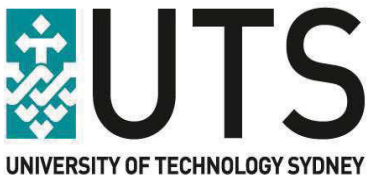

DIELECTRIC MODELLING OF HUMAN SKIN AND BREAST TISSUE IN TERAHERTZ FREQUENCIES: POTENTIAL APPLICATION TO CANCER DETECTION

CAO QUANG BAO TRUONG

THIS THESIS IS SUBMITTED
FOR THE REQUIREMENT OF DOCTOR OF PHILOSOPHY TO
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGY
UNIVERSITY OF TECHNOLOGY SYDNEY

December 2015

Acknowledgments

First of all, I would like to show my gratitude to my principal supervisor, Professor Tuan Hoang. Without his great support and enthusiastic supervision, my PhD study would not have been completed.

I am also deeply grateful to Professor Hung Nguyen, my co-supervisor, for his continuous encouragement, inspiration and valuable advice.

I would like to thank my research collaborators, Associate Professor Vincent Wallace and Associate Professor Anthony Fitzgerald for providing data and professional input into my PhD research.

Special thanks is extended to send my special thanks to my friends and colleagues in the Center for Health Technologies, University of Technology, Sydney. Not only were they helpful but they made my PhD study joyful and memorable.

Last but not least, I owe my deepest gratitude to my family. They have always given me so much support and I drew strongly upon this in order to overcome the most challenging periods of my PhD candidate.

Certificate of Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature _____

Name CAO QUANG BAO TRUONG

Contents

Abstract

Growing developments in the generation and detection of terahertz (THz) radiation over more than two decades have created a strong incentive for researchers to study the biomedical applications of terahertz imaging. Contrasts in the THz images of various types of cancer, especially skin and breast cancer, are associated with changes in the dielectric properties of cancerous tissues. In fact, dielectric models can explain the interaction between terahertz radiation and human tissue at a molecular level just as their parameters have the potential for becoming indicators of cancer. However, dielectric modelling of various forms of human tissue remains limited due to a number of factors, especially suboptimal fitting algorithms and tissue heterogeneity.

Thanks to the high water content of human skin, its dielectric response to terahertz radiation can be described by the double Debye model. The existing fitting method using a nonlinear least square algorithm can extract the model parameters which track their measurements accurately at frequencies higher than one THz but poorly at lower frequencies. However, the majority of dielectric contrast between normal and cancerous skin tissues has been observed in the low THz range. Accordingly, this research has developed two global optimization algorithms which are capable of globally accurate tracking thereby supporting the full validity of the double Debye model in simulating the dielectric spectra of human skin in the THz frequencies. Numerical results confirm their superiority over the conventional methods. Furthermore, the next goal of the study is to apply statistical analysis to the parameters of the double Debye model in

order to test their discrimination capability of skin cancer from normal tissue. Linear programming and support vector machine algorithms have also been employed using these parameters to classify normal skin tissue and basal cell carcinoma. By combining the double Debye parameters, the classification accuracy has shown significant improvement. The encouraging outcomes confirm the classification potential of the double Debye parameters.

The double Debye model, however, has been shown to be not suitable for simulating human breast tissue due to its low water content and heterogeneous structure, thus limiting the understanding of the THz dielectric response of breast tissue. To overcome this problem, this study proposes a new non-Debye dielectric model to fit the dielectric spectra of human breast tissue. Due to the mathematical complexity of the fitting procedure, a sampling gradient algorithm of non-smooth optimization is used to optimize the fitting solution. Simulation results confirm applicability of the non-Debye model through its exceptional ability to fit the examined data. Statistical measures have also been used to analyse the possibility of using the parameters of this model to differentiate breast tumours from healthy breast tissue. Based on the statistical analysis, popular classification methods such as support vector machines and Bayesian neural network have also been applied to examine these parameters and their combinations for breast cancer classification. The obtained classification accuracies indicate the classification potential of the model parameters as well as highlighting several valuable features of the parameter combinations.

List of Figures

List of Tables

Notation

THz	Terahertz
NMSC	Nonmelanoma Skin Cancer
BCC	Basal Cell Carcinoma
SCC	Squamous Cell Carcinoma
MMS	Moh's micrographic surgery
BCS	Breast conserving surgery
NS	Normal skin
$A \succeq 0$	A is A positive semi-definite matrix
$A \geq 0$	A is positive, i.e all its entries are positive
$x \geq p$	$x_i \geq p_i$ with $x \in R^n$, $p \in R^n$, $i = 1, 2, ..., n$
$x \in [p, q]$	$x_i \in [p_i, q_i]$ with $x \in R^n$, $p \in R^n$, $p \in R^n$, $i = 1, 2, ..., n$
$x \in R^n_+$	$\{x \in R^n \ : \ x_i \geq 0, \ i = 1, 2, ..., n\}$
$\tilde{x}$	x is a complex variable
$\Re(x)$	Real part of a complex variable x
$\Im(x)$	Imaginary part of a complex variable x
DD	Double Debye
THz-nDB	The non-Debye model for the human breast in the terahertz regime
NLS	Nonlinear Least Square
BB-GO	Global optimisation based on the branching and bounding method
BB-MNO	Minimax optimisation based on the branching and bounding method

QP	Quadratic program
R^2_{adj}	Adjusted R^2 indicates the goodness-of-fit
GOF	Goodness-of-fit
SVM	Support vector machine
HN	Havriliak-Negami relationship
CC	Cole-Cole equation
AVR	Average value
SE	Standard error
ROC	Receiver operating characteristic
AUROC	Area under the ROC curve
TPR	True positive rate
FPR	False positive rate
CV	Cross-validation
LOOCV	Leave-one-out cross-validation
RRSCV	Repeated random-subsampling cross-validation
RBF	Radial basic function
FDTD	Finite difference time domain
BNN	Bayesian neural network
NN	Neural network

Author's Publications

The contents of this thesis are based on the following papers that have been published, accepted, or submitted to peer-reviewed journals and conferences.

Journal Papers:

1. Bao C. Q. Truong, Hoang D. Tuan, H. H. Kha, and Hung T. Nguyen, "Debye parameter extraction for characterizing interaction of terahertz radiation with human skin tissue", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 6, pp. 1528-1537, June 2013.

2. Bao C. Q. Truong, Hoang D. Tuan, Anthony J. Fitzgerald, Vincent P. Wallace, and Hung T. Nguyen, "A dielectric model of human breast tissue in terahertz regime", *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 699-707, Feb 2015.

3. Bao C. Q. Truong, Hoang D. Tuan, Anthony J. Fitzgerald, Vincent P. Wallace, and Hung T. Nguyen, "The potential of the double Debye parameters to discriminate between basal cell carcinoma and normal skin", *IEEE Transactions on Terahertz Science and Technology*, vol. 5, no. 6, pp. 990-998, Nov 2015.

Conference Papers:

1. Bao C. Q. Truong, Hoang D. Tuan, H. H. Kha, and Hung T. Nguyen,"System identification for terahertz wave's propagation and reflection in human skin", *Fourth International Conference on Communications and Electronics*, Hue, Vietnam, pp. 364-368, Aug 2012.

2. Bao C. Q. Truong, Hoang D. Tuan, H. H. Kha, and Hung T. Nguyen,"Global optimization for human skin investigation in terahertz", *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, USA, pp. 5474-5477, Aug 2012.

3. Bao C. Q. Truong, Hoang D. Tuan, and Hung T. Nguyen,"Near-infrared parameters extraction: a potential method to detect skin cancer", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Osaka, Japan, pp. 33 - 36, July 2013.

4. Bao C. Q. Truong, Hoang D. Tuan, Anthony J. Fitzgerald, Vincent P. Wallace, and Hung T. Nguyen,"High correlation of double Debye model parameters in skin cancer detection", *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, USA, pp. 718-721, Aug 2014.

5. Bao C. Q. Truong, Hoang D. Tuan, Anthony J. Fitzgerald, Vincent P. Wallace, Tuan Nghia Nguyen, and Hung T. Nguyen,"Breast cancer classification using extracted parameters from a terahertz dielectric model of human breast tissue", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Milan, Italy, pp. 2804-2807, Aug 2015.

Chapter 1

INTRODUCTION

Microscopic examination in histopathology allows for exploration of the structure and functional information of processed specimens with the aim of manifesting their diseases while macroscopic images of tissue are provided by current medical imaging techniques such as magnetic resonance imaging (MRI) or X-ray computed tomography (CT). Combining macroscopic and microscopic imaging advances in terms of resolution to precisely delineate margins of cancer remains a unsolved problem of medical imaging. Terahertz imaging is a promising technique that shows potential to overcome the problem [Pickwell-MacPherson and Wallace, 2009]. This chapter begins with describing the background of this doctoral research. It is followed by addressing problems, objectives, and contributions of the research. The thesis structure is provided at the end of the chapter.

1.1 Background

Terahertz (THz) radiation, or T-rays, defines the electromagnetic (EM) waves generated in the typical region of spectrum from 100 GHz to 10 THz. This frequency regime is extended from just beyond the microwaves region up to the far-infrared and partly mid-infrared region. The attractive properties of THz radiation include non-ionisation, low photon energy (below 40 meV) and high water absorption [Siegel, 2002]. The available sources for microwaves including electronic power supplies are only suitable for frequencies of up to few hundred GHz as their electronic circuits are unresponsive at the higher range [Pickwell and Wallace, 2006]. Frequency multipliers can theoretically allow for reaching the terahertz region but they are not efficient. On the other hand, optical sources such as semiconductor laser diodes utilising the inter-band transitions between atomic and molecular states are capable of producing radiations whose frequencies only extend down to the order of 10 THz due to increasing difficulties with optical designs [Pickwell and Wallace, 2006]. The lack of efficient and coherent sources and detectors for THz systems has left a gap in the intermediate region in terms of the exploration of the THz spectra [Siegel, 2002].

Bridging the THz gap started with the development of a photoconductive emitter using intense ultrashort optical pulses to generate optical signals with a time resolution of few picoseconds (ps) [Auston, 1975]. More than one decade after, the free-space THz transmitters and detectors for terahertz time-domain spectroscopy (THz-TDS) were demonstrated by Smith et al. [1988] and these introduced subpicosecond photoconducting dipole antennas capable of emitting and detecting electrical pulses of less than 1 ps. Both the amplitude and phase of the THz pulses can be coherently measured with an excellent achievable signal-to-noise of up to 10000:1 [van Exter and Grischkowsky, 1990]. This results in the possibility of the spectroscopic measurements of optical and dielectric properties of materials such as liquids, gases, semiconductors, or dielectrics without considering the Kramers-Kronig relations [Grischkowsky et al.,

1990; Nuss et al., 1991; Pedersen and Keiding, 1992; Harde and Grischkowsky, 1991]. Although the initial THz imaging system proves to be acceptable in terms of spatial resolution, it is still impractical due to its long acquisition time of a few minutes for each pixel. Hu and Nuss [1995] demonstrated the first practical teraherz imaging technique which allows for reducing the acquisition time down to less than 5 ms. The reported images of this study together with the improvements in the generation and detection of coherent THz signals have greatly inspired further developments of THz-TDS and imaging techniques, launched in the 1990s, as well as their applications.

Recent developments in the broadband-pulse generation of terahertz (THz) radiation and its detection has accelerated biomedical applications in the frequency range. THz imaging, namely terahertz pulsed imaging (TPI), has now emerged as a potential medical imaging technique to identify the contrast between various types of tissue [Pickwell-MacPherson and Wallace, 2009]. Its advances are based on the nature of THz radiation. Having longer wavelengths than those in the visible/near-infrared domain, THz waves propagate inside biological tissues with less scattering. Significantly low power levels of THz radiation in typical time-domain systems satisfy the recommended safety threshold for the human body [Berry et al., 2003]. Another distinct asset is the high sensitivity of THz to water/high water-content materials, making THz imaging a feasible tool for cancer detection [Wallace, Taday, Fitzgerald, Woodward, Cluff, Pye and Arnone, 2004]. The high absorption of THz radiation by water is considered to be a dominant agent producing visible contrast features in THz images between normal and cancerous tissues due to the higher water content of the tumour in comparison with normal tissue [Woodward et al., 2003]. It is postulated that Protein, RNA and DNA content affect the interaction of THz radiation with tissue, along with water, hence they present themselves as potential sources of contrast [Smye et al., 2001; Fitzgerald et al., 2006].

Previous studies have shown a contrast in THz images, both *in vivo* and *ex vivo*, of various types of tissue. Early measurements reported contrast images between muscle

and fat of porcine tissue [Hu and Nuss, 1995]. Skin dermis phantoms and human teeth were imaged by Arnone et al. [1999] and these also displayed the visible imaging contrast due to differences in tissue moisture. Taylor et al. [2008] recently found remarkable differences in THz reflection between healthy and burned porcine skin specimens and this suggested the potential of real time THz imaging based on reflectivity. A THz imaging system based on reflection was used to non-invasively map cancerous regions of human skin while *ex vivo* images of this cancer also supported the possibility of distinguishing between normal and cancerous tissues using the THz imaging technique [Woodward et al., 2002, 2003]. With the introduction of small hand-held THz devices, there is now the motivation to speed up the translation of this technology into the clinical setting [Parrott et al., 2011]. In addition to the skin cancer application, the THz imaging might be beneficial to breast-conserving surgery (BCS) as Fitzgerald et al. [2006] investigated the possibility of using THz pulses to correctly map breast tumour margins. Intraoperative THz imaging devices such as the handheld THz imaging probe, the continuous-wave single frequency THz imaging, or THz near-field microscopy, have been developed for detecting the tumour margins during BCS [Ashworth et al., 2008; Chen, Lee, Huang, Chiu, Tsai, Tseng, Lu, Lai and Sun, 2011; St.Peter et al., 2013]. Recent advances in THz technology have also triggered applications in other cancers occurring in less accessible areas such as the cervix, colon, or liver [Wang and Mittleman, 2012; Reid et al., 2011; Sy et al., 2010].

1.2 Motivation of Thesis

Basal cell carcinoma (BCC), a form of non-melanoma skin cancer (NMSC), is the most common cancer in the Caucasian population [Mogensen and Jemec, 2007]. As the diagnosis of BCC depends on visual examination, biopsy is necessary if the diagnosis is uncertain. Moh's micrographic surgery (MMS) is a technique which is used to ensure complete excision of the cancer with minimal excision of healthy tissue [Shriner et al.,

1998]. MMS achieves the highest cure rate of 96–99%; however, MMS is expensive in terms of time and money. This treatment modality could be simplified if there was a system which is capable of accurately delineating the BCC margins preoperatively without the need for biopsy. The high correlation between THz images and histologic regions of BCC found by Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone [2004] supports further studies seeking to apply THz imaging to assist with MMS.

Changes in absorption coefficient and refractive index are among the most important factors explaining the contrast in the THz images of skin cancer. Therefore, the complex permittivity as a combination of these two optical properties is also worth considering as a promising component for the discrimination between healthy and diseased tissues. A large proportion of water in human skin is the foundation for studying human skin dielectric permittivity in the THz regime [Smye et al., 2001; Pickwell, Cole, Fitzgerald, Pepper and Wallace, 2004]. For analytical modelling, the double Debye model (DD) has been employed to characterise dielectric behavior of human skin since it was first introduced by Liebe et al. [1991] to describe the frequency-dependent permittivity of liquid water [Pickwell, Cole, Fitzgerald, Wallace and Pepper, 2004]. This model is determined by a frequency-dependent function with five driving parameters that describe the dispersion and relaxation times of two Debye relaxation processes in the THz regime. Extracting these five parameters of the double Debye model is important to understand the interaction between THz radiation and human skin as the DD parameters are believed to be the origin of contrast in THz images of skin cancer [Pickwell, Cole, Fitzgerald, Wallace and Pepper, 2004; Taylor et al., 2011; Pickwell et al., 2005]. Indeed, the extracted DD parameters were found to be capable of providing interesting contrast between normal skin and BCC [Pickwell et al., 2005]. In addition, Reid et al. [2010] reported the possibility of using the DD parameters to specify the biological tissue pathology with a limited resolution. Despite the advantages of the double Debye model, the existing extraction procedure using nonlinear least-squares methods (NLS) is far from being optimal and consistent. Liebe et al. [1991]; Barthel and Buchner

[1991]; Kindt and Schmuttenmaer [1996] have successfully used the two Debye relaxation processes to model the liquid water complex permittivity in the frequency ranges up to 1 THz. Based on this, Pickwell, Cole, Fitzgerald, Wallace and Pepper [2004]; Pickwell et al. [2005] has considered the double Debye model for complex permittivities of either water or human skin in the frequency band extending to 2 THz. However, this model fitting is inferior to its counterparts in the narrower band in terms of consistent performance over a whole examined spectrum. To be more specific, the discrepancies between the measured complex permittivity and that simulated by the double Debye model from [Pickwell et al., 2005] at the frequencies below 0.8 THz are remarkably higher than those published in [Kindt and Schmuttenmaer, 1996]. Liebe et al. [1991] suggested an addition of two Lorentzian resonant processes to the double Debye model in order to maintain a good fit when examining the frequency range extending to 2 THz. These extras certainly not only add unnecessary complexity to extraction but also raise serious doubts about the validity of the double Debye model. The question as to whether the performance of double Debye models is adequate for modelling human skin complex permittivity in the THz frequency range remains open.

Terahertz imaging has also shown its capability to produce images which show the essential contrast between healthy and cancerous breast tissue [Fitzgerald et al., 2006; Hassan et al., 2012]. Compared to conventional imaging systems, the nature of the THz radiation provides this technique with viable advantages to noninvasively and nondestructively image breast tumours with high resolution [Ashworth et al., 2009]. The THz images suggested the potential of the THz imaging technique for accurately detecting margins of breast cancer in BCS. Increases in absorption and refractive index of breast tumour compared with healthy tissue are associated with the imaging contrast but its origin is still not available [Ashworth et al., 2009]. Therefore, similarly to skin cancer, modeling the dielectric function of breast tissue not only explains the physical characteristics underpinning the contrast in the THz images but also possibly introduces some classifiers of breast tumour.

The free water content of other human body parts may be considerably less than that of skin, which means that the dielectric responses of that tissue may not comply with the simple Debye relaxation. Alternatively, the increasing content of biological water, or hydration water, which is constructed by chains of protein and their surrounding water molecules, is likely to cause non-Debye responses in the dielectric spectra of tissue [Fitzgerald et al., 2014; Feldman et al., 2002]. Indeed, the double Debye model fails to obtain a requisite fit for complex permittivity of breast tissue due to an increase observed in the real part of the complex permittivity spectra of the human breast in the range below 1 THz [Ashworth et al., 2009]. However, a suitable dielectric model for breast tissue in THz regime is yet to be achieved in the literature.

1.3 Objectives and Contribution

Considering the aforementioned research problems, the ultimate purpose of this doctoral research is to improve the dielectric modelling of human tissue and tumour, especially in the skin and the breast, in the terahertz regime. In addition to the modelling, this study targets identifying potential indicators of cancer from extracted parameters of the dielectric models. The outcomes of this research will not only contribute towards a better theoretical framework in terms of understanding the interaction between THz radiation and human tissue in general, but will also support the clinical applicability of the terahertz imaging to cancer-margin delineation in MMS and BCS. The main tasks of this research will be accomplished by achieving the objectives outlined below.

The first objective is to improve the fitting quality of the double Debye model through optimization algorithms. The conventional sum of square error (SSE) function, on which the fitting is based, is definitely a complex nonlinear and nonconvex function of the DD parameters, thus the NLS-based approach can only locate a local solution which will satisfy some necessary optimality conditions. Therefore, several global optimization (GO) algorithms will be introduced in this thesis to find a globally optimal solution

for the fitting problem. The first proposed optimization algorithm employs grinding-based technique that consecutively searches and updates the optimal solutions at each grinding point of a finite two-dimensional space which is created by the constraints of the problem variables. This method leads to the optimised double Debye model for a very accurate description of skin complex permittivity in the frequency range [0.2–2] THz. Simulation results prove the superiority of the grinding-based GO algorithm over the NLS. The second algorithm is based on applying a brand-and-bound method of global optimization (BB-GO) in order to improve the computational efficiency and intensity of the fitting procedure while still assuring convergence of the optimal solution. Simulation outcomes indicate that the BB-GO not only is significantly more effective in computational time than the grinding-based algorithm but also guarantees the same fitting quality. An error analysis of the fitting still shows lower fitting errors at high THz frequencies than those at the low range. Thus, a modification of the BB-based method, which takes advantage of a minimax optimization (MNO), is also proposed to control the fitting-error balance between low and high parts of the examined THz domain. Despite the fact that this modified algorithm does not improve the overall performance, its simulations provide a better estimation at the low THz frequencies.

The second objective of the PhD study is to explore the cancer discrimination ability of the double Debye model through statistical measures and classification. A combination of statistical measures including t-test, normalised values of difference in mean, Pearson correlation, and receiver operating characteristics (ROC) curve has been used to examine the DD parameters and several frequency-dependent parameters related to the DD model. The statistical outcomes allow for accurate selection of the most informative parameters for skin cancer discrimination. linear programing (LP) and support vector machine (SVM) classifiers have also been employed to classify the skin samples using the best parameters. Positive outcomes of the classification provide the first concrete assessment about the potential of the double Debye model for skin cancer classification.

The third objective of my research is to propose a new non-Debye dielectric model

for human breast tissue (abbreviated by THz-nDB) in the terahertz region as well as analyse this model in terms of its applicability to breast cancer detection. A robust sampling gradient algorithm for the nonsmooth optimization problem has been employed to fit this model. According to simulation results, this model can accurately describe complex permittivities of both healthy and cancerous samples. Furthermore, good performance of this model also facilitates a deeper insight into the contrast mechanism at the molecular level. For the cancer discrimination potential of the THz-nDB, its eight model parameters have been investigated by statistical tools. The obtained results suggest some of these paramters are indicators of breast tumours due to its high correlation with the tumour percentage of the samples, the statistical significance of the difference between normal tissue and tumours, and the high values of the area under the ROC curve. The selected parameters and their combinations have been used with the SVM and Bayesian neural networks (BNN) classifiers to classify the breast samples. Both classification methods present high classification accuracies for the applied data, thus proving the classification potential of these parameters. Furthermore, BNN demonstrates better overall performance as compared to the SVM in terms of using the parameters of the THz-nDB model for breast cancer classification.

1.4 Structure of Thesis

The thesis includes six chapters together with references. The first chapter introduces the background in terms of how I developed my research, the current problems in the area, and the research question raised. It also clarifies the specific objectives and provides an overview of the contributions of this doctoral research. The contents of the following chapters are briefly outlined below.

Chapter 2 presents a review of THz imaging technologies. Firstly, it introduces the distinct properties of terahertz radiation. The generation and detection techniques of THz radiation are carefully pointed out in the next section. These are followed by a

section describing terahertz time-domain spectroscopy as a key spectroscopic technique and the foundation for developments in THz imaging. After this, a variety of terahertz imaging techniques are reviewed. Last but not least, important medical applications of terahertz imaging in skin cancer and breast cancer, which are also the focus of this thesis, are provided in detail.

Chapter 3 firstly outlines its main content. This is followed by a secton that introduces the double Debye model and the proposed optimization problem to fit this model. The next section sequentially demonstrates two global optimization algorithms based on grinding and brand-and-bound techniques to fit the double Debye model to its measured data of human skin. Simulation results and dicussion concerning them are presented in another section. This is followed by concluding statements.

Chapter 4 starts with a brief introduction of the research targets. The next section provides information on the applied data and extraction procedure. This is followed by a section providing statistical analysis of the extracted DD paramters. After this, the classification ability of these parameters is examined using LP and SVM algorithms. Classification results and concluding statements are respectively presented in the last two sections.

Chapter 5 proposes a new non-Debye dielectric model to describe the dielectric properties of human breast tissue. It begins with reviewing current dieletric models in literature and then introducing the modified non-Debye model. The next section demonstrates a robust gradient sampling algorithm to tackle the complexity of fitting this dieletric model. This is followed by a section that provides a statistical analysis with the aim of assessing the cancer discrimination potential of the model parameters. The performance of these parameters and their combinations in the SVM and BNN classification may be viewed in the next two sections. The contributions of this chapter are summarised in the final section.

Chapter 6 includes two sections. Firstly, the summary section focuses on summarizing

the studies in this doctoral research and their respective contributions. The final section addresses limitations and raises essential questions for future research in this area.

Chapter 2

LITERATURE REVIEW

2.1 Properties of Terahertz Radiation

Terahertz (1 THz = 10^{12} Hz) radiation, or T-ray, has a typical frequency range of between 100 GHz – 10 THz [Mittleman et al., 1996]. 1 THz corresponds to the wavelength of 300μm and the photon energy of 4.14 meV. The energy level of THz radiation is much less than that of X-rays, thus making this radiation non-ionised. Therefore, the safe T-ray for human body and is promising in terms of medical application [Berry et al., 2003]. It also means that frequently screening and monitoring patients by THz imaging may be possible if the imaging technology becomes available [Pickwell-MacPherson and Wallace, 2009]. The safety level of exposure has been studied on human keratinocyte and blood leukocytes [Clothier and Bourne, 2003; Zeni et al., 2007]. Sensitivity to water is another important property of THz radiation. Molecular resonances of intermolecular bonds, especially hydrogen bonds, extending into the THz region explain the high absorption of this radiation by water. This distinct property of T-ray is particularly beneficial for biomedical applications as biological tissue commonly has high water content.

2.2 Terahertz Generation and Detection

Development in optics and electronics have been significantly improving THz sources and detectors. A generated THz signal can take one of two forms including broadband or continuous-wave (CW) pulses. This section aims to provide brief descriptions of the popular techniques used for terahertz generation and detection.

2.2.1 Generation

Generating terahertz radiation is based on sending incident electromagnetic waves to a nonlinear medium which facilitates nonlinear conversion of these waves [Lee, 2008]. In

particular, frequency down conversion utilises the interaction of two optical photons at two different frequencies with a nonlinear crystal to create a terahertz photon. The first type of this nonlinear optical process is optical rectification that allows femtosecond laser pulses to have a broad spectrum with the bandwidth being up to 10 THz to generate broadband THz pulses [Carrig et al., 1995; Huber et al., 2000; Lee et al., 2001]. Difference frequency generation is another process to produce CW THz pulses from optical beams [Kawase et al., 1999]. For the up frequency conversion, diodes with nonlinear characteristics are used for frequency multiplication of microwaves to convert incident waves from solid-state microwave-based sources into continuous-wave THz pulses which are harmonic forms of these microwaves.

Another technique to generate terahertz radiation employs exciting a biased photoconductive (PC) antenna by femtosecond laser beams [Hamster et al., 1993; Zhang and Xu, 2009]. This is based on utilizing varying photocurrents and accelerating charges in time to radiate electromagnetic waves. A PC antenna is constructed by two metal electrodes attached to a semiconductor substrate. The femtosecond lasers focus on the gap between the electrodes to create electron-hole pairs in the semiconductor after transient polarization due to electrical shortening and applied-voltage reduction. The generated electrons and holes act as free photocarriers that are accelerated by a static bias field. This photocurrent, which is proportional to the intensity of the incident optical beams, results in generating the broadband THz pulses. To generate CW THz radiation, a photomixing technique can be applied to mix two frequency-offset laser beams in order to form an optical beat which is then radiated to the PC antenna [Verghese et al., 1997; Preu et al., 2011]. Ultrafast photoconductive material, such as low-temperature GaAs, are commonly used to make the photomixer and its operating mechanism involves conductivity modulation and conversion into a current under an applied electric field.

A free-electron laser (FEL) is also able to produce extremely coherent THz waves [Doria et al., 2002; Grosse, 2002]. It employs relativistic electrons and electron accelerators to

stimulate photon emission. An electron beam triggered by a femtosecond laser source is transmitted through an array of magnets which create a transverse and periodic magnetic field to force the electrons motion into a sinusoidal oscillation. The incident electrons are accelerated due to this oscillation and are able to emit photons of the THz radiation. The wavelength of the generated THz pulses depends on the applied magnetic field and the electron beam energy. A backward wave oscillator (BWO) is another free-electron-based source which can be considered as a laboratory version of FEL [Dobroiu et al., 2004; Chen et al., 2012]. However, it is used to generate tunable CW THz radiations. This device shares a similar operating mechanism to FEL but it uses a grating instead of a magnet array to change the speed of the electron beam. The electrons generated by a heated cathode are accelerated during their travel to the anode inside a vacuum tube. This is done through the impact of an external magnetic field and is thus able to radiate THz waves.

Compact solid-state THz sources such as diode-based frequency multipliers and quantum cascade lasers (QCL) are also available. The Schottky diodes can multiply incident microwaves to generate THz radiations [Porterfield et al., 1999]. The recent GaAs Schottky allows for tripling the frequency of the output waves [Porterfield, 2007]. QCLs have been recently developed thanks to advances in nanotechnology [Ajili et al., 2002; Williams, 2007]. They employ semiconductor heterostructures which are periodically made of consecutive layers of different semiconductors. Intersubband transitions due to electron propagation between these structures result in emitting THz radiations.

2.2.2 Detection

THz detection techniques are widely categorised into coherent and incoherent signal detection [Lee, 2008]. The former targets measuring amplitude and phase of the THz signals whereas the latter only focuses on the measurement of the signal intensity. Generally, coherent detection techniques have many similarities to generation techniques in

terms of the operating mechanism and its supporting components. The same light source can be used for either generation or detection.

Photoconductive detection and electrooptic (EO) sampling techniques are the most popular methods to detect THz signals. A photoconductive detector for broadband THz pulses shares the used substrate material and antenna structures with the generators [Jepsen et al., 1996; Zhang and Xu, 2009]. Its detection mechanism is based on using a current amplifier to read the dc photocurrent induced by cross correlation between the THz waves and the photocarriers. The EO sampling technique utilises the Pockets effect of an EO crystal to induce birefringence under the static electric field created by the THz waves [Wu and Zhang, 1995; Nahata et al., 1996]. Measuring the birefringence in terms of the delay between the THz waves and the incident optical beam enables the determination of the THz waveform.

Heterodyne receivers use frequency beat and down conversion to detect the amplitude and phase of CW THz waves [Hubers, 2008]. Their structure includes a local oscillator, a mixer and a spectrometer. The local oscillator, a CW THz source, produces a reference signal that together with the THz signal is converted into an intermediate frequency signal in the microwave range by the mixer. This microwave signal can be detected by the spectrometer after being amplified.

Thermal detectors such as bolometers, pyroelectric detectors, and Golay cells are commonly used for incoherent detection of THz signals [Chasmar et al., 1956; Porter, 1981; Golay, 1947]. They all contain a radiation absorber whose temperature will increase due to the incident THz wave. A thermometer is used to measure this temperature change. Each type of thermal detector has a specific thermometer structure. The response of these detectors is slow compared to the coherent detectors due to the equilibrium condition of the temperature measurement.

2.3 Terahertz Time-Domain Spectroscopy

Terahertz time-domain spectroscopy (THz-TDS) was first introduced by Auston and Cheung at AT & T Bell Laboratories in 1985 [Auston and Cheung, 1985]. The breakthrough advances in the generation and detection of THz radiation in the following decades ensured that this technique became one of the most powerful tools for THz spectroscopy. THz-TDS provides coherent measurements of the THz electric field in terms of its amplitude and phase information [Ferguson and Zhang, 2002]. Implementation of THz-TDS in the imaging system can map a number of compositions of the sample materials. This advantage of THz-TDS facilitates its application in real-time material detection and biomedical devices [Ferguson and Zhang, 2002]. Furthermore, since the coherent measurement of THz signals make it possible to recover the complex optical or dielectric constants of the sample, THz-TDS can provide structural information of the sample at the molecular level [Globus et al., 2003; Naftaly and Miles, 2007]. Applications of THz-TDS in material characterization cover a wide range of various fields including astronomy, illegal drugs and explosive detection, industrial quality control, and especially THz imaging [Naftaly and Miles, 2007; Beard et al., 2001; Kawase et al., 2003; Shen et al., 2005; Yasui et al., 2005; Hu and Nuss, 1995; Pickwell and Wallace, 2006].

Fig. 2.1 shows the structure of a typical THz-TDS system set up in the transmission geometry [Ferguson and Zhang, 2002]. This system employs a femtosecond laser source to produce a series of pulses typically being 1-2 ps in duration with a repetition rate of up to 100 MHz. These femtosecond laser pulses are used for the generation and detection of THz pulses. In particular, the laser beam is split by a beam splitter into two paths. One beam is used to generate the THz signal probing the sample whereas the other is utilised to gate the detector. Then, the THz electric field as a function of time can be captured at a optical delay between the two beams. Assuming that the electric field is unchanged, this measurement is repeated with different time delays. As

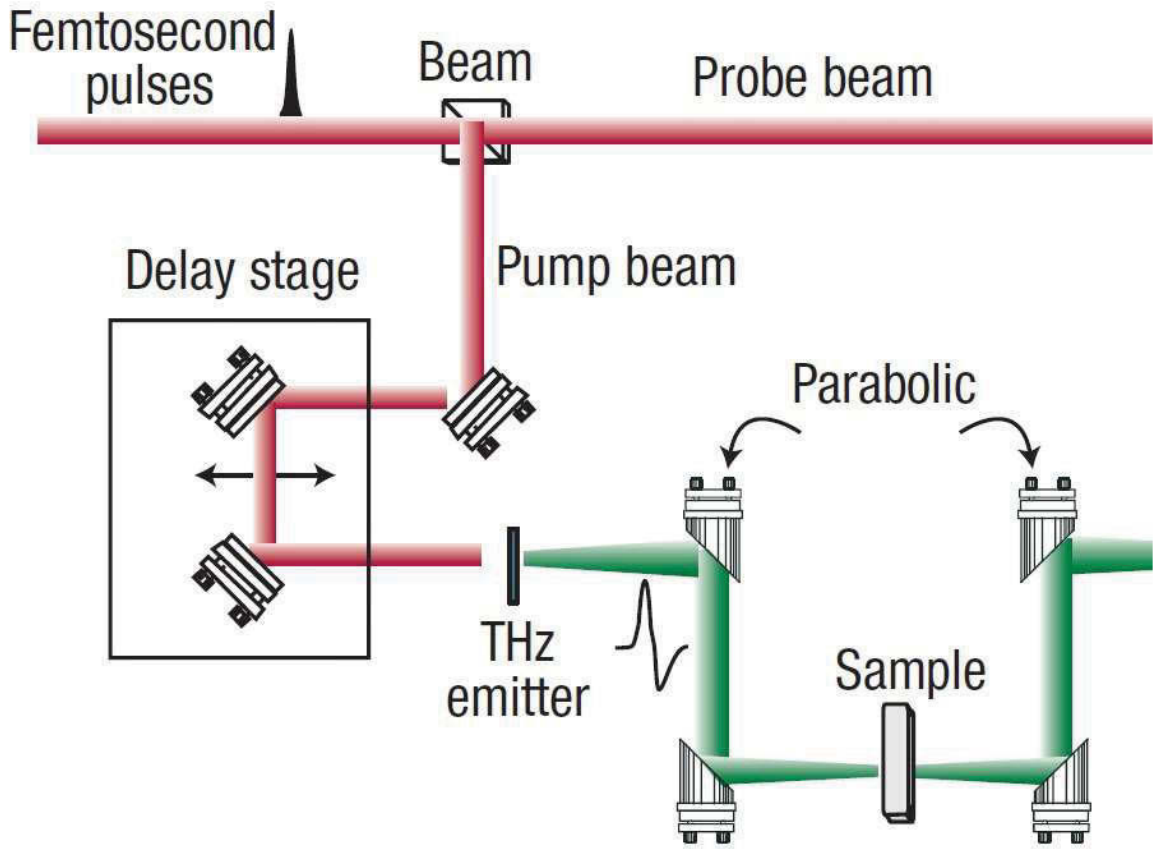

Figure 2.1: A THz-TDS system in transmission mode [Ferguson and Zhang, 2002].

a result, the time-domain waveform of the THz signal is obtained by connecting the multiple measurements in a time line of the optical delays. The spectrum of the THz pulses is obtained using the Fourier transform. The direct linkage of the THz amplitude and phase with the absorption coefficient and the refractive index of the sample enables the calculation of the real and imaginary parts of the complex permittivity.

THz-TDS possesses several remarkable advantages over other THz spectroscopic techniques such as far-infrared Fourier transform spectroscopy (FTS) [Beard et al., 2002; Chan et al., 2007]. The ability of THz-TDS to directly obtain the complex dielectric/optical constants results in eliminating the use of Kramers-Kronig relations whose indirect calculation in FTS potentially contains sources of error. Additionally, exploiting the time delays of THz waves passing through the sample in THz-TDS is beneficial

in terms of identifying the variation of the sample thickness [Mittleman et al., 1996]. THz-TDS is also more power-efficient, with the SNR $> 10^6$, than FTS as it only has around 300 in SNR [Beard et al., 2001]. Furthermore, employing highly coherent and short-pulse sources allows THz-TDS to achieve higher optical intensities in the order of magnitude as compared to the CW sources of FTS, thus making THz-TDS preferable for the measurements of absorption coefficients in transmission geometry [Beard et al., 2001]. In terms of the frequency resolution, THz-TDS can reach to approximately 1 GHz, which is comparable to the best of the FTS devices [Naftaly and Miles, 2007].

2.4 Terahertz Imaging

Hu and Nuss [1995] introduced the THz imaging system based on the THz-TDS technique in 1995. The operating mechanism of this initial system has been fundamental in terms of the development of subsequent THz imaging up until today. A THz-TDS-based imaging system (Fig. 2.2) commonly uses photoconductive antennas together with the EO generator and detector. Its image acquisition is based on raster-scanning techniques that raster the sample through the THz beam focus to acquire the imaging data at all pixels. Generally, the earliest system took approximately 10 minutes to obtain a 100×100 pixel image. The image acquisition rate has been enhanced in modern THz imaging systems by using more advanced methods to increase the speed of varying the optical delay. However, the variation speed in the delay line must be synchronised to the raster scan of the sample in order to identify the pixel location of the currently-obtained waveform. For image formation, various information collected from the measured THz electric fields in both the time and frequency domains can be used to form two-dimensional false-colour THz images. In particular, the time-domain pulses provide amplitude and phase information whereas their Fourier transformation offers many useful spectral components. Each part of the data can produce various images that demonstrate different information about the sample.

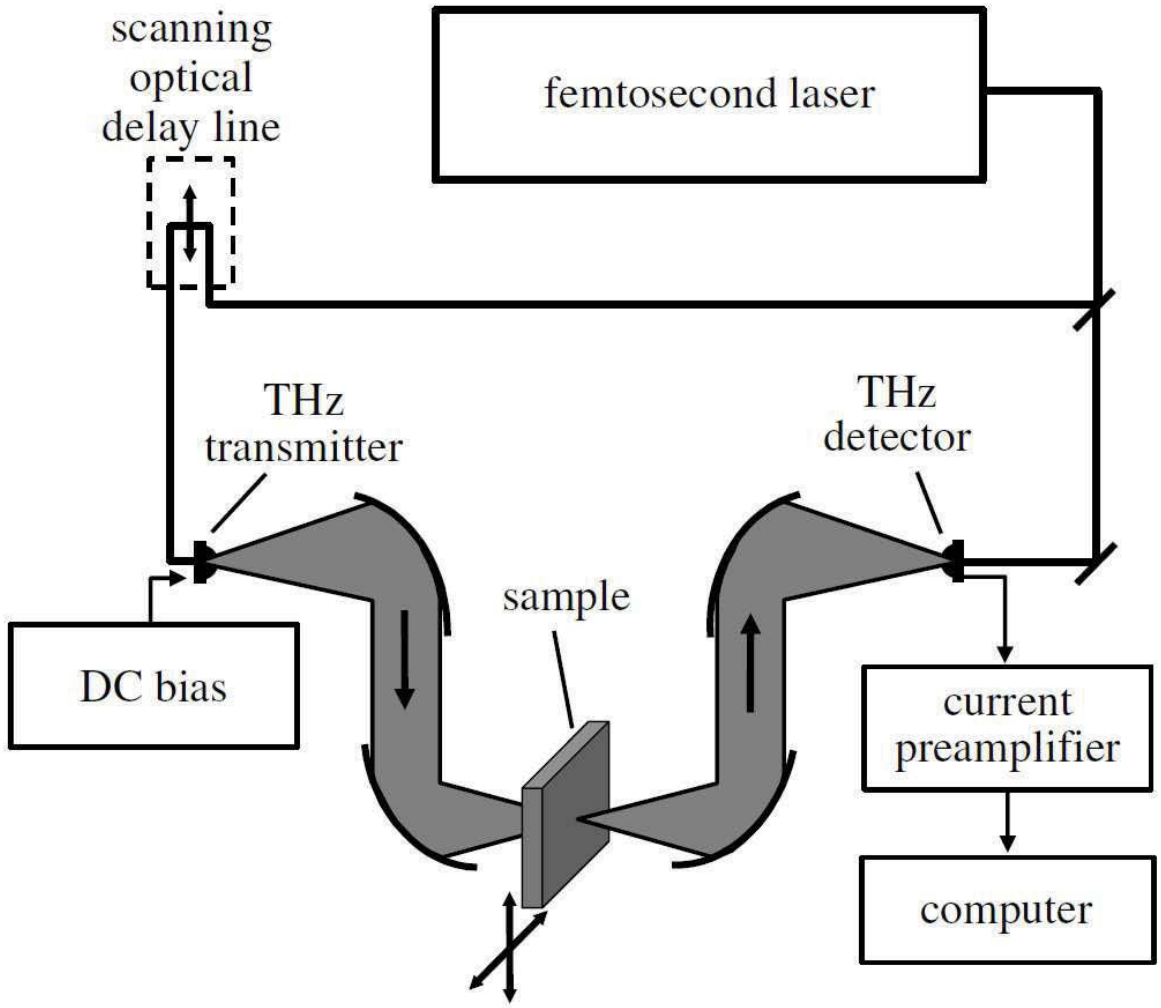

Figure 2.2: A typical time-domain imaging system using raster scan [Chan et al., 2007].

Recent expansion of the THz imaging leads to a number of developments in new imaging techniques [Siegel, 2002]. For instance, time-of-flight imaging and various imaging approaches based on tomographic and synthetic aperture configurations manipulate the low temporal coherence of THz pulses in the reflection geometry to form three-dimensional images of the sample [Mittleman et al., 1997; McClatchey et al., 2001; O'Hara and Grischkowsky, 2004; Ferguson and Zhang, 2002; Wang and Zhang, 2003]. Jiang and Zhang [1999] have employed the free-space EO sensing techniques in their pioneering studies of building a video-rate terahertz imaging system to overcome the limitation of the raster-scanning method on the image acquisition rate. The authors

have also introduced a chirped-pulse imaging scheme that allows for measuring the a full THz waveform with only a single femtosecond pulse from the laser source [Jiang and Zhang, 1998, 2000]. A variety of methods have been applied to implement the near-field imaging techniques in order to obtain sub-wavelength resolution [Hunsche et al., 1998; Mitrofanov et al., 2000; Chen et al., 2003; Schade et al., 2004; Wang et al., 2004]. Using CW THz radiation for imaging has been accompanied by developments in CW THz sources such as the invention of THz quantum cascade lasers or THz parametric oscillators [Darmo et al., 2004; Lu et al., 2005; Kim et al., 2006; Nguyen et al., 2006]. Numerous studies have reported the applications of THz imaging in a variety of areas ranging from security and medical diagnosis to monitoring processes and packing inspection in the food, material, and chemical industries [Wallace et al., 2006; Fitzgerald et al., 2006; Pickwell and Wallace, 2006; Mittleman et al., 1999; Shen et al., 2005; Kemp et al., 2003; Gowen et al., 2012; Zeitler et al., 2007; Tonouchi, 2007].

2.5 Medical Applications of Terahertz Imaging

The aforementioned advantages of THz radiation in terms of non-ionization, safe levels of emission energy, and sensitivity to hydrated materials such as biological tissues make THz imaging a very promising medical tool for cancer diagnosis. Current terahertz imaging systems such as THz pulsed imaging have been typically developed based on the THz-TDS technique [Yu et al., 2012]. In these systems, coherent THz pulses are used to characterise tissue composition as well as to produce either two-dimensional or three-dimensional images of biological samples. The THz imaging systems can be designed to detect the transmitted THz signal through the sample or the reflection off the sample as can be seen in Fig. 2.3 [Pickwell-MacPherson and Wallace, 2009]. The transmission mode requires well-sliced samples with a sufficiently-small thickness to ensure the measurement of the transmitted THz wave as this radiation is strongly absorbed by water inside the biological tissue. However, the structure of tissue can be

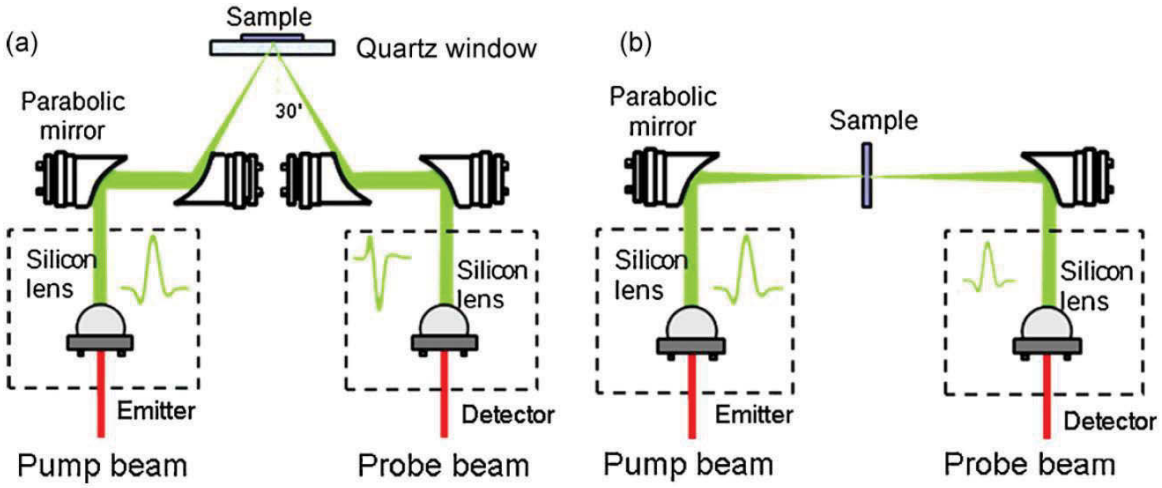

Figure 2.3: A schematic of the terahertz imaging system operating in the reflection (a) and transmission (b) geometries [Pickwell-MacPherson and Wallace, 2009].

damaged by slicing whereas the thin sample dehydrates more easily. These problems may affect the measurement accuracy. On the contrary, measurements in the reflection geometry prefer samples that are sufficiently thick to eliminate multiple reflections within the samples. Assuming that the sample is homogeneous and has an appropriate thickness, both the measurement modes are able to determine the optical/dielectric properties of the samples [Png et al., 2008; Huang, Ashworth, Kan, Chen, Wallace, ting Zhang and Pickwell-MacPherson, 2009; Huang, Wang, Yeung, Ahuja, Zhang and Pickwell-MacPherson, 2009]. By investigating the dielectric properties, intermolecular vibrations between molecules inside the samples can be revealed [Pickwell-MacPherson and Wallace, 2009]. These molecular interactions occur when attractive intermolecular forces such van der Waals forces between neighbouring molecules are weak. Since the structure and function of biomolecules, such as proteins, are heavily dependent on these forces, the signatures of various protein can be read by probing the THz spectra [Fischer et al., 2002]. THz imaging has been considered as a potential tool for a number of medical applications ranging from imaging skin burns to cancer diagnosis [Mittleman et al., 1999; Brun et al., 2010; Yu et al., 2012; Zaytsev et al., 2014; Fan et al., 2014].

2.5.1 Skin Cancer Detection

2.5.1.1 Overview of Skin Cancer

According to the Cancer Council Australia, skin cancer is one of the most common diagnosed cancers, and this accounts for approximately 80% of all newly diagnosed cancers. The statistics in 2007 show that more than 444,000 people need treatments for various types of skin cancer each year and the disease causes more than 1,700 deaths in Australia [AIHW & AACR, 2012]. Common types of skin cancer are divided into two groups: cutaneous malignant melanoma and non-melanoma skin cancer (NMSC) [Diepgen and Mahler, 2002]. Melanoma is the fourth most common cancer in Australia and New Zealand and has the highest mortality [Diepgen and Mahler, 2002; Elder, 1995].

NMSC is the most prevalent skin cancer in the light-skinned population and includes basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). The incidence rate of NMSCs are rapidly increasing due to a variety of causes, i.e. increased exposure to UV light. UV radiation can directly damage DNA, which contributes to skin mutation. The tumorigenesis is affected by not only solar exposure but also the constitutional predisposition, e.g. skin color, freckle spread, genetic history related to melanoma, and age.

The key factor in dealing with skin cancer is early detection of the disease. This prevents the further penetration of cancerous tissue into deeper skin layers and thus enables early treatments such as surgical excision [Rigel et al., 2010]. The current statement of skin cancer diagnosis is dependent on evaluating the clinical examination of biopsy specimens to explore the histopathological details of the patient [Stockfleth et al., 2010]. Despite the fact that the biopsy is considered as a gold standard for diagnosis, it may be time-consuming, expensive, and cause pain for patients [Hallock and Lutz, 1998]. This diagnostic procedure is augmented by advances in non-invasive imaging techniques and

computerisation of cellular figures. Furthermore, either traditional treatment, whether based on surgery or new non-invasive treatment strategies depends on identifying the tumour margin to guarantee complete removal of cancerous cells. Repeated biopsy examination may be necessary if the correct margin is not detected. This fact alone calls for non-invasive diagnostic tools. Medical imaging techniques are among the promising diagnostic tools for non-invasively detecting cancerous cells and mapping the tumour margin.

2.5.1.2 Existing Imaging Techniques of Skin Cancer Detection

There are various non-invasive imaging systems which have been employed to investigate unique features of cancerous areas in skin but the majority of them are limited by unexpected drawbacks [Mogensen and Jemec, 2007]. Dermatoscopy, a non-invasive tool for skin cancer diagnosis, has been widely applied in clinical examinations for over 20 years [Argenziano and Soyer, 2001]. It also has the largest clinical impact in terms of the new skin cancer diagnostic methods. A diagnosis process using the dermatoscopy includes visualization of the cellular structure of skin and is based on several criteria to make final decisions in pathology [Argenziano and Soyer, 2001]. As a consequence, this method is not only dependent on pathologists' experiences but also very time-consuming. Ultrasound skin imaging systems enable the production of biological images of skin based on measuring the reflection of oscillating-sound waves from the examined surface [Vogt et al., 2003]. However, various chemical contents and subtypes of the skin tumour are not distinguishable by ultrasonography, thereby limiting the applicability of this technique [Fornage et al., 1993]. Magnetic resonance imaging (MRI) needs whole body scanning to produce skin images and it is considered more suitable for prognosis than diagnosis [Williams et al., 2001]. Near-infrared (NIR) imaging includes two main techniques: optical coherence tomography (OCT) and reflectance confocal microscopy (RCM). Being capable of constructing high-resolution images, OCT allows visualization of micromorphological structure of skin with an axial resolution of about

15 μm [Fujimoto et al., 2000; Gambichler et al., 2007]. RCM can provide cellular details of skin with a higher resolution of 1–2 μm [Rajadhyaksha et al., 1995]. However, OCT cannot provide details about the morphology of single cells while RCM is restricted owing to low penetration of up to 300 μm into the skin [Rigel et al., 2010]. Fluorescence spectroscopy using autofluorescence or endogenous and exogenous fluorescent techniques is another skin cancer detection tool [Borisova et al., 2014]. Although the most important advantage of this diagnostic tool is its ability to discriminate between malignant and nonmalignant lesions, distinguishing *in vivo* between nonmelanoma and other benign or malignant skin tissues remains fairly limited. The disadvantages of the aforementioned imaging techniques leave the field open to research on new imaging technologies. Terahertz imaging, which has a number of advantages over the existing methods, has been displaying potential for application to skin cancer [Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004]. Its performance in skin cancer detection is thoroughly reviewed in the following section.

2.5.1.3 THz Imaging for Skin Cancer

THz imaging has long been used to study a number of applications for human skin. Cole et al. [2001] conducted the first *in vivo* measurements of the skin arm and hand, and consequently found that THz imaging in reflection geometry could detect changes in the hydration level of stratum corneum. Their further investigation of the reflections showed the possibility of determining the thickness of the stratum corneum layer through analysing the waveforms of the THz reflections. This finding could be beneficial to cosmetic applications of THz imaging.

Woodward et al. [2002] investigated the reflected THz signals off scar/inflammation and BCC tissues *in vitro* and *in vivo* together with applying time-domain techniques such as minimum peak and time-post-pulse (TPP) to imaging these samples. Accordingly, scar tissue, which is known for its decreasing water content as compared to its surrounding

normal tissue, creates negative changes in contrast to the normal and inflammation regions in false-colour THz images of the skin samples. On the other hand, positive changes in the contrast between BCC and its adjacent normal tissue were observed. This is possible due to a rise in water content as carcinogenesis leads to an increase in tumour cells and their size and composition. More insights into the significant differences in the time domain pulses between normal and cancerous skin were provided through frequency-domain analysis techniques [Woodward et al., 2003]. This study highlighted the complexity of obtaining the spectral information of skin tissue in the reflection mode.

Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone [2004] employed a portable terahertz pulsed imaging (TPI) system to image 18 BCC *ex vivo* and 5 *in vivo* samples with the aim of examining the capability of this imaging system to delineate tumour margins in MMS. The THz waveforms shown in Fig. 2.4 are among those used to generate THz images of the tissue samples. Particularly, maximum amplitudes or normalised amplitudes at the optical delay t giving the maximum difference in amplitude between waveforms of normal and cancerous tissues were employed to explore either the surface or depth information of the samples (e.g. Fig.2.5 from [Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004]). Accordingly, the contrast shown in the THz images was considered to be sufficient to map tumour margins to a certain extent, thus supporting the potential of the THz imaging in the tumour margin detection.

Another recent study by Joseph et al. [2014] proposed a combination of polarised terahertz and optical imaging to delineate nonmelanoma skin cancer. Their results indicate that cross-polarised terahertz imaging can provide a correct detection of the cancer area due to the lower reflectivity of tumourous area with respect to normal ones, thus improving the resolution of the optical imaging. As a result, this approach was suggested to be capable of microscopically accessing tissue morphology as well as accurately delineating tumour margins.

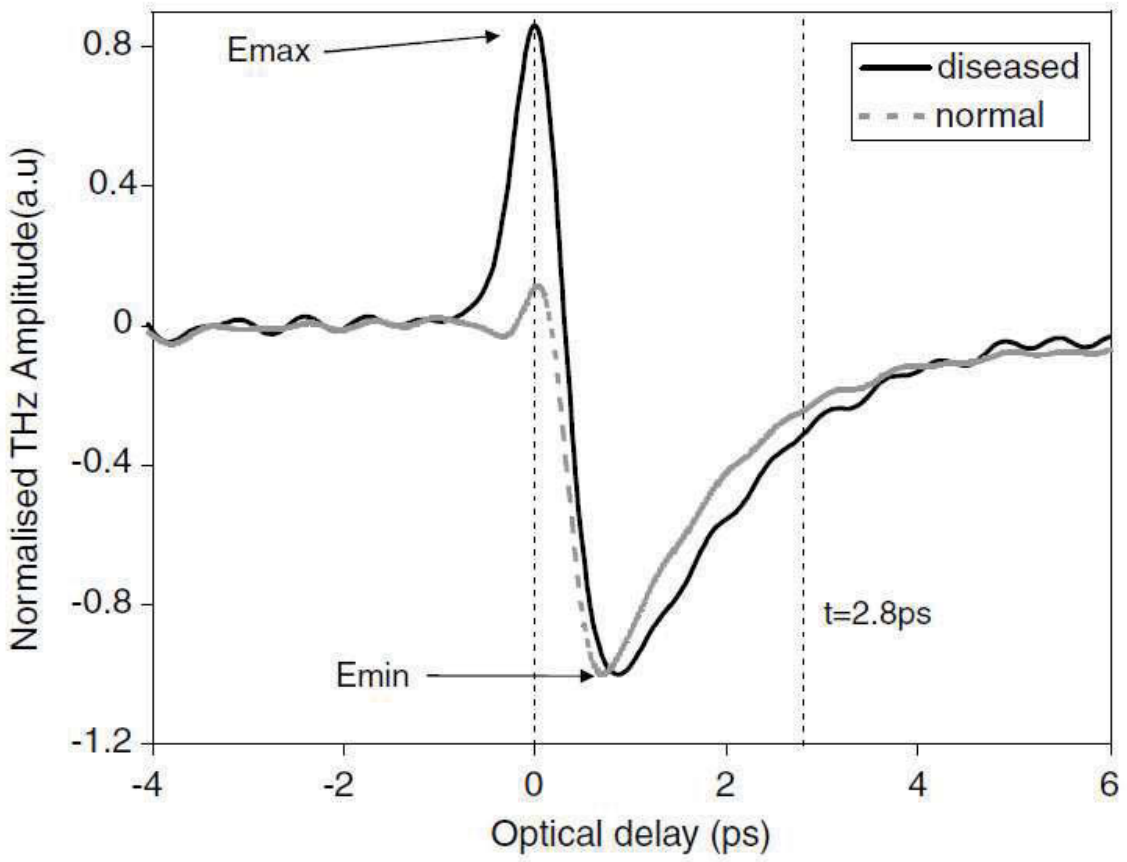

Figure 2.4: Averaged THz pulses reflected off a tissue sample [Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004].

A number of other studies looked at the spectroscopic sides of the THz images, such as the optical/dielectric properties, to understand the imaging contrast. Pickwell, Cole, Fitzgerald, Wallace and Pepper [2004] developed a finite-difference-time-domain (FDTD) model to simulate the interaction between THz radiation and human skin and, hence, found the correlation between the skin hydration changes and the refractive index of skin. This model was successfully applied to *ex vivo* data of skin cancer [Pickwell et al., 2005]. This study also indicates that the refractive index and absorption coefficient of BCC are higher than those of the neighbouring healthy tissues. It was followed by a conventional work of Wallace et al. [2006] who concentrated on the statistical analysis of the skin cancer optical properties. The authors found significant

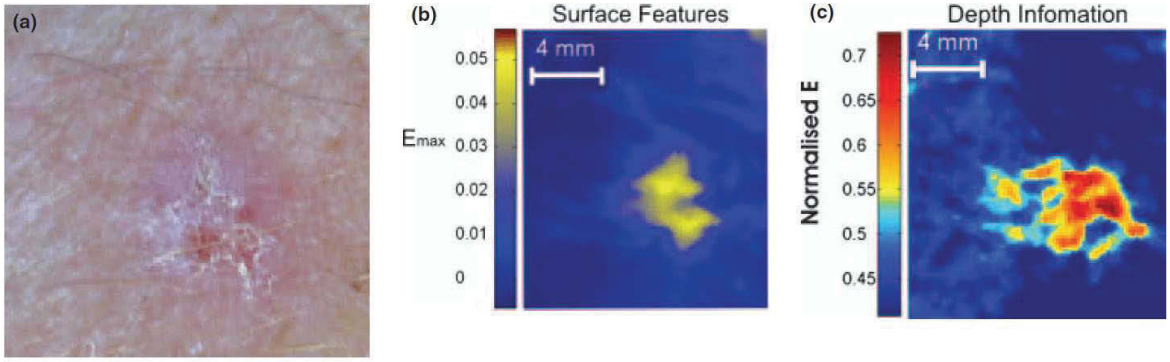

Figure 2.5: (a) A clinical photo of an invasive nodular BCC; (b) The corresponding THz image with the maximum amplitude E_{max} for showing surface features; (c) The corresponding THz images with a normalised amplitude at $t = 2.8\ ps$ for mapping the tumour at depth [Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004].

differences between BCC and normal samples in the refractive index and the absorption coefficient over the range 0.25–0.90 THz and 0.2–2.0 THz respectively. In fact, the differences in the fundamental properties partially elaborate the contrast features between healthy and cancerous regions in THz images of BCC [Woodward et al., 2003; Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004]. Furthermore, THz complex permittivity, which can be directly calculated from the complex refractive index, has been recently found to be capable of distinguishing dysplastic nevi, an early stage of melanoma, from non-dysplastic skin nevi [Zaytsev et al., 2015]. All the achievements have pointed to the potential of THz imaging to aid skin cancer delineation in surgical treatments. However, the aforementioned studies have only considered normal tissue and its BCC form. Therefore, we have yet to know whether THz imaging can distinguish between skin cancer types or between malignant and benign manifestations. Further investigation of skin disorders might serve to find an answer to these questions.

2.5.2 Breast Cancer Detection

2.5.2.1 Overview of Breast Cancer

Excluding nonmelanoma skin cancer, current statistics shows that breast cancer is the most commonly diagnosed cancer among females [American Cancer Society, 2011]. It is also one of the most fatal forms of cancer worldwide [Jemal et al., 2008]. According to the statistics, there were an estimated 1.4 million new cases diagnosed in 2008, which results in approximately 458,400 deaths from breast cancer. In general, the breast cancer incidence rates have been increasing since the late 1990s. The mortality rate of breast cancer has only stabilised or decreased in some parts of the world, e.g. North American and some European countries.

Early detection and advanced treatments have been attributing to mortality reduction in the West while the westernised lifestyle and inadequate effective programs of screening breast cancer have been leading to an increase in the mortality rates in many Asian countries [Canfell et al., 2008; Coughlin and Ekwueme, 2009]. Numerous studies have shown that early detection increases the patients' chance of survival due to more treatment options and their higher effectiveness [American Cancer Society, 2011]. Available methods for breast cancer early detection are based on mammography and clinical breast examination [Boyd et al., 2007; Anderson et al., 2008]. Mammography can detect the breast cancer at its early stage when treatment may be more effective but it may be prohibitively expensive for low- and middle-income countries. Thus, clinical breast examination to detect early signs and symptoms of the cancer is recommended for developing countries [Anderson et al., 2008]. There are a variety of treatment methods including lumpectomy, mastectomy, radiation therapy, hormone therapy, chemotherapy, or targeted biologic therapy [American Cancer Society, 2011].

2.5.2.2 Existing Techniques of Breast Cancer Detection

Breast-conserving surgery (BCS) including lumpectomy and irradiation therapy is the standard treatments for clinical stages I and II breast cancer, which is undergone by approximately 70% of diagnosed cases [Pleijhuis et al., 2009; Fitzgerald et al., 2012]. A number of studies have shown difference in survival rates between BCS and traditional mastectomy [Blichert-Toft et al., 2008]. From the psychological perspective, BCS based on surgical removal of breast tumour with clear margins is preferable with better cosmetic results and lower risk of wound infection than the mastectomy with full removal of the breast. However, BCS has an important disadvantage associated with the risk of local cancer recurrence (LR), thus requiring additional surgery [Kreike et al., 2008]. This problem depends on the completeness level of removing the breast tumours. In fact, in 20–40% of patients exposed to BCS, positive margins with tumour cells at around the cut edge of the surgical specimen were detected after resecting the primary tumour according to previous studies [Pleijhuis et al., 2009; Miller et al., 2004; Jacobs, 2008]. In addition to the positive margins, close margins defined by the presence of cancer cells within 2 mm from the cut edge need to be taken into account for possible re-excision due to the increasing risk of LR and the tumorous cells reaching the cut edge [Kunos et al., 2006]. For instance, Zavagno et al. [2008] recently conducted a trial in which 1520 patients undergoing BCS had their surgical margins evaluated for re-excision. Close and positive margins found in 431 cases corresponded to LR rates of 51.8% and 34.1% respectively. Unsuccessful removal of all cancerous cells with sufficiently large margins leads to an increased risk of local recurrence and thus, additional surgeries are required to completely remove the remaining tumours. These surgical operations may also cause unexpected cosmetic results, higher risk of wound infection, lower chance of survival, and increased cost of the treatment program [Fitzgerald et al., 2012]. Therefore, accurate localization of breast tumours plays a crucial role in obtaining tumour-free margins with the least number of secondary surgeries and minimal removal of healthy tissue. It also optimises the cosmetic side of the surgery.

A variety of techniques have been clinically applied to localise the tumour both preoperatively and intraoperatively. Although current preoperative mammography using radiographic X-ray is considered to be adequate for detecting breast cancer, it has low specificity and does not provide functional and quantitative information of breast tissue [Pleijhuis et al., 2009; Ikeda et al., 2003]. Combining ultrasound with the mammography also underestimates the tumour size and developing pattern of diffuse and multinodular tumours [Faverly et al., 2001]. Magnetic resonance imaging (MRI) has become the most preferable clinical imaging modality for preoperative tumour localization and assessment due to its high sensitivity and ability to deal with the shortcomings of mammography [Pleijhuis et al., 2009]. However, this technique is limited by real-time capability of margin assessment as well as not improving examination of margin status after resection [Pengel et al., 2009]. Wire-guided localization (WGL) using a wire guided by MRI, X-ray, or ultrasound was a clinical standard technique for intraoperatively localizing the tumour after resection but this has recently been found to cause unacceptable rates of positive margins [Pleijhuis et al., 2009; Kelly and Winslow, 1996]. Although intraoperative ultrasound-guided resection in which ultrasound is used to assess the tumour location and margins before and during BCS can lead to improvements in identifying margin status as compared to WGL, its detection of nonpalpable tumours and ductal carcinoma in situ is very limited [Smith et al., 2000; Klimberg, 2003]. Another technique using ultrasound is the cryoprobe-assisted localization and this has been found to be comparable to WGL with some improvements in reducing the amount of excised healthy tissue as well as excision time and complication [Tafra et al., 2006]. Radiographic X-ray mammography is also used for intraoperative assessment but this has low resolution and sensitivity [Huynh et al., 1998]. Frozen section analysis (FSA) and intraoperative touch preparation cytology (ITPC) are approaches which employ microscopic examination to assess tumor margins. FSA offers a safe and inexpensive procedure together with minimizing the re-excision rate but it is not suitable for small tumours, very time-consuming, and dependent on pathologists' experiences [Pleijhuis et al., 2009; Riedl et al., 2009]. Despite the fact that ITPC is an efficient technique in

terms of time, cost, accuracy, and tumour size as well as reportedly achieving significantly lower LR rates than the above methods, it is not commonly applied due to not providing any information in terms of close margins [Pleijhuis et al., 2009; Klimberg et al., 1998].

The aforementioned clinical techniques generally have a high possibility of leaving positive margins after resection and consequently lead to an increased risk in terms of LR and cosmetic results, thus it is essential to seek innovative approaches in relation to improving the intraoperative margin assessment. For instance, ^{18}F-fluoro-2-deoxy-D-glucose (^{18}F-FDG) positron emission tomography imaging (PET), a powerful tool for malignant diagnosis, has been investigated for breast cancer due to potentially increasing the ^{18}F-FDG in breast cancerous cells [Brown et al., 1996; Hall et al., 2007]. However, this imaging modality possesses several critical disadvantages such as its limitation in terms of partial solution (< 1 cm), low specificity, high cost, and resulting high radiation exposure to patients during the intraoperative assessment [Hall et al., 2007; Wahl, 2001; Strong et al., 2008]. Another technique under research is the radioguided occult lesion localization (ROLL) that employs a hand-held gamma probe together with a radio tracer (i.e. radioisotope or radioactive iodine) injected into the examined tumour [Luini et al., 1999; Gray et al., 2001]. ROLL proves to be more advanced in terms of simplicity, accuracy, and safety than the conventional WGL but still requires further research to be clinically applied [Lovrics et al., 2011; van Esser et al., 2008]. Optical imaging modality such as near-infrared fluorescence (NIRF) or fluorescence reflectance imaging (FRI) have been recently investigated as well. They provide some biomarkers of breast cancer as well as having a number of technical advantages such as ease of use, high lateral resolution, cost efficiency, nonionization, noninvasive and safety [Ke et al., 2003; Luker and Luker, 2008; Tagaya et al., 2008; Ntziachristos, 2006]. However, the intrinsic optical properties of near-infrared light results in limitations of the imaging systems including reduced sensitivity and low depth resolution due to light absorption and scattering [Tagaya et al., 2008; Ntziachristos, 2006].

2.5.2.3 THz Imaging for Breast Cancer

Breast tissue and its cancerous forms have also shown differences in terahertz properties. [Fitzgerald et al., 2006] conducted a pilot study on imaging *ex vivo* breast samples, including invasive and *in situ* carcinomas excised from 22 female patients. Their generated THz images were generally comparable with the corresponding histopathological photomicrographs with sufficiently high correlation. For instance, Fig. 2.6 compares

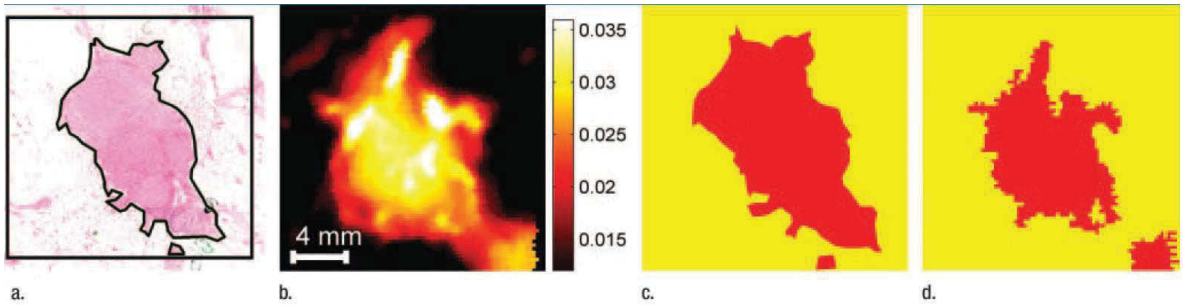

Figure 2.6: (a) Photomicrograph of an invasive ductal carcinoma sample; (b) the corresponding THz image using the minimum amplitude of the reflection impulses; The cancerous regions delineated from the photomicrograph (c) and the THz image (d) [Fitzgerald et al., 2006].

the microphotograph and THz image of one sample containing invasive ductal carcinoma. Additionally, the THz contrast created by the presence of the *in situ* form of breast tumours is crucial as non-calcified tumours are impalpable and hence are often missed during BCS. However, the outcomes of this study are limited by several factors such as tissue deformation due to histopathological preparation, alignment of the image window, or the small number of specimens which do not include benign tissues. In addition to these limitations, the origin of the imaging contrast is not well understood.

The positive results of Fitzgerald et al. [2006] have led to the development of a intro-operative THz imaging system using a portable scanning probe for medical applications by Teraview Ltd. (Cambridge, UK) [Ashworth et al., 2008]. Then, Ashworth et al. [2009] executed a thorough investigation of human breast cancer through analysing the terahertz properties extracted from reflections. 74 samples excised from 20 consecutive

female patients were used for this study as well as classified into three groups based on their dominant composition, namely healthy fat, fibre, and tumour. As can be seen in

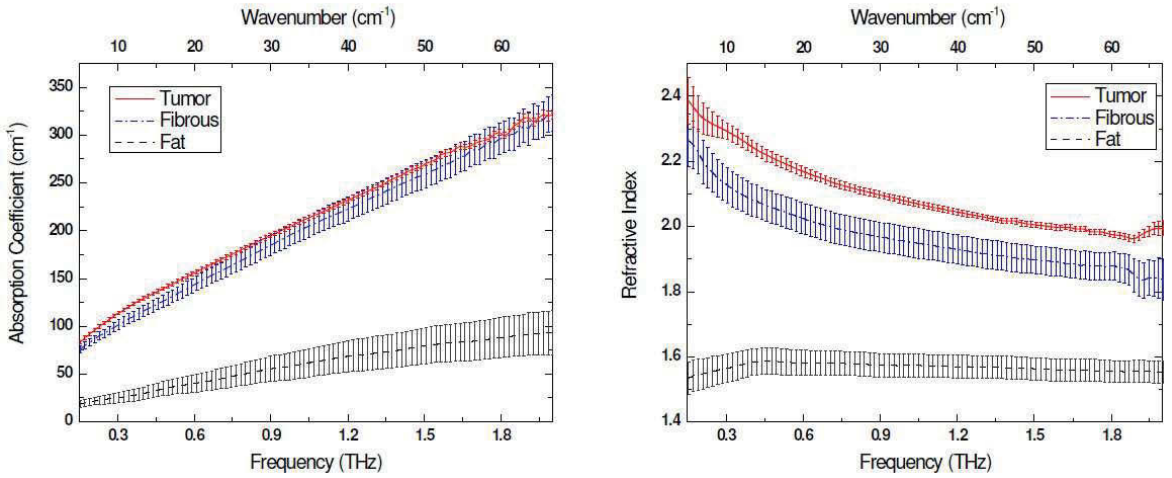

Figure 2.7: Averaged refractive indices and absorption coefficients with the corresponding standard errors of three breast tissue types: fat, fibre, and tumour [Ashworth et al., 2009].

Fig. 2.7, the breast tumour possesses a higher refractive index than both fat and fibrous tissue over the frequency range 0.2–2.0 THz while its absorption is found to slightly increase at around 0.32 THz. Furthermore, simulated impulse functions from these optical properties reveal a difference in peak of approximately 60% between healthy and tumorous breast samples, thus partly explaining the contrast sources in the THz images.

Imaging techniques used for breast tumour in the aforementioned studies are indeed not only limited by measurement factors but also the selection of imaging features to generate the THz images. To be more specific, it is not yet known which features offer the best contrast or an official procedure for the feature selection. A recent large-scale study of employing THz imaging for breast cancer detection was carried out by Fitzgerald et al. [2012] in order to seek the best classification features for imaging breast cancer. The THz database included reflected THz pulses measured from 20×20 mm samples freshly excised at suspected regions from 51 female patients. Ten heuristic parameters

and their combinations were investigated through principle component analysis and support vector machine classifiers. By using all ten of these parameters, the authors could classify the breast cancer samples with an accuracy of 92%. However, in spite of the encouraging results, the applicability of the THz probe during surgery has been challenged by a number of factors such as the presence of blood and other fluids in the imaged regions, or the variation of tissue hydration and temperature *in vivo*. Consequently, it still requires further systematic studies of the contrast mechanism and potential sources of difference between various types of breast tissue, both healthy and diseased, to make THz imaging an established medical tool.

2.5.3 Dielectric Modelling-Based Approach

Various features extracted from terahertz signals in both the time and frequency domains have been used to form terahertz images of human skin and breast tissue [Woodward et al., 2003; Fitzgerald et al., 2002, 2012; Joseph et al., 2009]. However, since the construction of images is based on arbitrary choices of the imaging features, it potentially involves signal noise and destructive information that can negatively affect the ability of the images to discriminate the disease [Brun et al., 2010; Chan et al., 2007]. Furthermore, it has been reported that changes in the absorption coefficient and refractive index stimulate the variation of these features, thus producing contrast in the THz imaging. The optical properties underline the physical mechanism that describes the interaction between THz radiation and human tissue. However, the root of the increase in the absorption and refractive index between tumours and healthy tissue still requires more thorough studies to be conducted [Ashworth et al., 2009; Parrott et al., 2011]. Apart from this, a solid understanding of the contrast mechanism also has a positive impact on determining informative features for cancer classification. As optical properties are directly related to dielectric permittivities, a concrete understanding of dielectric properties of human tissue provides a deeper insight into the contrast mechanism. In fact, a dielectric model can offer a theoretical framework characterizing the

complex permittivity of tissue, which reflects interactions between molecules and the THz radiation [Smye et al., 2001]. Therefore, modelling dielectric properties plays a vital role in not only explaining the contrast between healthy tissue and tumours in THz images but also potentially introducing more classifiers for cancer detection.

Considering the dielectric properties of human skin, Pickwell, Cole, Fitzgerald, Wallace and Pepper [2004], for the first time, introduced application of the double Debye model to describe the complex permittivities of skin tissue. This introduction was based on the high radiation absorption of water in the terahertz frequency as well as increased water content of skin tumour as compared to normal skin. The authors used the non-linear least square method to extract the five DD parameters which were subsequently incorporated into their developed FDTD program in order to simulate impulse functions of samples. This approach was tested with the *in vivo* measured impulse function of the skin on the volar forearm in the frequency range 0.2–2 THz. The simulation of that study shows a good match between the measurement and simulated value with a correlation of 0.97. Further investigation of the double Debye model into the skin on different sites such as dorsal forearm and palm was successfully conducted by Pickwell et al. [2005] with the similar agreements, hence confirming the applicability of this dielectric model. As a result, Pickwell et al. [2005] with the same approach carried out the study on BCC, both *ex vivo* and *in vivo*, and extracted the DD parameters from normal skin tissues and the BCCs. Significant differences in these parameters between normal and BCC cases were also found, hence suggesting their potential for the cancer discrimination. Their simulation results were very encouraging but still showed some discrepancies between the measured data and simulated values. Therefore, the study in this thesis analysed this fitting problem and realised that the fitting algorithm used to extract the DD parameters could be the source of the errors. Particularly, it is found that the fitting quality with the NLS method is not consistent, especially very poor at frequencies below 0.8 THz of the examined range 0.2-2 THz, despite that this extraction algorithm proves to be sufficient for the range only up to 1 THz [Liebe et al., 1991]. This

unexpected performance should be not a surprise because the NLS algorithm aims at local solutions that are not optimal for the fitting problem of the double Debye model. Therefore, there is an apparent need for an extraction algorithm capable of fishing the optimal solutions.

Fitzgerald et al. [2014] recently employed the double Debye model to simulate the impulse function of breast tissue in the terahertz region. Some correlation of the simulated data with its measurements was found for both healthy breast tissue and tumour. However, the authors pointed out the challenge of applying the double Debye model to breast tissue based on the limited fitting quality. Admittedly, breast tissue has more heterogeneous structure than skin tissue, which especially contains low water content and a considerable proportion of adipose tissue (fat or lipid). The low hydration indeed disproves the main motive for the use of the double Debye model. Furthermore, measured dielectric permittivities of various *ex vivo* breast samples by Ashworth et al. [2009] demonstrate an increase of the real spectra in the vicinity of 0.32 THz. This spectral response evidently does not obey the Debye relaxation theory. To sum up, other dielectric models should be developed to address the modelling challenges for breast tissue.

Chapter 3

PARAMETER EXTRACTION FOR THE DOUBLE DEBYE MODEL OF HUMAN SKIN TISSUE

3.1 Introduction

This chapter is concerned with parameter extraction for the double Debye model, which is used for determining the frequency-dependent complex permittivities of human skin. The model parameters are considered as the origin of contrast in THz images of skin cancer. However, the existing extraction methods by nonlinear least-squares (NLS) algorithms, such as Levenberg-Marquardt algorithm, could generate Debye models capable of tracking their measurements accurately at frequencies higher than one terahertz but poorly at lower frequencies where the majority of contrast between healthy and diseased skin tissues is actually observed. Adding the Lorentzian resonant process terms or more Debye relaxation processes as suggested by Liebe et al. [1991] might improve the fitting. However, these extras complicate the applicable model unnecessarily whereas the model complexity could be a potential source of over-fitting error.

To deal with the aforementioned problem, the study in this thesis targets at improving the fitting algorithm as well as optimizing the extraction of the double Debye parameters. As the conventional sum of square error function, on which the extraction is based, is a complex nonlinear and nonconvex function of the DD parameters, the NLS-based approach could locate a solution to satisfy only some necessary optimality conditions. In view of this suboptimal issue, our approach aims to locate the globally optimal solution of the problem, which leads to an optimal double Debye model for a very accurate description of human skin permittivity in the entire THz frequencies. As a result, we consecutively introduce two global optimization-based algorithms to not only optimize the fitting problem but also ensure the reliability and computational tractability of the optimized double Debye model for modelling dielectric permittivity of human skin. The first algorithm is based on grinding technique that continuously solves the optimisation problem at each point of a finite-dimensional space defined by the constraints of the problem variables and chooses the optimal solution accordingly.

Despite that this initial approach is effective in terms of optimising the fitting, it is computational and time-consuming. As a result, the second optimisation algorithm based on a branch-and-bound method was introduced in order to improve the efficiency of the fitting procedure. Specific details of the two optimisation algorithms are presented in the following sections.

This chapter is structured as follows. Section 3.2 is devoted to the problem formulation whereas the next section 3.3 introduces the global optimization algorithms to solve the proposed problem. Numerical results together with discussion on them are given in section 3.4. Finally, section 3.5 summarises the contributions in this chapter.

3.2 Problem Formulation

3.2.1 The Double Debye Model

The double Debye model has been shown by previous studies to be capable of describing the dielectric response of human skin at the THz frequencies [Pickwell, Cole, Fitzgerald, Wallace and Pepper, 2004; Wallace et al., 2006; Truong et al., 2013]. Since this model targets the mechanism of interaction between THz radiation and water molecules, human skin which always contains a large proportion of water can share this approach. Particularly, the well-known simple Debye equation is generally believed to be adequate for simulating the permittivity of pure water at frequencies up to 100 GHz [Alison and Sheppard, 1990]. A second Debye relaxation process added to the single Debye equation is proposed for fitting the permittivity of water in the transition range 0.1–1 THz [Barthel and Buchner, 1991; Kindt and Schmuttenmaer, 1996]. This incorporation reflects the impact of an external electric field on water molecules [Ronne and Keiding, 2002; Wallace et al., 2006]. The double Debye model is characterised by the frequency-dependent dielectric function

$$\epsilon_{\mathrm{DD}}(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_2}{1 + \jmath\omega\tau_1} + \frac{\varepsilon_2 - \varepsilon_\infty}{1 + \jmath\omega\tau_2}. \tag{3.1}$$

Here ε_s is the static permittivity at low frequency, ε_∞ is the high-frequency limit of permittivity, and ε_2 is the intermediate dielectric constant for the description of the transitional state between the two relaxation processes. Therefore, $\varepsilon_s - \varepsilon_2$ and $\varepsilon_2 - \varepsilon_\infty$ represent the dispersion in amplitude of the slow and fast relaxation processes respectively. It is suggested that in liquid water the molecules are in the form of tetrahedral structure; when water is excited by incident THz radiation, this structure is perturbed and reorients, for this to happen the structure must break. Thus, four hydrogen bonds need to be broken which is a slow process (τ_1). Subsequently, after approximately τ_1, the single water molecule will reorient and move (a fast process), in τ_2, to a new tetrahedral site [Rønne et al., 1997].

3.2.2 Optimization Problem

For mathematical simplification, the common form of the double Debye model in (3.1) can be rewritten without losing its generality as follows

$$\epsilon_{\mathrm{DD}}(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon_1}{1 + \jmath\omega\tau_1} + \frac{\Delta\varepsilon_2}{1 + \jmath\omega\tau_2}, \tag{3.2}$$

where $\Delta\varepsilon_1 = \varepsilon_s - \varepsilon_2$ and $\Delta\varepsilon_2 = \varepsilon_2 - \varepsilon_\infty$. This form of the double Debye model is used throughout the following sections.

The complex optical properties $\tilde{n}_m$ of human skin, both normal and cancerous, were measured by a number of studies through terahertz spectroscopy systems [Pickwell et al., 2005; Wallace et al., 2006]. Thus, the measured complex permittivity $\tilde{\varepsilon}_m$ at frequency ω can be calculated using the following relationship,

$$\epsilon_m(\omega) = (n_m)^2 = \big[n(\omega) - \jmath\kappa(\omega)\big]^2. \tag{3.3}$$

The square error function between the relative complex permittivity $\tilde{\epsilon}_{\mathrm{DD}}(\omega)$ in (3.2) and its measured value $\tilde{\epsilon}_m(\omega)$ in (3.3) is determined by

$$E(\omega_i) \;=\; |\epsilon_{\mathrm{DD}}(\omega_i) - \epsilon_m(\omega_i)|^2 = \left|\varepsilon_\infty + \frac{\Delta\varepsilon_1}{1 + \jmath\omega_i\tau_1} + \frac{\Delta\varepsilon_2}{1 + \jmath\omega_i\tau_2} - \tilde{\epsilon}_m(\omega_i)\right|^2. \tag{3.4}$$

Accordingly, the five double Debye parameters $(\varepsilon_\infty, \Delta\varepsilon_1, \Delta\varepsilon_2, \tau_1, \tau_2)$ in (3.2) can be extracted from the following sum-of-square-error minimisation

$$\min_{(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty, \tau_1, \tau_2)} \sum_{i=1}^{N} E(\omega_i) \;\; : \;\; \varepsilon_\infty \geq 1, \Delta\varepsilon_1 \geq 0, \;\; \Delta\varepsilon_2 \geq 0, \tau_1 > 0, \tau_2 > 0, \tag{3.5}$$

where N is the number of discretely-sampled frequencies $\omega_i = 2\pi f_i$ with $f_i \in [0.2, 2]$ THz. The constraints in (3.5) cover most practical cases of dispersive materials and can be explained from the physical viewpoint [Barthel and Buchner, 1991]. However, according to previous extraction results of Barthel and Buchner [1991]; Kindt and Schmuttenmaer [1996]; Pickwell et al. [2005] for human skin, they can be reasonably narrowed down to

$$1 \leq \tau_1 \leq 20, \;\; 0.01 \leq \tau_2 \leq 0.5 \tag{3.6}$$

The next two sections present the two proposed optimization algorithms to solve (3.5) with respect to (3.6).

3.3 Global optimization-based Algorithms

3.3.1 Grinding-based optimization

It is seen from (3.4) that each $E(\omega_i)$ is a highly nonlinear and nonconvex function in $(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty, \tau_1, \tau_2)$. Consequently, (3.5) is minimisation of sum of N nonlinear and nonconvex functions subject to convex constraints, which is a hard optimization problem. However, $E(\omega_i)$ is obviously convex in $(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)$ with τ_1 and τ_2 held fixed. This means that only two variables τ_1 and τ_2 are complicating variables which make (3.5) non-convex [Tuy, 1998]. Thus, we can rewrite (3.5) into the following parametric optimization,

$$\min_{(\tau_1, \tau_2)} \;\; F(\tau_1, \tau_2) \quad \text{subject to} \quad (3.6) \tag{3.7}$$

with

$$F(\tau_1, \tau_2) := \min_{(\varepsilon_s, \varepsilon_2, \varepsilon_\infty)} \;\; \sum_{i=1}^{N} E(\omega_i) : \;\; \varepsilon_\infty \geq 1, \;\; \Delta\varepsilon_1 \geq 0, \;\; \Delta\varepsilon_2 \geq 0, . \tag{3.8}$$

Firstly, our focus is to solve (3.8) as it is a convex problem. Let $x = (\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)^{\mathrm{T}} = (x_1, x_2, x_3)^{\mathrm{T}}$ and

$$a_i = \begin{bmatrix} \frac{1}{1+\jmath\omega_i\tau_1(i)} \\ \frac{1}{1+\jmath\omega_i\tau_2(k)} \\ 1 \end{bmatrix},$$

$$A_i = a_i^* a_i^{\mathrm{T}},\ b_i^{\mathrm{T}} = -2\mathrm{Re}(c_i a_i^{\mathrm{H}}), d_i = |c_i|^2,$$

$$A = \sum_{i=1}^{N} A_i, b^{\mathrm{T}} = \sum_{i=1}^{N} b_i^{\mathrm{T}}, d = \sum_{i=1}^{N} d_i,$$

Consequently, the objective function of (3.8) is the following convex quadratic function of x,

$$\begin{aligned}
\sum_{i=1}^{N} E(\omega_i) &= \sum_{i=1}^{N} \left|a_i^{\mathrm{T}} x - c_i\right|^2 \\
&= \sum_{i=1}^{N} (a_i^{\mathrm{T}} x - c_i)^* (a_i^{\mathrm{T}} x - c_i) \\
&= \sum_{i=1}^{N} x^{\mathrm{H}} a_i^* a_i^{\mathrm{T}} x + |c_i|^2 - c_i x^{\mathrm{H}} a_i^* - c_i^* a_i^{\mathrm{T}} x \\
&= \sum_{i=1}^{N} x^{\mathrm{H}} a_i^* a_i^{\mathrm{T}} x + [-2Re(c_i a_i^{\mathrm{H}})] x + |c_i|^2 \\
&= \sum_{i=1}^{N} (x^{\mathrm{T}} A_i x + b_i^{\mathrm{T}} x + d_i) \\
&= x^{\mathrm{T}} (\sum_{i=1}^{N} A_i) x + (\sum_{i=1}^{N} b_i^{\mathrm{T}}) x + \sum_{i=1}^{N} d_i \\
&= x^{\mathrm{T}} A x + b^{\mathrm{T}} x + d.
\end{aligned}$$

Thus, (3.8) is the following convex quadratic program in three scalar variables only

$$\begin{array}{ll} \min\limits_{x \in \boldsymbol{R^3}} & x^{\mathrm{T}} A x + b^{\mathrm{T}} x + d \\ \text{subject to} & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 1 \end{array}. \tag{3.9}$$

This optimization problem can be solved easily by any existing quadratic programming solver such as SeduMi [Sturm, 1999].

Since the objective function $F(\tau_1, \tau_2)$ consists of two variables only, the optimal solution of (3.7) can be easily located through repeatedly examining the value of F at each point $(\tau_1(k), \tau_2(k))$ identified by

$$\tau_1(k) = l_1 + \frac{u_1 - l_1}{M}k,\ \tau_2(k) = l_2 + \frac{u_2 - l_2}{M}k, \tag{3.10}$$

where $k = 0, 1, 2...M$, $l_1 = 1$,$u_1 = 20$, $l_2 = 0.01$, $u_2 = 0.5$ according to (3.6). M can be reasonably chosen from at least 80 to sufficiently avoid local optimums. Increasing values of M often lead to unnecessary trade-offs between optimal results and computational cost.

3.3.2 BB-Based optimization

Since the value $F(\tau_1, \tau_2)$ in (3.8) is very sensitive to (τ_1, τ_2), the set of the grinding points must be very dense for good performance. Thus, the grinding-based method is computationally demanding. Instead, a branching and bounding (BB) method of global optimization (BB-GO) is proposed to improve the computational efficiency.

The explicit form of the objective function in (3.5) is derived in a different way as follows. Set

$$0 < a_i := \Re\big(\epsilon_m(\omega_i)\big),\ 0 < b_i := -\Im\big(\epsilon_m(\omega_i)\big).$$

Thus

$$E(\omega_i) \ = \ (\varepsilon_\infty + \frac{\Delta\varepsilon_1}{1+\omega_i^2\tau_1^2} + \frac{\Delta\varepsilon_2}{1+\omega_i^2\tau_2^2} - a_i)^2 + (\frac{\omega_i\Delta\varepsilon_1\tau_1}{1+\omega_i^2\tau_1^2} + \frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2} - b_i)^2$$

Define

$$\begin{aligned} E_1(\omega_i) \ &:= \ (\varepsilon_\infty + \frac{\Delta\varepsilon_1}{1+\omega_i^2\tau_1^2} + \frac{\Delta\varepsilon_2}{1+\omega_i^2\tau_2^2} - a_i)^2 \\ &= \ (\varepsilon_\infty + \frac{\Delta\varepsilon_1}{1+\omega_i^2\tau_1^2} + \frac{\Delta\varepsilon_2}{1+\omega_i^2\tau_2^2})^2 + a_i^2 - 2a_i(\varepsilon_\infty + \frac{\Delta\varepsilon_1}{1+\omega_i^2\tau_1^2} + \frac{\Delta\varepsilon_2}{1+\omega_i^2\tau_2^2}) \\ &= \ \boldsymbol{x}^{\mathrm{T}}\Gamma_i(\tau_1, \tau_2)\boldsymbol{x} + \beta_i^{\mathrm{T}}(\tau_1, \tau_2)\boldsymbol{x} + a_i^2 \end{aligned} \tag{3.11}$$

for

$$\begin{aligned}\Gamma_i(\tau_1,\tau_2) &= \begin{pmatrix} \frac{1}{1+\omega_i^2\tau_1^2} \\ \frac{1}{1+\omega_i^2\tau_2^2} \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{1+\omega_i^2\tau_1^2} \\ \frac{1}{1+\omega_i^2\tau_2^2} \\ 1 \end{pmatrix}^{\mathrm{T}}, \\ \beta_i(\tau_1,\tau_2) &= -2a_i \begin{pmatrix} \frac{1}{1+\omega_i^2\tau_1^2} \\ \frac{1}{1+\omega_i^2\tau_2^2} \\ 1 \end{pmatrix}, \ \boldsymbol{x} = (\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)^{\mathrm{T}}. \end{aligned} \tag{3.12}$$

Analogously,

$$\begin{aligned} E_2(\omega_i) &:= (\frac{\omega_i\Delta\varepsilon_1\tau_1}{1+\omega_i^2\tau_1^2} + \frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2} - b_i)^2 \\ &= (\frac{\omega_i\Delta\varepsilon_1\tau_1}{1+\omega_i^2\tau_1^2} + \frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2})^2 + b_i^2 - 2b_i(\frac{\omega_i\Delta\varepsilon_1\tau_1}{1+\omega_i^2\tau_1^2} + \frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2}) \\ &= \boldsymbol{x}^{\mathrm{T}}\Lambda_i(\tau_1,\tau_2)\boldsymbol{x} + \chi_i^{\mathrm{T}}(\tau_1,\tau_2)\boldsymbol{x} + b_i^2 \end{aligned} \tag{3.13}$$

for

$$\Lambda_i(\tau_1,\tau_2) = \begin{pmatrix} \frac{\omega_i\tau_1}{1+\omega_i^2\tau_1^2} \\ \frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{\omega_i\tau_1}{1+\omega_i^2\tau_1^2} \\ \frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2} \\ 0 \end{pmatrix}^{\mathrm{T}}, \ \chi_i(\tau_1,\tau_2) = -2b_i \begin{pmatrix} \frac{\omega_i\tau_1}{1+\omega_i^2\tau_1^2} \\ \frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2} \\ 0 \end{pmatrix}. \tag{3.14}$$

Therefore, under the definition

$$\begin{aligned} &\Gamma(\tau_1,\tau_2) := \sum_{i=1}^{N}\Gamma_i(\tau_1,\tau_2), \ \Lambda(\tau_1,\tau_2) = \sum_{i=1}^{N}\Lambda_i(\tau_1,\tau_2), \\ &\beta(\tau_1,\tau_2) = \sum_{i=1}^{N}\beta_i(\tau_1,\tau_2), \ \chi(\tau_1,\tau_2) = \sum_{i=1}^{N}\chi_i(\tau_1,\tau_2), \ \gamma = \sum_{i=1}^{N}(a_i^2+b_i^2), \end{aligned} \tag{3.15}$$

problem (3.5) is rewritten as

$$\begin{aligned} \min_{\boldsymbol{x},\tau_1,\tau_2} \quad & [\boldsymbol{x}^{\mathrm{T}}(\Gamma(\tau_1,\tau_2)+\Lambda(\tau_1,\tau_2))\boldsymbol{x} + (\beta(\tau_1,\tau_2)+\chi(\tau_1,\tau_2))^{\mathrm{T}}\boldsymbol{x} + \gamma] \\ \text{subject to} \quad & x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 1, (3.6). \end{aligned} \tag{3.16}$$

From (3.12), (3.14) and (3.15), the functions $\Gamma(\tau_1,\tau_2)$ and $\Lambda(\tau_1,\tau_2)$ are recognised as sums of N fourth order polynomial fractional functions in (τ_1,τ_2), while $\beta(\tau_1,\tau_2)$ and

$\chi(\tau_1, \tau_2)$ are sums of N quadratic fractional functions in (τ_1, τ_2). Like (3.7)-(3.8), the objective function in (3.16) is obviously convex and quadratic in $\boldsymbol{x} = (\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)^{\mathrm{T}}$ if τ_1 and τ_2 are constant. Therefore, (3.16) is also transferred to the parametric optimization as follows,

$$\min_{\tau=(\tau_1,\tau_2)} \; F(\tau_1, \tau_2) \;\; (3.6) \tag{3.17}$$

with

$$F(\tau_1, \tau_2) = \min_{x} \; [\boldsymbol{x}^{\mathrm{T}}(\Gamma(\tau_1, \tau_2) + \Lambda(\tau_1, \tau_2))\boldsymbol{x} + (\beta(\tau_1, \tau_2) + \chi(\tau_1, \tau_2))^{\mathrm{T}}\boldsymbol{x} + \gamma] \;\; : \\ \varepsilon_\infty \geq \varepsilon_0, \;\; \Delta\varepsilon_1 \geq 0, \;\; \Delta\varepsilon_2 \geq 0. \tag{3.18}$$

Note that $\Gamma(\tau_1, \tau_2) \succeq 0$ and $\Lambda(\tau_1, \tau_2) \succeq 0$ for every (τ_1, τ_2) so the computation for $F(\tau_1, \tau_2)$ is a simple convex quadratic problem in just three dimensional variable $x \in R^3_+$. One can call any existing quadratic solver such as SeduMi [Sturm, 1999] for this computation. The following BB technique allow solving nonconvex problem (3.17).

Theoretical framework of BB: The BB method involves two basic operations:

- *Branching.* The initial rectangle $[p^0, q^0]$ with $p^0 = (l_1, l_2)$ and $q^0 = (u_1, u_2)$ is partitioned into finitely many rectangles $[p, q]$. At each iteration, a partition rectangle $M = [p, q]$ is subdivided further into subrectangles by bisection $\tau_{i_0} = (q_{i_0} - p_{i_0})/2$ with $i_0 = \arg\max_{i=1,2}\{(q_i - p_i)/10^{2-i}\}$.

- *Bounding.* Given a partition rectangle $M = [p, q]$ one has to compute a number $L(M)$ such that

$$L(M) \leq \min_{(\tau_1,\tau_2)\in M} F(\tau_1, \tau_2). \tag{3.19}$$

  Let M_κ be the candidate for further partition at iteration κ (M_κ is the one with minimal $L(M)$). To ensure convergence, branching and bounding must be consistent in the sense

$$\min_{(\tau_1,\tau_2)\in M_\kappa} F(\tau_1, \tau_2) - L(M_\kappa) \to 0 \quad \text{as} \quad \kappa \to +\infty. \tag{3.20}$$

Thus, the crucial step in the BB method is bounding. Since function $F(\tau_1, \tau_2)$ is highly nonconvex with unknown partial convexity structure such as d.c. (difference of two convex functions) [Tuy, 1998], its bounding is not so easy. We use the ideas of monotonic optimization [Tuy, 2000] to develop a bounding technique for realisation of BB algorithm for problem (3.17).

3.3.2.1 Monotonic optimization-Based Bounding

Before going to monotonic optimization based bounding for function $F(\tau_1, \tau_2)$ on a rectangle $[p, q]$ we recall some basic concepts of monotonic functions and properties that will be frequently used in our development. A function $f : R_+^2 \to R$ is said to be increased if $f(\tau) \geq f(\tau')$ for all $\tau \geq \tau' \in R_+^2$. Function $f := f_1 - f_2$ with increasing functions f_1 and f_2 is called d.m. (difference of two monotonic functions) [Tuy, 2000]. Obviously, a lower bound of such d.m. function f on $[p, q]$ is

$$f_1(p) - f_2(q).$$

Suppose that $\alpha \geq \alpha' \in R_+^n$. Then $\alpha\alpha^{\mathrm{T}} - \alpha'(\alpha')^{\mathrm{T}} \geq 0$ and so whenever $x \in R_+^n$,

$$\alpha^{\mathrm{T}}\boldsymbol{x} \geq (\alpha')x \quad \& \quad x^{\mathrm{T}}(\alpha\alpha^{\mathrm{T}})x \geq x^{\mathrm{T}}(\alpha'(\alpha')^{\mathrm{T}})x.$$

Now, using the above monotonic concepts, it is easily seen that for $E_1(\omega_i)$ defined by (3.11) and $(\tau_1, \tau_2) \in [p, q]$,

$$\begin{array}{rcl} E_1(\omega_i) & \geq & F_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty) \\ & := & \boldsymbol{x}^{\mathrm{T}}\Upsilon_i(q)\boldsymbol{x} + \xi_i^{\mathrm{T}}(p)\boldsymbol{x} + a_i \end{array} \tag{3.21}$$

for

$$\Upsilon_i(q) = \begin{pmatrix} \frac{1}{1+\omega_i^2 q_1^2} \\ \frac{1}{1+\omega_i^2 q_2^2} \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{1+\omega_i^2 q_1^2} \\ \frac{1}{1+\omega_i^2 q_2^2} \\ 1 \end{pmatrix}^{\mathrm{T}}, \ \xi_i(p) = -2a_i \begin{pmatrix} \frac{1}{1+\omega_i^2 p_1^2} \\ \frac{1}{1+\omega_i^2 p_2^2} \\ 1 \end{pmatrix}$$

Considering $E_2(\omega_i)$, the monotonicity of its terms are based on the sign of their gradients defined by

$$\frac{d}{d\tau_j}\frac{\tau_j}{1+\omega_i^2\tau_j^2} = \frac{1-\omega_i^2\tau_j^2}{(1+\omega_i^2\tau_j^2)^2}, j = 1, 2. \tag{3.22}$$

As $1-\omega_i^2\tau_1^2 < 0$ for all ω_i and τ_1 in their definition domain, the function $\tau_1/(1+\omega_i^2\tau_1^2)$ is decreasing and

$$\frac{\tau_1}{1+\omega_i^2\tau_1^2} \geq \frac{q_1}{1+\omega_i^2 q_1^2}$$

as well as,

$$-\frac{\tau_1}{1+\omega_i^2\tau_1^2} \geq -\frac{p_1}{1+\omega_i^2 p_1^2}.$$

Therefore, whenever $(\tau_1, \tau_2) \in [p, q]$,

$$\begin{aligned} E_2(\omega_i) &\geq \left(\frac{\omega_i\Delta\varepsilon_1 q_1}{1+\omega_i^2 q_1^2} + \min_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2}\right\}\right)^2 + b_i^2 \\ &\quad -2b_i\left(\frac{\omega_i\Delta\varepsilon_1 p_1}{1+\omega_i^2 p_1^2} + \max_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\Delta\varepsilon_2\tau_2}{1+\omega_i^2\tau_2^2}\right\}\right) \\ &:= G_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty) \\ &= \boldsymbol{x}^{\mathrm{T}}\Psi_i(p,q)\boldsymbol{x} + \zeta_i(p,q)^{\mathrm{T}}\boldsymbol{x} + b_i^2 \end{aligned} \tag{3.23}$$

with

$$\begin{aligned} \Psi_i(p,q) &= \begin{pmatrix} \dfrac{\omega_i q_1}{1+\omega_i^2 q_1^2} \\ \min\limits_{\tau_2\in[p_2,q_2]}\left\{\dfrac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} \\ 0 \end{pmatrix} \times \begin{pmatrix} \dfrac{\omega_i q_1}{1+\omega_i^2 q_1^2} \\ \min\limits_{\tau_2\in[p_2,q_2]}\left\{\dfrac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} \\ 0 \end{pmatrix}^{\mathrm{T}}, \\ \zeta_i(p,q) &= -2b_i \begin{pmatrix} \dfrac{\omega_i p_1}{1+\omega_i^2 p_1^2} \\ \max\limits_{\tau_2\in[p_2,q_2]}\left\{\dfrac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} \\ 0 \end{pmatrix} \end{aligned} \tag{3.24}$$

For each ω_i, min and max functions in (3.24) admit the following analytical forms.

- Whenever $\frac{1}{\omega_i} \leq p_2$, the gradient of the function $\tau_2/(1+\omega_i^2\tau_2^2)$ defined by (3.22) is non-positive, so it is decreasing and

$$\min_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \frac{\omega_i q_2}{1+\omega_i^2 q_2^2}, \quad \max_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \frac{\omega_i p_2}{1+\omega_i^2 p_2^2}\ .$$

- Whenever $\frac{1}{\omega_i} \geq q_2$,the gradient of the function $\tau_2/(1+\omega_i^2\tau_2^2)$ is non-negative, so it is increasing and

$$\min_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \frac{\omega_i p_2}{1+\omega_i^2 p_2^2},\ \max_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \frac{\omega_i q_2}{1+\omega_i^2 q_2^2}\ .$$

- Whenever $\frac{1}{\omega_i} \in [p_2,\ q_2]$, the gradient of the function $\tau_2/(1+\omega_i^2\tau_2^2)$ is non-negative on $[p_2, 1/\omega_i]$ and non-positive on $[1/\omega_i, q_2]$. This means the function is increasing on $[p_2, 1/\omega_i]$ and decreasing on $[1/\omega_i, q_2]$. Consequently,

$$\min_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \min_{\tau_2\in\{p_2,q_2\}}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\},$$
$$\max_{\tau_2\in[p_2,q_2]}\left\{\frac{\omega_i\tau_2}{1+\omega_i^2\tau_2^2}\right\} = \frac{\omega_i/\omega_i}{1+\omega_i^2(1/\omega_i)^2} = 0.5\ .$$

In summary, a lower bound of $F(\tau_1,\tau_2)$ in $[p,q]$ is

$$L(p,q) := \min_{x=(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)} \sum_{i=1}^{N}(F_i(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)+G_i(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)) : \\ \varepsilon_\infty \geq \varepsilon_0,\ \Delta\varepsilon_1 \geq 0,\ \Delta\varepsilon_2 \geq 0. \tag{3.25}$$

By (3.21) and (3.23),

$$\sum_{i=1}^{N}(F_i(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)+G_i(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)) = \\ \boldsymbol{x}^{\mathrm{T}}(\Upsilon(q)+\Psi(p,q))\boldsymbol{x} + (\xi(p)+\psi(p,q))^{\mathrm{T}}\boldsymbol{x}+\gamma, \tag{3.26}$$

where

$$\begin{aligned}\Upsilon(q) &= \sum_{i=1}^{N}\Upsilon_i(p),\ \Psi(p,q)=\sum_{i=1}^{N}\Psi_i(p,q),\\ \xi(p) &= \sum_{i=1}^{N}\xi_i(p),\ \psi(p,q)=\sum_{i=1}^{N}\psi_i(p,q).\end{aligned} \tag{3.27}$$

Since both $\Upsilon(q)$ and $\Psi(p,q)$ are obviously positive definite matrices, the lower bound $L(p,q)$ in (3.25) can be solved by the optimal value of the following simple three-dimensional convex quadratic problem

$$L(p,q) = \min_{x=(\Delta\varepsilon_1,\Delta\varepsilon_2,\varepsilon_\infty)} \boldsymbol{x}^{\mathrm{T}}(\Upsilon(q)+\Psi(p,q))\boldsymbol{x} + (\xi(p) \\ +\psi(p,q))^{\mathrm{T}}\boldsymbol{x}+\gamma\ : \\ \varepsilon_\infty \geq \varepsilon_0, \Delta\varepsilon_1 \geq 0,\ \Delta\varepsilon_2 \geq 0. \tag{3.28}$$

It is worth mentioning that as $p \nearrow q$,

$$F_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty) \nearrow E_1(\omega_i),\ G_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty) \nearrow E_2(\omega_i),$$

that leads to

$$L(p, q) \nearrow F(q) \quad \text{as} \quad p \nearrow q \tag{3.29}$$

which verifies the consistence condition (3.20) for the convergence of the corresponding BB algorithm.

3.3.2.2 Branching and Bounding Algorithm

With the bounding developed in the previous subsection, the implementation of the BB algorithm is as follows

- *Initialisation.* Start with $M_i = [p^i, q^i]$, $i = 1, 2, ..., L$. Set $\mathcal{S}_1 = \mathcal{N}_1 = \{M_1, M_2, ..., M_L\}$. Set $\kappa = 1$. Set $\mu = \min_{i=1,2,...,M}\ F((p^i + q^i)/2)$.

- *Step 1.* For each $M = [p, q] \in \mathcal{N}_\kappa$ compute $F((p + q)/2)$ by (3.18). Update current best value (CBV) $\mu \to F((p + q)/2)$ and current best solution τ_{opt} if $F((p + q)/2) < \mu$. Solve (3.28) to obtain $L(M) = L(p, q)$.

- *Step 2.* Delete all M such that $L(M) \geq \mu - \epsilon$ ($\epsilon > 0$ is a given tolerance). Let $\mathcal{R}_\kappa$ be the set of remaining rectangles. If $\mathcal{R}_\kappa = \emptyset$, terminate: μ is the optimal value of (3.17) with tolerance ϵ, i.e. $0 \leq \mu - F^* \leq \epsilon$, where F^* is the global optimal value of (3.17).

- *Step 3.* Choose $M_\kappa \in \arg\min\{L(M) : M \in \mathcal{R}_\kappa\}$ and divide it into two smaller rectangles $M_{\kappa,1}, M_{\kappa,2}$ according to the above bisection rule. Let $\mathcal{N}_{\kappa+1} = \{M_{\kappa,1}, M_{\kappa,2}\}$, $\mathcal{S}_{\kappa+1} = (\mathcal{R}_\kappa \setminus M_\kappa) \cup \mathcal{N}_{\kappa+1}$.
  Set $\kappa \to \kappa + 1$ and go back to Step 1.

The global convergence of the above algorithm is shown in the following theorem.

Theorem 1 *The above algorithm will terminate after finitely many iterations, yielding the optimal value of (3.17) with tolerance ϵ.*

Proof. The proof is similar to that for [Tuan and Apkarian, 2000, Th. 2]. Suppose that the algorithm is infinite. By [Tuy, 1998, Th. 5.5], the above bisection rule with the consistence condition (3.29) guarantee that the algorithm generates a sequence of nested rectangles $M_\nu = [p^\nu, q^\nu] \to M^* = \{p^*\}$ such that $L(M_\nu) \nearrow F(p^*) \geq F^*$. But by construction procedure in Step 2, we have $L(M_\nu) < F^* - \epsilon$, so $F(p^*) \leq F^* - \epsilon$, a contradiction. □

3.3.3 Minimax optimization

Most contrasts between the normal and cancerous skin permittivities occur in the interval [0.3, 0.8] THz of the refractive index [Bennett et al., 2011], [Wallace et al., 2006]. A drawback of the sum-of-square-error formulation (3.5) is that it does not always control the error distribution of the optimized DD model, i.e the errors between different frequency ranges are not equal. Particularly, the performance of the simulation at the low frequencies, where important contrast between normal and cancerous skin tissue is present, is worse than that at the higher spectral range. To address this issue, a more flexible minimax optimization also using the BB algorithm (BB-MNO), which can control the error quality well in all frequencies, is introduced. Thus, it is expected that the extracted DD parameters using this method should be more meaningful.

Derivation of the BB-MNO starts with dividing the frequency index set $\mathcal{N} = \{1, 2, ..., N\}$ into two disjoint subsets $\mathcal{N}_1$ and $\mathcal{N}_2$ such that $\omega_i \leq \nu, i \in \mathcal{N}_1$ and $\omega_i > \nu, i \in \mathcal{N}_2$ given a threshold ν. In other words, $\{\omega_i, i \in \mathcal{N}_1\}$ covers a lower frequency range, while the set $\{\omega_i, i \in \mathcal{N}_2 := \mathcal{N} \setminus \mathcal{N}_1\}$ covers a higher frequency range. On the other hand, the frequency-dependent complex permittivity exhibits a larger variation scale of its imaginary part than its real part. Consequently, there is a difference in the error level between these two index parts which could not be incorporated in (3.5). Therefore, the

so-called weight factors must be introduced to balance the error level between the real and imaginary components. Accordingly we modify $F(\tau_1, \tau_2)$ into

$$F(\tau_1, \tau_2) = F_1(\tau_1, \tau_2) + F_2(\tau_1, \tau_2) \tag{3.30}$$

with

$$\begin{aligned} F_1(\tau_1, \tau_2) = &\sum_{i \in \mathcal{N}_1} k_{11} E_1(\omega_i) + k_{21} E_2(\omega_i), \\ F_2(\tau_1, \tau_2) = &\sum_{i \in \mathcal{N}_2} k_{12} E_1(\omega_i) + k_{22} E_2(\omega_i), \end{aligned}$$

where k_{11}, k_{21}, k_{12}, k_{22} are weight factors for the real and imaginary parts. Instead of (3.16), the following minimax problem is considered,

$$\min_{\tau=(\tau_1,\tau_2)} F_{MN}(\tau_1, \tau_2) := \max_{j=1,2} F_j(\tau_1, \tau_2) \ : \ (3.6). \tag{3.31}$$

Note that for each (τ_1, τ_2), $F_{MN}(\tau_1, \tau_2)$ is still computed by the following convex quadratic problem

$$\begin{aligned} \min_{\boldsymbol{x},t} \ t \ : \ & (3.5b), \ \boldsymbol{x}^{\mathrm{T}}(\Gamma_j(\tau_1, \tau_2) + \Lambda_j(\tau_1, \tau_2))\boldsymbol{x} + (\beta_j(\tau_1, \tau_2) \\ & + \chi_j(\tau_1, \tau_2))^{\mathrm{T}}\boldsymbol{x} + \gamma_j \le t, \ j = 1, 2, \end{aligned} \tag{3.32}$$

for

$$\begin{aligned} \Gamma_j(\tau_1, \tau_2) &:= \sum_{i \in \mathcal{N}_j} k_{1j} \Gamma_i(\tau_1, \tau_2), \\ \Lambda_j(\tau_1, \tau_2) &= \sum_{i \in \mathcal{N}_j} k_{2j} \Lambda_i(\tau_1, \tau_2), \\ \beta_j(\tau_1, \tau_2) &= \sum_{i \in \mathcal{N}_j} k_{1j} \beta_i(\tau_1, \tau_2), \\ \chi_j(\tau_1, \tau_2) &= \sum_{i \in \mathcal{N}_j} k_{2j} \chi_i(\tau_1, \tau_2), \\ \gamma_j &= \sum_{i \in \mathcal{N}_j} (k_{1j} a_i^2 + k_{2j} b_i^2), j = 1, 2 \end{aligned} \tag{3.33}$$

The above BB algorithm is still directly applicable for the global optimal solution of minimax problem (3.31) with the lower bounding $\bar{F}(p, q)$ in (3.28) replaced by the

following convex quadratic problem

$$\begin{aligned}\bar{F}(p,q) = \min_{\boldsymbol{x},t}\ t\ :\quad & \boldsymbol{x}^{\mathrm{T}}(\Upsilon_j(q)+\Psi_j(p,q))\boldsymbol{x} + (\xi_j(p) \\ & +\psi_j(p,q))^{\mathrm{T}}\boldsymbol{x} + \gamma_j \le t,\ j=1,2,\end{aligned} \tag{3.34}$$

where

$$\begin{aligned}\Upsilon_j(q) = \sum_{i\in\mathcal{N}_j} k_{1j}\Upsilon_i(p),\ \Psi_j(p,q) = \sum_{i\in\mathcal{N}_j} k_{2j}\Psi_i(p,q), \\ \xi(p) = \sum_{i\in\mathcal{N}_j} k_{1j}\xi_i(p),\ \psi(p,q) = \sum_{i\in\mathcal{N}_j} k_{2j}\psi_i(p,q), j=1,2.\end{aligned} \tag{3.35}$$

3.4 Numerical Results and Discussion

In this section, simulation results obtained from the proposed algorithms are presented. The data used for the simulation was extracted from the references [Pickwell et al., 2005; Wallace et al., 2006] for the frequency band approximately between $[0.2, 2]$ THz. This data includes measured complex permittivities of human skin samples, both normal and BCC. The published data is also necessary for result comparison in order to show the algorithm improvement in optimizing the model-fitting procedure.

Each grinding-based optimization process uses 6400 grid points, corresponding to $M = 80$. Fig. 3.1-3.2 demonstrate the measured complex permittivities of normal skin (NS) and BCC samples from [Pickwell et al., 2005], the simulated values with the fitted DD model by this reference using NLS, and those by the grinding-based method. It is obvious that the grinding-based technique provides a better fitting performance than the reference, especially at the frequency range below 1 THz. Extracted values of the DD parameters using the grind-based method are recorded in Table. 3.1.

As regards the BB-GO algorithm, although each term $F_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)+G_i(\Delta\varepsilon_1, \Delta\varepsilon_2, \varepsilon_\infty)$ is a tight lower bound for $E(\omega_i)$ (see (3.21) and (3.18)), their sum defined by (3.26) is obviously much looser lower bound for the objective function $F(\tau_1, \tau_2)$ of (3.5) on the rectangle $[p, q]$. This fact should not be a surprise since $F(\tau_1, \tau_2)$ is implicitly a fraction of high order polynomials. Due to the convergence of the lower bound function in

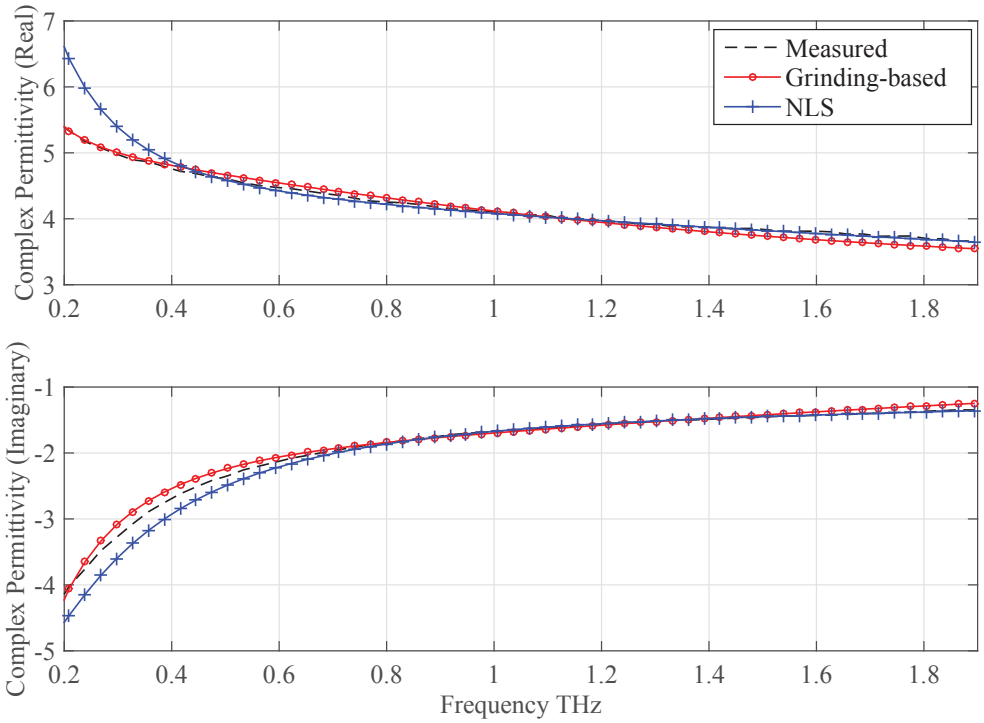

Figure 3.1: The real and imaginary part of the relative complex permittivities of normal skin plotted in the range [0.15-1.68] THz from: the measured data of Pickwell et al. [2005], the simulated spectra using the DD models obtained by Pickwell et al. [2005] and the grinding-based algorithm

(3.29), the first step for the implementation of BB algorithm is to narrow down the area for its branching and bounding. Thus, $[p, q]$ is divided into few smaller rectangles. The BB process is initialised with these rectangles. A rectangle $M_\kappa = [p_\kappa, q_\kappa]$ selected for the next iteration must simultaneously contain the minimum lower bound $L(M_\kappa)$ and the current best solution. Then, the BB algorithm restarts with the extended rectangle $M = [\frac{3}{4}p, \frac{5}{4}q]$.

Fig. 3.3-3.4 depict the measured complex permittivities of normal skin and BCC samples from [Pickwell et al., 2005] and compare them with the DD simulations by the existing NLS and BB-GO. Overall, the global-optimization-based modelling clearly tracks the measured values far better than the NLS-based approach of Pickwell et al. [2005]. The fitting quality of the double Debye model is especially improved significantly at

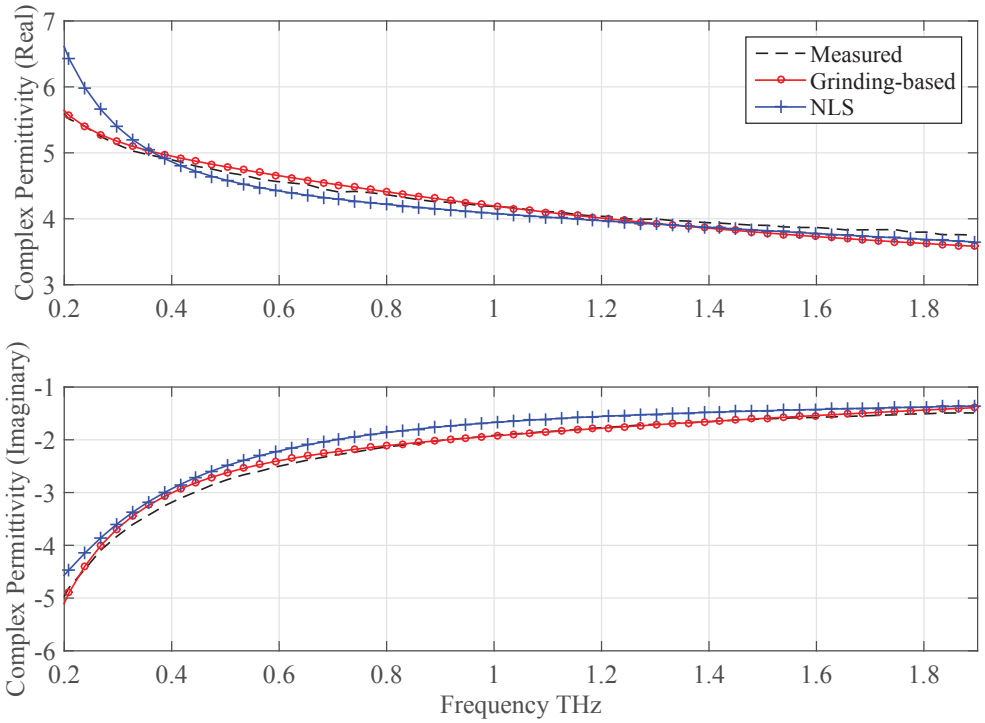

Figure 3.2: The real and imaginary part of the relative complex permittivities of BCC plotted in the range [0.15-1.68] THz from: the measurement of Pickwell et al. [2005], the simulated spectra using the DD models obtained by Pickwell et al. [2005] and the grinding-based algorithm

the lower frequencies by the proposed BB-GO. A consistent performance of the BB-GO has been also seen in other skin samples from [Wallace et al., 2006] according to Fig. 3.5-3.6.

The extracted values of five DD parameters for various skin samples from the references [Pickwell et al., 2005; Wallace et al., 2006] are presented in Table 3.1. In addition, sum-of-square-error values (SSE) indicating the optimal values of the objective function in (3.5) are also provided in this table to compare the fitting errors of the conventional NLS, grinding-based and BB-GO methods. The DD parameters extracted by Pickwell et al. [2005]; Wallace et al. [2006] using the NLS method are not quite optimal as their simulations exhibit a large deviation from the measured complex permittivities at low frequencies. The superiority of both the BB-GO and grinding-based methods over the

Case	Method	$(\Delta\varepsilon_1, \boldsymbol{\varepsilon}_2, \varepsilon_\infty)$	(τ_1, τ_2)	TSEV	QPs
NS-1	NLS	$(10.5, 1.6, 2.6)$	$(1.5, 0.06)$	3.6	-
NS-1	Grinding	$(22.6, 1.7, 2.9)$	$(4.1, 0.11)$	0.3	6400
NS-1	BB-GO	$(21.8, 1.7, 2.9)$	$(3.9, 0.11)$	0.3	3970
BCC-1	NLS	$(13.4, 1.6, 2.7)$	$(1.6, 0.06)$	5.3	-
BCC-1	Grinding	$(31.6, 1.8, 3.0)$	$(4.8, 0.11)$	0.5	6400
BCC-1	BB-GO	$(31.9, 1.8, 3.0)$	$(4.9, 0.12)$	0.5	3880
NS(Case 266)	Grinding	$(20.2, 1.7, 3.0)$	$(3.9, 0.11)$	0.6	6400
NS(Case 266)	BB-GO	$(19.9, 1.7, 3.0)$	$(3.8, 0.11)$	0.6	510
BCC(Case 266)	Grinding	$(29.4, 2.0, 3.0)$	$(4.1, 0.11)$	0.5	6400
BCC(Case 266)	BB-GO	$(29.5, 2.0, 3.0)$	$(4.1, 0.12)$	0.5	690
NS(Average)	Global	$(25.7, 1.8, 2.9)$	$(4.8, 0.10)$	-	-
NS(Average)	BB-MNO	$(26.8, 1.7, 3.0)$	$(4.9, 0.11)$	-	-
BCC(Average)	BB-GO	$(72.3, 2.0, 3.0)$	$(11.0, 0.13)$	-	-
BCC(Average)	BB-MNO	$(55.3, 1.8, 3.1)$	$(8.8, 0.14)$	-	-

Table 3.1: The double Debye parameters extracted from the measured complex permittivities of the skin samples (NS-1 and BCC-1 from [Pickwell et al., 2005]; NS(Case 266), BCC(Case 266), NS(Average), BCC(Average) from [Wallace et al., 2006]) by the grinding-based algorithm and BB-GO, and those from these references.

NLS is also revealed by their SSEs. Particularly, the SSEs obtained by the global optimization-based methods are orders of magnitude lower than those by the NLS. The difference in SSE between the BB-GO and grinding-based methods is very minor, indicating their similar performance in terms of fitting error. However, it is observed that even very small calibration of SSE often results in considerable differentiation of the corresponding five Debye parameters. This fact makes global optimization highly desirable.

The number of called quadratic problems (QP) corresponding to the simulations of the grinding-based and BB-GO algorithm are listed in the last column of Table 3.1. As mentioned earlier, it is found that the grinding-based algorithm requires computation of the quadratic objective function of (3.5) at at least $80\times80 = 6400$ grid points to ensure the optimality of the solution. Thus, each simulation with the grinding-based technique involves 6400 QPs as can be seen in Table 3.1. It is also apparent that the BB-GO approach is much more computationally efficient than the grinding-based

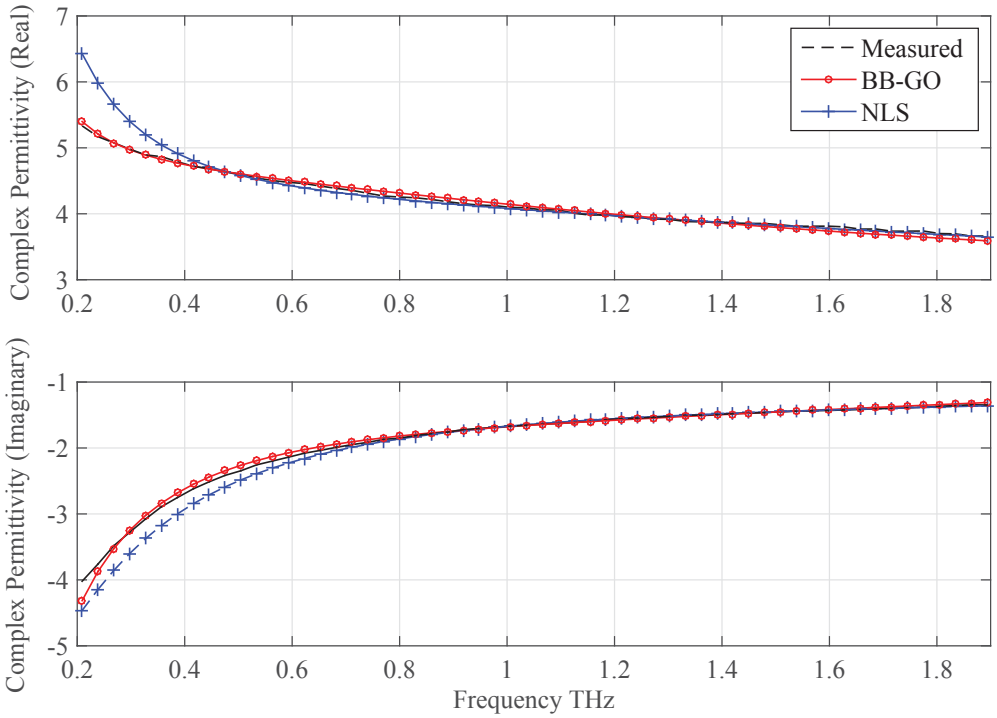

Figure 3.3: The real and imaginary part of the relative complex permittivities of normal skin: the measurement of Pickwell et al. [2005] and the simulated spectra using the DD models obtained by Pickwell et al. [2005] and the BB-GO

one. Indeed, the number of required quadratic problem solvers for implementation of the former is approximately a half or quarter as much as that for implementation of the latter. For instance, the average computational time for each QPs is about 1/7 second if the computer with Intel Core i5-2400 CPU 3.10 GHz processor and 4.00 GB RAM is utilised for the simulation. Comparing between normal and BCC skin samples, the values of the five double Debye parameters corresponding to these samples are far different, thus suggesting that the DD model can be potentially used for supporting skin cancer detection. This application of the optimal double Debye modelling still requires further studies.

To accurately assess the fitting performance of the global optimization-based algorithms, Table 3.2 provides the numerical values of the so called adjusted R-square

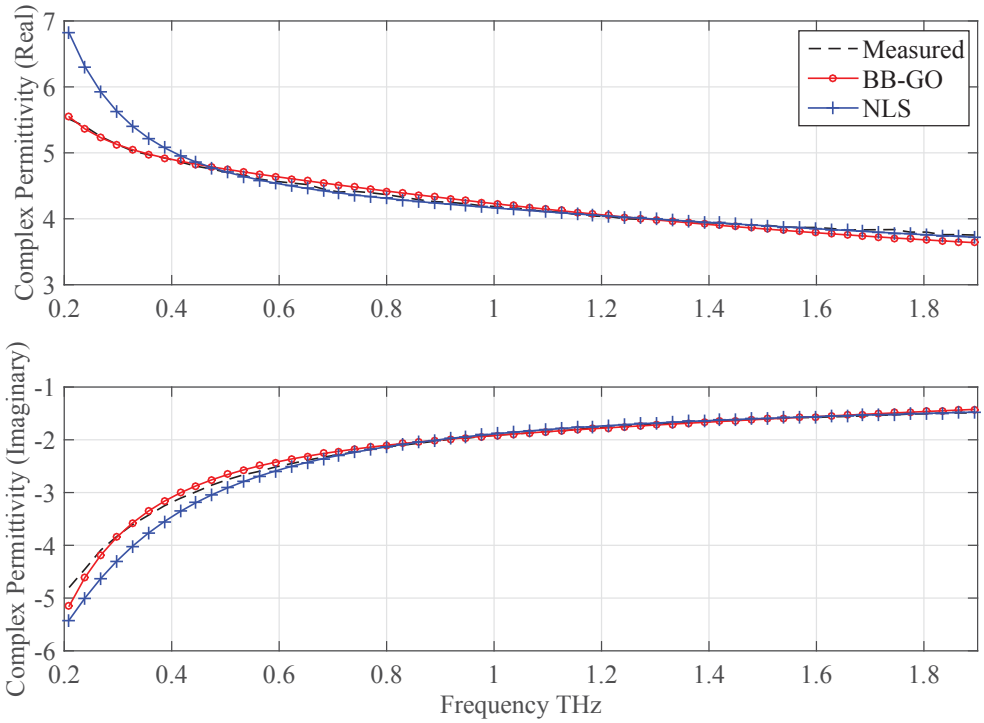

Figure 3.4: The real and imaginary part of the relative complex permittivities of BCC: the measurement of Pickwell et al. [2005] and the simulated spectra using the DD models obtained by Pickwell et al. [2005] and the BB-GO

(R^2_{adj}) (see [MATLAB, 2012]), which presents the goodness-of-fit (GOF) of numerical models, for the simulations of various skin samples. The value $R^2 = 1$ corresponds to the error-free (perfect) fitting. All simulations by the BB-GO provide excellent fitting with the corresponding R^2_{adj}s of around 0.99. One can also see that R^2_{adj}s for the DD models extracted by Pickwell et al. [2005] are considerably smaller than that by the BB-GO, especially for the real part $\Re(\tilde{\epsilon}_{\text{DD}})$. Furthermore, the distributions of the fitting errors for the DD simulation with the BB-GO also undergoes the Kolmogorov-Smirnov test by computing its critical p-values (see e.g. [SPSS, 2011; Lilliefors, 1967]). In most cases, these values are smaller than the significance level of 0.05, which means that these distributions are not normal. To analyze the fitting error property of the BB-GO simulations, the Breusch-Pagan and Koenker test under the null hypothesis

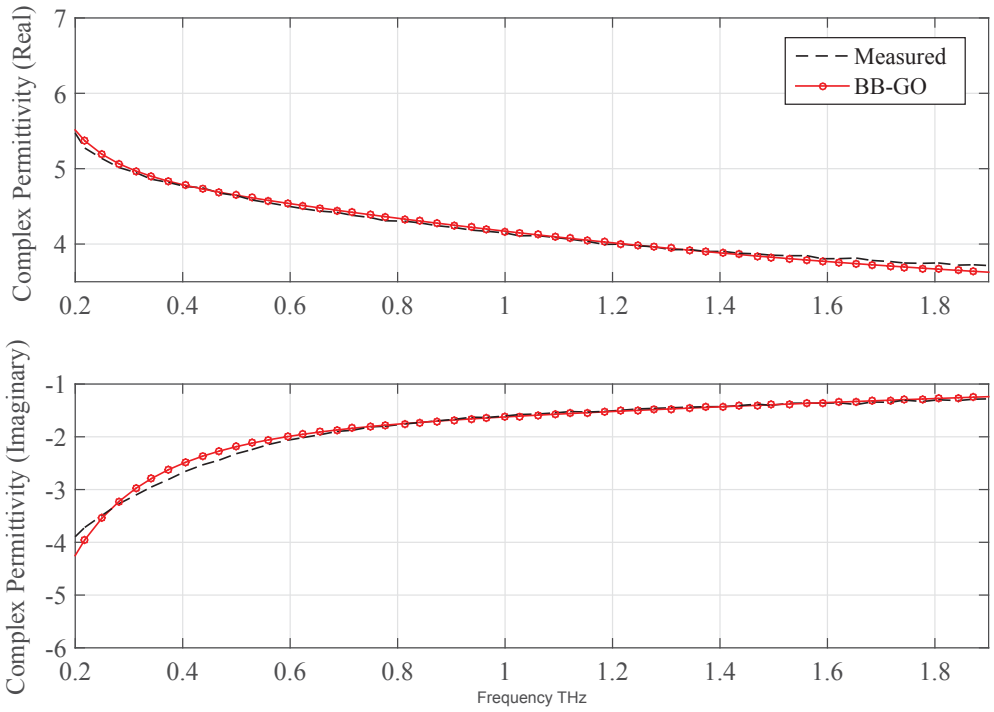

Figure 3.5: The real and imaginary part of the relative complex permittivities of normal skin (case 266) of Wallace et al. [2006]: the measurement and the simulated spectra using the DD models obtained by the BB-GO

of homoscedasticity [Breusch and Pagan, 1979; Koenker, 1981] is carried out by using SPSS Statistics software [SPSS, 2011]. Its p-values corresponding to both the real ($\Re(\tilde{\epsilon}_{\mathrm{DD}})$) and imaginary ($\Im(\tilde{\epsilon}_{\mathrm{DD}})$) parts of the optimal DD models are provided in Table 3.3. The fitting errors for $\Re(\tilde{\epsilon}_{\mathrm{DD}})$ are likely homoscedastic as their corresponding p-values are larger than the significance level of 0.05. The fitting errors are thus likely multiplicative in these cases. In contrast, the fitting errors for $\Im(\tilde{\epsilon}_{\mathrm{DD}})$ are likely heteroscedastic because their corresponding p-values are less than the significance level of 0.05. The fitting errors are then likely additive.

The frequency range 0.3–0.8 THz is of great interest as it provides the most contrast in complex permittivities between normal and cancerous skin tissues. Thus, it is desirable to have the computed model strictly track its measurement in this spectral range.

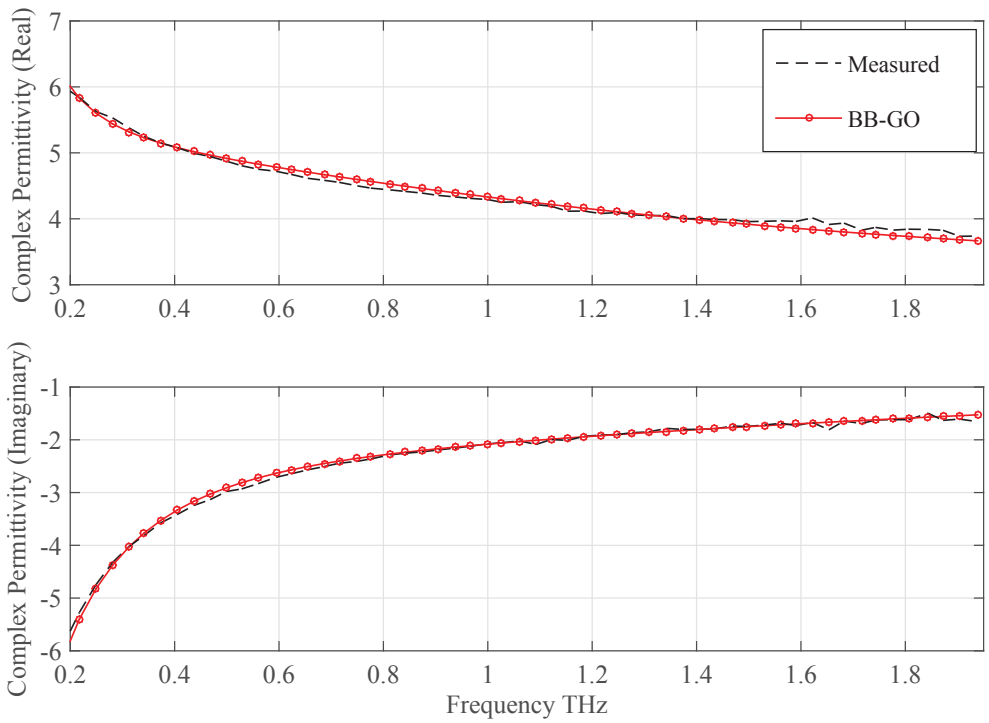

Figure 3.6: The real and imaginary part of the relative complex permittivities of BCC (case 266) of Wallace et al. [2006]: the measurement and the simulated spectra using the DD models obtained by the BB-GO

Therefore, unsatisfactory tracking may lead to loss of useful information for recognizing changes in tissue composition, especially free water, due to carcinogenesis. For instance, Figure 3.7, demonstrates a clear mismatch between the Debye model computed by total square minimisation and its measurement at the interested frequency band [0.3–0.8] THz. To conduct the BB-MNO simulation, values of the weights $k_{ij}, i, j \in [1, 2]$ and the frequency threshold ν for dividing the frequency index set are chosen based on the mismatched area of practical cases. Particularly, BB-MNO requires the following adjustments of these parameters: $\nu = 0.85$; $k_{11} = k_{12} = 4$ and $k_{21} = k_{22} = 1$ for the case 266 of [Wallace et al., 2006]; $k_{11} = k_{12} = 3$ and $k_{21} = k_{22} = 1$ for the average case of [Wallace et al., 2006]. The plots of double Debye models simulated by the BB-MNO in Fig. 3.7-3.8 demonstrate an essentially better tracking quality for their measured values at the frequencies lower than 1 THz. The contrast between normal skin and BCC

Case	Method	R^2_{adj}	
		$\Re(\tilde{\epsilon}_{\text{DD}})$	$\Im(\tilde{\epsilon}_{\text{DD}})$
NS-1	BB-GO	0.99	0.99
NS-1	NLS	0.73	0.95
BCC-1	BB-GO	0.98	0.99
BCC-1	NLS	0.67	0.95
NS(Case 266)	BB-GO	0.99	0.98
BCC(Case 266)	BB-GO	0.98	0.99
NS(Average)	BB-GO	0.98	0.99
BCC(Average)	BB-GO	0.96	0.98

Table 3.2: Values of R^2_{adj} corresponding to fitting the real and imaginary parts of the measured complex permittivities (NS-1 and BCC-1 from [Pickwell et al., 2005]; NS(Case 266), BCC(Case 266), NS(Average), BCC(Average) from [Wallace et al., 2006]) are compared between the BB-GO and NLS.

Case	p-value	
	$\Re(\tilde{\epsilon}_{\text{DD}})$	$\Im(\tilde{\epsilon}_{\text{DD}})$
NS-1	0.40	9.0E-06
BCC-1	0.11	1.2E-07
NS (Case 266)	0.55	2.0E-06
BCC (Case 266)	0.45	5.1E-03
NS (Average)	0.72	1.3E-05
BCC (Average)	0.25	3.2E-06

Table 3.3: p-value of the Breusch-Pagan and Koenker test for the real and imaginary parts of the extracted DD models using the BB-GO for the skin samples (NS-1 and BCC-1 from [Pickwell et al., 2005]; NS(Case 266), BCC(Case 266), NS(Average), BCC(Average) from [Wallace et al., 2006]).

is obviously revealed in these figures, especially in the frequency band [0.3–0.8] THz, hence paving the way for potential application of the BB-MNO in skin classification. Furthermore, as can be seen in Figure 3.9, in the whole usable range up to 2 THz, although the performance of the BB-MNO may not be as good as that of BB-GO, the former still maintains an acceptable tracking. In short, the proposed BB-GO method aims at improvement of the tracking quality in a selected spectral range in which the most information of skin dielectric response can be explored.

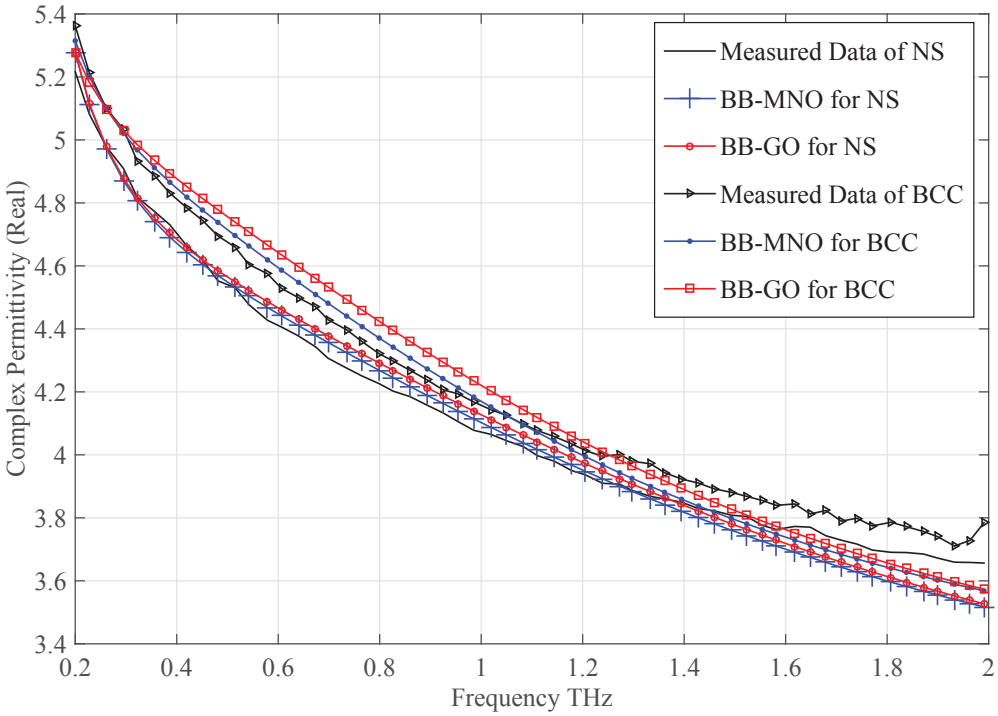

Figure 3.7: The real part of the relative complex permittivities of normal skin and BCC (average) from [Wallace et al., 2006]: the measurement and the simulated spectra using the DD models obtained by the BB-GO and the BB-MNO

3.5 Conclusion

Dielectric properties of human skin tissue in the terahertz frequencies have been described well by the double Debye model. Accurately fitting the measured complex permittivities to this model and extracting the model parameters remains a sufficient condition to guarantee the precise transformation of information from the extremely high-dimensional data to the 5-dimension space of the DD model. Apparently, this transformation should be potential for cancer classification as the contrast in the extracted DD parameters is found between normal skin and BCC. However, this suggestion requires further studies with a larger database for a confirmation. At the current stage, the proposed optimization algorithms, such as the grinding-based technique and BB-GO, for the DD parameter extraction are essential for better understanding and

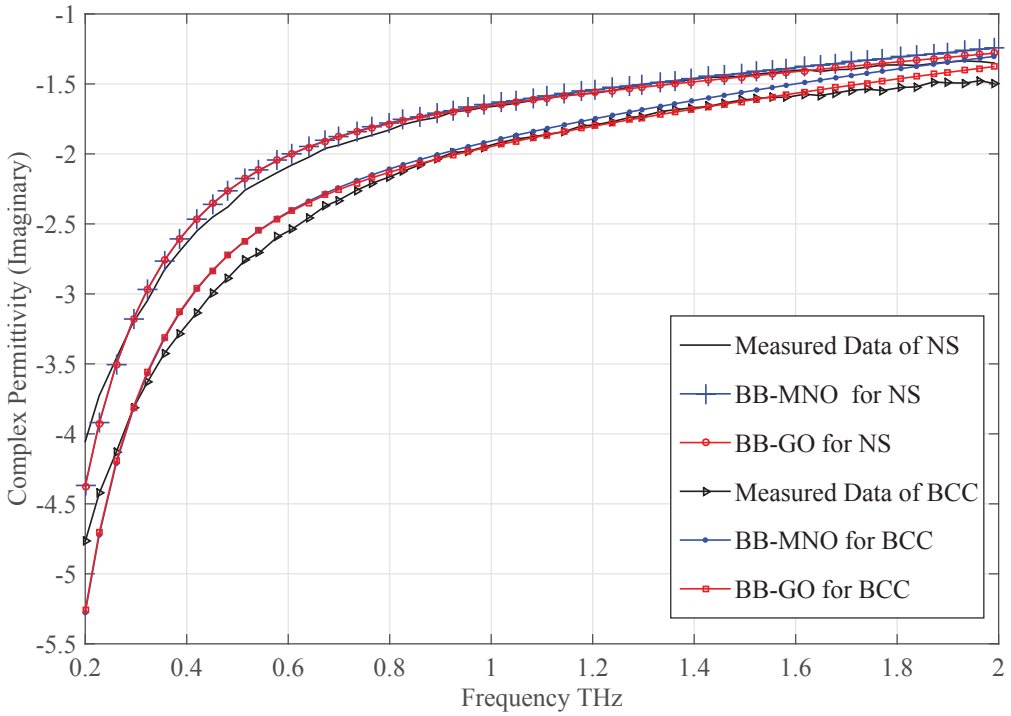

Figure 3.8: The imaginary part of the relative complex permittivities of normal skin and BCC (average) from [Wallace et al., 2006]: the measurement and the simulated spectra using the DD models obtained by the BB-GO and the BB-MNO

exploiting dielectric properties of human skin. It has been shown that the dielectric response of THz waves in skin tissue can be approximated well by the double Debye model if its five deterministic parameters are optimally extracted. The optimal parameter extraction is itself a challenging task but has been handled efficiently by the developed algorithms of global optimization.

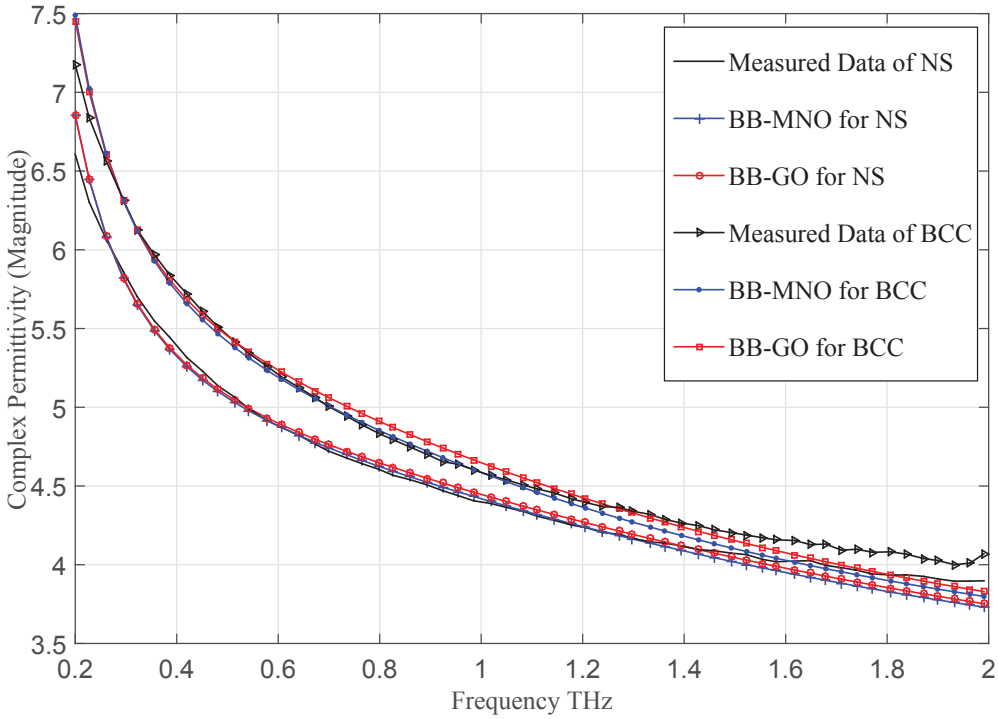

Figure 3.9: The magnitude of the relative complex permittivities of normal skin and BCC in the average case of Wallace et al. [2006] plotted by: the measurement and simulations by the BB-GO and the BB-MNO

Chapter 4

THE POTENTIAL OF THE DOUBLE DEBYE PARAMETERS TO DISCRIMINATE BETWEEN BASAL CELL CARCINOMA AND NORMAL SKIN

4.1 Introduction

The first aim of the study in this chapter is to provide comprehensive statistical analysis of the double Debye parameters in terms of differences between normal skin and BCC. In particular, the optimal values of these parameters are obtained by applying the BB-GO algorithm introduced in section 3.3.2 to fitting the complex permittivities of human skin samples, both normal and BCC. A tighter constraint than that previously applied for this algorithm is proposed to avoid overfitting of the model, which, in turn, reduces the variance of the extracted parameters. They are then statistically analysed through the two-tailed unpaired Student's t-test, normalized mean difference and correlation analysis. The second aim is to pinpoint the potential parameters among them for skin cancer classification. Analysis of the receiver operating characteristic (ROC) curves of these classifiers allows us to obtain those having the best capability of discriminating between normal skin and BCC. Support vector machines (SVM) is employed to classify the data using the most potential classification parameters of the double Debye model. The classification accuracy is assessed by leave-one-out cross-validation and random-subsampling cross-validation. The positive outcomes of this study provide the first glance at the potential of the double Debye parameters for improving the contrast and resolution of THz imaging in BCC.

The structure of this chapter is as follows. Section 4.2 introduces the applied data and fitting procedure. Section 4.3 presents statistical analysis of the double Debye parameters including difference analysis and correlation between normal skin and NMSC. Discrimination capability of these parameters and their classification performance in the SVM classification are covered in Section 4.4. Finally, the main contributions are summarized in section 4.6.

4.2 Model Parameters Extraction

4.2.1 Examined Data

The data used in this study has been previously described in [Wallace et al., 2006]. In particular, skin tissue samples were excised from ten patients undergoing MMS at the Dermatology Department, Addenbrooke's Hospital, Cambridge, United Kingdom. There are thirteen samples containing BCC and their percentages of BCC were estimated by an expert histopathologist. Ten normal skin samples were obtained from the adjacent areas to the tumour site. The TPI spectra1000 (Teraview Limited, Cambridge, United Kingdom) was used to conduct all spectroscopy measurements. This instrument was described in detail by Taday and Newnham [2004]. The TPI system was operated in transmission mode to collect the time domain data with 30 scans per second. Fourier transformation allows one to obtain the frequency domain data from the received signals. The spectral data facilitates retrieving both amplitude and phase information which is used to calculate the frequency-dependent refractive index and absorption coefficient of the sample. The reader is referred to [Wallace et al., 2006] for more details of the sample preparation and measurement procedure.

4.2.2 Extraction Procedure

The numerical value $\epsilon_m(\omega)$ of the frequency-dependent complex permittivity is obtained by using the formula

$$\epsilon_m(\omega) = (n(\omega) - \jmath\frac{c\alpha(\omega)}{2\omega})^2, \tag{4.1}$$

where $n(\omega)$ and $\alpha(\omega)$ are the measured frequency-dependent refractive index and frequency-dependent absorption coefficient. To fit the measured data $\epsilon_m(\omega)$ with its double Debye modelling in (3.1), we minimize the following sum of square errors between $\epsilon_m(\omega_i)$ and

their predicted values $\epsilon_{\mathrm{DD}}(\omega_i)$ according to the equation (3.1):

$$E = \sum_{i=1}^{N} |\epsilon_{\mathrm{DD}}(\omega_i) - \epsilon_m(\omega_i)|^2. \tag{4.2}$$

Here N is the number of data points used to fit the model over the frequency range (0.2–1.8) THz. This optimization problem is rewritten as

$$\begin{aligned} \min_{\varepsilon_s,\varepsilon_2,\varepsilon_\infty,\tau_1,\tau_2} E \ : \ & \varepsilon_\infty \geq 1, \varepsilon_s \geq \varepsilon_2, \ \varepsilon_2 \geq \varepsilon_\infty, \\ & p_1 \leq \tau_1 \leq q_1, \ p_2 \leq \tau_2 \leq q_2. \end{aligned} \tag{4.3}$$

Here $[p_1, q_1]$ and $[p_2, q_2]$ are chosen to cover all practical values of τ_1 and τ_2. The BB-GO algorithm introduced in section 3.3.2 is applied to achieve the optimal solution. The following brief description of this algorithm is basic to proposing a change in the applied constraint. The problem (4.3) is transferred into a parametric optimization problem in two variables (τ_1, τ_2)

$$\begin{aligned} \min_{\tau_1,\tau_2} \ F(\tau_1, \tau_2) \quad : \quad & p_1 \leq \tau_1 \leq q_1 \\ & p_2 \leq \tau_2 \leq q_2. \end{aligned} \tag{4.4}$$

Here, $F(\tau_1, \tau_2)$ is the optimal values of the following convex quadratic optimization

$$\begin{aligned} \min_{x=(\varepsilon_s,\varepsilon_2,\varepsilon_\infty)} \quad & \boldsymbol{x}^{\mathrm{T}}(\Gamma(\tau_1,\tau_2) + \Lambda(\tau_1,\tau_2))\boldsymbol{x} + (\beta(\tau_1,\tau_2) + \chi(\tau_1,\tau_2))^{\mathrm{T}}\boldsymbol{x} + \gamma \quad : \\ & \varepsilon_\infty \geq \varepsilon_0, \ \varepsilon_s \geq \varepsilon_2, \ \varepsilon_2 \geq \varepsilon_\infty. \end{aligned} \tag{4.5}$$

Here, the coefficient matrices of this objective quadratic function are specified in section 3.3.2. For fixed (τ_1, τ_2), (4.5) can be effectively solved by an existing quadratic solver such as SeduMi [Sturm, 1999]. The branch-and-bound algorithm described in section 3.3.2.2 is used for performing a global search for the optimal solution (τ_1, τ_2) of (4.4).

More details of the extraction procedure are thoroughly described in section 3.3.2. However, since the optimization problem (4.4) depends on the constraints of (τ_1, τ_2), it is essential to constrain their ranges. The previous results of Pickwell et al. [2005]; Truong et al. [2013] used $(p_1, q_1) = (1, 20)$ ps for τ_1 and $(p_2, q_2) = (50, 150)$ fs for τ_2 for *ex vivo* data, which lead to the high variance and numerically ill-posedness of the

extracted values ε_s. Taking this undesired consequence into account, we found that the overfitting occurs in cases that the spectra of ϵ_m diverge from the standard form of the human skin complex permittivity in some parts of the examined frequency range. Particularly, because ε_s highly correlates with the variation of τ_1, a large value of the latter will result in increasing the former. In fact, it is observable that the values of τ_1 higher than five picoseconds only provide minor improvements in terms of fitting quality, leading to the high variance of ε_s. Therefore, in the present study, we constrain τ_1 to the narrow range [1–5] ps. The extracted values of the double Debye parameters for normal skin and BCC samples were averaged with standard deviations and recorded in Table 4.1. The difference in the fitting quality between using $\tau_1 \in [1, 5]$ in this study

Sample	ε_s	ε_2	ε_∞	$\tau_1(ps)$	$\tau_2(ps)$
Normal	24.97 ± 2.54	4.63 ± 0.16	2.89 ± 0.14	3.82 ± 0.49	0.11 ± 0.01
BCC	33.33 ± 5.48	4.80 ± 0.20	2.90 ± 0.15	4.35 ± 0.67	0.11 ± 0.01

Table 4.1: The double Debye parameters.

and $\tau_1 \in [1, 20]$ in [Truong et al., 2014] is only 0.2% on average based on comparing their adjusted R-square values for the goodness of fit. However, the variance of ε_s is significantly smaller as compared with that in [Truong et al., 2014]. Further statistical analysis presented in the following sections will also support the superiority in the modified extraction approach with the stricter constraint of τ_1.

4.3 Statistical Analysis

In this section, statistical methods are used to analyse the double Debye parameters including ε_s, ε_2, ε_∞, τ_1, τ_2, the imaginary parts of the complex permittivity estimated by the double Debye model (denoted by $\Im(\epsilon_{\mathrm{DD}})$), and the imaginary parts of the two Debye relaxation terms (respectively denoted by $\Im(D_1)$ and $\Im(D_2)$ with $D_1 = (\varepsilon_s - \varepsilon_2)/(1 + \jmath\omega\tau_1)$ and $D_2 = (\varepsilon_2 - \varepsilon_\infty)/(1 + \jmath\omega\tau_2)$).

4.3.1 Student t-test

Two-tailed independent samples Student's t-test is used to determine the statistical differences between normal skin and BCC samples for the eight aforementioned parameters of the double Debye model and the measured complex permittivities [Altman, 1991; Salomatina et al., 2006]. For this test, null hypothesis refers to the statement that statistical difference in means of two sample groups is not significant. The unknown variances of these subjects are also taken into account by the t distribution. The so-called p value with a critical value, which is taken to be 0.05 in most practical cases, allows assessment of the null hypothesis. The p-value represents the probability that we observe no difference between two groups in the data. The t-test outcome giving a p-value below the threshold is sufficient to reject the null hypothesis as well as to admit the statistical significance of the difference.

Fig. 4.1 presents the variation in p-values for the real and imaginary part of the measured complex permittivities ϵ_m over the frequency range from 0.2 to 1.8 THz when we compare normal skin with BCC. The p-values for the imaginary part $\Im(\epsilon_m)$ is consistently lower than the critical value 0.05 over the frequency range while those corresponding to the real part $\Re(\epsilon_m)$ slightly vary around 0.05. The interesting differences in average values of the double Debye parameters can be observed in Table 4.1. To analyse the statistical significance of these differences, Table 4.2 shows the p-values corresponding to the five parameters of the double Debye model. Accordingly, ε_s can

DD	ε_s	ε_2	ε_∞	τ_1	τ_2
p-value	< 0.001	0.05	0.78	0.06	0.55

Table 4.2: The p-values for the double Debye parameters to illustrate the significance of difference between normal skin and BCC samples.

provide the statistical significance of difference between normal skin and BCC samples although such the difference is not observed in the remaining parameters. This statistical result agrees well with the previous analysis in literature. Particularly, Wallace

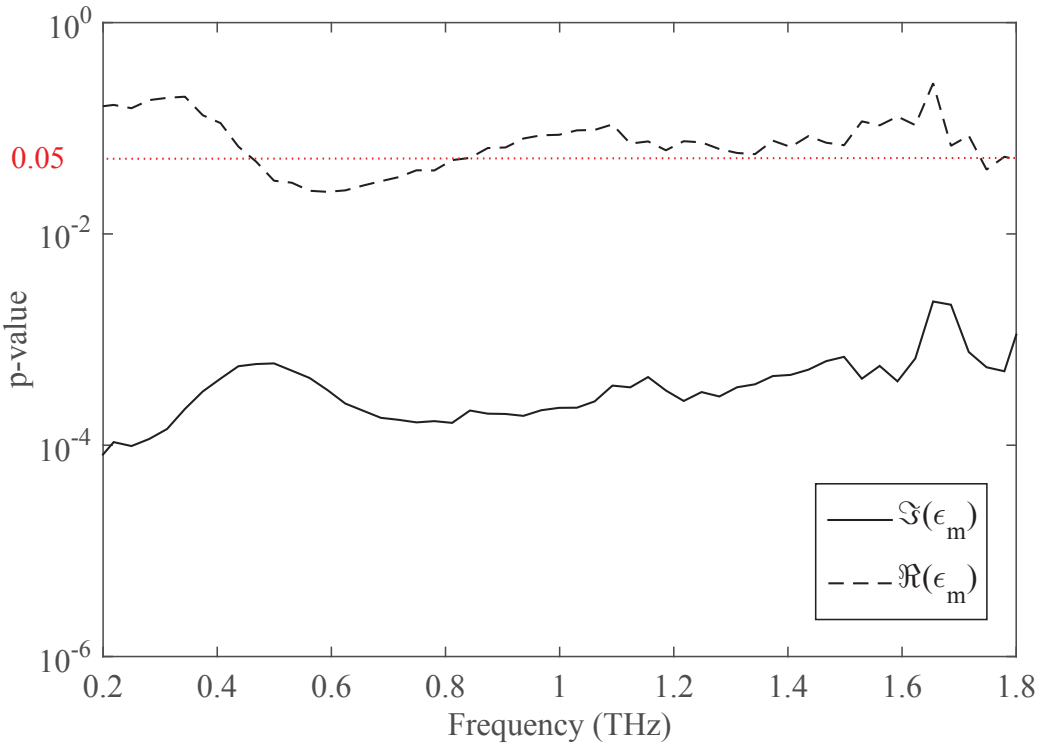

Figure 4.1: The p-values for comparing normal skin and BCC by using the real and imaginary part of the measured complex permittivity ϵ_m over the frequency range of 0.2-1.8 THz. The dotted horizontal red line indicates a p-value equal to 0.05.

et al. [2006] found that the absorption coefficient α in the frequencies below 1 THz offers the highest difference between diseased and normal tissue while ε_s as the low frequency limit parameter is mainly influenced by the change in α. This provides a comprehensive explanation for the significant difference found in ε_s [Pickwell et al., 2005]. Moreover, τ_1 and τ_2 are also so found dependent on α but their small values may affect their discrimination ability. Conversely, the t-test outcomes of ε_2 and ε_∞ may be interpreted by their less sensitivity to α than the refractive index n.

In Fig. 4.2, the p-values for $\Im(\epsilon_{\mathrm{DD}})$, $\Im(D_1)$, and $\Im(D_2)$ are plotted versus the frequency range $[0.2 - 1.8]$ THz. Accordingly, all these Debye parameters can provide significant differences between normal skin and BCC due to their p-values below 0.05 over this whole frequency range. Several bumps in the p-values of $\Im(\epsilon_m)$ (in Fig 4.1) are not seen

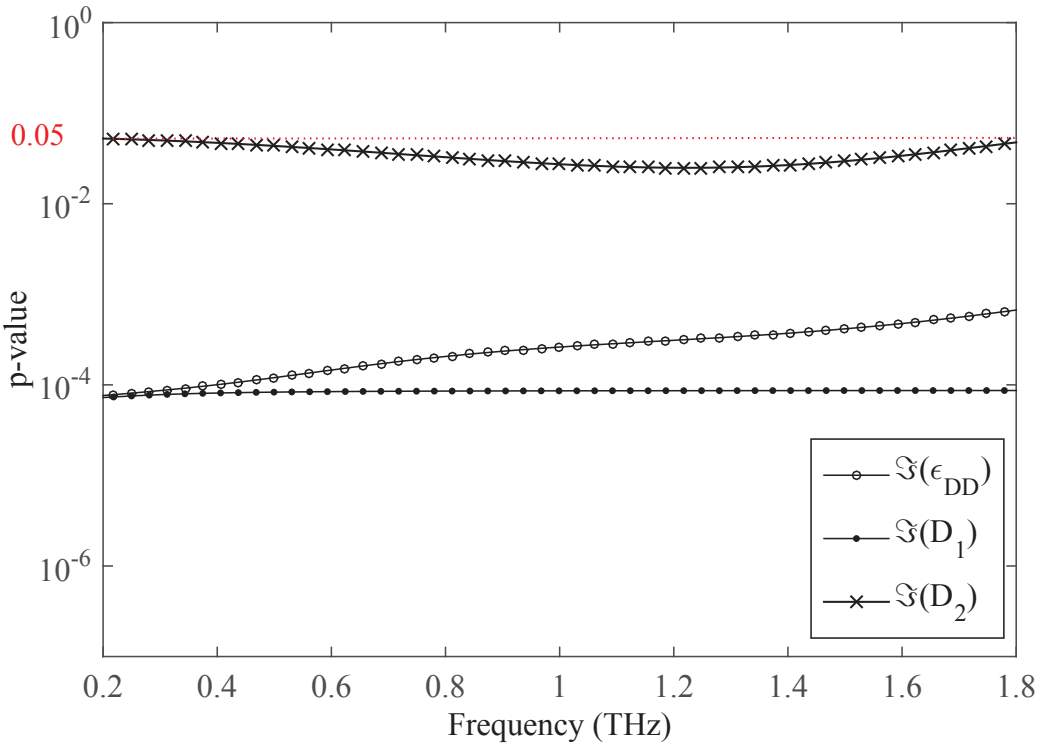

Figure 4.2: The p-values for comparing normal skin and BCC by using the values of $\Im(\epsilon_{\mathrm{DD}})$, $\Im(D_1)$ and $\Im(D_2)$ respectively over the frequency range of 0.2-1.8 THz. The dotted horizontal red line indicates a p-value equal to 0.05.

in that of $\Im(\epsilon_{\mathrm{DD}})$ because the double Debye model may smooth out some variations in $\Im(\epsilon_m)$. The similar response of $\Im(\epsilon_m)$ in Fig. 4.1 and $\Im(\epsilon_{\mathrm{DD}})$ in Fig. 4.2 indicates the capability of the double Debye model to maintain the significant difference between normal skin tissue and BCC in complex permittivity.

4.3.2 Normalized Percentage Difference in Mean

While the t-test can facilitate the analysis of potential differences between normal skin and BCC in the DD parameters, the normalized percentage difference in means is capable of assessing the level of these differences. The normalized percentage differences in ε_s, ε_2 , ε_∞, τ_1, τ_2 are presented in Table 4.3. ε_s shows 33.5% of the difference which is the highest among the examined DD parameters. This result indicate that ε_s is

DD	ε_s	ε_2	ε_∞	τ_1	τ_2
Difference(%)	33.49	3.64	0.63	13.96	2.91

Table 4.3: The normalized percentage difference in means of the double Debye parameters between the normal skin and BCC samples.

the most sensitive feature for skin cancer discrimination. Fig. 4.3 demonstrates the

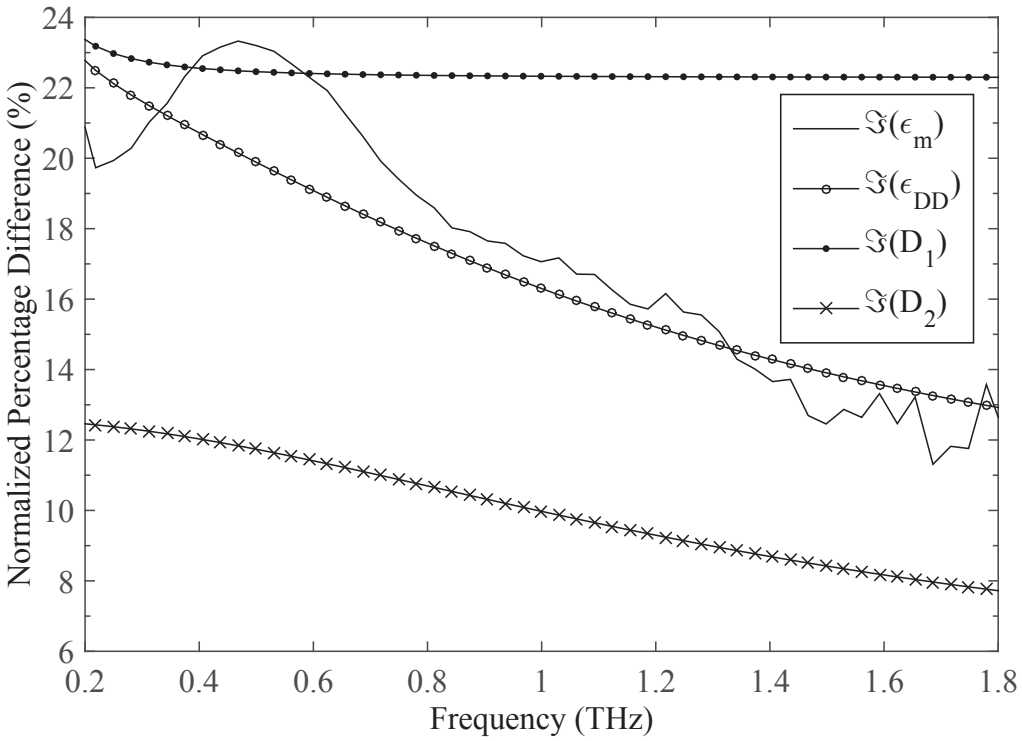

Figure 4.3: Normalized percentage difference between normal and cancerous skin tissues in mean of the measured absorption coefficient, the refractive index, and complex permittivity.

normalized percentage differences in $\Im(\epsilon_m)$, $\Im(\epsilon_{\mathrm{DD}})$, $\Im(D_1)$ and $\Im(D_2)$ between healthy and BCC samples over the frequency range [0.2–1.8] THz. According to this figure, $\Im(\epsilon_{\mathrm{DD}})$, $\Im(D_1)$ and $\Im(D_2)$ obtain the highest percentage differences of 22.5%, 23.5%, 12.5% respectively at 0.2 THz while the difference in $\Im(\epsilon_m)$ reaches its peak of 23.2% at 0.5 THz. The percentage differences in $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(D_2)$ notably reduce at higher frequencies while $\Im(D_1)$ practically retains its contrast level throughout the spectral range.

4.3.3 Correlation Analysis

With the differences found in subsections 4.3.1 and 4.3.2, it is interesting to investigate how the parameters of the double Debye model vary with the percentage tumour content in these samples. The Pearson's correlation coefficients measuring the linear dependence between the double Debye model parameters $\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1, \tau_2$ and the BCC percentage are recorded in Table 4.4. Accordingly, ε_s shows a good correlation $r = 0.70$ with

DD	ε_s	ε_2	ε_∞	τ_1	τ_2
Correlation	0.70	0.51	-0.11	0.37	0.15

Table 4.4: The Pearson's coefficient of correlation between each of the double Debye parameters and the sample tumour content.

the percentage tumour content of samples while the correlations for the remaining parameters are trivial. These findings agree well with the previous analysis of difference in ε_s.

The Pearson's correlation coefficients of $\Im(\epsilon_m)$, $\Im(\epsilon_{\mathrm{DD}})$, $\Im(D_1)$ and $\Im(D_2)$ with the percentage tumour content were calculated at each examined frequency. The correlation responses of these parameters over the range from 0.2 to 1.8 THz are presented in Fig. 4.4. Accordingly, $\Im(D)$ and $\Im(D_1)$ shows fairly stable responses over the examined frequencies with the minimum correlations of 0.78 and 0.77 respectively while the correlation of $\Im(D_2)$ increases with frequency and reaches the maximum of only 0.67. The correlation of $\Im(\epsilon_m)$ is generally similar to that of $\Im(D)$ and $\Im(D_1)$ in the low frequency range (below 1.2 THz) but disperses strongly at higher frequencies. Furthermore, the high correlation of these parameters with tumour content justifies the analysis of difference between normal and BCC in previous subsections. The low correlation of $\Im(D_2)$ also reflects its analysis results in the t-test and normalized percentage difference to a certain extent.

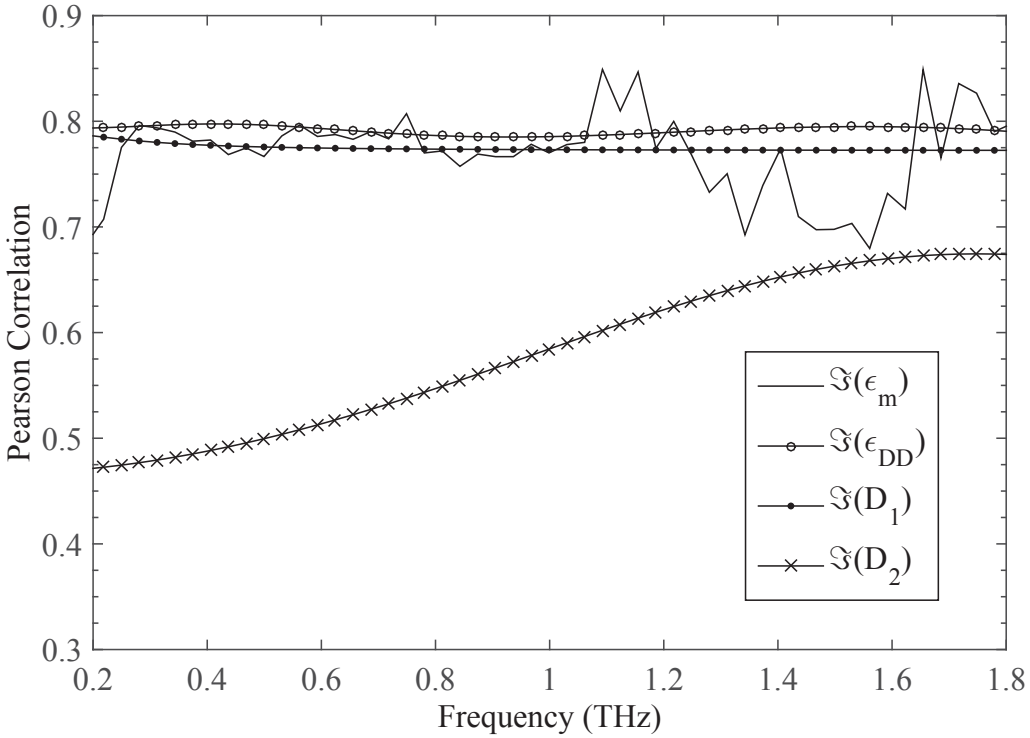

Figure 4.4: The Pearson correlations of $\Im(D_1)$, $\Im(D_2)$, $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(\epsilon_m)$ with the tumour percentage content in skin samples.

4.3.4 ROC analysis

The parameters of the double Debye model demonstrate significant differences between normal and diseased skin tissue as well as correlate well with the tumour percentage content of skin samples. In this section, a further test of their potential for classification between diseased and healthy tissue is carried out. ROC analysis is a common tool to investigate the classification accuracy in medical community [Fawcett, 2004]. Regarding its mechanism, a threshold value chosen by averaging two consecutive instances of original data is used to divide the data into two separate sets. Then, this classification is compared with tissue pathology to determine the sensitivity and specificity. In the ROC curve, 1-specificity is plotted versus sensitivity to describe the trade-offs in classification accuracy of a classifier/variable. For instance, Fig. 4.5 shows the ROC curves of ε_s, ε_2, ε_∞, τ_1 and τ_2. According to this figure, only the ROC curve of ε_s is completely above the

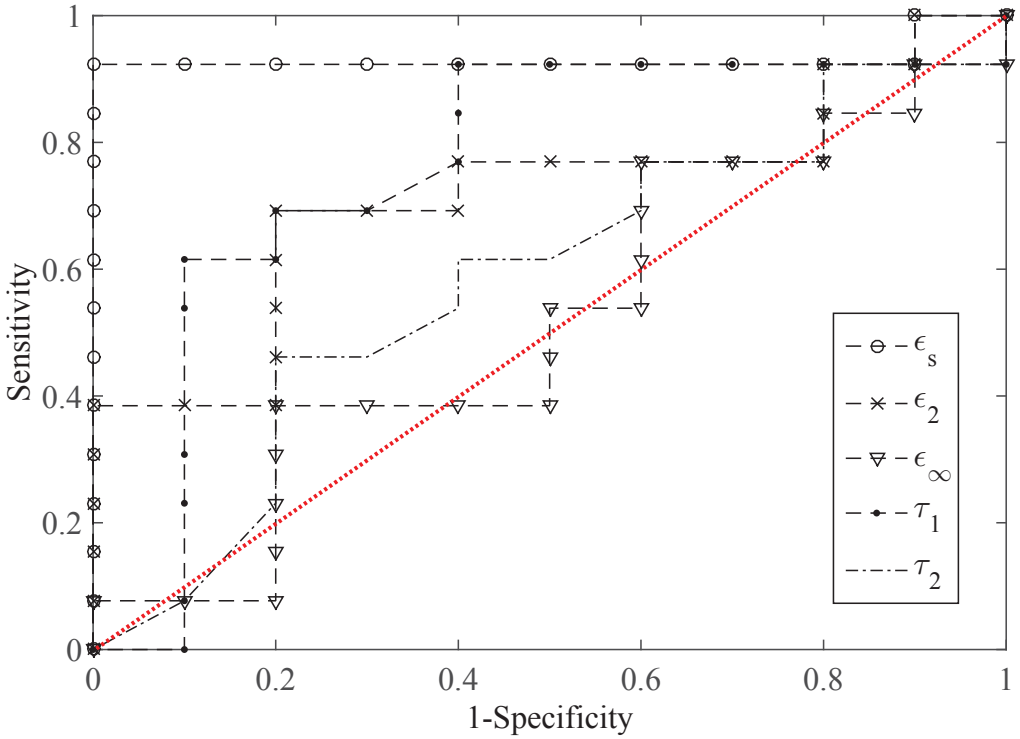

Figure 4.5: ROC curves of ε_s, ε_2, ε_∞, τ_1 and τ_2. The dotted line in red highlights the no-discrimination line.

diagonal line, thus confirming the discrimination capability of this classifier. In contrast, the interception of the ROC curves for ε_2, ε_∞, τ_1, τ_2 and the no-discrimination line at some points indicates their discrimination deficiency. This result fully matches with the statistical analysis of difference presented in Section 4.3.

To compare the classifiers, we employ the area under the ROC curve (AUC). AUC = 1 represents perfect discrimination while the AUC of any realistic classifier should not be lower than 0.5 (random chance). Table 4.5 provides the calculated AUCs of the five double Debye parameters. The averaged AUCs of $\Im(D_1)$, $\Im(D_2)$, $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(\epsilon_m)$ over the examined frequency range are also recorded in Table 4.5. While $\boldsymbol{\varepsilon}_r$ and $\Im(D_1)$ shows the highest discrimination potential among the investigated parameters, the rest are still prominent for classification due to the slight differences in AUC between these parameters. Furthermore, the AUC values of $\Im(D_1)$, $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(\epsilon_m)$ were plotted

DD	ε_s	ε_2	ε_∞	τ_1	τ_2	$\Im(D_1)$	$\Im(D_2)$	$\Im(\epsilon_{\mathrm{DD}})$	$\Im(\epsilon_m)$
AUC	0.93	0.72	0.51	0.76	0.58	0.93	0.79	0.91	0.89

Table 4.5: The AUCs for the five DD parameters and the AUCs averaged over the entire frequency range from 0.2 to 1.8 THz for $\Im(D_1)$, $\Im(D_2)$, $\Im(\epsilon_{\mathrm{DD}})$, $\Im(\epsilon_m)$.

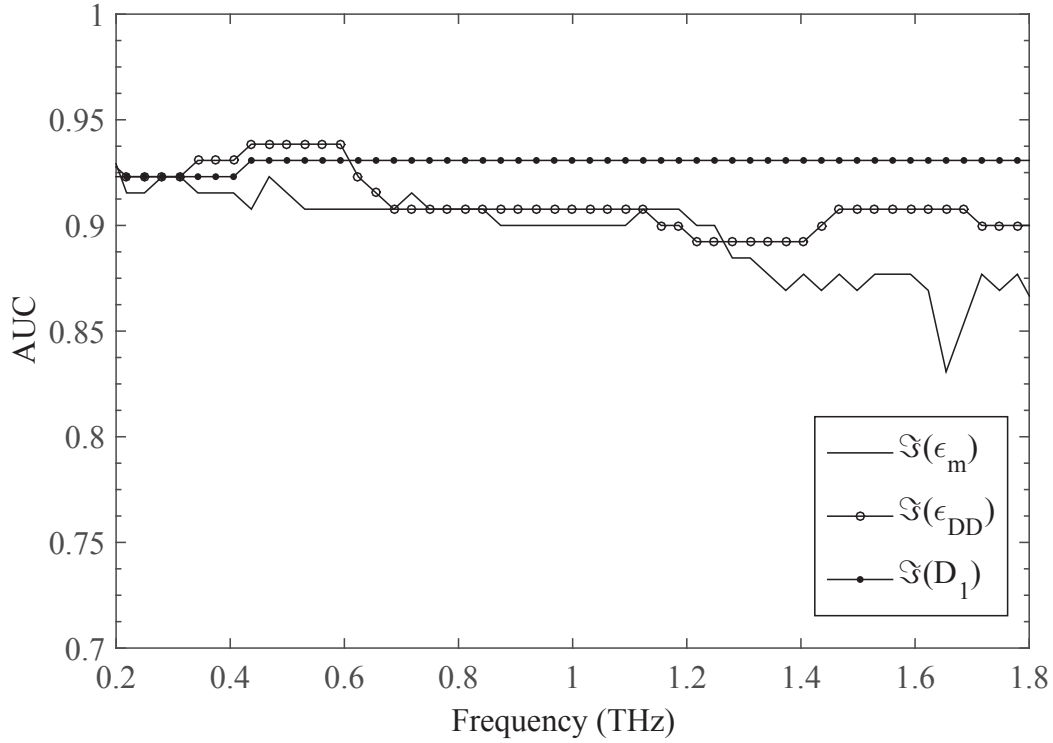

Figure 4.6: The areas under the ROC curves of $\Im(\epsilon_m)$, $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(D_1)$ over the frequency range $(0.2 - 1.8)$ THz.

over the range 0.2–1.8 THz in Fig. 4.6. According to this figure, the AUCs of $\Im(D_1)$, $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(\epsilon_m)$ simultaneously achieve their corresponding highest values at 0.5 THz.

4.4 Classification Methodologies

As the aforementioned statistical and ROC analysis facilitate the identification of potential classification variables, it is essential to examine them with a practical classification procedure that includes creating a rule for a training set of data and using the rule to

classify a testing set of data. Two different methods, namely pattern recognition via linear programming (LP) and support vector machine (SVM) classification algorithm, were employed to perform the classification with the double Debye parameters in our study. Then, the classification accuracies were estimated by the cross-validation (CV) technique.

4.4.1 Linear Programming

The linear programming-based diagnosis system is based on determining a linear discriminant function to separate the two distinct groups of the data [Bennett and Mangasarian, 1992; Mangasarian et al., 1995]. This method targets at finding a plane defined by $x^{\mathrm{T}}w = \gamma$ to separate the data points $x \in \mathbb{R}^n$ with n indicating the number of applied features. $w \in \mathbb{R}^n$ and γ are the constants of this linear function. Let $\mathbb{H} \subset \mathbb{R}^n$ be a set of H data points for diseased tissue and $\mathbb{M} \subset \mathbb{R}^n$ be a set of M data points for normal tissue. This plane will strictly separate these two groups if and only if

$$x^{\mathrm{T}}w \geq \gamma + 1 \ \forall x \in \mathbb{H} \quad \text{and} \quad x^{\mathrm{T}}w \leq \gamma - 1 \ \forall x \in \mathbb{M}. \tag{4.6}$$

However, these inequalities are not always achievable due to the linear inseparability of the tested data. Therefore, instead of completely-separating planes, we seek a plane which has minimal violations between the two data groups. Let $y = (y_1, y_2, ..., y_H) \in \mathbb{R}^+$ and $z = (z_1, z_2, ..., z_M) \in \mathbb{R}^+$ respectively represent these violations of $\mathbb{H}$ and $\mathbb{M}$. As a result, the optimal approximately-separating plane can be obtained through minimizing the average total violations as being described in the following optimization problem,

$$\begin{aligned} \min_{y,z,w,\gamma} \quad & \tfrac{1}{H}\textstyle\sum_{h=1}^{H} y_h + \tfrac{1}{M}\sum_{m=1}^{M} z_m \ : \\ & x_h^{\mathrm{T}}w - y_h \geq \gamma + 1 \ \forall x_h \in \mathbb{H}, \\ & x_m^{\mathrm{T}}w - z_m \leq \gamma - 1 \ \forall x_m \in \mathbb{M}, \\ & y_h \geq 0, \ z_m \geq 0, \ w \geq 0. \end{aligned} \tag{4.7}$$

Obviously, $y = 0 \in \mathbb{R}^H$ and $z = 0 \in \mathbb{R}^M$ in (4.7) means the plane $x^{\mathrm{T}}w = \gamma$ separates these two sets $\mathbb{H}$ and $\mathbb{M}$. The constraint $w \geq 0$ is added to the original problem

in [Mangasarian et al., 1995] due to the fact that the values of all the double Debye parameters for BCC samples are expected to be higher than those for normal samples. The problem (4.7) can be easily solved by the solver SeduMi [Sturm, 1999]. Then, the obtained optimal linear plane will be tested with the testing data for its capability of separating the cancer and healthy groups. A data point which satisfies the inequality $x^{\mathrm{T}}w > \gamma$ are labelled as a diseased case or vice versa.

4.4.2 Support Vector Machine

The support vector machine (SVM) is among the best supervised learning algorithm. The SVMs identifies the decision boundary separating two classes of data in a multi-dimensional feature space. The separating hyperplane is optimized with maximal geometric margin of support vectors between the classes. More details of SVM are thoroughly presented in [Burges, 1998]. For this study, the popular SVM package introduced by Chang and Lin [2011] with the Matlab implementation is utilised to perform the classification. Among kernel functions used to fit the hyperplane with non-linear decision surfaces to the data, the Gaussian radial basic function (RBF) is chosen to perform our classification. In fact, this function provides a less complicated approach compared to the other kernels since it only requires searching one kernel parameter γ. In addition, as the data in this study contains non-separable cases, there is a need to optimize the cost parameter C, also called soft margin parameter, which controls the trade-off between complexity and the number of nonseparable cases. Grid-search was applied to select the optimal values of the two parameters (C, γ) in order to achieve the highest classification accuracy. Eventually, the trained SVM model using the RBF with respect to the chosen (C, γ) was assessed with the testing data to predict the accuracy of the classifier.

4.4.3 Accuracy Estimation

Cross-validation (CV) is currently one of the most common methods to predict the true accuracy of a classier as well as select a classification model in supervised machine learning algorithms [Kohavi, 1995]. Indeed, CV reveals an estimation of the true error rate of the classifier thanks to the mechanism that keeps holding out a part of data set for testing and training the model with the rest of the data a limited number of times [Fisher and van Belle, 1993]. While different types of CV methods offer different trade-offs between bias and variance of the accuracy, the purpose of using CV is to decide upon the most appropriate method. Particularly, leave-one-out cross-validation (LOOCV) and repeated random subsampling cross-validation (RRSCV) were chosen to assess the classification accuracy of the proposed classifiers. LOOCV is the most computational form of the well-known k-fold CV, which requires the k training-testing times equivalent to the size of data. This method uses most of data to train the model and testing with only one left-out data point, thus providing an un-biased estimation of accuracy [Cawley and Talbot, 2003]. On the other hand, RRSCV allows leaving more data for testing and, as a result, reduces the variance of the predicted accuracy. Nonetheless, this method makes inefficient use of the data that is usually less than that in our expectation. The common ratio that 2/3 of the data is randomly held out for training and the remaining 1/3 of the data is for testing is chosen to estimate the RRSCV accuracy. The output of LOOCV and RRSCV includes the average values of cross-validation accuracy, sensitivity, and specificity. The number of training-testing times is equivalent to data size in the case of leave-one-out while it needs to be sufficiently large to guarantee a stable estimation in the case of random-subsampling.

4.5 Results and Discussion

The five double Debye parameters (ε_s, ε_2, ε_∞, τ_1, τ_2) are investigated in terms of their classification performance with the applied LP and SVM. In addition, the intensive statistical analysis and the ROC analysis in section 4.3 support selection of the most novel frequency for each of the frequency-dependent parameters including $\Im(D_1)$, $\Im(D_2)$, $\Im(\epsilon_{\mathrm{DD}})$, and $\Im(\epsilon_m)$ to classify the skin cancer. To be more specific, $\Im(D_1)$ and $\Im(D_2)$ are assessed at 0.2 THz (respectively denoted by $\Im(D_1)_{0.2}$ and $\Im(D_2)_{0.2}$) while the frequency 0.5 THz is considered for $\Im(\epsilon_{\mathrm{DD}})$ and $\Im(\epsilon_m)$ (respectively denoted by $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ and $\Im(\epsilon_m)_{0.5}$). These parameters together with the DD parameters are now considered as classification parameters.

For the RRSCV, the accuracy measures are averaged with respective standard deviations over 1000 of repeated training-testing times with 2/3 of the dataset (15 samples) held out for training and the remaining data (8 samples) used for testing. The training set and testing set are stratified to contain the same proportion of normal and diseased samples as the stratification in cross-validation can generally improve the robustness in terms of variance and bias as compared to regular cross-validation [Kohavi, 1995]. Each classification parameter is considered as an indicator of BCC in the LP and SVM classification process. Based on the cross-validation, we can compare the classification capability between the classification parameters as well as between LP and SVM classifiers.

According to Table 4.6, ε_s is the best classification variable with the LP due to its highest LOOCV and RRSCV accuracies among the proposed single parameters. Combination of ε_s and ε_2 in the LP classification offers an improvement in LOOCV up to 95.7%. However, incorporating more DD parameters, especially ε_∞, τ_1 and τ_2, into this combination causes overtraining, hence leading to a deterioration in classification accuracy. For example, combination of the three features $(\varepsilon_s, \varepsilon_2, \varepsilon_\infty)$ in the LP classification only provides the LOOCV of 91.3% and RRSCV of 83.5% which are considerably lower

Feature	LOOCV	RRSCV
ε_s	87.0±34.4	89.2±7.7
ε_2	69.6±47.1	69.4±11.6
ε_∞	47.8±51.1	48.3±8.6
τ_1	69.6±47.05	65.6±12.7
τ_2	56.5±50.5	43.3±9.9
$\varepsilon_s, \varepsilon_2$	95.7±20.9	89.4±9.1
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty$	91.3±28.8	83.5±12.8
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1$	91.3±28.8	82.5±13.0
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1, \tau_2$	91.3±28.8	82.8±12.4
$\Im(D_1)_{0.2}$	78.3±42.2	80.6±12.0
$\Im(D_2)_{0.2}$	78.3±42.2	76.3±12.0
$\Im(\epsilon_{\mathrm{DD}})_{0.5}$	82.6±38.8	84.2±10.4
$\Im(\epsilon_m)_{0.5}$	87.0±34.4	85.6±10.6

Table 4.6: The estimated accuracies (%) with their standard deviations by LOOCV and RRSCV for applying the double Debye parameters with the LP algorithms to classify the normal skin and BCC samples.

than 95.6% and 89.4% of (ε_s, ε_2). $\Im(D_1)_{0.2}$ and $\Im(D_2)_{0.2}$, $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ and $\Im(\epsilon_m)_{0.5}$ also offer good classification accuracies. The fairly close value of LOOCV to RRSCV of $\Im(\epsilon_m)_{0.5}$ confirms the true accuracy of the skin cancer classification using this feature.

For the SVM classification, Table 4.7 demonstrates the LOOCV and RRSCV values with their standard deviations for these parameters corresponding to applying the SVM classification algorithm. The cross-validation values obtained from the LOOCV are higher than those from the RRSCV in most cases. Furthermore, the higher standard deviation of the LOOCVs than that of RRSCVs is due to the fact that more data is left out in the leave-one-out method than its counterpart. As LOOCV makes the most efficient use of the data for training, these outcomes indicate that more training data can improve the prediction accuracy. Therefore, the true classification accuracy in this case may be closer to the LOOCV if the size of the training data is sufficiently large. For each classification variable, the best SVM parameters C, γ selected by the grid-search corresponding to each cross-validation method were presented in Table 4.7. The highest LOOCV at 95.7% provided by $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ indicates that this parameter is

Parameters	LOOCV	(C, γ)	RRSCV	(C, γ)
ε_s	91.3±28.8	(0.25, 2)	90.9±8.8	(0.5, 2)
ε_2	69.6±47.1	(1, 1)	61.1±12.1	(64, 0.125)
ε_∞	60.9±49.9	(32, 16)	50	(0.25, 1)
τ_1	73.9±44.9	(8, 0.25)	65.6±16.3	(64, 0.25)
τ_2	39.1±49.9	(0.25, 0.125)	50	(0.25, 1)
$\varepsilon_s, \varepsilon_2$	95.7±20.9	(2, 1)	86.9±12.4	(8, 0.0625)
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty$	95.7±20.9	(32, 0.5)	88.0±11.6	(16, 0.5)
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1$	95.7±20.9	(64, 0.125)	86.6±13.3	(64, 0.125)
$\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1, \tau_2$	87±34.4	(32, 0.125)	80.9±13.9	(64, 0.125)
$\Im(D_1)$ at 0.2 THz	91.3±28.8	(2, 32)	85.2±11.9	(2, 32)
$\Im(D_2)$ at 0.2 THz	82.6±38.8	(64, 0.5)	80.4±12.2	(64, 0.5)
$\Im(\epsilon_{\mathrm{DD}})_{0.5}$ at 0.5 THz	95.7±20.9	(64, 16)	88.0±12.4	(64, 16)
$\Im(\epsilon_m)_{0.5}$ at 0.5 THz	87±34.4	(16, 4)	86.3±10.4	(16, 4)

Table 4.7: The estimated accuracies (%) with their standard deviations by LOOCV and RRSCV for applying the double Debye parameters with the SVM algorithm to classify the normal skin and BCC samples.

potentially the best for the applied SVM in case of having adequate data in the training set. Less training data used in the random-subsampling approach consequently reduces the prediction accuracy as can be seen through the lower RRSCV (88.0%) of $\Im(\epsilon_{\mathrm{DD}})_{0.5}$. A similar outcome can be observed with $\Im(D_1)_{0.2}$ whose CV reduces from 91.3% in the LOOCV to 85.5% in the RRSCV. Combinations of ε_s and the other parameters including ε_2, ε_∞, τ_1 also provide the diagnosis accuracy estimated by LOOCV up to 95.7% while adding τ_2 to the classification weakens the prediction with the LOOCV of only 87%. In contrast, their performance assessed by RRSCV shows a remarkable reduction in classification accuracy. This may be explained by insufficient data used for training in the random-subsampling method. The stand-alone ε_s obtains the LOOCV accuracy of 91.3% while its RRSCV reaches the peak of 90.9% among all the applied parameters in the recent study. The slight differences between the LOOCVs and RRSCVs for ε_s and $\Im(\epsilon_m)$ not only indicate the sufficient data used for training in the RRSCV but also support the estimation of their true classification accuracies.

The SVM algorithm with its advances in terms of dealing with nonlinear inseparability

of the data demonstrates its superiority over the LP approach in a majority of cases. However, the level of the improvement varies between different applied features. In particular, the applied SVM remarkably improve the classification accuracies of $\Im(D_1)_{0.2}$, $\Im(D_2)_{0.2}$, and $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ as compared to the LP algorithm. For instance, the LOOCV of $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ increases from 82.6% by the LP to 95.7% by the SVM, which highlights the suitable classifier for this parameter. Furthermore, $\Im(D_1)_{0.2}$ is also another potential variable for the SVM classification as its LOOCV accuracy achieves 91.30% that is far higher than 78.26% for the LP. On the other hand, ε_s consistently presents an impressive performance with both the LP and SVM classifier, giving the cross-validation accuracies ranging approximately from 87.0% and 91.3%. Such an outcome marks this parameter as the most potential indicator for detecting BCC. Additionally, in terms of dealing with multi-feature classification, the performance of SVM is overall better than that of LP. Particularly, $(\varepsilon_s, \varepsilon_2, \varepsilon_\infty)$ and $(\varepsilon_s, \varepsilon_2, \varepsilon_\infty, \tau_1)$ in the SVM classification obtain higher LOOCV and RRSCV accuracies as compared with the LP method.

4.6 Conclusion

The complex permittivity of 23 samples of human skin including either completely healthy or containing a certain proportion of BCC are fitted by the double Debye model using the global optimization-based algorithm with the proposed strict constraint of τ_1. The extracted parameters of this model are statistically analysed to prove their capability of skin cancer discrimination. The differences between normal skin tissue and BCC are assessed by the Student's t-test and normalized percentage difference of the mean values. Among five parameters of the double Debye model, only ε_s can offer the statistical significance of difference which is supported by 33% of the mean difference between normal and BCC samples. The variation of this parameter has a high correlation of $r = 0.7$ with the tumour percentage content. This result indeed highlights the advantage of applying the appropriate constraint of τ_1 over the prior

study [Truong et al., 2014]. We also observe the significant difference in the imaginary part of the complex permittivity, both measured ($\Im(\epsilon_m)$) and predicted by the double Debye theory ($\Im(\epsilon_{\mathrm{DD}})_{0.5}$), and the two Debye terms ($\Im(D_1)_{0.2}$ and $\Im(D_2)_{0.2}$) over the examined frequency range from 0.2 to 1.8 THz. The normalized percentage differences in these parameters reach the top values for $\Im(D_1)_{0.2}$ and $\Im(D_2)_{0.2}$, and at 0.5 THz for $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ and $\Im(\epsilon_m)$. Furthermore, the high correlations of these parameters with the BCC content of the samples are also spotted over the frequencies. These analysis results of $\Im(D_1)_{0.2}$, $\Im(D_2)_{0.2}$ and $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ provide a solid explanation for the findings of the empirical combinations of the double Debye parameters in [Truong et al., 2014].

The applied SVM and LP classification procedure indeed offer encouraging results in terms of using the various parameters from the double Debye model to classify the BCC. Among the examined double Debye parameters, ε_s is the most novel classification feature based on its excellent discrimination capability with the AUC of 0.93 as well as its robust classification accuracy in both the LP and SVM approach with the RRSCVs of 89.2% and 90.9% respectively. While the classification ability of ε_2, ε_∞, τ_1, and τ_2 is rather weak with their estimated accuracies only up to 69.6%, some of their combinations with ε_s can provide some improvements. For instance, the pair of $(\varepsilon_s, \varepsilon_2)$ enhances the LOOCV accuracy from 91.3% to 95.7% in the SVM. However, the performance of each combination also depends on the applied inducer (i.e. LP or SVM), hence choosing the correct method will be key to explore the potential of the combinations. The three frequency-dependent parameters of the double Debye model, namely $\Im(\epsilon_{\mathrm{DD}})_{0.5}$, $\Im(D_1)_{0.2}$, and $\Im(D_2)_{0.2}$ also demonstrate the high ability of BCC discrimination through their impressive frequency-averaged AUC values which vary from 0.79 to 0.93. Furthermore, the SVM is significantly more efficient for $\Im(D_1)_{0.2}$, $\Im(D_2)_{0.2}$ and $\Im(\epsilon_r^D)_{0.5}$ as compared to the LP. $\Im(\epsilon_{\mathrm{DD}})_{0.5}$ with the SVM classifier provides the LOOCV accuracy of 95.7% and RRSCV accuracy of 88%, which are the best among the frequency-dependent parameters as well as comparable to ε_s and its combinations. In general, the SVM approach with its advance in dealing with linearly inseparable

data can indeed provide better classification accuracies for the examined double Debye parameters than the LP. This suggests that the SVM algorithm is more suitable for the DD parameters to classify the skin cancer. The classification procedure performed in this study was carried out in a small scale of the data, thus we can only confirm the potential of the proposed classifiers from the double Debye model and estimate their true classification accuracies. Therefore, we suggest that future studies produce THz images of BCC with these classifiers in order to confirm the accuracy of THz imaging in detecting tumour margins for MMS.

Chapter 5

A NEW DIELECTRIC MODEL OF HUMAN BREAST TISSUE IN TERAHERTZ REGIME

5.1 Introduction

The double Debye model has been used to understand the dielectric response of different types of biological tissues at terahertz frequencies but fails in accurately simulating human breast tissue. This leads to limited knowledge about the structure, dynamics, and macroscopic behaviour of breast tissue and hence constrains the potential of THz imaging in breast cancer detection. Therefore, the first goal of this study is to propose a new dielectric model capable of mimicking the spectra of human breast tissue's complex permittivity in THz regime. Namely, a non-Debye relaxation model is combined with a single Debye model to produce a mixture model of human breast tissue. A sampling gradient algorithm of non-smooth optimization is applied to locate the optimal fitting solution. Samples of healthy breast tissue and breast tumour are used in the simulation to evaluate the effectiveness of the proposed model. The second goal is to confirm the potential of using the parameters of the proposed dielectric model to distinguish breast tumour from healthy breast tissue, especially fibrous tissue. Statistical measures are employed to analyse the discrimination capability of the model parameters while support vector machines and Bayesian neural networks are applied to assess the possibility of using the combinations of these parameters for higher classification accuracy.

This chapter has the following structure. Section 5.2 is to introduce the mixture model of breast tissue, propose the fitting algorithm to extract the parameters of the proposed models, describe the tested data and discuss on simulation results. Section 5.3 provides the statistical analysis of the extracted model parameters. Then, these parameters are used to classify the breast samples with SVM and BNN in section 5.4. Eventually, the chapter is summarised in section 5.5.

5.2 Dielectric Properties Modelling

5.2.1 Dielectric Model of Human Breast Tissue

Human tissue is commonly considered as a dispersive material whose dielectric properties including relative permittivity and conductivity depend on frequency. The increase of frequency leads to a remarkable decrease in the relative permittivity but conversely causes an upturn in conductivity. Complex permittivity of human tissue in a very low frequency range is well-described by the single Debye relaxation model while its responses in higher frequencies above 0.1 THz require extra Debye relaxation processes [Liebe et al., 1991] :

$$\epsilon = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_n}{1 + \jmath\omega\tau_n}, \tag{5.1}$$

where $\Delta\epsilon_n$ indicates the permittivity dispersion of the n^{th} Debye relaxation process. The underlined fact of the multiple Debye model is the higher water content of various human tissues [Foster and Schwan, 1989; Gabriel, Lau and Gabriel, 1996]. The double-Debye model was initially used as a dielectric function of water and then has inspired a significant number of applications for highly-hydrated mixtures [Kindt and Schmuttenmaer, 1996; Reid et al., 2010; Pickwell, Cole, Fitzgerald, Wallace and Pepper, 2004; Smye et al., 2001]. For instance, the complex permittivity of human skin, which contains around 70% of water, has been accurately predicted by the two Debye relaxation processes [Pickwell et al., 2005; Truong et al., 2013]. Muscle with higher water content than skin employs the summation of five Debye dispersions in addition to a frequency-dependent conductivity term [Hurt, 1985]:

$$\epsilon = \epsilon_\infty + \sum_{n=1}^{5} \frac{\Delta\epsilon_n}{1 + \jmath\omega\tau_n} + \frac{\sigma}{\jmath\omega\epsilon_0}. \tag{5.2}$$

However, dielectric properties of biological tissues with low water content and more complicated structure and composition exhibit broader dispersion that may involve superposition of several relaxation processes or non-first-order kinematics of molecular

structure. The broadening of the dispersion could be resolved by adding a distribution parameter, and thus the Cole-Cole (CC) equation has been introduced as an alternative of the Debye equation [Gabriel, Lau and Gabriel, 1996]. Eventually, the experimental dielectric spectra of various human tissues may be effectively described by multiple CC relaxation model [Gabriel, Gabriel and Corthout, 1996; Gabriel, Lau and Gabriel, 1996; Said and Varadan, 2007]:

$$\epsilon = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_n}{1 + (\jmath\omega\tau_n)^{1-\alpha_n}} + \frac{\sigma}{\jmath\omega\epsilon_0}, \tag{5.3}$$

where α_n is the distribution parameter measuring the broadening of the dispersion n^{th}. The non-exponential relaxation processes described by the CC equation constitute a special case of non-Debye dielectric relaxation. The Havriliak-Negami (HN) relationship generalizes modelling the non-Debye relaxation processes by introducing two empirical exponents α and β [Havriliak and Negami, 1967]:

$$\tilde{\epsilon}(\omega) = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{[1 + (\jmath\omega\tau)^\alpha]^\beta}, \tag{5.4}$$

In this regard, human breast is also composed of inhomogeneous structures of fat cells and proteins. The fatty (adipose) tissues have a low water content and therefore play a substantial role in regulating the dielectric responses of breast tissue. Indeed, the spectra of real permittivity of breast tissue increase at the low frequencies and pose quite a flat response over the higher frequency range, which is similar to that of pure adipose tissue. Particularly, Fig. 5.1 presents the average measured complex permittivities versus the frequencies for breast tumour, healthy fibrous breast tissue (fibrous), and healthy adipose tissue (fat), which were published by [Ashworth et al., 2009]. The error bars show the standard errors of the mean values with 95% confidence interval. Further details of the data are thoroughly described in subsection 5.2.3. Accordingly, while the most noticeable feature in dielectric properties of the breast tissues is the peak at around 0.32 THz, a similar increase is also spotted at 0.5 THz in dielectric spectrum of fat tissue. The physical origin of this dielectric response is yet to be understood

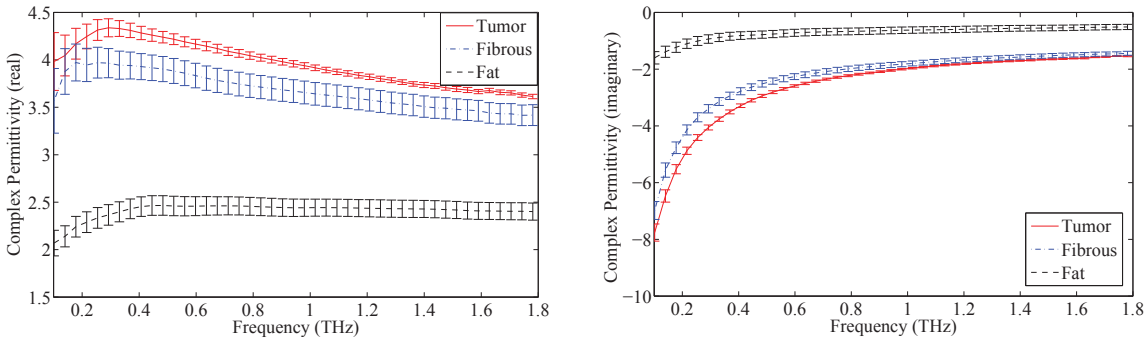

Figure 5.1: The average complex permittivities with the corresponding error bars representing 95% confidence intervals for breast tumour, healthy fibrous breast tissue, and healthy fat (adipose) breast tissue from [Ashworth et al., 2009]

.

[Ashworth et al., 2009; Khan et al., 2007]. Therefore, the increase of permittivity of breast tissue in the low range of THz regime is considered as a non-Debye response. Apart from this, breast tissue still contains a relative-high proportion of water, which is the most prevalent source of absorption in THz frequencies, and, thus, its complex permittivity like that of human skin may be also contributed by dielectric properties of water. In fact, the increase of THz frequency results in greater contribution of water absorption, thus the Debye relaxation process may be more transparent in the range above 1 THz. Considering the two aforementioned factors driving the dielectric spectra of breast tissue, an empirical mixture model based on a combination of non-Debye and Debye relaxation processes is proposed as follows,

$$\tilde{\epsilon}(\omega) = \epsilon_\infty + \frac{\omega\tau_1\Delta\epsilon_1 + \Delta\epsilon_2}{1 + (\jmath\omega\tau_1)^\alpha} + \frac{\Delta\epsilon_3}{1 + \jmath\omega\tau_2} + \frac{\sigma}{\jmath\omega}. \tag{5.5}$$

Here the term $(\omega\tau_1\Delta\epsilon_1 + \Delta\epsilon_2)$ produces the peak in the real part of the complex permittivity of breast tissue at frequencies below 1 THz. $\Delta\epsilon_1$ and $\Delta\epsilon_2$ underline the existence of two dielectric dispersions occurring in slow relaxation process characterized by time constant τ_1. However, it is important to emphasize that proposing these two parameters is more empirical rather than underlying any physical process, and thus they should be simply regarded as necessary parameters to provide accurate fits. The dispersive amplitude of fast relaxation mode with the time constant τ_2 is ϵ_3 corresponding to high

frequency. ϵ_∞ is the limiting permittivity constant at very high frequency. $\sigma/\jmath\omega$ represents the impact of dc-conductivity on dielectric loss of tissue. For easy tracking of the non-Debye model for breast tissue, it is denoted by THz-nDB in the presentation of this thesis.

As the real and imaginary parts of the complex permittivity have different impacts on the values of the introduced parameters in (5.5), the significance of the difference in these two features between healthy breast tissues and tumour can suggest the cancer discrimination potential of the THz-nDB parameters. Therefore, the normalized dif-

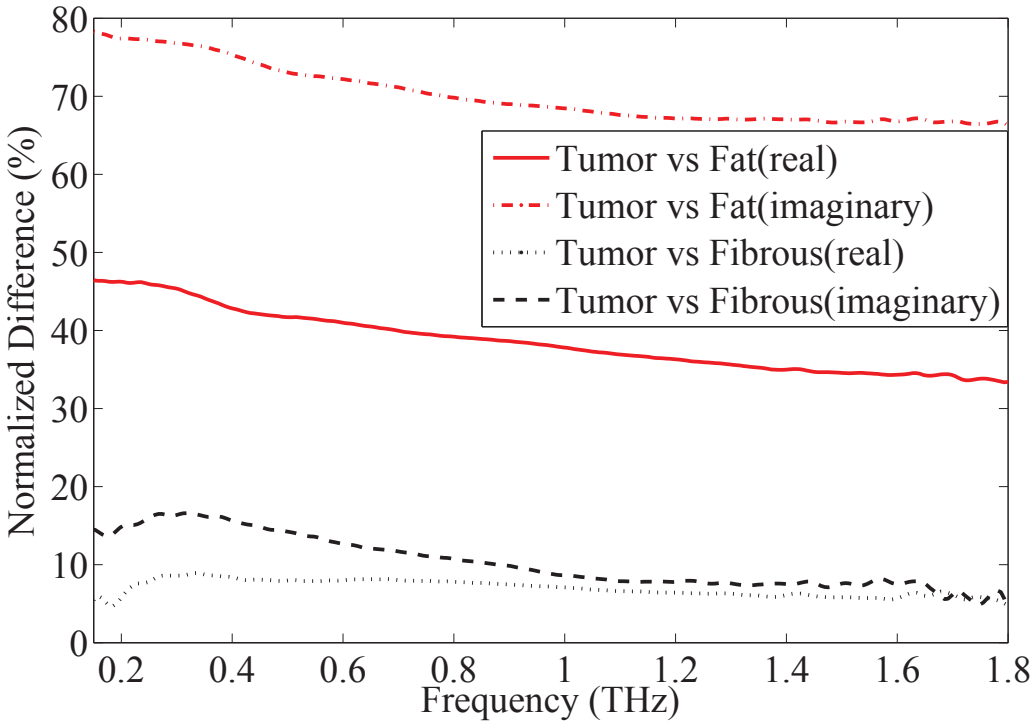

Figure 5.2: The normalized percentage difference in the average complex permittivities between the two healthy breast tissue groups (fibrous and fat) and the breast tumour.

ferences in the real and imaginary part of the average complex permittivities between tumour and healthy tissues, both fibrous and fat, were calculated and plotted over the examined frequency range in Fig. 5.2. Accordingly, the variation of the imaginary part

of the complex permittivity between the healthy and cancerous tissues is more significant than that of the real part in most of the spectral range. The fact that only σ is fully driven by the dielectric loss marks this parameter as the most prevalent indicator of breast cancer among the parameters of the model (5.5). The theoretical assessment will be statistically investigated in depth in the subsection 5.3.

5.2.2 Fitting Algorithm

In order to fit the data and extract the proper parameters of (5.5),the sum of squared error functions (SSE) which calculate the differences between the measured frequency-dependent complex permittivities and their predicted values by one of these models over a range of consecutive THz frequencies is employed. Set $\boldsymbol{y} = (\epsilon_\infty, \Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3, \sigma)^{\mathrm{T}} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{y}_5)^{\mathrm{T}}$. The optimization problem in order to fit the measured data is the following nonlinear least-square

$$\begin{aligned} \min_{\boldsymbol{y},\ \tau_1,\ \tau_2,\ \alpha} \quad & \sum_{i=1}^{M} |\tilde{\epsilon}(\omega_i) - \tilde{\epsilon}_m(\omega_i)|^2 \\ \text{subject to} \quad & \boldsymbol{y}_1 \geq 1, \boldsymbol{y}_2 \geq 0, \boldsymbol{y}_3 \leq 0, \boldsymbol{y}_4 \geq 0, \boldsymbol{y}_5 \geq 0 \\ & \boldsymbol{y}_1 + \boldsymbol{y_3} + \boldsymbol{y}_4 \geq 1 \\ & 0.5 \leq \tau_1,\ 0 \leq \tau_2 \leq 0.5,\ \alpha \geq 0. \end{aligned} \tag{5.6}$$

where $\tilde{\epsilon}(\omega_i)$ defined by (5.5) is a highly nonlinear function of 8-dimensional variable $(\boldsymbol{y}, \tau_1, \tau_2, \alpha)$, while $\tilde{\epsilon}_m(\omega_i)$ stands for the measured complex permittivity at a frequency ω_i. M is the number of sampling frequencies. The constraints of the model parameters in (5.6) are chosen to maintain the necessary physics properties of the complex dielectric model.

The problem (5.6) is seen as a very challenging optimization problem, for which the standard optimization tools and software are hardly suitable. Our first step is to analyse partially convex structures of (5.6) that are useful for computational purpose. It is seen from (5.5) that the objective function in (5.6) is convex quadratic in $\boldsymbol{y}$ when the three

variables τ_1, τ_2, α are held fixed. Thus, by setting $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = (\tau_1, \tau_2, \alpha)$, the total squared error function SSE in (5.6) is reexpressed by

$$f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^{\mathrm{T}} \sum_{i=1}^{M} A_i(\boldsymbol{x})\boldsymbol{y} + \sum_{i=1}^{M} b_i^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{y} + \sum_{i=1}^{M} d_i, \tag{5.7}$$

where

$$A_i(\boldsymbol{x}) = a_i^*(\boldsymbol{x}) a_i^{\mathrm{T}}(\boldsymbol{x}),\ b_i^{\mathrm{T}} = -2Re(c_i a_i^{\mathrm{H}}), d_i = |\tilde{\epsilon}_m(\omega_i)|^2,$$
$$A(\boldsymbol{x}) = \sum_{i=1}^{M} A_i(\boldsymbol{x}), b^{\mathrm{T}}(\boldsymbol{x}) = \sum_{i=1}^{M} b_i^{\mathrm{T}}(\boldsymbol{x}), d = \sum_{i=1}^{M} d_i,$$
$$a_i(\boldsymbol{x}) = \left[1, \frac{\omega_i \tau_1}{1 + (\jmath\omega_i\tau_1)^\alpha}, \frac{-1}{1 + (\jmath\omega_i\tau_1)^\alpha}, \frac{1}{1 + \jmath\omega_i\tau_2}, \frac{1}{\jmath\omega_i}\right]^{\mathrm{T}}.$$

Eventually, instead of the optimization problem (5.6) in eight decision variables, the following optimization in three-dimensional variable $\boldsymbol{x}$ is considered:

$$\min_{\boldsymbol{x} \in X} \ F(\boldsymbol{x}) \tag{5.8}$$

with

$$F(\boldsymbol{x}) = \min_{\boldsymbol{y} \in Y} \quad f(\boldsymbol{x}, \boldsymbol{y}). \tag{5.9}$$

Here X and Y are linear constraints of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively defined from (5.6):

$$X = \{\boldsymbol{x} \in R^3 \ : \ 0.5 \leq \boldsymbol{x}_1,\ 0 \leq \boldsymbol{x}_2 \leq 0.5,\ \boldsymbol{x}_3 \geq 0\}$$

and

$$Y = \{\boldsymbol{y} \in R^5 \ : \ \boldsymbol{y}_1 \geq 1, \boldsymbol{y}_2 \geq 0, \boldsymbol{y}_3 \leq 0, \boldsymbol{y}_4 \geq 0,$$
$$\boldsymbol{y}_5 \geq 0, \boldsymbol{y}_1 + \boldsymbol{y_3} + \boldsymbol{y}_4 \geq 1\}.$$

It is simple to calculate each value $F(\boldsymbol{x})$ by using any existing convex quadratic solver such as SeduMi [Sturm, 1999]. However, $F(\boldsymbol{x})$ is a $3D$ non-convex and nonsmooth function in the sense that almost everywhere there is only single $y(x) \in Y$ such that

$$f(x, y(x)) = F(x) \Leftrightarrow y(x) = \mathsf{arg}\min_{y \in Y} f(x, y) \tag{5.10}$$

for which,

$$\nabla F(x) = \frac{\partial f(x,y)}{\partial x}|_{y=y(x)} \tag{5.11}$$

For solution of (5.8), the robust gradient sampling algorithm introduced by [Burke et al., 2005] is employed. The beforehand selected parameters are:

- Backtracking reduction factor $0 < \gamma < 1$. We choose $\gamma = 0.5$.
- Armijo parameter in line search $0 < \beta < 1$. However, it only needs to choose a very small value of the Armijo parameter, which is practically equivalent to its nonstandard choice of 0.
- Sampling size $m > \dim(x) = 3$, thus set $m = 7$.
- Optimality tolerance reduction factor $\theta = 1$.
- Sampling radius reduction factor $\mu = 0.5$.

Initial parameters that will be iteratively updated are

- Sampling radius ratio $\epsilon_0 = 0.1$.
- Optimality tolerance $\nu_0 \geq 0$ is fixed to 10^{-6} throughout.

Iterations initialized from $x^{(0)} \in X$. For $\kappa = 0, 1, ...$

- **Step 1 (gradient sampling)**. Take $u^{(\kappa_1)}, ..., u^{(\kappa_m)}$ sampled uniformly from the unit ball $\{u \in R^3 \ : \ ||u|| \leq 1\}$ such that

$$\begin{aligned} &x^{(\kappa_0)} = x^{(\kappa)}, \\ &x_i^{(\kappa_j)} = x^{(\kappa)}(1 + \epsilon_\kappa u^{(\kappa_j)}) \in X, \ j = 1, 2, ..., m. \end{aligned} \tag{5.12}$$

  For each $x^{(\kappa_j)}$ solve the quadratic program (5.10) to find $y(x^{(\kappa_j)})$ and define the gradient $\nabla F(x^{(\kappa_j)})$ according to (5.11).

- **Step 2 (direction search)**. Solve the quadratic program

$$\min_{\tau_j, j=0,1,\ldots,m} ||\tau_j \nabla F(x^{(\kappa_j)})||^2 \ : \ \tau_j \geq 0, \sum_{j=0}^{m} \tau_j = 1 \tag{5.13}$$

to have its optimal solution τ_j for the direction definition

$$g^{(\kappa)} = -\sum_{j=0}^{m} \tau_j \nabla F(x^{(\kappa_j)}) \tag{5.14}$$

If $||g^{(\kappa)}|| = 0$ stop ($x^{(\kappa)}$ is the optimal solution). If $||g^{(\kappa)}|| \leq \nu_0$ set $t_\kappa = 0$, set the sampling radius $\epsilon_{\kappa+1} = \mu\epsilon_\kappa$ and go to Step 4. Otherwise keep $\epsilon_{\kappa+1} = \epsilon_\kappa$ and $d^{(\kappa)} = -g^{(\kappa)}/||g^{(\kappa)}||$ (normalized direction for line search), and go to Step 3.

- **Step 3 (step length)**. Use enumeration to solve

$$\begin{aligned} t_\kappa = \ & \max_{s=0,1,2,\ldots} \gamma^s \ : \\ & F(x^{(\kappa)} + \gamma^s d^{(\kappa)}) < F(x^{(\kappa)}) - \beta\gamma^s ||g^{(\kappa)}|| \end{aligned} \tag{5.15}$$

(it is finite according to Armijo rule) and go to Step 4.

- **Step 4 (update)**. Set

$$x^{(\kappa+1)} = x^{(\kappa)} + t_\kappa d^{(\kappa)} \tag{5.16}$$

and $\kappa + 1 = \kappa$. Go to Step 1.

5.2.3 Data for Simulation

Refractive indices and absorption coefficients of 74 breast samples are used to test the validity of the proposed model (5.5). The data were presented with group average values and error bars in [Ashworth et al., 2009]. Measurement procedures using THz spectroscopy in transmission geometry were also fully described by Ashworth et al. [2009]. In brief, breast specimens were excised from 20 non-consecutive female patients and preserved in refrigerated and humid condition. A THz time-domain spectrometer (TPIspectra1000, TeraView Ltd, Cambridge, UK) was used to measure the transmitted

THz pulses through the tissue samples prepared from the patient specimens. The THz properties of the samples were calculated by averaging the measured refractive indices and absorption coefficients. The percentage content of fat (healthy adipose tissue), fibrous (healthy fibrous tissue), and cancer/tumour for each sample was determined by a Consultant Breast Pathologist and this served as a true reference to evaluate the THz data. Then, each of these samples was categorized according to its predominant constituent which occupied more than 50% of the whole tissue. Complex permittivities calculated from these spectra represent the standard dielectric response of fat, fibrous tissue and tumour of human breast to THz radiation. More details of the calculations to obtain the data were thoroughly provided in [Ashworth et al., 2009].

5.2.4 Fitting Results

In this section, the capability of the model (5.5) to simulate the complex permittivity of human breast tissue is investigated through fitting this model to the data of pure adipose tissue, fibrous tissue and tumour. We employ the adjusted R-square value (R^2_{adj}), which is the square of the correlation between the response values and the predicted response values, to analyse the fitting quality [MATLAB, 2012].

Sample	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	$\tau_1(ps)$	$\tau_2(ps)$	α	SSE	R^2_{adj}
Adipose	2.50	3.54	-1.89	0.47	0.39	9.67	0.13	1.49	0.06	0.996
Normal	3.34	1.33	-3.05	0.71	5.27	1.91	0.11	1.20	0.41	0.998
BCC	2.86	28.64	-3.45	1.59	8.33	3.58	0.10	2.91	1.09	0.997

Table 5.1: The extracted parameters of the THz-nDB model (5.5) by fitting it to averaged complex permittivities of three types of human breast tissue.

It can be seen in Table 5.1 that R^2_{adj} in all three cases achieves the overall value from 0.99 . Fig. 5.3-5.5 demonstrate the comparison between complex permittivity simulated by (5.5) and the measured data for adipose tissue, healthy breast tissue and breast tumour respectively. The proposed dielectric model (5.5) mimics very well the dielectric properties of breast tissue not only at the high range of frequency beyond 1 THz but

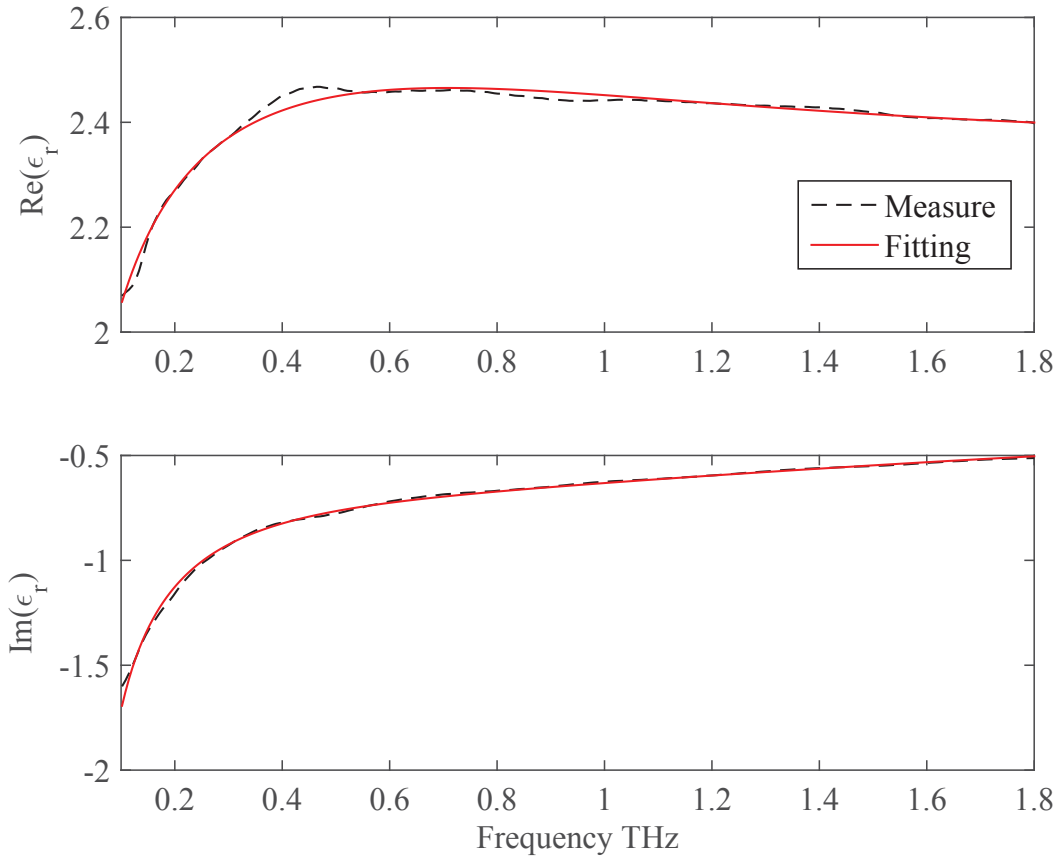

Figure 5.3: The measured average complex permittivities of healthy fat tissue from [Ashworth et al., 2009] and its fitting by the THz-nDB model.

also at the lower frequencies. It is worth emphasizing that the non-Debye response of the real part of the breast tissue complex permittivity can be characterized well by (5.5). This advance of the THz-nDB model (5.5) is practically essential for breast cancer detection as the maximum difference between healthy and cancerous tissue in optical properties was found at the low spectral range, especially 0.32 THz [Ashworth et al., 2009]. The fitting parameters of (5.5) and R^2_{adj}s for healthy tissue and tumour are recorded in Table 5.2. Accordingly, the contrast values in the THz-nDB parameters of (5.5) between healthy and cancerous tissue suggest the potential of treating these parameters as indicators to classify the breast Tumours.

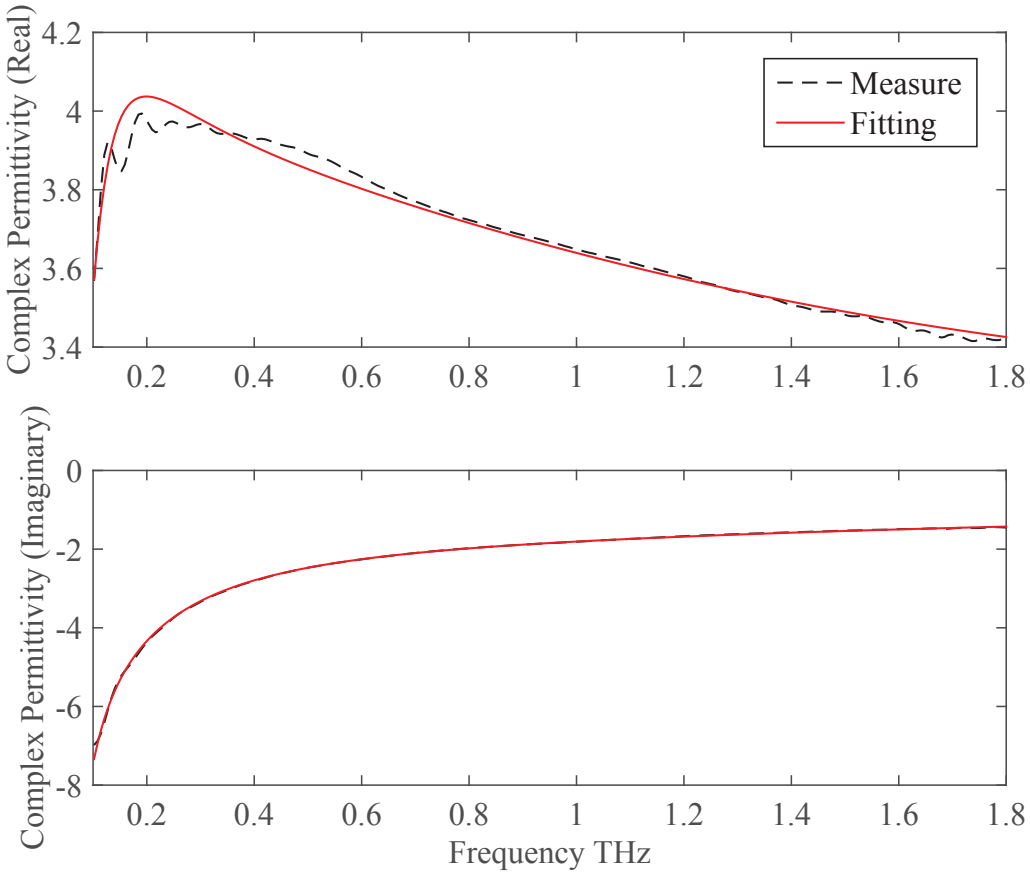

Figure 5.4: The measured complex permittivities of healthy fibrous breast tissue from [Ashworth et al., 2009] and its fitting by the THz-nDB model.

The fitting parameters of (5.5) and R^2_{adj}s for all samples are recorded in groups of fat, fibrous and tumour with their average values (AVR) and corresponding standard errors (SE) in Table 5.2. According to the statistical values of R^2_{adj}, the THz-nDB model (5.5) can simulate the measured complex permittivities of fibrous tissue and tumour better than those of fat tissue. This may be explained by the higher water content of the former, especially tumour. In fact, since THz radiation is very sensitive to water, the dielectric properties of tumour and fibre possess less non-smooth responses and noise both of which can negative impact on the predictability of the model (5.5). In general, the AVRs of ϵ_∞, $\Delta\epsilon_1$, $\Delta\epsilon_3$, τ_1, τ_2, α are not apparently different between breast tissue types whilst the large SEs of these parameters reveal the high possibility

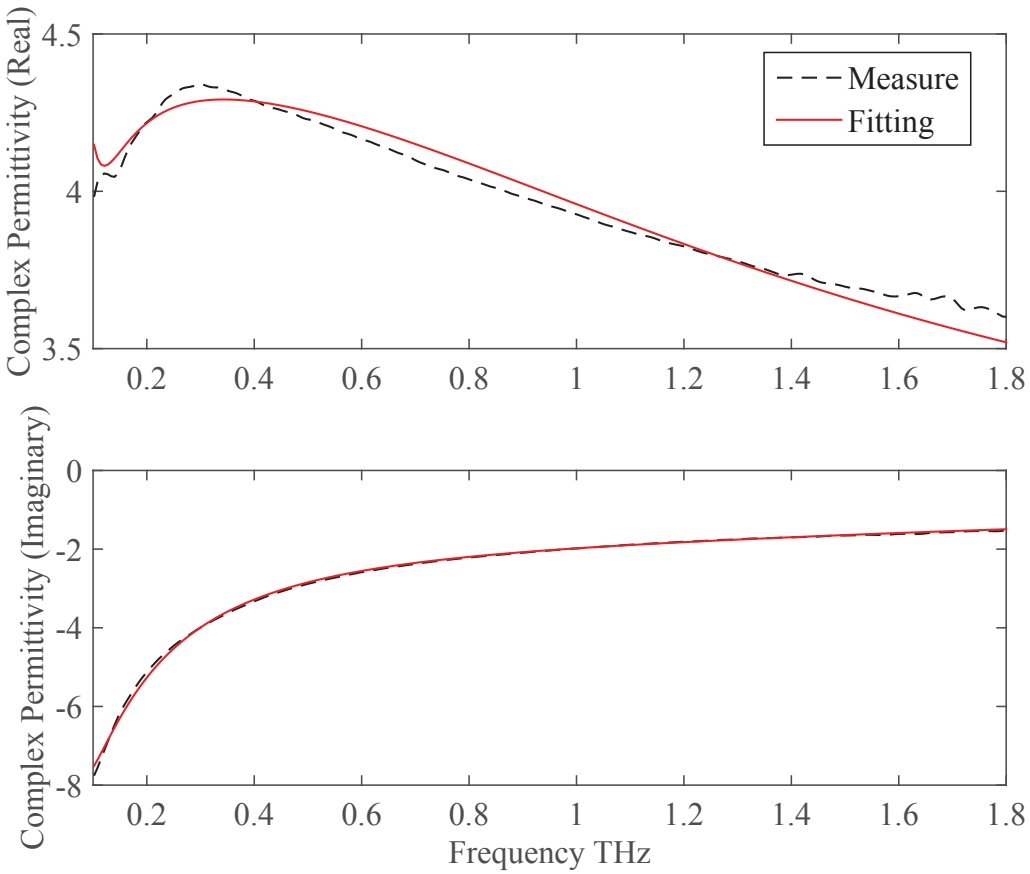

Figure 5.5: The measured complex permittivities of a breast tumour from [Ashworth et al., 2009] and its fitting by the THz-nDB model.

of overlaps of their values between tissue groups. On the contrary, $\Delta\epsilon_2$ and σ have obvious different AVRs between dominantly fat, fibrous and tumour tissues while their SE are rather small in comparison to these differences. This suggests $\Delta\epsilon_2$ and σ to be essential signatures of breast cancer applications, especially the cancer margin detection in THz imaging. An analysis in terms of the classification potential of the THz-nDB parameters is exposed in the next section.

Parameter	Fat	Fibrous	Tumour
ϵ_∞	2.41 ± 0.20	2.80 ± 0.31	3.15 ± 0.13
$\Delta\epsilon_1$	2.69 ± 2.15	38.22 ± 33.14	545.60 ± 980.60
$\Delta\epsilon_2$	-1.12 ± 0.56	-2.47 ± 0.67	-2.82 ± 0.43
$\Delta\epsilon_3$	0.61 ± 0.26	1.31 ± 0.25	1.34 ± 0.13
σ	1.12 ± 0.27	4.41 ± 0.88	7.89 ± 0.71
$\tau_1(ps)$	2.07 ± 0.61	3.50 ± 1.50	4.69 ± 4.25
$\tau_2(ps)$	0.12 ± 0.03	0.14 ± 0.04	0.10 ± 0.01
α	1.67 ± 0.30	1.91 ± 0.37	1.90 ± 0.28

Table 5.2: The group average of the extracted parameters with their standard errors and the corresponding average R^2_{adj} of the fittings for all 74 data samples.

5.3 Statistical Analysis

All 74 human breast samples are fitted by the THz-nDB model (5.5) using the described gradient sampling algorithm. The average value of each model parameter and its standard error in each breast tissue group is tabulated in Table 5.2. The contrast values in these parameters between tissue groups suggest their possible application for breast cancer classification. In the following subsections, we shall employ essential statistical measures to analyse the differences offered by these model parameters, and accordingly select potential parameters for breast cancer classification.

5.3.1 Correlation

Pearson correlations between the THz-nDB parameters and tissue histopathology are presented in Table 5.3. The parameter σ of the THz-nDB model (5.5) provide the

Component	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	$\tau_1(ps)$	$\tau_2(ps)$	α
Fat	-0.50	-0.09	-0.50	-0.61	-0.75	-0.122	-0.01	-0.09
Fibre tissue	-0.01	-0.05	0.14	0.25	-0.17	0.02	0.29	-0.16
Tumour	0.42	0.11	0.29	0.30	0.75	0.09	-0.21	0.08

Table 5.3: Pearson correlation between the parameters of the THz-nDB model (5.5) and the percentage of tissue components

highest correlation -0.75 with the fat percentage of tissue. This negative value of correlation implies that the reduction in the value of σ is proportional to the increase of fat percentage in the examined samples. Alternately, the correlation 0.75 of σ with tumour content explains the contribution of increasing neoplasm to the rise of this parameter. In fact, the similarity between fat and tumour in terms of high correlation with σ represents the nature that tumour has less adipose tissue and a more homogeneous structure in that respect as compared to fibrous tissue.

One of the recent challenges in breast cancer detection with THz imaging is that the dielectric properties/optical properties of fibrous tissue and tumour are similar [Ashworth et al., 2009]. In fact, fibrous tissue together with glands and connective duct have a high water content and hence respond to THz radiation as comparably as cancerous tissue does [Chen, Chen, Tseng, Lu, Kuo, Fu, Lee, Tsai, Huang, Chuang, Hwang and Sun, 2011]. As a result, it is clinically more apparent to detect tumours buried in fatty tissues rather than those in fibrous tissues. However, tumour is still expected to contain more water (up to 20% higher) than fibrous tissue, which could be essential for distinguishing between these two types of tissues. In addition, the similarity between fat and tumour in terms of high correlation with σ represents the nature that tumour has less adipose tissue and a more homogeneous structure in that respect as compared to fibrous tissue. Indeed, the significantly higher average value 7.89 of σ for tumours than the value 4.41 of fibrous tissue can be found in Table 5.2. In short, σ marks itself a highly prominent feature for differentiating tumour from fibrous tissue. Despite that we could not find the correlation between fibrous percentage and the parameters of the THz-nDB model (5.5), it is worth reminding that Pearson correlation only measures the linear relationship between two variables. Therefore, the relationship between the model parameters and fibrous content may be nonlinear and Pearson correlation is probably not a suitable statistical measure for such a case.

5.3.2 Significance of Difference

As the Pearson correlation is to reveal the physical meaning behind the values of the THz-nDB parameters rather than to directly analyse the contrast between tissue group, we use the independent 2-tailed t-test for these parameters. The specific p-values of the test corresponding to each pair of healthy tissue and tumour were provided in Table 5.3.2. The parameter σ demonstrates the statistically significant difference between

Pair of groups	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	τ_1	τ_2	α
Fat-Tumour	1.9E-6	0.29	4.16E-5	3.8E-5	3.9E-20	0.25	0.27	0.28
Fibrous-Tumour	0.05	0.33	0.40	0.84	4.6E-7	0.61	0.08	1.00
Normal-Tumour	6.1E-5	0.31	0.00	0.01	1.8E-14	0.42	0.04	0.58

Table 5.4: P-value of independent two-tailed sample t-test for each pair of sample groups

fibrous tissue and tumour with the p-value of approximately 0 which is considerably lower than the critical value 0.05 of the t-test. Similar statistical results are also seen between fat and tumour. While ϵ_∞, $\Delta\epsilon_2$, $\Delta\epsilon_3$, σ, τ_1, τ_2 can demonstrate significant differences between healthy breast tissue and tumour, only ϵ_∞, σ is capable to specify the difference between fibrous tissue and tumour.

5.3.3 ROC Analysis

Both the Pearson correlation analysis and independent two-tailed sample t-test introduce the two parameters $\Delta\epsilon_3$ and σ of (5.5) as an advantageous indicator which is highly sensitive to histopathology of breast samples. Therefore, an analysis from the perspective of classification is essential to accurately select rich features as well as confirm the potential of our model-parameter approach in breast cancer margin detection with THz imaging. A common technique to explore the sensitivity and specificity of classifiers in medical community is the receiver operating characteristic (ROC) plot [Fawcett, 2004]. Fig. 5.6 demonstrates the ROC plot for σ as a classifier with its threshold

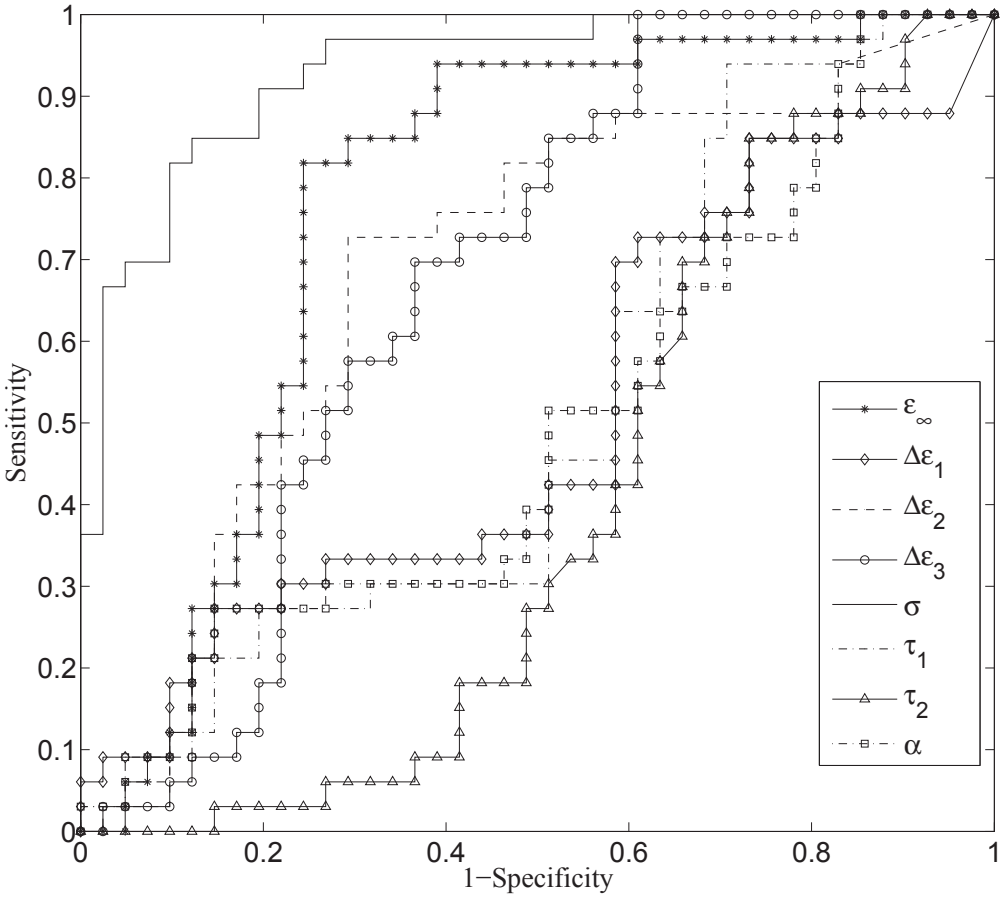

Figure 5.6: ROC plot for various thresholds of σ.

values chosen by averaging two consecutive test values of σ. For the purpose of evaluating σ regarding its merit of classification, the area under the ROC curve was used. The AUROC values of all eight THz-nDB parameters were calculated and recorded in Table 5.5. Particularly, the AUROC of σ is the highest with the value of 0.93. To

Parameter	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	τ_1	τ_2	α
AUC	0.77	0.51	0.69	0.67	0.93	0.52	0.40	0.50

Table 5.5: AUROC values obtained by ROC of each THz-nDB parameter of (5.5)

the best of our knowledge, this outcome is remarkably higher than the best AUROC

value of up to 0.78 in literature [St.Peter et al., 2013]. Therefore, σ is selected to be the principal classification feature. The remaining parameters of the THz-nDB model (5.5) can be considered as supporting features to improve the classification capability. However, combination of well-correlated parameters does not facilitate improvement of classification capability, thus it is necessary to find the independence parameters for the combination. As a result, linear correlations between all eight THz-nDB parameters were calculated to determine the independent parameters for classification.

5.4 Classification Potential

5.4.1 Support Vector Machine

The support vector machine has been emerging as the newest and among the most popular learning algorithms for pattern recognition [Burges, 1998]. The statistical significance of difference in the parameters of the THz-nDB model (5.5) between healthy breast tissue and tumour logically suggests the applicability of SVM as this classification algorithm aims to search for a hyperplane separating two data classes in a multi-dimensional space. Similarly to section 4.4.2, the SVM toolbox of Chang and Lin [2011] in the Matlab environment was used for classifying breast tumours with the parameters of nDB. Based on a number of trial simulations, the Gaussian radial basic (RBF) kernel function was applied due to its best classification performance for the data. This kernel function requires adjusting its parameter (C, γ) to suitable values in order to optimize the classification performance. Therefore, the grid-search was used to simultaneously find the optimal (C, γ).

5.4.2 Bayesian Neural Networks

The initial results with the SVM presented later in section 5.4.4 are encouraging but not able to fully reflect the classification potential of the model parameters in terms

of combining them. In fact, more parameters incorporated into the classification could not improve the classification accuracy. This motivates taking further steps towards using other applicable methods to improve the classification performance of these parameter combinations. Therefore, this section explores the limitation of the support vector machine classifier as well as the applicability of the Bayesian neural networks to improve the classification accuracy using combinations of the THz-nDB parameters for breast cancer detection. The Bayesian learning algorithm for neural networks not only enhances the their generalization but also makes the best use of data, thus making it preferable to small data with increasing noisy information and complexity [Bishop, 1995].

A traditional learning method such as SVM and regular neural networks commonly encounters crucial issues of generalisation that is defined by how well an obtained prediction model can detect new cases excluding from training data. The generalization loss leads to either underfitting or overfitting the data structure. The problem can be solved by determining appropriate complexity of the prediction model through globally searching its design parameters. This approach is very intensive and requires using a part of data for validation of the parameter search, thus not being able to optimize the use of data source [Hagan et al., 2014].

Bayesian neural networks (BNN) have been seen as a practical and powerful tool to improve the generalization and performance of neural networks since they were introduced by [MacKay, 1995]. The Bayesian framework applied in this method allows the learning process to overcome the aforementioned challenge. Particularly, based on the Bayes' theorem a probability distribution of network parameters is obtainable in the Bayesian learning. By that it means that uncertainty and noisy information of data can be taken into consideration to improve the prediction performance. In addition, the learning process using the Bayesian regulation facilitates automatic adjustment of network hyper-parameters, which are regulation constants controlling the complexity of the prediction model, to the most appropriate values. This allows the elimination of

using the validation set of data, thus maximizing data resource for training. As a result, BNN is of great interest to handling our small data. Besides, by viewing our multiparameter problem in this paper from the advantages of BNN, we can find it a probable solution to dealing with the increasing complexity of the prediction model when more model parameters are incorporated into the classification. Indeed, this complexity issue is directly concerned with adjusting more regulation constants, which is considered as an important advantage of the Bayesian approach [MacKay, 1995].

5.4.3 Accuracy Estimation

Both LOOCV and RRSCV are applied to validate the classification accuracy. Despite that LOOCV provides an unbiased estimation of accuracy, its prediction has not only high variance but also possibility of being too optimistic. On the contrary, the RRSCV with a significant proportion of data left out for testing does not make the best use of data for training but offers a better balance between bias and variance when estimating the classification accuracy. Combing the two validation methods is necessary for accurately justifying the classification performance. For the study in this section, it was chosen to use 80% of the data (59 samples) for training and 20% (15 samples) for testing in the RRSCV with 1000 repetitions of the training-testing process for each classifier.

5.4.4 Results

The LOOCV values of the eight THz-nDB parameters are presented in Table 5.6. As expected, σ has the highest LOOCV accuracy of 86.49% among the THz-nDB

Parameters	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	τ_1	τ_2	α
LOOCV	81.1	68.9	74.3	63.5	86.5	67.6	63.5	63.5

Table 5.6: SVM Classification Accuracy (%) in LOOCV corresponding to each THz-nDB parameter.

parameters. As expected, ϵ_∞ obtains the second highest LOOCV value which is followed by the classification accuracy of $\Delta\epsilon_3$. The LOOCV accuracies of the other THz-nDB parameters are far lower. In fact, these results reflect the earlier statistical analysis of these model parameters in section 5.3.

According to the LOOCVs of the THz-nDB parameters, their ten classification combinations are chosen to investigate and annotated by C1 to C10 respectively as can be seen in both Table 5.7 and 5.8. In fact, based on the statistical analysis in section 5.3, it is possible to form a variety of potential combinations for the classification. However, by analysing the LOOCV and RRSCV accuracy simultaneously with either the SVM or BNN a number of the combinations were filtered, and hence, only the best ten combinations were selected to present. They are not only able to achieve the highest classification accuracies but also optimal in terms of low dimension and complexity.

Combination Parameter	C	γ	LOOCV(%)	RRSCV(%)
C1. σ	1	0.25	86.5	85.0 $\pm$ 8.1
C2. $(\epsilon_\infty, \sigma)$	512	0.125	91.9	86.6 $\pm$ 8.9
C3. (σ, τ_1)	16	0.125	91.9	88.3 $\pm$ 8.5
C4. $(\epsilon_\infty, \Delta\epsilon_1, \sigma)$	4	0.0625	87.8	81.0 $\pm$ 9.5
C5. $(\epsilon_\infty, \Delta\epsilon_2, \sigma)$	1	0.03125	85.1	85.6 $\pm$ 8.5
C6. $(\epsilon_\infty, \Delta\epsilon_3, \sigma)$	256	0.25	93.2	87.6 $\pm$ 8.8
C7. $(\epsilon_\infty, \sigma, \tau_2)$	512	0.125	90.5	86.4 $\pm$ 8.6
C8. $(\epsilon_\infty, \sigma, \alpha)$	512	0.125	89.2	84.2 $\pm$ 9.6
C9. $(\epsilon_\infty, \Delta\epsilon_2, \Delta\epsilon_3, \sigma)$	1	0.03125	85.1	85.3 $\pm$ 8.3
C10. $(\epsilon_\infty, \sigma, \tau_2, \alpha)$	512	0.125	89.2	84.6 $\pm$ 9.6

Table 5.7: The estimated accuracies (%) by LOOCV and RRSCV for applying the double Debye parameters with the SVM to classify the healthy breast tissue and the breast tumour.

Table 5.7 shows the LOOCVs and the average classification accuracies with their standard deviations estimated by RRSCV for using SVM classifiers with combinations of the THz-nDB parameters in (5.5). The optimal kernel parameters including C and γ for each set of the model parameters are accountable for the best LOOCV of the combination. Under the impact of the smaller training set, C6 obtains the highest LOOCV

(93.2%) but a far lower RRSCV accuracy (87.6%). Despite that C1 only contains σ, its classification performance is still better than the higher-dimension combinations such as C5 and C9. In fact, σ has been considered as the most potential parameter of the THz-nDB model (5.5) for breast cancer classification [Truong et al., 2015]. C3 should be the most suitable for the SVM method thanks to its high and stable accuracies predicted by LOOCV (91.9%) and RRSCV (88.3%). However, the combinations with more THz-nDB parameters such as C4–C10 do not improve or even weaken the classification performance using SVM. As mentioned earlier, this remains the challenge of applying the SVM approach for the data, which motivates further investigation into the applicability of BNN.

Combination Parameter	LOOCV(%)	RRSCV(%)
C1. σ	86.5	86.5 ± 8.7
C2. $(\epsilon_\infty, \sigma)$	91.9	92.9 ± 6.6
C3. (σ, τ_1)	94.6	88.6 ± 7.4
C4. $(\epsilon_\infty, \Delta\epsilon_1, \sigma)$	94.6	93.1 ± 6.5
C5. $(\epsilon_\infty, \Delta\epsilon_2, \sigma)$	96.0	92.3 ± 6.5
C6. $(\epsilon_\infty, \Delta\epsilon_3, \sigma)$	94.6	93.0 ± 5.8
C7. $(\epsilon_\infty, \sigma, \tau_2)$	93.2	93.4 ± 6.5
C8. $(\epsilon_\infty, \sigma, \alpha)$	96.0	92.2 ± 6.6
C9. $(\epsilon_\infty, \Delta\epsilon_2, \Delta\epsilon_3, \sigma)$	97.3	92.4 ± 6.4
C10. $(\epsilon_\infty, \sigma, \tau_2, \alpha)$	97.3	93.6 ± 6.4

Table 5.8: The estimated accuracies (%) by LOOCV and RRSCV for applying the double Debye parameters with the BNN to classify the healthy breast tissue and the breast tumour.

According to Table 5.8, the problem of SVM indeed can be overcome by BNN structured by 10 hidden nodes. To be more specific, the overall classification performance of the combinations is improved whenever an extra THz-nDB parameter is added to the classification. Accordingly, the best accuracy obtained with the four-parameter combinations including C9 and C10 is 97.3% in LOOCV and 93.6% in RRSCV. The highest LOOCVs of the three- and two-parameter combinations including C2-C8 are 96.0% and 94.6% respectively. However, although the average accuracies of C2-C10

in RRSCV achieve very high values from 92.2%, the impact of increasing the number of input parameters on the classification is not significant. The similarity in C1,C2,C4 between LOOCVs and RRSCV accuracies suggests BNN can learn the data structure of these combinations very well regardless of the smaller data set for training in RRSCV. Conversely, the impressive LOOCVs of the rest are not achievable in RRSCV due to the shortage of training data. By and large, classifying the breast tumour using the combinations of the model parameters in (5.5) with BNN offers better overall performance than that with SVM.

5.5 Discussion and Conclusion

The THz-nDB model for complex permittivity of breast tissue at THz frequencies as given in equation (5.5) is proposed for the first time. This dielectric model is developed based on combining non-Debye relaxation theory and the Debye relaxation process due to the non-exponential dielectric response of breast tissue at low THz frequencies and the impact of a considerable proportion of water in breast tissue. Our fitting procedure employed the gradient sampling algorithm for non-optimization to extract optimal parameters of the THz-nDB model. Simulations prove that our proposed models can accurately mimic the measured data of fat/adipose tissue, healthy fibrous and cancerous breast tissue. A variety of statistical measures was used to investigate the potential of the THz-nDB parameters in terms of unveiling the tissue histopathology and cancer classification. Our analysis demonstrates two encouraging results: the parameter σ of the THz-nDB model highly correlates with fat tissue and tumour percentage of examined breast tissue samples, which marks this parameter as an indicator for pathology diagnosis; the ROC analysis confirms the more remarkable discrimination ability of σ to classify breast cancer as compared to previous studies.

The classification results with the SVM classifier generally indicate its ability to improve the classification performance with the combinations of up to only two THz-nDB

parameters, such as C2 and C3. Especially, C3 provide the LOOCV of 91.9% together with the RRSCV of 88.3% which is also the best performance among the SVM classifiers. Thus, the SVM method is limited in terms of efficiently learning the data structure of the combinations of the THz-nDB parameters for classification. Therefore, we revised the problem and successfully applied the BNN classifier to improve performance of the combinations. Particularly, ten parameter combinations C1-C10 were introduced for investigation with both SVM and BNN. Using the BNN, the best LOOCV is enhanced to 97.3% with the four-parameter combinations as compared to 93.2% with the SVM. The advance of BNN in classifying the data is also expressed over the estimated accuracies in RRSCV. The average accuracies vary between about 92.17–93.57% corresponding the different combinations, which are also by far higher than 88.3% (RRSCV) with the SVM. Apart from that, the classification accuracies predicted by cross validation methods in this study may be statistically insufficient for making a confirmation of the true classification accuracies in practice due to the used small data. However, our encouraging results should be basic to future studies that examine larger data such as THz images of breast tumour in order to improve the cancer-margin detection in BCS. Further developments of this study will also include selection of the best parameter combinations and improvement of classification methodologies.

The empirical THz-nDB model also requires more specific elaboration on physical process underlying the non-Debye dielectric response of breast tissue in the future. Knowing the best model to describe the dielectric properties allows applying the FDTD techniques using the proposed model to simulate the interaction of THz radiation with breast tissue. Thus, it becomes possible to recognize the theoretical differences in pulse shape between normal and tumour tissue as well as improve our understanding of how normal tissue changes its response due to tumorigenesis. This will be beneficial to deal with heterogeneous tissue which includes both characteristics of tumour and normal.

Furthermore, successfully modelling the dielectric response of breast tissue with the model THz-nDB implies that the low THz frequency range is the most accountable for

producing the data to classify breast cancer. However, applying the low frequencies for a practical THz imaging system, mainly THz Pulse Imaging (TPI) or Single Frequency THz Imaging, leads to a sacrifice in spatial resolution since the minimum size of tumour identifiable by THz needs to be larger. For instance, 1 THz generally has a spatial resolution of around 300 microns. Consequently, the application of THz imaging is limited to sampled areas larger than this size.

Chapter 6

Conclusion

6.1 Summary

This thesis demonstrates the studies of the terahertz dielectric properties of human skin and breast tissue with the aim of not only providing a better understanding of the contrasts due to carcinogenesis in their THz images but also introducing potential features for cancer classification. The database used, consisting of refractive indices and absorption coefficients measured from the *ex vivo* specimens of these tissues, was widely published in previous studies [Kindt and Schmuttenmaer, 1996; Pickwell et al., 2005; Wallace et al., 2006; Ashworth et al., 2009]. The method of the investigation is generally based on modelling the dielectric spectral responses of the examined tissues and then analysing the extracted parameters of the dielectric models in terms of their cancer discrimination capability. As skin and breast tissues have relatively different compositions, their respective dielectric spectra in the terahertz regime display distinct features. As a result, two separate dielectric models are taken into account for skin and breast tissue in this thesis.

Numerous studies have indicated that the double Debye model is capable of describing the dielectric properties of human skin in terahertz frequencies due to its high water content. However, it is shown that the model parameters obtained by previous studies using the nonlinear least square method cannot facilitate fitting the data well over an entire spectral range of up to 2 Hz. This means that the existing parameter extraction method is not optimal. Therefore, two global-based optimization algorithms have been introduced to improve the fitting quality and thus extract the optimal model parameters in this thesis. They focus on handling the proposed parametric optimization problem derived from a sum-of-square-error function between the dielectric data and the double Debye model. The first algorithm employs a grinding-based technique that locates the optimal solution through comparing values of the objective function at each point in its variable space. The outcomes prove that this approach is robust and able to significantly improve the fitting quality throughout the examined frequency domain as

compared to the initial method in literature. However, the grinding-based optimization algorithm poses an intensive computation, thus making it less practical to be used for THz imaging. Therefore, the second algorithm based on the BB technique of global optimization has been developed. The simulation results show that this method considerably improves the computation efficiency by reducing the number of QPs from thousands in the grinding-based approach to few hundreds while still satisfying the optimality of the extracted parameters. The optimal results of these proposed algorithms not only confirm the applicability of the double Debye model in the frequency band 0.2–2 THz but also facilitate precise transformation of information from the measured dielectric properties to this modelling.

To investigate the potential of the double Debye model in terms of cancer discrimination, the model parameters were extracted from the complex permittivity of 23 skin samples using the GO based algorithms. These tissue samples, which are either healthy or cancerous with a certain percentage of BCC, were excised from ten patients undergoing MMS. Statistical analysis of the model parameters indicates that ϵ_s is the only Debye parameter showing a significant difference between normal and BCC samples with 33% difference in mean. It also found that this parameter highly correlates with the BCC percentage in the samples. Furthermore, the frequency-dependent parameters including $\Im(\epsilon_r^m(\omega))$, $\Im(\epsilon(\omega)_r^D)$, $\Im(D_1(\omega))$ and $\Im(D_2(\omega))$ are significantly different between normal tissue and BCC while their variations have high correlations with the BCC content. As regards cancer discrimination, ϵ_s and these four combination parameters show their strong capability through their average AUCs ranging between 0.79–0.93. The LP and SVM classifiers were used to classify the BCC samples with all these parameters whereas the classification accuracy was estimated by both LOOCV and RSCV. The results indicate that ϵ_s is the best among the DD parameters as it obtains high and consistent classification accuracies with the SVM in both cross-validation estimations (91.3% LOOCV and 90.9% RSCV). The combination of ϵ_s and ϵ_2 improves the classification accuracy that, as a result, reaches the LOOCV of 95.7% with both

LP and SVM method. In addition, $\Im(\epsilon)_r^D(0.5\ THz)$ is the most valuable among the frequency-dependent parameters, which provides the LOOCV of 95.7% and RSCV of 88%. The SVMs generally perform better than the LP classifiers in most cases by virtue of the advantages of SVM to deal with linearly inseparable data. All the outcomes confirm the cancer classification potential of the dielectric approach based on the double Debye model.

As breast tissue is constructed by a significant proportion of fat cells and proteins as well as containing less free water compared to skin tissue, its terahertz dielectric spectra might pose non-Debye responses. This results in the failure of the double Debye model to describe the dielectric properties of breast tissue. Therefore, the non-Debye model (THz-nDB) for breast tissue has been developed to replace the double Debye model. The gradient sampling algorithm for nonsmooth optimization has been applied to fit this model to the measured data. The simulation results of 74 excised breast samples indicate that the proposed modelling can successfully fit the measured complex permittivities of all the breast samples. The statistical analysis of the extracted model parameters suggests σ as the most promising parameter for breast cancer discrimination. This parameter not only varies significantly between normal and cancerous breast tissue according to the t-test but also correlates well with fat and, more importantly, the tumour content of the tested samples. In addition to the statistical assessment, the AUROC value of 0.93 for σ in the ROC analysis is completely distinct from the remaining model parameters, thus confirming the cancer discrimination capability of this parameter.

Recent popular classifiers such as SVM and BNN were employed to examine the actual classification performance of the THz-nDB parameters. Both the methods obtain the same classification accuracy of 86.5% with σ, which also represents the best performance among all the model parameters. Furthermore, the essential advantage of the advance classification algorithms is their ability to handle high-dimensional data and multi-features. Therefore, ten classification parameters C1–C10 including σ and its

combinations with the remaining parameters of the THz-nDB model were considered for the classification improvement. The outcomes of the SVM classification suggest C3, consisting of only σ and τ_1, as the best classification feature for the SVM. This combination provides the LOOCV of 91.2% and RRSCV of 88.3%. However, the combinations with more than two model parameters in the SVM classification do not improve the performance or even make it worse. On the contrary, the BNN method applied for the combinations of four model parameters is capable of classifying the breast tumour with LOOCV and RSSCV accuracies of up to 97.3% and 93.6% respectively. The highest obtained accuracies correspond to the combination C10. In addition to this remarkable improvement, the BNN surpasses the SVM in terms of overall performance as it has been shown that the accuracies, approximated by both LOOCV and RRSCV, for the BNN classification with the majority of the combinations, are higher than those for the SVM. According to all the results and analysis, BNN proves to be more appropriate for the THz-nDB parameters and their combinations than SVM. Last but not least, a much larger amount of data is of particular importance in confirming the pilot study of using the non-Debye model THz-nDB for breast cancer classification.

6.2 Future Research

Optical data applied in the study of this thesis was collected from *ex vivo* measurements in the transmission setup. It has been shown that *in vivo* and *in vivo* data are not exactly the same because *in vivo* measurements are performed in the reflection geometry. In addition, excisions in *ex vivo* experiments potentially cause tissue deformation and shrinkage [Wallace, Fitzgerald, Shankar, Flanagan, Pye, Cluff and Arnone, 2004]. Therefore, *in vivo* data will need to be investigated with the dielectric models in order to fully understand their applicability to the THz imaging *in vivo*.

Future study will conduct classification on THz images in order to confirm the accuracy of cancer margin detection in both MMS for skin cancer and BCS for breast

cancer. Since the applied dielectric models such as the double Debye model or the non-Debye model reflect the molecular interaction between THz interaction and biological molecules, it is reasonable to expect the model parameters to be the origin of the terahertz contrast between normal and diseased tissue. In fact, this means that direct comparison between these parameters and the existing time- and frequency-domain parameters for imaging should be verified. In addition to the feature selection, the classification efficiency can be improved through testing and developing classification algorithms.

The calculation of the refractive index and absorption coefficient is not error-free due to the imperfection of measurements [Withayachumnankul et al., 2008]. The random and systematic errors are mostly related to the sample thickness and signal alignment. Furthermore, the applied transmission/reflection models for calculating the optical constants from the measured THz pulses are unable to take these error sources into account [Duvillaret et al., 1996, 1999; Dorney et al., 2001; Pupeza et al., 2007]. As a result, the uncertainties of determining the optical properties of the sample need to be considered in further studies because can potentially affect the outcomes of the dielectric-modelling approach.

A THz intraoperative probe has been developed that can perform measurements in reflection geometry during surgical processes such as MMS and BCS [Parrott et al., 2011]. The outcomes of the dielectric studies presented in this thesis will facilitate choosing classification features from the reflected signal collected using this probe. However, the currently-applied extraction procedure for the dielectric models is often performed in the transmission mode. It is also relatively time-consuming due to the optimization steps. This makes the dielectric approach quite impractical for real-time implementation in the THz imaging probe system. Therefore, solutions for improving the speed of the parameter extraction procedure will be considered in future works.

Bibliography

AIHW & AACR [2012], 'Cancer in Australia: an overview 2012', *Canberra: AIHW* **Cancer series no. 74**(Cat. no. CAN 70).

Ajili, L., Scalari, G., Hofstetter, D., Beck, M., Faist, J., Beere, H., Davies, G., Linfield, E. and Ritchie, D. [2002], 'Continuous-wave operation of far-infrared quantum cascade lasers', *Electronics Letters* **38**(25), 1675–1676.

Alison, J. M. and Sheppard, R. J. [1990], 'A precision waveguide system for the measurement of complex permittivity of lossy liquids and solid tissues in the frequency range 29 GHz to 90 GHz I.The liquid system for 29 to 45 GHz; an investigation into water', *Measurement Science and Technology* **1**(10), 1093.

Altman, D. G. [1991], *Practical Statistics for Medical Research*, Chapman & Hall.

American Cancer Society [2011], *Global Cancer Facts & Figures 2nd Edition*, Atlanta: American Cancer Society.

Anderson, B. O., Yip, C.-H., Smith, R. A., Shyyan, R., Sener, S. F., Eniu, A., Carlson, R. W., Azavedo, E. and Harford, J. [2008], 'Guideline implementation for breast healthcare in low-income and middle-income countries', *Cancer* **113**(S8), 2221–2243.

Argenziano, G. and Soyer, H. P. [2001], 'Dermoscopy of pigmented skin lesions-a valuable tool for early diagnosis of melanoma.', *The Lancet Oncology* **2**(7), 443–449.

Arnone, D. D., Ciesla, C. M., Corchia, A., Egusa, S., Pepper, M., Chamberlain, J. M.,

Bezant, C., Linfield, E. H., Clothier, R. and Khammo, N. [1999], Applications of terahertz (THz) technology to medical imaging, *in* 'Proc. SPIE', Vol. 3828, pp. 209–219.

Ashworth, P. C., Pickwell-MacPherson, E., Provenzano, E., Pinder, S. E., Purushotham, A. D., Pepper, M. and Wallace, V. P. [2009], 'Terahertz pulsed spectroscopy of freshly excised human breast cancer', *Opt. Express* **17**(15), 12444–12454.

Ashworth, P., O'Kelly, P., Purushotham, A., Pinder, S., Kontos, M., Pepper, M. and Wallace, V. [2008], An intra-operative THz probe for use during the surgical removal of breast tumors, *in* 'Infrared, Millimeter and Terahertz Waves, 2008. IRMMW-THz 2008. 33rd International Conference on', pp. 1–3.

Auston, D. H. [1975], 'Picosecond optoelectronic switching and gating in silicon', *Applied Physics Letters* **26**(3), 101–103.

Auston, D. H. and Cheung, K. P. [1985], 'Coherent time-domain far-infrared spectroscopy', *J. Opt. Soc. Am. B* **2**(4), 606–612.

Barthel, J. and Buchner, R. [1991], 'High frequency permittivity and its use in the investigation of solution properties', *Pure and Applied Chemistry* **63**(10), 1473–1482.

Beard, M. C., Turner, G. M. and Schmuttenmaer, C. A. [2001], 'Subpicosecond carrier dynamics in low-temperature grown GaAs as measured by time-resolved terahertz spectroscopy', *Journal of Applied Physics* **90**(12), 5915–5923.

Beard, M. C., Turner, G. M. and Schmuttenmaer, C. A. [2002], 'Terahertz spectroscopy', *The Journal of Physical Chemistry B* **106**(29), 7146–7159.

Bennett, D. B., Li, W., Taylor, Z. D., Grundfest, W. S. and Brown, E. R. [2011], 'Stratified media model for terahertz reflectometry of the skin', *IEEE Sensors Journal* **11**(5), 1253–1262.

Bennett, K. P. and Mangasarian, O. L. [1992], 'Robust linear programming discrimination of two linearly inseparable sets', *Optimization Methods and Software* **1**(1), 23–34.

Berry, E., Walker, G. C., Fitzgerald, A. J., Zinovev, N. N., Chamberlain, M., Smye, S. W., Miles, R. E. and Smith, M. A. [2003], 'Do in vivo terahertz imaging systems comply with safety guidelines?', *Journal of Laser Applications* **15**(3), 192–198.

Bishop, C. M. [1995], *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Blichert-Toft, M., Nielsen, M., Dring, M., Mller, S., Rank, F., Overgaard, M. and Mouridsen, H. T. [2008], 'Long-term results of breast conserving surgery vs. mastectomy for early stage invasive breast cancer: 20-year follow-up of the Danish randomized DBCG-82TM protocol', *Acta Oncologica* **47**(4), 672–681.

Borisova, E., Angelova, L. and Pavlova, E. [2014], 'Endogenous and exogenous fluorescence skin cancer diagnostics for clinical applications', *IEEE Journal of Selected Topics in Quantum Electronics* **20**(2), 211–222.

Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., Jong, R. A., Hislop, G., Chiarelli, A., Minkin, S. and Yaffe, M. J. [2007], 'Mammographic density and the risk and detection of breast cancer', *New England Journal of Medicine* **356**(3), 227–236.

Breusch, T. S. and Pagan, A. R. [1979], 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica* **47**(5), 1287–1294.

Brown, R., Leung, J., Fisher, S., Frey, K., Ethier, S. and Wahl, R. [1996], 'Intratumoral distribution of tritiated-FDG in breast carcinoma: correlation between Glut-1 expression and FDG uptake', *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **37**(6), 1042–1047.

Brun, M.-A., Formanek, F., Yasuda, A., Sekine, M., Ando, N. and Eishii, Y. [2010], 'Terahertz imaging applied to cancer diagnosis', *Physics in Medicine and Biology* **55**(16), 4615.

Burges, C. J. [1998], 'A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery* **2**(2), 121–167.

Burke, J., Lewis, A. and Overton, M. [2005], 'A robust gradient sampling algorithm for nonsmooth, nonconvex optimization', *SIAM Journal on Optimization* **15**(3), 751–779.

Canfell, K., Banks, E., Moa, A. M. and Beral, V. [2008], 'Decrease in breast cancer incidence following a rapid fall in use of hormone replacement therapy in australia', *Medical Journal of Australia* **188**(11), 641–644.

Carrig, T. J., Rodriguez, G., Clement, T. S., Taylor, A. J. and Stewart, K. R. [1995], 'Scaling of terahertz radiation via optical rectification in electro-optic crystals', *Applied Physics Letters* **66**(2), 121–123.

Cawley, G. C. and Talbot, N. L. [2003], 'Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers', *Pattern Recognition* **36**(11), 2585 – 2592.

Chan, W. L., Deibel, J. and Mittleman, D. M. [2007], 'Imaging with terahertz radiation', *Reports on Progress in Physics* **70**(8), 1325.

Chang, C. C. and Lin, C. J. [2011], 'LIBSVM: A library for support vector machines', *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chasmar, R. P., Mitchell, W. H. and Rennie, A. [1956], 'Theory and performance of metal bolometers', *J. Opt. Soc. Am.* **46**(7), 469–477.

Chen, G., Pei, J., Yang, F., Zhou, X. Y., Sun, Z. and Cui, T. J. [2012], 'Terahertz-wave

imaging system based on backward wave oscillator', *IEEE Transactions on Terahertz Science and Technology* **2**(5), 504–512.

Chen, H., Chen, T. H., Tseng, T. F., Lu, J. T., Kuo, C. C., Fu, S. C., Lee, W. J., Tsai, Y. F., Huang, Y. Y., Chuang, E. Y., Hwang, Y. J. and Sun, C. K. [2011], 'High-sensitivity in vivo THz transmission imaging of early human breast cancer in a subcutaneous xenograft mouse model', *Opt. Express* **19**(22), 21552–21562.

Chen, H., Lee, W.-J., Huang, H.-Y., Chiu, C.-M., Tsai, Y.-F., Tseng, T.-F., Lu, J.-T., Lai, W.-L. and Sun, C.-K. [2011], 'Performance of THz fiber-scanning near-field microscopy to diagnose breast tumors', *Opt. Express* **19**(20), 19523–19531.

Chen, H.-T., Kersting, R. and Cho, G. C. [2003], 'Terahertz imaging with nanometer resolution', *Applied Physics Letters* **83**(15), 3009–3011.

Clothier, R. and Bourne, N. [2003], 'Effects of THz exposure on human primary keratinocyte differentiation and viability', *Journal of Biological Physics* **29**(2-3), 179–185.

Cole, B. E., Woodward, R. M., Crawley, D. A., Wallace, V. P., Arnone, D. D. and Pepper, M. [2001], Terahertz imaging and spectroscopy of human skin in vivo, *in* 'Proc. SPIE', Vol. 4276, pp. 1–10.

Coughlin, S. S. and Ekwueme, D. U. [2009], 'Breast cancer as a global health concern', *Cancer Epidemiology* **33**(5), 315 – 318.

Darmo, J., Tamosiunas, V., Fasching, G., Kröll, J., Unterrainer, K., Beck, M., Giovannini, M., Faist, J., Kremser, C. and Debbage, P. [2004], 'Imaging with a terahertz quantum cascade laser', *Opt. Express* **12**(9), 1879–1884.

Diepgen, T. and Mahler, V. [2002], 'The epidemiology of skin cancer.', *British Journal of Dermatology* **146**, 1–6.

Dobroiu, A., Yamashita, M., Ohshima, Y. N., Morita, Y., Otani, C. and Kawase, K. [2004], 'Terahertz imaging system based on a backward-wave oscillator', *Appl. Opt.* **43**(30), 5637–5646.

Doria, A., Gallerano, G. and Giovenale, E. [2002], 'Free electron broad-band THz radiator', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **483**(12), 461–465.

Dorney, T. D., Baraniuk, R. G. and Mittleman, D. M. [2001], 'Material parameter estimation with terahertz time-domain spectroscopy', *J. Opt. Soc. Am. A* **18**(7), 1562–1571.

Duvillaret, L., Garet, F. and Coutaz, J.-L. [1996], 'A reliable method for extraction of material parameters in terahertz time-domain spectroscopy', *IEEE J. of Selected Topics in Quantum Electronics* **2**(3), 739–746.

Duvillaret, L., Garet, F. and Coutaz, J.-L. [1999], 'Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy', *Applied Optics* **38**(2), 409–415.

Elder, D. E. [1995], 'Skin cancer. melanoma and other specific nonmelanoma skin cancers', *Cancer* **75**(S1), 245–256.

Fan, S., He, Y., Ung, B. S. and Pickwell-MacPherson, E. [2014], 'The growth of biomedical terahertz research', *Journal of Physics D: Applied Physics* **47**(37), 374009.

Faverly, D. R. G., Hendriks, J. H. C. L. and Holland, R. [2001], 'Breast carcinomas of limited extent', *Cancer* **91**(4), 647–659.

Fawcett, T. [2004], 'Roc graphs: Notes and practical considerations for researchers', *ReCALL* **31**(HPL-2003-4), 1–38.

Feldman, Y., Puzenko, A. and Ryabov, Y. [2002], 'Non-Debye dielectric relaxation in complex materials', *Chemical Physics* **284**(1-2), 139 –168.

Ferguson, B. and Zhang, X.-C. [2002], 'Materials for terahertz science and technology', *Nature Materials* **1**, 26–33.

Fischer, B. M., Walther, M. and Jepsen, P. U. [2002], 'Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy', *Physics in Medicine and Biology* **47**(21), 3807.

Fisher, L. D. and van Belle, G. [1993], *Biostatistics: A Methodology for the Health Sciences*, John Wiley & Sons.

Fitzgerald, A. J., Berry, E., Zinovev, N. N., Walker, G. C., Smith, M. A. and Chamberlain, J. M. [2002], 'An introduction to medical imaging with coherent terahertz frequency radiation', *Physics in Medicine and Biology* **47**(7), R67.

Fitzgerald, A. J., Pickwell-MacPherson, E. and Wallace, V. P. [2014], 'Use of finite difference time domain simulations and Debye theory for modelling the terahertz reflection response of normal and tumour breast tissue', *PLoS ONE* **9**(7), e99291.

Fitzgerald, A. J., Pinder, S., Purushotham, A. D., OKelly, P., Ashworth, P. C. and Wallace, V. P. [2012], 'Classification of terahertz-pulsed imaging data from excised breast tissue', *Journal of Biomedical Optics* **17**(1), 016005–1–016005–10.

Fitzgerald, A. J., Wallace, V. P., Jimenez-Linan, M., Bobrow, L., Pye, R. J., Purushotham, A. D. and Arnone, D. D. [2006], 'Terahertz pulsed imaging of human breast tumors', *Radiology* **239**(2), 533–540.

Fornage, B. D., McGavran, M. H., Duvic, M. and Waldron, C. A. [1993], 'Imaging of the skin with 20-MHz us.', *Radiology* **189**(1), 69–76.

Foster, K. R. and Schwan, H. P. [1989], 'Dielectric properties of tissues and biological materials: a critical review.', *Crit. Rev. Biomed. Eng.* **17**(1), 25.

Fujimoto, J. G., Pitris, C., Boppart, S. A. and Brezinski, M. E. [2000], 'Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy', *Neoplasia* **2**(12), 9 – 25.

Gabriel, C., Gabriel, S. and Corthout, E. [1996], 'The dielectric properties of biological tissues: I. Literature survey', *Physics in Medicine and Biology* **41**(11), 2231.

Gabriel, S., Lau, R. W. and Gabriel, C. [1996], 'The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues', *Physics in Medicine and Biology* **41**(11), 2271.

Gambichler, T., Regeniter, P., Bechara, F. G., Orlikov, A., Vasa, R., Moussa, G., Stcker, M., Altmeyer, P. and Hoffmann, K. [2007], 'Characterization of benign and malignant melanocytic skin lesions using optical coherence tomography in vivo', *Journal of the American Academy of Dermatology* **57**(4), 629–637.

Globus, T., Woolard, D., Khromova, T., Crowe, T., Bykhovskaia, M., Gelmont, B., Hesler, J. and Samuels, A. [2003], 'THz-spectroscopy of biological molecules', *Journal of Biological Physics* **29**(2-3), 89–100.

Golay, M. J. E. [1947], 'Theoretical consideration in heat and infrared detection, with particular reference to the pneumatic detector', *Review of Scientific Instruments* **18**(5), 347–356.

Gowen, A., OSullivan, C. and ODonnell, C. [2012], 'Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring and quality control', *Trends in Food Science & Technology* **25**(1), 40–46.

Gray, R., Salud, C., Nguyen, K., Dauway, E., Friedland, J., Berman, C., Peltz, E., Whitehead, G. and Cox, C. [2001], 'Randomized prospective evaluation of a novel technique for biopsy or lumpectomy of nonpalpable breast lesions: Radioactive seed versus wire localization', *Annals of Surgical Oncology* **8**(9), 711–715.

Grischkowsky, D., Keiding, S., van Exter, M. and Fattinger, C. [1990], 'Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors', *J. Opt. Soc. Am. B* **7**(10), 2006–2015.

Grosse, E. [2002], 'THz radiation from free electron lasers and its potential for cell and tissue studies', *Physics in Medicine and Biology* **47**(21), 3755.

Hagan, M. T., Demuth, H. B., Beale, M. H. and Jes, O. D. [2014], *Neural Network Design*, Martin Hagan.

Hall, N., Povoski, S., Murrey, D., Knopp, M. and Martin, E. [2007], 'Combined approach of perioperative 18F-FDG PET/CT imaging and intraoperative 18F-FDG handheld gamma probe detection for tumor localization and verification of complete tumor resection in breast cancer', *World Journal of Surgical Oncology* **5**(1), 143.

Hallock, G. G. and Lutz, D. A. [1998], 'Prospective study of the accuracy of the surgeon's diagnosis in 2000 excised skin tumors', *Plastic & Reconstructive Surgery* **101**, 1255–61.

Hamster, H., Sullivan, A., Gordon, S., White, W. and Falcone, R. W. [1993], 'Subpicosecond, electromagnetic pulses from intense laser-plasma interaction', *Phys. Rev. Lett.* **71**, 2725–2728.

Harde, H. and Grischkowsky, D. [1991], 'Coherent transients excited by subpicosecond pulses of terahertz radiation', *J. Opt. Soc. Am. B* **8**(8), 1642–1651.

Hassan, A., Hufnagle, D., El-Shenawee, M. and Pacey, G. [2012], Terahertz imaging for margin assessment of breast cancer tumors, *in* 'Microwave Symposium Digest (MTT), 2012 IEEE MTT-S International', pp. 1–3.

Havriliak, S. and Negami, S. [1967], 'A complex plane representation of dielectric and mechanical relaxation processes in some polymers', *Polymer* **8**(0), 161 – 210.

Hu, B. B. and Nuss, M. C. [1995], 'Imaging with terahertz waves', *Optics Letters* **20**(16), 1716–1718.

Huang, S., Ashworth, P. C., Kan, K. W., Chen, Y., Wallace, V. P., ting Zhang, Y. and Pickwell-MacPherson, E. [2009], 'Improved sample characterization in terahertz reflection imaging and spectroscopy', *Opt. Express* **17**(5), 3848–3854.

Huang, S. Y., Wang, Y. X. J., Yeung, D. K. W., Ahuja, A. T., Zhang, Y.-T. and Pickwell-MacPherson, E. [2009], 'Tissue characterization using terahertz pulsed imaging in reflection geometry', *Physics in Medicine and Biology* **54**(1), 149.

Huber, R., Brodschelm, A., Tauser, F. and Leitenstorfer, A. [2000], 'Generation and field-resolved detection of femtosecond electromagnetic pulses tunable up to 41 THz', *Applied Physics Letters* **76**(22), 3191–3193.

Hubers, H.-W. [2008], 'Terahertz heterodyne receivers', *IEEE Journal of Selected Topics in Quantum Electronics* **14**(2), 378–391.

Hunsche, S., Koch, M., Brener, I. and Nuss, M. [1998], 'THz near-field imaging', *Optics Communications* **150**(16), 22 – 26.

Hurt, W. [1985], 'Multiterm Debye dispersion relations for permittivity of muscle', *IEEE Transactions on Biomedical Engineering* **BME-32**(1), 60–64.

Huynh, P. T., Jarolimek, A. M. and Daye, S. [1998], 'The false-negative mammogram.', *RadioGraphics* **18**(5), 1137–1154.

Ikeda, D. M., Birdwell, R. L., OShaughnessy, K. F., Brenner, R. J. and Sickles, E. A. [2003], 'Analysis of 172 subtle findings on prior normal mammograms in women with breast cancer detected at follow-up screening', *Radiology* **226**(2), 494–503.

Jacobs, L. [2008], 'Positive margins: The challenge continues for breast surgeons', *Annals of Surgical Oncology* **15**(5), 1271–1272.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T. and Thun, M. J. [2008], 'Cancer statistics, 2008', *CA: A Cancer Journal for Clinicians* **58**(2), 71–96.

Jepsen, P. U., Jacobsen, R. H. and Keiding, S. R. [1996], 'Generation and detection of terahertz pulses from biased semiconductor antennas', *J. Opt. Soc. Am. B* **13**(11), 2424–2436.

Jiang, Z. and Zhang, X.-C. [1998], 'Single-shot spatiotemporal terahertz field imaging', *Opt. Lett.* **23**(14), 1114–1116.

Jiang, Z. and Zhang, X.-C. [1999], 'Terahertz imaging via electrooptic effect', *Microwave Theory and Techniques, IEEE Transactions on* **47**(12), 2644–2650.

Jiang, Z. and Zhang, X.-C. [2000], 'Measurement of spatio-temporal terahertz field distribution by using chirped pulse technology', *Quantum Electronics, IEEE Journal of* **36**(10), 1214–1222.

Joseph, C. S., Patel, R., Neel, V. A., Giles, R. H. and Yaroslavsky, A. N. [2014], 'Imaging of ex vivo nonmelanoma skin cancers in the optical and terahertz spectral regions optical and terahertz skin cancers imaging', *Journal of Biophotonics* **7**(5), 295–303.

Joseph, C. S., Yaroslavsky, A. N., Al-Arashi, M., Goyette, T. M., Dickinson, J. C., Gatesman, A. J., Soper, B. W., Forgione, C. M., Horgan, T. M., Ehasz, E. J., Giles, R. H. and Nixon, W. E. [2009], Terahertz spectroscopy of intrinsic biomarkers for non-melanoma skin cancer, *in* 'Proc. SPIE', Vol. 7215, pp. 72150I–72150I–10.

Kawase, K., Mizuno, M., Sohma, S., Takahashi, H., Taniuchi, T., Urata, Y., Wada, S., Tashiro, H. and Ito, H. [1999], 'Difference-frequency terahertz-wave generation from 4-dimethylamino-N-methyl-4-stilbazolium-tosylate by use of an electronically tuned Ti:sapphire laser', *Opt. Lett.* **24**(15), 1065–1067.

Kawase, K., Ogawa, Y., Watanabe, Y. and Inoue, H. [2003], 'Non-destructive terahertz imaging of illicit drugs using spectral fingerprints', *Opt. Express* **11**(20), 2549–2554.

Ke, S., Wen, X., Gurfinkel, M., Charnsangavej, C., Wallace, S., Sevick-Muraca, E. M. and Li, C. [2003], 'Near-infrared optical imaging of epidermal growth factor receptor in breast cancer xenografts', *Cancer Research* **63**(22), 7870–7875.

Kelly, P. and Winslow, E. [1996], 'Needle wire localization for nonpalpable breast lesions: sensations, anxiety levels, and informational needs', *Oncology nursing forum* **23**(4), 639645.

Kemp, M. C., Taday, P. F., Cole, B. E., Cluff, J. A., Fitzgerald, A. J. and Tribe, W. R. [2003], Security applications of terahertz technology, *in* 'Proc. SPIE', Vol. 5070, pp. 44–52.

Khan, U., Al-Moayed, N., Nguyen, N., Obol, M., Korolev, K., Afsar, M. and Naber, S. [2007], High frequency dielectric characteristics of tumorous and non-tumorous breast tissues, *in* '2007 IEEE/MTT-S International Microwave Symposium', pp. 1341–1344.

Kim, S. M., Hatami, F., Harris, J. S., Kurian, A. W., Ford, J., King, D., Scalari, G., Giovannini, M., Hoyler, N., Faist, J. and Harris, G. [2006], 'Biomedical terahertz imaging with a quantum cascade laser', *Applied Physics Letters* **88**(15), 153903.

Kindt, J. T. and Schmuttenmaer, C. A. [1996], 'Far-infrared dielectric properties of polar liquids probed by femtosecond terahertz pulse spectroscopy', *The Journal of Physical Chemistry* **100**(24), 10373–10379.

Klimberg, V. S. [2003], 'Advances in the diagnosis and excision of breast cancer', *The American surgeon* **69**(1), 11–14.

Klimberg, V., Westbrook, K. and Korourian, S. [1998], 'Use of touch preps for diagnosis and evaluation of surgical margins in breast cancer', *Annals of Surgical Oncology* **5**(3), 220–226.

Koenker, R. [1981], 'A note on studentizing a test for heteroscedasticity', *Journal of Econometrics* **17**(1), 107–112.

Kohavi, R. [1995], A study of cross-validation and bootstrap for accuracy estimation and model selection, *in* 'Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2', IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1137–1143.

Kreike, B., Hart, A. A., van de Velde, T., Borger, J., Peterse, H., Rutgers, E., Bartelink, H. and van de Vijver, M. J. [2008], 'Continuing risk of ipsilateral breast relapse after breast-conserving therapy at long-term follow-up', *International Journal of Radiation Oncology Biology Physics* **71**(4), 1014 – 1021.

Kunos, C., Latson, L., Overmoyer, B., Silverman, P., Shenk, R., Kinsella, T. and Lyons, J. [2006], 'Breast conservation surgery achieving ≤ 2 mm tumor-free margins results in decreased local-regional recurrence rates', *The Breast Journal* **12**(1), 28–36.

Lee, Y. S. [2008], *Principles of Terahertz Science and Technology*, New York: Springer, New York.

Lee, Y. S., Meade, T., Norris, T. B. and Galvanauskas, A. [2001], 'Tunable narrow-band terahertz generation from periodically poled lithium niobate', *Applied Physics Letters* **78**(23), 3583–3585.

Liebe, H. J., Hufford, G. A. and Manabe, T. [1991], 'A model for the complex permittivity of water at frequencies below 1 THz', *International Journal of Infrared and Millimeter Waves* **12**, 659–675.

Lilliefors, H. W. [1967], 'On the Kolmogorov-Smirnov test for normality with mean and variance unknown', *Journal of the American Statistical Association* **62**(318), 399–402.

Lovrics, P., Cornacchi, S., Vora, R., Goldsmith, C. and Kahnamoui, K. [2011], 'Systematic review of radioguided surgery for non-palpable breast cancer', *European Journal of Surgical Oncology (EJSO)* **37**(5), 388 – 397.

Lu, J.-Y., Chang, H.-H., Chen, L.-J., Tien, M.-C. and Sun, C.-K. [2005], 'Optoelectronic-based high-efficiency quasi-CW terahertz imaging', *Photonics Technology Letters, IEEE* **17**(11), 2406–2408.

Luini, A., Zurrida, S., Paganelli, G., Galimberti, V., Sacchini, V., Monti, S., Veronesi, P., Viale, G. and Veronesi, U. [1999], 'Comparison of radioguided excision with wire localization of occult breast lesions', *British Journal of Surgery* **86**(4), 522–525.

Luker, G. D. and Luker, K. E. [2008], 'Optical imaging: Current applications and future directions', *Journal of Nuclear Medicine* **49**(1), 1–4.

MacKay, D. J. [1995], 'Bayesian neural networks and density networks', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **354**(1), 73 – 80.

Mangasarian, O. L., Street, W. N. and Wolberg, W. H. [1995], 'Breast cancer diagnosis and prognosis via linear programming', *Operations Research* **43**(4), 570–577.

MATLAB [2012], *Curve Fitting Toolbox*, The MathWorks Inc.: Natick, MA, U.S.A.

McClatchey, K., Reiten, M. T. and Cheville, R. A. [2001], 'Time resolved synthetic aperture terahertz impulse imaging', *Applied Physics Letters* **79**(27), 4485–4487.

Miller, A. R., Brandao, G., Prihoda, T. J., Hill, C., Cruz, A. B. and Yeh, I.-T. [2004], 'Positive margins following surgical resection of breast carcinoma: Analysis of pathologic correlates', *Journal of Surgical Oncology* **86**(3), 134–140.

Mitrofanov, O., Brener, I., Harel, R., Wynn, J. D., Pfeiffer, L. N., West, K. W. and Federici, J. [2000], 'Terahertz near-field microscopy based on a collection mode detector', *Applied Physics Letters* **77**(22), 3496–3498.

Mittleman, D., Gupta, M., Neelamani, R., Baraniuk, R., Rudd, J. and Koch, M. [1999], 'Recent advances in terahertz imaging.', *Applied Physics B: Lasers & Optics* **68**(6), 1085.

Mittleman, D., Jacobsen, R. and Nuss, M. [1996], 'T-ray imaging', *Selected Topics in Quantum Electronics, IEEE Journal of* **2**(3), 679–692.

Mittleman, D. M., Hunsche, S., Boivin, L. and Nuss, M. C. [1997], 'T-ray tomography', *Opt. Lett.* **22**(12), 904–906.

Mogensen, M. and Jemec, G. B. E. [2007], 'Diagnosis of nonmelanoma skin cancer/keratinocyte carcinoma: A review of diagnostic accuracy of nonmelanoma skin cancer diagnostic tests and technologies', *Dermatologic Surgery* **33**(10), 1158–1174.

Naftaly, M. and Miles, R. E. [2007], 'Terahertz time-domain spectroscopy for material characterization', **95**(8), 1658–1665.

Nahata, A., Weling, A. S. and Heinz, T. F. [1996], 'A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling', *Applied Physics Letters* **69**(16), 2321–2323.

Nguyen, K. L., Johns, M. L., Gladden, L., Worrall, C. H., Alexander, P., Beere, H. E., Pepper, M., Ritchie, D. A., Alton, J., Barbieri, S. and Linfield, E. H. [2006], 'Three-dimensional imaging with a terahertz quantum cascade laser', *Opt. Express* **14**(6), 2123–2129.

Ntziachristos, V. [2006], 'Fluorescence molecular imaging', *Annual Review of Biomedical Engineering* **8**(1), 1–33.

Nuss, M. C., Mankiewich, P. M., O'Malley, M. L., Westerwick, E. H. and Littlewood, P. B. [1991], 'Dynamic conductivity and coherence peak in $YBa_2Cu_3O_7$ superconductors', *Phys. Rev. Lett.* **66**, 3305–3308.

O'Hara, J. and Grischkowsky, D. [2004], 'Quasi-optic synthetic phased-array terahertz imaging', *J. Opt. Soc. Am. B* **21**(6), 1178–1191.

Parrott, E. P. J., Sy, S. M. Y., Blu, T., Wallace, V. P. and Pickwell-MacPherson, E.

[2011], 'Terahertz pulsed imaging in vivo: measurements and processing methods', *Journal of Biomedical Optics* **16**(10), 106010.

Pedersen, J. and Keiding, S. [1992], 'THz time-domain spectroscopy of nonpolar liquids', *Quantum Electronics, IEEE Journal of* **28**(10), 2518–2522.

Pengel, K., Loo, C., Teertstra, H., Muller, S., Wesseling, J., Peterse, J., Bartelink, H., Rutgers, E. and Gilhuijs, K. [2009], 'The impact of preoperative MRI on breast-conserving surgery of invasive cancer: a comparative cohort study', *Breast Cancer Research and Treatment* **116**(1), 161–169.

Pickwell, E., Cole, B. E., Fitzgerald, A. J., Pepper, M. and Wallace, V. P. [2004], 'In vivo study of human skin using pulsed terahertz radiation', *Physics in Medicine and Biology* **49**(9), 1595–1607.

Pickwell, E., Cole, B. E., Fitzgerald, A. J., Wallace, V. P. and Pepper, M. [2004], 'Simulation of terahertz pulse propagation in biological systems', *Applied Physics Letters* **84**(12), 2190–2192.

Pickwell, E., Fitzgerald, A. J., Cole, B. E., Taday, P. F., Pye, R. J., Ha, T., Pepper, M. and Wallace, V. P. [2005], 'Simulating the response of terahertz radiation to basal cell carcinoma using ex vivo spectroscopy measurements', *Journal of Biomedical Optics* **10**(6), 064021.

Pickwell, E. and Wallace, V. P. [2006], 'Biomedical applications of terahertz technology', *Journal of Physics D: Applied Physics* **39**(17), R301.

Pickwell-MacPherson, E. and Wallace, V. P. [2009], 'Terahertz pulsed imaginga potential medical imaging modality?', *Photodiagnosis and Photodynamic Therapy* **6**(2), 128 – 134.

Pleijhuis, R., Graafland, M., de Vries, J., Bart, J., de Jong, J. and van Dam, G. [2009], 'Obtaining adequate surgical margins in breast-conserving therapy for patients

with early-stage breast cancer: Current modalities and future directions', *Annals of Surgical Oncology* **16**(10), 2717–2730.

Png, G. M., Choi, J. W., Ng, B. W.-H., Mickan, S. P., Abbott, D. and Zhang, X.-C. [2008], 'The impact of hydration changes in fresh bio-tissue on THz spectroscopic measurements', *Physics in Medicine and Biology* **53**(13), 3501.

Porter, S. G. [1981], 'A brief guide to pyroelectric detectors', *Ferroelectrics* **33**(1), 193–206.

Porterfield, D. [2007], High-efficiency terahertz frequency triplers, *in* '2007. IEEE/MTT-S International Microwave Symposium', pp. 337–340.

Porterfield, D., Crowe, T., Bradley, R. and Erickson, N. [1999], 'A high-power fixed-tuned millimeter-wave balanced frequency doubler', *Microwave Theory and Techniques, IEEE Transactions on* **47**(4), 419–425.

Preu, S., Dhler, G. H., Malzer, S., Wang, L. J. and Gossard, A. C. [2011], 'Tunable, continuous-wave terahertz photomixer sources and applications', *Journal of Applied Physics* **109**(6), 061301.

Pupeza, I., Wilk, R. and Koch, M. [2007], 'Highly accurate optical material parameter determination with THz time-domain spectroscopy', *Optics Express* **15**(7), 4335–4350.

Rajadhyaksha, M., Grossman, M., Esterowitz, D., Webb, R. H. and Anderson, R. R. [1995], 'In vivo confocal scanning laser microscopy of human skin: Melanin provides strong contrast', *Journal of Investigative Dermatology* **104**(3), 946–952.

Reid, C. B., Fitzgerald, A., Reese, G., Goldin, R., Tekkis, P., OKelly, P. S., Pickwell-MacPherson, E., Gibson, A. P. and Wallace, V. P. [2011], 'Terahertz pulsed imaging of freshly excised human colonic tissues', *Physics in Medicine and Biology* **56**(14), 4333.

Reid, C. B., Pickwell-MacPherson, E., Laufer, J. G., Gibson, A. P., Hebden, J. C. and Wallace, V. P. [2010], 'Accuracy and resolution of THz reflection spectroscopy for medical imaging', *Physics in Medicine and Biology* **55**(16), 4825.

Riedl, O., Fitzal, F., Mader, N., Dubsky, P., Rudas, M., Mittlboeck, M., Gnant, M. and Jakesz, R. [2009], 'Intraoperative frozen section analysis for breast-conserving therapy in 1016 patients with breast cancer', *European Journal of Surgical Oncology (EJSO)* **35**(3), 264–270.

Rigel, D. S., Russak, J. and Friedman, R. [2010], 'The evolution of melanoma diagnosis: 25 years beyond the ABCDs', *CA: A Cancer Journal for Clinicians* **60**(5), 301–316.

Ronne, C. and Keiding, S. R. [2002], 'Low frequency spectroscopy of liquid water using THz-time domain spectroscopy', *Journal of Molecular Liquids* **101**(13), 199 – 218. Molecular Liquids. Water at the New Millenium.

Rønne, C., Thrane, L., strand, P.-O., Wallqvist, A., Mikkelsen, K. V. and Keiding, S. R. [1997], 'Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation', *The Journal of Chemical Physics* **107**(14), 5319–5331.

Said, T. and Varadan, V. [2007], Modeling the effective complex permittivity of heterogeneous breast tissue and comparison with relaxation models at millimeters wave frequencies, *in* '2007 IEEE Region 5 Technical Conference', pp. 157–162.

Salomatina, E., Jiang, B., Novak, J. and Yaroslavsky, A. N. [2006], 'Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range.', *J. Biomed. Opt.* **11**, 064026.

Schade, U., Holldack, K., Kuske, P., Wstefeld, G. and Hbers, H.-W. [2004], 'THz near-field imaging employing synchrotron radiation', *Applied Physics Letters* **84**(8), 1422–1424.

Shen, Y. C., Lo, T., Taday, P. F., Cole, B. E., Tribe, W. R. and Kemp, M. C. [2005], 'Detection and identification of explosives using terahertz pulsed spectroscopic imaging', *Applied Physics Letters* **86**(24), 241116.

Shriner, D. L., McCoy, D. K., Goldberg, D. J. and Wagner, Richard F., J. [1998], 'Mohs micrographic surgery', *Journal of the American Academy of Dermatology* **39**, 79–97.

Siegel, P. H. [2002], 'Terahertz technology', *IEEE Transactions on Microwave Theory and Techniques* **50**(3), 910–928.

Smith, L. F., Rubio, I. T., Henry-Tillman, R., Korourian, S. and Klimberg, V. [2000], 'Intraoperative ultrasound-guided breast biopsy', *The American Journal of Surgery* **180**(6), 419 – 423.

Smith, P., Auston, D. and Nuss, M. [1988], 'Subpicosecond photoconducting dipole antennas', *Quantum Electronics, IEEE Journal of* **24**(2), 255–260.

Smye, S. W., Chamberlain, J. M., Fitzgerald, A. J. and Berry, E. [2001], 'The interaction between terahertz radiation and biological tissue', *Physics in Medicine and Biology* **46**(9), R101–R112.

SPSS [2011], *Version 20*, IBM Corp.

Stockfleth, E., Rosen, T. and Schumack, S. [2010], *Managing Skin Cancer*, Springer.

St.Peter, B., Yngvesson, S., Siqueira, P., Kelly, P., Khan, A., Glick, S. and Karellas, A. [2013], 'Development and testing of a single frequency terahertz imaging system for breast cancer detection', *IEEE Transactions on Terahertz Science and Technology* **3**(4), 374–386.

Strong, V., Humm, J., Russo, P., Jungbluth, A., Wong, W., Daghighian, F., Old, L., Fong, Y. and Larson, S. [2008], 'A novel method to localize antibody-targeted cancer deposits intraoperatively using handheld PET beta and gamma probes', *Surgical Endoscopy* **22**(2), 386–391.

Sturm, J. F. [1999], 'Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones', *Optimization Methods and Software* **11**(1), 625–653.

Sy, S., Huang, S., Wang, Y.-X. J., Yu, J., Ahuja, A. T., ting Zhang, Y. and Pickwell-MacPherson, E. [2010], 'Terahertz spectroscopy of liver cirrhosis: investigating the origin of contrast', *Physics in Medicine and Biology* **55**(24), 7587.

Taday, P. F. and Newnham, D. A. [2004], 'Technological advances in terahertz pulsed systems bring far-infrared spectroscopy into the spotlight', *Spectroscopy Europe* pp. 20–24.

Tafra, L., Fine, R., Whitworth, P., Berry, M., Woods, J., Ekbom, G., Gass, J., Beitsch, P., Dodge, D., Han, L., Potruch, T., Francescatti, D., Oetting, L., Smith, J. S., Snider, H., Kleban, D., Chagpar, A. and Akbari, S. [2006], 'Prospective randomized study comparing cryo-assisted and needle-wire localization of ultrasound-visible breast tumors', *The American Journal of Surgery* **192**(4), 462 – 470.

Tagaya, N., Yamazaki, R., Nakagawa, A., Abe, A., Hamada, K., Kubota, K. and Oyama, T. [2008], 'Intraoperative identification of sentinel lymph nodes by near-infrared fluorescence imaging in patients with breast cancer', *The American Journal of Surgery* **195**(6), 850 – 853.

Taylor, Z. D., Singh, R. S., Bennett, D. B., Tewari, P., Kealey, C. P., Bajwa, N., Culjat, M. O., Stojadinovic, A., Lee, H., Hubschman, J. P., Brown, E. R. and Grundfest, W. S. [2011], 'THz medical imaging: In vivo hydration sensing', *IEEE Transactions on Terahertz Science and Technology* **1**(1), 201–219.

Taylor, Z. D., Singh, R. S., Culjat, M. O., Suen, J. Y., Grundfest, W. S., Lee, H. and Brown, E. R. [2008], 'Reflective terahertz imaging of porcine skin burns', *Opt. Lett.* **33**(11), 1258–1260.

Tonouchi, M. [2007], 'Cutting-edge terahertz technology', *Nature Photonics* **1**, 97–105.

Truong, B. C. Q., Tuan, H. D., Fitzgerald, A. J., Wallace, V. P. and Nguyen, H. T. [2015], 'A dielectric model of human breast tissue in terahertz regime', *IEEE Transactions on Biomedical Engineering* **62**(2), 699–707.

Truong, B. C. Q., Tuan, H. D., Kha, H. H. and Nguyen, H. T. [2013], 'Debye parameter extraction for characterizing interaction of terahertz radiation with human skin tissue', *IEEE Transactions on Biomedical Engineering* **60**(6), 1528–1537.

Truong, B., Tuan, H., Fitzgerald, A., Wallace, V. and Nguyen, H. [2014], High correlation of double Debye model parameters in skin cancer detection, *in* '36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)', pp. 718–721.

Tuan, H. D. and Apkarian, P. [2000], 'Low nonconvexity-rank bilinear matrix inequalities: algorithm and applications in robust controller and structure design', *IEEE Trans. Automatic Control* **45**(11), 2111–2117.

Tuy, H. [1998], *Convex Analysis and Global Optimization*, Kluwer Academic.

Tuy, H. [2000], 'Monotonic optimization: Problems and solution approaches', *SIAM Journal on Optimization* **11**(2), 464–494.

van Esser, S., Hobbelink, M., Peeters, P., Buskens, E., van der Ploeg, I., Mali, W., Rinkes, I. and van Hillegersberg, R. [2008], 'The efficacy of 'radio guided occult lesion localization' (ROLL) versus 'wire-guided localization' (WGL) in breast conserving surgery for non-palpable breast cancer: A randomized clinical trial - ROLL study', *BMC Surgery* **8**(1), 9.

van Exter, M. and Grischkowsky, D. [1990], 'Characterization of an optoelectronic terahertz beam system', *Microwave Theory and Techniques, IEEE Transactions on* **38**(11), 1684–1691.

Verghese, S., McIntosh, K. and Brown, E. [1997], 'Highly tunable fiber-coupled photomixers with coherent terahertz output power', *Microwave Theory and Techniques, IEEE Transactions on* **45**(8), 1301–1309.

Vogt, M., Paul, B., Scharenberg, S., Scharenberg, R. and Ermert, H. [2003], Development of a high frequency ultrasound skin imaging system: Optimization utilizing time domain reflectometry and network analysis, *in* 'IEEE Symposium on Ultrasonics', Vol. 1, pp. 744–747.

Wahl, R. L. [2001], 'Current status of PET in breast cancer imaging, staging, and therapy', *Seminars in Roentgenology* **36**(3), 250 – 260.

Wallace, V., Fitzgerald, A., Shankar, S., Flanagan, N., Pye, R., Cluff, J. and Arnone, D. [2004], 'Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo', *British Journal of Dermatologyy* **151**(2), 424–432.

Wallace, V. P., Fitzgerald, A. J., Pickwell, E., Pye, R. J., Taday, P. F., Flanagan, N. and Ha, T. [2006], 'Terahertz pulsed spectroscopy of human basal cell carcinoma.', *Applied Spectroscopy* **60**(10), 1127–1133.

Wallace, V. P., Taday, P. F., Fitzgerald, A. J., Woodward, R. M., Cluff, J., Pye, R. J. and Arnone, D. D. [2004], 'Terahertz pulsed imaging and spectroscopy for biomedical and pharmaceutical applications', *Faraday Discussions* **126**, 255–263.

Wang, K., Barkan, A. and Mittleman, D. M. [2004], 'Propagation effects in apertureless near-field optical antennas', *Applied Physics Letters* **84**(2), 305–307.

Wang, K. and Mittleman, D. M. [2012], 'Metal wires of terahertz wave guilding', *Nature* **430**, 376–379.

Wang, S. and Zhang, X.-C. [2003], 'Tomographic imaging with a terahertz binary lens', *Applied Physics Letters* **82**(12), 1821–1823.

Williams, B. S. [2007], 'Terahertz quantum-cascade lasers', *Nature Photonics* **1**, 517–525.

Williams, L. S., Mancuso, A. A. and Mendenhall, W. M. [2001], 'Perineural spread of cutaneous squamous and basal cell carcinoma: CT and MR detection and its impact on patient management and prognosis', *International Journal of Radiation Oncology, Biology, Physics* **49**(4), 1061–1069.

Withayachumnankul, W., Fischer, B. M., Lin, H. and Abbott, D. [2008], 'Uncertainty in terahertz time-domain spectroscopy measurement', *J. Opt. Soc. Am. B* **25**(6), 1059–1072.

Woodward, R. M., Cole, B. E., Wallace, V. P., Pye, R. J., Arnone, D. D., Linfield, E. H. and Pepper, M. [2002], 'Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue', *Physics in Medicine and Biology* **47**(21), 3853–3863.

Woodward, R. M., Wallace, V. P., Pye, R. J., Cole, B. E., Arnone, D. D., Linfield, E. H. and Pepper, M. [2003], 'Terahertz pulse imaging of ex vivo basal cell carcinoma.', *The Journal of Investigative Dermatology* **120**(1), 72–78.

Wu, Q. and Zhang, X. C. [1995], 'Free-space electro-optic sampling of terahertz beams', *Applied Physics Letters* **67**(24), 3523–3525.

Yasui, T., Yasuda, T., ichi Sawanaka, K. and Araki, T. [2005], 'Terahertz paintmeter for noncontact monitoring of thickness and drying progress in paint film', *Appl. Opt.* **44**(32), 6849–6856.

Yu, C., Fan, S., Sun, Y. and Pickwell-MacPherson, E. [2012], 'The potential of terahertz imaging for cancer diagnosis: A review of investigations to date', *Quantitative Imaging in Medicine and Surgery* **2**(1), 33–45.

Zavagno, G., Goldin, E., Mencarelli, R., Capitanio, G., Bianco, P. D., Marconato, R., Mocellin, S., Marconato, G., Belardinelli, V., Marcon, F. and Nitti, D. [2008], 'Role

of resection margins in patients treated with breast conservation surgery', *Cancer* **112**(9), 1923–1931.

Zaytsev, K. I., Kudrin, K. G., Karasik, V. E., Reshetov, I. V. and Yurchenko, S. O. [2015], 'In vivo terahertz spectroscopy of pigmentary skin nevi: Pilot study of non-invasive early diagnosis of dysplasia', *Applied Physics Letters* **106**(5), 053702.

Zaytsev, K. I., Kudrin, K. G., Koroleva, S. A., Fokina, I. N., Volodarskaya, S. I., Novitskaya, E. V., Perov, A. N., Karasik, V. E. and Yurchenko, S. O. [2014], 'Medical diagnostics using terahertz pulsed spectroscopy', *Journal of Physics: Conference Series* **486**(1), 012014.

Zeitler, J. A., Taday, P. F., Newnham, D. A., Pepper, M., Gordon, K. C. and Rades, T. [2007], 'Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting - a review', *Journal of Pharmacy and Pharmacology* **59**(2), 209–223.

Zeni, O., Gallerano, G. P., Perrotta, A., Romano, M., Sannino, A., Sarti, M., DArienzo, M., Doria, A., Giovenale, E., Lai, A., Messina, G. and Scarfi, M. R. [2007], 'Cytogenetic observations in human peripheral blood leukocytes following in vitro exposure to THz radiation: a pilot study', *Health Physics* **92**, 349–357.

Zhang, X. C. and Xu, J. [2009], *Introduction to THz Wave Photonics*, New York: Springer, New York.