral-oblique view is shown on the left, and the cranio-caudal view on the right. In the image, the widespread white areas correspond to the dense breast tissue, which may obscure small masses, while the dark areas correspond to the fatty tissues.

the specificity[2] between 83.8% and 93.8% [9]. Additionally, mammography relies on ionising radiation, calling for additional guidelines on the safe level of human exposure to radiation, and requires an uncomfortable level of compression in the breast during the imaging procedure to obtain good quality images. Despite its limitations and known variations in performance, mammography is the most popular breast imaging method due to its cost and performance in the overall population.

Ultrasound and Magnetic Resonance Imaging (MRI) are two other leading breast imaging modalities, with recommended uses different to mammography. Ultrasound breast imaging relies on sound waves to obtain images of the breast and the breast structures. Historically, ultrasound was mainly used to distinguish cysts from solid masses; more recently, advances in

[2]Specificity measures the proportion of negative results (i.e., not having cancer) that are correctly identified as such; specificity can also be referred to as true negative rate. Imaging methods with high specificity yield low levels of false positives.

technology, such as 3D ultrasound, allow for a fine level of characterisation of benign and malignant solid masses, as well as the evaluation of suspicious findings in women with dense breasts and negative mammograms [10]. Additionally, ultrasound is recommended for guiding biopsies and other interventional breast procedures. Ultrasound does not require ionising radiation and is relatively low-cost and readily available; however, the performance of ultrasound imaging is highly operator-dependent, limiting its widespread use for the detection of small, non-palpable tumours [11].

MRI uses the natural magnetic properties of the human body to produce images, by using strong magnetic fields and radio waves. Although MRI produces high resolution images that are highly sensitive, its use in breast cancer care is only recommended for a particular group of women with whom MRI has shown promising results. For example, MRI is recommended for the screening of patients with over 20% lifetime risk of breast cancer, for the evaluation of size and extent of cancer in recently diagnosed patients, and for the monitoring of treatment response in patients undergoing neoadjuvant chemotherapy [12]–[14]. The main limitation of breast MRI is the low specificity of the images, reported to range from 37% to 97% [14], which leads to an increased number of unnecessary biopsies. In addition, MRI may be uncomfortable, with patients reporting severe claustrophobia during examinations, is time consuming and costly, and has limited availability due to the specific hospital settings required for its operation.

Between mammography, ultrasound and MRI, only mammography has been shown to perform well enough in asymptomatic populations to be used on its own for the screening of breast cancer, albeit not in every country worldwide. There is, however, significant uncertainty on whether the benefits of screening mammography in terms of mortality reduction outweigh the risks, even in countries with the health system to support it. The risks of screening mammography include an increase in the number of overdiagnosed cancers[3] and false positives, unnecessary examinations and treatments (such as invasive biopsies and mastectomies), and anxiety and psychological distress subsequent to a false positive result.

A number of different studies published since the 1960s indicate that there may have been a reduction in breast cancer mortality up to 20% due to the implementation of screening mammography, although, questions exist as to whether the observed decrease in mortality is in fact due to screening mammography or the better treatment options that have been developed since. Additionally, any reduction in breast cancer mortality comes at a cost,

[3]Overdiagnosis is the term used when a condition is diagnosed that would otherwise not go on to cause symptoms or death [15].

with studies reporting increased rates of overdiagnosed cancers, unnecessary biopsies and false positive results for each breast cancer death avoided [7], [11], [16]–[21]. Overall, a need exists for new imaging modalities that are less prone to false positives and overdiagnosing breast cancer. Microwave breast imaging is one such modality that has shown clinical promise in recent years, and this is the primary technology investigated in this research.

1.3 Research Motivation

The detection of breast cancer using microwave technology relies on contrast in dielectric properties between healthy and cancerous tissues at microwave frequencies. Microwave breast systems use low-power microwaves that are non-ionising, do not require uncomfortable levels of breast compression — a notable advantage over mammography — and may be comparably low-cost. In addition, microwave systems for breast cancer detection may be integrated into portable platforms, and may thus potentially be used in geographical regions without well-developed healthcare systems.

At the time of writing, at least seven microwave prototypes for breast cancer detection have been developed, and their performance has been evaluated in pilot clinical trials with healthy or cancerous patients [22]–[37]. The results of the trials performed to date are encouraging, and demonstrate the potential of microwave breast imaging in identifying abnormal breast findings; for example, one such trial has shown sensitivity up to 86% in detecting breast cancer when imaging patients with dense breasts [31]. With the initial proof-of-concept established for microwave breast imaging, the question now arises, are microwave breast systems able to accurately differentiate breast cancer from abnormal breast findings that are benign?

In the context of today's breast screening programmes, answering this question is relevant from a clinical perspective as it means exploring mechanisms that allow accurate differentiation between positive and negative results. In fact, many automated breast diagnosis systems have already been proposed to work in conjunction with mammography, ultrasound or MRI, and are used by clinicians to identify features in a signal or image that may otherwise be missed through visual inspection alone [38], [39]. In the context of microwave systems, dedicated detection and diagnosis software could play a key role in establishing the potential of the technology in terms of breast monitoring and screening.

While microwave breast analysis systems have been primarily focused on imaging, a small number of studies has suggested that both detection and diagnosis of breast abnormalities with microwave technology is potentially

feasible by using the backscattered signals collected during a scan [40]–[53]. Differences exist in the backscattered signals captured from breasts with healthy tissue only, breasts with a benign abnormality, or breasts with a malignant finding, and machine learning algorithms may be able to capture such differences in an automated way. However, the majority of these studies have been tested with numerical data, only representative of relatively simple scenarios (for example, low density breasts), and it is unclear whether the performance of machine learning-based platforms is adequate in clinically-complex scenarios.

This thesis examines the topic of diagnosing breast abnormalities using microwave technology, and investigates the potential of such systems in clinically-realistic settings. Primarily, this thesis explores mechanisms to capture the difference between benign and malignant breast abnormalities in breasts with complex backgrounds, such as the case of dense breasts, as well as the use of machine learning algorithms towards the development of an automated diagnosis platform capable of low false positive rates. To achieve this, multiple research avenues were pursued:

1) Development of new feature sets to capture the differences in microwave signals from breasts with different types of abnormal findings.
2) Evaluation of suitable ways to develop an integrated platform using machine learning to diagnose breast cancer based on microwave signals.
3) To overcome the limitations with data available to validate such a diagnosis platform, clinically-realistic breast and tumour models were also explored.

The primary contributions of the work developed in this thesis are outlined in the following section, as well as the journal and conference publications arising from each work package.

1.4 Thesis Contributions

The contributions of the work arise from two primary work packages that are interconnected: 1) the development of breast models that include realistic representations of breast abnormalities; 2) the development of an automated diagnosis platform for breast cancer using microwave technology and machine learning, which was tested using the clinically-realistic models produced. Each work package involved both numerical and experimental elements. The specific novel contributions are summarised as follows:

1) Development of numerical breast tumour phantoms, clinically validated to be representative of tumours of different shapes and levels of malignancy;
2) Design and fabrication of experimental breast phantoms, including increasing levels of breast density, and experimental breast tumour phantoms, representative of tumours of different shapes and levels of malignancy;
3) Fundamental analysis of the impact of microwave signal acquisition in the quality of diagnostic information contained in backscattered signals;
4) Study of machine learning practices in the field of microwave breast diagnosis;
5) Proposal of a new set of features that capture the difference in microwave signals between benign and malignant breast tumours;
6) Development of an automated breast diagnosis platform based on microwave signals, and validation of the diagnosis platform with both the numerical and experimental sets of phantoms developed with this thesis.

Further contributions arising from collaborative work developed for this thesis include:

1) Design, building and testing of an experimental microwave imaging prototype system that builds on the experience gained with existing prototype systems;
2) Development of open-source and extensible imaging software in line with best practice, which can be used for image reconstruction with many types of system design.

These contributions are described in detail in the following chapters. The journal and conference papers already published, as well as other outputs arising from this research, are listed in the following sub-sections.

1.4.1 Journal Publications

[1] **Oliveira, B. L.**, O'Halloran, M., Conceição, R. C., Glavin, M., Jones, E., "Development of Clinically Informed 3-D Tumor Models for Microwave Imaging Applications", *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 520–523, 2016.

[2] O'Loughlin, D., **Oliveira, B. L.**, Elahi, M. A., Glavin, M., Jones, E., Popović, M., O'Halloran, M., "Parameter Search Algorithms for Microwave Radar-Based Breast Imaging: Focal Quality Metrics as Fitness Functions", *Sensors*, vol. 17, no. 12, p. 2823, 6th Dec. 2017.

[3] **Oliveira, B. L.**, O'Loughlin, D., O'Halloran, M., Porter, E., Glavin, M., Jones, E., "Microwave Breast Imaging: Experimental tumour phantoms for the evaluation of new breast cancer diagnosis systems", *Biomed. Phys. Eng. Express*, vol. 4, no. 2, p. 025 036, 2018.

[4] **Oliveira, B. L.**, Godinho, D., O'Halloran, M., Glavin, M., Jones, E., Conceição, R. C., "Diagnosing Breast Cancer with Microwave Technology: Remaining challenges and potential solutions with machine learning", *Diagnostics*, vol. 8, no. 2, pp. 1–22, 2018.

[5] **Oliveira, B. L.**, O'Loughlin, D., O'Halloran, M., Conceição, R. C., Glavin, M., Jones, E., "Experimental Validation of a Machine Learning Platform for Diagnosing Breast Cancer with Microwave Technology (In Preparation)",

1.4.2 Conference Publications

[1] **Oliveira, B. L.**, Glavin, M., Jones, E., O'Halloran, M., Conceição, R. C., "Avoiding unnecessary breast biopsies: Clinically-informed 3D breast tumour models for microwave imaging applications", in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Memphis, TN, USA, 6th–11th Jul. 2014, pp. 1143–1144.

[2] **Oliveira, B. L.**, Shahzad, A., O'Halloran, M., Conceição, R. C., Glavin, M., Jones, E., "Combined Breast Microwave Imaging and Diagnosis System", in *Proceedings of Progress In Electromagnetics Research Symposium*, vol. 1, Prague, Czech Republic, 6th–9th Jul. 2015, pp. 274–278.

[3] **Oliveira, B. L.**, O'Halloran, M., Glavin, M., Jones, E., "Breast Cancer Diagnosis: Development of a Dedicated Computer-Aided Detection and Diagnosis System for Microwave Breast Imaging", in *22nd Bioengineering in Ireland (BINI)*, Galway, Ireland: Royal Academy of Medicine in Ireland, 22nd–23rd Jan. 2016, p. 119.

[4] **Oliveira, B. L.**, Conceição, R. C., Shahzad, A., O'Halloran, M., Glavin, M., Jones, E., "Breast Cancer Diagnosis Revisited: A non-invasive classification system for Microwave Breast Imaging", in *6th Postgraduate Research Day NUIG-UL Alliance*, Limerick, Ireland, 29th Apr. 2016.

[5] O'Loughlin, D., **Oliveira, B. L.**, Glavin, M., Jones, E., O'Halloran, M., "Adaptive Microwave Breast Imaging: Experimental Performance Evaluation", in *24th Bioengineering in Ireland (BINI)*, Meath, Ireland: Royal Academy of Medicine in Ireland, 27th Jan. 2018.

[6] O'Loughlin, D., Elahi, M. A., Porter, E., Shahzad, A., **Oliveira, B. L.**, Glavin, M., Jones, E., O'Halloran, M., "Open-Source Software for Microwave Radar-Based Image Reconstruction", in *12th European Conference on Antennas and Propagation (EuCAP)*, London, the UK: IEEE, 9th–13th Apr. 2018.

[7] O'Loughlin, D., **Oliveira, B. L.**, Glavin, M., Jones, E., O'Halloran, M., "Effects of Interpatient Variance on Microwave Breast Images: Experimental Evaluation", in *40th Annual International Conference of the 40th IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA: IEEE, 17th–21st Jul. 2018.

1.5 Thesis Outline

The remainder of this thesis is organised as follows.

Chapter 2 provides a detailed review of the background to this thesis. The anatomy of the breast is first described, including a characterisation of benign and malignant breast findings. The guidelines for the clinical assessment of breast cancer are also discussed, both for symptomatic and asymptomatic populations. Next, the core principles of microwave breast systems are discussed: the dielectric properties of the breast, signal acquisition, and image reconstruction; the state-of-the-art microwave breast prototypes are also described. Finally, a review of the literature on the topic of machine learning applied to microwave breast systems is also presented, including both the detection and diagnosis of breast cancer based on backscattered signals.

Chapter 3 presents the two phantom sets developed in this work. The numerical tumour phantom set is produced from a clinically-representative modelling algorithm, which allows for several shapes and levels of spiculation. The procedure to validate the generated tumour models with input from clinicians is also described. Next, the design and fabrication of the experimental breast and tumour phantom set is presented. The experimental phantom set includes physical implementations of the numerical tumour set, varying both in shape and level of spiculation, as well as breast phantoms that model the variation in breast tissue composition observed in the population.

Chapter 4 proposes a machine learning-based platform, designed to diagnose the level of malignancy of a tumour based on the information contained in backscattered signals. The methodology employed to develop this platform is first described and emphasises the need for good machine learning practice. The proposed diagnosis platform is verified by testing with the numerical phantom set described in the previous chapter. Here, the use of the numerical dataset is intended to provide a thorough investigation of the complexities in diagnosing breast cancer with backscattered signals, when no other confounding factors are present.

Chapter 5 re-assesses the diagnosis platform proposed in Chapter 4 by means of the experimental breast and phantom set presented in Chapter 3. The experimental prototype used to acquire backscattered signals from the experimental phantom set is first described. An in-depth analysis of the results of experimental testing is also presented, which includes a sensitivity and specificity evaluation of the diagnosis process.

Finally, Chapter 6 summarises this thesis, and discusses the main results and conclusions. Future work is identified to further improve the microwave breast diagnosis platform proposed in this thesis. Future work avenues for the long-term clinical acceptance of microwave breast detection and diagnosis systems are also discussed.

CHAPTER 2

Background and Literature Review

2.1 Introduction

An understanding of the anatomy and other physical properties of the female breast is important in the development of microwave breast detection and diagnosis systems. Elements such as the size and shape of the breast influence the design of microwave systems, while the internal composition of the breast and the dielectric properties of breast tissues influence their operation. In this chapter, an overview of the anatomy of the breast and formation of breast cancer is first presented in Section 2.2, which includes a review of the risk factors for breast cancer and a discussion of guidelines for the clinical assessment of symptomatic and asymptomatic breast cancer.

Next, the fundamental principles behind microwave imaging systems are presented in Section 2.3, from three perspectives: the dielectric properties of breast tissues; the typical setup of microwave breast systems; and the algorithms commonly used for processing, particularly image reconstruction. In this section, an overview is also presented of some microwave breast imaging prototype systems that have already been tested in a clinical environment.

In Section 2.4, a review of the use of machine learning in combination with microwave technology to detect and diagnose breast cancer is presented. Feature extraction and classification methods previously used in machine learning platforms are considered. Concluding remarks are presented in Section 2.5, highlighting the main research questions addressed in this thesis.

2.2 Breast Anatomy, Cancer Development and Clinical Practice

2.2.1 Anatomy of the Breast

The female breast is a superficial structure protruding from the thoracic wall presenting with a wide range of shapes and sizes. The breast can assume different shapes, such as round, hemispherical or conical, and can appear more firm or pendulous. Breast size can also vary greatly. Studies identify that breasts can vary between 100 mL to over 2,000 mL in volume, and 9.7 cm and 18.2 cm in diameter [54]–[56]; the wide variation in breast volume and dimensions is a key factor to consider in the development of new microwave breast systems. Among others, the different breast shapes and sizes are determined by genetic, ethnic and dietary factors [57].

The breast is composed of different types of tissues, namely, glandular and supporting fibrous tissue embedded within a matrix of adipose (fatty) tissue, and surrounding skin. A simplified diagram of the breast's anatomy is shown in Fig. 2.1. The glandular tissue forms 15 to 20 lobes radiating from the nipple, which resembles a cone-type structure. Each lobe is subdivided into lobules, the milk-producing glands. Lobules are, in turn, a cluster of small sacs called acini, which are connected to the nipple through branching ducts. The functional units of the breast are the terminal duct lobular units, which encompass the lobules and the intralobular and extralobular terminal ducts, as identified in the inset in Fig. 2.1. The circular pigmented area of skin surrounding the nipple is termed the areola. The glandular tissue within the breast is surrounded by the stroma, which is composed of fat and fibrousconnective tissue. The latter is responsible for providing structural support to the breast in the form of fibrous bands or suspensory ligaments (named Cooper's ligaments), connected both to the skin and the underlying fascia. Together, the fibrousconnective tissue and the glandular tissue correspond to the dense areas of the breast.

When describing its anatomy, it is also important to note that the upper-lateral region of the breast extends into the axilla, forming an axillary tail. In fact, the axillary lymph nodes are the main site where the lymph from the breast drains to, and therefore a location where cancer cells are often found [57]–[59]. In addition to the axillary tail, six other sub-sites are used by clinicians to describe the locations of masses found in the breast: upper-outer quadrant, upper-inner quadrant, lower-outer quadrant, lower-inner quadrant, central portion and nipple. An overview of the location system is shown in Fig. 2.2. Alternatively, the time on the face of a clock and

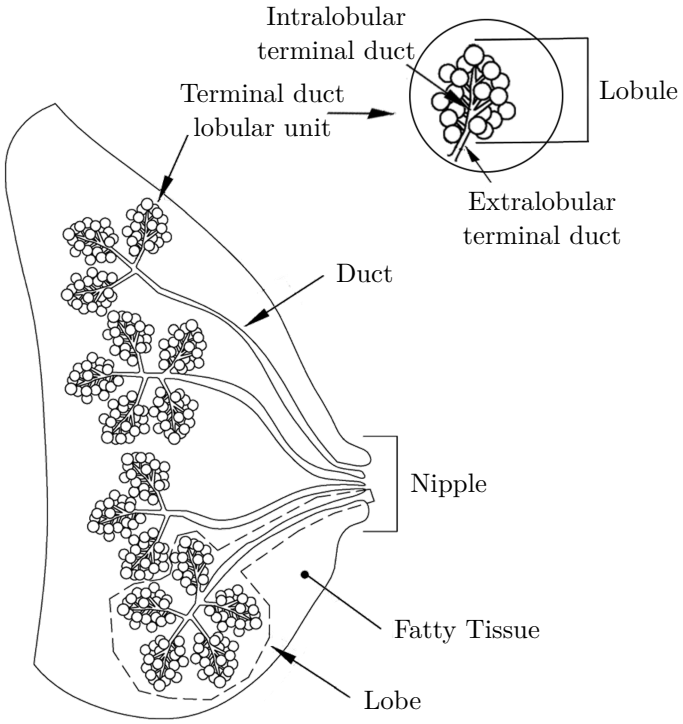


Figure 2.1: Simplified diagram of the anatomy of the breast, including the nipple, lobes, lobules and ducts, which are embedded within a matrix of fatty tissue. The terminal duct lobular unit is the functional unit of the breast and is shown in the inset. Adapted from [60].

the distance in centimetres from the nipple can be used for describing the location of masses found in the breast.

2.2.2 Changes in Breast Tissue

Cell behaviour includes a series of aspects that determine how the organism operates as a whole, namely cell metabolism, growth, multiplication, differentiation, survival and death. Normal cell behaviour is ensured by a series of regulatory mechanisms, which rely on signalling molecules that exist in all multicellular organisms. At times, this regulation of normal cell behaviour is lost, and cell growth and multiplication happen in an uncontrolled manner, originating masses of cells, which are termed as tumours or neoplasms [62]. Depending on whether they invade nearby organs or metastasise through the body, tumours may be classified as benign or malignant.

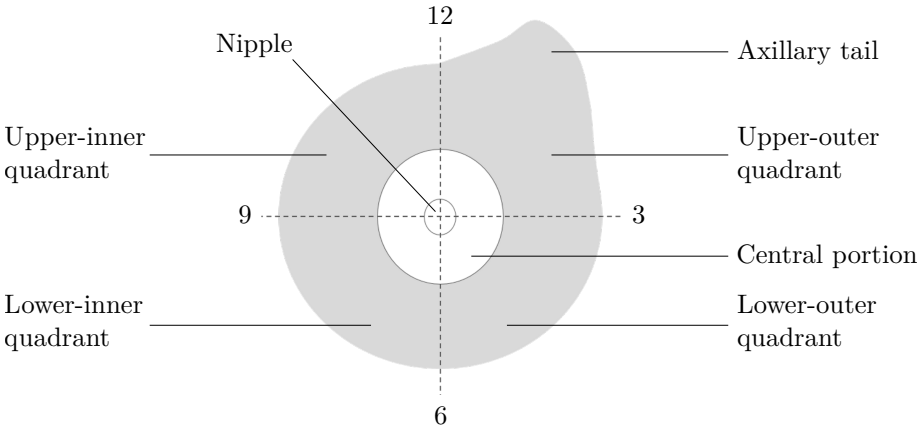


Figure 2.2: Schematic representation of the human breast (left breast displayed, viewed from the front). Seven sub-sites are used to identify specific portions of the breast. The time on the face of a clock and the distance in centimetres from the nipple can also be used for describing the location of masses found in the breast [61].

2.2.2.1 Benign Changes

Benign masses remain confined to their original location, and generally do not invade nearby tissues or metastasise to distant body sites; they are usually surrounded by a fibrous capsule and tend to grow along the normal structures of the breast [62], [63]. Benign changes in the breast usually present as palpable lumps that may move upon touch, and may result from normal ageing of the breast, increased hormone sensitivity or other factors. The most common types of benign masses in the female breast are fibroadenomas (epithelial tumours containing fibrous tissue) and cysts (closed cavities in the body containing a liquid or semi-solid material). Importantly, they usually have distinct architectural features, showing round or lobular shapes with smooth and well delineated margins. Other forms of benign breast abnormalities include [59], [64]:

- Fibrocystic disease, which refers to a breast with lumps that may feel painful;
- Duct ectasia, which refers to blockage of the lactiferous ducts;
- Duct papillomas, which are tumours that form in a duct;
- Epithelial hyperplasia, which is an increase in the number of cells lining the terminal duct lobular unit. Atypical duct and lobular hyperplasia are common examples of benign breast diseases, which involve an increase of cells that look abnormal under a microscope.

2.2.2.2 Malignant Tumours: Invasive and *In Situ*

In contrast to benign changes in breast tissue, malignant tumours are capable of invading surrounding normal tissue and metastasising through the body via the circulatory or lymphatic system; they tend to grow in irregular directions, disrupting existing anatomical patterns [62], [63]. The vast majority of breast cancer cases found in women correspond to breast carcinomas, which are malignant tumours found in the epithelial cells of the terminal duct lobular units of the breast [65], [66].

Depending on whether the cancer cells invade nearby tissues, tumours may be classified as invasive or *in situ*. Invasive carcinomas are the most prevalent form of breast cancer, accounting for up to 80% of the diagnosed cancer cases in the United States in 2017 [6]. With invasive breast cancers, cells spread outside of the basement membrane of the ducts and lobules into the surrounding tissue. Invasive tumours tend to grow in irregular directions with irregular shapes and angular margins, resulting in protrusions referred to as spicules; malignant invasive tumours usually present with many microlobules or spicules at their margins. According to the WHO, the types of invasive breast cancers are [67], [68]:

- Infiltrating ductal carcinomas;
- Lobular carcinomas;
- Special subtypes of invasive carcinoma including, tubular, cribriform, papillary, medullary or mucinous;
- No special type (referred to as not otherwise specified).

When cancer cells are confined within the basement membrane of the terminal duct lobular units of the breast, the tumour is classified as *in situ*; with *in situ* cancer forms, there is no evidence of further tissue invasion. Although less aggressive, *in situ* breast cancer has the potential to evolve into invasive cancer and is considered a true cancer precursor, as will be discussed in Section 2.2.3. According to the WHO [67], [68], the two sub-types of *in situ* breast cancer are:

- Intraductal carcinoma, also referred to as Ductal Carcinoma *In Situ* (DCIS). This is the most common form of a non-invasive breast carcinoma, corresponding to approximately 83% of all diagnosed *in situ* cases [6], [64]. DCIS refers to a condition where cancer cells replace the normal epithelial cells in the breast ducts, potentially expanding the ducts and lobules;

- Locular Carcinoma *In Situ* (LCIS), which is the development of an epithelial tumour in the lobules of the breast, and may consequently expand the lobules.

2.2.2.3 Calcifications

Calcifications are calcium deposits in the breast, usually resulting from natural ageing. Calcifications are usually not felt by palpation, but they are detectable by current breast imaging modalities, particularly, mammography, due to the high contrast in X-ray attenuation. Although calcifications do not represent abnormal changes in the breast per se, they usually appear co-located with other suspicious findings within the breast tissue, and can be helpful in detecting regions of the breast where suspicious breast abnormalities might appear.

While macrocalcifications, typically coarse and larger in size, are usually no reason for concern, microcalcifications have been shown to appear in areas of rapidly dividing cells. Microcalcifications are highly correlated with the appearance of DCIS [64], which may otherwise not be detected in mammograms. The size and density of microcalcifications may vary and they may appear clustered. Determining the location, morphology and number of calcifications in a cluster may be helpful in determining the malignancy of a suspicious region detected in a mammogram.

2.2.2.4 Location of Breast Abnormalities

Based on the localisation system identified in Fig. 2.2, a number of clinical studies have investigated the incidence of breast cancer based on primary tumour site and breast laterality, reaching similar conclusions, e.g. [69]–[72]. From the identified studies, the most comprehensive analysis was performed by the Surveillance, Epidemiology, and End Results (SEER) Program of the United States National Cancer Institute [72], containing data for over 1 million breast cancer cases.

Table 2.1 details the results from the SEER database regarding primary breast tumour site and laterality, for the United States population in the period between 1973 and 2011 [72]. The highest breast cancer incidence is observed in the upper-outer quadrant of the breast, for a total of 32.7% of the cases. The left breast is also slightly more common for cancer occurrence, totalling 50.2% of all cases versus 48.5% corresponding to the right breast. Furthermore, it should also be considered that tumours arise predominantly within radiologically dense breast tissue, i.e., breasts with a higher ratio of glandular and fibrousconnective tissue [73]–[75].

Table 2.1: Results from the Surveillance, Epidemiology, and End Results (SEER) database detailing the relationship between breast cancer occurrence and breast tumour site and laterality, for the United States population between 1973 and 2011 [72].

		% of cases
Primary tumour site within the breast	Upper-outer quadrant	32.7
	Borderline between quadrants	30.0
	Upper-inner quadrant	9.5
	Lower-outer quadrant	6.4
	Central portion	5.7
	Lower-inner quadrant	5.0
	Nipple	0.8
	Axillary tail	0.8
	Not otherwise specified	19.1
Breast laterality	Left breast	50.2
	Right breast	48.5
	Bilateral	0.1
	Not otherwise specified	1.2

2.2.3 Risk Factors for Breast Cancer

Age has been identified as one of the most important risk factors for breast cancer across all populations and ethnicities. For example, for the United States population between 2010 and 2014, the rate of breast cancer incidence was shown to approximately double every 10 years until the age of 65; smaller increases were then observed until the age of 80 [6], [64], [76], [77].

In decreasing order of relative risk, other non-modifiable risk factors for breast cancer include: genetic predisposition, high breast density, personal history of breast cancer, family history, high levels of endogenous hormones in post-menopausal women, early start of menstruation, height (where tall women have been found to have a higher risk of breast cancer than shorter women), late age at first full-term pregnancy and late menopause. The risk factors are discussed in detail in [6]. Some breast cancer risk factors result from modifiable behaviours, such as exposure to radiation, alcohol consumption, never having breastfed, post-menopausal weight gain and obesity, recent use of menopausal hormones and recent use of oral contraceptives [6].

Importantly, the occurrence of *in situ* carcinomas and other benign diseases of the breast have also been linked to a higher incidence of invasive

breast cancer later in life [6], [64], [78]–[81]. DCIS is often considered both as a potential precursor to invasive breast cancer, and as an indicator of developing a new invasive breast cancer. In contrast, LCIS rarely becomes invasive cancer, but it is an indicator of increased risk. In relative terms, the risk associated with a diagnosis of DCIS or LCIS ranks in the same category as high breast density.

From the group of benign breast diseases, developing atypical hyperplasia of the ducts or the lobules indicates risk of further developing invasive breast cancer, ranking higher than DCIS or LCIS in terms of relative risk. Complex fibroadenomas and ductal papillomas also serve as indicators of increased breast cancer risk, although the relative risk is much lower than that associated with atypical hyperplasia. Other benign diseases, such as duct ectasia or simple cysts, are associated with little to no increased breast cancer risk.

2.2.4 Clinical Assessment of Breast Cancer

The clinical assessment of breast cancer can be perceived as happening simultaneously for two types of populations — asymptomatic, through screening programmes, and symptomatic, usually involving a diagnostic decision through medical testing. Whether presenting or not with symptoms of a developing cancer, a key determinant of breast cancer outcomes in any type of population is detecting the disease in an early stage. Early detection of breast cancer allows the effective deployment of early treatment therapies, ultimately contributing to improving breast cancer control in terms of patient mortality [3]. Here, breast cancer care for symptomatic and asymptomatic populations are discussed.

2.2.4.1 Symptomatic Breast Cancer

With symptomatic populations, the need is to confirm the diagnosis of a suspicious finding, rather than to detect it exists in the first place. Historically, a breast tumour would have to be completely excised and histologically examined to determine its diagnosis. Currently, other procedures are also used by medical experts to assist in the diagnosis of a tumour, for example:

- Physical examination, which refers to the act of physically examining the size and location of breast lumps that may exist;
- Breast imaging, where the most common clinical imaging tools are mammography, ultrasound and MRI;

- Fine-needle aspiration, a diagnostic tool to examine suspicious tissue; a biopsy of the tissue is performed by sampling a mass of cells with a hollow needle, which is then histologically analysed.

Initially developed in the 1970s [82], and later updated in 1998 [83], a procedure termed the triple test has exhibited great effectiveness in diagnosing breast tumours. The triple test consists of a combined evaluation of a suspicious finding using physical examination, mammography, and fine-needle aspiration; each independent examination scores the finding with 1 point if benign, 2 if suspicious and 3 if malignant. A combined score less than or equal to 4 determines the finding as benign, and greater than or equal to 6 as malignant. The triple test has shown comparable accuracy to excision and histological analysis of the tumour [84]. Importantly, the triple test has also shown higher sensitivity and specificity when compared to deploying either of the single procedures independently, sometimes as high as 100% [85], [86]. Due to its high diagnostic ability, the triple test has become standard practice for diagnosing breast cancer in many countries [87].

Particularly regarding the clinical assessment of breast cancer through imaging, a set of guidelines have been created by the American College of Radiology that aims at standardising the reporting practices of breast imaging within the medical community. This set of common imaging descriptors and assessment guidelines is called the Breast Imaging Reporting and Data System (BI-RADS) [8], [88]–[91], and includes detailed versions for mammography [8], ultrasound [90] and MRI [91]. Particularly, BI-RADS defines a set of terms to describe tumour shape, the internal composition of the breast in terms of dense tissue, the morphology and distribution of calcifications, and also a categorisation scheme to assess the likelihood of malignancy of a tumour. Ranging from benign to malignant, six assessment categories are linked to higher likelihoods of malignancy, coupled with a clinical management recommendation that determines whether follow-up imaging or tissue diagnosis is required.

Once the diagnosis of a malignant tumour has been confirmed, whether it is *in situ* or invasive breast cancer, the extent of the disease should be assessed and the tumour staged. Clinicians usually refer to the TNM Staging System, created by the American Joint Committee on Cancer supported by the American Cancer Society [92]. This system is based on the size and extent of the Tumour (T), spread of the tumour to the lymph Nodes (N), and presence of Metastases (M). Notably, an extra class is created for *in situ* tumours when assessing the T size of the diagnosed malignancy.

Based on the combined TNM information, anatomic stage/prognostic groups are assigned, where Stage I cancers are the least advanced and often

have a better prognosis. Interestingly, the determination of tumour size appears to be relevant when staging the tumour, but not necessarily a key factor in determining its diagnosis.

2.2.4.2 Asymptomatic Screening of the Breast

Asymptomatic populations do not present with symptoms of a developing cancer, for example, palpable lumps in the breast, and usually do not present with risk factors either, such as the ones detailed in Section 2.2.3. As asymptomatic breast cancer may still be present, a number of countries have implemented nation-wide mammography screening programmes, with the aim of detecting breast cancer early. Lately, however, the value of breast cancer screening mammography has come into question, with recent studies suggesting that screening mammography may have little effect in mortality reduction [7], [11], [16]–[21].

The benefit of screening mammography is the reduction in mortality due to breast cancer, while the risks include an increase in the number of overdiagnosed cancers and false positives, and anxiety and psychological distress subsequent to a false positive result. An increase in the number of overdiagnosed cancers is particularly concerning as it refers to detected cancers that may not cause any symptoms during a woman's lifetime and are, therefore, treated unnecessarily. An increase in the number of false positives may also result in unnecessary invasive procedures to confirm a diagnosis, such as biopsies.

The available evidence for screening mammography is based on the results from randomised controlled trials, and in particular a number of trials performed in North American and European countries since the 1960s. The data from many of these trials have been reviewed extensively over the years (particularly, the influential systematic Cochrane review by Gøtzsche and Jørgensen [18]), and noticeably, no definite conclusion has been reached amongst the different studies [7], [11], [16]–[21]. This is due both to the variability of the trial methodology and collected data, and also due to the analysis methodology implemented by the many review studies available in the literature. Particularly, lead-time and length-time biases are two recognised biases that may impact the evaluation of the efficacy of screening programmes, and ultimately, conclusions drawn from the existing trials [17]:

- Lead-time bias refers to the length of time between the detection of cancer with screening, and its natural presentation had it not been screened (when it becomes symptomatic); i.e., the cancer is diagnosed early, but survival has not been prolonged;

- Length-time bias refers to the perception that screening leads to better outcomes, primarily based on the difference between slow and fast-growing tumours in terms of their likelihood of mortality; by definition, slow-growing tumours are less likely to recur and less likely to be fatal. Screening mammography tends to detect more slow-growing tumours, simply because they have longer asymptomatic phases when compared to fast-growing tumours, resulting in the wrong perception that cancers detected by screening are more likely to be treated and cured. Overdiagnosis can be seen as an extreme form of length bias.

Amongst the different trials, a reduction of up to 20% in breast cancer mortality due to screening mammography has been identified; however, questions do exist on whether the observed decrease in mortality is, in fact, due to screening mammography, or to better treatment options that have been developed since the start of formalised trials. Additionally, any reduction in breast cancer mortality comes at a cost; studies have reported that the ratio of breast cancer deaths avoided to number of overdiagnosed cancers varies between 1:2 and 1:5, with at least 152 false positives and 31 unnecessary biopsies per breast cancer death avoided [11], [18], [21].

Based on these results, several international breast cancer organisations have published position papers on screening mammography. For example, the Swiss Medical Board has found that the risks outweigh the benefits and has recommended that all screening programmes be abandoned [20]. However, the WHO provides a conditional recommendation for screening mammography depending on age and resource setting [7]:

- Screening is always recommended, at an interval of two years, for women aged 50–69 in countries with strong health systems;
- Screening is not recommended in countries with limited resources and weak health systems, with prompt access to effective diagnosis and treatment options favoured.

The uncertainty surrounding the results of the mammography trials discussed in this section casts doubt on the long-term value of screening programmes for breast cancer. New trials are needed to effectively assess the true value of screening mammography, but many years of trial data will be needed before their results can be discussed. In this context, a need exists for new imaging modalities that may succeed in the areas where mammography has been shown to struggle.

In the following section, microwave breast imaging is introduced as one such modality that has shown clinical promise in recent years. The

fundamental principles behind microwave systems for breast cancer detection are reviewed, and the typical setup of microwave breast systems and the algorithms commonly used for image reconstruction are also considered. A review is also presented of the main microwave breast prototypes tested in a clinical environment.

2.3 Microwave Technology for Breast Cancer

Microwave systems for human applications originated from the field of ground-penetrating radar, historically used for the purpose of subsurface applications such as detecting land mines and examining archaeology sites; all radio and microwave applications usually involve transmitting signals and collecting their radar returns in order to detect an object.

In the context of breast cancer, active microwave systems rely on the difference that breast tissues exhibit in their dielectric properties at microwave frequencies; typical microwave systems range from 0.5 GHz to 9 GHz. An advantage from the point of view of safety is that microwave breast imaging systems are non-ionising, unlike mammography. Besides active microwave systems, there are other applications for microwaves currently under development that could prove useful in the detection of cancer. For example, passive microwave systems use radiometers to identify tumours based on temperature differences between normal and cancerous tissues [93]–[95]. Hybrid systems have also been researched, for example [96], [97], where active microwave systems are typically combined with thermoacoustic modalities to create images of the breast.

Nowadays, one of the main applications for active microwave systems is to image the breast, for the purpose of detecting breast cancer; in fact, an extensive body of literature exists on the topic, exploring ways to improve microwave images of the breast. In the remainder of this section, a review of the existing research on dielectric properties is presented in Section 2.3.1, also addressing the uncertainty about the dielectric contrast between healthy and cancerous breast tissues; microwave system setup and signal acquisition is described in Section 2.3.2, and an overview of microwave image reconstruction methodologies is provided in Section 2.3.3. Additionally, a review of the some microwave imaging prototypes that have been proposed is also presented in Section 2.3.4.

2.3.1 Dielectric Properties of Breast Tissues

The study of the physiology of the breast in terms of the dielectric properties of breast tissues, i.e. relative permittivity and conductivity, is hugely important in the development of microwave breast systems. Specifically focussing on the human breast, a series of historical studies performed *ex vivo* measurements and report a significant contrast in the dielectric properties between healthy (fat and glandular) and cancerous tissues in the microwave frequency band [98]–[101]. However, the dielectric contrast reported across the historical studies varies greatly. This variation could be explained due to the different experimental methodologies employed across studies, where properties were measured under different temperature and frequency settings, and also due to the intrinsic heterogeneity of breast tissues and across patients [102].

Aiming to address the method limitations discussed above, Lazebnik *et al.* published one of the most comprehensive examinations to date of the dielectric properties of breast tissues in 2007 [103], [104]. Two studies were performed using a large sample size, a variety of tissue types and a consistent experimental methodology. The first study [103] used only tissue samples from breast reduction surgeries, to ensure healthy breast tissues were measured, and a later one [104] used tissue samples from cancer surgeries. In both studies, dielectric properties were measured over a wide frequency range, from 0.5 GHz to 20 GHz. All tissue samples from both studies were quantified through a histological analysis in terms of the percentage of fat, glandular, and fibrousconnective tissues, and also the percentages of tissue corresponding to benign or cancerous findings. In addition, a statistical analysis of the resulting data was performed.

Tissues were categorised into three groups depending on the percentage of fat content, with Group 1 containing less than 30% fat tissue (high water content), Group 2 containing 31% to 84% fat tissue, and Group 3 containing samples with more than 85% fat tissue (low water content). In addition, a group of cancerous samples was also analysed, containing 30% or more of malignant tissue (DCIS, LCIS, and infiltrating ductal and lobular carcinomas). The main findings of the two studies were:

- There is significant dielectric heterogeneity in healthy breast tissue;
- There is an inherent variability in the dielectric measurement, which is higher in the more heterogeneous samples (Group 2);
- Samples with increasing glandular content (high water content) show higher dielectric properties;

- Samples with increasing fat content (low water content) show lower dielectric properties;
- When comparing healthy tissues (samples with glandular tissue, with no more than 10% fat content) to cancerous tissues, a contrast of 8% was found in the relative permittivity and 10% in the conductivity;
- When the same comparison was performed but adjusting for the fibrousconnective tissue content, there were no statistically significant differences between the healthy and cancerous tissues.

The *ex vivo* studies by Lazebnik *et al.* [103], [104] provide a standard set of dielectric properties of healthy and malignant breast tissues. However, while the dielectric properties of cancerous tissues are in good agreement with the historical studies [98]–[101], the healthy-to-cancerous dielectric contrast found in the 2007 studies is much smaller than previously reported. In fact, the contrast may be nearly non-existent, considering the statistical analysis that adjusts both for glandular and fibrousconnective tissue content.

More recently, clinical trials have produced results that imply the situation may not be as challenging as implied in the studies by Lazebnik *et al.*, specifically in terms of the dielectric contrast between healthy and cancerous tissues. Clinical studies performed with a number of microwave prototype systems have been able to produce useful images of breast abnormalities, suggesting that a larger healthy-to-cancerous contrast ratio may actually exist [22], [25], [27], [28], [32], [105], [106]. Furthermore, new dielectric properties measurement studies have started to highlight possible reasons for the conflicting findings, where some of the identified factors include:

- Differences in the measurement of *in vivo* and *ex vivo* tissues, discussed in [107]–[110];
- The heterogeneity of breast tissue samples, and the impact of quantification of the cancerous cells in a tissue sample in the dielectric properties, as discussed in [111];
- Differences in the sensing depth and volume techniques used in the measurements, addressed in [112], [113].

Despite the open debate, the properties reported by Lazebnik *et al.* [103], [104] remain one of the most widely used set of dielectric properties in microwave breast studies. Furthermore, these properties represent a worst-case scenario, ultimately leading to the creation of more challenging test cases.

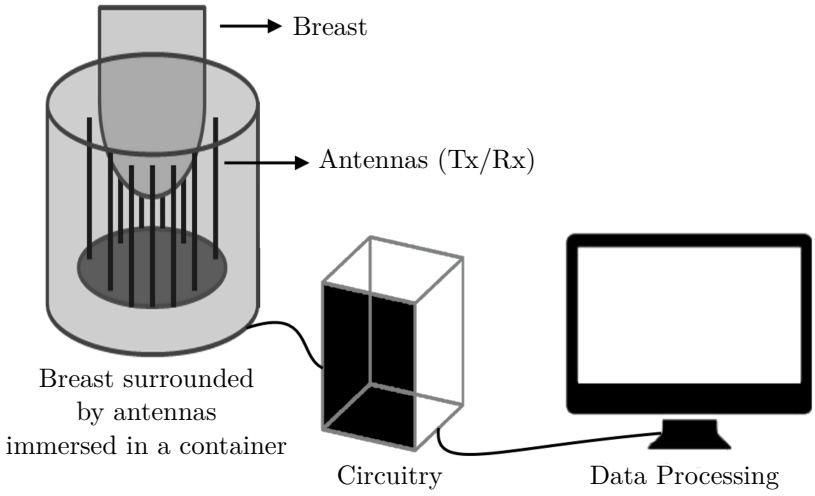


Figure 2.3: Simplified diagram depicting the operation of a microwave breast prototype. The breast is surrounded by an array of antennas (transmitting antennas, Tx, and receiving antennas, Rx), while immersed in a container. Backscattered signals are collected by data acquisition circuitry, which includes signal conditioning and sampling, and a data processing unit is then used to form an image of the breast.

2.3.2 System Architecture and Signal Acquisition

In a typical microwave breast setup, the breast is illuminated with pulses from antennas placed at a number of different locations surrounding the breast, where each antenna may act both as a transmitter (Tx) and a receiver (Rx). A simplified diagram of the operation of a microwave system for breast imaging or diagnosis is shown in Fig. 2.3. The interaction between the microwave signals and the breast tissues results in scattering of the signals, changing their speed, propagation path, phase, polarisation and strength; the surface at which the scattered field is sampled is referred to as the acquisition surface [114]. In a monostatic configuration, the transmitting and receiving antennas are co-located and backscattered (i.e., reflected) signals are collected at the point where they were generated. In a multistatic configuration, for any given measurement, one antenna transmits and multiple antennas receive; in this type of setup, both backscattered and transmitted signals are collected.

Several other factors are involved in the design of a prototype. Design choices may have an impact in the type of scan, scan duration and results obtained. Some of the design considerations include:

- Portability, microwave systems can be fixed (integrated on an examination table), or portable (for example, wearable or handheld);
- Positioning of the patient, prone or supine on an examination table, or seated;
- Coupling medium, systems may or may not require a coupling medium to interface between the antenna array and the breast; typically, the breast is in contact either with the air, a liquid immersion medium, or a solid coupling shell;
- Antennas, type of antenna, design of antenna array, and whether it operates in a monostatic or multistatic configuration;
- Acquisition, in the time- or frequency-domain, which determines hardware design;
- Choice of signal processing algorithms.

In microwave systems integrated into an examination table, the patient may lie prone and the breast hangs through an opening in the surface of the table. Typically, an array of antennas surrounds the breast, resulting in many transmission paths through the breast. With these systems, the chest area may suffer reduced visibility due to the thickness of the mat the patient lies on, and the axillary region may also not be seen in the images. In contrast, with systems requiring the patient to be seated or lie supine, the surface of the breast quadrants may not be entirely covered.

The choice of medium to interface between the antennas and the breast is also important. Systems where the breast hangs freely in the air are subject to breast movement, due to patient breathing or patient movement. Systems using an immersion liquid are also prone to breast movement; additionally, the surface of the breast may become naturally deformed by the liquid. The antenna array can be fitted to a biocompatible coupling shell, which is in direct contact with the breast and does not require any other immersion liquids. Although more straightforward, systems using coupling shells may be prone to poor contact between the shell and the breast.

Designing an antenna array includes the choice of antenna, the number of antennas, and their relative location to each other. Microwave systems may be stationary, where the antenna array is fixed to one unique position throughout the scan, or moving. The choice of antenna array design may impact the calibration of the system, as well as the interaction between different hardware components.

Signal acquisition may happen in the time- or frequency-domain. In a frequency-domain system, a network analyser is used to perform a frequency

sweep within a fixed frequency band. A switching circuit is also often used to allow multiple automated measurements, for monostatic or fully multistatic acquisition. In contrast, with time-domain systems, a pulse generator is used to produce an ultra-wideband time-domain pulse. Short pulses are transmitted sequentially by selecting antennas with a switching circuit, and backscattered signals are recorded at the receiving antennas.

The information contained in the backscattered signals can be used and processed in multiple ways. The creation of an image of the breast is usually the first approach employed to gain insight about the composition of the breast. With microwave imaging algorithms, the goal is to create an image of the breast that reflects the dielectric profile of the breast as a function of location. A variety of reconstruction algorithms has been proposed in the literature, operating in the time- or frequency-domain. A review of some of these algorithms is presented next.

2.3.3 Image Reconstruction

Microwave imaging algorithms aim at creating an image of the breast that reflects the dielectric profile of the breast as a function of location. Two main approaches are used to create microwave images of the breast, either 1) by computing a spatial distribution of the actual dielectric properties of breast tissues, or 2) by identifying areas of the breast with strong scatterers, i.e., areas of the breast where the dielectric properties differ from the normal surrounding tissue, suggesting the presence of tumours.

Microwave tomography algorithms aim to quantify the relative permittivity and the conductivity of the breast based on the scattered field measured by multiple antennas; tomographic algorithms are formulated as inverse problems, and may operate in the time- or frequency-domain. Examples of tomography algorithms proposed in the context of microwave breast imaging include Born linear approximations [115], non-linear gradient-based methods [116], [117], non-linear global methods based on evolutionary algorithms [118], or algorithms incorporating *a priori* information for example, from MRI scans [119].

Radar-based imaging algorithms operate by synthetically focusing the radar return signals to create an image of dielectric scatterers present within the breast. Most radar-based imaging algorithms are based on the Delay-And-Sum (DAS) beamformer [120], where all backscattered signals are time-shifted and summed to a synthetic focal point, resulting in a profile of backscattered energy in the breast. At points where a tumour exists, the backscattered signals add coherently, resulting in high energy regions of the image. On the contrary, reflections from normal heterogeneous breast tissue

result in low energy regions of an image. More sophisticated beamformers exist, for example, to compensate for attenuation and phase effects [121], reward the coherence of reflections from particular synthetic points [122], or compensate for time-shifting effects [123]–[128]. Recently, a comparative study of the performance of different radar-based image reconstruction algorithms with clinical data was published [129]. The results of this study show that the conventional DAS algorithm can accurately recover the clinical information contained in the images.

Prior to reconstruction, artefact removal algorithms are used to pre-process the signals. Backscattered signals are typically dominated by a number of artefacts resulting from the difference in relative permittivity between the immersion medium, the skin, and the interior of the breast. Of these, the most significant reflections arise in the early time of the signal, from the immersion medium-skin interface, and the skin-breast interior interface. This early time-time artefact can be orders of magnitude higher than the reflection from interior breast tissues, which may consequently mask internal scatterers (e.g., the tumour) and generate significant clutter in the resultant energy profiles. Current artefact removal algorithms include rotational subtraction [130], averaging subtraction [131], or adaptive filtering [132].

2.3.4 Clinical Prototypes

In recent years, several microwave imaging prototypes have been developed and trialled in clinical settings, albeit with small-to-medium sample sizes. A number of factors are involved in the design of a prototype, and, as investigations with microwave systems are only preliminary, no particular design choice has been deemed ideal thus far [133]. Among the largest trials published to date were those performed with a tomography-based system developed by the Dartmouth College, USA [22]–[26], and with the Multistatic Array Processing for Radiowave Image (MARIA®) system created by the University of Bristol, UK (now in commercial development) [27]–[31]. The main findings from these trials are summarised below.

Dartmouth College, United States, developed a system that has been used in a number of clinical trials since 2000, operating at a total scan time of 5 min per patient. With their table-based system, the patient lies prone, and a liquid immersion medium is required to interface between the antenna array and the breast. Their antenna array is composed of 16 monopole antennas in a multistatic configuration. The Dartmouth system is tomography-based, and uses a multi-frequency Gauss-Newton iterative algorithm for image reconstruction. The results from trials using the Dartmouth prototype with a healthy female population between the ages

of 40 and 79 (the largest of the healthy trials counted 43 volunteers) [22]–[24] suggested that the dielectric properties of breast tissues may be higher than reported in previous *ex vivo* studies, for example [103], [104]. The studies with healthy women also showed that water content in breast tissues dictates their dielectric properties, and observed a frequency-dependent decrease in conductivity and relative permittivity as the fat content increases. An additional study looked at the differences between benign and malignant tissues, by examining the same prototype with 53 healthy women and 97 women with abnormal findings, all aged between 35 and 81 [25]; this study noted that a difference exists in the dielectric properties of malignant tissues (masses greater than $1\,\mathrm{cm}^3$) compared to benign tissues, particularly, the conductivity. The most recent study using the Dartmouth prototype imaged 8 patients undergoing neoadjuvant chemotherapy for locally advanced breast cancer, where each patient was imaged 5 to 8 times over the course of the treatment [26]. In this study, conductivity was shown to be well correlated to the pathological response after 30 days of treatment, i.e., the decrease in tumour size due to the chemotherapy treatment was well correlated with a decrease in the conductivity of the tumour region. Once again, this result suggests that conductivity might be a key indicator for differentiating benign and malignant tissues.

The MARIA® system is integrated in an examination table where the patient lies prone and the system uses a coupling shell. The antenna array rotates to gather different views of the breast and uses 60 slot antennas in a multistatic configuration, for a total scan time of 10 s. The system operates in the frequency domain, and uses a radar-based image reconstruction algorithm, in combination with a rotational subtraction artefact removal algorithm. Several studies have been published detailing the results of the clinical evaluation of the MARIA® system with healthy patients and patients presenting with abnormal findings [27]–[31]. The latest study reports the results for a population of 223 women aged 32 to 89, where 80 had confirmed diagnoses of cancer [31]. Overall, this study reports a 75% sensitivity in detecting cancer, but interestingly, the sensitivity is as high as 86% when considering only the population of women who presented with dense breasts. This result is a positive indication towards the use of microwave breast imaging, as the current *de facto* breast imaging modality, mammography, is known to underperform with denser breasts. Additionally, results from trialling the MARIA® system also show that it is potentially possible to automatically discriminate between benign and malignant breast tissues based on their frequency-dependent dielectric response [29].

Other microwave breast systems have been developed by research groups worldwide. Although only trialled with a small number of patients (maximum

of 13 patients), the results are in line with the Dartmouth and MARIA® findings. Some examples include:

- The Tissue Sensing Adaptive Radar (TSAR) system developed by the University of Calgary, Canada [32]: this is a table-based system requiring an immersion medium, which operates in the frequency-domain and uses Vivaldi antennas in a monostatic configuration. This system operates at a scan time of 30 min per patient, and uses a DAS algorithm to compute images of the breast. The group from the University of Calgary has also developed a complementary system to estimate the bulk permittivity of breast tissues at microwave frequencies [134]; accurate and patient-specific permittivity estimation is of great interest as it may lead to more accurate images of the breast;
- The system developed by the Southern University of Science and Technology, China [33]: a table-based system requiring an immersion medium, which operates in the frequency-domain and uses horn antennas in a multistatic configuration. This system operates at a total scan time of 4 min, and uses adaptive filtering for artefact removal in combination with the DAS algorithm to create an image of the breast;
- The system from Shizuoka University, Japan [34], [35]: this is also a table-based system, but uses a coupling shell. This system operates in the frequency-domain and uses an array of antennas in a multistatic configuration. The total scan time with this system is 3 min, and it uses a rotational subtraction artefact removal in combination with DAS to create an image of the breast;
- The hand-held system of Hiroshima University, Japan [36]: this system was designed to be used with a patient lying supine on an examination table, and it uses a coupling shell to interface between the antennas and the breast. The Hiroshima system is radar-based, using the DAS algorithm in combination with averaging subtraction artefact removal, and operates in the time-domain;
- The wearable system of McGill University, Canada [37]: the wearable system is integrated in a bra, and uses an antenna array in a multistatic configuration. The total scan time of this system is 5 min, and it uses a differential artefact removal algorithm in combination with DAS, operating in the time-domain. In contrast to the previous systems, the McGill wearable prototype was designed as a breast monitoring tool, where patients are repeatedly scanned to identify changes in the breast tissues over time. In their trial, scans were found to be repeatable, although sources of variability were found, such as patient positioning.

The results from the prototypes discussed in this section are indicative of the potential behind microwave technology to image the breast. In most of the studies published to date using microwave technology, the main aim has been imaging the breast to detect cancer, i.e., to identify the presence of malignant tumours within the breast. However, as the field progresses, it becomes important to understand whether the technology has the potential to differentiate between true positives and true negatives, namely, can microwave systems discern between benign findings, such as cysts, and invasive tumours? Can imaging alone achieve such a goal, or are other signal processing applications necessary, such as machine learning?

In Section 2.4, the differences between benign and malignant tumours from the perspective of microwave systems are summarised, followed by a review of the state-of-the-art in one the core topic of this thesis: the use of machine learning algorithms to identify the presence and type of breast tumours with different levels of malignancy.

2.4 Machine Learning Applied to Microwave Breast Technology

As discussed in Section 2.3.4, microwave images generated from clinical data in small scale pilot trials have suggested it may be possible to detect breast cancer, even in dense breasts [31]. However, determining the malignancy of detected tumours has not been a primary concern in these clinical studies, leaving an open question as to the potential of microwave technology to diagnose abnormal breast findings. In principle, a number of avenues could be explored to distinguish breast tumours according to their level of malignancy, including:

- The dielectric properties of tumours may provide indications that a contrast between benign and malignant tumours exists;
- The detection of microcalcifications, as these have been shown to appear in areas of rapidly dividing cells;
- The shape of tumours, as there is extensive clinical evidence indicating that the shape of benign and malignant tumours differs substantially, particularly, the spiculation (i.e. prominent spike-like protrusions) at the margin of a tumour.

Additionally, from a microwave perspective, the diagnosis of breast tumours could rely on:

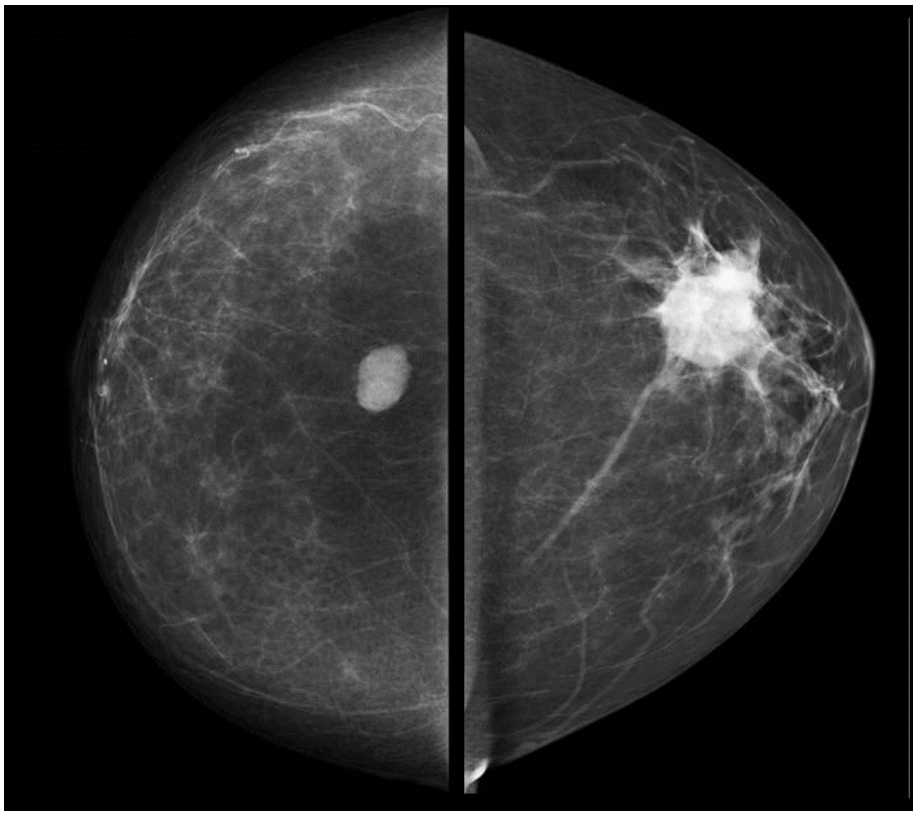
- Images (e.g., radar or tomography-based);
- Backscattered signals.

Concerning the use of dielectric properties for the diagnosis of breast tissues, there are positive preliminary indications from small patient studies. One study by Dartmouth College, USA, used their microwave tomography system to image women with diagnosed cancer, fibrocystic disease, fibroadenomas and other benign abnormalities [25]; the resultant tomographic images suggest that a difference does exist in the dielectric properties of benign and malignant breast tumours, particularly the conductivity. More recently, the radar-based MARIA® system has also been used with women with diagnosed cancer and women with fibroadenomas and cysts, and results also show that the dielectric properties of benign and malignant breast abnormalities might differ [29].

However, further investigations characterising the range of properties of benign and malignant tumours are needed before microwave diagnosis systems solely based on dielectric properties are viable. As discussed in Section 2.3.1, the standard set of dielectric properties for breast tissues is based on the studies by Lazebnik *et al.* [103], [104]. These studies refer to measurements performed in healthy (i.e., non-cancerous) or cancerous tissue samples (DCIS, LCIS, and infiltrating ductal and lobular carcinoma); tissue samples of benign abnormalities were not analysed in Lazebnik's studies. To date, no standard set of measurements exists that explores the differences between the relative permittivity and conductivity of benign breast abnormalities and *in situ* and invasive abnormalities.

Another possibility to identify malignancies within the breast is the presence of microcalcifications. As shown in Section 2.2.2.3, microcalcifications are correlated with the appearance of malignant findings, particularly DCIS. However, while microcalcifications are visible in mammographic images, they are not well characterised in the microwave frequency range. To date, most clinical studies have not reported the detection of microcalcifications in radar or tomography images; additionally, the dielectric properties of microcalcifications are not known.

Finally, the shape and spiculation of tumours are widely recognised markers for malignancy [88], [135]–[137]. Benign tumours are roughly elliptical and usually have well circumscribed margins, and malignant tumours have irregular shapes and are surrounded by a radiating pattern of spicules [63], [135]–[137]. In Fig. 2.4, examples of tumours detected in mammograms are shown [138]. In Fig. 2.4a, a well-circumscribed, palpable, benign tumour is shown in the lower quadrants of the right breast of an 83 year-old woman. In Fig. 2.4b, a malignant tumour detected in the upper-outer quadrant of



(a) (b)

Figure 2.4: Example of breast tumours detected in mammograms, cranio-caudal view. In (a), a well-circumscribed, palpable, benign tumour is shown in the lower quadrants of the right breast of an 83 year-old woman. In (b), a malignant tumour detected in the upper-outer quadrant of the left breast of a 63 year-old women is shown; this tumour is irregularly shaped and spiculated. From [138].

the left breast of a 63 year-old women is shown; this tumour is irregularly shaped and spiculated.

In principle, shape information could be captured by reconstructed microwave images. However, image reconstruction may involve filtering of the backscattered signals, resulting in a loss of the phase information contained in the acquired backscattered signals. Additionally, images may not be able to recover the shape information of the targets, with the level of resolution required for an accurate differentiation of benign and malignant tumours according to their shape and level of spiculation. Overall, with the current knowledge and state-of-the-art in microwave breast systems, backscattered signals are expected to preserve more information to inform

on the level of malignancy of a tumour.

Previous studies have shown how microwave backscattered signals change when tumours are present within the breast, and also when tumours of different sizes or shapes are encountered [40]–[53]. The same set of studies, though tested with relatively simple datasets, have also demonstrated that classification and machine learning algorithms are able to capture and learn from these shape differences in backscattered signals, without the need to reconstruct images. Machine learning techniques may be directly used with microwave backscattered signals and may provide more detailed information about the tumour, although it is yet to be determined whether the performance of such algorithms is adequate in clinically-complex scenarios.

In the remainder of this section, the use of machine learning with microwave signals is reviewed from two perspectives, first, Section 2.4.1 looks at tumour *detection* (i.e. identifying if a tumour is present), and second, Section 2.4.2 examines *diagnosis* (i.e. determining if a tumour is benign or malignant).

2.4.1 Detection of Breast Cancer

Support Vector Machines (SVMs) were used in [139] to create probabilistic breast images indicative of tumour presence. With this method, the breast is divided into cells and each one is classified for the presence of a tumour; the output is then transformed to *a posteriori* probability of tumour presence. The method was tested with numerical, adipose breast models and cubical tumours of constant size. The results from this study showed promise in localising tumours within breast tissue by means of classification. A similar concept was also implemented later in [140] to create probabilistic maps of the breast. In this study, Linear Discriminant Analysis (LDA) and SVMs were used to classify the breast cells as containing or not containing tumour, and the method was tested on experimental breast phantoms containing adipose and glandular tissue, including tumour phantoms of different sizes and shapes.

In [141], a system combining Principal Component Analysis (PCA) for dimensionality reduction with LDA or SVMs was proposed to classify backscattered signals for the presence of a tumour. The method was tested with numerical breast models, containing adipose tissue and limited inclusions of glandular tissue, and tumour models of several sizes and shapes placed in three different locations within the breast. This study showed promise in detecting tumours by classifying backscattered signals, but detection accuracy was shown to decrease when analysing breasts containing more glandular tissue or smaller tumours. In [142], the same concept and setup

was used to detect differences between successive microwave breast scans arising from tumour growth. By using differential signals, tumour detection performance was increased.

In [143], a similar setup to previous studies was designed to classify backscattered signals for the presence of a tumour. Numerical breast models both with and without glandular tissue content were used, and combined with spherical tumour models of varying size. This study compared the performance of k-Nearest Neighbours (KNN) and SVM, with results confirming that the use of SVM as the classification algorithm leads to higher detection performance.

PCA in combination with LDA or SVMs were again used in [144] to classify backscattered signals collected with an experimental time-domain microwave breast system. The experimental dataset included adipose breast phantoms, and tumours of spherical shape and constant size, placed in two different locations within the breast. This study also showed the potential of classifying backscattered signals to detect tumours, but reported a lower detection rate, which is most likely due to the use of real experimental data, as opposed to numerical models. Interestingly, this study notes that using only a subset of the available channels (i.e., signals arising from selected transmit-receive antenna pairs) might improve detection performance, particularly, the subset of channels in the quadrant closest to the tumour.

In [145]–[147], PCA was combined with ensemble classifiers subject to a cost function that retrieve only the most informative channels to decide if a tumour is present in a scan; here, SVMs are optimised to minimise the false negative rates while simultaneously constraining the rate of false positives. The method was tested on data collected with the same experimental time-domain microwave breast system as [144], including breast phantoms containing adipose and glandular tissue, and spherical tumours of constant size. The method was also tested on a quasi-clinical dataset, where simulated ideal tumour signatures were combined with backscattered signals collected from healthy volunteers. The study found that improved detection performance can be achieved when using only the channels with the most meaningful information.

The detection system proposed in [148] expanded previous work to include multiple feature extraction methods, some based on time-frequency transformations to decompose the original data; the feature extraction methods used include empirical mode decomposition, discrete wavelet transform, dual time complex wavelet transform, discrete cosine transform, and PCA for comparison purposes. Here, the same cost-sensitive SVMs that retrieve only the information of the most informative channels were used, and tested on the same quasi-clinical dataset. In this analysis, the detection performance

using time-frequency features outperformed classification using PCA, which has been one of the most widely used dimensionality reduction methods in microwave tumour classification studies.

In summary, tumour detection studies based on machine learning using microwave systems performed to date indicate that there is likely to be much useful information in the backscattered signals to inform decisions about the presence or absence of a tumour. Some of these studies show that detection performance can be improved by: using differential signals which highlight the tumour signature; extracting time-frequency features of the signals ahead of the classification; and selecting the channels with the most meaningful classification information. However, most previous detection studies have been based on numerical data, or experimental breast phantoms with limited inclusions of glandular tissue and limited variety of tumour sizes and shapes.

2.4.2 Diagnosis of Breast Cancer

In [42]–[47], several feature extraction methods (PCA, discrete wavelet transforms, and independent component analysis) were used in combination with different classifiers (LDA, Quadratic Component Analysis (QDA) and SVM) to diagnose breast tumours with microwave backscattered signals. The analysis was based on numerical breast models containing adipose tissue and limited inclusions of glandular tissue. Tumours were modelled with several sizes and shapes to represent benign and malignant tumours, and were located in the centre of the breast or in slightly off-centred positions. These studies showed promise in using backscattered signals to diagnose tumours, and suggested that classifying tumour size ahead of tumour shape may improve diagnostic performance.

The suitability of neural networks to classify backscattered signals was also assessed. A combination of genetic algorithms and neural networks with discrete wavelet transforms was proposed in [48], [49], and tested on a similar numerical dataset to the studies above. As before, diagnostic performance was improved by separating tumours based on their size before classifying them as benign or malignant, and by investigating which channels provide the most useful information.

The same numerical dataset was also used in [50] to investigate the potential of self-organising maps to track the development of a tumour from a benign state to different levels of malignancy. This study showed potential in distinguishing between different shapes of tumours.

The effect of signal pre-processing on the diagnostic performance when dealing with more complex breast models has also been investigated [53].

In this study, MRI-derived clinically-realistic breast models [149] were used, and tumour models of several sizes and shapes were included and located in various positions within the breast. The hypothesis behind this study was that diagnostic performance might improve if the backscattered signal is windowed to contain only the tumour signature. The classification framework relied on PCA in combination with SVMs. Results from this study showed that the windowing methodology helped improve diagnostic performance when examining more complex breast models.

Experimental datasets have also been used to assess the performance of diagnosis systems, by using PCA in combination with SVMs, LDA and QDA. In [51], [52], tumour phantoms with various sizes and shapes were immersed in a breast phantom with dielectric properties matching those of adipose tissue. Importantly, the experimental results presented in these studies are in general agreement with previous numerical data.

The breast tumour diagnosis studies summarised in this section indicate that the shape of a breast tumour influences its signature within a backscattered signal, potentially providing an approach to allow machine learning models to learn how to distinguish between benign and malignant tumours. These studies have also looked at the effect of intelligently using the most informative channels. In addition, these studies have concluded that it is beneficial to separate tumours according to size before final diagnosis, and also, that further signal pre-processing methodologies should be explored when dealing with more complex breast models, for example, breast models with increased content of glandular tissue.

2.5 Conclusions

The topic of screening mammography for breast cancer in asymptomatic populations is timely and of much current debate. Many different studies have been published on the issue, with the purpose of determining whether screening mammography offers an acceptable compromise between overall reduction in breast cancer mortality, and increase in the rates of false positives and overdiagnosed cancers, with studies now suggesting that the long-term benefit of screening with mammography might be minimal. In this context, and considering that early detection is widely accepted as the cornerstone of breast cancer care, a need exists for new modalities that are capable of overcoming the limitations of mammography.

Microwave imaging for breast cancer detection has shown great promise, and in fact, preliminary clinical studies have shown increased sensitivity of microwave imaging systems in detecting breast cancer in dense breasts.

However, furthering the field of microwave technology for breast cancer now means understanding if the technology also has the potential to accurately differentiate between true positive breast tumours (i.e., breast cancer) and true negative findings (for example, cysts). Recently, small scale studies have demonstrated the potential of automated software using machine learning algorithms to classify backscattered signals and diagnose breast cancer.

In this chapter, the breast anatomy was reviewed, including a description of breast and malignant abnormal breast findings. This was followed by a summary of the guidelines for the clinical assessment of breast cancer, whether in symptomatic or asymptomatic populations. The core principles of microwave systems were described next from a number of different perspectives. Finally, the literature on the topic of microwave breast diagnosis using machine learning was reviewed.

The work described in this thesis attempts to address some of the limitations in current practice. The first objective of this research is the development of clinically-realistic breast and tumour phantoms for use in system development and evaluation. The next chapter describes the development of numerical and experimental tumour models in detail.

CHAPTER 3

Development of Numerical and Experimental Phantoms

3.1 Introduction

Prior to clinical translation and acceptance, any new technology designed to investigate the human body must be thoroughly assessed. Evaluation of medical systems often relies on the use of phantoms, which are fabricated to mimic the properties of interest of relevant body structures and should respond in a similar manner to how human tissues will interact with the system.

For breast cancer detection and diagnosis systems using microwave technology, phantoms should include an accurate representation of the breast and abnormal breast findings. Suitable phantoms should include:

1) Breast models of several shapes and sizes;
2) Varying distributions of the different tissues within the breast model;
3) Tumour models representing different breast abnormalities (benign and malignant);
4) Tumour models of different sizes and shapes;
5) Accurate representation of the locations of tumours within the breast;
6) Accurate dielectric properties of all breast and tumour tissues.

Calcifications are not usually included in phantoms designed to test microwave breast systems, as their response is not well characterised in the microwave frequency range. Calcifications will not be considered in the remainder of this work.

In the research described in this thesis, two sets of phantoms were designed that meet the above criteria for designing clinically meaningful breast and tumour models: firstly, a numerical tumour model set was developed, and secondly, an experimental breast phantom set that includes physical implementations of the numerical tumour models from the numerical set.

The numerical phantom set consists of tumour models derived from clinical information, where a clinical procedure was employed to validate the applicability of the models. The numerical tumour phantom set is described in detail in Section 3.2, and resulted in the original research article titled "Development of Clinically Informed 3-D Tumor Models for Microwave Imaging Applications", published in *IEEE Antennas and Wireless Propagation Letters* (2015).

An experimental evaluation phantom set was also developed — the Breast Imaging and Diagnosis (BRIGID) phantom set, which is described in detail in Section 3.3. The BRIGID phantom set consists of breast phantoms with varying compositions of internal breast tissue, and tumour phantoms that differ in their shape, size, and level of spiculation. The phantom set was fabricated with a modular design, allowing for 156 combinations between breast and tumour models, to represent a variety of breast cancer scenarios. The BRIGID tumour phantoms were developed based on the numerical tumour models described in Section 3.2, which were verified by clinicians, thus confirming the applicability of the experimental phantoms to the microwave breast systems testing. The BRIGID phantom set was originally presented in the journal publication titled "Microwave Breast Imaging: Experimental Tumour Phantoms for the Evaluation of New Breast Cancer Diagnosis Systems" published in *Biomedical Physics and Engineering Express* (2018).

3.2 Breast Tumour Model Development: Numerical Approach

In this section, the numerical tumour models designed to test microwave breast cancer detection and diagnosis systems are described. A brief overview of other works in the field is first presented in Section 3.2.1, followed by a description of the mathematical model used to produce the clinically-informed tumour phantoms in Section 3.2.2. Finally, the procedure followed to validate the proposed models is described in Section 3.2.3.

3.2.1 Overview of the Field

For the numerical testing of microwave systems for breast cancer detection and diagnosis, three main elements are required: an electromagnetic simulation platform, breast models, and tumour models. The electromagnetic simulation software should accurately model the propagation of electromagnetic signals in the biological tissues. Additionally, the breast and tumour models should be anatomically-accurate and representative of clinical cases.

The Finite-Difference Time-Domain (FDTD) numerical method is an electromagnetic simulation approach that has been widely used in microwave breast imaging studies [150]. The frequency dependent nature of biological tissue properties can be incorporated into the FDTD numerical model of the breast through a Debye formulation, allowing for accurate modelling of the scattering of electromagnetic signals.

In the context of breast modelling, earlier studies relied on describing the breast with hemispherical or conical shapes, usually including simplified representations of the distributions of breast tissues. However, more recent studies now often make use of the numerical breast phantom repository developed by the University of Wisconsin Computational Electromagnetics Laboratory (UWCEM) [149], [151].

The UWCEM repository contains volumetric anatomically-realistic breast models derived from MRI images of patients, with a variety of shapes and sizes and distributions of fat and glandular tissues within the breast. The breast models are also mapped to the dielectric properties of normal and malignant breast tissues as measured by Lazebnik *et al.* [103], [104], and can be directly used with the FDTD numerical simulation method.

In the context of tumour modelling, three main methods are used in numerical studies to represent abnormal findings of the breast:

- Simplified approaches, by representing tumours as spheres, e.g. as with the numerical analyses presented in [152], [153];
- Gaussian Random Spheres (GRS) [41], [43];
- Polygonal approximations [40], [61], [136], [154].

GRS are stochastic shape models that can be used to create irregularly shaped objects. The two governing properties of the method are the mean and the covariance function of the radius of the GRS; by altering these parameters, tumour models with several sizes, shapes and margin types can be created. The application of the GRS method to produce breast tumour models was initially described in [41], which resulted in the creation of models with smooth margins to mimic benign tumours, and microlobulated

and spiculated models to represent malignant tumours. The GRS method was also used in [43], where a database of breast tumours of four types was developed: smooth and macrolobulated models to represent the properties of benign tumours, and spiculated and microlobulated to represent malignant tumours. Despite representing accurate anatomical information, to date, tumour models created by the GRS method have not been validated in a clinical environment.

Another approach proposed in the literature consists of describing breast tumours as polygonal approximations. In a study analysing tumours found in mammograms [136], it was observed that the tumour boundaries hand-drawn by radiologists on the mammograms are well approximated by irregular polygons. This initial concept was later further explored in [40], resulting in the proposal of an algorithm to model breast tumours; with this algorithm, tumours are modelled as ellipses, and an irregularity is created at the margin to mimic different tumour shapes. The algorithm proposed in [40] by Chen *et al.* allows the creation of models with irregular shapes and margin types, to represent malignant tumours, and models with regular shapes and smooth margins, to represent benign tumours. Despite being derived from mammographic information, the tumour models proposed with this method were two-dimensional; additionally, the applicability of the models in a clinical setting was not validated.

Finally, the location of the tumours within the breast tissue is also a concern in designing clinically-accurate breast and tumour models to use with microwave systems. Once more, the tendency in many numerical studies has been to simplify the models by placing tumours in the centre of the breast, or in slightly off-centred positions. Accurate breast models should include tumours appropriately localised within the breast tissue as per the findings described in Table 2.1.

In summary, while tumour models have been proposed that mimic the smooth or spiculated nature of benign and malignant tumours respectively, little information has been provided as to their clinical suitability. The tumour models existing in the literature have generally not undergone a validation procedure based on clinical practice, potentially limiting their applicability for clinically-realistic microwave breast studies. Given the limitations with current approaches to tumour modelling, in the following section a new tumour modelling method is presented.

3.2.2 Mathematical Modelling

The tumour modelling method proposed in this research extends previous work by Chen *et al.* from 2008, which generated accurate tumour models by

means of polygonal approximation [40]. In turn, Chen's method relies on a previous study, which had shown that tumours delineated in mammograms are well matched to polygonal approximations [136].

The proposed algorithm is based on the principle that the shape of a tumour resembles an ellipsoid:

$$\frac{d^2cos^2(\theta)sin^2(\varphi)}{a^2} + \frac{d^2sin^2(\theta)sin^2(\varphi)}{b^2} + \frac{d^2cos^2(\varphi)}{c^2} = 1 \qquad (3.1)$$

where (d, θ,φ) correspond to the spherical coordinates that describe the ellipsoid; here, the distance of each vertex to the centre of the ellipsoid, d, is a function of the two angles (θ,φ). The ellipsoidal behaviour is controlled by pre-specifying the lengths of each semi-axis, (a,b,c).

The ellipsoid is modelled as a polygon that consists of a series of triangular faces interlinked together (each face thus containing three vertices). There is a 1:1 relationship between the number of faces of the polygon and the number of desired spicules of the resultant tumour model. It should be noted that the number of pre-specified triangular faces also influences the width of each spicule. Then, an extended version of Chen's criterion of distortion is applied as described below. For each vertex of the polygon, $d(\theta,\varphi)$ is altered according to the value of s, which is a parameter controlling the degree of spiculation at the surface of the tumour:

$$d'(\theta, \varphi) = n.[d(\theta, \varphi).(1 + \mu(\theta, \varphi))] \qquad (3.2)$$

where $\mu \in U[-s, +s]$, d' is the new distance to the centre after the above modification is applied, and U denotes the uniform distribution from which s is randomly chosen. The degree of spiculation satisfies $0 \leq s \leq 1$, such that $s = 0$ produces a perfectly smooth border and $s = 1$ produces the maximum degree of spiculation. A novel element of this method is the parameter n, which is defined as the proportion of the tumour's surface area to be covered with spicules; here, $0 \leq n \leq 1$. The algorithm can be selectively applied to different portions of the tumour, which further allows for greater flexibility in generating different models with specific characteristics.

Figure 3.1 shows examples of tumour models generated with the proposed method, where each model is shown in the Head-to-Toe (H-T), Left-to-Right (L-R) and Back-to-Front (B-F) axes; the examples shown in this figure highlight the versatility of the proposed method in generating several types of tumour models. The main features of the tumour models are listed below:

1) Even distribution of spicules on the surface of the tumour model, as shown in Fig. 3.1a, Fig. 3.1c, Fig. 3.1e;

2) Spicules limited to one portion of the surface of the tumour, as shown in Fig. 3.1b,
3) Spicules located at non-adjacent portions of the surface of the tumour, as shown in Fig. 3.1d;
4) A variety of degrees of spiculation at different locations on the tumour surface, as represented in Fig. 3.1f.

This procedure is applied to the tumour volume as a whole (as opposed to stacks of slices of the polygon), in an attempt to mimic the natural growth of a tumour within the breast. The algorithm outputs a logical array wherein each voxel lies inside or outside the tumour, thus facilitating ease of integration into a full numerical breast model.

The following subsection discusses the methodology implemented to validate clinical applicability of the proposed tumour models.

3.2.3 Clinical Validation

One significant novelty of this work is the clinical validation of the tumour models generated with the proposed method. This validation study involved two independent approaches. Firstly, tumours found in MRIs of the breast were reproduced with the proposed algorithm, and the Structural Similarity (SSIM) metric was used to evaluate the result [155], by estimating the degree to which structural elements in the tumour are preserved. Secondly, breast cancer radiologists assessed the accuracy of the tumour models generated with the proposed method, while also labelling the models according to their level of malignancy.

Breast MRIs were obtained through The Cancer Imaging Archive [156] and are a part of the Reference Image Database to Evaluate Therapy Response (RIDER). This database has the benefit of containing 5 breast MRIs where tumours have been previously identified and segmented, thus facilitating their use for the purpose of this study.

For each of the 5 tumour models extracted from RIDER (these models are referred to as "original"), the degree of spiculation at each point on the surface of the tumour was computed, by solving Eq. (3.2) for the degree of spiculation s:

$$s(\theta, \varphi) = \frac{d'(\theta, \varphi)}{d(\theta, \varphi)} - 1 \tag{3.3}$$

Then, the proposed modelling algorithm was used to replicate the tumours extracted from RIDER. A slice-by-slice Structural Similarity (SSIM) analysis was performed between all original and replicated models; here, the

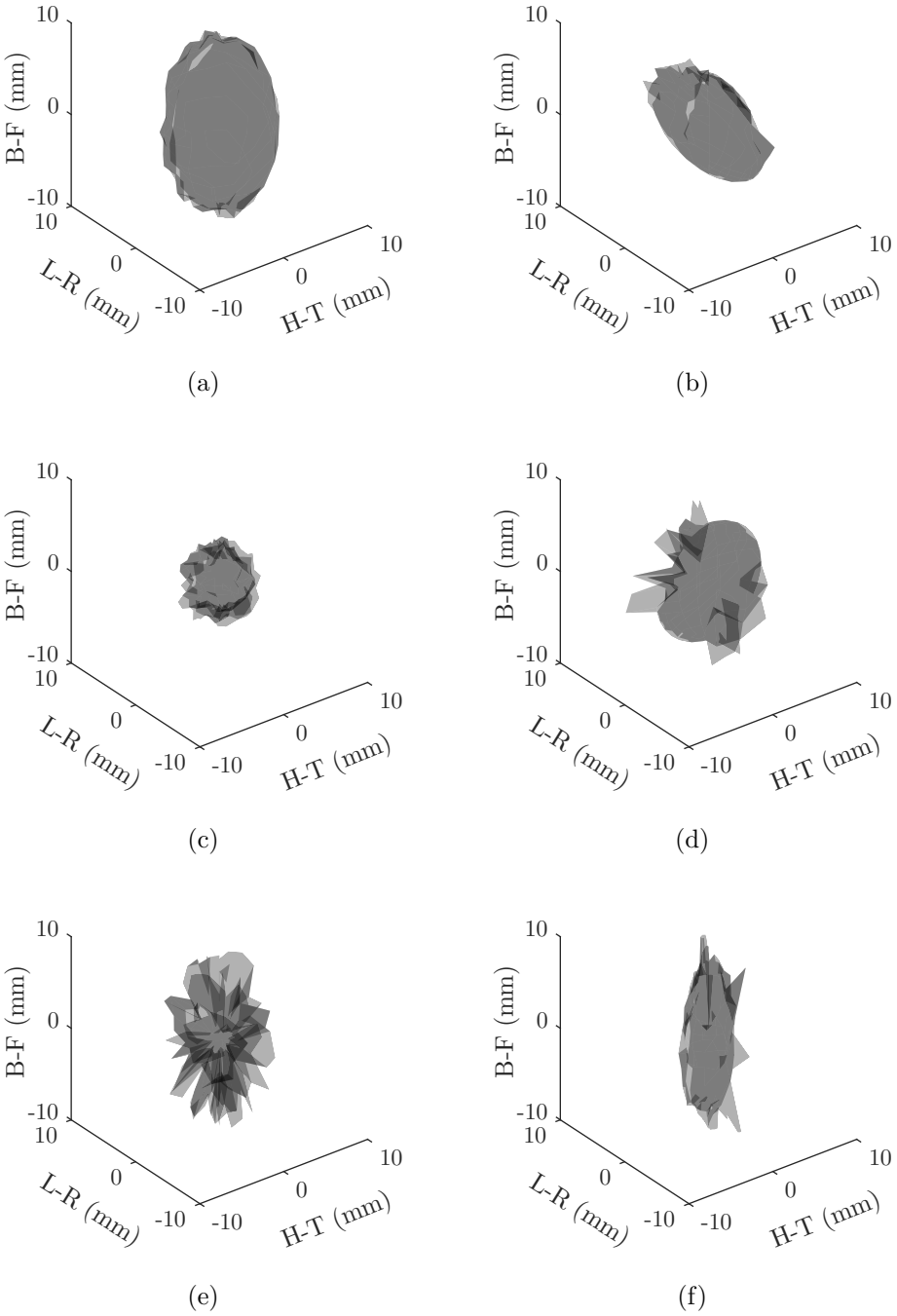


Figure 3.1: Tumour models generated with the proposed algorithm for varying sizes, shapes and degrees of spiculation (s). Mean radii for the models vary between 3 mm and 10 mm. Degrees of spiculation: (a) $s = 0.08$, (b) $s = 0.2$, (c) $s = 0.3$, (d) $s = 0.6$ (e) $s = 0.8$, (f) $s = 0.2$ and $s = 1$. Each tumour model is shown in the Head-to-Toe (H-T), Left-to-Right (L-R) and Back-to-Front (B-F) axes.

average original/replicated SSIM index obtained for all slices was 0.77. The slice-by-slice SSIM indices were also computed between all original models and the baseline ellipsoids used to initialise the algorithm in each case. This time, the average original/ellipsoid SSIM index obtained for all slices was 0.47. The improvement in the original/ellipsoid (0.47) to original/replicated (0.77) SSIM indices suggests that, firstly, simplistic ellipsoids are poor representations of breast tumours found in clinical practice, and secondly, that the proposed modelling algorithm can accurately generate tumours found in clinical practice.

As an example, Fig. 3.2a and Fig. 3.2b show a 2D coronal representation of one of the original tumours and the replicated 3D model, respectively, where the brightness of each pixel represents the quantity of tissue in that particular vertical region of the model. The baseline ellipsoid is displayed in Fig. 3.2c, and the results of the slice-by-slice SSIM analysis are displayed in Fig. 3.2d.

In addition, 3 independent radiologists examined 35 different models produced by the proposed method. The radiologists were shown images such as the ones in Fig. 3.2; this representation was chosen since it attempts to recreate the general properties of mammograms, which are the most frequent breast images seen in clinical practice and therefore of a form quite familiar to radiologists. The radiologists were asked to evaluate the shape and margin of the models and label them as benign (BI-RADS 1 and 2), suspicious (BI-RADS 3 and 4) or malignant (BI-RADS 5); an extra class labelled "other" was included to ensure the radiologists had the option to indicate that the tumour models in this study were not representative of clinical findings. The results of this clinical validation study are described below:

1) For 29% (10) of the tumour models, the decision between the three radiologists was unanimous. For 63% (22) models, two out of three radiologists agreed on the label. Finally, there was no consensus between the radiologists' assessment for the remaining 8% (3) models;
2) Based on a majority vote, a total of 6 models were labelled as benign, 16 models were labelled as suspicious, and 10 models were deemed malignant.

It should be noted that none of the radiologists labelled any of the models as "other", confirming the models are clinically representative. However, labelling the level of malignancy of the models proved to be more difficult. This result is due to the fact that patient outcome, i.e. level of malignancy of a tumour, is not typically assessed by clinicians from one single feature

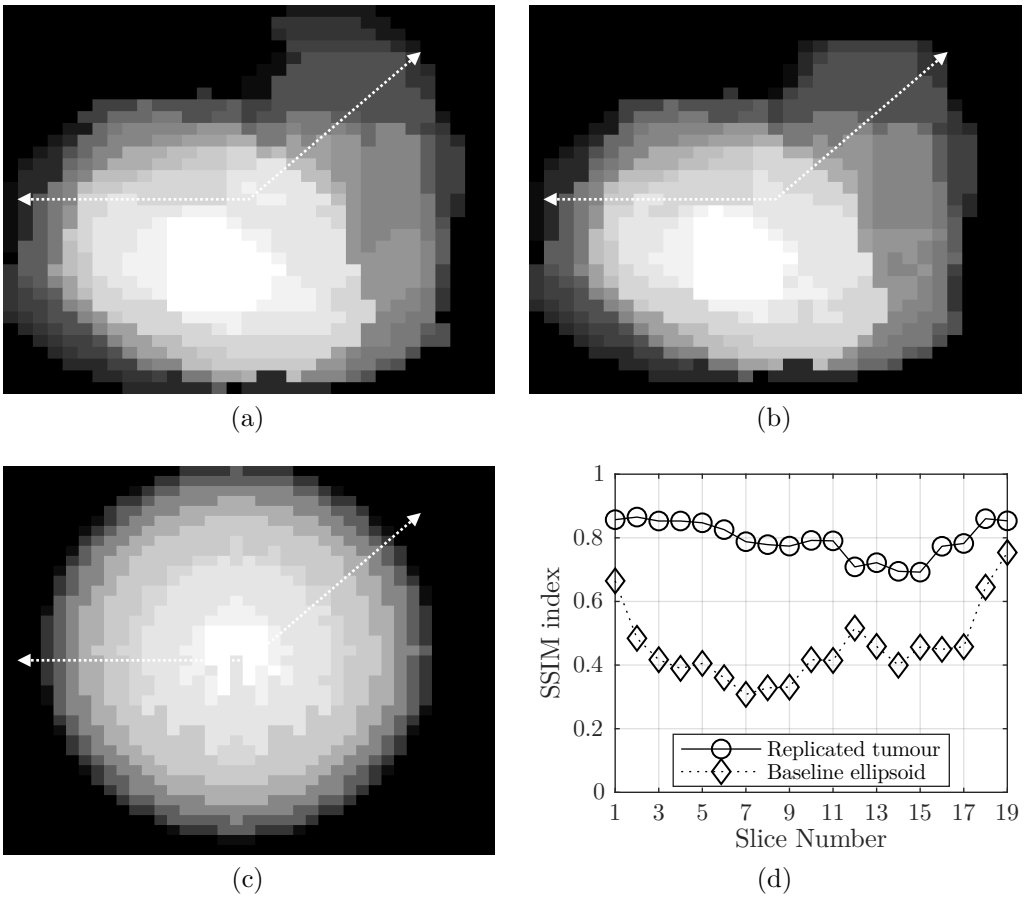


Figure 3.2: Coronal representations of numerical tumour models, where the brightness of each pixel represents the quantity of tissue in that particular vertical region of the model. (a) Tumour extracted from the Reference Image Database to Evaluate Therapy Response (RIDER) database. (b) Tumour model replicated with the proposed method. (c) Baseline ellipsoid used to initialise the modelling algorithm. The white dashed arrows indicate areas of similarity between images. The images are presented as generated, with no post-processing to improve quality. In (d) the slice-by-slice Structural Similarity (SSIM) comparison for this tumour is shown.

identified on a breast image, but rather from a large combination of features, including personal history of the patient and morphological assessment of the images.

Further discussion also revealed that any spiculation at the margin of a tumour is usually considered as a reason for suspicion by clinicians; patients who present with breast images revealing any suspicious features are usually forwarded for biopsy. These results are significant since they suggest a clear distinction between establishing the need for tissue biopsy and patient

outcome, thus providing great scope for further investigation on classification schemes for breast cancer.

3.3 BRIGID Set: Experimental Breast and Tumour Phantoms

In this section, the BRIGID experimental phantom set is introduced, which includes breast and tumour phantoms suitable for testing microwave breast cancer detection and diagnosis systems. The BRIGID tumour phantoms were based on the numerical tumour models described in Section 3.2, which were subjected to a clinical validation process. The experimental phantoms are thus expected to be clinically-realistic and applicable for the testing of microwave breast systems.

An overview of the experimental phantoms currently used in the literature is first presented in Section 3.3.1. The fabrication protocol is then presented in Section 3.3.2, followed by a description of the dielectric properties measurement method in Section 3.3.3. Section 3.3.4 describes how the different phantom tissues were fabricated and characterised, and finally, the BRIGID platform of breast and tumour phantoms is described in Section 3.3.5.

3.3.1 Overview of the Field

In addition to being anatomically and clinically-representative, experimental phantoms should be repeatably and cheaply fabricated, dielectrically and mechanically stable over time, and flexibly but reliably reconfigured for a variety of test scenarios. With this in mind, several phantoms have been recently proposed to represent the shapes and dielectric properties of breast tissues. The available phantoms mostly differ in the type of Tissue-Mimicking Material (TMM) used to model the breast and tumour tissues, as well as the shape chosen to represent the breast.

Oil-in-gelatin mixtures have been used in [157]–[159] to produce breast phantoms with varied shapes and interiors. Target dielectric properties can be controlled by mixing different percentages of oil in the gelatine matrix. However, these phantoms are very sensitive to environmental exposure and have a limited shelf-life, with [158] reporting a significant change in their dielectric properties over a six month period.

Triton X-100-based liquid mixtures have also been proposed as TMMs [160], [161], where the target dielectric properties of breast tissues can be reproduced by varying the proportions of the Triton X-100 in water or saline.

Triton X-100-based phantoms are dielectrically stable over temperature and longer periods of time when compared to the oil-in-gelatin mixtures.

Liquid-based phantoms require the use of external and internal shells to separate the target breast tissues. For example, in [162], [163] 3D-printed, MRI-derived skin and glandular plastic shapes are used in combination with liquid Triton X-100 mixtures to produce anatomically-realistic phantoms. However, one recent study has determined that the plastic shell may have a substantial effect on the measurements, ultimately impacting the fidelity of the backscattered signals and the images produced [164].

Polyurethane has also been proposed as a TMM, where the dielectric properties are controlled by adding carbon black and graphite powders to the rubber mixture [165]. This material was used in the development of modular phantoms in [166]. Here, the interior of the phantom is solid but a hole is left for inclusion of a target. With this modular design, targets can be reliably and repeatably inserted in the same position. Furthermore, the long-term stability of this phantom material has been previously assessed in [166] and other imaging modalities have been used to characterise the consistency of the material [167].

More recently, early-stage studies have proposed new TMMs to be used in the construction of multi-modality phantoms. In [168], new TMMs are described that are able to match the dielectric and acoustic properties of fat and glandular tissues, skin, muscle and tumour tissue, with the aim of tailoring suitable synthetic materials for microwave and ultrasound applications simultaneously.

Besides the choice of TMM, the experimental phantoms described above also differ in the dielectric properties obtained for each of the breast tissues. In Table 3.1 and Table 3.2, a summary is presented of the relative permittivity and conductivity, respectively, of fat, skin, glandular and tumour tissues obtained in each of the experimental phantoms available in the literature. In both tables, the median dielectric properties measured by Lazebnik *et al.* [103], [104] are also included for comparison. Both tables indicate that the absolute dielectric properties of breast tissues vary across experimental phantoms; however, in relative terms, all breast phantoms show increasing relative permittivity from fat through skin, glandular and tumour tissues.

Despite the advances in producing TMMs with reasonable dielectric properties and anatomically-derived skin and glandular shapes most studies still only model tumours using basic shapes such as spheres or ellipsoids, similarly to the numerical case addressed in Section 3.2. As discussed in Section 2.4, the spiculation at the border of a tumour is widely accepted as an indicator of its level of malignancy, and this has often not been included in

Table 3.1: Summary of the range of relative permittivity (ϵ_r) for common experimental phantoms for microwave testing. The median relative permittivity measured by Lazebnik *et al.* [103], [104] is also included for comparison. Four breast tissues are shown: fat, skin, glandular and tumour, at a frequency of 3 GHz.

	Fat	Skin	Glandular	Tumour
	ϵ_r	ϵ_r	ϵ_r	ϵ_r
Lazebnik *et al.* [103], [104]	5	38	47	67
Triton-X-100 [163]	5	—	35–46	52
Oil-in-Gelatin [158]	13	36	32	54
Oil-in-Gelatin [159]	7	38	38–48	—
Rubber-solid [165]	5	24	36	>36
Rubber-solid [166]	6–10	35	36–39	55–65

Table 3.2: Summary of the range of conductivity (σ) for common experimental phantoms for microwave testing. The median conductivity measured by Lazebnik *et al.* [103], [104] is also included for comparison. Four breast tissues are shown: fat, skin, glandular and tumour, at a frequency of 3 GHz.

	Fat	Skin	Glandular	Tumour
	σ (S/m)	σ (S/m)	σ (S/m)	σ (S/m)
Lazebnik *et al.* [103], [104]	0.1	1.8	2.1	3.1
Triton-X-100 [163]	0.1	—	1.4–2.0	2.1
Oil-in-Gelatin [158]	0.6	2	1.4	2.5
Oil-in-Gelatin [159]	0.2	1.6	1.6–2.0	—
Rubber-solid [165]	0.2	1.6	3.0	>3.0
Rubber-solid [166]	0.1–0.5	3.0	3.0–4.0	4.0–6.4

the development of experimental phantoms for microwave systems. Accurate tumour modelling should address the issue of shape and degree of spiculation and is essential for the thorough evaluation of the potential of microwave systems not only as a detection method, but also as a way to diagnose the early occurrences of breast cancer.

To address some of the gaps identified in the literature in the fabrication of experimental phantoms, new breast and tumour models were designed based on a mixture of polyurethane, carbon-black and graphite; the suitability of this TMM to the microwave problem has already been established in previous studies, for example, [165]–[167]. The polyurethane mixture

provides flexibility in creating a multitude of breast tissues and tumour shapes, producing phantoms that are robust and easy to use. In this work, the dielectric properties measured by Lazebnik *et al.* [103], [104] are used as a target for the phantoms' properties. As discussed in Section 2.3.1, these properties represent a worst-case scenario, ultimately leading to the creation of more challenging test cases.

In the following subsections, the fabrication process of the new breast and tumour models is described, including the methodology applied to validate the dielectric properties of the produced tissues.

3.3.2 Fabrication Protocol

The polyurethane, carbon-black and graphite mixture described in [165], [166] was chosen as a TMM in this study as it provides great flexibility both in recreating the target dielectric properties of breast tissues, and in producing tumour phantoms with different shapes and levels of spiculation.

A polyurethane base (VytaFlex 20, Smooth-On, Easton, PA, USA) was first prepared by mixing equal masses of the two liquid precursors to polyurethane as per the manufacturer's instructions. Immediately after, graphite powder (general purpose powder from Fischer Scientific, Loughborough, Leicestershire, United Kingdom) and carbon-black powder (acetylene 50% compressed, 99.9%+ from Alfa-Aesar, Ward Hill, MA, USA) were slowly incorporated and blended with the polyurethane base. Although the curing process takes 12–16 hours, it is important to incorporate the powders when the mixture is most malleable.

The greater the mass percentage of powder, the higher the values of the dielectric properties of the sample. However, the blended mixture becomes increasingly difficult to mix as the mass percentage of powder increases. Small amounts of acetone were also added to the mixture as a thinning agent, firstly to ensure each sample was evenly mixed, and secondly to allow for samples with higher relative permittivity to be created. As previously discussed in [166], it should be noted that acetone has an effect on the dielectric properties, especially the conductivity, which means mixture quantities had to be adjusted accordingly to achieve the desired dielectric property values.

Cuboidal reference samples were prepared and used to determine the optimal dielectric properties of each tissue type before preparing the final phantom set. The resulting target mass percentages obtained for each tissue type are shown in Table 3.3. The following subsection details the methodology employed to measure the dielectric properties of tissues, and the validation procedure.

Table 3.3: Summary of the percentages of the constituent parts by mass used in the fabrication of each of the tissues layers.

	Fat	Skin	Glandular	Tumour
Polyurethane part-A (%)	40.0	32.0	27.5	25.75
Polyurethane part-B (%)	40.0	32.0	27.5	25.75
Graphite (%)	20.0	30.0	41.0	42.0
Carbon-black (%)	0	6.0	4.0	6.5
Acetone (mL/100 g)	0	3.0	4.0	6.0

3.3.3 Dielectric Properties Measurement Method

All dielectric measurements were performed with the Keysight E5063A ENA Series Network Analyser (Wokingham, Berkshire, United Kingdom) using the performance probe, where calibration was completed using the standard three-load procedure. The relative permittivity and conductivity were extracted from the reflection coefficients automatically using the Keysight Materials Measurement Suite software, as used in, for example, [169], [170]. Measurements were recorded at a total of 100 linearly spaced frequency points between 1 and 8.5 GHz. The chosen frequency resolution is widely used in dielectric properties measurement studies [109], [112], [113], [170]–[172].

The dielectric measurement system was validated post-calibration through repeated measurements on a standard liquid, 0.1 M saline solution. Measurements were conducted at 21.2 °C, and compared to the known model of saline from the literature [173]. In this study, accuracy is defined by the average percentage variation between the known model and the mean of the measured data, over all frequency points. For the relative permittivity, the accuracy of the measurements relative to the literature model was found to be within 1.0%, while for the conductivity, the accuracy was within 3.7%.

To ensure a good contact between sample and probe, all cuboidal reference samples were first sliced. Visual inspection of the interior of each sample was performed to assess its consistency. The internal cross-section of the sample was then sanded smooth.

To assess the repeatability of the probe-sample contact, 10 measurements were performed, at a fixed location, on a sample of each of the breast layers modelled in this study: fat, skin, glandular and tumour tissues. Here, repeatability was assessed as the maximum percentage variation within measured data, averaged over all frequency points. In terms of relative permittivity, the percentage variation of the measurements was found to

be 1.0% for fat, 3.0% for skin, 3.5% for glandular tissue, and 1.9% for tumour tissues. Regarding the conductivity, the percentage variation of the measurements was determined to be 7.9% for fat, 4.7% for skin, 5.6% for glandular tissue, and 5.8% for tumour tissues.

Based on the validation of the measurement setup and the probe-sample contact repeatability test, the reference samples were characterised. The dielectric properties were measured at a minimum of 20 scattered points within the sample. This serves as an indication of the consistency of each tissue sample. Fig. 3.3 displays the dielectric properties for each of the breast layers modelled in this study. The curves are plotted such that each coloured region in the graph represents the raw measured data from all 20 locations in each reference sample. The median dielectric properties of each tissue type as measured by Lazebnik *et al.* [103], [104] are also displayed by dashed lines.

Considering any remaining systematic dielectric measurement error, remaining variability in the quality of the probe-sample contact, and the inherent heterogeneity of breast tissues, it is found that the phantom materials used in this study display a reasonable match to the target dielectric properties in many respects. Particularly in terms of relative permittivity, Fig. 3.3a shows that the contrast achieved between the different tissue types is similar to the target ones. For example, the glandular-to-tumour contrast ratio in this study is 1:1.5, while for the target properties the ratio is 1:1.4.

Concerning the skin, Fig. 3.3a shows that the experimental relative permittivity is lower than the target one. However, previous studies have already demonstrated that, even despite the lower properties, skin tissue produced with this method results in backscattered reflections similar to human tissue [165].

As shown in Fig. 3.3b, the conductivity measured for the tumour tissue produced in this study is higher than the target. This is a result of the use of acetone during the fabrication of the tissues. However, conductivity mainly influences losses in the medium, and due to the typical size of tumours, the higher loss is not expected to adversely impact the backscattered energy.

The following subsection describes the development of each one of the breast tissues (skin, glandular, tumour and fat) using the fabrication protocol from Section 3.3.2, as well as modular phantom assembly.

3.3.4 Breast Phantom Development

Many of the current microwave breast analysis systems in commercial development are conformal in their design, which means that the breast is in direct contact with the surface of the prototype. As a result, the breast

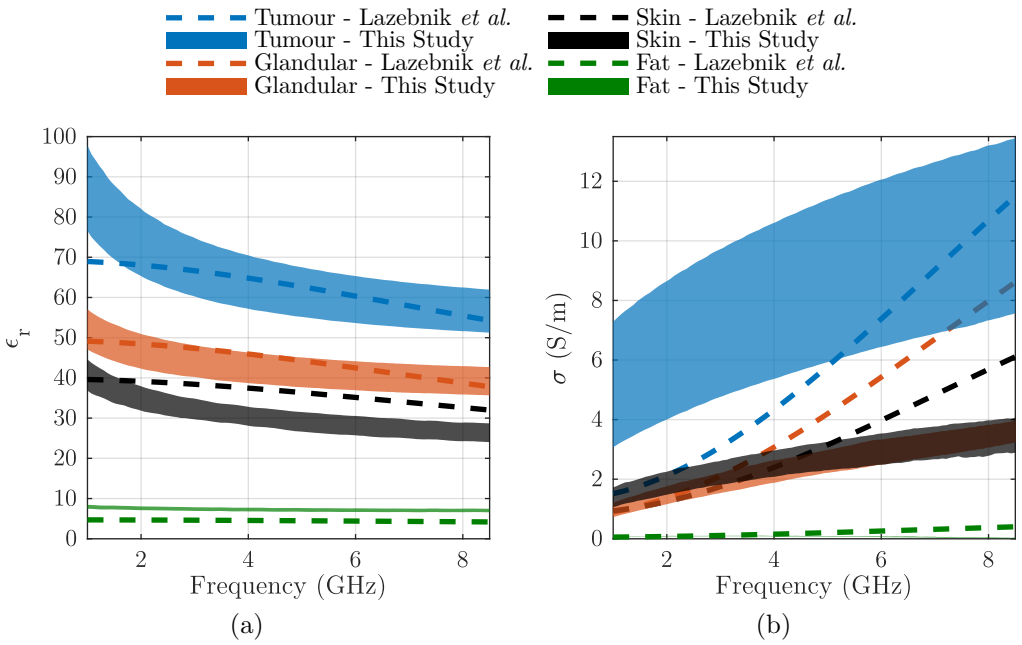


Figure 3.3: Dielectric Properties of Fat (green), Skin (black), Glandular (orange) and Tumour (blue). (a) Measured relative permittivity (ϵ_r); (b) Conductivity (σ). Each of the coloured regions represents the raw data measured in this study of each reference sample, for all locations. Also shown are the median dielectric properties measured by Lazebnik *et al.* [103], [104], by the dashed lines.

conforms to the shape of the prototype during a scan. With this type of design, the system can be used without an immersion medium, which is advantageous in a clinical setting as there is no immersion medium to be changed between patients. At the same time, the experimental evaluation of conformal systems requires phantoms that mimic the same type of breast deformation. Although anatomically-realistic, MRI-derived skin shapes do not mimic the same type of breast deformation, rendering them unsuitable for the evaluation of conformal microwave breast systems. For example, the systems presented in [27], [105] make use of a hemispherical shell to which the breast is fitted during the scan.

In this study, a 7 cm hemispherical shell was designed and this was used to inform mould design for the phantom production. In addition, the breast was modelled using four tissue types, firstly an outer skin layer, internal glandular structures, tumours and finally the fatty background tissue, with dielectric properties as described in Fig. 3.3.

To assist in the phantom making process, a variety of moulds representing different breast shapes were 3D-printed using the Ultimaker 2+ Extended

(Ultimaker B.V., Geldermalsen, The Netherlands) in PolyLactic Acid (PLA) filament. Prior to pouring the mixtures described in Section 3.3.2 in the moulds, each mould was coated in non-silicon mould release (Formula 7, Ambersil, Bridgwater, England), to allow for each mould and phantom to be easily separated post-curing. The use of 3D-printed moulds allowed for greater flexibility while creating new phantoms, ultimately contributing to the diversity of the breast phantom set. Importantly, all recipes and processes developed in this study can be accurately reproduced with any other 3D-printed moulds. The following subsections provide a detailed protocol of phantom fabrication.

3.3.4.1 Skin Layer

The thickness of human skin around the breast is known to vary between 0.5 mm and 3 mm, depending on many factors such as, the quadrant of the breast, breast size and hormonal changes [174]–[176]. With this in mind, the skin was modelled as a layer of material with average thickness 2 mm.

A 3D-printed mould and counter-mould were used to make the skin layer. The mould was first evenly coated in a layer of the polyurethane mixture and pressure applied to the counter-mould. The mould and counter-mould for the skin layer can be seen in Fig. 3.4; the top view of one of the resulting skin layers is also shown in the same figure.

For each skin layer, the mass was measured, the interior volume recorded, and the density calculated. The mean density for all skins produced in this study is 0.98 g/mL, with an associated standard deviation of 0.09. The low standard deviation is a positive indicator regarding the consistency and reproducibility of the phantom making process.

The thickness of each skin layer was also measured at 24 points evenly distributed along the skin surface. As detailed in Table 3.4, the mean skin thickness ranges between 1.98 mm and 2.99 mm, which is consistent with the reports of human skin thickness variation. In addition, knowledge of the skin thickness as a function of location in laboratory phantoms allows for imaging algorithms to be assessed with regards to any variation that may exist in the skin surface, and for the efficacy of skin artefact removal algorithms to be evaluated.

The dielectric properties of each skin layer were also measured at the same locations as the thickness. As demonstrated in Table 3.4, the low standard deviation in the dielectric properties is evidence of the consistency and repeatability of the fabrication and measurement protocol.

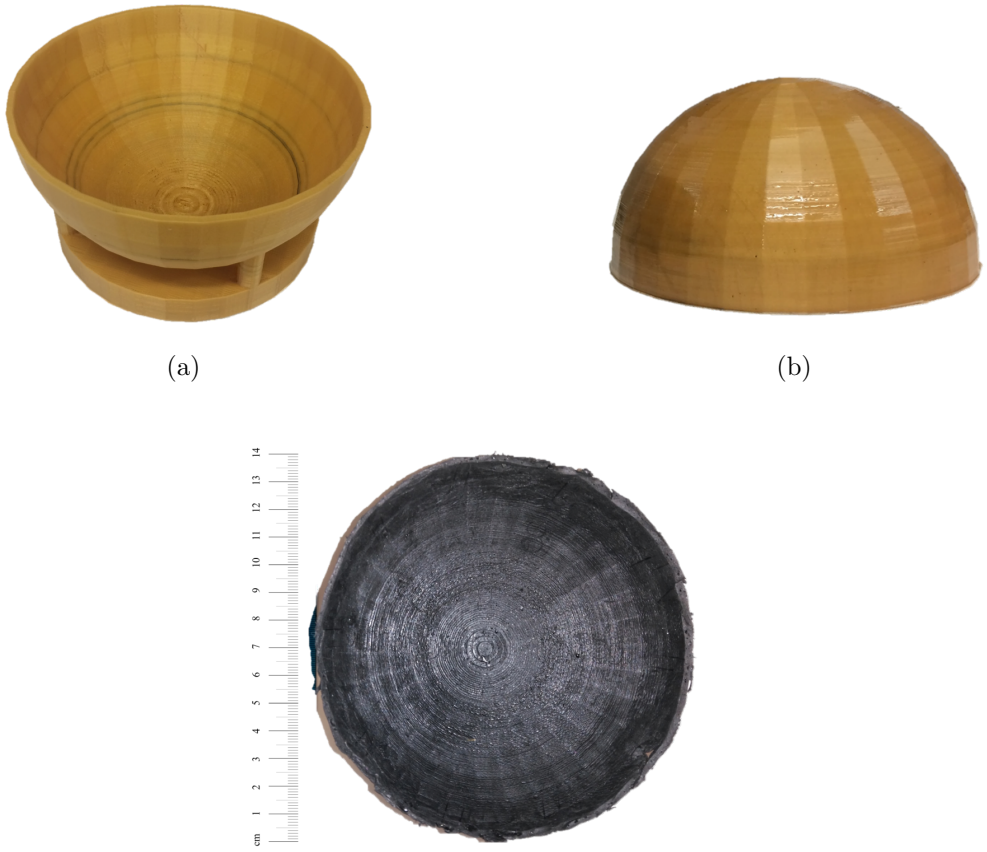


Figure 3.4: (a) Mould, and (b) counter-mould used to make the skin layers for the experimental phantoms. (c) Top-view of an example skin produced in this study.

3.3.4.2 Glandular Tissues

The glandular tissues of the breast were modelled as conical structures that radiate from the nipple area. These were fabricated in half-cylindrical moulds, and the resulting half-cylinders were then separated into cones. The mould and sample glandular pieces from before and after division are shown in Fig. 3.5.

In total, 3 batches of glandular mixture were made, resulting in 48 conical structures. For each glandular structure, the mass and volume were measured and the density calculated. As detailed in Table 3.5, the mean glandular density for each batch varies between 1.19 g/mL and 1.22 g/mL, with a standard deviation of 0.02 and 0.03, respectively. Once again, this

Table 3.4: Summary of thickness and dielectric properties for each skin sample, where M stands for Mean, and SD for Standard Deviation. Dielectric properties (relative permittivity, ϵ_r, and conductivity, σ) are shown at a frequency of 3 GHz.

	Thickness (mm)		ϵ_r		σ (S/m)	
	M	SD	M	SD	M	SD
Skin A	2.99	0.63	30.3	2.2	1.9	0.4
Skin B	2.53	0.31	31.4	1.2	1.9	0.3
Skin C	2.20	0.42	31.5	1.2	2.0	0.2
Skin D	2.31	0.71	30.1	1.6	2.0	0.4
Skin E	1.98	0.63	29.8	1.2	1.9	0.2
Skin F	2.06	0.36	30.4	1.8	2.1	0.4
Skin G	2.17	0.51	29.4	1.2	2.0	0.3

Figure 3.5: Glandular tissue construction. The half-cylindrical mould used to fabricate the glandular tissue is displayed on the left. An example of an uncut glandular piece is then shown in the middle, and the two resultant cones are shown on the right.

Table 3.5: Summary of density and dielectric properties for each batch of glandular tissue, where M stands for Mean, and SD for Standard Deviation. Dielectric properties (relative permittivity, ϵ_r, and conductivity, σ) were measured at a frequency of 3 GHz.

	Density (g/mL)		ϵ_r		σ (S/m)	
	M	SD	M	SD	M	SD
Batch 1	1.19	0.02	41.8	1.7	1.9	0.4
Batch 2	1.22	0.03	43.1	1.9	2.3	0.3
Batch 3	1.19	0.02	45.9	3.5	2.7	0.7

serves as a positive indication of the reproducibility of the phantom making process, and the consistency of the resulting glandular tissue.

The dielectric properties were also measured for each batch of glandular mixture, using a similar procedure to that used for the skin layer. This resulted in a mean relative permittivity ranging between 41.8 and 45.9, which is well matched to the target dielectric properties (Table 3.1 and Table 3.2).

3.3.4.3 Tumour Tissues

The shape of breast tumours is an important factor in clinical diagnosis. Models of tumours were developed based on the ideas presented in Section 3.2, and a series of approximately spherical, non-spiculated targets was fabricated to model benign tumours, and a series of spiculated models was fabricated to model malignant tumours.

After fabrication, each tumour model was photographed, the dimensions recorded and the mass measured as shown in Table 3.6. Approximately spherical targets (labelled as low spiculation) were moulded by hand to the desired dimensions and shape before the polyurethane mixture had cured. Spiculated targets were fabricated in two ways: some were carved from spheres post-curing (labelled as intermediate spiculation), while others were assembled from pieces of the polyurethane mixture once it had cured (labelled as high spiculation). The top view of all tumour models produced in this study is shown in Fig. 3.6.

Appropriate contact between dielectric probe and tissue, and the presence or absence of air gaps in the interface are known to have a significant impact on the accuracy of dielectric properties measurements [103], [104], [177], [178]. The sizes and complex shapes of the tumour models used in the present study meant that a suitable contact between probe and tumour

Table 3.6: Summary of size and mass of all tumour models. The dimensions of each tumour model were recorded in the Head-to-Toe (H-T), Left-to-Right (L-R) and Back-to-Front (B-F) axes.

	Spiculation Label	H-T (mm)	L-R (mm)	B-F (mm)	Avg Size (mm)	Mass (g)
L1	Low	5.22	6.96	3.69	5.29	0.2
L2	Low	8.36	8.83	6.18	7.79	0.5
L3	Low	10.86	10.95	10.84	10.88	1.5
L4	Low	13.66	13.78	11.95	13.13	2.1
L5	Low	20.36	20.70	19.44	20.17	6.4
L6	Low	11.28	20.48	8.20	13.32	1.9
L7	Low	21.70	13.10	9.11	14.64	2.1
L8	Low	11.87	12.35	20.85	15.02	3.2
I9	Intermediate	11.73	15.04	11.73	12.83	0.9
I10	Intermediate	15.16	16.63	9.66	13.82	1.3
I11	Intermediate	13.17	22.05	17.54	17.59	1.5
I12	Intermediate	18.83	14.07	13.63	15.51	1.7
I13	Intermediate	14.38	26.96	8.80	16.71	1.7
I14	Intermediate	16.58	10.54	22.38	16.50	1.5
H15	High	14.11	17.18	15.51	15.60	0.8
H16	High	20.53	19.53	14.09	18.05	0.9
H17	High	15.79	18.97	3.73	12.83	0.3
H18	High	23.17	17.16	13.36	17.90	0.9
H19	High	23.26	8.73	23.47	18.49	0.9
H20	High	21.32	24.51	21.00	22.28	1.5
H21	High	25.79	12.52	33.00	23.77	1.1
H22	High	21.02	22.80	34.44	26.09	0.8

could not be guaranteed. For this reason, the dielectric properties of each individual tumour phantom could not be measured accurately.

3.3.4.4 Fat Tissue and Modular Phantom Assembly

The final step in creating a rubber solid phantom is assembling the different tissues. A skin layer was placed inside a breast mould, and a pre-determined number of glandular structures were placed inside the skin (Fig. 3.7a), to produce phantoms with varying contents of glandular tissue. The interior of the phantom was then filled with fat tissue, which binds all of the different

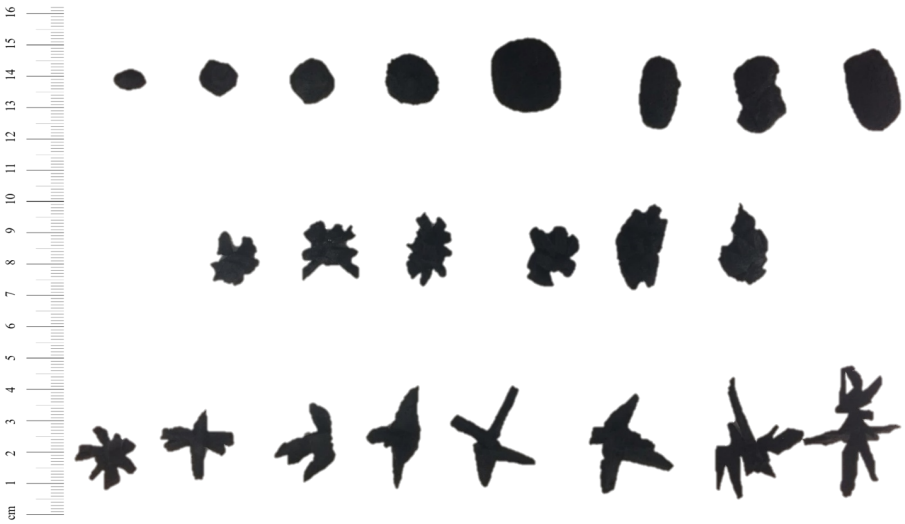


Figure 3.6: Top view of the tumour models produced in this study. In the top row, the low spiculation models are shown, followed by the intermediate spiculation models in the middle row, and the high spiculation models in the bottom row.

(a) (b)

Figure 3.7: Phantom assembly (top-view): (a) glandular structures inside the skin; (b) fat layer is used to fill the phantom, leaving a hole for inclusion of a target.

breast tissues together (Fig. 3.7b). While assembling the breast phantom, a hole was left for inclusion of a target. The hole was created by placing a 3D-printed cylindrical mould (with an external diameter of 2.5 cm) inside the interior of the phantom, at a depth between 3.5 cm and 4.5 cm. Once the fat layer set, the phantom was complete. Table 3.7 summarises the location of the centre of the bottom of the hole in each BRIGID phantom.

Table 3.7: Coordinates for the centre of the bottom of the hole left for inclusion of a target in all Breast Imaging and Diagnosis (BRIGID) breast phantoms. The coordinates are shown in the Head-to-Toe (H-T), Left-to-Right (L-R) and Back-to-Front (B-F) dimensions.

Breast Phantom	H-T (mm)	L-R (mm)	B-F (mm)
BRIGID 0	15.00	0	38.15
BRIGID 10U	35.00	0	43.15
BRIGID 10E	30.00	0	33.15
BRIGID 15U	35.00	0	33.15
BRIGID 15E	30.00	0	33.15
BRIGID 20E	28.30	0	40.65
BRIGID 30E	25.50	0	33.85

Additionally, the tumours described in Section 3.3.4.3 were assembled into cylindrical plugs that can easily be inserted in any breast phantom that follows a modular design. This allows the tumour models produced in this study to be used in the experimental evaluation of any microwave breast system, regardless of the prototype shape or design.

The creation of the tumour plugs was achieved by using a 2-part 3D-printed mould consisting of a base and a cylinder which has an internal diameter of 2.5 cm. The tumour model was placed at the centre of the bottom of the plug, and fat mixture was poured into the plug. All tumour models were placed at a pre-defined angle inside the plug (denoted as "rotation 0"). Examples of some tumour plugs can be seen in Fig. 3.8. The combination of a base breast phantom platform with different tumour plugs allows the generation of many phantom configurations for test purposes.

3.3.5 Test Platform

From the procedures described in the previous sections, the BRIGID phantom set was created, for the testing of microwave systems for breast cancer detection and diagnosis.

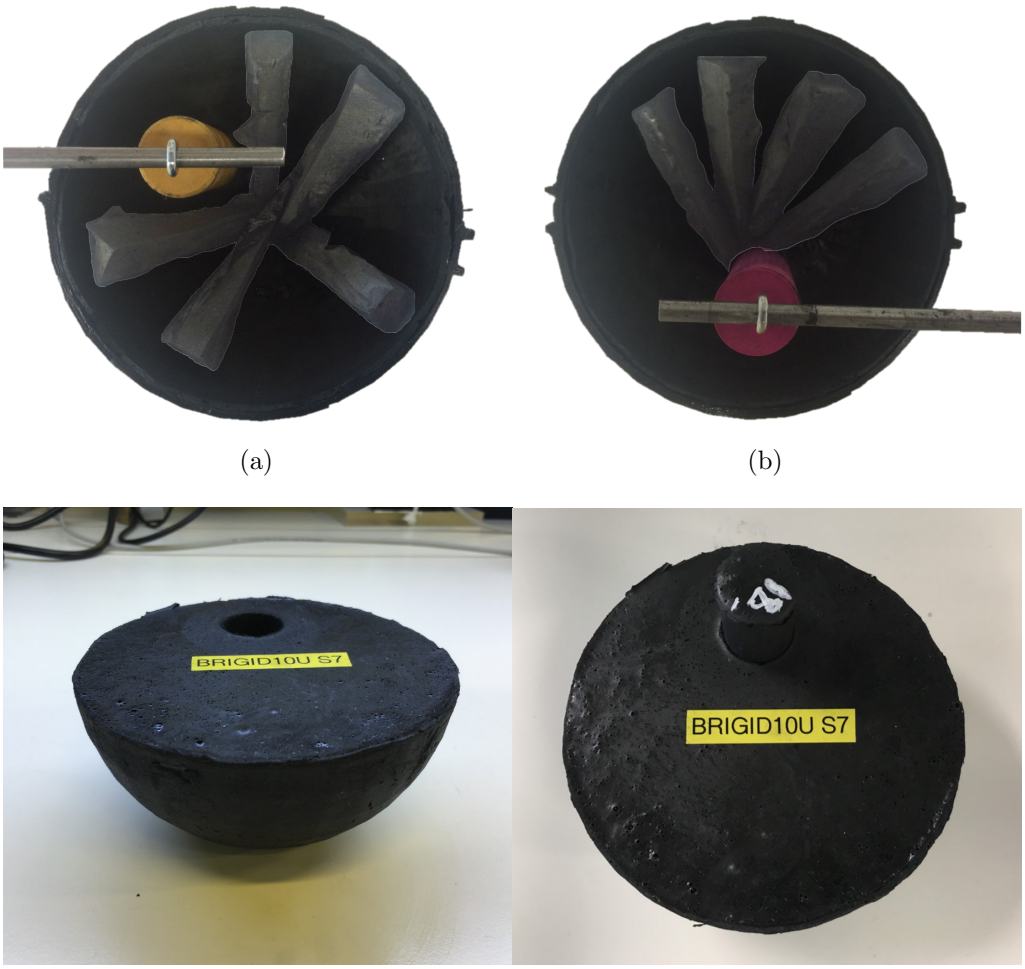
The BRIGID phantom set consists of 22 tumour models with 3 levels of spiculation, different shapes, and sizes ranging from 5 mm to 26 mm. All tumour models were fabricated and assembled into a cylindrical plug format, according to the procedure described in Section 3.3.4.4. The unique tumour phantoms enable, for the first time, a comprehensive study into the potential of microwave systems both for the detection and diagnosis of breast cancer.

Figure 3.8: Example of some of the tumour plugs to be used in combination with the breast phantoms. All plugs have an indication of "rotation 0" on the top.

The BRIGID phantom set also contains 7 breast phantoms, with varying glandular tissue content, between 0% and 30% in volume. All breast phantoms contain a skin layer which mimics the thickness variation of human skin, fatty background tissue, and a hole for the inclusion of a tumour. The distribution of glandular content within each breast is such that either:

1) The glandular tissue is distributed evenly within the breast, and the tumour is embedded within the glandular content (referred to as "even" distribution). An example is shown in Fig. 3.9a;
2) The glandular tissue is distributed asymmetrically within the breast; in this case, the tumour is near the skin and there is no glandular tissue in its vicinity (referred to as "uneven" distribution). An example is shown in Fig. 3.9b.

Due to the modular design of the BRIGID phantom set, a total of 154 unique combinations of breast and tumour phantoms are possible, allowing for a wide variety of breast cancer scenarios to be tested in any microwave breast system. In Table 3.8, a summary of the properties of the breast phantoms is shown for completeness. As noted previously, the BRIGID Phantom Set is stable-in-time as well as robust and easy to manipulate. By way of example, in Fig. 3.9c, the perspective view of the BRIGID 10U phantom is shown; Fig. 3.9d shows the top view of the same phantom when combined with a tumour plug.



(a) (b)

(c) (d)

Figure 3.9: Example of a Breast Imaging and Diagnosis (BRIGID) phantom. (a) shows the top view of the BRIGID 10E phantom during the assembly phase; the glandular tissue is evenly distributed in the breast, and a hole is left for the inclusion of a target, which is surrounded by the glandular tissue. (b) shows the top view of the BRIGID 10U phantom during the assembly phase, where the glandular tissue is distributed asymmetrically within the phantom; here, a hole is left for the inclusion of a target and there is no glandular tissue surrounding it. (c) shows the perspective view of the complete BRIGID 10U phantom, and (d) shows the top view of the same phantom when combined with a tumour plug.

Table 3.8: Summary of all relevant properties for each one of the 7 Breast Imaging and Diagnosis (BRIGID) breast phantoms. All values shown are means from all measurements taken. The dielectric properties (relative permittivity, ϵ_r, and conductivity, σ) were measured at a frequency of 3 GHz.

	Fat		Skin					Glandular				
	ϵ_r	σ (S/m)	Name	Thickness (mm)	Density (g/mL)	ϵ_r	σ (S/m)	Volume (%)	Distri-bution	Density (g/mL)	ϵ_r	σ (S/m)
BRIGID 0	9.00	0.18	A	2.99	1.05	30.30	1.90	0	—	—	—	—
BRIGID 10E	9.00	0.18	F	2.06	0.89	30.40	2.10	10	even	1.15	42.21	2.06
BRIGID 10U	8.79	0.14	G	2.17	1.00	29.40	2.00	10	uneven	1.23	43.44	2.16
BRIGID 15E	7.77	0.11	E	1.98	1.01	29.80	1.90	15	even	1.02	46.23	2.88
BRIGID 15U	7.80	0.11	D	2.31	1.07	30.10	2.00	15	uneven	1.16	45.46	2.57
BRIGID 20E	7.63	0.10	C	2.20	0.98	31.50	2.00	20	even	1.28	43.08	2.26
BRIGID 30E	8.25	0.12	B	2.53	1.07	31.40	1.90	30	even	1.17	41.26	1.81

3.4 Conclusions

Numerical and experimental evaluation of microwave prototypes using phantoms is an important testing stage before moving to *in vivo* and clinical testing. Phantoms enable evaluation of the data acquisition system, pre-processing methods and final algorithms in a stable and repeatable environment.

In this chapter, a numerical tumour modelling algorithm was first introduced, which resulted in a journal paper "Development of Clinically Informed 3-D Tumor Models for Microwave Imaging Applications", published in *IEEE Antennas and Wireless Propagation Letters*.

The numerical tumour models generated with the proposed modelling algorithm result in anatomically-realistic tumours models. The proposed algorithm was used to replicate tumours found in MRIs of the breast, achieving high levels of similarity in the produced images. In addition, a clinical validation study was conducted, resulting in the broad agreement that the proposed method allows for clinically-representative tumour models to be produced. The flexibility of the method in generating models of varying sizes, shapes and degrees of spiculation allowed for the creation of a database of tumour models, labelled by the consulted radiologists according to BI-RADS.

The proposed tumour models can be directly combined with the breast models from the UWCEM database, for FDTD simulation purposes. With the combined models, a high degree of clinical representation is achieved when further evaluating any microwave breast cancer diagnosis algorithms.

Based on the ideas described with the numerical tumour models, an experimental tumour phantom set was also designed, as well as a set of experimental breast phantoms — the BRIGID phantom set. This work resulted in an original journal contribution titled "Microwave Breast Imaging: experimental tumour phantoms for the evaluation of new breast cancer diagnosis systems", published in *Biomedical Physics and Engineering Express*.

The BRIGID phantom set consists of a set of experimental breast and tumour phantoms, and exhibits a number of characteristics that render it suitable for characterisation of prototype microwave systems regarding their potential to detect and diagnose breast cancer. In particular, the 22 tumour phantoms included in the BRIGID phantom set have varying shapes and levels of spiculation, to emulate benign and malignant breast tumours. The modelling of the level of spiculation in experimental tumour phantoms is a significant novelty over tumour phantoms typically used in previous experimental studies.

The tumour phantoms can be dynamically combined with a variety of breast phantoms that were fabricated to mimic the natural variation in the density of the female breast. All breast phantoms include a skin layer, different internal distributions of glandular and fat tissues, and are suitable for testing any hemispherical conformal microwave prototype, whether an immersion medium is required or not. The tumour plugs produced in this study can be combined with any breast phantom that follows a modular design.

Concerning the remainder of the work developed in this thesis, the BRIGID phantom set is used for the evaluation of the feasibility of machine learning-based platforms for the detection and diagnosis of breast cancer with microwave backscattered signals. Chapter 4 addresses this topic through the development of such a platform using numerical models, while Chapter 5 furthers this work using data acquired using the BRIGID phantom set.

CHAPTER 4

Machine Learning for Microwave Breast Diagnosis: Numerical Study

4.1 Introduction

To date, the majority of microwave breast systems have been developed with the primary goal of detecting the presence of tumours inside the breast, usually through the creation and analysis of images. As discussed in Section 2.4, a small number of studies have demonstrated the potential of using machine learning and classification algorithms to analyse microwave backscattered signals in more detail with the purpose of both detecting tumours and diagnosing their malignancy. These studies have been tested with relatively simple datasets, usually involving breast models with limited inclusions of glandular tissue, and tumour models that may or may not include a range of levels of spiculation to account for tumour malignancy. In addition, the datasets used to date usually locate the tumour only in the centre of the breast, which, as noted in Section 2.2.2.4, only accounts for 5.7% of the breast cancer cases found in clinical practice [72].

Notwithstanding the simplicity of the datasets used for testing, the detection and diagnosis studies described in Section 2.4.1 and Section 2.4.2 have suggested that there may be sufficient information contained in the microwave backscattered signals to inform both on the presence of tumours within the breast [140]–[142], [144]–[148], and on the differences in the shape between benign and malignant tumours [42]–[45], [48]–[53], [179]. Particularly, the same collection of studies has indicated that:

1) Signal pre-processing algorithms may be useful in highlighting the response of a tumour, while decreasing the response of the background noise due to the presence of glandular tissue;

2) Time-frequency representations of the original data may also be used to extract meaningful features contained within the backscattered signal;
3) There is a subset of channels in a microwave scan that contain the most relevant information towards the classification of backscattered signals, and the use of other channels may actually be detrimental towards the performance of the classification algorithms.

Despite the positive indications from previous studies, the optimal conditions for the realisation of machine learning-based platforms for the classification of microwave backscattered signals have yet to be fully understood.

In this chapter, a machine learning platform for microwave breast diagnosis is presented, and is analysed in detail using a numerical dataset of breast and tumour models that incorporates anatomical and clinical information. The use of a numerically-generated dataset is intended to provide a thorough investigation of the inherent complexities of automated microwave diagnosis platforms in controlled situations, when no other confounding factors are present such as, added sources of noise from experimental systems. This chapter also investigates whether a relationship exists between the predictive power of backscattered signals and the distribution of antennas in a microwave scan. In addition, this chapter explores the topic of appropriate machine learning methodology and how this may impact the fidelity of the results. The findings of this chapter are further explored in Chapter 5, using an experimental dataset obtained using laboratory phantoms.

Section 4.2 of this chapter identifies the main challenges still to be addressed in diagnosing breast tumours with microwaves. Section 4.3 then details the methodology used in this study, by introducing a three-stage diagnosis system for addressing some of the primary challenges. The results are presented in Section 4.4, while Section 4.5 discusses and analyses these results. Finally, Section 4.6 concludes the chapter.

The work described in this chapter has resulted in an original journal article titled "Diagnosing Breast Cancer with Microwave Technology: remaining challenges and potential solutions with machine learning", published in *Diagnostics* (2018).

4.2 Challenges in Microwave Breast Diagnosis System Design

In this section, some remaining challenges, as well as potential solutions, in the development of microwave breast diagnosis systems are discussed from two perspectives: addressing and exploiting the complexity of backscattered

signals gathered in clinically-realistic conditions (Section 4.2.1); and developing a validation methodology for the classification models (Section 4.2.2).

4.2.1 Characteristics of Clinically-Realistic Data

Benign and malignant tumours may present a wide range of sizes, shapes and spiculations at their margin, which can change the backscattered energy received at a given antenna. In addition, the shape of the human breast changes from person to person, as does the distribution of adipose and glandular tissues inside the breast, which effectively alters the attenuation along each propagation path. This anatomical diversity leads to equally diverse backscattered signals, making the design of a single platform for diagnosis a complex task. Some of the challenges related to breast and tumour composition can be summarised as follows:

1) Difficulty in capturing the tumour signature from the backscattered signal due to:
    i) Presence of skin, the response from which can be orders of magnitude larger than the tumour signature;
    ii) Presence of glandular tissue clusters, which can be confused with tumour tissue, due to similarities in composition (water content and generally higher dielectric properties);
    iii) Tumours can occur in different locations within the breast, embedded in various breast structures; much existing work has ignored the location of tumours.
2) Differences in the tumour signature for a given transmit-receive antenna pair due to:
    i) Tumours of different shapes, meaning antennas in different locations have a different view of the tumour;
    ii) Various angles between transmit and receive antennas, which can affect the phase of the tumour signature;
    iii) Varying distances between the antennas and the edge of the tumour.

Particularly regarding 1), a number of strategies have already been proposed in previous studies. Artefact removal algorithms have been proposed, which deal with large skin reflections and decrease the glandular tissue influence on the backscattered signals [180], [181]. Previous studies have also proposed: pre-processing signals by means of windowing to highlight and time-align the tumour signature [53]; extracting features based on

time-frequency representations of the data to further capture the characteristics of the tumour signature while disregarding the background noise [148]; and classifying a dataset according to tumour size before attempting at classification based on the level of malignancy [42]–[44], [48].

Concerning 2), while some studies have observed an improvement in diagnostic performance by restricting classification to the signals captured with the most informative transmit-receive antenna pairs [48], [141], [144]–[147], no thorough investigation of optimal antenna topology and optimal use of the information from each channel was found in the literature.

A further set of challenges exists in translating microwave breast diagnosis systems to experimental and clinical evaluation including: patient positioning and movement; intra-patient variation due to menstrual cycle and hormonal changes; and inter-patient variation in breast size, shape and composition. However, these are beyond the scope of this thesis.

4.2.2 Challenges in Building Robust Classification Models

For the work presented in this thesis, the creation of a robust machine learning based system can be considered to encompass three phases [182]:

- Training of the classification model, i.e., determining a set of parameter values to a model;
- Model selection/refinement, i.e., finding the subset of hyperparameters that leads to optimal model performance;
- Model validation, which should provide a reliable indication of the model's expected performance.

Ideally, a machine learning algorithm trained with a particular dataset should present with low bias and low variance, meaning it is generalisable to new, unseen datasets. Common practice is that a model should first be trained on a subset of the data, and then tested on another unseen subset of the data. The training set should be as large as possible, to minimise the variance in training the model, but the unseen subset of the data should also be representative of the original dataset so the performance evaluation is meaningful [182], [183].

However, performance evaluation commonly observed in the literature is prone to variations in approach, and often some degree of error, leading to overly-optimistic performance reports that may only be representative of a specific problem. Poor model validation is often due to:

1) Overfitting of the classification model during the training phase;
2) Overfitting during model selection/refinement;
3) Contamination of the information across the dataset.

Regarding 1) and 2), cross-validation has long been regarded as a good method to prevent overfitting of the model during training, and it is widely used as the basis for model selection/refinement. However, it has also been shown that using the performance obtained from cross-validation during model selection as the overall performance of the model might be overly-optimistic, and not generalisable. This effect is often referred to as selection bias [183], [184].

Poor model validation due to contamination of information (as mentioned in 3)) is often not addressed in previous studies and requires some explanation. With a dataset suffering from contamination of information, a machine classification model might learn from observations (and from relationships between observations) that it should not have been exposed to, ultimately leading to results that may be overly-optimistic or even unreproducible. For example, common feature transformation methods (such as, e.g. PCA) require the calculation of scores or weights between multiple observations; good machine learning practice would require that any such methods are first applied to the training set only, and the same transformations then projected onto the test and validation sets. Any pre-processing (e.g., normalisation) or feature-based methods that involve calculating weights across a dataset should be subject to the same practice.

Many of the issues listed above have not been explicitly addressed in previous studies. Implementing careful and consistent methodologies for model validation and performance evaluation should, however, become best practice. Ultimately, creating classification models without proper validation methodologies could compromise the usability of microwave breast diagnosis systems.

In the following section, the methodology designed for this study is explained, which addresses the challenges discussed above.

4.3 Methods

In this study, a 3-stage platform has been implemented with the purpose of diagnosing breast tumours using backscattered signals. The proposed methodology comprises data acquisition, data processing and diagnosis. The overall diagnostic architecture is depicted in Fig. 4.1.

Stage 1 consists of the microwave breast scan. To address some of the issues in dealing with clinically-realistic datasets (as highlighted in Section 4.2.1), a data processing stage was implemented next (Stage 2), comprising a Tumour Windowing (TW) approach to isolate signal segments of interest, and Feature Extraction (FE). The relative benefits of both algorithms are analysed by comparing the diagnostic performance of applying one of the following: only TW; only FE; TW in combination with FE, i.e. feature extraction performed after the tumour signature is windowed from the original backscattered signal.

Stage 3 consists of diagnosis and encompasses classification of the dataset through a range of techniques, including the random forest classification algorithm, antenna grouping, and final decision as benign or malignant. The concept that some channels (transmit-receive antenna pairs) might be more useful to improve diagnostic performance is also explored, by implementing three learning designs, where each classification model makes different use of the information from each channel. The three classification model types will be described in greater detail in Section 4.3.3.2.

With the algorithms implemented in Stage 3, this study aims to understand:

1) If the angle between the transmit and receive antennas in a channel (defined as channel angle for the remainder of this work) relates to its predictive power;
2) If the distance between the tumour and the channel has an impact on diagnostic performance;
3) how to best use the information from each channel while adhering to best machine learning practices.

In addition, a careful model validation methodology was implemented in Stage 3, to prevent issues like those detailed in Section 4.2.2. The three stages of the proposed microwave breast diagnosis platform will be described in greater detail in the following sub-sections.

4.3.1 Numerical Simulation

The numerical dataset of breast and tumour models used in this study was created with the FDTD formulation, as explain in Section 3.2.1.

MRI-derived breast models were taken from the repository created by the UWCEM laboratory [149]. All breast models in the repository are mapped to the dielectric properties of normal and malignant breast tissues as established by Lazebnik *et al.* [103], [104]. In total, 3 heterogeneous

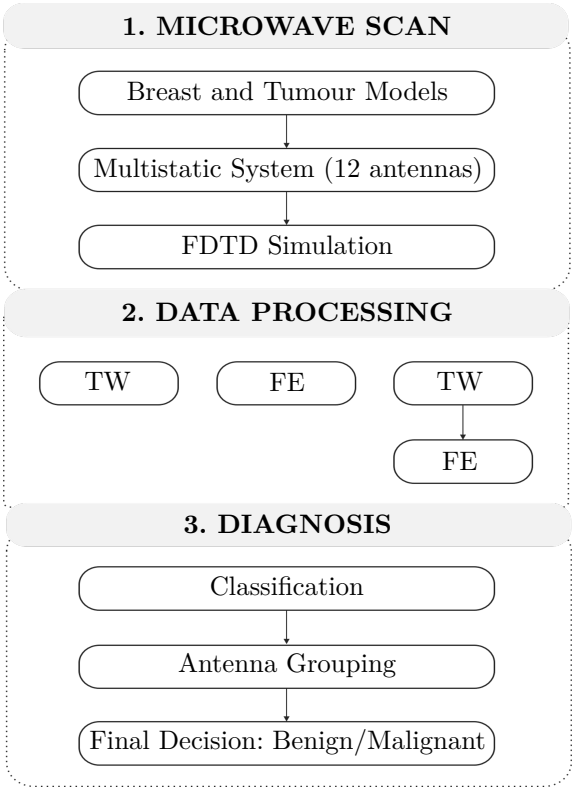


Figure 4.1: 3-stage diagnosis platform implemented in this study. Stage 1 consists of data collection from numerical models. Stage 2 consists of data processing by means of Tumour Windowing (TW) and Feature Extraction (FE); the relative importance of each algorithm is compared by applying TW, FE, and their combination. Stage 3 is the diagnosis stage, which uses random forests as the classifier, includes an antenna grouping algorithm, and ends with a final diagnosis of benign or malignant.

breast models were used in this study. The breast models range between 1% to 27% in terms of glandular tissue by volume of breast, with the remainder percentage of tissue corresponding to adipose tissue.

For the creation of tumour models, the clinically-informed tumour modelling algorithm described in the previous chapter was used to generate 72 unique tumour models, with average diameters ranging from 6 mm to 20 mm. Several degrees of spiculation were used to create tumours grouped into two distinct classes: smooth borders to represent benign tumours (with $0 \leq s \leq 0.25$), and spiculated borders for malignant tumours (with $0.50 \leq s \leq 0.90$), where s is the spiculation parameter with $0 \leq s \leq 1$. The tumours were placed in 5 different positions within the breast as described in Fig. 2.2, corresponding to locations in the four breast quadrants and the central portion.

The electromagnetic measurement system was modelled with a concentric ring of 12 equally-distanced Hertzian dipole antennas around the breast in a fully multistatic configuration (where the channel angle between two adjacent antennas is equal to 30°). Each antenna element is modelled as an electric current source. The antennas were immersed in a medium with dielectric properties equivalent to those of adipose tissue. The FDTD simulations were performed using a differentiated Gaussian pulse with centre frequency of 6 GHz and a bandwidth of 6 GHz. The spatial resolution of the system is 1 mm, and the sampling frequency is 600 GHz. Additionally, a reference simulation was also performed. This reference signal is later used to remove antenna effects in the backscattered signals from simulations of the full breast with tumours.

Figure 4.2 displays a schematic representation of the acquisition setup designed for FDTD simulations in this study, where the antennas are represented by the black diamonds surrounding the breast. A coronal slice of one of the breast models used in the study is shown, including glandular tissue in the interior, and a malignant tumour in one of the lower quadrants (the spiculated shape in black). To aid the visualisation of the setup, examples of paths from a transmitting antenna (Tx) to the tumour and from the tumour to a receiving antenna (Rx) are shown in dash and dot-dash lines. The channel shown in blue is an example of a "reflection channel" (i.e., a channels where the transmit and receive antennas are located on the same side of the target, thus generating signals with reflected information), and the channel shown in orange is an example of a "transmission channels" to mean channels where the transmit and receive antennas are located on opposite sides of the tumour, thus producing "through-channel" signals with transmitted information).

With the proposed setup, one microwave breast scan is composed of backscattered signals collected from 78 independent channels. In total, 1,080 microwave scans were performed (3 breast models each combined with 72 tumour models in 5 different positions within the breast). Therefore, a dataset containing a total of 84,240 signals is used in this study.

4.3.2 Data processing

This section describes the processing methods used to prepare the data ahead of classification. Two methods are used to process the backscattered signals acquired in Stage 1 of the 3-stage diagnosis platform (Fig. 4.1): tumour windowing (Section 4.3.2.1), and feature extraction (Section 4.3.2.2).

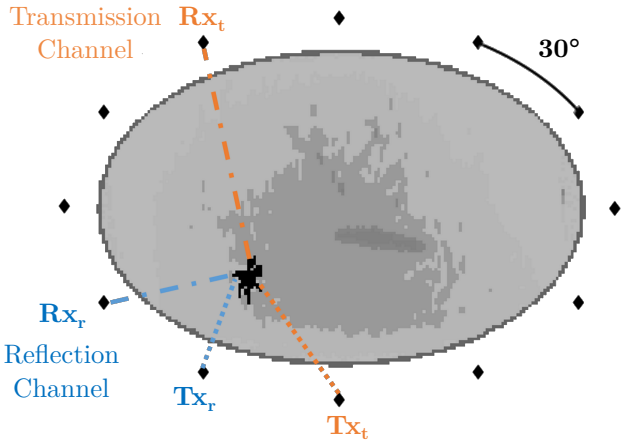


Figure 4.2: Representation of the acquisition setup designed for this study, where the antennas are represented by the black diamonds surrounding the breast. A coronal slice of one of the breast models is shown; the breast has glandular tissue in the interior, and a malignant tumour in one of the lower breast quadrants (represented by the black spiculated shape). With this acquisition configuration, the channel angle between two adjacent antennas is 30°. Examples of paths from a transmitting antenna (Tx) to the tumour and from the tumour to a receiving antenna (Rx) are shown in the dash and dot-dash lines: antenna pair (Tx_r, Rx_r) shown in blue is an example of a reflection channel, and antenna pair (Tx_t, Rx_t) shown in orange is an example of a transmission channel.

4.3.2.1 Tumour windowing

The Tumour Windowing (TW) algorithm presented in this thesis relies on first identifying the location of the tumour. A variety of methods could be used for this purpose, for example, image reconstruction through microwave tomography as briefly presented in [53]. An investigation of methods of tumour location estimation is beyond the scope of this work, and the ideal tumour location will henceforth be used.

Once the tumour location is identified, the round-trip propagation delay between the tumour and each channel is calculated, based on the average propagation speed through three media: immersion medium, skin and interior of the breast; the estimated tumour response is then windowed from the backscattered signal. The approximate window length was decided empirically. Visual assessment of a subgroup of backscattered signals gathered with different tumour models embedded in breast models with varying background contents found that a window length of 2.5 times the pulse width is appropriate to extract the full tumour response from the signals.

The propagation delay is highly dependent on the average dielectric

properties of each medium; consequently, reflections from different tumours propagating through different paths will be difficult to align. To compensate for this effect, the windowing algorithm looks for the peak energy in each backscattered signal, and time-aligns the tumour responses on this basis. Each windowed tumour response is finally downsampled to a sampling frequency of 30 GHz. After downsampling, the window length of the tumour signatures consisted of 60 time samples corresponding to 2 ns.

The proposed tumour windowing algorithm, operating in the time-domain, has the following benefits: a high level of clutter resulting from the glandular clusters is potentially removed, and signals collected from different channels are time-aligned. As a result, the tumour response is isolated, potentially simplifying the task given to the classification algorithm. Prior to windowing, an artefact removal step was introduced in this study to compensate for antenna effects in the signals and to eliminate the reflection from the skin.

Figure 4.3 shows the effect of tumour windowing within the 3-stage diagnosis platform. The effect of tumour windowing is shown for a set of signals collected in one microwave scan with a benign tumour embedded in an adipose-only breast model. As indicated on the image on the top left of Fig. 4.3, signals are first collected from a microwave scan in the time-domain; artefact removal is then applied to remove the skin response (image on the top right). Then, the tumour windowing algorithm is applied in two steps. First (image on the bottom left of Fig. 4.3), a window is selected in the backscattered signals that contains the tumour responses. Second (image on the bottom right of Fig. 4.3), the peak responses of each tumour signal are time-aligned.

When only tumour windowing is applied during Stage 2 of the 3-stage diagnosis platform (Fig. 4.1, "TW"), the windowed time-domain signatures are treated as independent observations, which are then passed as input to the classification algorithm.

4.3.2.2 Feature extraction

Feature Extraction (FE) is frequently applied to capture meaningful information embedded in a signal, and is helpful in reducing the dimensionality of a classification problem when compared to using the original data.

Visual analysis of backscattered signals reveals that benign tumours result in signals that tend to preserve the original morphology of the Gaussian peak, while malignant tumours result in more irregular signals, due to increased reflections from tumour spicules. Therefore, the use of a set of features that capture signal morphology and frequency content for diagnosis is also

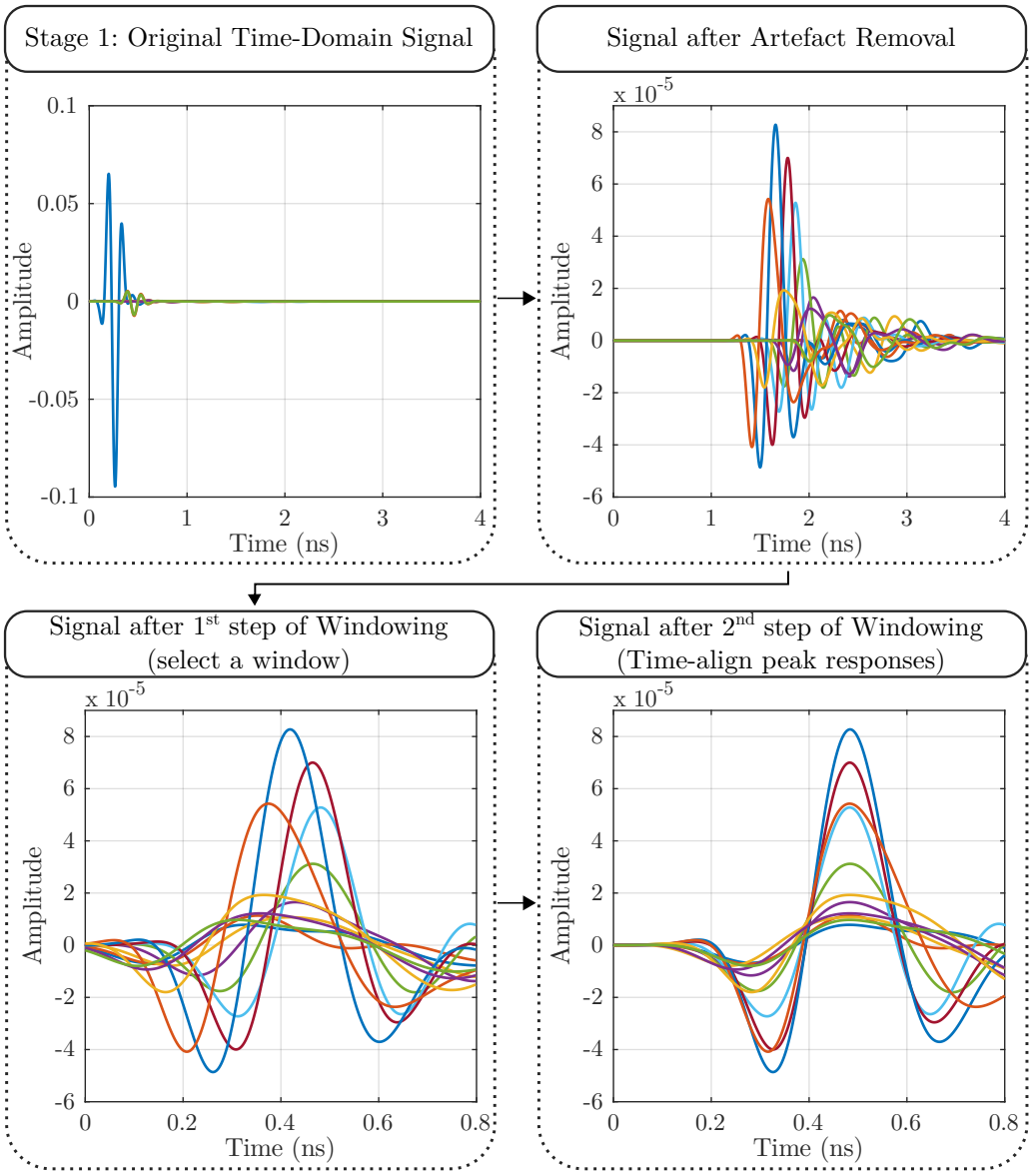


Figure 4.3: Effect of tumour windowing on the subset of signals collected from one microwave scan with a benign tumour embedded in an adipose-only breast model. The process of tumour windowing within the 3-stage diagnosis platform is shown. Signals from a microwave scan are collected in the time-domain, in Stage 1 of the diagnosis platform (image on the top left). Artefact removal is then applied to remove the skin response and to compensate for antenna effects (image on the top right). Then, the tumour windowing algorithm is applied in two steps: first, a window is selected in the backscattered signals that contains the tumour responses (image on the bottom left); second, the peak responses of each signal are time-aligned (image on the bottom right). Each windowed and time-aligned tumour response is treated as an independent observation and passed as input to the classification algorithm.

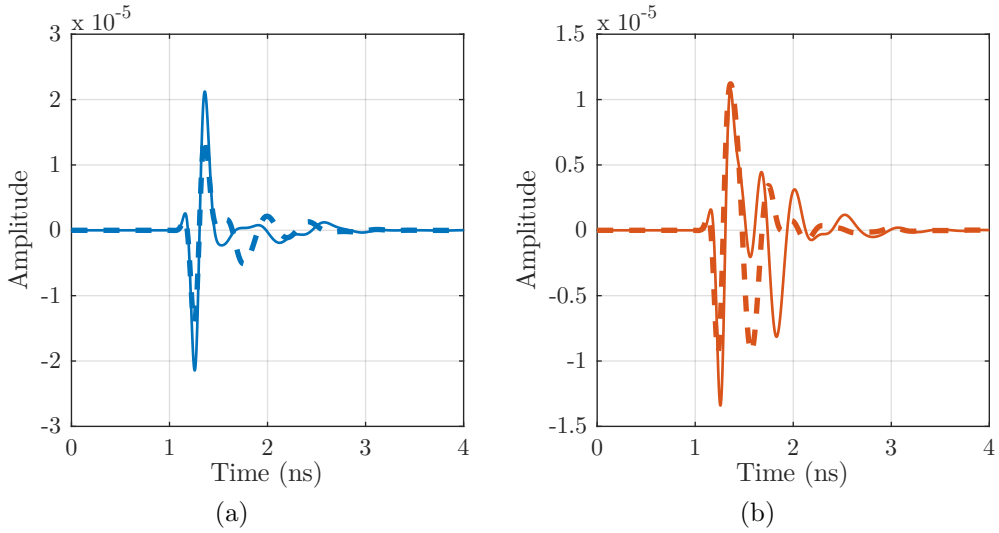


Figure 4.4: Example of tumour signatures from (a) benign tumour models and (b) malignant tumour models, captured in ideal conditions, in a fully adipose breast model. In (a), the benign tumour signatures are smooth, and the shape of the Gaussian curve is largely preserved. In (b), the malignant tumour signatures are subject to a greater degree of distortion, exhibiting an increased number of peaks.

examined here. The proposed feature extraction method relies on peak analysis of different time and frequency representations of the original data, where each group of features is calculated for the signal collected by each channel of each scan. As the extraction of features is done independently on each observation, no calculations are made across the dataset and between tumour signatures, which prevents accidental data contamination issues, such as those described in Section 4.2.2.

By way of example, Fig. 4.4 displays some of the differences identified by visual analysis of benign (Fig. 4.4a) and malignant signatures (Fig. 4.4b). The signals were collected under ideal conditions to highlight the expected differences between types of tumours, with an adipose-only breast model; for both tumour types, two tumour models were simulated with different sizes and shapes. The resultant signals have been time-aligned and windowed. As observed in Fig. 4.4a, the backscattered signals from the benign tumour models exhibit little distortion and the original Gaussian shape is preserved well; conversely, in Fig. 4.4b, the malignant tumour models result in backscattered signals with a higher level of waveform distortion.

In total, 30 features were extracted from each signal, divided into four sub-groups, as shown in Table 4.1.

Table 4.1: Description of all 30 features used in this work, divided into four sub-groups: Time-domain features, Autocorrelaton features, Power Spectral Density (PSD) features based on Welch's method, and PSD features using the periodogram method.

Time-domain features, calculated from the windowed signals
#1 – #2 Amplitude and location of the maximum positive peak
#3 – #4 Amplitude and location of the maximum negative peak
#5 Variance
#6 Root-mean-squared error
#7 – #8 Number of positive and negative peaks
#9 – #10 Mean amplitude of the positive and negative peaks
#11 – #12 Mean Full-Width Half-Maximum (FWHM) of the positive and negative peaks
#13 – #14 Mean separation between positive and negative peaks
#15 Number of zero crossings
#16 Integral of the signal
#17 Integral of the absolute value of the signal
#18 Positive percentage area of the signal
#19 Negative percentage area of the signal
Autocorrelation features, which involves calculating the autocorrelation sequence of each signal [185], [186]. The following features are then extracted from the autocorrelation sequence
#20 Mean value of the autocorrelation sequence
#21 Number of peaks in the autocorrelation sequence
#22 Mean amplitude of the peaks
#23 Mean FWHM of the peaks
#24 Mean separation between the peaks
PSD features, estimate of the psd of the signal, using Welch's method [187]
#25 Mean value of the Welch estimate
PSD features, estimate of the psd of the signal, using the periodogram method [188], [189]
#26 Mean value of the periodogram estimate
#27 Number of peaks in the periodogram
#28 Mean amplitude of the peaks
#29 Mean FWHM of the peaks
#30 Mean separation between the peaks

In Fig. 4.5, Fig. 4.6, Fig. 4.7 and Fig. 4.8, the feature extraction process within the 3-stage diagnosis platform is shown. By way of example, the responses of a benign (blue) and a malignant (orange) tumour model embedded in an adipose-only breast model are shown. As shown in Fig. 4.5, after the microwave scan is collected, features #1 to #19 (from Table 4.1) are calculated from the time-domain signals, after artefact removal was performed to remove the skin response and to compensate for antenna effects. These features largely capture the morphological characteristics of tumour signals. Next, the autocorrelation sequence is calculated from the time-domain signals (after artefact removal), and features #20 to #24 are calculated from this (Fig. 4.6). These features are intended to capture inherent periodicity in the signals. Finally, the power spectral density based on Welch's method and the periodogram method are calculated from the time-domain signals (after artefact removal), and features #25 and #26 to #30 are extracted (Fig. 4.7 and Fig. 4.8, respectively). Power spectral density features indicate how the distribution of power changes depending on the level of malignancy.

When only feature extraction is performed on the original backscattered signals during Stage 2 of the 3-stage diagnosis platform (Fig. 4.1), the method is referred to as "FE". If feature extraction is performed after the backscattered signals have been processed with the tumour windowing algorithm, the method is referred to as "TW + FE"; in this case, the windowed signals are used to calculate the 30 features, as opposed to the original backscattered signals as shown in Fig. 4.5, Fig. 4.6, Fig. 4.7 and Fig. 4.8.

4.3.3 Computer aided diagnosis

This section describes, Stage 3 (Diagnosis) of the 3-stage microwave diagnosis platform described in Fig. 4.1. An overview of the random forests classification algorithm is first provided in Section 4.3.3.1. Section 4.3.3.2 describes the three learning designs implemented in this study. The antenna grouping algorithm is detailed in Section 4.3.3.3. The validation methodology is described in Section 4.3.3.4, and the metrics to assess diagnostic performance are discussed in Section 4.3.3.5.

4.3.3.1 Classification algorithm: Random forests

In this study, random forests [190] were implemented to classify backscattered signals.

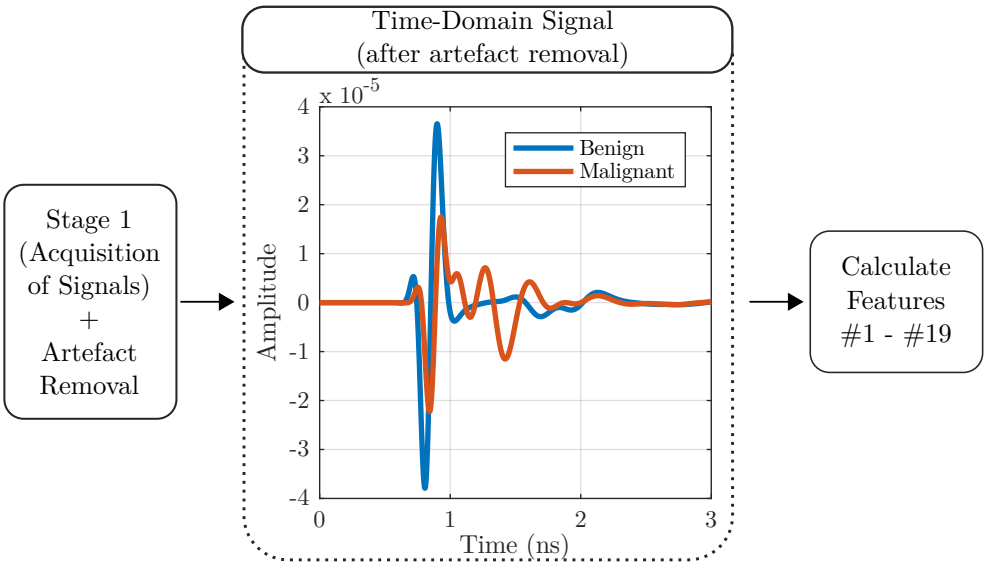


Figure 4.5: Process of feature extraction within the diagnosis platform. The responses of a benign (blue) and a malignant (orange) tumour model embedded in an adipose-only breast model are shown. After the acquisition of signals in Stage 1 (Microwave Scan), features #1 to #19 are calculated from the time-domain signals after artefact removal.

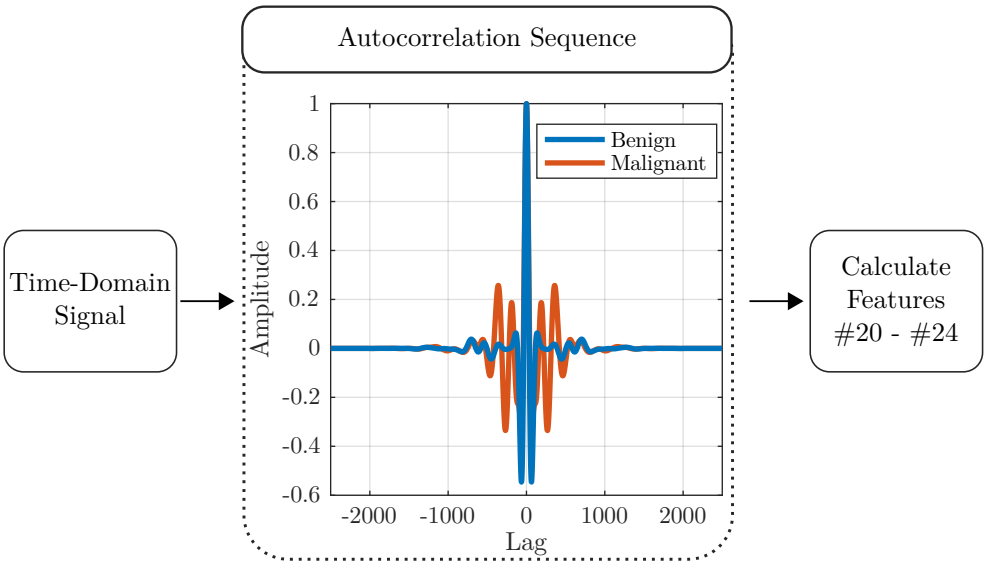


Figure 4.6: Process of feature extraction within the diagnosis platform. The responses of a benign (blue) and a malignant (orange) tumour model embedded in an adipose-only breast model are shown. The autocorrelation sequence is calculated from the time-domain signals after artefact removal and features #20 to #24 are then calculated.

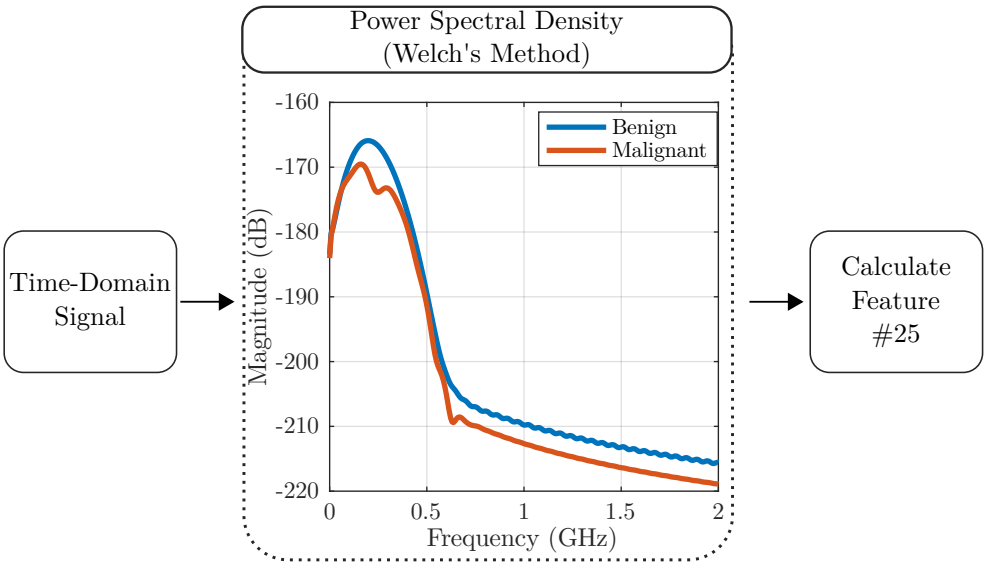


Figure 4.7: Process of feature extraction within the diagnosis platform. The responses of a benign (blue) and a malignant (orange) tumour model embedded in an adipose-only breast model are shown. The power spectral density based on Welch's method is calculated from the signals after artefact removal and feature #25 is then calculated.

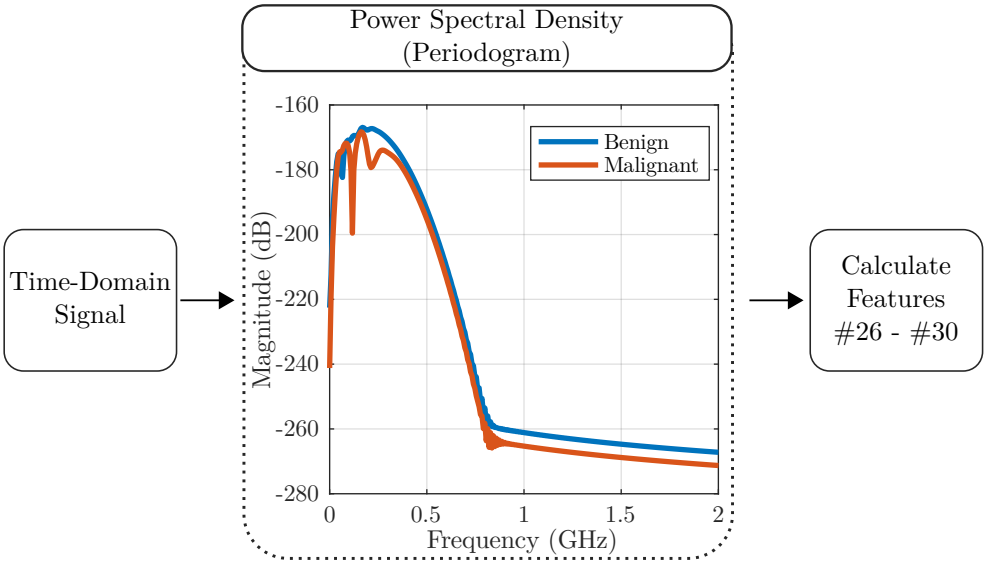


Figure 4.8: Process of feature extraction within the diagnosis platform. The responses of a benign (blue) and a malignant (orange) tumour model embedded in an adipose-only breast model are shown. The power spectral density based on the periodogram method is calculated from the signals after artefact removal and features #26 to #30 are then calculated.

The method of random forests is an ensemble method that essentially works by generating many single classification trees [191] and outputting the class that is the mode of the classes of all individual trees. Each tree is grown (i.e., trained) using a randomly sampled subset of observations and features from the entire dataset. Due to the inherent randomness in the process, the generated trees are uncorrelated, which ultimately contributes to the low bias and low variance of the algorithm. Random forests provide generalisable models that tend not to overfit, are quick to run and are easy to interpret [190].

For the operation of a random forest, one-third of the observations in the original dataset are left out when training each tree. These observations are referred to as out-of-bag and are used as a separate set to assess the performance error of each tree. The out-of-bag error provides a measurement of the generalisation ability of the process, which is useful when optimising the internal parameters of the random forest. Random forests also allow measuring the importance of each feature in the training of each tree. In the context of diagnosing backscattered signals as benign or malignant, a measure of feature importance could provide the means to further refine classification models.

In this study, the following hyperparameters of the random forest were optimised to ensure good trained models: number of trees, number of features, leaf size. A Bayesian optimisation algorithm was implemented to perform the search for the best hyperparameters. The best hyperparameters were deemed to be those yielding the smallest out-of-bag misclassification error (that is, the hyperparameters yielding the highest accuracy).

4.3.3.2 Design of classification models

Although all channels in a given scan may contain information about a tumour, the tumour signature varies between channels depending on the location of the tumour relative to the antennas in a channel, and also on the channel angle (i.e., the angle between transmit and receive antennas in a channel).

The variance in the tumour signatures between channels may impact the performance of the classification model, as the variance between channels may be as large as the variance between the signatures of benign and malignant tumours. To explore the significance of inter-channel variance, three types of classification models were designed, which differ in the way signals from different channel angles are utilised by the classification algorithm. The three types of classification models are shown in Fig. 4.9. Differences in the performance of the three types of classification models may help to identify

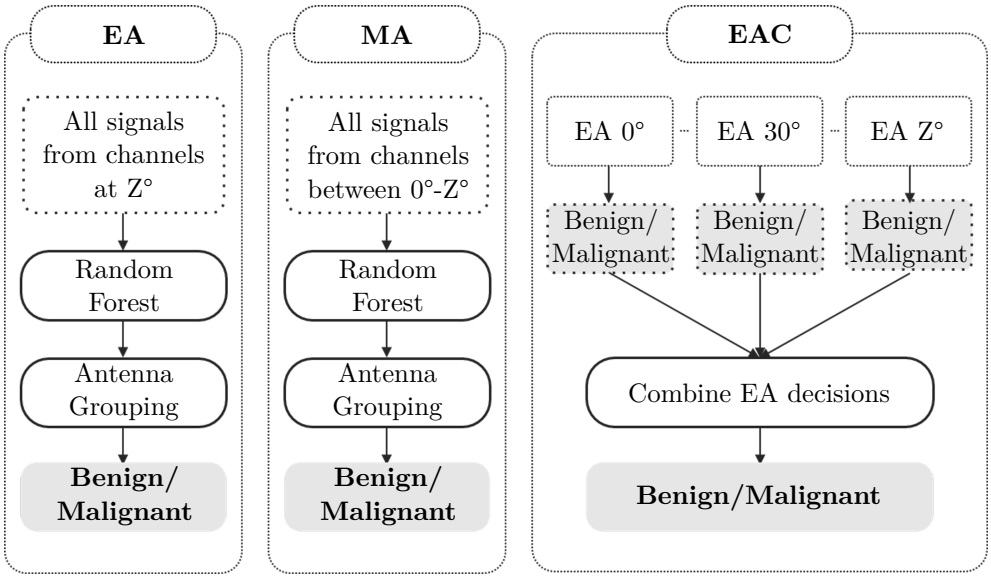


Figure 4.9: Description of the three types of classification models implemented: EA (Equal Angle), MA (Multiple Angle) and EAC (Equal Angle Combined). The classification models vary in the way signals from different channel angles are utilised by the classification algorithm, where Z represents the channel angle (Z varies between 0° and 180°, and increases in steps of 30°). EA models only classify signals from a single channel angle. MA models simultaneously classify signals collected at multiple channel angles. Through majority voting, EAC models combine the predictions from multiple EA models at different channel angles to produce a final diagnosis.

if an optimal antenna pair topology exists in terms of the channel angle, which can ultimately contribute to improving diagnostic performance.

The assumed system architecture is as described in Section 4.3.1, with one ring of 12 antennas equally distributed around the breast. Let Z be the channel angle; here, $Z \in [0, 180]°$, and Z increases in steps of 30°.

Equal Angle (EA) classification models only receive information from channels with equal channel angles. Seven EA models were built to assess if channels with equal angles contribute to improved diagnostic performance.

Multiple Angle (MA) models use information from multiple channel angles simultaneously. If such a model underperforms, it will serve as an indication that the information captured by channels with different channel angles varies significantly, and that the classification model cannot adequately learn the similarities within benign and malignant tumours across signals collected at different angles. In total, six MA models were built using antenna pairs in the interval $[0, i.Z]°$, where $Z = 30°$ and $i = 1, 2, ..., 6$, until all antenna pairs were used.

Equal Angle Combined (EAC) models use all possible EA models (one

Table 4.2: Summary of the classification models built for this study, and the channel angles used in each model. In Equal Angle (EA) models, only signals from channels at the specified angle are used in the process. In Multiple Angle (MA) models, all signals from channels in the specified range are used. In Equal Angle Combined (EAC) models, individual EA models in the specified range are combined through majority voting to produce a final diagnosis.

	EA	**MA**	**EAC**
#1	0°	–	–
#2	30°	0° – 30°	0° – 30°
#3	60°	0° – 60°	0° – 60°
#4	90°	0° – 90°	0° – 90°
#5	120°	0° – 120°	0° – 120°
#6	150°	0° – 150°	0° – 150°
#7	180°	0° – 180°	0° – 180°

for each channel angle), and the predictions from each one are combined (through majority voting) at the end to produce a final diagnosis. By combining the predictions from each individual model, models which yield an incorrect result are likely to be disregarded, ultimately contributing to an increase in diagnostic performance. As before, six EAC models were built using antenna pairs in the interval $[0, i.Z]^\circ$, with $Z = 30^\circ$, until all antenna pairs were used. Table 4.2 summarises the models of each type, in particular the range of angles considered in this study.

4.3.3.3 Antenna grouping

Each patient scan is comprised of signatures collected from 78 different channels (as per the system architecture described in Section 4.3.1), which are classified independently. However, in a realistic, clinical diagnostic system, a diagnosis is given based on a full scan, and not on the basis of a single signature. This means that the independent channel predictions need to be combined to form a final diagnosis. In the existing literature, either the procedure in determining the final diagnosis is not thoroughly discussed, or the diagnostic performance is reported based on the results from the independent channels.

To address this, an antenna grouping algorithm is implemented in this study, by which the predictions of the independent channels are grouped, and a majority vote is completed to determine if a scan is benign or malignant. The advantages of implementing such an algorithm are two-fold. Firstly,

with microwave diagnosis systems, the possibility should be considered that a signal contains lower quality information about the tumour shape, which could result in incorrect predictions about its malignancy (e.g., signals from channels that may have poor signal-to-noise ratios). By implementing the antenna grouping algorithm, a mechanism is created that allows incorrect predictions from lower quality channels to be reduced in importance. Secondly, channels closer to a tumour should intuitively produce more useful information for its diagnosis; here, the proximity between tumour and channel is defined as the proximity of the tumour to a straight line path between the Tx and Rx antennas. By implementing a ranked version of the antenna grouping algorithm, it is possible to investigate if the proximity between tumour and channel translates into higher diagnostic performances. The ranked version of the antenna grouping algorithm operates as follows. Let W be the number of channels used to perform antenna grouping, ordered by proximity to the tumour. Antenna grouping is performed by increasing W in steps of 1, until all available channels are used. For example, if $W = 3$, the majority vote is taken from the signals collected by the 3 channels closest to the tumour, before deciding on the final diagnosis.

4.3.3.4 Validation methodology

In this study, a validation methodology based on the idea of nested cross-validation [183] has been implemented to assess diagnostic performance, and mitigate sources of contamination when optimising the classification model. It has been shown that nested cross-validation helps prevent overly optimistic reports of model performance [183], [184]. An overview of the process is shown in Fig. 4.10, and can be summarised as follows:

- The entire dataset is divided into k stratified folds, containing equal representations of each class. In this study, $k = 5$ outer folds was chosen as it offers a good compromise between a statistically robust performance analysis and speed of implementation. All signals from one breast scan are kept together when splitting each fold into training and test.
- For each outer fold, the model is trained and the classifier hyperparameters optimised. As previously detailed, random forests directly provide the out-of-bag error, which serves as an unbiased estimate of the model performance when optimising its hyperparameters. When using other classifiers another inner cross-validation loop can be implemented at this stage.
- The predictive power of the model is then reported as the average performance obtained in the test sets across all outer folds.

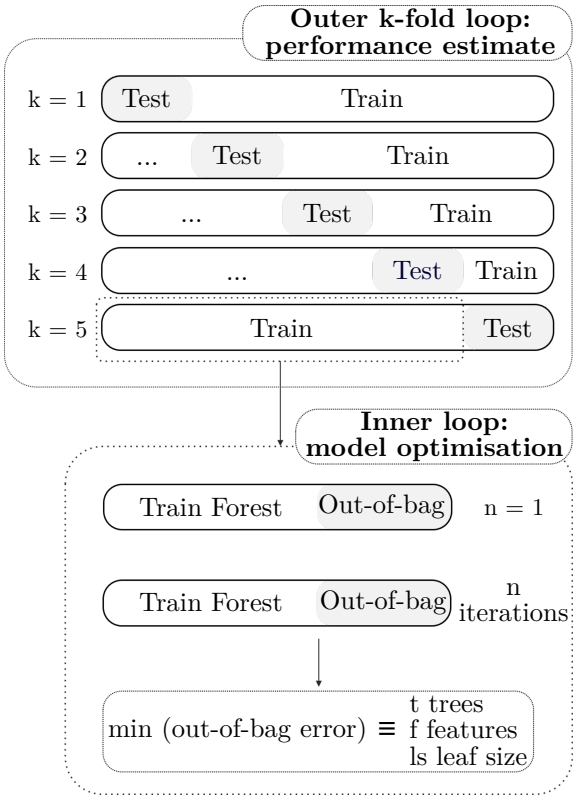


Figure 4.10: Nested cross-validation methodology implemented in this study to perform model optimisation and estimate model performance. In each fold, the random forest model is optimised on the train set, and new predictions are made on the test set. The predictive power of the model corresponds to the average performance obtained in the test sets across all outer folds.

4.3.3.5 Performance metrics

In this study, the performance of a classification model is assessed by plotting the Receiver Operating Characteristic (ROC) curves. ROC curves are created by plotting the false positive rate achieved by the classification model in the horizontal axis, against the true positive rate in the vertical axis [192]. ROC curves provide with a simple graphical representation of the diagnostic ability of the classification model, by varying the decision threshold that is used in producing the final binary decision, i.e., whether breast tumours are benign or malignant.

The Area Under the ROC Curve (AUC) is also used as a measure of classification performance. Generally, the higher the AUC, the better the classifier performs.

4.4 Results

This section is divided into three sub-sections. Section 4.4.1 discusses the issue of antenna topology and antenna grouping. Here, a relationship between the channel angle (angle between transmit and receive antennas) and predictive power is investigated, resulting in the proposal of a method to use the information from several multistatic scan channels. In Section 4.4.2, the effect of increasing tissue heterogeneity on overall diagnostic performance is discussed. Section 4.4.3 identifies possible avenues to expand on the knowledge gained with the extraction of features.

4.4.1 Design of classification models

This section details the analysis of the optimal design of classification models, tested in a breast model containing 5% of glandular tissue by volume.

Three types of classification models were defined in Section 4.3.3.2: EA, MA, EAC. Figure 4.11a, Fig. 4.11b and Fig. 4.11c detail the diagnostic performance achieved by all models produced, for each of the processing methods under analysis, TW, FE and TW+FE, respectively. The effect of antenna grouping (as defined in Section 4.3.3.3) is also investigated in Fig. 4.11, by comparing diagnostic performance before antenna grouping (solid lines) and after antenna grouping is applied (dashed lines), using all available channels in the majority vote.

Firstly, the positive impact of antenna grouping is clearly noticeable. The diagnostic performance when antenna grouping is applied is always superior (as shown by the dashed lines in Fig. 4.11). By taking the majority vote of all individual decisions from one single breast scan, a minority of incorrect predictions are cancelled by a majority of correct classifications. A more in-depth analysis of the effect of the ranked version of the antenna grouping algorithm reveals that at least 3 channels are necessary to achieve reliable diagnostic performance; however, above 3 channels, the performance stabilises and only minor improvements are observed (these results are not shown in Fig. 4.11). This result is seen across all classification model types (EA, MA and EAC), and by applying either of the pre-processing methods (TW, FE and TW+FE).

In Fig. 4.11, it is also noticeable that EA and EAC models generally seem to outperform MA models. This result confirms the hypothesis that classification models perform better when dealing with signals collected under the same conditions:

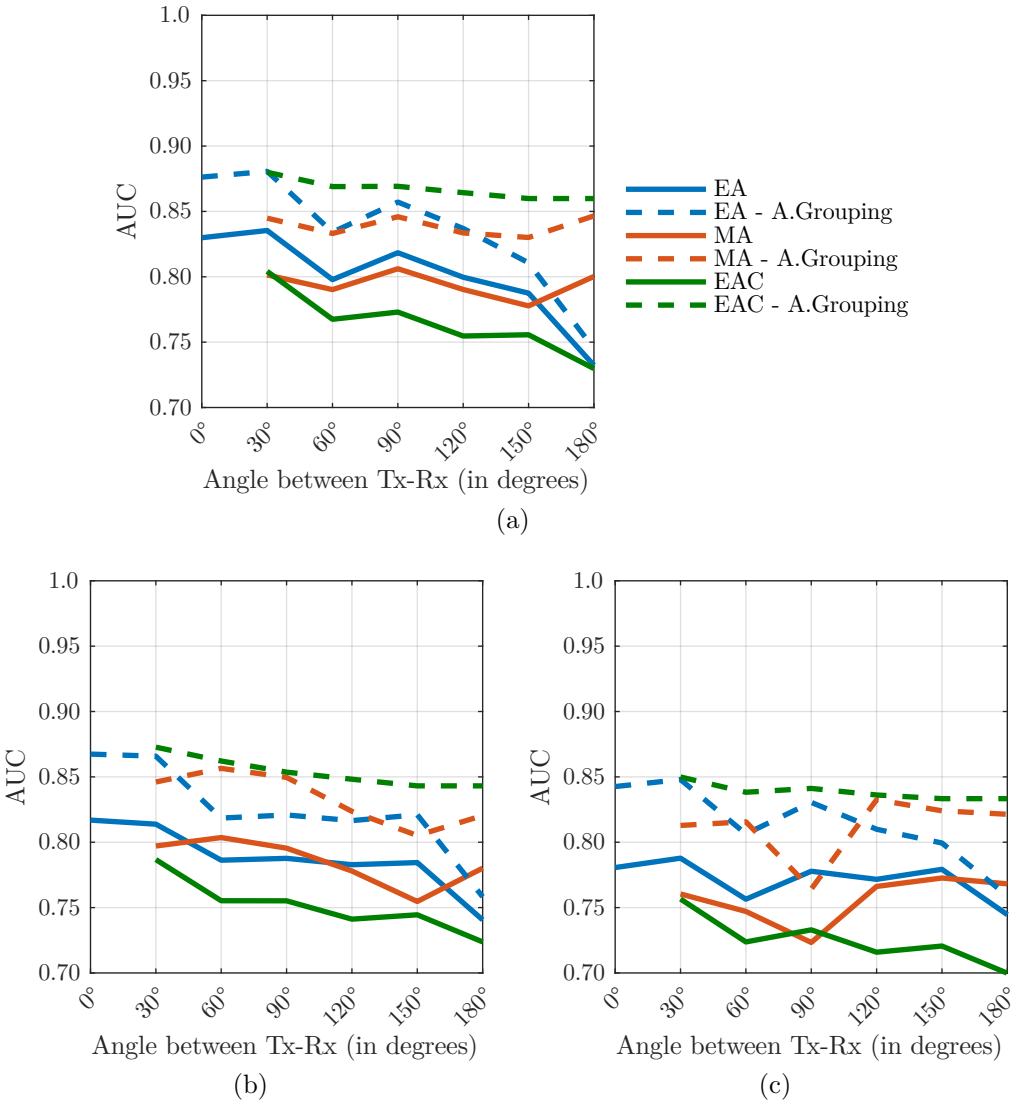


Figure 4.11: Diagnostic performance for the EA (Equal Angle, in blue), MA (Multiple Angle, in orange) and EAC (Equal Angle Combined, in green) models produced when using: (a) TW (Tumour Windowing), (b) FE (Feature Extraction), (c) TW+FE (feature extraction performed after tumour windowing). "A.Grouping" refers to the use of the antenna grouping algorithm (using all available channels towards the majority vote). The solid lines correspond to the diagnostic performance when antenna grouping was not applied, and the dashed lines when antenna grouping was applied. The vertical axis shows the diagnostic performance, reported in terms of AUC (Area Under the ROC Curve). The horizontal axis shows the channel angles used to build each model; for the MA and EAC models, the models contain all channel angles between 0° and the angle shown in the horizontal axis.

- With the TW pre-processing method (Fig. 4.11a), tumour windowing and time alignment of the signals have been performed; however, it is likely that the TW processing is not sufficient to completely neutralise the inherent differences from channels at different angles, especially considering that the inter-channel variability is likely to increase when noisy experimental or clinical data is used. One additional factor to consider with the TW processing is that knowledge of tumour location is fundamental, and localisation errors might also impact accurate time-alignment of tumour signals from different channel angles;
- In the FE and TW+FE pre-processed datasets (Fig. 4.11b and Fig. 4.11c, respectively), comparable behaviour is observed. Models classifying signals from the same channel angles perform better. In addition, the dataset pre-processed only with FE, which does not require previous knowledge of the tumour location, slightly outperforms the TW+FE pre-processed dataset.

It is also interesting to note that channel angles below 90° in the EA models lead to higher diagnostic performance when compared to channels at higher angles, which might indicate that signals from reflection channels keep more information about tumour shape than signals from transmission signals. EAC models seem to benefit from this; when combining information from individual EA models, the predictions made by the EA models at lower channel angles dominate, ultimately contributing to the disregarding of incorrect predictions made at higher channel angles. Regardless of the pre-processing method, the best results seem to be achieved with EAC 0° – 30°.

In summary, optimal diagnostic performance is achieved when EAC models were used, particularly when combining channels with reflected backscattered signals. Antenna grouping is needed to achieve one final diagnosis per scan, and it helps increase diagnostic performance of the system as it provides the means to disregard random incorrect predictions. Using all channels in the antenna grouping algorithm provides the best performance, although, perhaps not surprisingly, the most relevant information appears to be contained in the channels closest to the tumours.

4.4.2 Effect of tissue heterogeneity

Increasing tissue heterogeneity is a concern when designing platforms for the diagnosis of breast cancer based on microwave backscattered signals. As glandular and tumour tissues are both characterised by higher dielectric

properties, the response due to glandular clusters in the breast might sometimes be confused with the response of a tumour, causing an increased rate of false positives. In this study, the proposed windowing and time-alignment methodology is analysed to determine if it is sufficient to handle breast heterogeneity, and if the extraction of the above-mentioned features provides meaningful information.

Separate classification models were used to classify signals arising from breasts with different glandular contents. It should be noted that this construction is an ideal one, that might not be practically realisable in experimental or clinical conditions. In fact, there is a growing interest in the development of automated methods for quantifying the background tissue of the breast, as shown, for example, by the following publications [134], [193]–[197]. However, these ideal classifiers are helpful in breaking down the problem of classifying backscattered signals into multiple simpler tasks, allowing an understanding of optimal conditions for the operation of automated diagnosis platforms to be developed.

From Section 4.4.1, one of the best performing antenna topologies was that of the EAC model at 0° – 30°, using all available channels when performing antenna grouping, where the TW and FE processing methods performed the best among all considered tests. The effect of increasing tissue heterogeneity is shown in Fig. 4.12, by plotting ROC curves obtained for the TW dataset (Fig. 4.12a), and FE dataset (Fig. 4.12b) using the optimal antenna topology. Separate classification models were built to diagnose scans from breast models with 1% (blue line), 5% (orange line) and 27% (green line) of glandular tissue by volume.

The random forest classifier appears to be robust to tissue heterogeneity. The average performance across breast models with increasing glandular tissue content is comparable, when using either the TW or the FE methods during the pre-processing stage of the system. However, in an experimental or clinical setup, the performance of the TW pre-processed dataset is likely to decrease as tissue heterogeneity increases; noisier experimental backgrounds lead to an increased number of reflections, and the localisation of the tumour signature in the backscattered signal will be affected. Conversely, the extracted features are able to capture the differences between benign and malignant tumours, even with signals recorded in more heterogeneous breast models.

Finally, the ROC curves indicate that diagnostic performance may also be optimised by varying the decision threshold. The range of optimal decision thresholds investigated in this study range between 0.36 and 0.52 (not shown in Fig. 4.12 for conciseness).

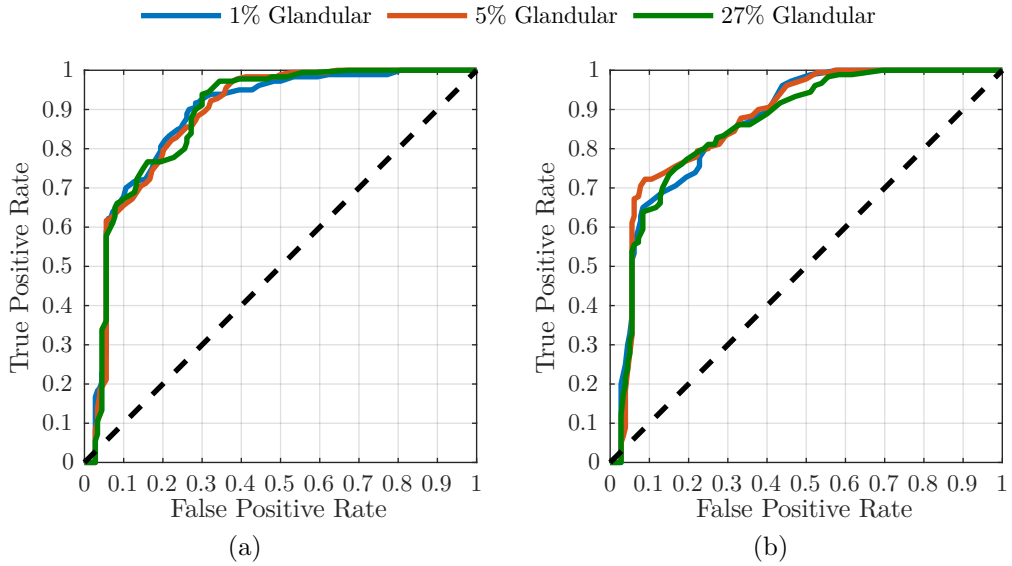


Figure 4.12: ROC curves showing diagnostic performance with the EAC (Equal Angle Combined) 0° – 30° classification model using all channels in antenna grouping, for (a) TW (Tumour Windowing) dataset, (b) FE (Feature Extraction) dataset. The blue line corresponds to the performance of a breast model with 1% glandular tissue by volume, orange line 5%, and green line 27%. The black dotted line represents the null hypothesis in the ROC curve.

4.4.3 Relative feature contribution

Previous sections examined the effect of antenna grouping, and the impact of tissue heterogeneity on the best performing system from initial baseline tests. In this section, an analysis of feature selection is presented, by means of the relative feature contribution map provided as one of the outputs of the random forest classifier. Investigating which features mostly contribute to the training of each tree inside a random forest could help refine the classification models, increasing their performance in complex scenarios, such as in experimental systems prone to high noise levels.

In Fig. 4.13a and Fig. 4.13b, the relative feature contribution map is shown, for the breast model with 27% glandular tissue, for the TW and FE pre-processed datasets respectively. Classification was performed with the EAC 0° – 30° model, which uses all channels in the antenna grouping algorithm. This model is shown as an example, although similar feature contributions were observed across all breast and classification models.

Firstly, Fig. 4.13a (TW) shows that the classification model using tumour windowing is heavily reliant on one single feature. This feature is time sample 34 in the example shown. All classification models used in this study display

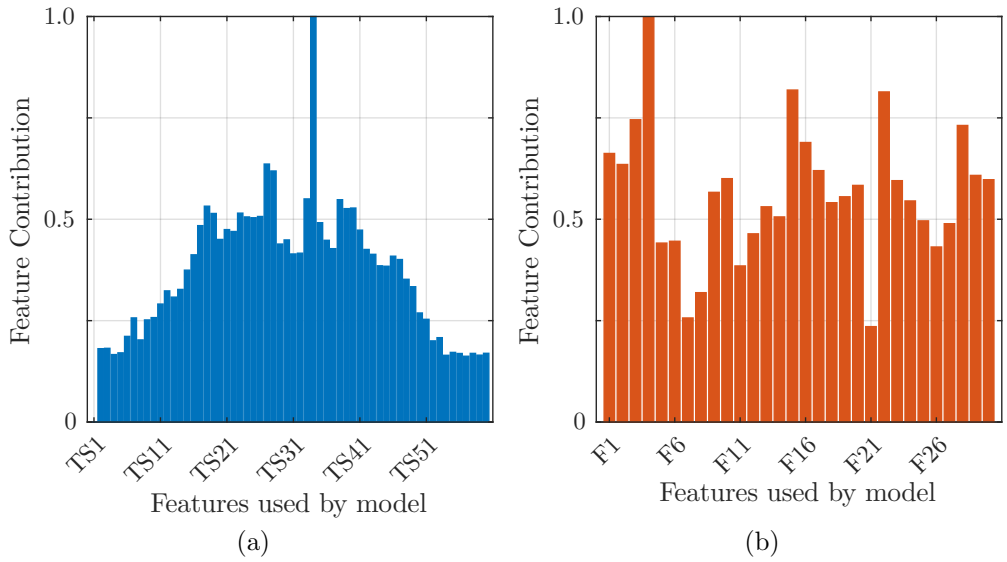


Figure 4.13: Map of relative feature contribution calculated during the training of the random forest model for the breast model with 27% glandular content by volume. The EAC (Equal Angle Combined) 0° – 30° model, with all channels in antenna grouping, is used. (a) refers to the dataset pre-processed with TW (Tumour Windowing) method; the horizontal axis shows the time samples (TS) which make up the time-domain tumour responses, (b) refers to the dataset pre-processed with the FE (Feature Extraction) method, where F1 to F30 shown in the horizontal axis correspond to features 1 through 30.

the same reliance on one feature, which varies between time sample 33 and 36. This result suggests that any errors in the tumour windowing and alignment algorithm could indeed have a large impact in the performance of the classification.

Regarding classification with the FE method (Fig. 4.13b), a larger number of features appear to contribute to the performance of the random forest model. Feature #4 (location of the maximum negative peak in the tumour signature) ranks highest, which is visible across all classification and breast models used in this study. Nine other extracted features are also identified as being particularly important in the training of the classification models. In decreasing order of contribution: number of zero crossings of the tumour signature (feature 15 shown in Fig. 4.13b), mean amplitude of the peaks in the autocorrelation curve (#22), amplitude of the maximum negative peak in the tumour signature (#3), mean amplitude of the peaks of the periodogram (#28), integral of the tumour signature (#16), amplitude of maximum positive peak in the tumour signature (#1), location of maximum positive peak in the tumour signature (#2), integral of the absolute value of

the tumour signature (#17), mean FWHM of the peaks of the periodogram (#29).

The features listed above are some of the features that mostly contributed to the classification of shape and spiculation of benign and malignant tumours. Particularly, the contribution of the features derived from the autocorrelation and power spectral density analysis should be noted, which reflect information otherwise not available in the time-domain tumour signatures.

4.5 Discussion

With the results presented in this chapter, antenna grouping was identified as a key step for an increased diagnostic performance. Individual signals which compose one breast scan are independent observations, and are classified accordingly, receiving a label of benign or malignant. By performing antenna grouping, those individual predictions are grouped into one final diagnosis (majority vote) for each scan. The results of this study showed the benefit behind this approach, as, by doing so, a mechanism is created to disregard minor incorrect predictions. In addition, results showed that a relatively small number of antennas closest to the tumour are needed to ensure the correct prediction (in the case of this study, 3 channels closest to the tumour were found to yield correct diagnosis). However, using the information from all channels results in similar performances compared to using only the channels closest to the tumour.

The results also showed how signals from different channel angles have to be appropriately used by the classification model. By building individual classification models which only classify signals from channels with the same angle, diagnostic performance is increased. The predictions of individual classification models can later be combined into a fused-type model, which once again contributes to increasing the diagnostic performance. In addition, the results also showed that signals from reflection channels lead to better performance than signals from transmission channels. In this study, the optimal channel angle was 0° – 30°.

Data pre-processing was also shown to have an impact on diagnostic performance. When dealing with time-domain signals, knowledge of tumour location is required. With this information, a tumour windowing and time-alignment algorithm can be implemented to isolate the tumour response, while decreasing the influence of the background. A new set of 30 features was also investigated, which are extracted per backscattered signal; these features mostly rely on peak analysis of the time-domain signal, and of the frequency content of the signal. Both methods performed comparably,

however, in practice, additional factors come into play which will impact performance:

- With currently available algorithms, localisation of the tumour signature in the response could be prone to error, which would impact the performance of the tumour windowing and could ultimately decrease diagnostic performance. This factor should not be neglected when building systems to classify time-domain signals.
- The extraction of features performed well, even when the time-domain signal was not pre-processed by windowing. This method appears as an alternative when exact tumour location is not available to the user. In addition, a reduced set of features of maximised contribution was identified, which could lead the way into finding an optimal set of features towards more robust classification models for microwave systems.

Finally, good machine learning practice is extremely important when designing microwave breast diagnosis systems. Without adequate feature processing methods and model validation strategies, reports of performance could be overly optimistic and not reproducible, ultimately impeding clinical acceptance of microwave diagnosis tools.

4.6 Conclusion

In this chapter, a comprehensive analysis of a machine learning platform applied to the field of microwave breast systems was presented, specifically with the goal of diagnosing the malignancy of breast tumours. A set of goals was identified as potential enablers to the operation of machine learning systems, which had not been thoroughly assessed in the literature up until now. To achieve the aim of the study, certain simplified assumptions were undertaken, so the true potential of the automated platforms can be assessed when no other confounding factors are present, such as the inherent complexity of experimental systems.

The primary conclusions from this chapter primarily indicate that:

- Feature extraction may be more flexible in real systems as it does not depend on accurate localisation of backscattered signals;
- Feature extraction may be able to retrieve meaningful information about the difference in the backscattered signals of benign and malignant tumours; accurate knowledge of the tumour location may thus not be an essential step to classification process;

- Careful design of the classification models in terms of antenna topology is essential to diagnostic performance, even in the simplified scenarios; particularly, classification models work best when dealing only with signals from the same channel angle.

Further investigations are needed to assess the robustness of microwave breast diagnosis systems given the complexities of experimental systems. In the following chapter, the same machine learning diagnosis platform is used to classify experimental data obtained using laboratory phantoms from the BRIGID phantom set presented in Section 3.3, in order to further validate the outcome of the investigations using the numerical models.

CHAPTER 5

Machine Learning Applied to Microwave Breast Diagnosis: Validation with an Experimental Dataset

5.1 Introduction

The previous chapter presented a numerical study on the potential of using machine learning to classify breast tumours as benign or malignant, based on the differences in the shape and level of spiculation captured in the backscattered signals. The numerical analysis in Chapter 4 was intended to provide a thorough investigation of the potential of microwave breast diagnosis systems based on machine learning in a controlled situation, when no real-life confounding factors are present. The analysis was based on a numerical dataset of breast and tumour models, which was built to represent clinical scenarios. The results from Chapter 4 indicate that classifying backscattered signals to diagnose tumours shows promise, but the study also revealed the importance of carefully designing appropriate classification systems.

The results from Chapter 4 mainly show that the performance of classification models may be impacted by the variance within the signals collected by different channels; a machine learning design including multiple "Equal Angle" models (where each model only classifies signals collected by channels with the same channel angle), was shown to be helpful in dealing with the inter-channel variance. Additionally, the results from the previous chapter indicate that extracting features may lead to similar diagnosis performances when compared to other signal pre-processing algorithms, such as tumour windowing, without the added complexity required to determine the tumour location. Finally, results from the previous chapter also showed that careful

validation of machine classification models is key in obtaining reliable and reproducible results.

In this chapter, the findings of the numerical study are further explored by means of an experimental dataset — the BRIGID phantom set, which was described in Section 3.3. Testing with experimental datasets allows further analysis of the impact of additional sources of variability and noise that arise in real-world conditions. The analysis based on experimental phantoms presented in this chapter is not intended to quantitatively validate the results obtained with the numerical models. Rather, the analysis in Chapter 4 provided a baseline set of operating conditions for microwave breast diagnosis systems based on machine learning; in this chapter, the BRIGID phantom set is now used to further verify the potential of microwave breast diagnosis systems based on machine learning, but with the additional factor of testing in more realistic conditions.

To date, experimental microwave diagnosis studies found in the literature used a variety of shapes to represent benign and tumour models, but did not account for the content of glandular tissue in the female breast [51], [52]. This gap in the literature is addressed in this chapter by analysing a database of backscattered signals collected when the BRIGID tumour phantoms were combined with all 7 BRIGID breast phantoms, which vary in the content and distribution of glandular tissue. An in-house experimental microwave prototype was used for signal acquisition, as described in Section 5.2.

Section 5.3 addresses some of the major assumptions with current approaches to automated microwave diagnosis platforms:

1) A set of sample signals and radar images produced from the BRIGID phantoms is presented in Section 5.3.1. The analysis of the signals and radar images justifies both the applicability of the BRIGID phantom set to the classification problem, as well as the diagnostic value of backscattered signals;
2) A suitable design for the operation of classification models is discussed from a theoretical perspective in Section 5.3.2, and justifies the continued use of channel angle to produce high-similarity groups of signals.

The methodology used in this experimental analysis is then discussed in Section 5.4; the methodology is based on the core principles described in Chapter 4, and includes data pre-processing by tumour windowing and feature extraction, random forests for classification, and antenna grouping for decision making. The results are listed in Section 5.5, while Section 5.6 discusses the results and concludes the chapter.

The work presented in this chapter will be captured in a journal paper titled "Experimental Validation of a Machine Learning Platform for Diagnosing Breast Cancer with Microwave Technology" (currently in preparation). This work was completed in collaboration with Declan O'Loughlin, who led the hardware development of the experimental prototype system described in Section 5.2.

5.2 Data Acquisition

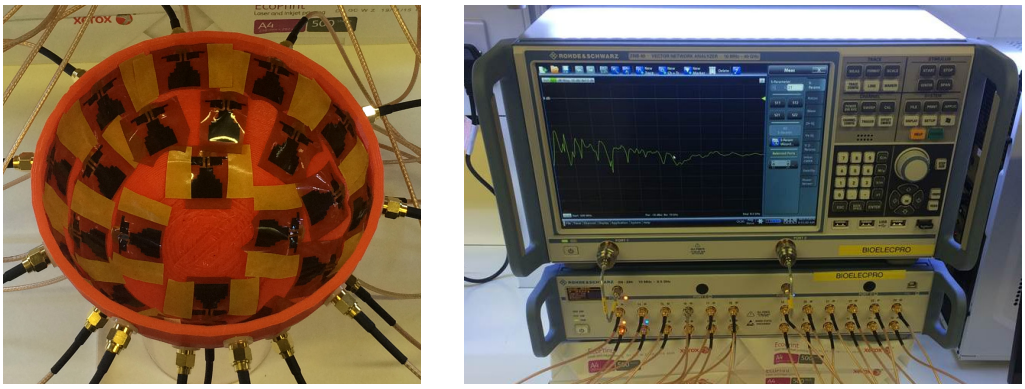
The BRIGID breast and tumour phantoms were used in this study to assess the potential of automated microwave breast diagnosis systems. Backscattered signals were acquired with an in-house experimental microwave prototype [198].

The review of the operational systems presented in Section 2.3.4 informed the design of the experimental system used in this study. The choice of design matches that of the MARIA® system [27]–[31], where the patient lies prone on an examination table with the breast pendant through an opening. With this choice of design, a coupling medium is necessary for patients with breasts of smaller volumes, and the coverage of the axillary tail of the breast and the chest wall is limited.

The system designed for this study features a hemispherical radome, which houses the antennas; as the size of the BRIGID phantoms matches the size of the system radome, no immersion medium is required for data acquisition in this study. A photograph of the radome and antenna array is show in Fig. 5.1a.

The hemispherical radome has a diameter of 14 cm and was manufactured using fused deposition modelling. The radome was printed with the same setup as described in Section 3.3.4, using PLA filament with the Ultimaker 2+ Extended (Ultimaker, Geldermasen, the Netherlands). The radome has 24 openings which securely house the connectors attached to each antenna. The antennas used in this system are flexible microstrip antennas [199] from the wearable system designed by McGill University, Canada [105] (described in Section 2.3.4). The antennas were optimised to operate from 2 GHz to 4 GHz. The antennas are 2 cm by 2 cm square and 24 antennas are located inside the radome for a total of 276 independent transmit-receive channels.

For analysis purposes, the antennas can be seen as belonging to one of three rings of antennas that are equally spaced in the front-to-back plane. Given the hemispherical shape of the radome, the propagation paths for channels within each ring are larger in the ring close to the chest ("top ring"), and smaller in the ring close to the nipple ("bottom ring").



(a) (b)

Figure 5.1: In-house prototype of a microwave breast system used for signal acquisition. In (a), the radome and antenna array are shown. In (b) the ZNB40 2-port VNA and ZN-Z84 24-port switching matrix (Rohde and Schwartz GmbH, Munich, Germany) are shown.

Data were acquired in the frequency domain using a stepped frequency sine-wave at 50 frequency points linearly spaced between 2 GHz and 4 GHz. A ZNB40 2-port VNA and ZN-Z84 24-port switching matrix (Rohde and Schwartz GmbH, Munich, Germany) acquired all 276 independent multistatic channels in 30 s (shown in Fig. 5.1b). Twenty-four coaxial cables 457 mm in length connected each port of the switching matrix to surface-mounted connectors on each antenna, which were housed in the radome (cables and connectors both manufactured by Cinch Connectivity Solutions, Waseca, MN, USA).

Prior to use, a reference scan with BRIGID 0 (homogeneous breast phantom with a skin layer) was taken. This reference scan was used to compensate for differences between the antennas, and compensation was applied to all scans before artefact removal and imaging.

5.3 Design of Microwave Breast Diagnosis Systems

In this section, the design choices assumed in the numerical study — value of backscattered signals, and use of channel angle to produce high-similarity groups of signals — are revisited in the context of data acquired from experimental phantoms instead of numerical models. The use of the data from the BRIGID phantom set for evaluation purposes is addressed in

Section 5.3.1, followed by a discussion of the suitability of classification models based on channel angle in Section 5.3.2.

5.3.1 Diagnostic Value of Backscattered Signals

Section 2.4 discussed various approaches that could be explored to distinguish breast tumours according to their level of malignancy: dielectric properties, presence of microcalcifications, or shape of tumours. The small number of studies characterising the dielectric properties of benign and malignant tumours and microcalcifications in the microwave range limits the viability of such approaches and fosters interest in alternatives. Additionally, the shape and spiculation of tumours are widely recognised markers for their malignancy, and this information was shown to be preserved within the backscattered signals (as demonstrated with numerical simulations in Chapter 4). In this section, a brief analysis is presented of both backscattered signals and radar images reconstructed from the BRIGID phantom set, to demonstrate that the signals produced using the BRIGID phantom set preserve similar diagnostic information to the numerical models, and may therefore be suitable to experimentally test diagnosis platforms.

Tumours with different levels of spiculation embedded in breasts with different proportions of glandular tissue are likely to result in backscattered signals with different morphologies. To verify this effect, tumour plugs from the BRIGID phantom set were added to the BRIGID breast phantoms in different combinations, and signals were acquired using the system described in Section 5.2. All recorded signals were pre-processed by means of rotational artefact removal as described in [125], [130], [200]. By way of example, Fig. 5.2 shows the backscattered signals from breast phantoms BRIGID 0, BRIGID 10E, BRIGID 20E, when combined with tumour models with increasing level of spiculation: L4 tumour phantom, of low spiculation and approximately spherical shape; I12 phantom with intermediate spiculation; and H15 phantom with high spiculation:

- The top image in Fig. 5.2 shows the result of combining the 3 tumour phantoms in turn with the BRIGID 0 model. This breast model is fully homogeneous, and therefore represents the simplest case with a near-ideal tumour response, i.e., when sources of clutter due to glandular tissue are not present and the reflections are largely determined by the tumours. Here, it can be observed that tumours with different levels of spiculation result in backscattered signals with different amplitudes and phases;

- The remaining images in Fig. 5.2 show the reflections when the same tumour plugs are combined with the other breast models with increasing percentages of background glandular tissue. Here, the differences in the shape and amplitude of the reflections are indicative of how the delay of propagation through several heterogeneous background media also impacts the amplitude and phase of backscattered signals.

These observations suggest that the morphology of signals collected during a microwave scan depends on the shape and level of spiculation of a tumour. In addition, the interaction between the tumour model and the surrounding fat and glandular tissues also influences the shape of the reflection. These results are in line with the points discussed in Section 4.3.2.2 regarding the extraction of features based on the morphology and frequency content of the backscattered signals.

Additionally, images were reconstructed using the multistatic the Delay-And-Sum algorithm [125], [126], [180], [201], [202]. By way of example, reconstructed radar images for the same tumour plugs combined with the BRIGID 10E and BRIGID 20E breast phantoms are shown in Fig. 5.3:

- In all images, the maximum response of the image is within the tumour area;
- Increased clutter can be seen in images from the breast phantom with more glandular tissue (BRIGID 20E);
- Quantitatively, the signal-to-clutter ratio of the images from BRIGID 20E is, on average, 3 dB lower than images reconstructed with the same tumour models in a phantom with less glandular tissue (BRIGID 10E);
- The maximum intensity varies significantly, both between tumour models and between phantoms;
- Images from BRIGID 20E are between 5–8 dB lower in intensity than those from BRIGID 10E;
- Additionally, images using the L4 tumour model are up to 3 dB lower in intensity than those using the H15 tumour model.

Despite the differences in intensity and signal-to-clutter ratio, the images of all three tumour models in each breast model are visually similar. With the current state-of-the-art in image reconstruction, radar-based imaging is not expected to recover the shape information of the targets, with the level of resolution required for classification.

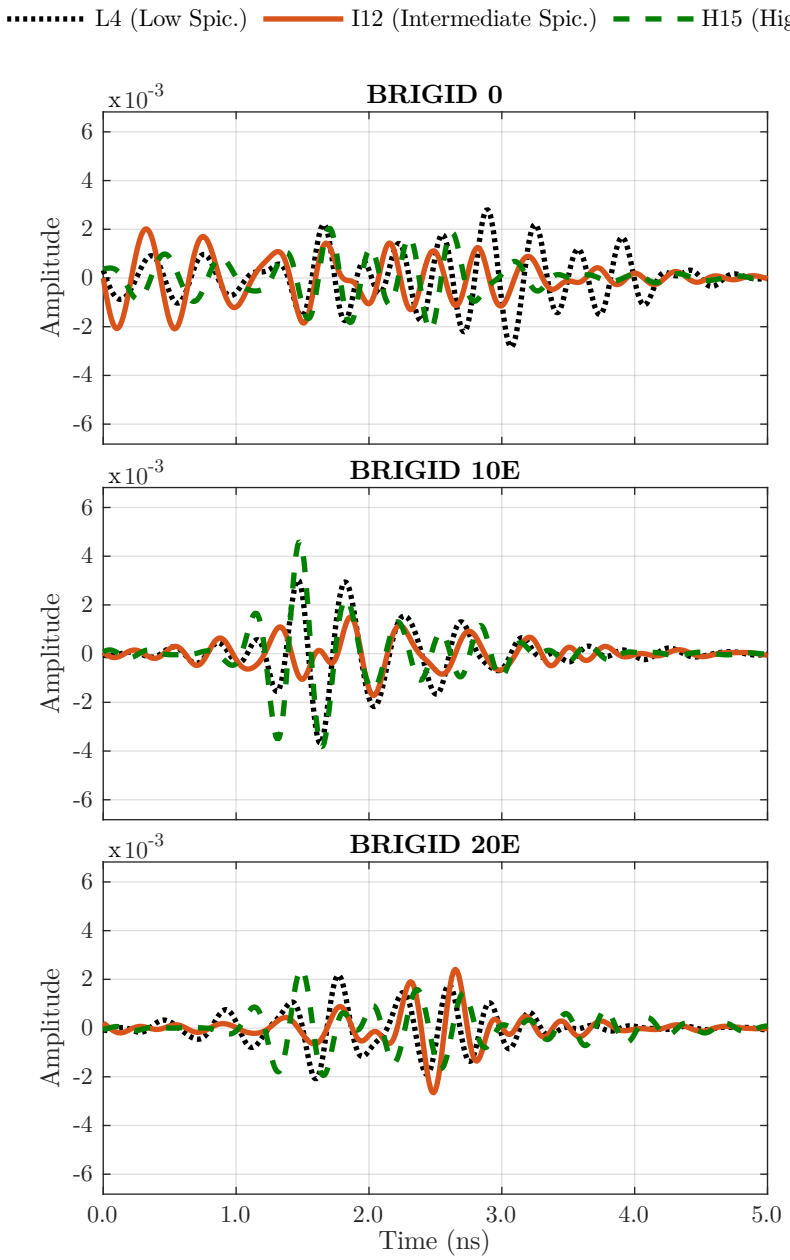


Figure 5.2: Sample signals collected with the Breast Imaging and Diagnosis (BRIGID) phantom set with tumour plugs L4 (low spiculation, dotted black line), I12 (intermediate spiculation, orange line), H15 (high spiculation, dashed green line). From top to bottom, the tumour plugs were combined with the homogeneous breast model BRIGID 0, and heterogeneous models BRIGID 10E, BRIGID 20E.

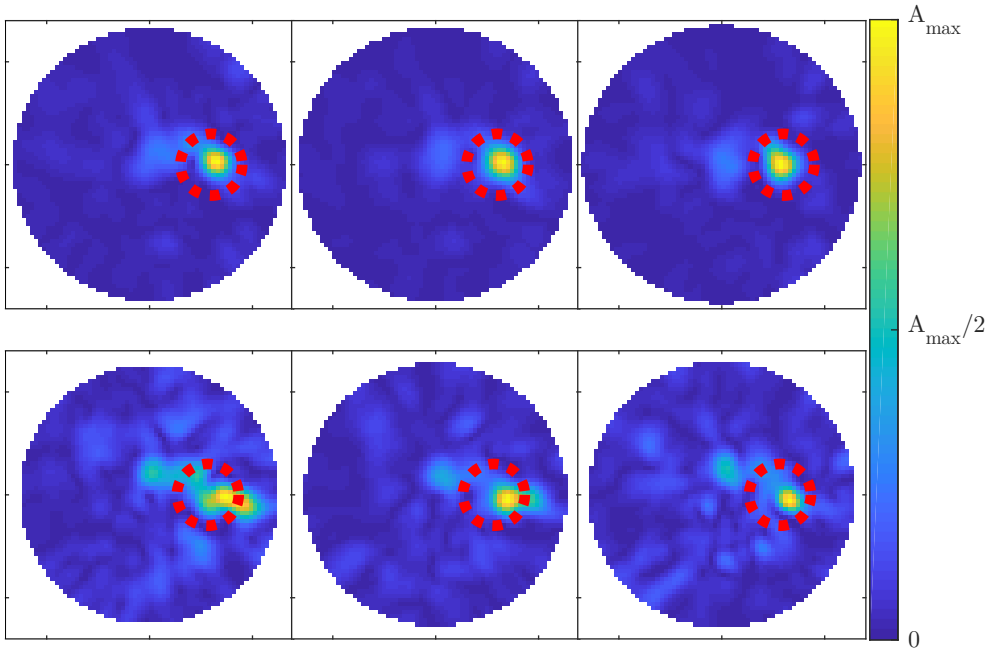


Figure 5.3: Representative backscattered energy plots for example Breast Imaging and Diagnosis (BRIGID) phantoms: BRIGID 10E phantom (top row), BRIGID 20E phantom (bottom row). Both phantoms were combined with the L4 (low spiculation, image on the left), I12 (intermediate spiculation, image in the centre) and H15 (high spiculation, image on the right) tumour plugs. The actual tumour location is at the centre of the red circle in each image. For visualisation purposes, the colourbar is such that A_{max} represents the maximum intensity of each individual image.

5.3.2 Design of classification models

Section 4.2 of the previous chapter discussed in detail many of the challenges in classifying the level of malignancy of tumours based on the shape information contained in backscattered signals. Although it was demonstrated that there may be sufficient differences in the backscattered signals arising from benign and from malignant tumours, the variance from signals captured from the same tumour by the many available antennas in a system was shown to be not negligible; indeed, in extreme cases, the inter-channel variance could exceed the differences between benign and malignant tumours. Additionally, one other source of variance in the backscattered signals that should not be overlooked is due to the varying contents of background glandular tissue that may, at times, shadow the tumour response. Fundamentally, unaddressed sources of variance in the backscattered signals could hinder the usability of microwave diagnosis systems based on classification models.

Ultimately, a classification model will perform better when the differences in the data to be classified arise primarily from the differences between classes (in this study, benign and malignant), which would require that the following are examined prior to classification: 1) reducing the influence of the glandular content, and 2) addressing the inter-channel variance. Artefact removal and tumour windowing are two of the primary algorithms used to decrease the influence of the glandular content in the backscattered signal, while preserving the tumour signature. However, the variance in backscattered signals arising from different channels is not often discussed in microwave studies using machine learning [42]–[45], [48]–[53], [140]–[142].

The issue of inter-channel variance can be addressed at a classification level, i.e., when designing the classification models. In the previous chapter, different classification models for each channel angle were empirically chosen to group similar signals; this approach resulted in an increased diagnostic performance when compared to models that learn and classify all signals simultaneously. Here, the assumption that grouping by channel angle results in high similarity signals is further explored from a theoretical standpoint. Intuitively, using the length of propagation path to a target would be one of the most logical approaches to group signals, and is worthy of discussion.

Separating backscattered signals by the length of propagation path to a target is likely to produce high-similarity groups of signals, since the attenuation along each path significantly affects the amplitude and shape of the response: shorter propagation paths are expected to result in high-amplitude signals that preserve the shape of the tumour response, while longer propagation paths may result in signals with diminished energy. However, the operation of such an approach is computationally expensive as it would require *a priori* knowledge of the tumour location to a high degree of accuracy, which must usually be obtained through image reconstruction.

In a dataset of signals collected in a microwave scan, different channels are expected to generate signals with different propagation paths mostly depending on:

1) Channel angle, where a distinction is made between:
    i) "Reflection channels", meaning channels where the transmit and receive antennas are located on the same side of the target (thus originating signals with reflected information);
    ii) "Transmission channels", meaning channels where the transmit and receive antennas are located on opposite sides of the tumour (thus producing "through-channel" signals);
2) Proximity between the channel and the tumour.

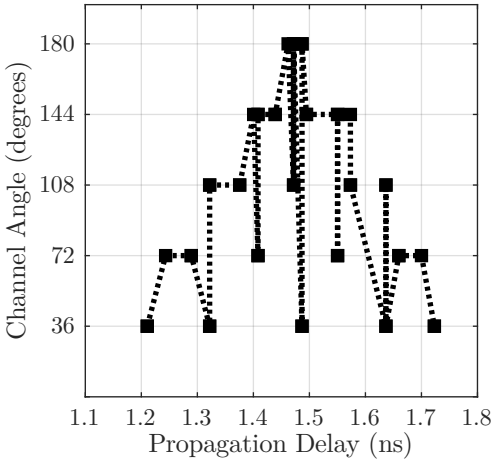


Figure 5.4: Distribution of channel angles with increasing propagation paths, shown for signals acquired with the Breast Imaging and Diagnosis (BRIGID) phantom set. The theoretical propagation delays were calculated when tumour plug L4 (low spiculation) is embedded in breast model BRIGID 0, assuming a relative permittivity of 9.0 at a frequency of 3 GHz. Propagation paths for channels in the top ring of antennas are shown by way of example. The reflection channels with shorter distances to the target have shorter propagation delays; all the transmission channels are next; finally, the reflection channels with the longest distances to the target have the longest propagation delays.

To verify the distribution of channels per propagation path in the in-house microwave prototype system, the theoretical propagation delays were calculated for BRIGID 0 assuming a relative permittivity of 9.0 at a frequency of 3 GHz (as detailed in Table 3.8). By way of example, Fig. 5.4 shows the propagation paths at each channel when BRIGID 0 is combined with tumour plug L4 (low spiculation); in this figure, the propagation paths for the top ring of antennas of the microwave prototype are the only on display. The general trend displayed in the plot of Fig. 5.4 shows that the shortest propagation delays correspond to the reflection channels where transmit and receive antennas are close to the target ("near reflection channels"), followed by the transmission channels, and finally by the reflection channels with a longer distance to the target ("far reflection channels"). Additionally, reflected and transmitted channels naturally exhibit different information about the propagation of signals, as the view from the tumour in a reflection or transmission channel is different.

To further aid in the understanding of the types of "channel", Fig. 5.5 shows a simplified representation of the near reflection, transmission and far reflection channels in the top ring of antennas of the microwave prototype used in this study (represented in blue, orange and green, respectively). In

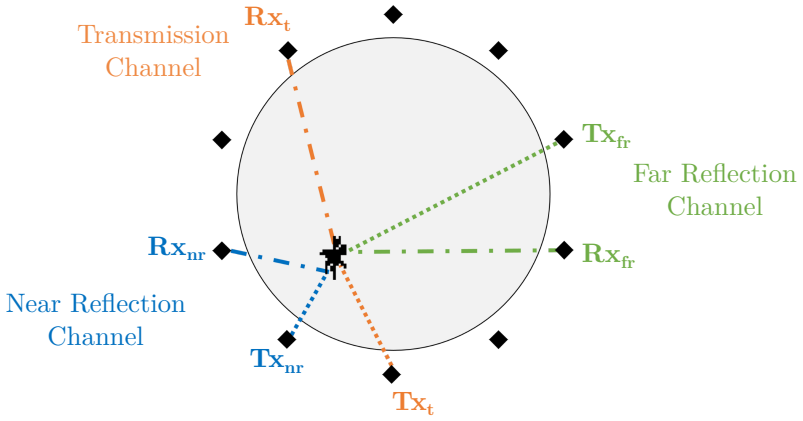


Figure 5.5: Simplified representation of the channels in a microwave scan. The top view of the microwave prototype used in this system is shown, highlighting the channels in the top ring of antennas. The antennas are represented by the black diamonds surrounding the breast; the breast interior is shown in grey with a malignant tumour in one of the lower breast quadrants (represented by the black spiculated shape). Antenna pair (Tx$_{nr}$, Rx$_{nr}$) displayed in blue is an example of a "near reflection channel"; antenna pair (Tx$_t$, Rx$_t$) displayed in orange represents a "transmission channel"; and antenna pair (Tx$_{fr}$, Rx$_{fr}$) shown in green represents a "far reflection channel".

this figure, the black diamond shapes represent the antennas.

Given the above explanation, grouping signals in different classification models by length of propagation path effectively corresponds to grouping signals in three zones of angles: 1) near reflected channels, 2) transmission channels, and 3) far reflection channels. Additionally, when separating signals by angle, each classification model is provided with a group of signals incorporating a full view surrounding the tumour. Using the channel angle information available in a microwave prototype is a simple and intuitive way of grouping signals; as demonstrated in Chapter 4, classification based on separate models for each channel angle increased the diagnostic performance.

In terms of expected performance, it is also plausible to assume that the quality of information preserved in signals with increasing propagation path may decrease, as the attenuation along the path has a higher effect; particularly, the far reflection channels are expected to underperform.

5.4 Methodology

The experimental analysis presented in this chapter follows the same core methodology designed for the numerical study presented in Chapter 4, consisting of three stages: data acquisition, data processing and diagnosis.

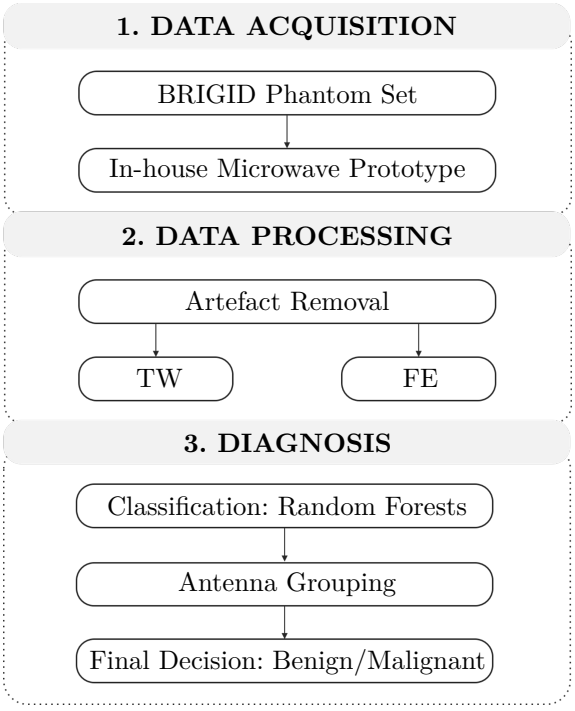


Figure 5.6: Updated 3-stage diagnosis platform. Stage 1 consists of data collection in a microwave breast prototype, using the Breast Imaging and Diagnosis (BRIGID) phantom set. Stage 2 consists of data processing by means of artefact removal, followed by Tumour Windowing (TW) or Feature Extraction (FE). Stage 3 is the diagnosis stage, which uses random forests as the classifier and includes an antenna grouping algorithm to produce a final diagnosis as benign or malignant.

The main findings from the previous chapter are used to refine the diagnostic platform, which is depicted in Fig. 5.6. The updated architecture is intended to reduce the overall operational cost and complexity of the diagnosis system.

Stage 1 consists of the signal acquisition for the BRIGID phantom set with the in-house experimental prototype, as described in Section 5.2. Particularly, the dataset analysed in this study is based on the low spiculation tumour phantoms (plugs 1–8, described in Table 3.6) to represent benign tumours, and the high spiculation tumour phantoms (plugs 15–22) to represent malignant tumours. This choice of input data enables the creation of a classifier with equally distributed observations amongst the two class labels, which simplifies the operation of the classification models. The tumour phantoms were each combined with all 7 BRIGID phantoms, totalling 112 test cases. Contrary to Section 4.4, where breast models with different glandular content where classified separately, there was no separation by breast model prior to the classification in the current study. This choice

of input data allows an analysis as to whether the classification models are able to discern which differences in the signals arise from the level of malignancy of a tumour and which differences are due to reflections from increased glandular content.

Stage 2 consists of data processing, which starts by performing artefact removal with the rotational subtraction algorithm [125], [130]; next, the signals are pre-processed by tumour windowing or feature extraction. The tumour windowing algorithm was described in Section 4.3.2.1; the operation of the tumour windowing requires prior knowledge of the tumour location, which is available as described in Table 3.7, thus the windowing can be considered to be ideal. The feature extraction algorithm was described in Section 4.3.2.2. The relative benefits of both algorithms are once again analysed by comparing the diagnostic performance when only tumour windowing is applied, and only feature extraction is applied. The discussion presented in Section 4.5 showed how the extraction of features from backscattered signals performed well, even when the signal was not pre-processed by tumour windowing; performing feature extraction has the added benefit of not requiring prior knowledge of the tumour location, reducing the complexity of the system.

Stage 3 consists of diagnosis, and encompasses classification of the dataset through random forests, followed by antenna grouping and final decision as benign or malignant. The random forest algorithm was previously described in Section 4.3.3.1. The same antenna grouping algorithm described in Section 4.3.3.3 is once again utilised; Section 4.5 discussed how antenna grouping is important for increased diagnostic performance, as it allows for minor incorrect predictions from extraneous channels to be disregarded.

The choice of design for the random forest models is based on the discussion presented in Section 5.3.2. Individual classification models are built for the signals collected at each channel angle, within each ring of antennas (EA models, as defined in Section 4.3.3.2). Additionally, in Section 4.5, the diagnostic performance was shown to increase when the predictions of individual EA models were combined by majority voting into a fused-type model (EAC models). The relative benefit of EA and EAC models is assessed by comparing their diagnostic performance with the performance of a random forest model that classifies signals from all channels without regard for their angle — these models are defined as All Channels (AC) for the remainder of this chapter. Figure 5.7 summarises the operation of the three types of model design assessed in this chapter: EA, EAC, and AC.

Finally, the same validation practices discussed in Section 4.3.3.4 were adhered to in order to increase the reliability of the results, and prevent overly optimistic reports of model performance.

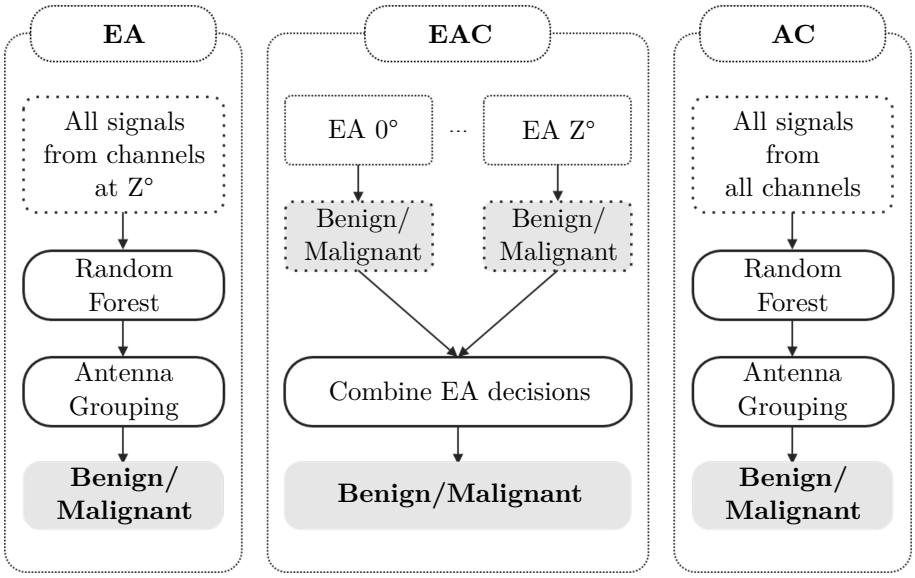


Figure 5.7: Illustration of the different classification models implemented in this study: EA (Equal Angle), EAC (Equal Angle Combined) and AC (All Channels), where Z represents the channel angle in the microwave prototype. EA models classify signals from independent channel angles; EAC models are fused models that combine the predictions from individual EA models by majority voting to produce a diagnosis; and AC models classify signals from all channels together regardless of their angle.

5.5 Results

The results section is divided into three separate sub-sections. Section 5.5.1 presents an investigation on the performance of the classification models designed for the diagnosis of tumours based on their backscattered responses, the usefulness of the antenna grouping algorithm, and the relative advantages of data pre-processing by means of tumour windowing or feature extraction. Section 5.5.2 then investigates the effect of tissue heterogeneity in diagnostic performance, and Section 5.5.3 identifies a subset of features with maximised relative contribution to the learning process.

5.5.1 Design of classification models

In this section, the aim is to study if grouping signals by channel angle ahead of the classification is indeed helpful towards increasing the performance of microwave breast diagnosis systems. The value of antenna grouping is also re-visited. To this extent, classification models of type AC and EA were used to classify the BRIGID dataset, using the tumour windowing and

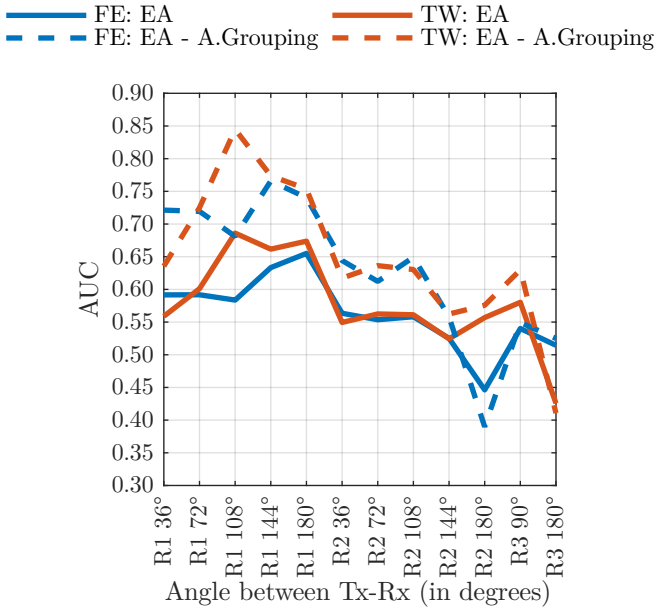


Figure 5.8: Diagnostic performance, plotted by means of the AUC (Area Under the ROC curve), when EA (Equal Angle) models are used to classify the dataset after FE (Feature Extraction, blue lines) and TW (Tumour Windowing, orange lines). The performance of 12 different EA models is compared, each using one channel angle within each ring of antennas: R1 stands for the top ring, R2 the middle ring, and R3 the bottom ring of antennas. Performance is also compared when the antenna grouping algorithm is used (A.Grouping, represented by the dashed lines), and when the antenna grouping is not applied (solid lines); here, the antenna grouping algorithm used all available channels towards the majority vote.

feature extraction methods.

Figure 5.8 plots the AUC for EA models using feature extraction (blue lines) and tumour windowing (orange lines). The solid lines and dashed lines in the plot indicate the AUC before and after applying antenna grouping, respectively.

The first noticeable observation is that antenna grouping greatly increases the performance of the classification, supporting the numerical results obtained in Section 4.4.1. By taking the majority vote of all individual decisions from one single breast scan, a minority of incorrect predictions are cancelled by a majority of correct classifications, which helps increase diagnostic performance.

Figure 5.8 also shows comparable diagnosis performances after tumour windowing and after feature extraction are applied. This result suggests that extracted features have the potential of capturing the essence of the tumour signature without requiring the tumour location to be known in

Table 5.1: Comparison of the diagnostic performance with the best performing EA (Equal Angle) model, the EAC (Equal Angle Combined) models for top, middle and bottom rings of antennas separately, the EAC model using all signals, and the AC (All Channels) model using all signals. Feature Extraction (FE) and Tumour Windowing (TW) were both used as pre-processing methods in the analysis.

Model Type	AUC [FE]	AUC [TW]
EA Best Model	0.77 (EA 108°)	0.84 (EA 144°)
EAC Top Ring	0.81	0.80
EAC Middle Ring	0.61	0.65
EAC Bottom Ring	0.55	0.50
EAC All Signals	0.73	0.74
AC All Signals	0.65	0.74

advance as in the ideal TW results.

In addition, a clear trend cannot be identified as to which channel angle yields the best diagnostic performance. In fact, Fig. 5.8 shows that EA models using channels from different antenna rings produce substantially different AUCs, which also appear to be highest for the channels within the top ring of antennas; the meaning of this result will be discussed below.

The value of EAC models, which fuse the individual predictions of EA models, is shown in Table 5.1. In this table, a comparison is drawn between the best EA model from Fig. 5.8, the EAC performances for the channels in each of the rings in the prototype, the EAC performance when using all signals, and the performance of using the AC model to classify all signals. In this analysis, antenna grouping was always used, as its value has already been demonstrated.

Once more, the diagnostic performance when using channels in the top ring seems to be the best amongst all models, with feature extraction or tumour windowing. However, and considering the clinical application of a microwave breast diagnosis platform, it is unrealistic to design a system that penalises a particular set of channels based on findings that are specific to a patient. When tumour location is not available *a priori*, the classification system should be designed in such a way that its performance does not depend on rewarding or penalising channels based on the channel-tumour distance. To this extent, the EAC model which uses all signals after feature extraction achieves a diagnostic performance that is 8% higher than the undifferentiated AC model.

For completeness, the ROC curves comparing the performance of the

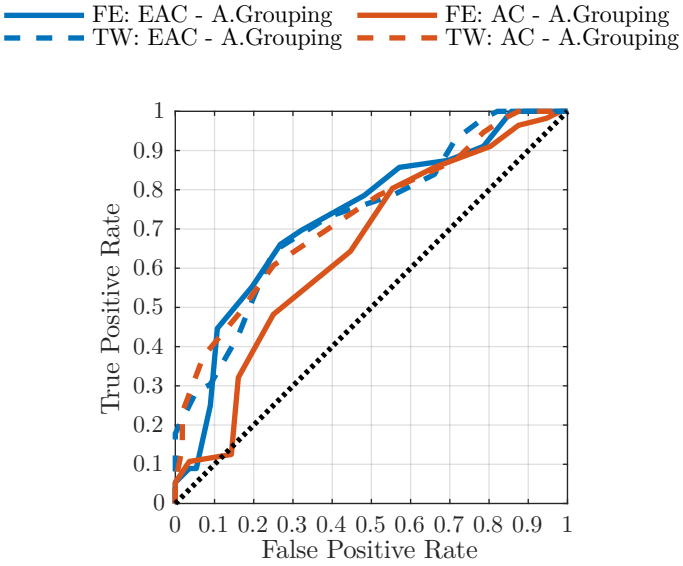


Figure 5.9: ROC curves comparing the diagnostic performance with the EAC (Equal Angle Combined, blue lines) and the AC (All Channels, orange lines) models, always applying the antenna grouping algorithm; pre-processing with FE (Feature Extraction) and TW (Tumour Windowing) are shown by the solid lines and dashed lines respectively. The black dotted line represents the null hypothesis in the ROC curve.

EAC (blue lines) and the AC (orange lines) models are shown in Fig. 5.9. Pre-processing with feature extraction and tumour windowing are shown by the solid lines and dashed lines respectively. As previously demonstrated, the added value of using the EAC models is noticeable, with the corresponding ROC curves exhibiting greater area. Additionally, the ROC curves for feature extraction and tumour windowing exhibit similar behaviour.

5.5.1.1 Performance Variations Across Channel Angles

In Fig. 5.8 and Table 5.1, an unexpected variation of AUC for EA models of different channel angles can be noted: the signals collected from channels in the top ring of antennas seemed to produce a higher diagnostic performance.

To understand this result, the relationship between diagnostic performance of EA models and propagation path length was inspected when each tumour model was combined with each of the BRIGID breast phantoms. By way of example, Fig. 5.10 details the analysis when tumour model H19 (high spiculation) was combined with BRIGID 0.

In this figure, diagnostic performance is expressed in terms of cumulative percentage of correct predictions, which was calculated for channels with

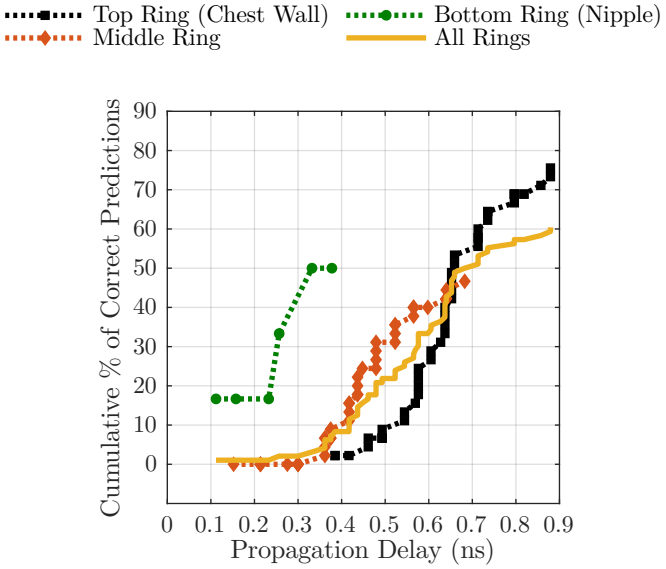


Figure 5.10: Diagnostic performance, expressed in terms of cumulative percentage of correct predictions, for channels with increasing propagation path length, for signals acquired with the Breast Imaging and Diagnosis (BRIGID) phantom set. The result is shown for tumour model H19 (high spiculation) combined with BRIGID 0. The predictions are presented for each ring of antennas individually: the top ring is shown in black, the middle ring in orange, and the bottom ring in green; additionally, the cumulative performance including all rings is shown in yellow for completeness. The middle ring of antennas suffers from the problem of tumour depth, where the antennas closest to tumours in the vicinity of skin see a deterioration in performance. Overall, the top ring of antennas contributes with the most to performance.

increasing length of propagation path. To make the analysis easier, the cumulative predictions are shown for each ring of antennas individually. The top ring is shown in black, the middle ring in orange, and the bottom ring in green; for completeness, the cumulative performance including all rings is shown in yellow.

As expected, the propagation paths for channels in the top ring are the largest, and for the bottom ring the smallest; this is due to the hemispherical shape of the radome (shown in Section 5.2). However, the cumulative performance in each of the rings is of similar shape: there is a small increase in performance for the channels with the shortest propagation paths, followed by a steep increase for channels with mid-range propagation paths, followed finally by another slow increase in performance for the channels with the largest propagation path lengths. This observation is consistent with the hypothesis of Section 5.3.2, and shows that the channels with mid-range propagation delays are the ones that mostly contribute to the classification;

according to Fig. 5.4, these channels are likely to correspond to the reflection channels close to the tumour as well as the transmission channels.

Figure 5.10 also shows that the middle ring of antennas yields the lowest diagnostic performance. The explanation of this result is likely due to the positioning of the tumour models.

As detailed in Table 3.7, tumours in the BRIGID phantoms are always located at an approximately equal distance between the chest wall and the nipple (B-F plane), which coincides with the plane of the middle ring of antennas; given the offset in the tumour positioning in the H-T plane, tumours will appear very close to the skin for a subset of antennas in the middle ring. The quality of the signal when the tumour is close to the skin is expected to deteriorate due to poor performance of the artefact removal process in such cases, consequently affecting the quality of the predictions. Ultimately, antennas nearest tumour models in the vicinity of skin are problematic for the classification.

5.5.2 Effect of Tissue Heterogeneity

To mimic the limitations of a clinical system, the classifiers used in this study had no knowledge of the glandular tissue content of each breast scan. The analysis presented in this section investigates the effect of tissue heterogeneity in the diagnostic performance by investigating the number of correct predictions per breast phantom.

Table 5.2 details the breakdown of correct benign and malignant predictions of the BRIGID phantom set, when feature extraction and tumour windowing are used. For both processing methods, the benign tumours are more often correctly classified for six of the seven breast phantoms; this result confirms that, as implemented, the diagnosis system maintains specificity, which, as discussed in Section 2.2.4.2 is a drawback of current screening mammography. In the final breast phantom (BRIGID 30E, with the highest content of background tissue), the opposite happens: the malignant tumours are more often correctly classified. This result could be explained by the high glandular content of the phantom; increased glandular content leads to larger and more widespread reflections in the backscattered signals, and it is possible that the classification model incorrectly labels these reflections as originating from a malignant tumour. Overall, using tumour windowing as the pre-processing method leads to a 5.3% increase in correct predictions.

Table 5.2 also shows that there is a substantial variation in the specificity and sensitivity of the system across breast models, in some instances surpassing 50%. This result may in part be explained by the small size of the dataset used; for each breast model, a relatively small dataset of 8

Table 5.2: Breakdown of correct predictions in each breast model of the Breast Imaging and Diagnosis (BRIGID) phantom set, when FE (Feature Extraction) and TW (Tumour Windowing) are used. The correct predictions expressed in percentages are shown in brackets.

	Number of Correct Predictions			
	FE		TW	
	Benign (%)	Malignant (%)	Benign (%)	Malignant (%)
BRIGID 0	6/8 (75.0)	4/8 (50.0)	7/8 (87.5)	7/8 (87.5)
BRIGID 10U	6/8 (75.0)	2/8 (25.0)	7/8 (87.5)	1/8 (12.5)
BRIGID 10E	8/8 (100)	3/8 (37.5)	8/8 (100)	3/8 (37.5)
BRIGID 15U	7/8 (87.5)	5/8 (65.5)	7/8 (87.5)	6/8 (75.0)
BRIGID 15E	6/8 (75.0)	1/8 (12.5)	7/8 (87.5)	4/8 (50.0)
BRIGID 20E	8/8 (100)	5/8 (65.5)	8/8 (100)	2/8 (25.0)
BRIGID 30E	3/8 (37.5)	7/8 (87.5)	3/8 (37.5)	6/8 (75.0)
Total	44/56 (78.6)	27/56 (48.2)	47/56 (83.9)	29/56 (51.8)

benign and 8 malignant tumours was available, creating fluctuations in the performance analysis between breast models. It should also be considered that the signals collected from breast models with high glandular content may, in some cases, be dominated by the glandular response; it is likely that the diagnosis platform has an upper limit on the amount of glandular content for meaningful results.

5.5.3 Relative Feature Contribution

The previous sections have examined the design of classifiers, and the impact of tissue heterogeneity on the correct predictions across all BRIGID breast phantoms. In this section, an analysis of feature selection is presented by means of the relative feature contribution map provided as one of the outputs of the random forest classifier.

In Fig. 5.11a and Fig. 5.11b, the relative feature contribution map is shown when tumour windowing and feature extraction are used respectively. The relative feature contribution is expressed as the average feature importance map retrieved from the operation of each EA classifier.

Regarding classification with tumour windowing (Fig. 5.11a), the relevant features identified by the feature contribution map appear to concentrate on the entire tumour signature.

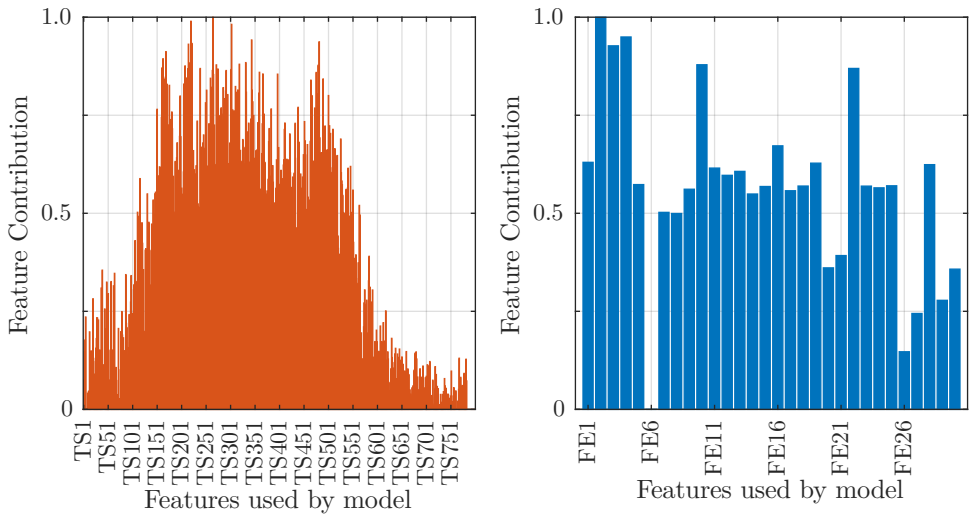


Figure 5.11: Map of relative feature contribution calculated during the training of the random forest model. (a) refers to the dataset pre-processed with TW (Tumour Windowing), and the horizontal axis shows the time samples (TS) which make up the time-domain tumour responses; (b) refers to the dataset pre-processed with FE (Feature Extraction); the horizontal axis shows features 1 through 30 used in the classification, indicated by F1 to F30.

Classification with the FE method seems to benefit from a large number of features (Fig. 5.11b). Feature #2 (Location of the maximum positive peak) ranks highest, but a number of other features were identified as relevant. The top-10 features that contributed the most to the operation of the random forests are: location of the maximum positive peak in the tumour signature (#2), location of the maximum negative peak in the tumour signature (#4), amplitude of the maximum negative peak in the tumour signature (#3), mean amplitude of the negative peaks in the tumour signature (#10), mean amplitude of the peaks in the autocorrelation curve (#22), integral of the tumour signature (#16), amplitude of the maximum positive peak in the tumour signature (#1), negative percentage area of the tumour signature (#19), mean amplitude of the peaks in the periodogram curve (#28), mean FWHM of the positive peaks in the tumour signature (#11).

The top-10 features were used to re-compute the classification results. The breakdown of the correct predictions per breast phantom when all 30 features are used are compared with results using the top-10 features in Table 5.3. The specificity of the system increased by 8.9%, which highlights

Table 5.3: Breakdown of correct predictions in each breast model of the Breast Imaging and Diagnosis (BRIGID) phantom set, when FE (Feature Extraction) is used, either with all 30 extracted features contributing to the classification, or only the top-10 best performing features. The correct predictions expressed in percentage are shown in brackets.

	Number of Correct Predictions			
	FE [30 Features]		FE [Top-10 Features]	
	Benign (%)	Malignant (%)	Benign (%)	Malignant (%)
BRIGID 0	6/8 (75.0)	4/8 (50.0)	7/8 (87.5)	4/8 (50.0)
BRIGID 10U	6/8 (75.0)	2/8 (25.0)	7/8 (87.5)	3/8 (37.5)
BRIGID 10E	8/8 (100)	3/8 (37.5)	8/8 (100)	3/8 (37.5)
BRIGID 15U	7/8 (87.5)	5/8 (65.5)	8/8 (100)	5/8 (65.5)
BRIGID 15E	6/8 (75.0)	1/8 (12.5)	7/8 (87.5)	2/8 (25.0)
BRIGID 20E	8/8 (100)	5/8 (65.5)	8/8 (100)	4/8 (50.0)
BRIGID 30E	3/8 (37.5)	7/8 (87.5)	4/8 (50.0)	6/8 (75.0)
Total	44/56 (78.6)	27/56 (48.2)	49/56 (87.5)	27/56 (48.2)

the importance of identifying a subset of features with maximum contribution to the diagnosis problem; additionally, the diagnosis becomes less time-consuming if a reduced feature set is used.

5.6 Discussion

Classification models tend to perform better when the differences in the dataset arise primarily from the differences in the class labels (benign and malignant in the case of this study). Given the diversity of breast cancer scenarios that occur clinically, it is difficult to guarantee that all sources of variance beside the level of malignancy have been removed from consideration; for example, the variance introduced by different levels of heterogeneity, and the variance introduced by the differences in the backscattered signals collected at different channels. Additionally, and considering the overall goal of reducing the operational cost and complexity of the diagnosis process, the classification platform should be designed in such a way that it does not require *a priori* information to ensure its accuracy. For example, requiring knowledge of the tumour location may impair the usability of automated diagnosis systems, as currently available algorithms may be subject to errors in detecting and localising targets in breasts with varying contents of

glandular tissue.

The results from this chapter demonstrated the relevance of appropriately designing classification models. Particularly, the EAC models proposed with this study rely on grouping together signals of the same channel angle; with this design, a full view of a tumour is offered to the learning process, and there is an inherent separation by tumour propagation path as a consequence. As the EAC models do not require tumour location to be known in advance, they allow for all of the information collected in a breast scan to be appropriately used, and ensures that channels providing incorrect predictions have less of an impact in the overall diagnostic performance, without the need to penalise certain channels based on patient-specific channel-tumour distance.

An in-depth analysis of the contribution of each channel to the overall diagnostic performance also showed that the channels with mid-range propagation path lengths preserve the most relevant information towards the classification process. Particularly, antennas nearest tumour models in the vicinity of skin were found to have the smallest propagation paths and are problematic for the classification; in addition, reflection channels that are far from the tumour were also found to underperform, as the signals in these scenarios are largely dominated by attenuation.

The results confirmed the usefulness of the antenna grouping algorithm, and also investigated the pre-processing methods by comparing feature extraction and tumour windowing algorithms. Overall, both pre-processing methods achieve similar performances, although feature extraction in certain cases suffers from more incorrect predictions. This result is, nevertheless, encouraging, as feature extraction does not rely on *a priori* knowledge of the tumour location, as is the case of tumour windowing.

The impact of tissue heterogeneity was also assessed. In its current design, the 3-stage diagnosis platforms offers better accuracy in diagnosing benign tumours when compared to the malignant tumours, resulting in an increased specificity, i.e., lower rates of false positives. This serves as a strong indicator of the potential of microwave breast diagnosis systems in countering the high rates of false positives currently linked to screening mammography (as discussed in Section 2.2.4.2). Nevertheless, the sensitivity in classifying the malignant tumours in the BRIGID phantom set was low, which can be improved by further optimisation of the random forests model, via the introduction of a cost value to the process.

It was also observed that there is a large variation in the sensitivity and specificity of the system across breast models. This result may, in part, be explained by the small size of the experimental phantom set, creating large fluctuations in the performance analysis between breast models. It is also likely that the diagnosis platform has an upper limit on the amount of

Table 5.4: Summary of the performance obtained in the numerical and experimental investigations of microwave diagnosis platforms, when using feature extraction in the data processing stage. The numerical results of performance are averaged across all three breast models that were each classified by a new random forests model. The experimental result corresponds to using the top-10 features for the diagnosis. Accuracy, Specificity, Sensitivity and AUC (Area Under the ROC Curve) are shown.

	Numerical	Experimental Top-10	Difference [Experimental - Numerical]
Accuracy (%)	75.4	67.8	-7.6
Specificity (%)	82.6	88.5	+5.9
Sensitivity (%)	68.2	51.3	-16.9
AUC (%)	83.9	80.9	-3.0

glandular content before returning random predictions. Larger datasets are needed to assess the true impact of high volumes of glandular content in the diagnosis process.

The relative contribution of the different features to the diagnostic performance was assessed. A majority of features that rely on peak analysis of the time-domain backscattered signals showed a particularly high contribution to the overall diagnostic performance with the BRIGID phantom set; in addition, some features from the autocorrelation and power spectral density analyses also ranked high in terms of classification contribution. It should be noted that the top-10 performing sets of features in the numerical and the experimental studies were not exactly the same, although they rely on similar numbers of features in the time- and frequency-domain. The difference in the result suggests that an optimal subset of features may need consider to the characteristics of each acquisition hardware system. Using the top-10 features that showed maximised contribution to the random forests algorithm with the BRIGID phantom set helped increased the specificity of the system by a further 9%, which indicates that there is an optimal number of refined features to be used with microwave diagnosis platforms.

Quantitatively, a comparison is now drawn between the results obtained for the numerical and experimental investigations. Table 5.4 compares the accuracy, specificity, sensitivity and AUC for EAC models using the features extracted from all signals available from the data collection; since the numerical investigation produced different classifiers for each breast background type, the results across all three numerical breast models are averaged. Additionally, the experimental result corresponds to using the

top-10 features for the diagnosis.

Overall, the experimental results show a decrease of 7.6% and 3.0% in the accuracy and AUC of the diagnosis, respectively. This result may be explained due to the non-ideal conditions that are inherent to an experimental acquisition, such as noise; additionally, it is likely that the size of the BRIGID phantom set is not sufficient to truly quantify the effect of acquiring data in an experimental or clinical setting. Nevertheless, the high specificity result achieved both in the numerical and experimental investigations indicates the potential of microwave breast diagnosis systems in the screening of breast cancer for asymptomatic populations.

5.7 Conclusion

In this chapter, the 3-stage microwave diagnosis platform was validated with the BRIGID experimental phantom set. The diagnosis platform consists of data acquisition with an experimental microwave prototype designed as part of collaborative work developed during this research. Data pre-processing allows highlighting the tumour response, whether by means of tumour windowing or feature extraction. And finally, diagnosis is performed by means of random forests in a carefully designed architecture, where multiple classifiers each classify the responses from a group of signals originating from channels with the same channel angle; furthermore, the predictions from the individual models are grouped in a fused-type model.

Ultimately, classification models function by finding similarities and differences across signals in diverse populations; the pre-grouping of signals by channel angle ensures that the classification models have a better chance at identifying the differences in the signals that arise from the differences in the level of malignancy, and not from other effects such as tissue heterogeneity or the inherent inter-channel variance. Overall, the proposed system architecture is shown to be a potentially valuable component of microwave diagnosis platforms, as it ensures that all data collected in a microwave scan can be used during the diagnosis process, without the need for *a priori* and specific patient information.

In line with the observations from the numerical investigation of Chapter 4, this chapter verified that:

1) The diagnosis process does not have to rely on accurate knowledge of tumour location;
2) Feature extraction is an efficient mechanism of accentuating the differences between benign and malignant tumours;

3) Using only a subset of features that contain the most relevant information about the signature of a tumour is beneficial to the diagnosis;
4) That machine learning architecture design, and good practice should always be adhered to.

Future work in the area of automated microwave breast diagnosis should continue on the identification of a set of features that captures relevant information about the shape of a tumour. The ideal subset of features may operate in the time- or frequency- domain, but should not be affected by increased content of glandular tissue and increased noise from experimental or clinical prototypes; furthermore, the optimal subset of features may need to be tuned to the characteristics of each acquisition hardware.

At the same time, it should also be considered that an automated diagnosis platform should not have to rely on complex classification algorithms to guarantee its accuracy; the identification of an optimal subset of features would ease the task of the classifier, ultimately contributing to the performance of the diagnosis.

Finally, larger experimental datasets are also needed to assess the true potential of the diagnosis platform in breast models with high volumes of glandular content.

CHAPTER 6

Conclusions

This chapter summarises the research presented in this thesis and presents opportunities for future work. A summary of the work is presented in Section 6.1. Section 6.2 summarises the main contributions of the work, while Section 6.3 presents a discussion of future work avenues.

6.1 Summary of the Research

Breast cancer still remains the most common cause of cancer death for women today. Early detection of cancer in populations showing no symptoms of the disease is key to survival, as breast cancers identified at an early stage allow for more effective deployment of medical treatments. To this extent, x-ray mammography is widely used as an imaging modality to regularly screen asymptomatic women; however, recent studies have questioned the value of screening mammography, with figures indicating that the ratio of avoided breast cancer deaths to number of overdiagnosed cancers may vary between 1:2 and 1:5, with at least 152 false positives and 31 unnecessary biopsies per breast cancer death avoided. In the context of current practices for breast cancer screening, a need exists for emerging breast imaging modalities that are less prone to false positives and unnecessary biopsies. **Chapter 1** introduced this topic by describing the societal context and current screening practices for breast cancer.

Chapter 2 provided a detailed literature review of the past works concerning the research topic of this thesis. The anatomy of the breast was described from the point of view of system development, including a characterisation of benign and malignant breast findings; the shape and spiculation of a tumour were identified as recognised markers for malignancy. The guidelines for the clinical assessment of breast cancer, both in symptomatic

and asymptomatic populations were also discussed, highlighting the uncertainty of the medical opinion regarding the value of screening mammography. Next, the core principles of microwave systems were described including the dielectric properties of the breast, signal acquisition and image reconstruction, and the state-of-the-art microwave breast prototypes. Lastly, a review of the literature on the topic of microwave breast diagnosis systems using machine learning was also presented. The reviewed studies presented small-scale analyses suggesting that there may be information preserved in a backscattered signal to inform on the shape and spiculation of a tumour, and that machine learning-based platforms may be able to learn from this information to classify tumours as benign or malignant.

Chapter 3 presented the two phantom sets developed in this work, which were subsequently used in investigating the diagnosis platform developed in this thesis.

The first phantom set discussed in Chapter 3 is based on a numerical tumour modelling algorithm that extended previous work from the literature. The tumour models produced may vary in size, shape and level of spiculation creating a wide range of breast cancer models. In addition, independent analyses by radiologists indicated that the tumour models are clinically-representative; the clinical validation of tumour models for testing of microwave systems is novel in the literature.

An experimental phantom set was also proposed in Chapter 3 — the BRIGID phantom set, which consists of a set of tumour and breast phantoms. The tumour phantoms include 3 classes of malignancy, represented in terms of increasing levels of spiculation at the border of the tumour. The set of tumour models can be dynamically combined with several breast phantoms, which have increasing percentages of glandular tissue to represent the diversity of breast density in women. The inclusion of different levels of tumour spiculation and different percentages of glandular tissue in experimental models is novel in the literature.

The following chapters investigated the potential of a machine learning platform to classify breast tumours as benign or malignant with microwave backscattered signals. The examination in **Chapter 4** was based on the numerical tumour models presented in the previous chapter, while the analysis in **Chapter 5** used a dataset of signals from the BRIGID phantom set, where signals were acquired with an experimental microwave prototype developed during this research through collaborative work.

A 3-stage automated platform was developed that encompasses data acquisition, data pre-processing (which includes a tumour windowing, TW, and a feature extraction, FE, algorithm), and final diagnosis by means of the random forests classification algorithm in a two-level architecture. The

classification architecture is such that similar signals are grouped together in different classification models; in this work, the channel angle available in a microwave prototype system was investigated as a suitable candidate for the division of backscattered signals. The diagnosis stage also includes an antenna grouping algorithm, which is applied after the classification predictions are made. With this algorithm, a majority vote of all available predictions respective to one scan is made to produce a final benign/malignant label. The validation methodology implemented in the diagnosis platform adheres to good machine learning practice, to ensure results are representative and reproducible.

The results from the numerical and experimental analyses suggest that feature extraction may be an efficient mechanism of retrieving the differences between benign and malignant tumours captured in backscattered signals; in addition, the feature extraction algorithm does not rely on accurate *a priori* knowledge of tumour location (as is the case with the tumour windowing approach). The analyses also indicate that dividing signals in separate classification models according to their channel angle may be an efficient way of improving diagnostic performance. In addition, the inclusion of the antenna grouping algorithm in the machine learning platform was shown to be a key step for an increased diagnostic performance. In its current design, the proposed diagnosis platform offers better accuracy in diagnosing benign tumours when compared to malignant tumours, resulting in an increased specificity, i.e., lower rates of false positives.

Overall, the numerical results shown in Chapter 4 were supported by the experimental conclusions in Chapter 5. Quantitatively, the investigation using experimental data revealed lower diagnostic performance when compared to the numerical investigation, which may be explained due to the non-ideal conditions inherent to experimental acquisitions; additionally, it is likely that the size of the BRIGID phantom set is not sufficient to truly quantify the effect of acquiring data in experimental or clinical settings.

6.2 Summary of the Contributions

The main contributions and outputs of the work presented in this thesis are restated here:

1) Development of clinically-validated breast tumour phantoms. This contribution resulted in the journal publication titled “Development of Clinically Informed 3-D Tumor Models for Microwave Imaging Applica-

tions", published in *IEEE Antennas and Wireless Propagation Letters* (2015);

2) Design and fabrication of experimental breast phantoms, including increasing percentages of breast density, and experimental breast tumour phantoms with increasing levels of spiculation. This contribution resulted in the journal publication titled "Microwave Breast Imaging: Experimental Tumour Phantoms for the Evaluation of New Breast Cancer Diagnosis Systems" published in *Biomedical Physics and Engineering Express* (2018);

3) Investigation of a feature set to capture the difference in backscattered signals between benign and malignant breast tumours;

4) Investigation of machine learning architecture design and machine learning validation methodologies, and development of a machine learning-based platform for the classification of breast tumours as benign or malignant based on their backscattered response, with novel elements as identified above;

5) Testing of the diagnosis platform by means of both a numerical dataset of clinically-representative tumours, and experimental data from the BRIGID phantom set. This work was published "Diagnosing Breast Cancer with Microwave Technology: remaining challenges and potential solutions with machine learning", *Diagnostics* (2018); a second paper on validation using the experimental data, titled "Experimental Validation of a Machine Learning Platform for Diagnosing Breast Cancer with Microwave Technology", is currently in preparation.

Additionally, contributions from collaborative work developed throughout this research include:

1) Design, building and testing of an experimental microwave prototype system. This work was led by Declan O'Loughlin and is described in two co-authored journal publications: O'Loughlin *et al.*, "Parameter Search Algorithms for Microwave Radar-Based Breast Imaging: Focal Quality Metrics as Fitness Functions", *Sensors* (2017); O'Loughlin *et al.*, "Sensitivity and Specificity Estimation using Patient-Specific Microwave Imaging in Diverse Experimental Breast Phantoms", *IEEE Transactions on Medical Imaging* (Accepted, 2018);

2) Development of open-source and extensible imaging software, which can be used for image reconstruction with many types of system design. This work was led by Declan O'Loughlin and resulted in the co-authored

conference publication O'Loughlin *et al.* , "Open-Source Software for Microwave Radar-Based Image Reconstruction", *12th European Conference on Antennas and Propagation (EuCAP)* (2018).

6.3 Discussion and Future Work

The results pertaining to the main goal of this thesis have demonstrated that there is merit in further investigating an automated platform for the diagnosis of breast cancer using microwave technology in conjunction with machine learning. A set of characteristics for the development of an automated diagnosis platform was identified, namely, microwave breast diagnosis systems should:

1) Adhere to good machine learning practices, to guarantee the reproducibility of the results;
2) Extract features that accurately characterise the shape information contained in the signature of a tumour, while de-emphasising the influence of the surrounding glandular tissue;
3) Use an appropriate classification architecture, such as, pre-grouping signals by reflection angle;
4) Implement an antenna grouping algorithm, to group all of the individual classification predictions from different channels into one final diagnosis for each tumour;
5) Implement an optimisation strategy that rewards system specificity, while conditioning the minimum sensitivity the diagnosis system should deliver.

To improve on the proposed microwave diagnosis platform, future work could consider an antenna grouping algorithm that emphasises the predictions of some channels, while de-emphasising others (in its current implementation, the antenna grouping algorithm is based majority vote from all predictions pertaining to one breast scan). For example, channels with lower angles between transmit and receive antennas were found to consistently yield more accurate predictions. One challenge with such an algorithm lies with the fact that the diagnosis platform should not rely heavily on previous knowledge of tumour location.

Future work could also consider integrating an additional layer of data processing to deal with breast heterogeneity. If knowledge of the average dielectric properties of a breast can be obtained, different classifiers can

be created and optimised for different ranges of breast glandular content; in this scenario, the classifier is likely to achieve higher performances, as demonstrated with the numerical study. One straightforward way to estimate the average dielectric properties of the breast could rely on the length of transmissions paths when data are acquired. Another, more complex method, could follow the example of the University of Calgary, Canada; the group have built a complementary system to their microwave radar prototype, which serves to obtain an average of the dielectric properties of each breast.

Considering the ultimate goal of translating such a system to the clinical practice, the diagnosis process should classify observations by means of algorithms that are computationally inexpensive and not time-consuming. To this extent, the investigation of a new feature set is a key enabler of the operation of diagnosis systems. Future work in the area of feature engineering should investigate features both in the time and frequency domains, that are 1) tuned to the hardware specifications of a microwave prototype, and 2) able to decompose the backscattered signal and differentiate which components arise from the tumour signature and from the glandular tissue. Ensuring the robustness of the extracted features in breasts with high contents of glandular tissue bears clinical relevance, as breast density is a recognised risk factor for the occurrence of breast cancer.

To be able to analyse the true potential of automated classification platforms, larger experimental datasets are needed. Future work in this area could include extending the BRIGID phantom set to include more breast and tumour phantoms. Additionally, as the BRIGID phantom set is suitable for testing with a variety of prototypes, it would be of interest to use the BRIGID phantom set to acquire experimental data with some of the state-of-the-art microwave breast prototypes. The operating conditions of current microwave breast prototypes have been validated with small scale clinical trials, creating an ideal scenario for the acquisition of a large database of high-quality experimental backscattered signals.

In the long-term, further enhancement of microwave diagnosis platforms may integrate more comprehensive knowledge of the dielectric properties of breast tissues to increase the accuracy of the diagnosis. Recent studies have indicated that the relative permittivity and conductivity of benign and malignant tumours might differ. For example, studies with the tomography-based prototype system developed by Dartmouth College have identified a change in the dielectric properties, particularly, the conductivity, between benign and malignant tumours. An additional study using the radar-based MARIA® system has also suggested that it is potentially possible to automatically discriminate benign and malignant breast tissues based on their frequency-dependent dielectric response. Using the estimated relative permit-

tivity and conductivity of detected tumours as features to the classification process may be key in further refining the diagnosis platform. However, a number of studies have recently highlighted the uncertainty regarding the measurement of the dielectric properties of human tissues. Further studies would be required to accurately determine the relative permittivity and conductivity of benign and malignant breast tissues, before accurate diagnosis based on these properties may be realisable.

In summary, clinical experience to date with modalities such as X-ray mammography have pointed to the need to develop new, more effective tools. To quantify the potential of microwave breast systems, future clinical trials with this technology would have to focus not only on identifying the presence of a tumour, but also, on the blind assessment of their level of malignancy. In addition, future trials would also need to consider that accurate validation of a machine learning-based diagnosis platform involves large samples of signals, including a fair representation of several breast tumour types, which means larger patient populations (for comparison, one of the largest trials performed to date reports detection results for 223 patients only).

Overall, however, the results of this work suggest that future microwave breast analysis systems, which include automated diagnosis platforms such as the one proposed in this work, could help to tangibly impact patient outcomes.

Bibliography

[1] American Cancer Society, *Cancer Facts & Figures 2018*. Atlanta, GA: American Cancer Society, Inc, 2018.

[2] M. Ervik, F. Lam, J. Ferlay *et al.* (). Cancer Today, [Online]. Available: `http://gco.iarc.fr/today` (visited on 20/05/2018).

[3] C.-H. Yip, R. A. Smith, B. O. Anderson *et al.*, "Guideline implementation for breast healthcare in low- and middle-income countries: Early detection resource allocation", *Cancer*, vol. 113, pp. 2244–2256, 8 Suppl 15th Oct. 2008.

[4] M. P. Curado, B. Edwards, H. R. Shin *et al.*, Eds., *Cancer Incidence in Five Continents Vol. IX*, vol. 160, Lyon, France: IARC Scientific Publications, 2007.

[5] R. Sankaranarayanan, R. Swaminathan and R. Lucas, *Cancer Survival in Africa, Asia, the Caribbean and Central America: Database and Attributes (SurvCan)*. Lyon, France: IARC Scientific Publications, 2011, vol. 162.

[6] American Cancer Society, *Breast Cancer Facts & Figures 2017-2018*. Atlanta, GA: American Cancer Society, Inc, 2017.

[7] World Health Organization, *WHO Position Paper on Mammography Screening*. Geneva, Switzerland, 2014.

[8] E. Sickles, C. D'Orsi and L. Bassett, *ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013.

[9] *Breast Cancer Surveillance Consortium*. NCI-funded Breast Cancer Surveillance Consortium (HHSN261201100031C), http://www.bcsc-research.org/statistics/performance/screening/2009/perf_age.html.

[10] R. J. Hooley, L. M. Scoutt and L. E. Philpotts, "Breast Ultrasonography: State of the Art", *Radiology*, vol. 268, no. 3, pp. 642–659, 1st Sep. 2013.

[11] A. Chetlen, J. Mack and T. Chan, "Breast cancer screening controversies: Who, when, why, and how?", *Clin. Imaging*, vol. 40, no. 2, pp. 279–282, 1st Mar. 2016.

[12] D. Saslow, C. Boetes, W. Burke *et al.*, "American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography", *CA Cancer J. Clin.*, vol. 57, no. 2, pp. 75–89, 1st Mar. 2007.

[13] W. DeMartini, C. Lehman and S. Partridge, "Breast MRI for cancer detection and characterization: A review of evidence-based clinical applications", *Acad. Radiol.*, vol. 15, no. 4, pp. 408–416, Apr. 2008.

[14] D. S. Salem, R. M. Kamal, S. M. Mansour *et al.*, "Breast imaging in the young: The role of magnetic resonance imaging in breast cancer screening, diagnosis and follow-up", *J. Thorac. Dis.*, vol. 5, S9–S18, Suppl 1 Jun. 2013.

[15] H. G. Welch and W. C. Black, "Overdiagnosis in Cancer", *J Natl Cancer Inst*, vol. 102, no. 9, pp. 605–613, 5th May 2010.

[16] M. McCartney, *The Patient Paradox*. London, UK: Pinter & Martin Publishers, 7th Nov. 2013, 338 pp.

[17] D. Berry, "Breast cancer screening: Controversy of impact", *Breast*, vol. 22, S73–S76, 0 2 Aug. 2013.

[18] P. C. Gøtzsche and K. J. Jørgensen, "Screening for breast cancer with mammography", *Cochrane Database Syst. Rev.*, no. 6, p. CD001877, 4th Jun. 2013.

[19] B. Heleno, M. F. Thomsen, D. S. Rodrigues *et al.*, "Quantification of harms in cancer screening trials: Literature review", *BMJ*, vol. 347, f5334, 16th Sep. 2013.

[20] N. Biller-Andorno and P. Jüni, "Abolishing Mammography Screening Programs? A View from the Swiss Medical Board", *N. Engl. J. Med.*, vol. 370, no. 21, pp. 1965–1967, 22nd May 2014.

[21] K. J. Jørgensen, M. Kalager, A. Barratt *et al.*, "Overview of guidelines on breast screening: Why recommendations differ and what to do about it", *Breast*, vol. 31, pp. 261–269, 1st Feb. 2017.

[22] P. Meaney, M. Fanning, D. Li *et al.*, "A clinical prototype for active microwave imaging of the breast", *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 11, pp. 1841–1853, Nov. 2000.

[23] S. P. Poplack, K. D. Paulsen, A. Hartov *et al.*, "Electromagnetic Breast Imaging: Average Tissue Property Values in Women with Negative Clinical Findings", *Radiology*, vol. 231, no. 2, pp. 571–580, 1st May 2004.

[24] P. M. Meaney, M. W. Fanning, T. Raynolds *et al.*, "Initial Clinical Experience with Microwave Breast Imaging in Women with Normal Mammography", *Acad. Radiol.*, vol. 14, no. 2, pp. 207–218, Feb. 2007.

[25] S. P. Poplack, T. D. Tosteson, W. A. Wells *et al.*, "Electromagnetic Breast Imaging: Results of a Pilot Study in Women with Abnormal Mammograms", *Radiology*, vol. 243, no. 2, pp. 350–359, May 2007.

[26] P. M. Meaney, P. A. Kaufman, L. S. Muffly *et al.*, "Microwave imaging for neoadjuvant chemotherapy monitoring: Initial clinical experience", *Breast Cancer Res.*, vol. 15, no. 2, R35, 24th Apr. 2013.

[27] A. W. Preece, I. J. Craddock, M. Shere *et al.*, "MARIA M4: Clinical evaluation of a prototype ultrawideband radar scanner for breast cancer detection", *J. Med. Imaging*, vol. 3, no. 3, pp. 033502-1-7, 20th Jul. 2016.

[28] P. Bannister, "A Novel Microwave Radar Breast Imaging System in a Symptomatic Breast Clinic", in *BSBR Breast Imaging Research Network Workshop*, Manchester, UK, 6th Nov. 2016.

[29] N. Ridley, A. Iriarte, L. Tsui *et al.*, "Automatic labelling of lesions using radiofrequency feature discrimination", in *European Congress of Radiology*, Vienna, Austria, 1st–5th Mar. 2017.

[30] N. Ridley, M. Shere, I. Lyburn *et al.*, "Cancer detection in dense tissue using radiofrequency imaging - a clinical evaluation", in *European Congress of Radiology*, Vienna, Austria, 1st–5th Mar. 2017.

[31] H. Massey, N. Ridley, I. Lyburn *et al.*, "Radiowave detection of breast cancer in the symptomatic clinic - a multi-centre study", in *International Cambridge Conference on Breast Imaging*, Cambrigde, UK, 3rd–4th Jul. 2017.

[32] E. C. Fear, J. Bourqui, C. Curtis *et al.*, "Microwave Breast Imaging With a Monostatic Radar-Based System: A Study of Application to Patients", *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 5, pp. 2119–2128, May 2013.

[33] F. Yang, L. Sun, Z. Hu *et al.*, "A large-scale clinical trial of radar-based microwave breast imaging for Asian women: Phase I", in *IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, San Diego, CA, USA, 9th–14th Jul. 2017, pp. 781–783.

[34] Y. Kuwahara, S. Miura, Y. Nishina *et al.*, "Clinical test of microwave mammography", in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Orlando, FL, USA, 7th–13th Jul. 2013, pp. 2028–2029.

[35] Y. Kuwahara, "Microwave Imaging for Early Breast Cancer Detection", in *Breast Imaging*, A. M. Malik, Ed., Rijeka: InTech, 4th Oct. 2017, Ch. 3.

[36] H. Song, S. Sasada, T. Kadoya *et al.*, "Detectability of Breast Tumor by a Hand-held Impulse-Radar Detector: Performance Evaluation and Pilot Clinical Study", *Sci. Rep.*, vol. 7, no. 1, p. 16 353, 27th Nov. 2017.

[37] E. Porter, H. Bahrami, A. Santorelli *et al.*, "A Wearable Microwave Antenna Array for Time-Domain Breast Tumor Screening", *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1501–1509, Jun. 2016.

[38] J. S. Suri and R. M. Rangayyan, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*. Washington, WA: SPIE, 2006, vol. 15, 1012 pp.

[39] T. Jinshan, R. M. Rangayyan, X. Jun *et al.*, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances", *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[40] Y. Chen, E. Gunawan, K. S. Low *et al.*, "Effect of Lesion Morphology on Microwave Signature in 2-D Ultra-Wideband Breast Imaging", *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 2011–2021, 2008.

[41] S. Davis, B. D. Van Veen, S. C. Hagness *et al.*, "Breast Tumor Characterization Based on Ultrawideband Microwave Backscatter", *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 237–246, 2008.

[42] R. Conceição, M. O'Halloran, D. Byrne *et al.*, "Tumor classification using radar target signatures", in *Proceedings of Progress in Research and Electromagnetics Symposium*, vol. 1, Cambridge, Massachusetts, USA, 5th–8th Jul. 2010, pp. 346–349.

[43] R. C. Conceição, M. O'Halloran, E. Jones *et al.*, "Investigation of classifiers for early-stage breast cancer based on radar target signatures", *Prog. Electromagn. Res.*, vol. 105, pp. 295–311, 2010.

[44] R. C. Conceição, M. O'Halloran, M. Glavin *et al.*, "Support Vector Machines for the Classification of Early-Stage Breast Cancer Based on Radar Target Signatures", *Prog. Electromagn. Res. B*, vol. 23, pp. 311–327, 2010.

[45] R. C. Conceição, M. O'Halloran, M. Glavin *et al.*, "Evaluation of features and classifiers for classification of early-stage breast cancer", *J. Electromagn. Waves Appl.*, vol. 25, no. 1, pp. 1–14, 2011.

[46] R. C. Conceição, R. M. Capote, B. L. Oliveira *et al.*, "Imaging and classification of breast cancer with multimodal PEM-UWB techniques", in *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Torino, Italy, 9th–13th Sep. 2013, pp. 421–424.

[47] R. C. Conceição, R. M. Capote, B. L. Oliveira *et al.*, "Novel multimodal PEM-UWB approach for breast cancer detection: Initial study for tumour detection and consequent classification", in *7th European Conference on Antennas and Propagation (EuCAP)*, Gothenburg, Sweden, 8th–12th Apr. 2013, pp. 630–634.

[48] M. O'Halloran, B. McGinley, R. C. Conceição *et al.*, "Spiking neural networks for breast cancer classification in a dielectrically heterogeneous breast", *Prog. Electromagn. Res.*, vol. 113, pp. 413–428, 2011.

[49] M. O'Halloran, S. Cawley, B. McGinley *et al.*, "Evolving Spiking Neural Network Topologies for Breast Cancer Classification in a Dielectrically Heterogeneous Breast", *Prog. Electromagn. Res. Lett.*, vol. 25, pp. 153–162, Jul. 2011.

[50] M. Jones, D. Byrne, B. McGinley *et al.*, "Classification and monitoring of early stage breast cancer using ultra wideband radar", in *8th International Conference on Systems (ICONS)*, Seville, Spain, 27th Jan. 2013, pp. 46–51.

[51] R. C. Conceição, H. Medeiros, M. O'Halloran *et al.*, "Initial classification of breast tumour phantoms using a UWB radar prototype", in *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Torino, Italy, 9th–13th Sep. 2013, pp. 720–723.

[52] R. C. Conceição, H. Medeiros, M. O'Halloran *et al.*, "SVM-based classification of breast tumour phantoms using a UWB radar prototype system", in *31st URSI General Assembly and Scientific Symposium (URSI-GASS)*, Beijing, China, 16th–23rd Aug. 2014, pp. 1–4.

[53] B. L. Oliveira, A. Shahzad, M. O'Halloran *et al.*, "Combined Breast Microwave Imaging and Diagnosis System", in *Proceedings of Progress In Electromagnetics Research Symposium*, vol. 1, Prague, Czech Republic, 6th–9th Jul. 2015, pp. 274–278.

[54] A. Ringberg, E. Bågeman, C. Rose *et al.*, "Of cup and bra size: Reply to a prospective study of breast size and premenopausal breast cancer incidence", *Int. J. Cancer*, vol. 119, no. 9, pp. 2242–2243, 1st Nov. 2006.

[55] D. E. McGhee and J. R. Steele, "Breast volume and bra size", *Int. J. Cloth. Sci. Technol.*, vol. 23, no. 5, pp. 351–360, 4th Oct. 2011.

[56] A. Markkula, A. Bromée, M. Henningson *et al.*, "Given breast cancer, does breast size matter? Data from a prospective breast cancer cohort", *Cancer Causes Control*, vol. 23, no. 8, pp. 1307–1316, 1st Aug. 2012.

[57] K. L. Moore, A. F. Dalley and A. M. R. Agur, *Clinically Oriented Anatomy*, 6th Edition. Philadelphia, PA: Lippincott Williams & Wilkins, 2009.

[58] R. L. Drake, A. . Vogl and A. W. M. Mitchell, *Gray's Anatomy for Students*, 2nd Edition. Philadelphia, PA: Churchill Livingstone/Elsevier, 2010.

[59] L. S. Bickley and P. G. Szilagyi, *Bates' Guide to Physical Examination and History Taking*, 11th Ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2012.

[60] S. Negin Mortazavi, D. Geddes and F. Hassanipour, "Lactation in the Human Breast From a Fluid Dynamics Point of View", *J. Biomech. Eng.*, vol. 139, no. 1, 1st Jan. 2017.

[61] B. L. Oliveira, M. O'Halloran, R. C. Conceição *et al.*, "Development of Clinically Informed 3-D Tumor Models for Microwave Imaging Applications", *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 520–523, 2016.

[62] G. M. Cooper and R. E. Hausman, *The Cell: A Molecular Approach*, 7th Edition. Sunderland, MA: Sinauer Associates, 2016.

[63] W. A. N. Dorland, *Dorland's Illustrated Medical Dictionary*, 32nd Edition. Philadelphia, PA: Saunders/Elsevier, 2012.

[64] J. M. Dixon, *ABC of Breast Diseases*, 3rd Edition. Wiley, 13th Feb. 2006.

[65] S. R. Wellings and H. M. Jensen, "On the Origin and Progression of Ductal Carcinoma in the Human Breast", *J. Natl. Cancer. Inst.*, vol. 50, no. 5, pp. 1111–1118, 1st May 1973.

[66] S. R. Wellings, "A Hypothesis of the Origin of Human Breast Cancer from the Terminal Ductal Lobular Unit", *Pathol. Res. Pract.*, vol. 166, no. 4, pp. 515–535, 1st Apr. 1980.

[67] World Health Organization, *ICD-10: International Statistical Classification of Diseases and Related Health Problems - 10th Revision*, 2nd Edition. Geneva: World Health Organization, 2004.

[68] ——, *ICD-O-3: International Classification of Diseases for Oncology - 3rd Edition*. Geneva: World Health Organisation, 2013.

[69] P. D. Darbre, "Recorded Quadrant Incidence of Female Breast Cancer in Great Britain Suggests a Disproportionate Increase in the Upper Outer Quadrant of the Breast", *Anticancer Res.*, vol. 25, pp. 2543–2550, 3C 5th Jan. 2005.

[70] V. Y. Sohn, Z. M. Arthurs, J. A. Sebesta *et al.*, "Primary tumor location impacts breast cancer survival", *Am. J. Surg.*, vol. 195, no. 5, pp. 641–644, May 2008.

[71] *Breast Cancer Incidence, Mortality, Treatment and Survival in Ireland: 1994-2009*. Cork, Ireland: National Cancer Registry, 2012.

[72] (). Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2013 Sub (1973-2011 varying) - Linked To County Attributes - Total U.S., 1969-2012 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2014, based on the November 2013 submission.

[73] G. Ursin, L. Hovanessian-Larsen, Y. R. Parisky *et al.*, "Greatly increased occurrence of breast cancers in areas of mammographically dense tissue", *Breast Cancer Res.*, vol. 7, no. 5, R605–R608, 2005.

[74] S. M. P. Pereira, V. A. McCormack, S. M. Moss *et al.*, "The spatial distribution of radiodense breast tissue: A longitudinal study", *Breast Cancer Res.*, vol. 11, no. 3, R33, 2009.

[75] S. M. P. Pereira, V. A. McCormack, J. H. Hipwell *et al.*, "Localized Fibroglandular Tissue as a Predictor of Future Tumor Location within the Breast", *Cancer Epidemiol. Biomarkers Prev.*, vol. 20, no. 8, pp. 1718–1725, 8th Jan. 2011.

[76] C. E. DeSantis, J. Ma, A. G. Sauer *et al.*, "Breast cancer statistics, 2017, racial disparity in mortality by state", *CA Cancer J. Clin.*, vol. 67, no. 6, pp. 439–448, 1st Nov. 2017.

[77] R. M. Tamimi, D. Spiegelman, S. A. Smith-Warner *et al.*, "Population Attributable Risk of Modifiable and Nonmodifiable Breast Cancer Risk Factors in Postmenopausal Breast Cancer", *Am. J. Epidemiol.*, vol. 184, no. 12, pp. 884–893, 15th Dec. 2016.

[78] M. A. Lopez-Garcia, F. C. Geyer, M. Lacroix-Triki *et al.*, "Breast cancer precursors revisited: Molecular features and progression pathways", *Histopathology*, vol. 57, no. 2, pp. 171–192, 1st Aug. 2010.

[79] M. Morrow, S. J. Schnitt and L. Norton, "Current management of lesions associated with an increased risk of breast cancer", *Nat. Rev. Clin. Oncol.*, vol. 12, no. 4, pp. 227–238, Apr. 2015.

[80] T. A. King, M. Pilewskie, S. Muhsen *et al.*, "Lobular Carcinoma in Situ: A 29-Year Longitudinal Experience Evaluating Clinicopathologic Features and Breast Cancer Risk", *J. Clin. Oncol.*, vol. 33, no. 33, pp. 3945–3952, 20th Nov. 2015.

[81] S. W. Dyrstad, Y. Yan, A. M. Fowler *et al.*, "Breast cancer risk associated with benign breast disease: Systematic review and meta-analysis", *Breast Cancer Res. Treat.*, vol. 149, no. 3, pp. 569–575, Feb. 2015.

[82] C. Johnsén, "Breast disease. A clinical study with special reference to diagnostic procedures", *Acta. Chir. Scand. Suppl.*, vol. 454, pp. 1–108, 1975.

[83] A. Morris, R. F. Pommier, W. A. Schmidt *et al.*, "Accurate evaluation of palpable breast masses by the triple test score", *Arch. Surg.*, vol. 133, no. 9, pp. 930–934, Sep. 1998.

[84] C. Hermansen, H. Skovgaard Poulsen, J. Jensen *et al.*, "Diagnostic reliability of combined physical examination, mammography, and fine-needle puncture ("triple-test") in breast tumors. A prospective study", *Cancer*, vol. 60, no. 8, pp. 1866–1871, 15th Oct. 1987.

[85] K. T. Morris, R. F. Pommier, A. Morris *et al.*, "Usefulness of the Triple Test Score for Palpable Breast Masses", *Arch. Surg.*, vol. 136, no. 9, pp. 1008–1013, 1st Sep. 2001.

[86] S. Kharkwal, Sameer and A. Mukherjee, "Triple Test in Carcinoma Breast", *J. Clin. Diagn. Res.*, vol. 8, no. 10, NC09–NC11, Oct. 2014.

[87] Cancer Australia, *The Investigation of a New Breast Symptom: A Guide for General Practitioners 2017, Summary of the Development Process and Methodology*. Surry Hills, NSW: Cancer Australia, 2017.

[88] C. D'Orsi, E. Sickles, E. Mendelson *et al.*, *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013.

[89] E. Sickles and C. D'Orsi, *ACR BI-RADS® Follow-up and Outcome Monitoring. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013.

[90] E. Mendelson, M. Böhm-Vélez and W. Berg, *ACR BI-RADS® Ultrasound. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013.

[91] E. Morris, C. Comstock and C. Lee, *ACR BI-RADS® Magnetic Resonance Imaging. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013.

[92] S. B. Edge, D. R. Byrd, C. C. Compton *et al.*, *AJCC Cancer Staging Manual - 7th Edition*. New York, NY: Springer, 2010.

[93] A. H. Barrett, P. C. Myers and N. L. Sadowsky, "Detection of breast cancer by microwave radiometry", *Radio Sci.*, vol. 12, pp. 167–171, 6S Dec. 1977.

[94] B. Bocquet, J. C. Van De Velde, A. Mamouni *et al.*, "Microwave radiometric imaging at 3 GHz for the exploration of breast tumors", *IEEE Trans. Microw. Theory Techn.*, vol. 38, no. 6, pp. 791–793, Jun. 1990.

[95] F. Bardati and S. Iudicello, "Modeling the Visibility of Breast Malignancy by a Microwave Radiometer", *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 214–221, Jan. 2008.

[96] G. Ku and L. V. Wang, "Scanning microwave-induced thermoacoustic tomography: Signal, resolution, and contrast", *Med. Phys.*, vol. 28, no. 1, pp. 4–10, 1st Jan. 2001.

[97] B. Guo, J. Li, H. Zmuda *et al.*, "Multifrequency Microwave-Induced Thermal Acoustic Imaging for Breast Cancer Detection", *IEEE Trans. Biomed. Eng.*, vol. 54, no. 11, pp. 2000–2010, Nov. 2007.

[98] S. S. Chaudhary, R. K. Mishra, A. Swarup *et al.*, "Dielectric properties of normal and malignant human breast tissues at radiowave and microwave frequencies", *Indian J. Biochem. Biophys.*, vol. 21, no. 1, pp. 76–79, 1984.

[99] A. J. Surowiec, S. S. Stuchly, J. R. Barr *et al.*, "Dielectric properties of breast carcinoma and the surrounding tissues", *IEEE Trans. Biomed. Eng.*, vol. 35, no. 4, pp. 257–263, Apr. 1988.

[100] A. M. Campbell and D. V. Land, "Dielectric properties of female human breast tissue measured in vitro at 3.2 GHz", *Phys. Med. Biol.*, vol. 37, no. 1, p. 193, 1st Jan. 1992.

[101] W. T. Joines, Y. Zhang, C. Li *et al.*, "The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz", *Med. Phys.*, vol. 21, no. 4, pp. 547–550, Apr. 1994.

[102] L. Sha, E. R. Ward and B. Stroy, "A review of dielectric properties of normal and malignant breast tissue", in *IEEE SoutheastCon*, Columbia, SC, USA, 5th–7th Apr. 2002, pp. 457–462.

[103] M. Lazebnik, L. McCartney, D. Popovic *et al.*, "A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries", *Phys. Med. Biol.*, vol. 52, no. 10, pp. 2637–2656, 21st May 2007.

[104] M. Lazebnik, D. Popovic, L. McCartney *et al.*, "A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries", *Phys. Med. Biol.*, vol. 52, no. 20, pp. 6093–6115, 21st Oct. 2007.

[105] E. Porter, M. Coates and M. Popović, "An Early Clinical Study of Time-Domain Microwave Radar for Breast Health Monitoring", *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 530–539, Mar. 2016.

[106] A. Fasoula, S. Anwar, Y. Toutain *et al.*, "Microwave vision: From RF safety to medical imaging", in *11th European Conference on Antennas and Propagation (EuCAP)*, Paris, France, 19th–24th Mar. 2017, pp. 2746–2750.

[107] O. Casas, R. Bragós, P. J. Riu *et al.*, "In vivo and in situ ischemic tissue characterization using electrical impedance spectroscopy", *Ann. N. Y. Acad. Sci.*, vol. 873, pp. 51–58, 20th Apr. 1999.

[108] D. Haemmerich, O. R. Ozkan, J.-Z. Tsai *et al.*, "Changes in electrical resistivity of swine liver after occlusion and postmortem", *Med. Biol. Eng. Comput.*, vol. 40, no. 1, pp. 29–33, 1st Jan. 2002.

[109] R. J. Halter, T. Zhou, P. M. Meaney *et al.*, "The correlation of in vivo and ex vivo tissue dielectric properties to validate electromagnetic breast imaging: Initial clinical experience", *Physiol. Meas.*, vol. 30, no. 6, S121–S136, Jun. 2009.

[110] A. Shahzad, S. Khan, M. Jones *et al.*, "Investigation of the effect of dehydration on tissue dielectric properties in ex vivo measurements", *Biomed. Phys. Eng. Express*, vol. 3, no. 4, p. 045 001, 2017.

[111] T. Sugitani, S.-i. Kubota, S.-i. Kuroki *et al.*, "Complex permittivities of breast tumor tissues obtained from cancer surgeries", *Appl. Phys. Lett.*, vol. 104, no. 25, pp. 253702-1-5, 23rd Jun. 2014.

[112] P. M. Meaney, A. P. Gregory, N. R. Epstein *et al.*, "Microwave open-ended coaxial dielectric probe: Interpretation of the sensing volume re-visited", *BMC Med. Phys.*, vol. 14, no. 1, p. 3, 17th Jun. 2014.

[113] E. Porter and M. O'Halloran, "Investigation of Histology Region in Dielectric Measurements of Heterogeneous Tissues", *IEEE Trans. Antennas Propag.*, vol. 65, no. 10, pp. 5541–5552, Oct. 2017.

[114] N. K. Nikolova, "Microwave Biomedical Imaging", in *Wiley Encyclopedia of Electrical and Electronics Engineering*, J. G. Webster, Ed., Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014.

[115] M. Guardiola, S. Capdevila, J. Romeu *et al.*, "3-D Microwave Magnitude Combined Tomography for Breast Cancer Detection Using Realistic Breast Models", *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 1622–1625, 2012.

[116] A. Shahzad, M. O'Halloran, M. Glavin *et al.*, "A novel optimized parallelization strategy to accelerate microwave tomography for breast cancer screening", in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2014, pp. 2456–2459.

[117] Q. Fang, P. M. Meaney and K. D. Paulsen, "Microwave image reconstruction of tissue property dispersion characteristics utilizing multiple-frequency information", *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 8, pp. 1866–1875, Aug. 2004.

[118] A. Sabouni, S. Noghanian and S. Pistorius, "A global optimization technique for microwave imaging of the inhomogeneous and dispersive breast", *Can. J. Electr. Comp. Eng.*, vol. 35, no. 1, pp. 15–24, Win. 2010.

[119] A. H. Golnabi, P. M. Meaney and K. D. Paulsen, "Tomographic Microwave Imaging With Incorporated Prior Spatial Information", *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 5, pp. 2129–2136, May 2013.

[120] S. Hagness, A. Taflove and J. Bridges, "Two-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: Fixed-focus and antenna-array sensors", *IEEE Trans. Biomed. Eng.*, vol. 45, no. 12, pp. 1470–1479, Dec. 1998.

[121] M. O'Halloran, E. Jones and M. Glavin, "Quasi-Multistatic MIST Beamforming for the Early Detection of Breast Cancer", *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 830–840, Apr. 2010.

[122] M. Klemm, I. J. Craddock, J. A. Leendertz *et al.*, "Improved Delay-and-Sum Beamforming Algorithm for Breast Cancer Detection", *Int. J. Antennas Propag.*, vol. 2008, p. 761 402, 2008.

[123] M. Klemm, J. A. Leendertz, D. Gibbins *et al.*, "Microwave Radar-Based Breast Cancer Detection: Imaging in Inhomogeneous Breast Phantoms", *IEEE Antennas Wireless Propag. Lett.*, vol. 8, pp. 1349–1352, 2009.

[124] D. Byrne and I. J. Craddock, "Time-Domain Wideband Adaptive Beamforming for Radar Breast Imaging", *IEEE Trans. Antennas Propag.*, vol. 63, no. 4, pp. 1725–1735, Apr. 2015.

[125] D. O'Loughlin, B. L. Oliveira, M. A. Elahi *et al.*, "Parameter Search Algorithms for Microwave Radar-Based Breast Imaging: Focal Quality Metrics as Fitness Functions", *Sensors*, vol. 17, no. 12, p. 2823, 6th Dec. 2017.

[126] D. O'Loughlin, F. Krewer, M. Glavin *et al.*, "Focal quality metrics for the objective evaluation of confocal microwave images", *Int. J. Microw. Wirel. Technol.*, vol. 9, no. 7, pp. 1365–1372, Sep. 2017.

[127] D. O'Loughlin, F. Krewer, M. Glavin *et al.*, "Estimating average dielectric properties for microwave breast imaging using focal quality metrics", in *10th European Conference on Antennas and Propagation (EuCAP)*, 10th–15th Apr. 2016, pp. 1–5.

[128] D. O'Loughlin, M. Glavin, E. Jones *et al.*, "Optimisation of confocal microwave breast images using image focal metrics", in *22nd Bioengineering in Ireland (BINI)*, Galway, Ireland, 22nd–23rd Jan. 2016, pp. 22–23.

[129] M. A. Elahi, D. O'Loughlin, B. Lavoie *et al.*, "Evaluation of image reconstruction algorithms for confocal microwave imaging: Application to patient data", *Sensors*, vol. 18, no. 6, p. 1678, 2018.

[130] M. Klemm, I. J. Craddock, A. W. Preece *et al.*, "Evaluation of a hemispherical wideband antenna array for breast cancer imaging", *Radio Sci.*, vol. 43, no. 6, pp. 1–15, Dec. 2008.

[131] E. C. Fear and M. A. Stuchly, "Microwave detection of breast tumors: Comparison of skin subtraction algorithms", in *Subsurface Sensing Technologies and Applications II*, vol. 4129, International Society for Optics and Photonics, 6th Jul. 2000, pp. 207–218.

[132] E. J. Bond, X. Li, S. C. Hagness *et al.*, "Microwave imaging via space-time beamforming for early detection of breast cancer", in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Orlando, FL, USA, 13th–17th May 2002, pp. III-2909-2912.

[133] D. O'Loughlin, M. O'Halloran, B. M. Moloney *et al.*, "Microwave Breast Imaging: Clinical Advances and Remaining Challenges", *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2580–2590, 2018.

[134] J. Bourqui and E. C. Fear, "System for Bulk Dielectric Permittivity Estimation of Breast Tissues at Microwave Frequencies", *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 9, pp. 3001–3009, Sep. 2016.

[135] A. T. Stavros, D. Thickman, C. L. Rapp *et al.*, "Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions", *Radiology*, vol. 196, no. 1, pp. 123–134, 1st Jul. 1995.

[136] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels *et al.*, "Measures of acutance and shape for classification of breast tumors", *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 799–810, 1997.

[137] G. Rahbar, A. C. Sie, G. C. Hansen *et al.*, "Benign versus malignant solid breast masses: US differentiation", *Radiology*, vol. 213, no. 3, pp. 889–894, Dec. 1999.

[138] M. L. R. Delgado, "Breast Stromal Tumors and Carcinoma with Sarcomatous Metaplasia. Radiological-Pathological Correlation", presented at the European Congress of Radiology, Vienna, Austria, 7th–11th Mar. 2013.

[139] A. Kerhet, M. Raffetto, A. Boni *et al.*, "A SVM-based approach to microwave breast cancer detection", *Eng. Appl. Artif. Intell.*, Special Issue on Engineering Applications of Neural Networks - Novel Applications of Neural Networks in Engineering Special Issue on Engineering Applications of Neural Networks - Novel Applications of Neural Networks in Engineering, vol. 19, no. 7, pp. 807–818, Oct. 2006.

[140] R. C. Conceição, D. Byrne, J. A. Noble *et al.*, "Initial study for the investigation of breast tumour response with classification algorithms using a microwave radar prototype", in *10th European Conference on Antennas and Propagation (EuCAP)*, Davos, Switzerland, 10th–15th Apr. 2016, pp. 1–2.

[141] D. Byrne, M. O'Halloran, E. Jones *et al.*, "Support Vector Machine-Based Ultrawideband Breast Cancer Detection System", *J. Electromagn. Waves Appl.*, vol. 25, no. 13, pp. 1807–1816, 1st Jan. 2011.

[142] D. Byrne, M. O'Halloran, M. Glavin *et al.*, "Breast Cancer Detection Based on Differential Ultrawideband Microwave Radar", *Prog. Electromagn. Res. M*, vol. 20, pp. 231–242, 2011.

[143] J. Sacristán, B. L. Oliveira and S. Pistorius, "Classification of electromagnetic signals obtained from microwave scattering over healthy and tumorous breast models", in *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Vancouver, Canada, 15th–18th May 2016, pp. 1–5.

[144] A. Santorelli, E. Porter, E. Kirshin *et al.*, "Investigation of Classifiers for Tumor Detection with an Experimental Time-Domain Breast Screening System", *Prog. Electromagn. Res.*, vol. 144, pp. 45–57, 2014.

[145] Y. Li, A. Santorelli, O. Laforest *et al.*, "Cost-sensitive ensemble classifiers for microwave breast cancer detection", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, 19th–24th Apr. 2015, pp. 952–956.

[146] Y. Li, A. Santorelli and M. Coates, "Comparison of microwave breast cancer detection results with breast phantom data and clinical trial data: Varying the number of antennas", in *10th European Conference on Antennas and Propagation (EuCAP)*, Davos, Switzerland, 10th–15th Apr. 2016, pp. 1–5.

[147] Y. Li, E. Porter, A. Santorelli *et al.*, "Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation", *Biomed. Signal Process. Control*, vol. 31, pp. 366–376, Jan. 2017.

[148] H. Song, Y. Li and A. Men, "Microwave breast cancer detection using time-frequency representations", *Med. Biol. Eng. Comput.*, vol. 56, no. 4, pp. 571–582, 1st Apr. 2018.

[149] E. Zastrow, S. K. Davis, M. Lazebnik *et al.*, "Development of anatomically realistic numerical breast phantoms with accurate dielectric properties for modeling microwave interactions with the human breast", *IEEE Trans. Biomed. Eng.*, vol. 55, no. 12, pp. 2792–800, Dec. 2008.

[150] M. O'Halloran, R. C. Conceição, D. Byrne *et al.*, "FDTD modeling of the breast: A review", *Prog. Electromagn. Res. B*, vol. 18, pp. 1–24, 2009.

[151] E. Zastrow, S. K. Davis, M. Lazebnik *et al.* (). Database of 3D Grid-Based Numerical Breast Phantoms for use in Computational Electromagnetics Simulations, [Online]. Available: `http://uwcem.ece.wisc.edu/MRIdatabase/`.

[152] D. Flores-Tapia, M. O'Halloran and S. Pistorius, "A Bimodal Reconstruction Method for Breast Cancer Imaging", *Prog. Electromagn. Res.*, vol. 118, pp. 461–486, 2011.

[153] M. Sarafianou, I. J. Craddock and T. Henriksson, "Towards Enhancing Skin Reflection Removal and Image Focusing Using a 3-D Breast Surface Reconstruction Algorithm", *IEEE Trans. Antennas Propag.*, vol. 61, no. 10, pp. 5343–5346, Oct. 2013.

[154] B. L. Oliveira, M. Glavin, E. Jones *et al.*, "Avoiding unnecessary breast biopsies: Clinically-informed 3D breast tumour models for microwave imaging applications", in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Memphis, TN, USA, 6th–11th Jul. 2014, pp. 1143–1144.

[155] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[156] (). The Cancer Imaging Archive (TCIA) — A growing archive of medical images of cancer, [Online]. Available: `www.cancerimagingarchive.net/`.

[157] M. Lazebnik, E. L. Madsen, G. R. Frank *et al.*, "Tissue-mimicking phantom materials for narrowband and ultrawideband microwave applications", *Phys. Med. Biol.*, vol. 50, no. 18, pp. 4245–4258, 21st Sep. 2005.

[158] E. Porter, J. Fakhoury, R. Oprisor *et al.*, "Improved tissue phantoms for experimental validation of microwave breast cancer detection", in *4th European Conference on Antennas and Propagation (EuCAP)*, Barcelona, Spain, 12th–16th Apr. 2010, pp. 1–5.

[159] A. Mashal, F. Gao and S. C. Hagness, "Heterogeneous Anthropomorphic Phantoms with Realistic Dielectric Properties for Microwave Breast Imaging Experiments", *Microw. Opt. Technol. Lett.*, vol. 53, no. 8, pp. 1896–1902, Aug. 2011.

[160] S. Romeo, L. Di Donato, O. M. Bucci *et al.*, "Dielectric characterization study of liquid-based materials for mimicking breast tissues", *Microw. Opt. Technol. Lett.*, vol. 53, no. 6, pp. 1276–1280, 1st Jun. 2011.

[161] N. Joachimowicz, C. Conessa, T. Henriksson *et al.*, "Breast Phantoms for Microwave Imaging", *IEEE Antennas Wireless Propag. Lett.*, vol. 13, pp. 1333–1336, 2014.

[162] M. J. Burfeindt, T. J. Colgan, R. O. Mays *et al.*, "MRI-Derived 3-D-Printed Breast Phantom for Microwave Breast Imaging Validation", *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 1610–1613, 2012.

[163] N. Joachimowicz, B. Duchêne, C. Conessa *et al.*, "Easy-to-produce adjustable realistic breast phantoms for microwave imaging", in *10th European Conference on Antennas and Propagation (EuCAP)*, Davos, Switzerland, 10th–15th Apr. 2016, pp. 1–4.

[164] T. Rydholm, A. Fhager, M. Persson *et al.*, "Effects of the Plastic of the Realistic GeePS-L2S-Breast Phantom", *Diagnostics*, vol. 8, no. 3, 2018.

[165] J. Garrett and E. Fear, "A New Breast Phantom With a Durable Skin Layer for Microwave Breast Imaging", *IEEE Trans. Antennas Propag.*, vol. 63, no. 4, pp. 1693–1700, Apr. 2015.

[166] A. Santorelli, O. Laforest, E. Porter *et al.*, "Image classification for a time-domain microwave radar system: Experiments with stable modular breast phantoms", in *9th European Conference on Antennas and Propagation (EuCAP)*, Lisbon, Portugal, 13th–17th Apr. 2015, pp. 1–5.

[167] J. Moll, D. Wörtge, V. Krozer *et al.*, "Quality control of carbon-rubber tissue phantoms: Comparative MRI, CT, X-ray and UWB microwave measurements", in *11th European Conference on Antennas and Propagation (EuCAP)*, Paris, France, 19th–24th Mar. 2017, pp. 2723–2727.

[168] G. Fiaschetti, J. Browne, M. Cavagnaro *et al.*, "Tissue Mimicking Materials for Multi-modality Breast Phantoms", in *2nd URSI Atlantic Radio Science Conference (URSI AT-RASC)*, Gran Canaria, 28th May-1st June 2018.

[169] L. Farrugia, P. S. Wismayer, L. Z. Mangion *et al.*, "Accurate in vivo dielectric properties of liver from 500 MHz to 40 GHz and their correlation to ex vivo measurements", *Electromagn. Biol. Med.*, vol. 35, no. 4, pp. 365–373, 1st Oct. 2016.

[170] J. Bonello, L. Farrugia and C. V. Sammut, "Effects of preservative solutions on the dielectric properties of biological tissue", in *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Verona, Italy, 11th–15th Sep. 2017, pp. 1216–1219.

[171] T. Yilmaz, M. A. Kılıç, M. Erdoğan *et al.*, "Machine learning aided diagnosis of hepatic malignancies through in vivo dielectric measurements with microwaves", *Phys. Med. Biol.*, vol. 61, no. 13, pp. 5089–5102, 2016.

[172] E. Porter, A. L. Gioia, A. Santorelli *et al.*, "Modeling of the dielectric properties of biological tissues within the histology region", *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 5, pp. 3290–3301, Oct. 2017.

[173] A. Peyman, C. Gabriel and E. H. Grant, "Complex permittivity of sodium chloride solutions at microwave frequencies", *Bioelectromagnetics*, vol. 28, no. 4, pp. 264–274, May 2007.

[174] H. Ulger, N. Erdogan, S. Kumanlioglu *et al.*, "Effect of age, breast size, menopausal and hormonal status on mammographic skin thickness", *Skin Res. Technol.*, vol. 9, no. 3, pp. 284–289, 1st Aug. 2003.

[175] S. A. Willson, E. J. Adam and A. K. Tucker, "Patterns of breast skin thickness in normal mammograms", *Clin. Radiol.*, vol. 33, no. 6, pp. 691–693, 1st Jan. 1982.

[176] T. L. Pope, M. E. Read, T. Medsker *et al.*, "Breast skin thickness: Normal range and causes of thickening shown on film-screen mammography", *J. Can. Assoc. Radiol.*, vol. 35, no. 4, pp. 365–368, Dec. 1984.

[177] J. Baker-Jarvis, M. D. Janezic, P. D. Domich *et al.*, "Analysis of an open-ended coaxial probe with lift-off for nondestructive testing", *IEEE Trans. Instrum. Meas.*, vol. 43, no. 5, pp. 711–718, Oct. 1994.

[178] E. C. Burdette, F. L. Cain and J. Seals, "In Vivo Probe Measurement Technique for Determining Dielectric Properties at VHF through Microwave Frequencies", *IEEE Trans. Microw. Theory Techn.*, vol. 28, no. 4, pp. 414–427, Apr. 1980.

[179] B. L. Oliveira, D. Godinho, M. O'Halloran *et al.*, "Diagnosing Breast Cancer with Microwave Technology: Remaining challenges and potential solutions with machine learning", *Diagnostics*, vol. 8, no. 2, pp. 1–22, 2018.

[180] M. A. Elahi, A. Shahzad, M. Glavin *et al.*, "Hybrid Artifact Removal for Confocal Microwave Breast Imaging", *IEEE Antennas Wireless Propag. Lett.*, vol. 13, pp. 149–152, 2014.

[181] M. A. Elahi, M. Glavin, E. Jones *et al.*, "Adaptive artifact removal for selective multistatic microwave breast imaging signals", *Biomed. Signal Process. Control*, vol. 34, pp. 93–100, 1st Apr. 2017.

[182] I. Guyon, A. Saffari, G. Dror *et al.*, "Model Selection: Beyond the Bayesian/Frequentist Divide", *Journal of Machine Learning Research*, vol. 11, pp. 61–87, Jan 2010.

[183] G. C. Cawley and N. L. C. Talbot, "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation", *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Jul 2010.

[184] C. Ambroise and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 10, pp. 6562–6566, 14th May 2002.

[185] J. R. Buck, M. M. Daniel and A. Singer, *Computer Explorations in Signals and Systems Using MATLAB*. Upper Saddle River, NJ: Prentice Hall, 2002, 216 pp.

[186] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005, 490 pp.

[187] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms", *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[188] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method", *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[189] S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications", *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 360–371, Jan. 2006.

[190] L. Breiman, "Random Forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 1st Oct. 2001.

[191] L. Breiman, J. H. Friedman, R. A. Olshen *et al.*, *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall, 1984, 366 pp.

[192] K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning", in *Proceedings of the Sixth International Workshop on Machine Learning*, Ithaca, New York, 26th–27th Jun. 1989, pp. 160–163.

[193] B. R. Lavoie, M. Okoniewski and E. C. Fear, "Estimating the Effective Permittivity for Reconstructing Accurate Microwave-Radar Images", *PLoS One*, vol. 11, no. 9, e0160849, 9-Sep-2016.

[194] D. O'Loughlin, B. L. Oliveira, A. Santorelli *et al.*, "Sensitivity and Specificity Estimation using Patient-Specific Microwave Imaging in Diverse Experimental Breast Phantoms", *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 303–311, 2019.

[195] D. O'Loughlin, B. L. Oliveira, M. Glavin *et al.*, "Advantages and Disadvantages of Parameter Search Algorithms for Permittivity Estimation for Microwave Breast Imaging", in *13th European Conference on Antennas and Propagation (EuCAP)*, Krakow, Poland, 31st Mar.–5th Apr. 2019.

[196] D. O'Loughlin, M. Glavin, E. Jones *et al.*, "Evaluation of experimental microwave radar-based images: Evaluation criteria", in *IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, Boston, USA: IEEE, 8th–13th Jul. 2018, pp. 895–896.

[197] D. O'Loughlin, B. L. Oliveira, M. Glavin *et al.*, "Effects of Interpatient Variance on Microwave Breast Images: Experimental Evaluation", in *40th Annual International Conference of the 40th IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA: IEEE, 17th–21st Jul. 2018.

[198] B. L. Oliveira, D. O'Loughlin, M. O'Halloran *et al.*, "Microwave Breast Imaging: Experimental tumour phantoms for the evaluation of new breast cancer diagnosis systems", *Biomed. Phys. Eng. Express*, vol. 4, no. 2, p. 025 036, 2018.

[199] H. Bahramiabarghouei, E. Porter, A. Santorelli *et al.*, "Flexible 16 Antenna Array for Microwave Breast Cancer Detection", *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2516–2525, Oct. 2015.

[200] D. O'Loughlin, M. A. Elahi, E. Porter *et al.*, "Open-Source Software for Microwave Radar-Based Image Reconstruction", in *12th European Conference on Antennas and Propagation (EuCAP)*, London, the UK: IEEE, 9th–13th Apr. 2018.

[201] R. Nilavalan, A. Gbedemah, I. J. Craddock *et al.*, "Numerical investigation of breast tumour detection using multi-static radar", *Electronics Letters*, vol. 39, no. 25, pp. 1787–1789, 2003.

[202] M. A. Elahi, C. F. Curtis, E. Jones *et al.*, "Detailed evaluation of artifact removal algorithms for radar-based microwave imaging of the breast", in *USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, Vancouver, BC, Canada, 19th–24th Jul. 2015, p. 307.