UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

School of Health Care Sciences

A proposal submitted in fulfilment of PhD requirements

in the

Department of Clinical Anatomy

A diagnostic algorithm for accurate identification of breast carcinoma on ultrasound

Development of a novel imaging software algorithm

Student no.: 24004708

Kathryn Malherbe

PhD: Clinical Anatomy

Supervisor: Prof. Albert van Schoor

Co-supervisor: Dr Mable Kekana

Researcher contact details

Cell: 0716732188

Email address: kathryn.malherbe@up.ac.za

Postal address: 590 Peacehaven Road, Rietvalleirand, 0181

1 August 2021

EXECUTIVE SUMMARY

Introduction

Breast cancer remains the most common form of cancer among women. Due to its high incidence, technology that improves detection rates needs to be developed. While routine mammography remains the gold standard for breast cancer detection, considerable research is being done to alter breast cancer detection methods and diagnostic processes. The increased interest in ultrasound as a diagnostic tool for breast cancer detection has led to rapid developments in the application of computer-aided diagnosis (CAD) for breast ultrasound.

Aim

This study aimed to develop and test targeted diagnostic segmentation algorithms utilising imaging software for the ultrasound-based diagnosis of suspicious mass lesions of the breast.

Materials and method

The first tier of the study was a retrospective, cross-sectional study with a population of 1000 ultrasound images. This included images from Parklane Women's Imaging Centre, performed between January 2017 and December 2018. All malignant lesions as detected on ultrasound and subsequently confirmed with histology as lobular carcinoma (LC) and ductal carcinoma (DC) were included.

The second tier of the study was a prospective case-controlled study with a population estimate of 100 ultrasound images of all suspicious mass lesions detected on ultrasound performed in a radiology department from April 2020 to March 2021. The breast ultrasound images used in the study were samples of diagnostic cases obtained during routine clinical care at the radiology department.

The final tier compared the histological output to the MIPAR^{TM} software image segmentation analysis to determine if a clinically valid algorithm had been developed.

Conclusion

The use of MIPAR^{TM} software for the segmentation and morphological assessment of breast cancer masses depicted on ultrasound is limited. The proposed research study promoted the further development of CAD software for breast ultrasound.

DEFINITION OF TERMS

Key term	Definition
Algorithm	An objective specification to solve a set of problems.
Bayesian theorem	Bayesian statistics is a mathematical procedure that applies probabilities to statistical problems.
CAD	Computer-aided detection software used during radiology imaging and diagnosis.
Carcinoma	Abnormal cell proliferation arising in the epithelial tissue of the internal organs.
Carcinoma in situ	A region of abnormal cell proliferation situated only within the original site of onset.
Ductal carcinoma in situ	Cancerous cells within the lining of the milk ducts with no extravasation or invasion to surrounding tissues or organs.
Lobular carcinoma in situ	Abnormal cell proliferation situated within the lobules of the breast with no evidence of surrounding invasion.
MIPAR™	Materials Imaging Processing and Automated Reconstruction.

TABLE OF CONTENTS

EXECUTIVE SUMMARY ..... **II**

KEY TERMS ..... **III**

DEFINITION OF TERMS ..... **III**

TABLE OF CONTENTS ..... **IV**

LIST OF FIGURES ..... **VIII**

LIST OF TABLES ..... **IX**

LIST OF ANNEXURES ..... **X**

LIST OF ABBREVIATIONS ..... **XI**

LAYOUT OF THESIS CHAPTERS ..... **XII**

CHAPTER 1 : INTRODUCTION **1**

1.1 Aim 3

1.2 Objectives 3

1.3 Impact of PhD output in the South African context 4

1.4 Limitations of the study 4

1.5 Conclusion 5

CHAPTER 2 : LITERATURE REVIEW **6**

2.1 Computer-aided diagnosis 6

2.1.1 History of computer-aided diagnosis 6

2.1.2 Mammography and ultrasound-based diagnosis 7

2.2 Segmentation techniques 8

2.2.1 Image segmentation for ultrasound 9

2.2.2 Image processing 11

2.2.2.1 Pre-processing 11

2.2.2.2 Segmentation 12

2.2.2.3 Classification 12

2.3 MIPAR™ imaging software 12

2.3.1 Background information 12
2.3.2 MIPAR™ software applications 13
2.3.2.1 Alignment 13
2.3.2.2 Pre-processing 13
2.3.2.3 Segmentation 13
2.3.2.4 Visualisation 14
2.3.2.5 Quantification 14
2.3.3 Future developments 14
2.4 Tumour microenvironment 14
2.4.1 Basic overview of the tumour microenvironment 15
2.4.1.1 Fibroblasts and tumour progression 16
2.4.1.2 Dendritic cells and the role of oestrogen receptors 17
2.4.1.3 Fibroblasts, dendritic cells and artificial intelligence 17
2.4.1.4 Macrophages, lymphocytes and the role of oestrogen receptors 18
2.4.1.5 Macrophages, lymphocytes and artificial intelligence 18
2.4.1.6 Extracellular matrix, mast cells and neutrophils 19
2.4.1.7 Mast cells, extracellular matrix, neutrophils and artificial intelligence 19
2.4.2 Breast cancer and the local microenvironment 20
2.4.2.1 Artificial intelligence and the local microenvironment 20
2.4.3 Breast cancer and the metastatic microenvironment 21
2.4.3.1 Artificial intelligence and the metastatic environment 22
2.4.4 Breast cancer and the infiltrated immune cell environment 22
2.4.4.1 Artificial intelligence and molecular alterations to the microenvironment 22
2.4.5 Artificial intelligence and beyond 23
2.5 Summary of literature review 24
CHAPTER 3 : MATERIALS AND METHODS 26
3.1 Study design 26
3.2 Study setting 26
3.3 Study population and sampling 26

3.4 Sampling method 27
3.5 Sample size 28
3.6 Data collection process 28
3.7 Measurement tools 30
3.8 Measurement methods and techniques 30
3.9 Variables 30
3.10 Quality control 31
3.11 Statistical analysis 33
3.12 Ethical and legal considerations 34
3.13 Conclusion 34
CHAPTER 4 : DISSEMINATION OF RESULTS 36
4.1 Introduction 36
4.2 Overall sample population 36
4.3 Retrospective data output 36
4.3.1 Objective 1 38
4.3.2 Objective 2 42
4.3.3 Objective 3 43
4.4 Prospective data output 45
4.4.1 Objective 4 45
4.4.2 Objective 5 47
4.4.3 Objective 6 47
4.5 Summary of results 49
CHAPTER 5 : DISCUSSION OF RESULTS 50
5.1 Introduction 50
5.2 Literature comparison to results 51
5.2.1 Tumour microenvironment 52
5.2.2 Diagnostic imaging analysis 53
5.3 Various diagnostic processes of breast ultrasound 54
5.3.1 Contrast-enhanced ultrasound comparison to the current study 55

5.3.2 Machine learning models compared to the current study....55
5.4 Approach to developing software in the South African market....58
5.5 Conclusion....60
CHAPTER 6 : STUDY CONCLUSION....61
6.1 Introduction....61
6.2 Conclusion of study....61
6.3 Limitations to results....62
6.4 Positive attributes of the study....62
6.5 Recommendations for future studies....62
REFERENCES....63
ANNEXURES....73

LIST OF FIGURES

Figure 2.1: Automatic segmentation of lobular carcinoma 9

Figure 2.2: Automatic segmentation of ductal carcinoma 9

Figure 2.3: Image processing of CAD software[25,41] 10

Figure 3.1: Segmentation process algorithm[25] 31

Figure 4.1: Accuracy rate and positive predictor value of the diagnostic algorithm 37

Figure 4.2: Clustering graph for greyscale means of malignant findings 42

Figure 4.3: Greyscale means of malignant and surrounding soft tissue 43

Figure 4.4: ROC curve for logistic regression models 44

Figure 4.5: Grading of malignancy and greyscale means 45

Figure 4.6: Cluster analysis of accuracy rate and positive predictor value for diagnostic algorithm 48

Figure 4.7: ROC for the prospective data sample 48

Figure 5.1: Standard (a) and new (b) proposed breast cancer screening methodology in healthcare 57

LIST OF TABLES

Table 3.1: Data set findings for the mock study sample 32

Table 4.1: Retrospective sample frequency, means and medians 38

Table 4.2: Overall benign sample population histopathological results 39

Table 4.3: Sample population of malignant findings 40

Table 4.4: Invasive malignant sample population 41

Table 4.5: Grading of ductal carcinoma in the sample population 41

Table 4.6: Grading score of carcinoma histopathological findings 41

Table 4.7: Sample population of prospective study 46

Table 4.8: McNemar's test of agreement for the prospective study sample 47

Table 4.9: Cohen's Kappa as measurement of agreement for prospective data 47

Table 4.10: Overall mean greyscale levels between ductal and lobular carcinoma 49

Table 5.1: A comparison study between various breast imaging modalities[130] 55

Table 5.2: Metrics for theory of change 60

LIST OF ANNEXURES

Annexure A: Permission to access data at Parklane Women's Imaging Centre 73

Annexure B: Declaration against plagiarism 74

Annexure C: Principal investigator's declaration for the storage of research data and/or documents 74

Annexure D: Declaration of Helsinki 74

Annexure E: Pictorial review of MIPAR™ imaging software interfaces 74

Annexure F: Statistician support letter 74

Annexure G: Non-disclosure agreement Parklane Radiology Department Letter of intent and clinical advisory support from Helen Joseph Hospital and Milpark Breast Centre of Excellence of Prof. Carol-Ann Benn 74

Annexure H: Ethics committee letters of approval 74

Annexure I: Funding and commercialisation of PhD research 74

Annexure J: Published articles related to current thesis research output 74

Annexure K: Proof of language editing of final thesis 119

LIST OF ABBREVIATIONS

AAV	Adeno-associated virus
AI	Artificial intelligence
ANN	Artificial neural networking
BI-RADS	Breast Imaging Reporting and Data System
CAD	Computer-aided diagnosis
CAF	Cancer-associated fibroblasts
CEUS	Contrast-enhanced ultrasound
DC	Ductal carcinoma
DCIS	Ductal carcinoma in situ
DML	Deep machine learning
ECM	Extracellular matrix
ER	Oestrogen receptor
LC	Lobular carcinoma
MATLAB	Materials Laboratory
MHz	Megahertz
MIPAR	Materials Imaging Processing and Automated Reconstruction
ML	Machine learning
PACS	Picture Archiving and Communication System
ROC	Receiver operator curve
SVM	Support vector machine
TM	Trademark
WHO	World Health Organisation

LAYOUT OF THESIS CHAPTERS

Chapter 1: Introduction

This chapter provides an overview of the study and introduces the basic literature regarding breast cancer diagnosis and prognosis methods in practice. The background and key concepts of certain terminology used during the research are explained.

Chapter 2: Literature review

This chapter provides literary support for the research question and aids in the assessment of previous methodologies used in similar research studies to ascertain the correct protocol to follow.

Chapter 3: Methodology

This chapter provides details to prove certain hypotheses of the research and to provide the basis of data capturing and analysis methods used.

Chapter 4: Results

This chapter presents the results so that the problem statement can be resolved.

Chapter 5: Discussion

This chapter includes a discussion of the relevant objectives and research questions that were postulated in Chapter 1 and integrates the information gathered in Chapter 2 for further analysis of the findings.

Chapter 6: Conclusion and recommendations

This chapter assesses all the data assessments that were gathered and suggests future research on similar topics.

CHAPTER 1: INTRODUCTION

Abraham Maslow once said, “If your tool is a hammer, then everything is a nail.”[1] This aphorism is known as the law of the instrument, so-termed in 1966. This rings true of the perception of sonography in the health sciences environment as it is a popular tool for physicians, radiologists, sonographers and clinical enthusiasts; however, in practice, this tool’s technical and clinical efficacy combined is limited.[1]

Breast cancer remains the most common global form of cancer among women[2]. The high incidence rate related to the disease has enabled the early detection of breast cancer using the latest technology in high-frequency ultrasound. While routine mammography remains the gold standard for breast cancer detection, considerable research is being done to improve breast cancer detection methods and the diagnostic process. The increased interest in ultrasound as a diagnostic tool for breast cancer detection has led to rapid developments in the application of CAD for breast ultrasound.[3]

In South Africa, the latest National Cancer Registry information[4-6] is limited to the 1994–2009 period. Projections for the incidence of cancer by 2030 suggest a global shift to low- and middle-income countries due to increased population growth as well as the adoption of a western lifestyle. Breast cancer is the fifth leading cause of cancer death; 60% occurs in low-income countries[4-5]. This is mainly due to late diagnosis and poor access to treatment facilities.[7] Due to the lack of African data and recent statistics, breast cancer registries for Africa are limited to Zimbabwe, Uganda, Malawi and South Africa (Eastern Cape).

South Africa is considered a middle-income country with a dual healthcare system (both public and private access) and multiple racial demographics.[8-10] Factors in South Africa associated with healthcare inequality are uniquely associated with the lingering effects of post-apartheid policies and the socioeconomic discrepancies among the various racial groups. From 2005 data, the lifetime incidence is estimated at 25/100 000 and 1 in 36 women.[8,11-12] Between 1994 and 2009, 85 000 women in South Africa were diagnosed with breast cancer of which the highest incidence rates were reported among black women, followed by white, coloured and Asian groups.[5-6,13] This compares to the population demographics: 79% black, 8% white, 8% coloured and 2% Asian.

In comparison to global trends, the United States of America (USA) has a lifetime incidence of 92/100 000 and France 89/100 000. It seems disproportionate but true. Epidemiological trends and limitations of pathology-based diagnosis, especially in rural South Africa, are the main causative factors.[5,14]

When compared to other countries, South Africa has a low mortality rate but this is due to the low quality of data in 2005; 20% of deaths were ill-defined and 70% were accurately recorded.[13] A limited number of hospitals provide surgery, chemotherapy and radiation therapy modalities due

to a lack of quality healthcare professionals, no provision of human resources, theatres and imaging equipment despite high patient volumes.[5]

Other important factors unique to the South African context are the lack of anti-cancer medications, poor palliative care facilities, chemotherapy drugs such as etoposide, dacarbazine and calcium folinate not being procured for the South African market and the limited supply of anthracyclines. Despite the current high rate of cancer incidence and mortality rates and the lack of infrastructure in South Africa, the government's focus is on increased supplies and infrastructure for the COVID-19 pandemic. This need should not displace oncologic treatment in South Africa.[6,13,15]

Due to the increased needs during the COVID-19 pandemic, oncology services have been scaled down.[6] Most deaths in South Africa occur during the reproductive years. This reality compromises the World Health Organisation's (WHO) development goals for economic stability, poverty reduction and healthcare equality. The major non-communicable disease (NCD) burden consists of cardiovascular disease, lung disease, diabetes Type 2, cancer and depression. Breast cancer is the second leading cause of death in NCDs.[5-6,13,15-16]

Resources in South Africa are mainly focused on the human immunodeficiency virus (HIV), tuberculosis (TB) and the management of other infectious diseases. The increased risk of hypertension and obesity in South Africa also contributes to a higher breast cancer risk. Conflicting reports on the lack of breast cancer education in South Africa occur and those populations with lower education rates are reported to have a lower risk for developing breast cancer but this may be attributed to the lack of early intervention and the availability of reliable statistics.[5]

A cohort study in Soweto[5,13] found that a fifth of the population lives with HIV and 14% of breast cancer patients had at least one metabolic disease and depression. This confirms the double burden of the epidemiological overlap of NCD and infectious diseases in South Africa.[5,13] The need for adequate diagnosis and intervention is imperative.

One such method is the Breast Imaging Reporting and Data System (BI-RADS®) ultrasound lexicon that aids in the characterisation of benign and malignant masses.[17] The prediction rate of breast ultrasound in the detection of suspicious masses or solid mass lesions has been extensively researched and reported in the literature as ranging from 68%–96%.[18] Ultrasound imaging plays an integral role in the detection of suspicious mass findings in correlation to mammography and BI-RADS® scores.[4] Upon discovery of a suspicious mass, an ultrasound-guided biopsy is performed to aid in the surgical decision-making process and its use is associated with a low mortality and morbidity rate.[19-20]

Yet to be developed are non-invasive diagnostic methods to avoid the need for biopsy tissue sampling. There is a clear advantage to avoiding invasive biopsy, considering the economic, psychological and physical aspects related to excision core biopsy in affected women. When

considering ultrasound or any other non-invasive diagnostic tools as the means of differentiating between benign and malignant findings, it is important to consider the false-negative of each diagnostic method. Women should be well-informed about the probabilities and be given full knowledge of the possible diagnostic (such as possible false-negative rate) and post-care implications related to excision core biopsy.[21]

The increased use of ultrasound in breast cancer diagnosis has led to rapid development in the application of Computer Aided Detection (CAD) to breast ultrasound. Previous literature assessed the need to standardise the process and found that combining image information using averaging calculations aids in improving diagnostic performance in medical imaging.[8,22-23]

None of the literature disseminated above examined and analysed the predictive validity of retrospective ultrasound images employing a unique segmentation algorithm. This study has developed one such unique algorithm for use in prospective image analysis, such as CAD. The outcome will potentially lead to improved prediction validity through targeted lesion segmentation and classification using standardised algorithms.

1.1 Aim

This study aimed to develop and test targeted diagnostic segmentation algorithms utilising imaging software for the ultrasound-based diagnosis of suspicious mass lesions of the breast.

1.2 Objectives

Retrospective study objectives:

- Collection of retrospective ultrasound images (for the first-tier population study) for morphologic imaging analysis of B-mode ultrasound images associated with breast malignancy.
- Processing of retrospective ultrasound images (first-tier population sample) with Materials Image Processing and Automated Reconstruction (MIPAR^{TM}) software to develop an algorithm for segmentation of mass lesions through semi-automatic processing.
- Correlation between algorithm findings (using MIPAR^{TM} software) and the reported BI-RADS-US lexicon reporting output (of radiology practices) of retrospective ultrasound images to assess validity in assessment output and specificity and sensitivity of the algorithm in comparison to pathology-based diagnosis.
- Comparison of intensity mean (greyscale levels) of segmented images to assess differences between various types of mass lesions.

Prospective study objectives:

- Application of developed algorithm from first-tier study output to all suspicious mass lesions of the breast on ultrasound during prospective research.
- Comparison of algorithm prediction validity to the final histological assessment of suspicious mass lesions

1.3 Impact of PhD output in the South African context

The algorithm developed for this study has led to local government interest and stakeholder engagement to embed the algorithm into a working prototype for web-based software use. It will also be developed into a mobile application for clinical practice. The current market software tools assist radiologists by reducing the false-negative rate, detecting early-stage cancer and reducing the mortality rates for cancer.

This PhD thesis strives to improve the current diagnostic dilemma by employing a unique software algorithm able to identify, segment and predict breast cancer subtypes. This would allow on-site diagnosis without the need for core needle biopsy infrastructure on site, especially in rural clinics. During the algorithm's initial development, preliminary prototype testing has shown an 85%–95% accuracy rate with empirical data used for training a machine learning (ML) system as part of the software to be developed.

The market focus in South Africa is women's health imaging and the improvement of diagnostic detection rates. The current extent of breast cancer disease in South Africa is vast and remains the highest incidence of all cancers currently reported by the National Health Report of the Cancer Association of South Africa (CANSA).[6] The use of widely accessible, cost-effective software, such as the proposed PhD software, in rural settings may lead to the vast improvement of detection rates.

According to the global market forecast in PR Newswire, the current ultrasound market is estimated to reach R8.4 billion by 2023.[24] By 2030, 70% of the world's cancer burden will be in South Africa, with breast cancer the most commonly diagnosed. South Africa is committed to the United Nation's sustainable development goals. The Department of Health is currently developing a policy on diagnosing and managing breast cancer in South Africa.[5,13,15-16] Further development and market impact of the proposed software is discussed as an annexure to this thesis (see Annexure I).

1.4 Limitations of the study

A limitation of the study is the use of a selected radiology department that may cause intra-rater variability in comparison to other breast imaging centres with different protocols during the diagnosis of suspicious mass lesions.

The study is limited to the findings of invasive ductal carcinoma (DC) and lobular carcinoma (LC) ultrasound images. Other pre-malignant findings (in situ or atypical hyperplasia) are not included in this study.

An issue that can interfere with the image segmentation of data sets is a milling artefact known as curtaining.[25] It receives its name from surface modulations resulting from incomplete or excessive removal of material. Curtaining can be detrimental to image segmentation; it can however be avoided to a certain extent by proper sample preparation and milling parameter selection. However, in certain instances of sample geometries and microstructures, curtaining is unavoidable and post-acquisition filters are the only recourse.[25] MIPAR™ allows a graphical environment for image reconstruction, applying frequency domain filters (FFT) which include low pass, high pass, annular and custom filters.[25]

1.5 Conclusion

This chapter demonstrated the challenges of breast cancer diagnosis within the South African context and the need for accurate, cost-effective diagnostic software in clinical practice. This research was aimed at developing a unique algorithm able to identify, segment and predict breast cancer subtypes for the improvement of the current diagnostic dilemma that the South African population is facing.

The following chapter will focus on reviewing studies that have utilised similar or different approaches to assessing breast cancer morphology on ultrasound to determine the best methodology for use in this study.

CHAPTER 2: LITERATURE REVIEW

This chapter includes the history of computer algorithms and studies with study outcomes similar to those of the current research. The chapter also aims to support the research question, aim and methodology stated in subsequent chapters.

2.1 Computer-aided diagnosis

To improve detection and accuracy rates in radiology imaging departments, software engineers developed CAD systems – a software technique employed in computers to aid in the interpretation of images. It is an interdisciplinary method using both digital and radiological image processing techniques.[8,26-27]

CAD detects tumours on an ultrasound image by improving the image contrast, detecting lesions and reducing workload. A basic CAD system consists of three main processes: image acquisition, image pre-processing and image segmentation. Image pre-processing is further subdivided into noise reduction and contrast enhancement. The basic processing modules used are image denoising, image enhancement and image segmentation.[19]

2.1.1 History of computer-aided diagnosis

The first attempt at computerising medical images,[26] which is, to date, a major research topic in medical imaging with recent research delving into the artificial intelligence (AI) era for the medical field, occurred during the 1960s.[22,28-29] CAD serves as a diagnostic aid to support the role of the physician through the use of non-invasive and accurate computer systems.[26] CAD incorporates the quantitative analysis of images during the diagnostic process, proven from previous studies to increase the sensitivity of diagnosis by 21.2% and reduce the false-negative rate of diagnostic screening by 77%. Despite this figure, automated detection software is not widely used during breast screening.[30] A prospective study using CAD software during diagnosis has shown a 74% increase in cancer detection.[30] Certain technical advances in breast imaging – such as tissue harmonic imaging, compound imaging and extended field of view – have made its use integral during breast cancer diagnosis. Standardised CAD techniques used in conjunction with ultrasounds reduce the inter-observer variation.[29]

The detection rate of invasive cancers measuring less than 1 cm increases with the use of CAD systems. In conjunction with dedicated breast imagers, it has the potential to reduce false-negative rates from 31% to 19%.[11,31] The system assigns various sensitivity and specificity rates to cancers based on the lesion type. With CAD, the sensitivity for malignant calcifications is 86%–99%, with only 57% marked as amorphous calcifications.[32] The sensitivity for masses is estimated at 43%–

85%.[33] Further research is required to recognise suspicious asymmetries as they develop over time during serial imaging follow-up, as well as assessing the medico-legal implications of retained CAD-marked image information.

2.1.2 Mammography and ultrasound-based diagnosis

Despite the high incidence and mortality rates of breast cancer, the cause is still unknown. Currently, there is no effective way to prevent the occurrence of breast cancer. From this claim, it can be said that early detection plays a key role in breast cancer diagnosis and treatment.

Mammography remains the gold standard for diagnosing breast disease but still has certain technical disadvantages. These drawbacks include factors such as dense breast tissue and increased false-positive rates due to unnecessary biopsies for uncertain mass lesions depicted on images.[11,29] To aid in reducing these limitations, ultrasound was developed to provide adjunct information.[30,34]

Accurate diagnosis is dependent on sufficient training, experience and time constraints. Approximately 10% of all breast cancers are missed during radiological diagnosis. Among the reasons behind this dilemma are the reduced conspicuity of mammogram lesions, image noise (random variation in information of image due to brightness) as well as overlying or underlying breast anatomy obscuring the region of interest.[9]

The current study arose from the main researcher's interest regarding the unique morphometric properties of LC on ultrasound that was peaked during her study for her Master's degree where it was found that LC has a diffuse growth pattern, which relates to its low detection rate by mammography.[35] Infiltration of the breast tissue is diffuse and in a single row pattern of malignant cells, causing no destruction of the underlying normal breast tissue or reactive connective tissue. Thus, early-stage and even late-stage LC is rarely detected on mammograms.[36] Ultrasound plays an integral role in the improved sensitivity and diagnosis of LC.[20]

Ultrasound produces high-frequency sound waves, measured in megahertz (MHz), which 'bounce' from internal soft tissue structures at varying degrees of reflection to produce a digital image. This method can determine real-time imaging for biopsy guidance, giving it prevalence and efficacy within medical imaging technology.

Ultrasound has the advantage of non-reduced image quality from dense type breast tissue. It also boasts other modern-day uses such as zero radiation exposure and no need for a compression technique. All sections of breast tissue can be observed during dynamic imaging, enabling visualisation of small, early-stage cancers (hyperplasia).[30] Ultrasound is known for its ease of use, increased portability and cost-efficacy in low socioeconomic regions.[8,30] Ultrasound plays a significant role in the detection of breast cancers through its detection abilities of otherwise obscured mass lesions.[8,37] Limitations, such as low resolution, low contrast, speckle noise and

image blur are also reported. Ultrasound is wholly operator-dependent and thus repeatability in studies is limited. It can be said that digital image processing and pattern recognition software could alter operator competency in this department.

The concept of double-reading screening in breast ultrasound has received increased interest in the past few years. Double reading refers to two or more independent radiologists interpreting the same group of images. Discordant readings are assigned for further assessment, either by discussion among interpreting radiologists (consensus) or by an independent third radiologist who makes the decision (arbitration).[19-20,29]

The term 'double reading' is subclassified as a human-human and human-computer algorithm.[3,8,27,38] A *human-human double-reading system* uses a second radiologist during screening procedures to increase the perception of mass lesions. The latter can also aid in the subsequent interpretation thereof. The person-power differs based on independence/consensus among single to multiple radiologists working in various departments.

The second type is known as a *human-CAD double-reading system* and aids in the perception of suspicious mass lesions. It has proven to prospectively increase cancer detection. This latter method is useful for radiologists with less experience but, for breast specialists, it does not yield any interest.[19,39] The basic principle is that an algorithm allows the segmentation of a mass lesion from surrounding soft tissue. Non-spiculated masses or ill-defined lesions are likely to be marked with less than half of the architectural distortion cases detected.[3,26]

2.2 Segmentation techniques

A basic ultrasound image consists of multiple pixels which are a rectangular tiling of fundamental elements (pixels). A picture element is a small block that represents the amount of grey intensity to be displayed for that portion of an image. Most images have a pixel value or integer which ranges from 0 (very black) to 255 (very white)[25,27,40]. This basic integer information is used to allow the segmentation process to commence. The process of image segmentation is simplified in relevant software through the auto-segmentation of an image and can provide robust and fast automatic contouring of a breast lesion. This method can be used as a reliable second opinion for a radiologist in the department.

A segmentation software package such as MIPAR™ offers a simple yet powerful means of locally quantifying a variety of metrics at each point on a reconstructed surface. The metrics range from basic functions, such as positioning and curvature, to complex metrics, such as local thickness and surface roughness.[25]

Figures 2.1 and 2.2 display a representation of basic ultrasound images, consisting of basic pixels to which a segmentation process was applied for automatic contouring using the MIPAR™ imaging software processes. Figure 2.1 is a representation of LC with accurate segmentation and

separation from the surrounding soft tissue. Figure 2.2 is a representation of accurate segmentation from the surrounding soft tissue for DC. It is evident from these illustrations that there are definite morphological differences between DC and LC in an ultrasound.[41]

Figure 2.1: Automatic segmentation of lobular carcinoma

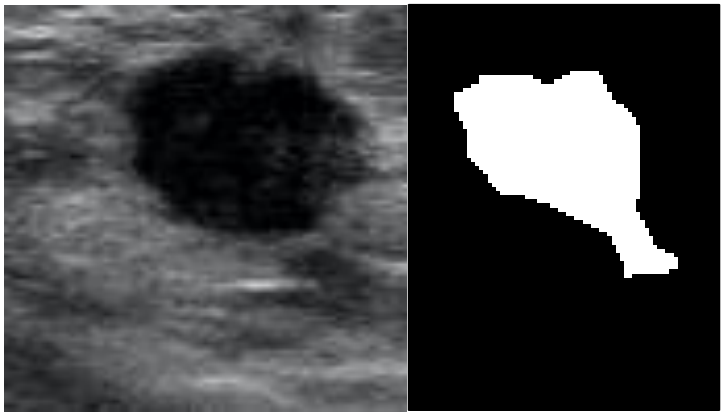


Figure 2.2: Automatic segmentation of ductal carcinoma

2.2.1 Image segmentation for ultrasound

Breast ultrasound techniques and the diagnosis of suspicious mass lesions are highly dependent on operator competency and a high inter-observer variation has been reported.[2,42] The process of image segmentation enables interpreters to assess the probability of a suspicious mass lesion being malignant, enhancing the performance competency of less-seasoned radiologists or sonographers. Figure 2.3 depicts the four basic steps for processing an image using CAD software.[8,43]

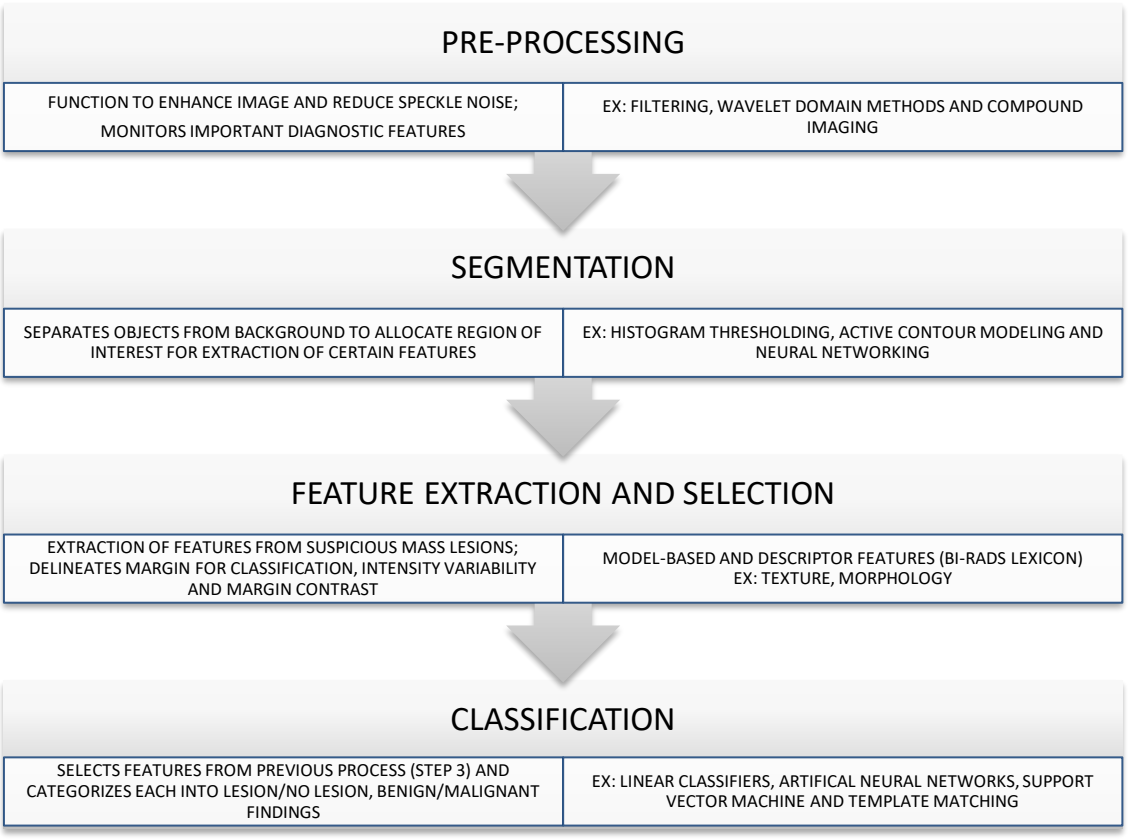


Figure 2.3: Image processing of CAD software[25,41]

The best practice for early cancer detection remains early detection, prompt diagnosis and effective treatment regimes. Breast cancer is treatable, depending on the onset (prognosis), locoregional onset and systemic treatment (predicting therapy response) used.[44] The treatment strategy depends on efficacy (prognosis), security (sensitivity) and node status.[26,44]

With breast cancer, many factors need to be accounted for to ensure accurate classification (diagnosis) and forecasting (prognosis). AI processing includes the element of evolutionary behaviour, much like biological systems with high-level connectivism between distributing processing elements. There are two main types of AI: artificial neural networking (ANN) and evolutionary computing.

ANN is the process of nonlinear mapping between set inputs to outputs.[43,45] It achieves biological performance using dense processing elements such as biological neurons. ANN can learn and generalise from examples given.[43,45] Success is measured if the relationship between variables is governed by complex linear functions.

Evolutionary computing consists of a collection of algorithms based on the evolution of the population towards the solution of a problem. It is subdivided into genetic algorithms, genetic programming and evolutionary algorithms. Successful use is measured using selecting features for the classification of mammogram calcifications.[8,45-46] The quality of features cannot be categorised due to the quality of a feature being dependent on its contribution to detection, classification and prognosis, as well as its pre-processing steps and classification measures.[8,45-

[47] There are various features, such as geometric features, which refer to factors such as size and shape.

The boundary is the starting point of extracting an object using AI. Various boundary methods such as binary sets – sets of pixels in a greyscale image with edge detection that defines an object by its edges – are used. Other geometric features include area, volume, contrast, counting pixels inside an object boundary, perimeters, and shape (no single shape descriptor can be used on its own to define an object).[8,47] Texture features refer to matrices and greyscale level means. Morphology refers to a collection of processing methods. Snakes refer to a model for an object boundary assumed to be continuous and wavelet theory provides a method to choose a pair of linear filters to analyse an image signal.[25,40]

2.2.2 Image processing

In ultrasound imaging, the intensity of an image relates to the pixel number associated with various elements in the imaging field. The intensity in diagnostic imaging can also be described as the state of quality being intense – a measure of strength.[9,27,43] For a data set of various pixel numbers or elements, the arithmetic mean is the central value of a discrete set of numbers. It is calculated as the sum of values divided by the number of values. The intensity mean is a representation of the average sum of pixel intensity numbers for a specific ultrasound image.

An algorithm or recipe which is applied to an ultrasound image is an unambiguous specification of how to solve a class of problems such as data processing or automated reasoning tasks.[8] During the application of MIPAR™ imaging software, batch processing of multiple images allows for various data sets to be extracted from the processed images, allowing assessment of the standard intensity mean and standard deviation and allows comparison to the original/standard image.[25]

2.2.2.1 Pre-processing

The most common form of pre-processing is the modification of greyscale images through greyscale histogram equalisation. The main purpose thereof is to change the intensity values of pixels to distribute intensity more uniformly. Thus, the distribution of grey levels is depicted more uniformly. Local histogram modification enhances image contrast in certain regions of interest. The image can, however, result in distortion of the area of interest due to the monotonous mapping of the greyscale values. Due to the known low contrast of ultrasound images, this technique is important to improve overall image contrast.[25,41,48]

2.2.2.2 Segmentation

The second step in CAD is the separation of suspicious mass lesions from the background parenchyma. Through segmentation, the ultrasound image is partitioned into several non-intersecting regions, extracted regions of interest and suspicious mass findings. Suspicious mass findings, with varying sizes and outlines, are typically darker than the surrounding soft tissue. Improved efficacy in this important step is dependent on the algorithm used for the user's specific needs. The current study developed novel segmentation algorithms to differentiate between DC and LC on ultrasound using morphology detection tools.[25,41,48]

2.2.2.3 Classification

ANN is a common tool used in CAD systems.[45] It consists of a multi-layer preceptor (MLP) neural network capable of approximating the relationship between input features and desired classes. Training data for specific benign versus malignant features are used to train the ANN. Classifiers based on Bayesian neural networks (BNN) are applied to classify the breast nodule. A decision tree consists of a decision node, branches and leaves; each branch leads to the next decision node. Decision trees are used to predict categorical variables. SVMs are another effective method used for pattern recognition, ML and data capture.[8,20,49]

2.3 MIPAR™ imaging software

The need for high-resolution post-processing techniques has resulted in the development of a robust software package capable of various post-acquisition image analysis tasks. MIPAR™ was developed by MATLAB™ but is used as stand-alone software using the power of processing algorithms from MATLAB™.[25] Three-dimensional characterisation software requires tools subject to various data set parameters. Few software packages have the capability represented by extensive toolsets such as alignment, pre-processing, segmentation, visualisation and quantification. In conclusion, MIPAR™ was written and developed in the MATLAB™ environment to allow MIPAR™ to be executed as a stand-alone software package on Macintosh, Windows and Linux platforms with its main focus on modular construction.[25]

2.3.1 Background information

Characterisation of images consists of five modules, of which an image processor, a batch processor and a 3D toolbox are used during image integration. Each module consists of five stages of image processing: alignment, pre-processing (raw data enhancement), segmentation,

visualisation and quantification. The image processor module consists of various contrast-enhancing and noise reducing filters. Many software programs have been developed for single 2D or 3D analysis; however, few software programs equip users with the tools for both 2D and 3D image characterisation. MIPAR™ is one of these few.[25]

2.3.2 MIPAR™ software applications

The image processor aids in developing a sequence of processing steps, known as a recipe. The goal of recipe formulation is to segment raw intensity pixels into a set of features. The process and parameter selections are performed in real-time, allowing for process removal and editing. The batch processor module is the second module that allows for the application of the recipe to a series of images for 2D analysis or to build a 3D structure.[25,41] Following 2D images, batch processing results open directly in the processed image editor, allowing for an accelerated review of a recipe's segmentation quality across all processed images. The image characterisation sequence is as follows.

2.3.2.1 Alignment

A process by which similar images are aligned with each other to maximise agreement of their spatial intensity distribution. Following the acquisition of data images, the alignment process is twofold: image transformation and similarity quantification. The most common pairing technique is cross-correlation that performs rigid image transformations to maximise the normal correlation coefficient of the two images.[41]

2.3.2.2 Pre-processing

A process includes all steps which alter the raw pixel intensities on a greyscale spectrum to ultimately improve the segmentation process. Pre-processing steps include factors such as brightness, contrast enhancement and noise reduction filters.[25]

2.3.2.3 Segmentation

This process is defined as the separation of data into fragmentary regions. It is the most crucial step in quantitative data analysis. Segmentation consists of various phases; the most important of these is phase segmentation whereby pixels are labelled in phases of various subtypes. Each of these subtypes is then assigned to a specific morphology. Multiple algorithms can be applied to assist the segmentation process.[25]

2.3.2.4 Visualisation

Visualisation is the ultimate purpose in image characterisation to maximise the characterisation potential of each image.[25]

2.3.2.5 Quantification

Quantification is the final process and its efficacy depends on the metric, corresponding algorithms and accuracy of the data's segmentation.[35] Quantification tools are subdivided into global, feature-by-feature and localised. Global quantification is the extraction of a single quantity such as a volume fraction or surface area density. Feature-by-feature quantification is the extraction of a specific metric – such as volume, surface area or diameter – from each feature of interest. Localised quantification is performed at each patient or vertex on a reconstructed surface.[43]

2.3.3 Future developments

MIPAR™ offers a simple yet powerful means of locally quantifying a variety of metrics at each point on a reconstructed surface. The metrics range from basic functions such as positioning and curvature to more complex functions such as local thickness and surface roughness. The software has made great strides in the field of aerospace materials.[41]

A critical gap in knowledge has been noted in breast cancer detection, prognosis and evaluation between tumour microenvironment and associated neoplasm.

AI has multiple subsets or methods for data extraction and evaluation, including ANN, which allows computational foundations, similar to neurons, to make connections and new neural pathways during data set training. Deep ML (DML) and AI hold great potential to accurately assess the tumour microenvironment models employing vast data management techniques. [24-25,50-54] [24-25,50-54] [24-25,50-54]

Despite the significant potential AI holds, there is still much debate surrounding the appropriate and ethical curation of medical data from Picture Archiving and Communication Systems (PACS). AI output's clinical significance holds its outcome based on its human predecessor's data training sets. Integration between biomarkers, risk factors and imaging data will allow the best predictor models for patient-based outcomes.[35-36,55-56]

2.4 Tumour microenvironment

The tumour microenvironment or surrounding stroma contains various vital components such as immune cells and extracellular matrices (ECMs) which act against antitumor immune cells. This

leads to tumour progression and ultimately metastasis.[57-60] The stromal environment contains many interesting signalling pathways and molecular structures related to the prognostic outcomes of breast cancer.[59]

The genetic alterations of cancer cells related to signalling pathways control both the processes of tumorigenesis and progression. These alterations are due to the overexpression of oncogenic mutations such as growth factor receptor tyrosine kinases and nuclear receptors such as oestrogen receptors (ERs). Due to the above complexities related to cancer signalling networks and inordinate signalling pathways, translating to pathway reactivation, the efforts to produce anti-cancer drugs are challenging. However, individual pathways, such as Ras-ERK, are strongly related to cancer mutations and promise targeted therapies in the future.[61]

The latest studies are now focusing on the tumour microenvironment as a critical element for determining tumour development, progression and treatment response.[53,58,62-63] In the same research interest groups, AI has multiple subsets or methods for data extraction and evaluation.[64-66] One such method is ANN[53,62,67-69] which allows computational foundations, similar to neurons, to make connections and new neural pathways during data set training (see Figure 2.1). A method used for quantitative biology is massive parallel reporter assay (MPRA), which assesses deoxyribonucleic acid (DNA).[52] This allows biologists the ability to predict molecular interactions and gene interactions. The mechanistic framework of gene regulation allows the possibility of new therapies to be developed.[24,57-58,63] There is a lack of congruence between biologists and ANN systems; the latest custom ANNs allow mathematical assumptions of common biological concepts so that the output is relayed as how biologists would interpret results.[54,62,67-69]

2.4.1 Basic overview of the tumour microenvironment

The breast cancer microenvironment can be subdivided into three main subsets: local, regional and distant. Each of these subsets contains cellular content such as fibroblasts, leukocytes, ECM, cytokines, growth factors and hormones,[57] described in detail below as the various cell type subsets related to breast cancer diagnosis and prognosis.

Dvorak stated that “tumours are much more than wounds that do not heal”.[70] Tumour cells undergo significant changes causing release from regulatory signals and promoting proliferation and invasion. The most crucial factor thereof is the overexpression of the Vascular Endothelial Growth Factor (VEGF), allowing surrounding stroma to be incorporated in its progression process.[70]

The use of AI technology to improve diagnostic detection rates and remote disease monitoring can reduce the overall time required for comprehensive patient treatment planning. Anti-VEGF agents and AI-generated prognosis have been studied using vision loss and the results could promote the prevention of vision loss before its occurrence.[71]

The angiogenesis process includes a complex interplay between tumour, endothelial and stromal cells, promoting tumour growth. A study in 2006[67] found a novel method of assessing angiogenesis employing chick embryos and their chorioallantoic membrane. An automated image analysis method able to quantify the micro vessel density and growth potential in images was developed. This shows the potential to be used for tumour growth detection in breast cancer imaging.[54,58,67,70,72] However, it lacks efficacy for extensive tumour series analysis of the tumour microenvironment. Other methods proposed for tumour microenvironment composition analysis are Gene Set Enrichment Analysis (GSEA) (San Diego), xCell (California) and TIminer (Russia), which allow immunogenic analysis and quantification of the immune infiltrate.[53,58-59,63,69,73-74]

GSEA[75] is a computational method able to define concordant differences between two biological states as a statistical output (see Figure 2.2). xCell[76] is a novel signature-based method used for 64 immune and stromal cell types. Utilising in silico analyses and cross-comparison to cytometry immunophenotyping, xCell shows excellent promise when compared to other methods.[76] TIminer[77] is a computational pipeline used for the assessment of tumour-to-immune cell interactions based on sequencing data.

In 2016, another computational method[78] reported a Microenvironment Cell Populations (MCP) counter that analyses the transcriptomic markers in single-cell populations; this method is robust compared to other utilised samples.[58]

The discussion below states the current body of knowledge and attributes of each factor/cell/protein related to breast cancer and the tumour microenvironment and is followed by the latest technology and insights related to AI and DML.

2.4.1.1 Fibroblasts and tumour progression

Most cells within the tumour microenvironment are fibroblasts which secrete various soluble factors modulating tumour stroma, growth and invasion properties. It was previously found[57,72] that cancer-associated fibroblasts (CAFs) have unique protein expression profiles. A bi-directional signalling pathway[59] has also been suggested between these unique fibroblasts and their adjacent cancer cells, suggesting a possible influence in the transcription of breast cancer cell profiles. That CAFs enhance tumour angiogenesis was also affirmed by Orimo et al.[72] The origin of these individual cells has been suggested as either bone marrow, normal fibroblasts and even epithelial-mesenchymal transition processes.[57]

The microcellular environment is maintained by fibroblasts using the remodelling of ECM.[59,70] Fibroblasts, associated with carcinoma, have unique characteristics that promote tumour progression, presenting as either heterogeneous or myofibroblasts with fibroblast activation protein. The potential of carcinoma-associated fibroblasts promoting tumour growth using secreted stromal-derived factor-1 (SDF-1), acting as a paracrine activator that increases tumour

cell proliferation through CXCR4, exists.[72] An interesting finding was a co-culture of fibroblasts in healthy breast tissue 'educating' fibroblasts to secrete hepatocyte growth factor (HGF) to promote tumour progression activities.[57,72]

The main question that arises from these studies is: from where are these CAFs derived? One hypothesis is that healthy fibroblasts undergo phenotype modification from constant aberrant signalling from adjacent tumour cells.[79]

2.4.1.2 Dendritic cells and the role of oestrogen receptors

Dendritic cells play an essential role in prohibiting neoplastic cell growth by presenting antigens to CD4+, CD8+ and T cells.[57-60,63,67,70,72-73] The maturation process of dendritic cells depends on the local microenvironment that determines its tolerance of immunosuppression of localised neoplastic invasion. In previous studies, surrounding tumour-associated stroma has shown dendritic cells with an inability to stimulate antitumor immunity. These so-called tumour-associated dendritic cells produce proangiogenic factors, enhancing endothelial cell migration and causing tumour progression.[54,58,67,70,72]

Dendritic cells have multiple roles in essential processes such as immunity, autoimmunity and differentiation of T cells. They are mainly activated by stress response or pathogen-induced damage, which causes the secretion of cytokines stimulating T lymphocytes and immune response. Oestrogen receptors play a crucial role in dendritic cell function.[57-60] When the dendritic cell ligands bind to ERs, migration processes are triggered. Recent studies have shown that treatment of E2 alongside mature dendritic cells and T cells could stimulate T-cell proliferation.[54,63]

2.4.1.3 Fibroblasts, dendritic cells and artificial intelligence

Fibroblasts play a role in mortality prediction of idiopathic pulmonary fibrosis (IPF). The use of AI to quantify prognostic histological features was studied and it was found that interstitial mononuclear inflammation and intra-alveolar macrophages proved as novel biomarkers in detecting IPF.[80] A group of researchers at Osaka University[81] developed an AI-based system to identify various cancer cells utilising microscopic images. A convolutional neural network was trained with 8 000 images of cells obtained from a phased contrast microscope. Following the data set training process, another 2 000 images were tested to distinguish mouse cancer cells from human cells and radioresistant cells from radiosensitive ones. This study holds much promise in developing a universal system able to identify and distinguish between all variants of cancer cells.[81]

Another researcher group based in Boston at the Tufts Medical Centre[82], developed multiple AI tools to detect and track dendritic cells. An AI algorithm was developed using in vivo confocal

microscopy (IVCM) analysis of the human cornea, typically done manually, making it a time-consuming process. The use of such AI models for analysis ensures high accuracy and reduced objectivity associated with human analysis.[82]

2.4.1.4 Macrophages, lymphocytes and the role of oestrogen receptors

Macrophages associated with tumour cells display unique phenotyping, promoting tumour growth, angiogenesis and tissue remodelling.[52-53,57-60,62-63,67-70]

The immune response includes the key role of macrophages to promote T-cell recruitment and activation. Their collaborative activation alongside T and B cells is due to cytokines and chemokines being released.[57,67,73] Despite their functional role in tumour defence, they are actively present in the tumour microenvironment leading to tumour progression and immunosuppression. Many articles report ER present in macrophage precursor cells during various stages of their differentiation process. E2 treatment has shown to change macrophage behaviour.[24,51-54,57-60,62,67-69,72-73]

A key player in recent research in the tumour microenvironment has been lymphocytes.[69] Lymphocytes are mostly T cells, CD4+ helper cells, Treg with CD4+ and CD 8+ cells.

Treg cells in the tumour microenvironment block its normal antitumor function and suppress other immune cells such as CD8+ T cells. Treg cells also produce a large amount of receptor activator of nuclear factor κβ ligand (RANKL) which promotes metastasis and RANK-expressing neoplastic cells[72,83]. A high concentration of Treg cells is associated with advanced type breast cancer. This is postulated as neoplastic cells recruiting Treg using prostaglandin E2 secretion and suppression effector cells, producing an immunosuppressive microenvironment.[52,57,59,68,70]

2.4.1.5 Macrophages, lymphocytes and artificial intelligence

Machine Learning (ML) can distinguish various cell and tissue types in a biopsy specimen based on a training set of ‘ground truth’ examples. A research study in 2018[84] made use of ML algorithms as a method to identify macrophages from digital scans of non-small cell lung carcinoma tissue slides. The study compared pathologist output to an ML algorithm and displayed improved accuracy compared to human-reader intervention and output.[84]

Working group collaboration with the Massive Analysis and Quality Control Consortium[83] works on ML algorithms to characterise tumour-infiltrating lymphocytes. Such methods will enhance the validity of prognostic prediction methods in pathology. Besides the clinically evident improved prediction rates of ML, it also permits changes to the current feature set used for ML analysis, thus improving accuracy and interpretation in current standard methods.[85]

2.4.1.6 Extracellular matrix, mast cells and neutrophils

The main proteins within the complex ECM are collagen (structural), fibronectin (glycoproteins) and chondroitin sulphate (proteoglycans). Recent studies have shown that ECM[59] is more versatile than initially thought, acting as a critical player in cell growth, proliferation and migration.

In cancer, ECM is typically disorganised in appearance, causing abnormal feedback regulatory mechanisms. This is mainly due to the metabolism of ECM being altered by CAFs and immune cells.[53,58-59,63,69,73] One of the main proteins within ECM, namely collagen, promotes cancer cell invasion using collagen IV degradation. ECM also promotes the passage of cytokines and growth factors, enabling intercellular communication. The alteration in protein activity as seen in cancer is associated with patient outcomes.

Mast cells (MCs) form part of the immune system and are associated with parasitic infections. Depending on the type of inflammatory stimulus, MCs release various inflammatory mediators. Mucosal MCs produce tryptase, whereas connective tissue MCs secrete tryptase, chymase and carboxypeptidases. All these enzymes, along with IL-8, TGF-β and TNF-α, have a strong association with angiogenesis and matrix metalloproteinases (MMP) modulation of various breast cancer phenotypes.[53,67-68]

ERα is present in MCs. The treatment of E2 (prostaglandins of the Estrogen series), has shown, in rat MC models, a release of histamine. This process discovery is exciting because histamine release plays a role in breast cancer promotion utilising its H3R and H4R receptors.[54,63]

Neutrophils are a fundamental component of the immune response, acting as a first-line defence mechanism against infection employing phagocytosis. Neutrophils work alongside other immune-fighting cells such as macrophages and dendritic cells.[53,58-59,63,69,73]

Neutrophils are known to have nuclear receptors and E2 and ER binding help regulate neutrophil survival and function. Several serine proteases, such as neutrophil elastase (NE), proteinase 3 (PR3) and cathepsin G (CG), essential for infectious agent elimination and inflammation modulation are secreted by neutrophils.[57,70]

2.4.1.7 Mast cells, extracellular matrix, neutrophils and artificial intelligence

One of the critical elements of MC granules is histamine as it has been shown to promote tumour cell proliferation and growth of mammary carcinomas through H2 receptors.[86] A study attempting to assess ML functional genomic networks discovered histamine hypersensitivity in response to a local inflammatory response, which begs the question of its underlying molecular and genetic traits and how ML could promote its prognostic indicator role in tumour progression.[87] Many research studies coin MCs as the most misunderstood cell type during breast cancer proliferation and

immune response since their discovery 140 years ago, making them a key focus of future research endeavours.[88]

Advances in 3D cell tissue engineering have led to the development of 'cancer on a chip' platforms that allow the tumour microenvironment model to produce improved analytic outputs, especially for discovering the role of ECM during tumour progression. The possibility of integrating AI for improved drug screening models is made possible through these chip platforms.[83,89]

Microscopy has reached the age of digitisation with outputs such as CellaVision, which classifies degenerated lymphocytes and web-like remnants. Ikemura[90] *et al* hypothesise that these remnants are Neutrophil Extracellular Traps (NETs). They aim to develop an AI platform able to detect NETs rapidly on blood smears.[90]

The use of computational models to screen endocrine-active compounds holds much promise as a cost-effective alternate method in practice. An ML algorithm[55] was applied to over 7500 compounds related to nuclear oestrogen receptor (ERα and ERβ) activity. The model's performance was evaluated using receiver-operating curve values obtained from fivefold cross-validation procedures that proved values ranging from 0.56–0.86.[54,91]

The following sections elaborate on the surrounding environments related to the breast cancer and tumour microenvironment. Each section will discuss the current trends and research and the latest AI technology being developed.

2.4.2 Breast cancer and the local microenvironment

Normal mammary gland development relies on appropriate cross-talk between epithelial and stromal cells, inhibiting abnormal cell growth and neoplasm formation. Myoepithelial cells have previously been known for their tumour suppression capabilities as they produce a base membrane barrier around luminal epithelial cells. The loss of such myoepithelial cells would promote in situ carcinoma to invasive type carcinoma.[52,57-58,60,72]

Two models have been suggested to explain this carcinoma invasion. The 'escape model' suggests genetic changes of tumour epithelial cells, allowing the invasion to adjacent ducts. The 'release model' suggests that the tumour microenvironment disrupts the basement membrane, allowing tumour cells to spread into the stroma. Both of these models prove the importance of both epithelial and stromal components in tumour progression.[73]

2.4.2.1 Artificial intelligence and the local microenvironment

In the last decade, many approaches have been used to quantify the non-cancerous cell populations acquired from tumour samples using computational algorithms with different statistical frameworks and data sets. The two most common algorithms used for tumour microenvironment

estimation are regression-based deconvolution algorithms and gene set enrichment methods.[58,83] The algorithms are dependent on the pre-acquired knowledge of data sets for accurate measurement, a statistical framework and a pre-determined signature for each cell type. The regression-based deconvolution algorithm determines the gene expression profile ratio in the total tumour expression profile. Gene set enrichment assigns scores to the various cell types as a function of its expression in each gene set.[47]

A recent study of the University of Eastern Finland[54] developed an AI model capable of predicting breast cancer risk based on demographic risk factors and genetic variants. The method used for the AI model is a gradient tree with adaptive iterative searching methods. The gene interaction map included ESR1 an FGFR1 genes[24,52-54,57-59,63,67-70,72,92] linked to oestrogen receptor subtype breast cancer. Since cancer incidence is a multifactorial process, the use of AI in predicting breast cancer risk through this novel method holds much promise for future disease incidence.[54,62,68]

2.4.3 Breast cancer and the metastatic microenvironment

During the complex metastasis process, tumour cells either have a dormant state or an active state to form micrometastases. During primary tumour recruitment, the cytokines select associated bone marrow cells to incur a premetastatic process before tumour mobilisation occurs. It has been noted that fibroblasts and cancer cells travel alongside one another during the metastatic process. Breast cancer cells promote RANKL through the active secretion of cytokines and growth factors.[52,54,57,59,63,70,72] This activates osteoclast formation and bone resorption. RANKL has more recently been noted in the formation of lung metastases, thus providing a hypothesis of specific immune cells participating in the formation of metastases.

A fundamental attribution to tumour progression and drug resistance is the tumour microenvironment. This implies that various ill-controlled cells all relate to cancer progression. This concept has been around since the 1880s when Steven Paget suggested the ‘seed and soil’ concept[93] where fertile soil (the tumour microenvironment) and the seed (cancer cells) work in harmony to promote growth.

The ‘seed planted in the soil’ concept confirms that malignant disease remains the foundation of a tumour progressing whereas the tumour microenvironment facilitates the invasion ability of these cancer cells. For this reason, research currently focuses on epithelial-mesenchymal transition (EMT), where specific mediators allow the progression of tumour cells to invasive type lesions. Examples of these mediators are IL-1, IL-6 and IL-8 and they allow tumour cell proliferation with EMT increasing their ability to metastasise.[52,57,69,72]

Both intrinsic and extrinsic inflammatory pathways promote an inflammatory microenvironment. Tumour cells promote inflammatory mediators that lead to the progression of cancer within the microenvironment using T cells, NK cells, macrophages and dendritic cells.[70]

2.4.3.1 Artificial intelligence and the metastatic environment

Recent AI insights allow assessing molecular subtypes and their therapeutic response utilising predictive image analysis of breast cancer phenotypes. In a research study of the TCGA Breast Phenotype Group,[53,68] multidisciplinary researchers phenotypically characterised 84 solid breast tumours to gain insights into underlying molecular characteristics and gene expression profiles. Significant similarities were noted between enhancement texture (entropy) and molecular subtypes (normal-like, luminal A, luminal B, HER2-enriched, basal-like) even after controlling for tumour size (P =.04 for lesions ≤2 cm; P =.02 for lesions from 2cm to ≤5 cm). [67-68,70,73,93]

Regarding treatment outcomes, a semi-manual delineation method of tumour volume using breast magnetic resonance imaging (MRI) proved a high prediction anomaly for low recurrence rates in patients, proving the potential for digital automation in its prediction outcomes.

2.4.4 Breast cancer and the infiltrated immune cell environment

Significant gene expression changes occur within myoepithelial cells, confirming a change during tumour progression in the microenvironment. An example of the overexpression of genes is chemokine CXCL14 binding to CXCR4 and promoting the proliferation and migration of tumour cells[57]. Others also confirmed that changes in the stroma and gene expression occur most frequently when healthy breast tissue transitions to DC in situ (DCIS).[57,72,92]

Since breast cancer is a heterogeneous disease, it has three main phenotypes: luminal, human epidermal growth factor receptor type and triple-negative type. Since breast cancer promotes an inflammatory microenvironment, immune filtration is presently based on ER presence.[57,63,72,94] A substantial proportion of natural killer cells and neutrophils within ER-positive breast cancer and cytotoxic and TCD4+ cells in smaller amounts exist. The presence of eosinophils, monocytes and B lymphocytes proves a good prognosis following chemotherapy.[53,58,73,93]

2.4.4.1 Artificial intelligence and molecular alterations to the microenvironment

Despite the significant potential AI holds, much debate surrounding the appropriate and ethical curation of medical data from PACS still exists. The clinical significance of AI output still holds its outcome based on its human predecessor's data training sets.[85]

The integration between biomarkers, risk factors and imaging data will allow the best predictor models for patient-based outcomes.[53,57,95]

State of the art research has found an ML approach, named CytoReason (version 1.0)[50], distinguishing between nivolumab responders and non-responders. Since adipocytes are

postulated to be involved in the tumour microenvironment, this study also showed evidence of their regulatory role in ipilimumab resistant nivolumab patients. The study requires extensive research on the role of adipocytes in tumour progression, leading to new immunotherapy methods. CytoReason[50] integrates genetics, proteomics, cytometry and literature with ML to help create disease models.

A key focus on T-cell subsets related to cancer immunology and therapy is imperative as a predictor of such subsets could promote advances in immunology research. A research group in China developed a method, Immune Cell Abundance Identifier (ImmuCellAI), that allows gene set signature-based algorithms to estimate the abundance of 24 immune cell types from gene expression data.[96] However, the method has limitations, such as measuring the abundance of cells being limited to the deviation from gene signatures. The method also did not include the spatiotemporal attributes of the immune cells.[96]

A wild-type adeno-associated virus (AAV) particle capsid[55,97] is currently the most used gene therapy method due to its established ability to deliver gene material to organs. However, a few naturally derived AAV capsids are deficient in the essential components required for gene therapy.

A group of researchers at Harvard developed ML technology to engineer new, improved capsids for therapeutic use.[68,98] Starting at Harvard in 2015, the authors set out to overcome the limitations of current capsids by developing new machine-guided technologies to engineer a suite of new, improved capsids rapidly and systematically for widespread therapeutic use to outperform AAVs generated by conventional random mutagenesis approaches. This demonstrates a powerful tool for sizeable broad-scale DNA synthesis and iterative machine-guided design to develop improved synthetic AAV capsids.[97,99]

The current and latest trends in local, metastatic and microenvironments hold much potential for breast cancer research and AI technology in the future. The next section discusses the most recent research and trials undertaken during 2020.

2.4.5 Artificial intelligence and beyond

A recent 2020 study in Italy[45,69] focused on predicting the disease and establishing a therapeutic plan and patient-focused follow-up sessions. In this regard, a multidisciplinary approach was encouraged during the development of the Multigene Signature Panels and Nottingham Prognostic Index. ML allows the cross-correlation of prognostic indicators to determine possible markers related to patient outcomes. Two ML methods, namely ANN and Support Vector Systems using SPSS IBM Modeller 18.1 software, were employed. Their accuracy, sensitivity and specificity were measured as 95.29%–96.86%, 0.35–0.64 and 0.97–0.99, respectively.[45] The study was limited to a select study population without long-term recurrence following 20 years of remission in breast cancer patients.

Breast cancer comprises a complex genetic background and an understanding of the intricate relationship between these cancer cells and surrounding stromal/immune cells is essential to ensure adequate treatment methods are implemented. In vitro cell culture systems lack dedicated physiological outputs during drug testing.[58,72] Mouse models are the ideal animal model for assessing drug tolerance, however, they are limited in testing the tumour microenvironment of humans. Various models have been proposed for tumour microenvironment studies, where the latest in vitro 3D models can study both cell-cell and cell-material interactions parametrically.[58,72]

The use of stromal-to-epithelial yield using the spatial extraction of features is also a novel approach to assessing disease progression. These studies allow further insights into the role of epithelial and stromal cells and an alternate tiered approach to DML.[52,59-60,67,69-70,72-74,78,93]

The main solution all therapists, pathologists and clinicians require is an improved prognosis method for breast cancer. Shimizu and Nakayama[100] developed a complete atlas of prognostic breast cancer genes, a computational framework and a prediction score applicable to all breast cancer subsets. The method is unique in its stratification of patients at the clinical stage and ER-negative subtypes.[100]

2.5 Summary of literature review

The chapter outlined the literature on how the increased use of ultrasound in breast cancer diagnosis has led to rapid developments in the application of CAD. Previous literature assessed the need to standardise the process and found that combining image information using averaging calculation aids in improving diagnostic competency. A large gap in the literature regarding the use of MIPAR™ software for segmentation, morphological assessment and accurate determinants related to breast cancer masses depicted on ultrasound exists.[25]

The use of tissue engineering[101] in cancer research allows an accurate representation of the tumour microenvironment in human studies.[53,60,68] Since there is currently vast recognition of the tumour microenvironment in tumour progression, it is now the focus of therapeutic research. New strategies to normalise the surrounding stroma, modulate the immune system and promote antitumor activity enhancement are evident. The critical role of E2 and its signalling pathways requires more research on the use of intertumoral therapy as part of an adjuvant therapy approach to immune response.

Despite some limitations in mouse models, the data support the role of the tumour microenvironment in the treatment of breast cancer.[52,54,57,59-60,63,68,94,102]

The various tumour microenvironment elements and their latest research endeavours using AI and DML show that improved prognostic and therapeutic methods are imperative. The role of unique cancer signalling pathways, targeted therapies and novel diagnostic trends will boast

significant strides when conjoined with AI in practice. The development of a comprehensive prognostic cancer gene mutation atlas will be a step into the future for pathologists, even more so as a multidisciplinary approach for developing the Multigene Signature Panels and Nottingham Prognostic Index.

Although most studies are experimental and undergoing clinical trials, the possibility of integrating such methods in clinical practice is inevitable in the future. The current study examined and analysed the validity and utility of the use of segmentation software for the analysis and accurate identification of breast carcinoma detected on ultrasound images.

The following chapter presents the methodology used to prove the assumptions stated in this chapter.

CHAPTER 3: MATERIALS AND METHODS

3.1 Study design

This study implemented a two-tiered methodology. The first tier was a retrospective quantitative study with a cross-sectional study design. The second tier was a prospective case-controlled study.

3.2 Study setting

This study was conducted in academic collaboration with the Helen Joseph Breast Care and Milpark Breast Centre of Excellence for the exclusivity of clinical advice, as per agreement with Prof. Carol-Ann Benn, who signed a letter of intent and mutual understanding to support the research project during its course (see Annexure G). The data acquisition centre which serves as a dedicated breast centre, boasting American College of Radiology accreditation, is Parklane Women's Imaging Centre with a sample population average of 1000 biopsies per annum. Signed non-disclosure agreements to uphold every patient's anonymity and privacy were obtained from all parties involved in the study (see Annexures A and G). The researcher of the current study also confirms a statement of plagiarism along with the Ethical standards of Helsinki were upheld during the course of the study's research data collection(See Annexure B & D).

The data acquired for ultrasound imaging analysis during the study were all breast ultrasounds performed by radiologists with over 30 years of experience at the Parklane Women's Imaging Centre. The radiologists at Parklane are responsible for performing and interpreting the high-frequency ultrasound examinations of the breast on the Toshiba Aplio (Tecmed 2012) with a high-resolution 13 MHz frequency probe. The Parklane department performs a minimum of 200 breast ultrasound studies per week and 1000 ultrasound-guided biopsies per annum. A standard breast ultrasound imaging protocol is used at the department for the evaluation of malignant type masses. The imaging protocol includes the use of standardised callipers for size measurements as well as the ACR and BI-RADS® method for radiological findings. Hard copies of patient files are kept at the department, and it has a single PACS database where all patient data are stored electronically.

3.3 Study population and sampling

The overall study population included the records of all ultrasound images of patients who were referred to the radiology department for routine or diagnostic screening.

First-tier population

The size of the study population included all 1000 retrospective ultrasound images recorded from January 2017 to December 2018 and positively diagnosed with LC and DC via a follow-up biopsy. The ultrasound images included in the study used radiologic investigation and were confirmed by histopathological tests as malignant findings. The unit of analysis for the study was the high-frequency ultrasound images of the patients with confirmed LC and DC.

No special ultrasound techniques – such as colour Doppler, elastography or volumetric ultrasound (3-dimensional imaging) – were considered for the primary aim of the study. The purpose of the study was to design an algorithm for breast lesion detection as a common tool to be used in the department. This will be most valuable in low socioeconomic regions lacking infrastructure and experience. The main objective of the first-tier population study was the morphologic analysis of B-mode ultrasound images associated with LC and DC.

Second-tier population

The second-tier population included all prospective ultrasound images which were flagged on ultrasound reports as suspicious mass lesions from January 2020 to April 2021. The population sample included all participants who were referred to the department/s for routine annual screening. Appropriate participant informed consent was obtained (see Annexure A). Those patients who were subsequently referred for ultrasound core or excision biopsy for histological assessment were included in the sample population. The interpreter was blind to patient identifiers such as name, age and socioeconomic factors. The interpreter was only privy to the patient accession number for the population sampling process. The radiology department has a paper-based and electronic register of all patients referred for ultrasound biopsies and exclusive access to the associated ultrasound images related to these procedures.

3.4 Sampling method

For both the first- and second-tier sample population, the sampling method consisted of a consecutive sampling of all histologically positive malignancies in the breast, diagnosed through ultrasound-guided core biopsies. The two time periods were from January 2017 to December 2018 for the retrospective analysis and from April 2021 to March 2022 for the prospective analysis. Of the identified cases, those related specifically to malignant histology findings were included for study purposes until the sample size was reached.[16,35]

The *retrospective* data set consisted of 1000 high-frequency ultrasound images during B-mode ultrasound scanning with an associated radiologic report, as well as the pathology results of the included patients. Sampling included the collection of filed retrospective hard copy histology reports which correlated to the ultrasound images of a positive diagnosis case of breast cancer. These images were retrieved from PACS and each data set was assigned a unique study number.

The *prospective* data set consisted of 100 high-frequency ultrasound images during B-mode ultrasound scanning that were reported as suspicious mass lesions and from which a subsequent core biopsy was utilised for histological assessment. The researcher was blind to the final histological assessment during the prospective sampling process to ascertain whether the algorithm provided a true reflection of the diagnosed histological outcome.

The interpreter of the ultrasound images had extensive experience in the analysis and interpretation of high-frequency ultrasound images. The interpreter was blind to the patient's age, sex and other socioeconomic risk factors. Interpreter recruitment was based on convenience and consideration of their diagnostic experience in the field of high-frequency breast ultrasound. The interpreter used the standard reporting technique based on the American College of Radiology's standards.[4,17]

3.5 Sample size

A biostatistician from the South African Medical Research Council (SAMRC) conducted a power calculation to determine the most appropriate sample size: N = 1000 ultrasound images were suggested for retrospective analysis and N = 100 for prospective analysis. Based on previous epidemiological studies conducted in this field of study, sample sizes of 1000 and 100, respectively, were deemed appropriate to obtain a 95% confidence interval with a 5% margin of error.[103-104]

Of the 100 images used for the prospective study, a random sample of each main breast cancer type namely LC and DC were used to test the prediction validity of the developed algorithm. Statistically, based on current population incidence rates, DC is more common than LC, thus it was expected that a ratio of 70:30 would be used for validation purposes.[23,105]

3.6 Data collection process

An information session was held at the radiology department regarding the purpose of the study and to orientate the potential interpreter to the data collection process and address any questions or concerns that might arise during the study.

First-tier data collection process

High-frequency ultrasound images and histology reports of women who were positively diagnosed with LC and DC from January 2017 to December 2018 were extracted from the radiology database and PACS by the researcher to form a single data set. Patient identifiers on the ultrasound images and histology reports were removed and only a unique study number was allocated. The original reports of ultrasound images were not considered during the study.

The interpreter was blind to the other corresponding information regarding patient age, histological findings and ultrasound BI-RADS® classification. The interpreter was requested to upload ultrasound images of the included study population with MIPAR^{TM} software to produce a semi-automatic algorithm that segments the tumour from surrounding soft tissue. The interpreter analysed all 1000 retrospective ultrasound images independently in weekly sittings of 20 patients per session so that the full analysis took approximately five to six weeks. A further in-depth description of the MIPAR^{TM} software's methodology was required as it was not a well-known software tool in the field of diagnostic imaging. A pictorial review of the interfaces is included in Annexure E.

The first module used with MIPAR^{TM} is the image processor to develop the initial sequence of processing known as the recipe. While there are various recipes to perform for different image outcomes, all have the common goal of segmenting a greyscale intensity into a binary image. The user's processing steps are recorded in real-time, allowing parameter editing and process removal/insertion. The real-time capability of MIPAR^{TM} to optimise image parameters is the most useful feature during the recipe steps.

Batch processing follows the completion of the initial recipe so that it can be saved and loaded into the second processing module. The batch processing module allows the automatic application of the initial recipe to a series of images of all cases in a total study group database. Tilt correction, image alignment and volume reconstruction can be performed. This step is depicted as the second pictorial interface in Annexure I[25] A 3D toolbox can also be utilised, along with feature set analysis (see Annexure E for further demonstration of the software capabilities).

Second-tier data collection process

A standard algorithm was derived from the first-tier data collection process that served as a method of tumour/mass lesion segmentation from the surrounding soft tissue on the ultrasound images provided. Whether LC and DC would produce the same algorithm during the initial segmentation process was uncertain. It was hypothesised that the algorithm would be different for each and that two separate algorithms would have to be available for the second-tier data collection process.

Standard care of the department would follow, thus, the data collection process for the prospective study did not contravene any standard protocol at the department. During the initial imaging of each patient, an ultrasound image with tumour/mass lesion evidence was processed through the MIPAR^{TM} software. The processing of a selected ultrasound image ran alongside subsequent core biopsy/surgical excision of the mass lesion with relevant histopathological findings. Patient identifiers on the ultrasound images and histology reports were removed and each was allocated a unique study number.

The outcome of the MIPAR^{TM} software algorithm applied to the mass lesion was compared to final histopathological results to assess whether any similarity was observed.

3.7 Measurement tools

The first part of the data collection tool to collect histological results as well as ultrasound images associated with malignant type findings was completed by the researcher. The images where suspicious mass findings were reported were allocated a study number to ensure anonymity. The ultrasound images which were confirmed through retrospective histopathological findings as LC or DC would be included in the study sample.

The images were converted to tagged image file format (.tiff) for processing purposes. MIPAR^{TM} software would be used to process the ultrasound images. The images would be processed with the segmentation recipe/algorithm to delineate the tumour/mass. Measurements such as intensity mean, pixel intensity and shape comparison were loaded using the MIPAR^{TM} software to allow the comparison of various types of mass lesions.

3.8 Measurement methods and techniques

Ultrasound images were retrieved from a local PACS and were analysed on a five-megapixel Barco workstation. All ultrasound images were interpreted according to the guidelines of the BI-RADS® lexicon. The interpreter was unaware of the histology results, ultrasound BI-RADS® classification and previous clinical history.

Tumours or suspicious mass lesions were identified according to the BI-RADS-US®, a lexicon for suspicious lesions. The lexicon includes all descriptors in the radiology reports such as speculated, irregular, ill-defined borders, hypoechoic, posterior acoustic shadowing, micro lobulated, asymmetric density and architectural distortion. Two unique algorithms for LC and DC, respectively, were utilised during the segmentation process of the collected data to ensure an accurate mechanism of measurement for the various breast cancer types.

3.9 Variables

Dependent variables included ultrasound morphology descriptors for suspicious mass findings for invasive LC and DC. The independent variables included the histopathological results. The variables of the proposed algorithm included the size and shape of the mass lesion, intensity mean and pixel value. Confounding variables included the possibility of inter-rater variability with the image analysis performed on the ultrasound images. The confounding variables of the proposed algorithm included the semi-automated segmentation process which reduces reproducibility.

3.10 Quality control

Validity refers to the extent to which a concept is accurately measured in a quantitative study.[10] This study is valid in its direct measure of morphometric properties of malignant ultrasound findings. The instrument is homogenous and the measure is a specific construct and has validity through convergence to other studies that used a similar measuring method.[34]

Reliability refers to the consistency of measure in a study.[106-107] The radiologists performing ultrasound imaging related to this study used the same output method and instrument tool during image analysis. The validity and reliability of this method were strengthened by previous studies done by Chappellier[108] and Constantini et al.[56] The interpreter was provided with the same data collection sheet for analysis to reduce inter-rater variability.[10,25] A small mock study sample was tested during the development of the research protocol.

The sample population included 12 randomly selected suspicious mass lesions reported by ultrasound. A senior application specialist from MIPAR™ assisted with the basic recipe (algorithm) which promised utility for image segmentation of a tumour within an ultrasound image.

The basic concept is demonstrated in Figure 3.1 which is reproduced from the work of Sosa et al.[25] The actual steps as outlined in this flow diagram were undertaken in this study.

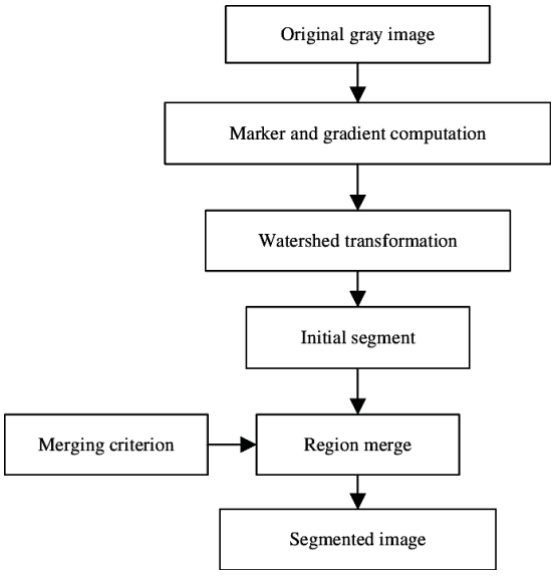


Figure 3.1: Segmentation process algorithm (adapted) [25]

Six of the images were confirmed through retrospective histology results as LC and the remaining six images were confirmed as DC. The images were loaded through batch processing for LC and differed for DC. An additional layer of segmentation was added to the primary recipe for effective segmentation of DC. The mean intensity of pixels was calculated for the various segmented layers – tumour and surrounding soft tissue as well as tumour segment DC.

The mathematical process used for the LC algorithm in this study is described below:

Algorithm for LC:

- Image intensity mean (pixel intensity) minus
- Tumour intensity mean (pixel intensity) minus
- Surrounding tissue.[19]

The mathematical process used for the DC algorithm in this study was in line with what is described by Sosa et al.[25] The mathematical process is described below:

Algorithm for DC:

- Image intensity mean (pixel intensity) minus
- Tumour intensity mean minus
- Surrounding tissue intensity mean minus
- Tumour segment DC.[25]

The values of the various intensity means were processed as a data set (Table 3.1).

Table 3.1: Data set findings for the mock study sample

Algorithm 1		
N = intensity mean (pixel)		
Data set	LC – T	LC – ST
LC 1	36.3966	105.7651
LC 2	34.1332	82.4882
LC 3	58.1069	87.4093
LC 4	45.8658	104.4974
LC 5	34.1181	95.8381
LC 6	24.7001	98.8514
Algorithm 2		
Data set 2	DC – T	DC – ST
DC 1	54.6937	111.7207
DC 2	71.8258	118.6943
DC 3	87.8457	102.7707
DC 4	61.3628	95.9202
DC 5	60.8028	80.1279
DC 6	89.7359	114.0227

LC – T: Invasive Lobular Carcinoma Tumour; LC – ST: Invasive Lobular Carcinoma Surrounding Soft Tissue

DC – T: Invasive Ductal Carcinoma Tumour; DC – ST: Invasive Ductal Carcinoma Surrounding Soft Tissue

It is evident from the mock study sample that both LC and DC have unique morphometric properties. This initial evaluation of the mock study sample provided sufficient data for the study to be used prospectively.

3.11 Statistical analysis

Data analysis was performed in consultation with a biostatistician and a p-value less than or equal to ($\leq$) 0.05 was considered significant using Fisher's exact tests for all correlations analysed (see Annexure F). Bayesian statistics were utilised during the prospective study portion to determine the probability of the algorithm's efficiency.

- Main objective 1: Data analysis included descriptive statistics and standard deviation of various collected ultrasound images.
- Main objective 2: Comparisons between pixel intensity measurements LC and DC were made using a paired T-test.
- Main objective 3: The strength of correlation between the collected measurements were made using a Pearson's correlation test.
- Main objective 4: Comparison between pixel intensity measurements of suspicious mass lesions for prospective histology results was undertaken.
- Main objectives 5 and 6: A linear regression formula showed whether a trend existed between the dependent and independent variables.

The descriptive statistics' mean, median, standard deviation and interquartile range were used to describe the continuous variables such as age and time to event. Frequencies and proportions were used to describe the categorical variables. Clustering techniques such as K-means and hierarchical clustering were considered to cluster the bivariate intensity means for breast and surrounding tissue.

A Gaussian mixture model was considered to model the groups, with three groups chosen to reflect the three groups of cancer types. Cohen's Kappa was calculated as a measure of agreement for comparing methods in classifying cancers as malignant or benign. The percentage of the agreement was reported. McNemar's test was used to test for agreement.

Multinomial logistic regression was used to predict the probability of the type of cancer regressing the specialist's classification on that of the algorithm's classification. Tests were evaluated at a 5% level of significance. All analysis was done using STATA 15.

3.12 Ethical and legal considerations

Permission was sought from the Parklane radiology department for use of their patient data (see Annexure G) and to gain access to the stored image data on their ultrasound imaging systems. The Breast Care centres of Prof Carol-Ann Benn were not used as part of data acquisition but served as academic support during the data gathering and research process which occurred over the stated period (See Annexures C).

The researcher was aware that the perusal of personal files could be an invasion of privacy and that the patient's consent should normally be sought for this activity. However, it was not possible to contact each patient to request permission to use the retrospective high-frequency ultrasound images and histological data, therefore, permission to use the data was sought from the medical departments where the data was stored (by contacting the custodian of the data). A unique user access code was provided for the researcher of this study to access the data storage system with access limited to accession numbers only and no identifying markers of patient age, gender or ethnicity included.

The researcher handled the data securely and confidentially. Patient confidentiality was upheld by maintaining the anonymity of personal information. No identifying information was used or reported on. Patient names were replaced with unique study numbers on the high-frequency ultrasound images presented to the interpreters.[42,109]

No direct contact was made with the patients concerning the research findings. With regards to the prospective aspect of the study, no harm was caused to the patients, nor was any risk of harm experienced by any of the participants involved. The algorithm applied to ultrasound images did not impede the standard care received during the screening procedure.

Since the research pertains to the future planning of preventive and prognostic care of breast carcinoma patients, the patients whose records are studied during follow-up and patient management may be benefitted. The protocol was submitted for internal departmental approval to the Faculty of Health Sciences Research Ethics Committee for ethical clearance (ethics approval no. 327/2020) as well as to the PhD Committee for final approval.

3.13 Conclusion

This study is a two-phased study, including both retrospective and prospective data analysis of breast ultrasound images and associated histopathological diagnosis. The data used is to assess possible trends in the various data set groups. There are confounding variables, such as inter-rater variability with the ultrasound image assessment process, that could possibly affect the statistical output.

The data was analysed by a medical biostatistician using Kappa statistics and one-way ANOVA using STATA 14 software. Permission was obtained from the relevant departments and ethical bodies for the use of their retrospective and prospective data. Patient anonymity and confidentiality were upheld throughout the research process. The study research was published in two peer review journals, which can be seen in Annexure J.

The following chapter focuses on the results and their dissemination during statistical analysis.

CHAPTER 4: DISSEMINATION OF RESULTS

4.1 Introduction

The chapter focuses on the presentation of statistical results as done in consultation with a statistician from the Medical Research Council.

The images from January 2017 to December 2018 were effortlessly obtained from Parklane Women's Imaging Centre for retrospective analysis and from December 2020 to June 2021 for prospective analysis.

The *retrospective* data set consisted of 1000 high-frequency ultrasound images during B-mode ultrasound scanning with associated radiologic and pathologic reports. The retrieved images from the local PACS system consisted of 1000 images. The *prospective* data set consisted of 100 high-frequency ultrasound images during B-mode ultrasound scanning which were reported as displaying suspicious mass lesions. The initial pathological results were blind to the researcher during the algorithm analysis process.

4.2 Overall sample population

During the retrospective data collection of the 1000 sample population, 784 image data sets had both radiological and pathological reports available on site to allow effective data set training and analysis using the developed algorithm of this study. Of the 784 sample population, 524 were reported by the pathological diagnosis as benign findings and 260 were reported as malignant. During the retrospective assessment using the developed algorithm, 564 were assessed using MIPAR^{TM}.

During the prospective data collection of the 100 sample population, 98 image data sets provided both radiological and pathological reports for blind analysis using concurrent algorithm testing on site.

4.3 Retrospective data output

The data labelled as retrospective included the ultrasound images, radiology reports and pathology reports. The images were segmented and the mean greyscale levels according to pixel width and length were tabulated to determine if the presence of a unique categorisation method for breast cancer subsets could be developed with the algorithm in use.

Various statistical measures and outputs were used for the assessment of the retrospective data. McNemar's test of agreement was used to test the null hypothesis of p1 = p.1, versus the

alternative of p1⊄p.1.[7] In this study, the agreement of the algorithm to the gold standard of histopathological findings was compared using the test statistic:

$Q = (n12-n21)^2/(n12+n21)$

The test statistic, Q, follows a chi-square distribution with one degree of freedom.

Cohen's Kappa was also used as an alternate to McNemar's test, where the degree of agreement in the data is placed on a scale from -1 to 1. A positive value of Kappa indicates an agreement in the data being compared, whereas negative values prove active disagreement. If Kappa is equal to zero, then no agreement is reported. The main difference between the use of McNemar's test and Cohen's Kappa statistics is the null hypothesis and the amount of data being used. Both tests, however, are designed to determine if there is an agreement or not.[110]

Cohen suggests a Kappa result be interpreted as follows: a value ≤ 0 is considered to have no agreement, 0.01–0.20 a slight agreement, 0.21–0.40 a fair agreement, 0.41–0.60 a moderate agreement, 0.61–0.80 a substantial agreement and 0.81–1.00 an almost perfect agreement (Figure 4.1). When considering per cent agreement, a value of 61% agreement poses a problem, considering that nearly 40% of the sample has inaccurate medical output. For this reason, most researchers prefer 80% agreement as the minimum acceptable inter-rater agreement.[111]

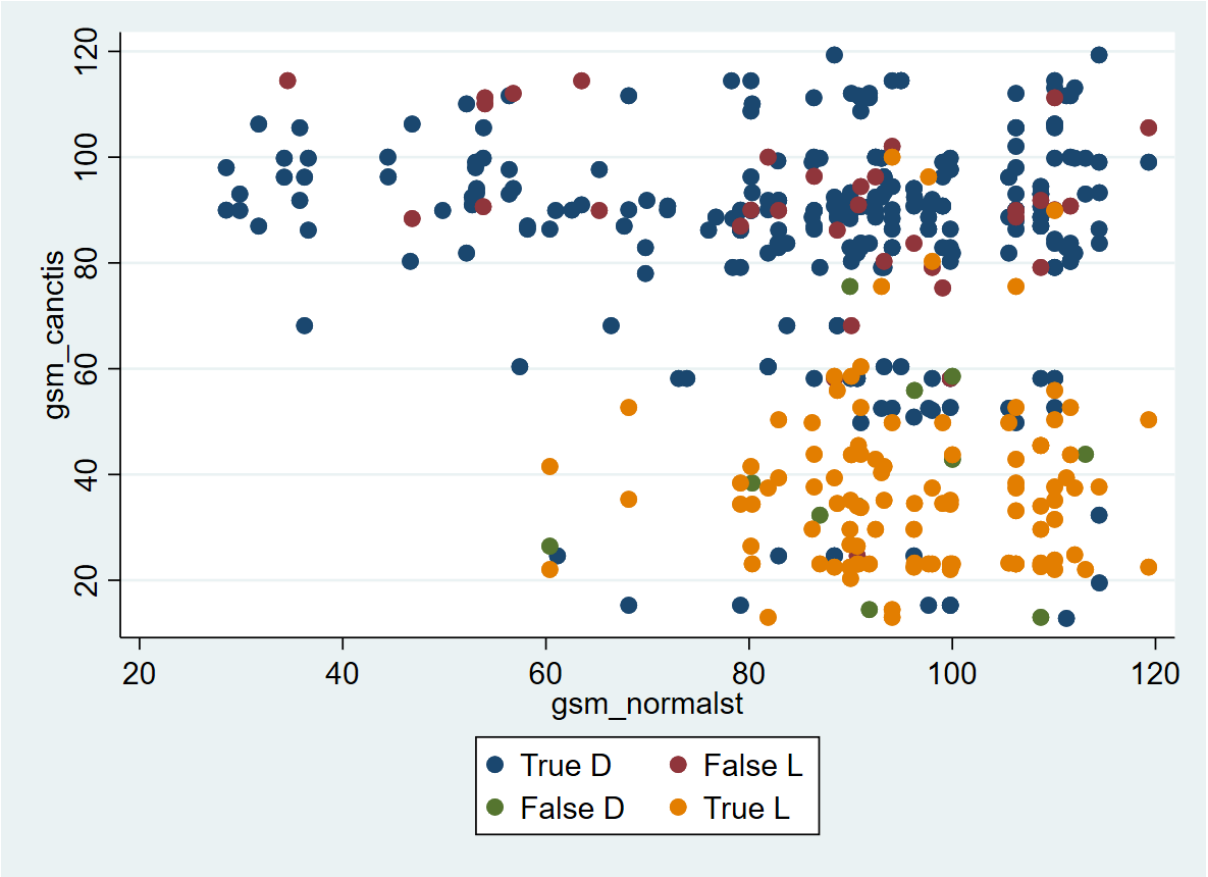


Figure 4.1: Accuracy rate and positive predictor value of the diagnostic algorithm

Legend key

D – Ductal carcinoma

L – Lobular carcinoma

Kappa statistics are frequently used to test inter-rater variability.[111] The study holds the importance of proving that the data collected in the study is a correct representation of the variables measured. Despite Kappa being a popular choice for agreement outputs, it does cause limitations with health-related research; Cohen's interpretation may not be acceptable for medical research outputs as a value as low as 0.41 is considered acceptable.[111]

For this study, one data collector at the Parklane Women's Imaging Centre was utilised, thus the question posed from a statistical perspective is: will a single data collector interpret the same output for each variable every time this data is presented to them? It would seem an obvious assumption that a single data collector would have the same view and output at every instance, however, previous research has shown a reliability coefficient as low as 0.15 when testing intra-rater variability.[111]

4.3.1 Objective 1

The first objective of the study was to collect retrospective ultrasound images for the first-tier population study. This was done through morphologic imaging analysis and relation to breast malignancy as stated by subsequent histopathological diagnosis.

The overall sample frequency, means and medians were assessed to provide an overall assessment of the sample population used for image evaluation. A total of 267 benign images and 525 malignant images were assessed (Table 4.1).

To allow assessment of the sensitivity and specificity of the algorithm tool used, the retrospective data was evaluated with 94% of the DC-type output and 86.3% of the LC type output being diagnosed as correct when compared to the pathological report of each.

Table 4.1: Retrospective sample frequency, means and medians

benign=0.malignant=1		
Presentation	0 1	Total
Known cancer	27 50	77
	35.06 64.94	100.00
	5.14 18.73	9.72
Lump/discharge/other	251 139	390
	64.36 35.64	100.00
	47.81 52.06	49.24
Screening	200 43	243
	82.30 17.70	100.00
	38.10 16.10	30.68
Unknown	47 35	82
	57.32 42.68	100.00
	8.95 13.11	10.35

Total	525 267	792
	66.29 33.71	100.00
	100.00 100.00	100.00

As seen in Table 4.1, a total of 792 patients had a full recorded patient file with information relevant to annual screening and biopsies performed for suspicious mass lesions.

The various clinical presentations of patients, entering the radiology department was included in the retrospective patient reports. Most patients (n = 200) were scheduled for their routine annual screening mammography with no presenting symptoms. It was also noted that 251 of the total sample population reported symptoms, either as a palpable mass, pain in the breast (mastalgia) or a discharge from the nipple.

The importance of educating patients on breast self-examination is imperative from these results as the asymptomatic group proved to show some malignant findings during the routine diagnostic workup.

In Table 4.2, 526 patients reported benign findings, either confirmed using radiological reporting or using histopathological results. Of the 526 benign type findings, 178 were reported as fibroadenoma, which is a common finding during sonographic evaluation and typical reasoning behind the patient's reported mastalgia symptoms.

Table 4.2: Overall benign sample population histopathological results

	benign=0.	malignant=1
Benign	0	Total
Benign breast tissue	67	67
	100.00	100.00
	12.74	12.74
Fibroadenoma	178	178
	100.00	100.00
	33.84	33.84
Fibroadenosis	67	67
	100.00	100.00
	12.74	12.74
Fibrocystic Change	21	21
	100.00	100.00
	3.99	3.99
Fibrocystic change	3	3
	100.00	100.00
	0.57	0.57
Papilloma	39	39
	100.00	100.00
	7.41	7.41

other	151	151
	100.00	100.00
	28.71	28.71
Total	526	526
	100.00	100.00
	100.00	100.00

Another common finding among the benign sample population is fibrocystic disease, a common diagnosis for breast pain, tenderness, and cyclical changes (n = 63). The group typed as ‘other’, was referred to as any other benign type findings such as phyllodes tumour, lipoma, fibroadenolipoma or other benign type findings reported during the diagnostic imaging output.

In Table 4.3, the overall malignant sample population was assessed. Of the total 268, the majority was categorised as invasive type carcinoma; 63,43% of the total malignant sample population was in this category. Another 76 (28,36%) were reported as invasive carcinomas with additional in situ carcinoma in the same biopsy specimen. This again serves to emphasise the importance of early-onset findings using diagnostic algorithms assisting as a second reader to improve the prognosis of such patients.

Table 4.3: Sample population of malignant findings

Cancer	Freq.	Percent	Cum.
Adenocarcinoma	2	0.75	0.75
Axillary lymph node: Metastatic	1	0.37	1.12
DCIS	12	4.48	5.60
Invasive	170	63.43	69.03
Invasive + DCIS	76	28.36	97.39
Metastatic	1	0.37	97.76
Metastatic carcinoma	1	0.37	98.13
Metastatic duct	1	0.37	98.51
Metastatic duct carcinoma	1	0.37	98.88
Non-Hodgkin's lymphoma	1	0.37	99.25
invasive	2	0.75	100.00
Total	268	100.00	

Of the invasive malignancy reported in the overall retrospective sample population (Table 4.4), the majority were found to have no specific primary tumour upon histopathological findings; 61 were reported as DC and 20 as LC. This range of frequency holds true for the overall reported malignancy in women in South Africa, with DC remaining the most predominant reported cancer, followed by LC (5%–10%).

Table 4.4: Invasive malignant sample population

Invasive Cancer	1	Total
Adenocarcinoma	9	9
	3.70	3.70
Ductal	61	61
	25.10	25.10
Lobular	20	20
	8.23	8.23
Mammary	12	12
	4.94	4.94
Metaplastic	1	1
	0.41	0.41
NST / NOS	127	127
	52.26	52.26
other	13	13
	5.35	5.35
Total	243	243

During histological analysis, specimens are also graded according to their rate of proliferation, more specifically in DCIS findings (Table 4.5). The majority of the DCIS sample population was of high-grade type (n = 37), followed by intermediate grade (n = 30) and low grade (n = 25).

Table 4.5: Grading of ductal carcinoma in the sample population

DCIS grade	Freq.	Percent	Cum.
High grade	37	40.22	40.22
Intermediate grade	30	32.61	72.83
Low grade	25	27.17	100.00
Total	92	100.00	

In Table 4.6, of the 226 reported as malignant during histopathological analysis, 41.15% were reported as Grade 2 findings and 37.61% as Grade 3.

Table 4.6: Grading score of carcinoma histopathological findings

Grade	Freq.	Percent	Cum.
1	48	21.24	21.24
2	93	41.15	62.39
3	85	37.61	100.00
Total	226	100.00	

4.3.2 Objective 2

The second objective of the study was to process the selected retrospective ultrasound images (first-tier population sample) with MIPAR^{TM} software to develop an algorithm for the segmentation of mass lesions through semi-automatic processing.

McNemar's test assessed the significance of the retrospective data population when the two subtypes – LC and DC – were being evaluated. However, the Kappa statistic measured as 0.7231 is considered as displaying no agreement between the two subtypes assessed for diagnostic validity. This statistical output poses a question as to how an ANN system improves over time as more data validation is done. This question is answered during the prospective data analysis section to follow in this chapter.

In Figure 4.2, the overall greyscale means reported from each image analysed by the diagnostic algorithm was plotted on a clustering graph to assess if the initial mock study sample hypothesis of various greyscale means for various carcinoma types holds true. It is evident from the graph that there is a definite difference in greyscale means between LC and DC. LC greyscale means, or mean pixel values, all lie typically under the range of 60, whereas DC greyscale means typically lie above the range of 75.

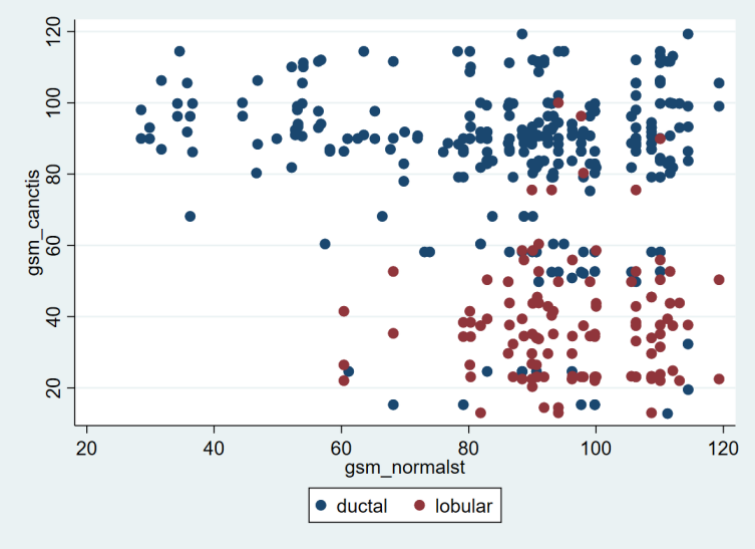


Figure 4.2: Clustering graph for greyscale means of malignant findings

Legend key:

Gsm: grey scale means

Canctis: Cancer tissue

Normalst: Normal soft tissue

4.3.3 Objective 3

The third objective was to assess the algorithm findings (using MIPAR^{TM} software) and reported BI-RADS-US lexicon reporting output (of radiology practices) of retrospective ultrasound images to assess the validity in assessment output (to assess the specificity and sensitivity of the algorithm in comparison to pathology-based diagnosis).

The algorithm provided a finding for each retrospective B-mode ultrasound image being assessed and these were cross-referenced with the pathological result. The sensitivity and specificity were reported as 86% and 89.71%, respectively, with an overall accuracy rate of 88.8% when the algorithm was compared to the histopathological reports.

Due to the limited difference in the agreement that was found with the Kappa statistical output, further assessment using a multivariate T-test was performed. The p-value of the T-test was less than 0.05, which proves a significant difference in greyscale means between LC and DC. The retrospective sample was larger than the prospective data group and this provided insights into the difference between both the health of surrounding soft tissue of both cancers and the tumours themselves in relation to the greyscale means.

In Figure 4.3, initial interest in whether the surrounding soft tissue or the tumour microenvironment also boasted a difference in greyscale means as was seen in Figure 4.2 was prompted. As seen in Figure 4.3, LC's tumour microenvironment has a limited range of greyscale levels when compared to the DC range.

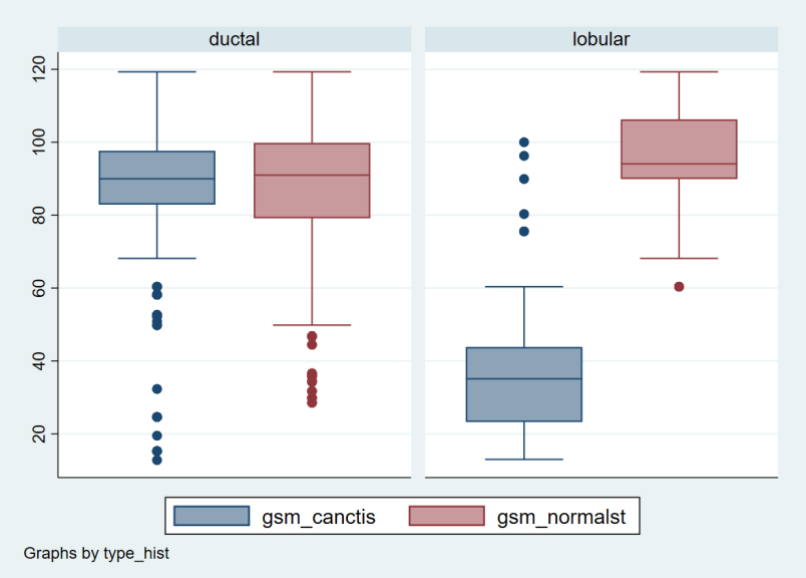


Figure 4.3: Greyscale means of malignant and surrounding soft tissue

Legend key:

gsm – grey scale means

type_hist – histological type of cancer

As discussed, during the literature review, the tumour microenvironment holds much promise as a possible target region for the treatment and diagnosis of breast cancers. A Ranksum nonparametric test that showed a partial significant difference between the greyscale means of both ductal and lobular tumour microenvironments was further performed.

Since the algorithm serves as part of the development of software for the healthcare industry and market, the sample was also assessed using a receiver operator curve (ROC) where any value above the value of 1 or close to the value of 1 is considered as an algorithm with the capability to discern accurately between malignant and benign findings. The ROC value for the retrospective data was reported as 0.936, with a probability index of 0.9701168, thus proving the algorithm has a strong accuracy when compared to histopathological findings. This is seen in Figure 4.4.

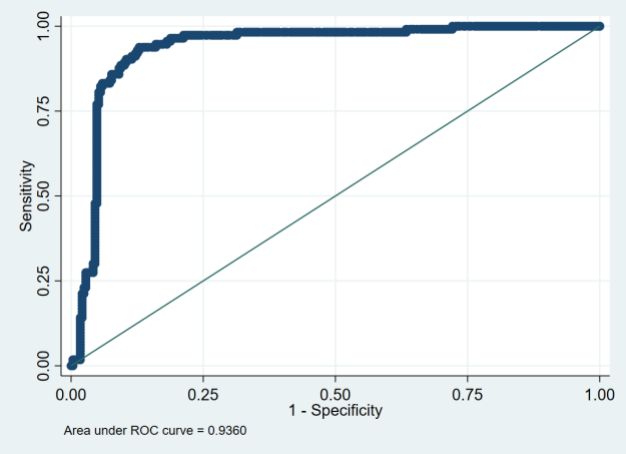


Figure 4.4: ROC curve is derived from logistic regression models

Legend key:

ROC – Receiver operating characteristic curve

Further analysis of the retrospective sample proved no comparison between malignant grading and the greyscale means.

Since a large sample population was used during retrospective data analysis, the statistical output assessed whether any other unique findings could be reported based on greyscale level means. It was found that the various malignant finding gradings did not hold any unique patterns or segments when plotted on the cluster graph (Figure 4.5).

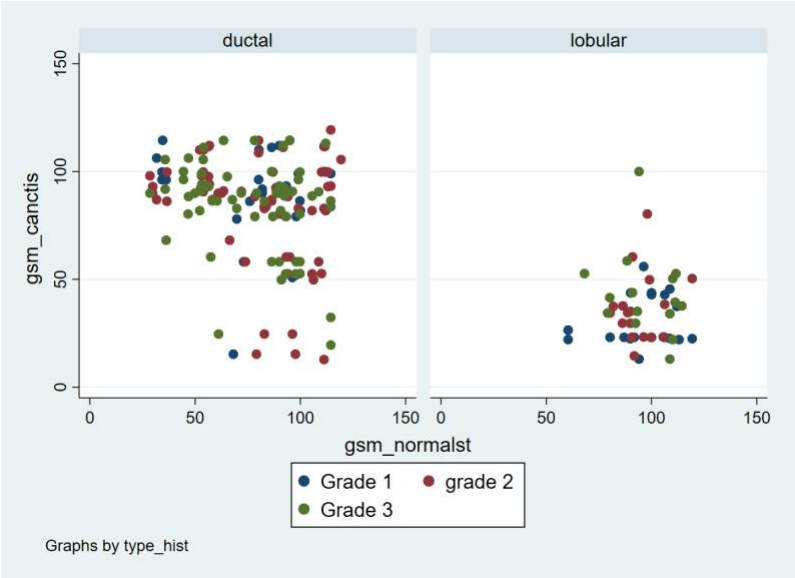


Figure 4.5: Grading of malignancy and greyscale means

4.4 Prospective data output

The data labelled as prospective included ultrasound images, radiology reports and pathology reports. The images were segmented and the mean greyscale levels according to pixel width and length were also tabulated to determine if the presence of a unique categorisation method for breast cancer subsets could be developed with the algorithm in use. The data were collected concurrently to the patient's standard care of practice; the output of the algorithm was not used to impede any treatment or diagnostic outputs.

4.4.1 Objective 4

The fourth objective was to compare the intensity mean (greyscale levels) of the segmented images to assess differences between various types of mass lesions.

The prospective sample population proved the means for McNemar's test and data clustering outputs to assess the novelty of the diagnostic algorithm when compared to the gold standard of pathological analysis.

Data clustering provided some interesting results with regards to greyscale levels for the two main subtypes of breast cancer. Both the tumour masses and the surrounding soft tissue showed unique differences when plotted on the graphs as a pixel value for each element represented on the B-mode ultrasound image. Data clustering proved that the surrounding soft tissue values of LC were lower in pixel value than DC subtypes.

In Table 4.7, the overall sample population consisted of 98 data sets, of which 72 were histologically confirmed as DC (in situ, high-grade or low-grade carcinoma) and 26 were confirmed as LC (high-grade or low-grade carcinoma).

Table 4.7: Sample population of prospective study

	algorithm	
type_hist	0	1	Total
Ductal	69	4	73
	94.52	5.48	100.00
	95.83	15.38	74.49
Lobular	3	22	25
	12.00	88.00	100.00
	4.17	84.62	25.51
Total	72	26	98
	73.47	26.53	100.00
	100.00	100.00	100.00

Cluster analysis is the task of grouping certain data inputs together to assess any similarities or differences in the variable output. It is the main function of exploratory data analysis when used for studies such as pattern recognition, image analysis and ML. The cluster analysis itself does not make use of a specific algorithm but rather has the main purpose of solving a general task. Cluster analysis was originally used in the field of anthropology by Driver and Kroeber[112] in 1932 and then was moved to the field of psychology by Joseph Zubin[113] in 1938.

Under the umbrella term of cluster analysis, various algorithms are used to provide various statistical outputs. Distribution models are used for statistical purposes, such as multivariate normal distributions, whereas graph-based models are a subset of nodes in a graph where every two nodes in the subset are connected by an edge considered as a prototypical cluster form.[114]

Clustering algorithms include various subtypes. One of the most used is connectivity-based clustering which is based on the core idea of objects being more related to nearby objects than those further away. These algorithms connect assorted variables to form clusters. Distribution based clustering, or Gaussian mixed models, are data sets with a fixed number of Gaussian distributions randomly selected with parameters iterated to better fit the data set in question.[114]

Another statistic measure ideal for the current study is the assessment of sensitivity and specificity for the algorithm developed when compared to histology reports. It tests the characteristics of the algorithm in question as a positive predictor value (PPV) and a negative predictor value (NPV) to produce the clinical relevance of the algorithm. The sensitivity of a test is the proportion of people who tested as positively malignant among all who were being evaluated. The specificity of a test

is the proportion of people who test negative among all who were evaluated using the algorithm.[8,115]

4.4.2 Objective 5

The fifth objective was to apply the developed algorithm from the first-tier study output to all suspicious mass lesions of the breast on ultrasound during the prospective research.

The diagnostic algorithm was applied to the prospective sample population and, using McNemar's test, an agreement of 1, with a probability value of 0.7055 ($p>0.05$), was achieved thus proving a strong agreement that the diagnostic algorithm is as effective as pathological outcomes for both LC and DC (Table 4.8).

Table 4.8: McNemar's test of agreement for the prospective study sample

```
  | Controls |
Cases | Exposed Unexposed | Total
-----------------+------------------------+-----------
  Exposed | 22 3 | 25
  Unexposed | 4 69 | 73
-----------------+------------------------+-----------
  Total | 26 72 | 98

McNemar's chi2(1) = 0.14 Prob > chi2 = 0.7055 (p-value)
Exact McNemar significance probability = 1.0000
```

4.4.3 Objective 6

The sixth objective was to compare the algorithm prediction validity to the final histological assessment of suspicious mass lesions.

The algorithm prediction of breast cancer subtypes was compared to the gold standard of pathological assessment and the Kappa correlation indicated 92.86 with a Kappa statistic of 0.81; a relatively strong agreement that the diagnostic algorithm and histological outputs agree in their findings was confirmed.

In Table 4.9, an overall agreement between the histological outputs and algorithm prediction, using Kappa statistics of 92.86%, with a Kappa value of 0.8145 (a good agreement of data outputs), was obtained. A sensitivity and specificity rate of 88% and 94.52%, respectively, were found with a PPV of 16.06 and NPV of 0.127. It is however noted that Kappa does not provide a unique representation of agreement level for medical biostatistics.

Table 4.9: Cohen's Kappa as measurement of agreement for prospective data

```
  Expected
```

```
Agreement Agreement Kappa Std. Err. Z Prob>Z
------------------------------------------------------------------------
  92.86% 61.50% 0.8145 0.1010 8.07 0.0000
```

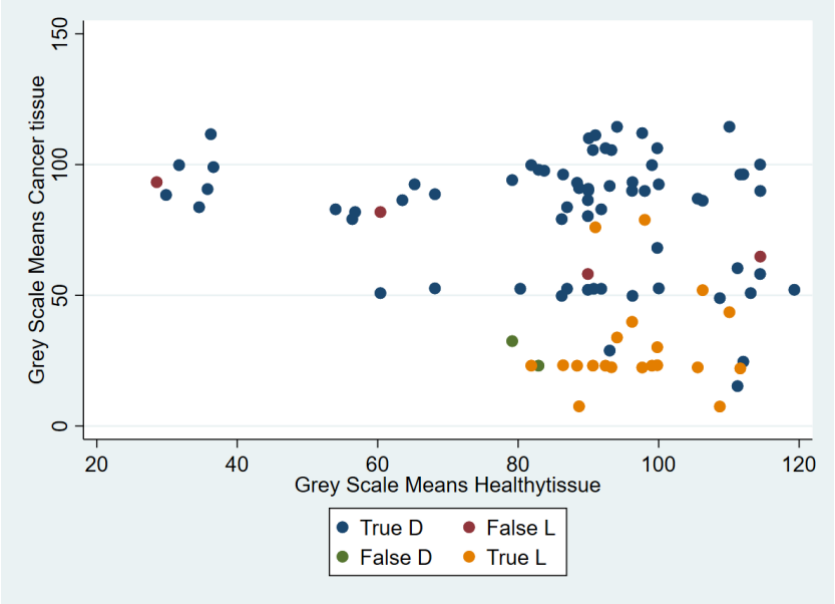


Figure 4.6: Cluster analysis of accuracy rate and positive predictor value for diagnostic algorithm

For further testing of the results displayed in Figure 4.6, a Doornik-Hansen multivariate test for normality was performed with a chi-square value of 51.091 and a probability value > chi-square of 0.0000. This confirms that the greyscale means of surrounding soft tissue is not normally distributed or of statistical significance.

A similar ROC curve analysis was performed for the prospective data sample for logistic regression analysis (Figure 4.7). The value was reported as 0.9468 and proves a strong accuracy rate for the algorithm to accurately predict a breast cancer subtype when compared to histopathological findings.

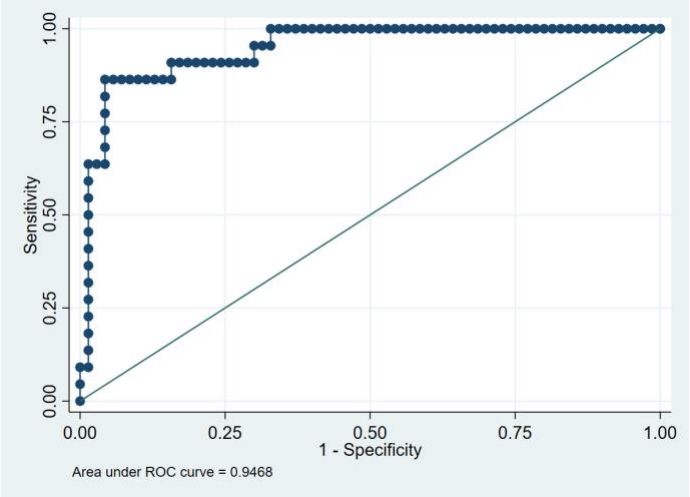


Figure 4.7: ROC for the prospective data sample

.

Note ROC curve comes from the logistic regression model, outcome = log(Odds of being Lobular), predictors = GSM values for cancer and normal tissue. Produces ROC curve with AUC = 0.9468

In Table 4.10, it is seen that the mean greyscale pixel value for DC was 81.00049 and for LC was 30.73843, with a 95% confidence interval. The surrounding soft tissue had a similar mean pixel value of 85.96938 for DC and 95.52764 for LC, respectively.

Table 4.10: Overall mean greyscale levels between ductal and lobular carcinoma

```
  Variable | Obs Mean Std. Err. [95% Conf. Interval]
-------------+---------------------------------------------------------------
 gsm_canctis | 70 81.00049 2.756073 75.50228 86.4987
gsm_normalst | 70 85.96938 2.788501 80.40648 91.53229
-------------------------------------------------------------------------------
-> type_hist = 1
  Variable | Obs Mean Std. Err. [95% Conf. Interval]
-------------+---------------------------------------------------------------
 gsm_canctis | 22 30.73843 3.865905 22.69884 38.77802
gsm_normalst | 22 95.52764 1.961269 91.44896 99.60633
-------------------------------------------------------------------------------
```

4.5 Summary of results

This chapter outlined the results obtained according to the two tiers of the study.

During the initial statistical assessment, it was found that relatively strong agreement occurred between the developed diagnostic algorithm and pathological outputs. However, the data displayed an imbalance in comparison for the retrospective and prospective sample populations which may have caused a skewed statistical output for comparison purposes.

The discussion of the above statistical results is presented in the next chapter.

CHAPTER 5: DISCUSSION OF RESULTS

5.1 Introduction

The two main histological subtypes of breast cancer are invasive LC and invasive DC. A total of 10%–15% of all breast cancers are characterised as LC with a unique discohesive single file pattern of round tumour cells and are renowned for their difficult detectability by digital imaging. In addition, LC is more often diagnosed at a late stage with associated poor prognostic outcomes. A study by Du et al. in 2018[116] found a unique metastatic pattern related to LC. This may be the reason for the poor response rates to chemotherapy and tamoxifen when compared to DC-type cancers.[116]

From a pathological basis, the main difference between these two main cancer types is the lack of E-cadherin protein expression. Invasive LC is also more strongly expressed with oestrogen receptors than DC. Along with this, LC has high rates of progesterone positivity but lower rates of epidermal growth factor receptor 2. LC has a higher rate of Human Epidermal Growth Factor HER2, Her3, PIK3CA and FOXA1 mutations. LC remains a poorly understood breast cancer subtype; recent research has focused its efforts on LC's unique immune signatures and lower metabolic rates. This researcher assessed the immune signature differences between the normal soft tissue of DC and LC and the distinct difference in immune profiling also holds in the current study's diagnostic outputs.[39,116-117]

Within LC, the T-reg cells and MCs were more pronounced whereas in DC the rate of natural killer cells (CD56dim) and activated dendritic cells were more. It was concluded that LC boasts many unique clinical and molecular features still largely unexplored by research communities.[116-117]

The result obtained from this thesis strengthens the current research output and proves that diagnostic imaging may also allow a unique method of classification for the two main breast cancer subtypes. Another novel treatment method currently being studied is targeted immunotherapy as discussed in Chapter 2. Infiltrating lymphocytes are a favourable prognostic indicator for predicted response to neoadjuvant chemotherapy. The issue with most studies to date is the small sample sizes used for the LC population groups. This may have contributed to non-congruent statistical outputs.

LC has been shown to present with later stage recurrence and is more prone to micrometastatic disease. This has previously been explained as tumour dormancy, growth arrest, persistence within the microenvironment and therapy resistance. To further assess this notion, immune assessment, tumour microenvironment and the angiogenesis process need to be studied to confirm the so-called 'dormant state'.[23,105]

5.2 Literature comparison to results

The research of Du et al.[116] was further strengthened by a similar study[118-119], confirming the unique metastatic pattern of invasive LC of the breast. The study included 761 patients: 88 (11.6%) with invasive LC and 673 (88.4 %) with invasive DC. The study confirmed that metastatic sites of invasive LC in patients were frequently found in the bone, gastrointestinal tract and ovarian metastatic disease. DC had a greater lung and liver metastatic involvement.[117,120] The current study may assist diagnostic imaging in follow-up sessions by allowing targeted assessment of these known metastatic sites for breast cancer subtypes.

During the clinical evaluation of patients in the breast imaging department, previous studies[104,121-123] confirmed that invasive LC does not present with the typical firm immobile lump that most clinicians are inclined to palpate to confirm the presence of possible malignancy. The more common signs of invasive LC include thickened skin, nipple inversion and skin dimpling. The adage 'if your tool is a hammer, then everything is a nail' proves true in the unique clinical presentation of breast cancer subtypes and poses many difficulties during annual clinical screening programmes. Due to the characteristic diffuse growth pattern without desmoplastic reaction of invasive LC, they are a difficult lesion to detect both clinically and mammographically.[124]

The sensitivity of mammography to detect invasive LC previously ranged from 57%–81%.[124] While MRI offers increased resolution capabilities, it is also limited in its ability to assess decision-making of breast-conserving surgery outcomes. Considering the above statement, the need for alternate diagnostic methods is valid.

The current study provides one such method, more so for invasive LC and best seen during sonographic evaluation, by deploying AI and DML to assist in visualising different lesions on ultrasound. The diagnostic algorithm developed during the current study boasts an 88% sensitivity and 94.52% specificity rate for malignant lesions when compared to the gold standard of the histopathological outcome. Furthermore, it had a Kappa statistic output of 0.81 for subtype accuracy of LC and DC when compared to histology results. Utilising ML systems to be trained on thousands of pre-categorised imaging datasets allows an ANN system to not only 'know' what invasive LC is but also to 'understand' the unique diagnostic underpinnings of this elusive disease, often missed by the human eye.

This proves that the use of algorithms in clinical practice to ensure that early-stage cancers are not missed during clinical or mammographic screening programmes is necessary. Future treatment methods for LC include focusing on its molecular profile such as fibroblast growth factor receptor 1 – a specific feature of invasive LC. More recent evidence[96] suggests immune receptor gene expression defines certain invasive LC subtypes and their survival rates. In comparison to invasive DC, invasive LC has shown a higher immune cell activity for most types involved.[124]

5.2.1 Tumour microenvironment

The tumour microenvironment holds much promise for clinical, therapeutic and diagnostic systems. The need for further research in this field is most required for metastatic diseases which display poor prognostic outcomes for patients affected by these diseases. In Chapter 2, an overall discussion was included on the early work of Dvorak in which he states that tumours are "wounds that do not heal".[70] For decades, pathologists globally appreciated this 'wound-like' characterisation, however, more recently, it has been redefined as the tumour microenvironment.

The flagship study on the tumour microenvironment was the early work of Allinen et al.[57], which proved a unique gene expression between normal cells, DC and DCIS. It was found that normal breast tissue and DCIS had a significantly higher rate of myoepithelial cells. Genetic aberrations were found most prominently in DCIS myoepithelial cells. Despite the great strides in current breast cancer research, it remains a major clinical challenge with a need for novel therapeutic and diagnostic strategies to reduce the overall mortality rates associated with breast cancer.[73,83]

Du et al.[116] proved that lymphocytic infiltration in invasive LC tumours relates to its poor neoadjuvant chemotherapy response. Another study by Desmeat et al. stated that despite the presence of lymphocyte infiltration in invasive LC, it was 21% lower in invasive DC.[125] The molecular subtyping and unique immune cell responses of DC and LC serve a future role in cross-examination employing targeted ultrasound algorithms where neural networks would be able to subcategorise immune response and targeted follow-up using trained imaging algorithms.[125]

Several in vivo imaging modalities have been utilised in assessing the tumour microenvironment of animal tumour models[101,126]. Computed tomography, using endogenous targeted microbubble contrast media, proved a 50–200 μm in-plane resolution, a 1–2mm positron emission tomography and ultrasound with a 50–500μm in-plane resolution. In comparison to traditional optical imaging microscopy with a resolution range of 2–200μm, the study confirms the diagnostic efficiency of conventional imaging methods to study the tumour microenvironment.[125]

Due to rapid cell growth, hypoxia is typically present in the tumour microenvironment and causes a significant reduction in oxygen from the surrounding blood supply. Both chronic and acute type hypoxia are found in the tumour microenvironment.[125] Using T2-weighted MRI imaging and gadolinium contrast media injection, a distinct morphological difference in tumour and muscle layers was found in mouse models. The study proved that non-invasive imaging methods may be used for the assessment of both cell types and physiological changes of the ECM. The conjoint use of the latest imaging technologies, such as contrast-enhanced ultrasound and AI algorithms, holds much promise to assess the mechanistic information for pre- and post-treatment methods, ensuring that optimal patient outcomes are achieved.[71,127]

5.2.2 Diagnostic imaging analysis

During the early 2000s, ultrasound provided limited diagnostic input to breast analysis due to the lower resolution units that were deployed at that time. Currently, the improvement to high-resolution systems has surpassed the notion of limited diagnostic input, leading to 3–4 more breast cancers per 1000 not previously visualised on mammographic imaging being detected.[34]

A study based in Austria in 2008,[34] used ultrasound not only as an adjunct to mammography but also as a standard protocol for the routine annual screening of breast cancer, more specifically in dense type breast tissue. The study found an improved detection and differentiation of breast cancer. The valued role of breast ultrasound is still contested among medical professionals as the skills and clinical objectives may vary between clinical institutions.

From a medico-legal perspective, the discussion of breast density during initial patient assessments has now become a matter of formal medical legislation in places such as Connecticut[42] where high-risk patients with dense type breast tissue are given informed decision-making regarding additional imaging modalities such as ultrasound and MRI to avoid a delay in possible future breast cancer diagnosis.

The therapeutic implications of invasive DC and invasive LC – more specifically with their pathological response rates and surgical margin status – have been much debated recently[23,83,93]. Due to most invasive LC being low-grade ER-positive tumours, little benefit from neoadjuvant chemotherapy is obtained. Thus, the need to determine reliable biomarkers from cytotoxic agents from which patients would benefit is critical.[98]

The algorithm developed during the current study can be developed for use in neoadjuvant chemotherapy stages to assess if patients with increased prognostic survival rates, five years post-diagnosis, pose any unique tumour microenvironment imaging findings. This may aid in developing imaging-based biomarkers in the future.[98,127]

The increased prevalence of breast cancer within breast screening centres has caused work overload for clinicians. The lack of sufficient qualified radiologists at public sector hospitals causes much strain in early and effective diagnosis input. For this reason, modern AI algorithms, such as the developed algorithm currently being used, may assist in early diagnosis on sites without relevant clinical experience or infrastructure.

A study based in the Czech Republic[128] assessed the prognostic feasibility of contrast-enhanced ultrasound (CEUS) for BI-RADS 3–5 lesions. CEUS allows real-time evaluation of microvascular architecture in suspected breast lesions. Malignant tumours in the study proved a 77.6% specificity and 52.7% sensitivity in comparison to benign tumours detected at a 74.6% sensitivity and 66.4% specificity rate. The current study provided a superior sensitivity and specificity rate of

88% and 94.2%, respectively. The conjoint use of CEUS alongside a trained algorithm may prove a more effective method for the diagnostic characterisation and prediction of breast lesions.[128]

5.3 Various diagnostic processes of breast ultrasound

Various subsets of ultrasound systems exist. These subsets, such as CEUS, elastography, 3-dimensional ultrasound, automated breast ultrasound and computer-aided ultrasound, strengthen the use of ultrasound during breast imaging. The basic B-mode ultrasound systems can detect morphology, orientation, internal echo-structure and margins of suspicious lesions in both fatty and dense type breast tissue.[129]

Elastography raised interest in the early 2010s when the physical properties of sonographic waves and dynamic compression, which produces so-called 'shear waves' or lateral sound waves, were assessed.[21,129] During the dynamic compression and deformation of underlying soft tissue during a sonographic examination, the internal structure's response and propensity to elasticity help elucidate various levels of tissue stiffness.

Elastography has many of its subsets, providing qualitative and/or quantitative results such as strain elastography, shear-wave elastography and transient elastography. The stiffness of malignant breast lesions is related to the desmoplastic reaction of the intra- and extra-nodular infiltration of interstitial soft tissue. There are, however, some clinical limitations to using ultrasound elastography such as lesions deeper than 4cm in the breast tissue; dynamic compression will not provide appropriate lateral sound waves during soft tissue displacement.[21,129]

CEUS focuses on the propensity of blood vessels present in malignant lesions. CEUS utilises intravenous gas microbubbles to improve backscattering from the venous system on ultrasound. CEUS provides both qualitative and quantitative analysis and characterisation of breast lesions. This can include factors such as enhancement patterns between benign and malignant lesions.[11,129-130]

Computer-aided ultrasound assists with the overarching lack of operator dependability and high inter-observer variation rates. CAD serves as a second reader to a radiologist. A basic CAD system consists of four main stages: pre-processing, segmentation utilising an active contour model, feature extraction and selection and classification of the breast lesion. Classification is met by various methods such as ANN, BNN or SVMs.[11,129-130]

Research such as this, confirms the need for imaging modalities to work in unison as a combination of methods may improve and strengthen the overall diagnostic outputs related to breast cancer.

5.3.1 Contrast-enhanced ultrasound comparison to the current study

A study by Sridharan *et al.*[130] compared the specificity and sensitivity of B-mode ultrasound, CEUS and mammography. In the study, the following values were reported (Table 5.1):

Table 5.1: A comparison study between various breast imaging modalities[130]

	B-mode ultrasound	CEUS	Mammography	Current study
Sensitivity	50%	75%	100%	88%
Specificity	92%	75%	20%	95%
Accuracy	81%	75%	38%	93%

The study, however, made use of a small sample size that may have led to a disproportionate statistical output. This study proved a sensitivity and specificity rate of 88% and 94.5%, respectively, which is between 4%–45 % more accurate than the study output of Sridharan *et al.*[129-130]

5.3.2 Machine learning models compared to the current study

Some advantages in pathological engineering such as circulatory tumour DNA and micro-RNAs hold much promise for early breast cancer screening, however, they do not have the appropriate application for use in low-income countries making use of large-scale screening programmes such as those in South Africa.[45,91,131]

A study by Huo et al.[131] set a cumulative collection of data from 1345 patients with a similar data collection method as the current study using both ultrasound and histopathological outcomes for the data validation of their ML models. Their study focused on small lesions, less than 2cm, typically missed on mammography.[131] The dependant variable was the diagnostic result of either benign or malignant tumours and the independent variable was the developed ML model. Seventy-five per cent of the total sample population of their study was used for data set training and the excess 25% was used for data validation.

This current study made use of 794 retrospective images for data set training of the proposed diagnostic algorithm and 98 for the data validation of the prospective sample population – a total of 892 images. Various ML models were tested in this study. The first model delivered an area under the curve (AUC) rate of 0.755 with a clinician accuracy rate of 0.913. The clinician's overall sensitivity and specificity for detecting malignant breast lesions were reported as 92.8% and 89.8%, respectively.[131]

The research of Huo *et al*.[131] provided independent variables for future use in ML models during breast ultrasound. The overall AUC values reported for their ML model were 0.741 and 0.880. The logistic regression model boasted 82.9% and 81.9% sensitivity and specificity rates.[131] This current study proved an AUC value of 94.7% and a sensitivity and specificity of 88% and 94.5% – a 2% improvement in clinician-only diagnostic outputs than recorded within their study. Their model assisted specialists to decide between short-term follow-up or immediate biopsy and intervention.

The current study made use of ANN from an algorithm developed on the MIPAR™ application software. A similar study by Asri et al.[46] used various algorithms, namely support vector machines (SVM), naïve Bayesian networks, SVM-radial basis function (RBF) kernels and RBF neural networks, for the classification and prediction of breast cancer outcomes. The purpose of their study was to evaluate the efficiency of their proposed algorithm through accuracy, sensitivity, specificity and precision.

Classification is considered a crucial component of ML and data mining. Of the four algorithms used during their study, the RBF kernel proved the most accurate at a rate of 96.8% utilising original Wisconsin breast cancer datasets.[44,46] The most promising algorithm to predict breast cancer survival rates was the SVM with an accuracy rate of 97%. The experimental output of their study proved SVM superior in classifiers at 97.1% with a Kappa statistic of 0.93. Neural networks proved the best at precision and recall of benign masses (95%–97%) in comparison to malignant masses (91%–94%). The study also made use of a ROC curve and confusion matrices to assess the efficacy of the classifiers in their ML models; SVM proved the best ROC at a rate of 99%.[44,46]

A study performed in KwaZulu-Natal in 2019[132] assessed ML algorithms in breast cancer screening and detection. The Breast Cancer Coimbra Dataset was used as it contained anthropometric blood analysis data – one of the most non-invasive methods to test for breast cancer. The study developed and proposed a new breast cancer screening tool, as part of the standard breast imaging protocol, utilising AI as a first approach to identifying breast cancer risk in patients.[132]

The schematic diagram in Figure 5.1 may better serve the unique two-phased hospital systems of South Africa (public and private sectors) as the possible reduction of costly additional diagnostic modalities by using AI systems during the initial point of contact in a screening centre is suggested. This may lead to a quicker turnaround time for surgical intervention and possible improvement in the overall prognosis of patients diagnosed with breast cancer. For each standard step during the diagnostic workup of a possible positive diagnosis for breast cancer, AI models can be utilised to ensure that fewer false-negative results occur.

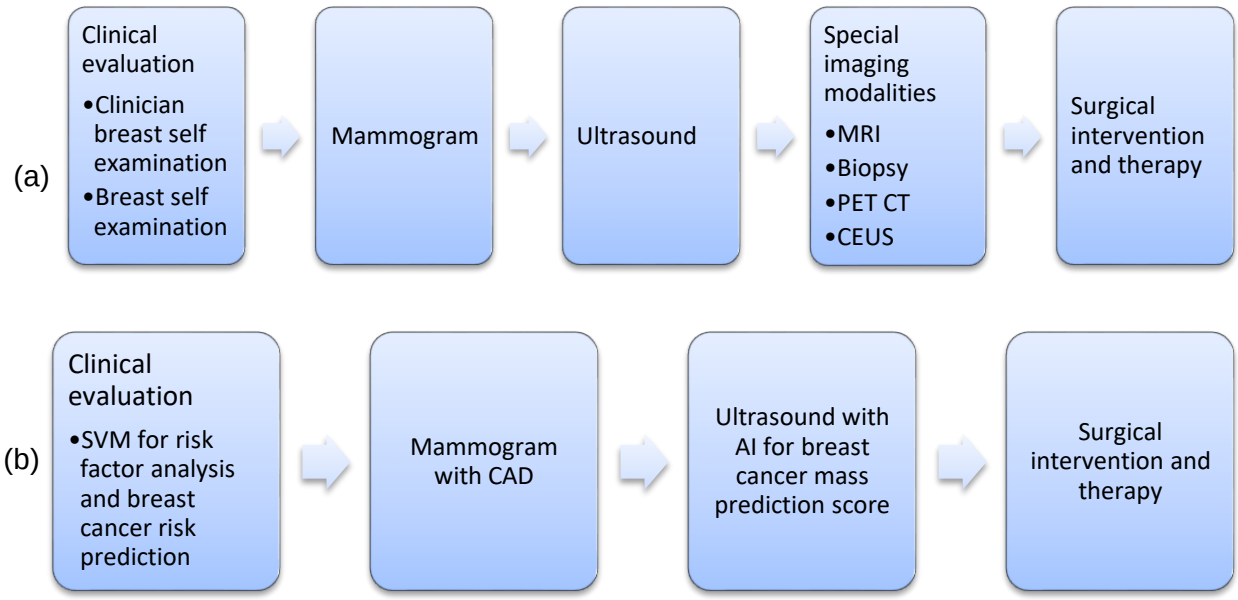


Figure 5.1: Standard (a) and new (b) proposed breast cancer screening methodology in healthcare

For the initial clinical evaluation, a support vector machine can be utilised that would allow patient history taking and throughput to be set on a scaling system, allowing before diagnostic imaging, a depiction of high-risk individuals and flagging for further processing. Standard mammographic imaging would then employ CAD measures to help visualise any possible suspicious mass lesions. Breast ultrasound would then deploy the algorithm of the current thesis during real-time scanning, allocating a prediction score for any suspicious mass lesions to aid the physician in determining if on-site biopsy intervention or direct referral to surgical intervention and treatment is required.

These additional ML processes would assist both the private and public sectors but more so for district hospitals without the infrastructure for on-site diagnostic measures such as stereotactic or ultrasound-guided biopsy. The accuracy, skill and diagnostic output of any clinician, sonographer or physician would be improved through the use of such AI systems, benefitting both patient and doctor.

Zheng et al.[55] assessed unique texture features within a certain region of interest (ROI) of a linear transducer array much like a specialist radiologist would assess a suspicious mass lesion during diagnostic workup. It was stated that typical mammogram findings included both direct and indirect signs such as architectural distortion, asymmetry or a developing density.

Computerised microcalcification detection programmes include similar primary steps to the initial radiologist assessment of mammogram images. The typical steps include the enhancement of the image's signal-to-noise ratio, extraction of the potential image signals, classification of the deployed signal based on rule-based methods and diagnostic output. The above method is the architectural

basis of a convolutional neural network that can differentiate between true and false signals. The output node serves the probability that the input ROI contains microcalcifications.[126]

The diagnostic algorithm developed during the current study has been supported by its statistical output to be able to distinguish between malignant subtypes DC and LC. The initial algorithm's training on retrospective images boasted a lower accuracy than the prospective data set, namely 88.8%. The initial sensitivity and specificity rates for the retrospective data set were 86% and 89.7%, respectively. The initial ROC value was 0.936. Following further imaging training, the prospective data set boasted a 2% improvement in sensitivity rate (88%) and over 5% improvement in specificity (94.5%). The ROC value also increased to 0.947, with a Kappa statistic of 0.81, confirming a substantial agreement between the proposed diagnostic algorithm and histopathological outcomes.

As the ANN was trained on more data during the prospective data analysis stage, the accuracy rate, which can be cross-referenced for diagnostic output, increased as the system's classification neural networks developed their unique sub-features within the detected ultrasound mass.

5.4 Approach to developing software in the South African market

The highest rate of cancer-related mortality in women in South Africa is breast cancer.[6] This is due to the increased stigma associated with the disease in rural settings and the lack of appropriate infrastructure in the healthcare sector to promote early detection and diagnosis. The challenge for AI is a collaboration between sector-related market leaders and passionate innovation drivers for change. The market focus is patient-based care with innovative specialist solutions to improve turnaround times and diagnostic efficiency rates; narrowing the gap between diagnosis and treatment regimens is also the aim. For this, the researcher of the current study proposes a go-to-market solution, utilising multiple hands-on teams to ensure an effective business ecosystem.[13]

The current market tools assist radiologists by reducing the false-negative rate, detecting cancer early-stage and reducing the mortality rates associated with cancer. This study's unique add-on to the current diagnostic dilemma is an individual patient-user interface integrated into the web-based system as an app. Patients of physicians using this product log in for free with a secure login interface. The patient-user system will have unique educational tools to aid patients during their breast cancer treatment process. The site will also allow patients individual membership of the Breast Cancer Support NPO for further emotional support.

The proposed product's value will be quantified through innovation, improved diagnostic efficiency for physicians and educational support for patients. AI in South African healthcare will empower staff, improve healthcare policy (through a better understanding of disease) and improve healthcare delivery (through the earlier detection of the disease status). The researcher of the current study postulates that the above measures would improve diagnostic detection and access for rural

communities via online application software to reach sustainable development goals for healthcare systems.

An increase in time spent on high management uses, improved patient turnaround and less patient anxiety will lead to increased use of AI products in the market. The proposed product can extend to international sales through a web-based approach and make use of the Google Ads web system to incorporate the advertising of health care products applicable to chemo and radiation therapy needs. Free patient access allows those without support in the treatment system to acquire information.

Practitioners using the platform will register at a subscription rate per month, with free use for the patients diagnosed at the practice. The proposed reporting template may then be shared with referring physicians and oncologists. AI for medical diagnosis will have key performance indicators such as improved computing power and advanced algorithms which will compete with thousands of images being reviewed within seconds of data capturing.

The primary outcomes that would be measured are speed to diagnosis, reduction in working hours and increased accuracy of results. Through this business metric, a return on investment can be attained for potential market use. The current study recommends that, for a valid theory of change to be implemented in the market, the strategy and metric alignment proposed in Table 5.2 need to occur. Four of the WHO's Sustainable Development Goals (SDG) would be implemented for post validation of the algorithm in the clinical setting.

Table 5.2: Metrics for theory of change

	METRIC 1	**METRIC 2**	**METRIC 3**	**METRIC 4**
Definition	Breast cancer survivors attend breast education workshops	Youth develop marketable skills in non-traditional jobs	Clinics and doctors have quicker, improved diagnosis and patient outcome rates	Patients have increased life expectancy and quality of life following diagnosis
Aligned with goal	SDG 4 Quality education	SDG 8 Decent work and economic growth	SDG 3 Good health and well-being SDG 9 Industry innovation and infrastructure	SDG 3 Good health and well-being SDG 9 Industry innovation and infrastructure
Outcome threshold?	Miss > 3 sessions	Graduated Yes/No	More than 10% of patients diagnosed per annum	Job loss count post-diagnosis
Data format	Attendance in %	Attendance in % Graduation in %	Statistic output 1 yr. post-diagnosis	Statistic output 1 yr. post-diagnosis

The need for adequate patient education at grassroots level is a key motivator for patients to attend their annual mammographic screening. The algorithm may potentially allow further employment of youth for data validation and training as the current education sector encourages non-traditional job creation, especially for roles within the Fourth Industrial Revolution. Early detection would promote good health and well-being as well as innovation in the diagnostic imaging industry. To measure the proposed theory of change as a social impact metric, the outcome threshold would be validated through improvement in early-stage cancer diagnosis as well as prognosis and life quality following the use of early detection measures in practice.

5.5 Conclusion

The statistical output of the current research study holds much promise as a potential algorithm to be embedded in a go-to-market product in the medical health sector. Various recent studies have been proven less effective than the current developed diagnostic algorithm, thus proving its validity and accuracy in use.

CHAPTER 6: STUDY CONCLUSION

6.1 Introduction

The study aimed to develop a targeted diagnostic algorithm to assist in the ultrasound-based diagnosis of suspicious mass lesions of the breast. The retrospective objectives included the processing of the retrospective data population employing the diagnostic algorithm; the artificial diagnostic output would then be compared to the gold standard of histopathological outcomes.

Following the sensitivity and specificity rates for the retrospective sample, the greyscale means of the two main breast cancer subtypes, DC and LC, would be compared to confirm the initial hypothesis – those different malignant lesions pose unique greyscale means on ultrasound imaging – based on the mock study sample.

For the prospective data set, the diagnostic algorithm tested during the retrospective stage of data analysis would be applied to case-controlled cases deemed suspicious in sonographic appearance to then determine if the algorithm can accurately determine the malignant output. Confirmation by subsequent histopathological diagnosis would then occur.

6.2 Conclusion of study

The initial statement from Maslow[1] – "if your tool is a hammer, then everything is a nail" – was proposed in this thesis. To develop algorithms, the ML model central to its classifying and training purposes, supports this statement. The overarching issue of the lack of clinical and sonographic experience of sonographers, mammographers and radiologists in the field of breast imaging can be addressed through the current study's output.

The main need in the healthcare system is to improve access to services and to deploy early detection methods during the initial diagnosis and eventual management of breast cancer. For this, specialists in the field of breast imaging should be encouraged to make use of algorithms, ML and intelligent software systems to defer to the 'tool is a hammer' concept in the current technological era.

Of all the literature examined and analysed in this study, very few works confirmed the predictive validity of retrospective ultrasound images employing a unique segmentation algorithm. The statistical outputs of studies that were compared to the statistical output of the current research proved less accurate in relation to histopathological findings. This study has proven that greyscale means among various breast cancer subtypes offer a unique and distinct difference in greyscale values, thus proving – with a confirmed accuracy rate of 94.7% – the hypothesis that prediction of suspicious mass lesions can be determined using targeted segmentation algorithms such as the developed model of the current study.

6.3 Limitations to results

The prospective data set was limited to 100 case-controlled images which caused certain masses, initially confirmed through radiological diagnosis as benign, to be missed and resulted in false-negative results.

The current study proved that DC has a greyscale mean value of 81 and LC of 30; the surrounding soft tissue of the two distinct subtypes also differed at 85 and 95, respectively. These findings have some validity in encouraging further studies on ultrasound algorithms and tumour microenvironments.

6.4 Positive attributes of the study

The study has led to a developed software program, funded by the Technology Innovation Agency of South Africa, and has concluded its first stage of pilot testing through the thesis. The opportunity for the algorithm to be applied to other fields of diagnostic imaging has also been considered as the main features of segmentation and categorisation remain the same.

6.5 Recommendations for future studies

Future studies should include larger sample populations and further assessment of other cancer subtypes such as HER2 neu positive, ER and PR types. Further studies in the field of diagnostic algorithms should include the assessment of the tumour microenvironment to ascertain if imaging modalities could also play a role in the latest research on the topic.

The use of first point-of-contact algorithms during the initial screening of patients at local breast centres also holds the potential for further research to assess the overall improvement in prognosis and costs to patients if diagnostic measures improve through targeted ML models in practice.

REFERENCES

1. Maslow AH. The psychology of science a reconnaissance: Harper & Row; 1966.
2. van Dyk B. Application of ultrasound characteristics in the accurate prediction of benign versus malignant solid breast nodules. Faculty Health Sciences. 2011; Master of Technology Radiography
3. Gruszauskas NP, Drukker K, Giger ML, Sennett CA, Pesce LL. Performance of breast ultrasound computer-aided diagnosis: Dependence on image selection. Acad Radiol. 2008; 15(10):1234-45. doi:10.1016/j.acra.2008.04.016
4. Radiology ACo [Internet]. National mammography database (nmd) : Nrdr - national radiology data registry. 2018 [21 September 2018]. Available from: https://www.acr.org/Clinical-Resources/Breast-Imaging-Resources.
5. Singh E, Joffe M, Cubasch H, Ruff P, Norris SA, Pisa PT. Breast cancer trends differ by ethnicity: A report from the south african national cancer registry (1994-2009). Eur J Public Health. 2017; 27(1):173-8. doi:10.1093/eurpub/ckw191
6. Cancer treatment challenges during covid-19 | cansa – the cancer association of south africa. South Africa2020. p. http://www.cansa.org.za.
7. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med. 2015; 13(6):554-61. doi:10.1370/afm.1865
8. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Comput Med Imaging Graph. 2007; 31(4-5):198-211. doi:10.1016/j.compmedimag.2007.02.002
9. Kim JY, Jung EJ, Park T, Jeong SH, Jeong CY, Ju YT, et al. Prognostic importance of ultrasound bi-rads classification in breast cancer patients. Jpn J Clin Oncol. 2015; 45(5):411-5. doi:10.1093/jjco/hyv018
10. Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs. 2015; 18(3):66-7. doi:10.1136/eb-2015-102129
11. Berg WA, Gutierrez L, NessAiver MS, Carter WB, Bhargavan M, Lewis RS, et al. Diagnostic accuracy of mammography, clinical examination, us, and mr imaging in preoperative assessment of breast cancer. Radiology. 2004; 233(3):830-49. doi:10.1148/radiol.2333031484
12. Mukhtar TK, Yeates DR, Goldacre MJ. Breast cancer mortality trends in england and the assessment of the effectiveness of mammography screening: Population-based study. J R Soc Med. 2013; 106(6):234-42. doi:10.1177/0141076813486779
13. Ayeni OA, Joffe M, Cubasch H, Rinaldi S, Taljaard C, Vorster E, et al. Prevalence of comorbidities in women with and without breast cancer in soweto, south africa: Results from the sabc study. S Afr Med J. 2019; 109(4):264-71. doi:10.7196/SAMJ.2019.v109i4.13465

14. Singh GK, Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the united states, 1950-2014: Over six decades of changing patterns and widening inequalities. J Environ Public Health. 2017; 2017:2819372. doi:10.1155/2017/2819372
15. van Rooyen M, Becker T, Westraadt J, Marx G. Creep damage assessment of ex-service 12% cr power plant steel using digital image correlation and quantitative microstructural evaluation. Materials (Basel). 2019; 12(19) doi:10.3390/ma12193106
16. Malherbe K, Bresser P. Association between ultrasound morphologic features and histopathological findings of lobular carcinoma. J Med Radiat Sci. 2019; 66(3):177-83. doi:10.1002/jmrs.336
17. D'Orsi C. Acr birads atlas. In: Springer, editor. Atlanta: American College of Radiology; 2013.
18. Gunhan-Bilgen I, Ustun EE, Memis A. Inflammatory breast carcinoma: Mammographic, ultrasonographic, clinical, and pathologic findings in 142 cases. Radiology. 2002; 223(3):829-38. doi:10.1148/radiol.2233010198
19. Calas MJG, Alvarenga AV, Gutfilen B, Pereira WCdA. Avaliação de parâmetros morfométricos calculados a partir do contorno de lesões de mama em ultrassonografias na distinção das categorias do sistema bi-rads. Radiologia Brasileira. 2011; 44(5):289-96. doi:10.1590/s0100-39842011000500006
20. Joycelyn Jie Xin L, Guek Eng L. A review of invasive lobular carcinoma of the breast: Should it be treated like invasive ductal carcinoma? Integrative Cancer Science and Therapeutics. 2016; 3(5) doi:10.15761/icst.1000211
21. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: Review of techniques and clinical applications. Theranostics. 2017; 7(5):1303-29. doi:10.7150/thno.18650
22. Castellino RA. Computer aided detection (cad): An overview. Cancer Imaging. 2005; 5:17-9. doi:10.1102/1470-7330.2005.0018
23. Biglia N, Mariani L, Sgro L, Mininanni P, Moggio G, Sismondi P. Increased incidence of lobular breast cancer in women treated with hormone replacement therapy: Implications for diagnosis, surgical and medical treatment. Endocr Relat Cancer. 2007; 14(3):549-67. doi:10.1677/ERC-06-0060
24. Tareen A, Kinney J. Finally, machine learning interprets gene regulation clearly. In: Laboratory CSH, editor. Science News: Science Daily; 2019.
25. Sosa JM, Huber DE, Welk B, Fraser HL. Development and application of mipar™: A novel software package for two- and three-dimensional microstructural characterization. Integrating Materials and Manufacturing Innovation. 2014; 3(1):123-40. doi:10.1186/2193-9772-3-10
26. Chen DR, Chien CL, Kuo YF. Computer-aided assessment of tumor grade for breast cancer in ultrasound images. Comput Math Methods Med. 2015; 2015:914091. doi:10.1155/2015/914091
27. Wu JY, Tuomi A, Beland MD, Konrad J, Glidden D, Grand D, et al. Quantitative analysis of ultrasound images for computer-aided diagnosis. J Med Imaging (Bellingham). 2016; 3(1):014501. doi:10.1117/1.JMI.3.1.014501

28. Qiu J, Wu Q, Ding G, Xu Y, Feng S. A survey of machine learning for big data processing. EURASIP Journal on Advances in Signal Processing. 2016; doi:10.1186/s13634-016-0355-x
29. Muralidhar GS, Haygood TM, Stephens TW, Whitman GJ, Bovik AC, Markey MK. Computer-aided detection of breast cancer - have all bases been covered? Breast Cancer (Auckl). 2008; 2:5-9. doi:10.4137/bcbcr.s785
30. Selinko VL, Middleton LP, Dempsey PJ. Role of sonography in diagnosing and staging invasive lobular carcinoma. J Clin Ultrasound. 2004; 32(7):323-32. doi:10.1002/jcu.20052
31. Berg WA, Gilbreath PL. Multicentric and multifocal cancer: Whole-breast us in preoperative evaluation. Radiology. 2000; 214(1):59-66. doi:10.1148/radiology.214.1.r00ja2559
32. Digital mammography: A holistic approach. Hogg P, Kelly J, Mercer C, editors. London: Springer;
33. Berg WA, Leung K. Diagnostic imaging: Breast. 3rd ed. Hooley D, Davis E, editors. Philadelphia: Elsevier; 2019.
34. Madjar H. Role of breast ultrasound for the detection and differentiation of breast lesions. Breast Care (Basel). 2010; 5(2):109-14. doi:10.1159/000297775
35. Kim SH, Seo BK, Lee J, Kim SJ, Cho KR, Lee KY, et al. Correlation of ultrasound findings with histology, tumor grade, and biological markers in breast cancer. Acta Oncol. 2008; 47(8):1531-8. doi:10.1080/02841860801971413
36. Thompson A, Brennan K, Cox A, Gee J, Harcourt D, Harris A, et al. Evaluation of the current knowledge limitations in breast cancer research: A gap analysis. Breast Cancer Res. 2008; 10(2):R26. doi:10.1186/bcr1983
37. Gwatkin RBL. Reproduction in mammals, . Gamete Research. 1989; 24(1):123-4. doi:10.1002/mrd.1120240115
38. Shin HJ, Kim HH, Huh MO, Kim MJ, Yi A, Kim H, et al. Correlation between mammographic and sonographic findings and prognostic factors in patients with node-negative invasive breast cancer. Br J Radiol. 2011; 84(997):19-30. doi:10.1259/bjr/92960562
39. Chen HD, Guo Y. Computer-aided detection of breast cancer using ultrasound images. Computer Science. 2010; 16:46-56.
40. Calas M, Alvarenga A, Gutfilen B, Coelho W, Pereira A. Evaluation of morphometric parameters calculated from breast lesion contours at ultrasonography in the distinction among bi-rads categories. Radiol Brasil. 2011; 44(5):289-96.
41. Smith TM, Bonacuse P, Sosa J, Kulis M, Evans L. A quantifiable and automated volume fraction characterization technique for secondary and tertiary γ′ precipitates in ni-based superalloys. Materials Characterization. 2018; 140:86-94. doi:10.1016/j.matchar.2018.03.051
42. Tagliafico AS, Calabrese M, Mariscotti G, Durando M, Tosto S, Monetti F, et al. Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: Interim report of a prospective comparative trial. J Clin Oncol. 2016; 34(16):1882-8. doi:10.1200/JCO.2015.63.4147

43. Shan J, Alam SK, Garra B, Zhang Y, Ahmed T. Computer-aided diagnosis for breast ultrasound using computerized bi-rads features and machine learning methods. Ultrasound Med Biol. 2016; 42(4):980-8. doi:10.1016/j.ultrasmedbio.2015.11.016
44. Jain A. Artificial intelligence techniques in breast cancer diagnosis and prognosis. Jain A, Jain S, Jain L, editors. Switzerland: World Scientific; 2019.
45. Boeri C, Chiappa C, Galli F, De Berardinis V, Bardelli L, Carcano G, et al. Machine learning techniques in breast cancer prognosis prediction: A primary evaluation. Cancer Med. 2020; 9(9):3234-43. doi:10.1002/cam4.2811
46. Asri H, Mousannif H, Moatassime HA, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science. 2016; 83:1064-9. doi:10.1016/j.procs.2016.04.224
47. Jimenez-Sanchez A, Cast O, Miller ML. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. Cancer Res. 2019; 79(24):6238-46. doi:10.1158/0008-5472.CAN-18-3560
48. Medicum K. Microscopy and microanalysis. Journal of Microscopy. 2018; 1:1-10.
49. Symmans WF. Tissue banking. Breast Cancer Research. 2010; 12(S1) doi:10.1186/bcr2489
50. CytoReason. Machine-learning driven findings uncover new cellular players in tumor microenvironment | e. In: EurekAlert, editor. Public Release: AAAS; 2017.
51. Web P. Machine learning-assisted prognostication based on genomic expression in the tumor microenvironment. In: Cision, editor. AI powered analysis provides quantitative measurements of human interpretable featurs intumor microenvironment. Boston: PRWEB; 2020.
52. Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Mod Pathol. 2018; 31(10):1502-12. doi:10.1038/s41379-018-0073-z
53. Koelzer VH, Sirinukunwattana K, Rittscher J, Mertz KD. Precision immunoprofiling by image analysis and artificial intelligence. Virchows Arch. 2019; 474(4):511-22. doi:10.1007/s00428-018-2485-z
54. Behravan H, Hartikainen JM, Tengstrom M, Kosma VM, Mannermaa A. Predicting breast cancer risk using interacting genetic and demographic factors and machine learning. Sci Rep. 2020; 10(1):11044. doi:10.1038/s41598-020-66907-9
55. Leong AS, Zhuang Z. The changing role of pathology in breast cancer diagnosis and treatment. Pathobiology. 2011; 78(2):99-114. doi:10.1159/000292644
56. Costantini M, Belli P, Bufi E, Asunis AM, Ferra E, Bitti GT. Association between sonographic appearances of breast cancers and their histopathologic features and biomarkers. J Clin Ultrasound. 2016; 44(1):26-33. doi:10.1002/jcu.22312
57. Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell. 2004; 6(1):17-32. doi:10.1016/j.ccr.2004.06.010

58. Tsai MJ, Chang WA, Huang MS, Kuo PL. Tumor microenvironment: A new treatment target for cancer. ISRN Biochem. 2014; 2014:351959. doi:10.1155/2014/351959
59. Soysal SD, Tzankov A, Muenst SE. Role of the tumor microenvironment in breast cancer. Pathobiology. 2015; 82(3-4):142-52. doi:10.1159/000430499
60. Mittal S, Stoean C, Kajdacsy-Balla A, Bhargava R. Digital assessment of stained breast tissue images for comprehensive tumor and microenvironment analysis. Front Bioeng Biotechnol. 2019; 7:246. doi:10.3389/fbioe.2019.00246
61. Sever R, Brugge JS. Signal transduction in cancer. Cold Spring Harb Perspect Med. 2015; 5(4) doi:10.1101/cshperspect.a006098
62. Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, et al. Artificial intelligence in lung cancer pathology image analysis. Cancers (Basel). 2019; 11(11) doi:10.3390/cancers11111673
63. Segovia-Mendoza M, Morales-Montor J. Immune tumor microenvironment in breast cancer and the participation of estrogen and its receptors in cancer physiopathology. Front Immunol. 2019; 10:348. doi:10.3389/fimmu.2019.00348
64. Jain A, Jain A, Jain S, Lakhami J. Artificial intelligence techniques in breast cancer diagnosis and prognosis. Bunke H, Wang PSP, editors. Singapore: World Scientific; 2019.
65. John EP, Ziyin ZMD, Ji-Bin LMD, Shuo WBS. Artificial intelligence in ultrasound imaging: Current research and applications. Advanced Ultrasound in Diagnosis and Therapy. 2019; 3(3) doi:10.37015/audt.2019.190811
66. Ranschaert ER, Orozov S, Algra PR. Artificial intelligence in medical imaging. EUSOMI, editor. Russia: SpringerLink; 2020.
67. Doukas CN, Maglogiannis I, Chatziioannou A, Papapetropoulos A. Automated angiogenesis quantification through advanced image processing techniques. Conf Proc IEEE Eng Med Biol Soc. 2006; 2006:2345-8. doi:10.1109/IEMBS.2006.260675
68. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin. 2019; 69(2):127-57. doi:10.3322/caac.21552
69. Reichling C, Taieb J, Derangere V, Klopfenstein Q, Le Malicot K, Gornet JM, et al. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage iii colon cancer outcomes in petacc08 study. Gut. 2020; 69(4):681-90. doi:10.1136/gutjnl-2019-319292
70. Dvorak HF. Tumors: Wounds that do not heal-redux. Cancer Immunol Res. 2015; 3(1):1-11. doi:10.1158/2326-6066.CIR-14-0209
71. Adamis AP, Brittain CJ, Dandekar A, Hopkins JJ. Building on the success of anti-vascular endothelial growth factor therapy: A vision for the next decade. Eye (Lond). 2020; 34(11):1966-72. doi:10.1038/s41433-020-0895-z

72. Orimo A, Gupta PB, Sgroi DC, Arenzana-Seisdedos F, Delaunay T, Naeem R, et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated sdf-1/cxcl12 secretion. Cell. 2005; 121(3):335-48. doi:10.1016/j.cell.2005.02.034
73. Petitprez F, Sun CM, Lacroix L, Sautes-Fridman C, de Reynies A, Fridman WH. Quantitative analyses of the tumor microenvironment composition and orientation in the era of precision medicine. Front Oncol. 2018; 8:390. doi:10.3389/fonc.2018.00390
74. Paeng K, Jung G, Lee S, Cho SY, Cho EY, Song SY. Abstract 2445: Pan-cancer analysis of tumor microenvironment using deep learning-based cancer stroma and immune profiling in h&e images. Bioinformatics, Convergence Science, and Systems Biology. 2019:2445-. doi:10.1158/1538-7445.Am2019-2445
75. Subramanian E, Tamayo A, Mootha M, S. L. Gene set enrichment analysis (gsea). San Diego: Broad Institute, 2005.
76. Aran D, Hu Z, Butte AJ. Xcell: Digitally portraying the tissue cellular heterogeneity landscape. Genome Biology. 2017; 18 doi:10.5281/zenodo.1004662
77. Tappeiner E, Finotello F, Charoentong P, Mayer C, Rieder D, Trajanoski Z. Timiner: Ngs data mining pipeline for cancer immunology and immunotherapy. Bioinformatics. 2017; 33(19):3140-1. doi:10.1093/bioinformatics/btx377
78. Becht E, Giraldo N, Lacroix L, Buttard B, Elarouci N, Petitprez F. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene ecpression. Genome Biology. 2016; 17(218)
79. Ohlund D, Elyada E, Tuveson D. Fibroblast heterogeneity in the cancer wound. J Exp Med. 2014; 211(8):1503-23. doi:10.1084/jem.20140692
80. Makela K, Mayranpaa MI, Sihvo HK, Bergman P, Sutinen E, Ollila H, et al. Artificial intelligence identifies inflammation and confirms fibroblast foci as prognostic tissue biomarkers in idiopathic pulmonary fibrosis. Hum Pathol. 2020; doi:10.1016/j.humpath.2020.10.008
81. Toratani M, Konno M, Asai A, Koseki J, Kawamoto K, Tamari K, et al. A convolutional neural network uses microscopic images to differentiate between mouse and human cell lines and their radioresistant clones. Cancer Res. 2018; 78(23):6703-7. doi:10.1158/0008-5472.CAN-18-0653
82. Vision R. Detection and segmentation of dendritic cells with deep learning. Tufts Medical Centre2015.
83. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting tumor microenvironment for cancer therapy. Int J Mol Sci. 2019; 20(4) doi:10.3390/ijms20040840
84. Weigel KJ, Paces W, Ergon EJ, Caldara J, Nguyen H, Luengo C, et al. Artificial intelligence assisted macrophages classification in tumor biopsies. In: Biosciences, editor. Flagship2018.
85. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials. Nature. 2013; 502(7471):317-20. doi:10.1038/nature12564

86. Aponte-Lopez A, Fuentes-Panana EM, Cortes-Munoz D, Munoz-Cruz S. Mast cell, the neglected member of the tumor microenvironment: Role in breast cancer. J Immunol Res. 2018; 2018:2584243. doi:10.1155/2018/2584243
87. Tyler AL, Raza A, Krementsov DN, Case LK, Huang R, Ma RZ, et al. Network-based functional prediction augments genetic association to predict candidate genes for histamine hypersensitivity in mice. G3 (Bethesda). 2019; 9(12):4223-33. doi:10.1534/g3.119.400740
88. Varricchi G, Marone G. Mast cells: Fascinating but still elusive after 140 years from their discovery. Int J Mol Sci. 2020; 21(2) doi:10.3390/ijms21020464
89. Fetah KL, DiPardo BJ, Kongadzem EM, Tomlinson JS, Elzagheid A, Elmusrati M, et al. Cancer modeling-on-a-chip with future artificial intelligence integration. Small. 2019; 15(50):e1901985. doi:10.1002/smll.201901985
90. Ikemura K, Barouqa M, Fedorov K, Kushnir M, Billett H, Reyes-Gil M. Artificial intelligence to identify neutrophil extracellular traps in peripehral blood smears Diagnostics and OMICS-Laboratory Diagnostics. New York: International Soceity on Thrombosis and Haemostasis; 2020.
91. Ciallella HL, Russo DP, Aleksunes LM, Grimm FA, Zhu H. Predictive modeling of estrogen receptor agonism, antagonism, and binding activities using machine- and deep-learning approaches. Lab Invest. 2020; doi:10.1038/s41374-020-00477-2
92. Molla M, Waddell M, Page D, Shavlik J. Using machine learning to design and interpret gene-expression microarrays. AI Magazine: Special Issue on BioInformatics. 2015.
93. Langley RR, Fidler IJ. The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. Int J Cancer. 2011; 128(11):2527-35. doi:10.1002/ijc.26031
94. Abdou Y. Machine learning assisted prognostication based on genomic expression in the tumor microenvironment of estrogen receptor positive and her2 negative breast cancer. Annals of Oncology. 2019; 30:55-98.
95. XTalks. Using ai and multiplex biomarker analysis for deeper insights into the tumor microenvironment – br. In: Braenne I, editor. Life Sciences, Biomarkers2019.
96. Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, et al. Immucellai: A unique method for comprehensive t-cell subsets abundance prediction and its application in cancer immunotherapy. Adv Sci (Weinh). 2020; 7(7):1902880. doi:10.1002/advs.201902880
97. Church G. Dyno advances ai-powered gene therapy. Nat Biotechnol. 2020; 38(6):661. doi:10.1038/s41587-020-0569-1
98. Barroso-Sousa R, Metzger-Filho O. Differences between invasive lobular and invasive ductal carcinoma of the breast: Results and therapeutic implications. Ther Adv Med Oncol. 2016; 8(4):261-6. doi:10.1177/1758834016644156
99. Europa H. Artificial intelligence has potential to transform gene therapy. Health Europa; November 2019.

100. Shimizu H, Nakayama KI. A 23 gene-based molecular prognostic score precisely predicts overall survival of breast cancer patients. EBioMedicine. 2019; 46:150-9. doi:10.1016/j.ebiom.2019.07.046
101. Bahcecioglu G, Basara G, Ellis BW, Ren X, Zorlutuna P. Breast cancer models: Engineering the tumor microenvironment. Acta Biomater. 2020; 106:1-21. doi:10.1016/j.actbio.2020.02.006
102. Place A, Huh S, Polyak K. The microenvironment in breast cancer progression: Biology and implications for treatment. Breast Cancer Researcdh. 2011; 13
103. Le Gal M, Ollivier L, Asselain B, Meunier M, Laurent M, Vielh P, et al. Mammographic features of 455 invasive lobular carcinomas. Radiology. 1992; 185(3):705-8. doi:10.1148/radiology.185.3.1438749
104. Pointon KS, Cunningham DA. Ultrasound findings in pure invasive lobular carcinoma of the breast: Comparison with matched cases of invasive ductal carcinoma of the breast. Breast. 1999; 8(4):188-90. doi:10.1054/brst.1999.0042
105. Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: Tumor characteristics and clinical outcome. Breast Cancer Res. 2004; 6(3):R149-56. doi:10.1186/bcr767
106. Calas. MJG, Alvarenga. AV, Bianca Gutfilen, Coelho W, Pereira dA. Evaluation of morphometric parameters calculated from
breast lesion contours at ultrasonography in the distinction
among bi-rads categories. Radiologia Brasileira. 2011; 44(5):289-96.
107. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. Tutorials in Quantitative Methods for Psychology. 2012; 8(1):23-34. doi:10.20982/tqmp.08.1.p023
108. Chapellier C, Balu-Maestro C, Bleuse A, Ettore F, Bruneton JN. Ultrasonography of invasive lobular carcinoma of the breast: Sonographic patterns and diagnostic value. Clinical Imaging. 2000; 24(6):333-6. doi:10.1016/s0899-7071(00)00234-5
109. Levine RJ. The belmont report: Ethical principles and guidelines for the protection of human subjects of research. Washington D.C: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1975 Contract No.: 78-0014.
110. Cohen J. A coefficient of agreement for nominal scales. Education and Psychological Measurement. 1960; 20:37-46.
111. McHugh ML. Interrater reliability: Tha kappa statistic. Biochemia Medica. 2012; 22(3):276-82.
112. Koeber D, Driver K. Quantitative expression of cultural relationships. University of California Publications in American Archeology and Ethnology. 1932;
113. zubin J. A technique for measuring like-mindedness. The Fournal of Abnormal and Social Psychology. 1938; 33(4):508-16.
114. Davis JA. Clustering and structural balance in graphs. Human Relations. Clustering and structural balance in graphs; 20:181-7.

115. Bhardwaj SS, Camacho F, Derrow A, Fleischer AB, Jr., Feldman SR. Statistical significance and clinical relevance: The importance of power in clinical trials in dermatology. Arch Dermatol. 2004; 140(12):1520-3. doi:10.1001/archderm.140.12.1520
116. Du T, Zhu L, Levine KM, Tasdemir N, Lee AV, Vignali DAA, et al. Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism. Sci Rep. 2018; 8(1):7205. doi:10.1038/s41598-018-25357-0
117. Mathew A, Rajagopal PS, Brufsky A. Distinct pattern of metastases in patients with invasive lobular carcinoma of the breast. Geburtshilfe und Frauenheilkunde. 2017; 77(6):660-6.
118. Menezes GL, van den Bosch MA, Postma EL, El Sharouni MA, Verkooijen HM, van Diest PJ, et al. Invasive ductolobular carcinoma of the breast: Spectrum of mammographic, ultrasound and magnetic resonance imaging findings correlated with proportion of the lobular component. Springerplus. 2013; 2:621. doi:10.1186/2193-1801-2-621
119. Mendelson EB, Harris KM, Doshi N, Tobon H. Infiltrating lobular carcinoma: Mammographic patterns with pathologic correlation. AJR Am J Roentgenol. 1989; 153(2):265-71. doi:10.2214/ajr.153.2.265
120. Pestalozzi BC, Zahrieh D, Mallon E, Gusterson BA, Price KN, Gelber RD, et al. Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: Combined results of 15 international breast cancer study group clinical trials. J Clin Oncol. 2008; 26(18):3006-14. doi:10.1200/JCO.2007.14.9336
121. Evans N, Lyons K. The use of ultrasound in the diagnosis of invasive lobular carcinoma of the breast less than 10 mm in size. Clin Radiol. 2000; 55(4):261-3. doi:10.1053/crad.1999.0291
122. Watermann DO, Tempfer C, Hefler LA, Parat C, Stickeler E. Ultrasound morphology of invasive lobular breast cancer is different compared with other types of breast cancer. Ultrasound Med Biol. 2005; 31(2):167-74. doi:10.1016/j.ultrasmedbio.2004.11.005
123. Izumori A, Takebe K, Sato A. Ultrasound findings and histological features of ductal carcinoma in situ detected by ultrasound examination alone. Breast Cancer. 2010; 17(2):136-41. doi:10.1007/s12282-009-0134-8
124. Luveta J, Parks RM, Heery DM, Cheung KL, Johnston SJ. Invasive lobular breast cancer as a distinct disease: Implications for therapeutic strategy. Oncol Ther. 2020; 8(1):1-11. doi:10.1007/s40487-019-00105-0
125. Ramamonjisoa N, Ackerstaff E. Characterization of the tumor microenvironment and tumor-stroma interaction by non-invasive preclinical imaging. Front Oncol. 2017; 7:3. doi:10.3389/fonc.2017.00003
126. Abadjian MZ, Edwards WB, Anderson CJ. Imaging the tumor microenvironment. Adv Exp Med Biol. 2017; 1036:229-57. doi:10.1007/978-3-319-67577-0_15
127. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med. 2020; 288(1):62-81. doi:10.1111/joim.13030

128. Janu E, Krikavova L, Little J, Dvorak K, Brancikova D, Jandakova E, et al. Prospective evaluation of contrast-enhanced ultrasound of breast bi-rads 3-5 lesions. BMC Med Imaging. 2020; 20(1):66. doi:10.1186/s12880-020-00467-2
129. Guo R, Lu G, Qin B, Fei B. Ultrasound imaging technologies for breast cancer detection and management: A review. Ultrasound Med Biol. 2018; 44(1):37-70. doi:10.1016/j.ultrasmedbio.2017.09.012
130. Sridharan A, Eisenbrey JR, Dave JK, Forsberg F. Quantitative nonlinear contrast-enhanced ultrasound of the breast. AJR Am J Roentgenol. 2016; 207(2):274-81. doi:10.2214/AJR.16.16315
131. Huo L, Tan Y, Wang S, Geng C, Li Y, Ma X, et al. Machine learning models to improve the differentiation between benign and malignant breast lesions on ultrasound: A multicenter external validation study. Cancer Manag Res. 2021; 13:3367-79. doi:10.2147/CMAR.S297794
132. Salod Z, Singh Y. Comparison of the performance of machine learning algorithms in breast cancer screening and detection: A protocol. Durban: University of KwaZulu-Natal; 2019.

ANNEXURES

Annexure A: Permission to access data at Parklane Women's Imaging Centre

Permission to access Records / Files / Data base at the Parklane Hospital

Re: Permission to do research at Parklane Radiology in **Johannesburg**

Prof Albert van Schoor and I are researchers working at the Clinical Anatomy Department of Health Sciences at University of Pretoria.

I am requesting permission on behalf of all of us to conduct a study on the Parklane Hospital grounds that involves access to patient records.

The request is lodged with you in terms of the requirements of the Promotion of Access to Information Act. No. 2 of 2000.

The title of the study is: ***A Diagnostic Algorithm for accurate detection of breast carcinoma on ultrasound***

The researchers request access to the following information:

Access to the clinical files, record book of histology reports and the imaging data base.

We intend to publish the findings of the study in a professional journal and/ or at professional meeting like symposia, congresses, or other meetings of such a nature.

We intend to protect the personal identity of the patients by assigning each patient a random code number.

We undertake not to proceed with the study until we have received approval from the Faculty of Health Sciences Research Ethics Committee, University of Pretoria.

Yours sincerely

Permission to do the research study at this radiology practice and to access the information as requested, is hereby approved.

Signed:
Date: 4/12/19
Place: Milpark

Prof Carol Ann Benn
Specialist Surgeon and Head of the Helen Joseph and Netcare Breast Care Centre

by signing this document I am providing clinical assistance, ultrasound images of minimum 40 000 and advice during the research development process of SMART as part of the Technology Innovation Agency Initiative.

Signed: Kyara Bergstrom
Date: 02-Dec-19 | 07:09:08 SAST
Place: Milpark

Kyara Bergstrom
Head of Research and Complementary health Netcare Breast Care Centre of Excellence

Dr Peter Schoub, Head radiologist at Parklane Radiology Women's Wellness Centre

Date: 19/12/19 at Park Lane (place)

Signature:

Annexure B: Declaration against plagiarism

Declaration regarding plagiarism

1. I understand what plagiarism is and am aware of the University's policy in this regard.
2. I declare that this Ph.D. thesis/ dissertation is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.
3. I have not used work previously produced by another student or any other person to hand in as my own.
4. I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

Signed *KMalherbe*

Id nr: 85060200212083

Annexure C: Principal investigator’s declaration for the storage of research data and/or documents

Principal Investigator's Declaration for the storage of research data and/or documents

I, the Principal Investigator(s), Kathryn Malherbe of the following study titled:

IMAGE SEGMENTATION: A DIAGNOSTIC ALGORITHM FOR ULTRASOUND DETECTION OF LOBULAR CARCINOMA.

will be storing all the research data and/or documents referring to the above-mentioned study at the Department of Clinical Anatomy, Basic Medical Sciences Building.

I understand that the storage for the abovementioned data and/or documents must be maintained for a minimum of 15 years from the end of this trial/study.

START DATE OF TRIAL/STUDY:1 Jan 2020 End: Dec 2035

Name: Kathryn Malherbe

Signature: *KMalherbe*

Date: 20/12/2019

Annexure D: Declaration of Helsinki

World Medical Association Declaration of Helsinki

ETHICS PRINCIPLES for Medical Research Involving Human Subjects World Medical Association Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the: 29th WMA General Assembly, Tokyo, Japan, October 1975.

The World Medical Association (WMA) has developed the Declaration of Helsinki as a statement of ethical principles for medical research involving human subjects, including research on identifiable human material and data.

The Declaration is intended to be read as a whole and each of its constituent paragraphs should be applied with consideration of all other relevant paragraphs.

Consistent with the mandate of the WMA, the Declaration is addressed primarily to physicians. The WMA encourages others who are involved in medical research involving human subjects to adopt these principles.

The Declaration of Geneva of the WMA binds the physician with the words, “The health of my patient will be my first consideration,” and the International Code of Medical Ethics declares that “A physician shall act in the patient's best interest when providing medical care.”

It is the duty of the physician to promote and safeguard the health, well-being and rights of patients, including those who are involved in medical research. The physician's knowledge and con- science are dedicated to the fulfilment of this duty.

Medical progress is based on research that ultimately must include studies involving human subjects.

The primary purpose of medical research involving human sub- jects is to understand the causes, development and effects of diseases and improve preventive, diagnostic and therapeutic interventions (methods, procedures and treatments). Even the

Medical research is subject to ethical standards that promote and ensure respect for all human subjects and protect their health and rights.

While the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects.

It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy and confidentiality of personal information of research subjects. The responsibility for the protection of research subjects must always rest with the physician or other health care professionals and never with the research subjects, even though they have given consent

Physicians must consider the ethical, legal and regulatory norms and standards for research involving human subjects in their own countries as well as applicable international norms and standards. No national or international ethical, legal or regulatory requirement should reduce or eliminate any of the protections for research subjects set forth in this Declaration.

11. Medical research should be conducted in a manner that mini- mises possible harm to the environment.

12. Medical research involving human subjects must be conducted only by individuals with the appropriate ethics and scientific education, training and qualifications. Research on patients or healthy volunteers requires the supervision of a competent and appropriately qualified physician or other health care professional. Clinical Review & Education

JAMA Published online October 19, 2013 **E1**

Annexure E: Pictorial review of MIPAR™ imaging software interfaces

IMAGE PROCESSOR INTERFACE OF MIPAR™ IMAGING SOFTWARE

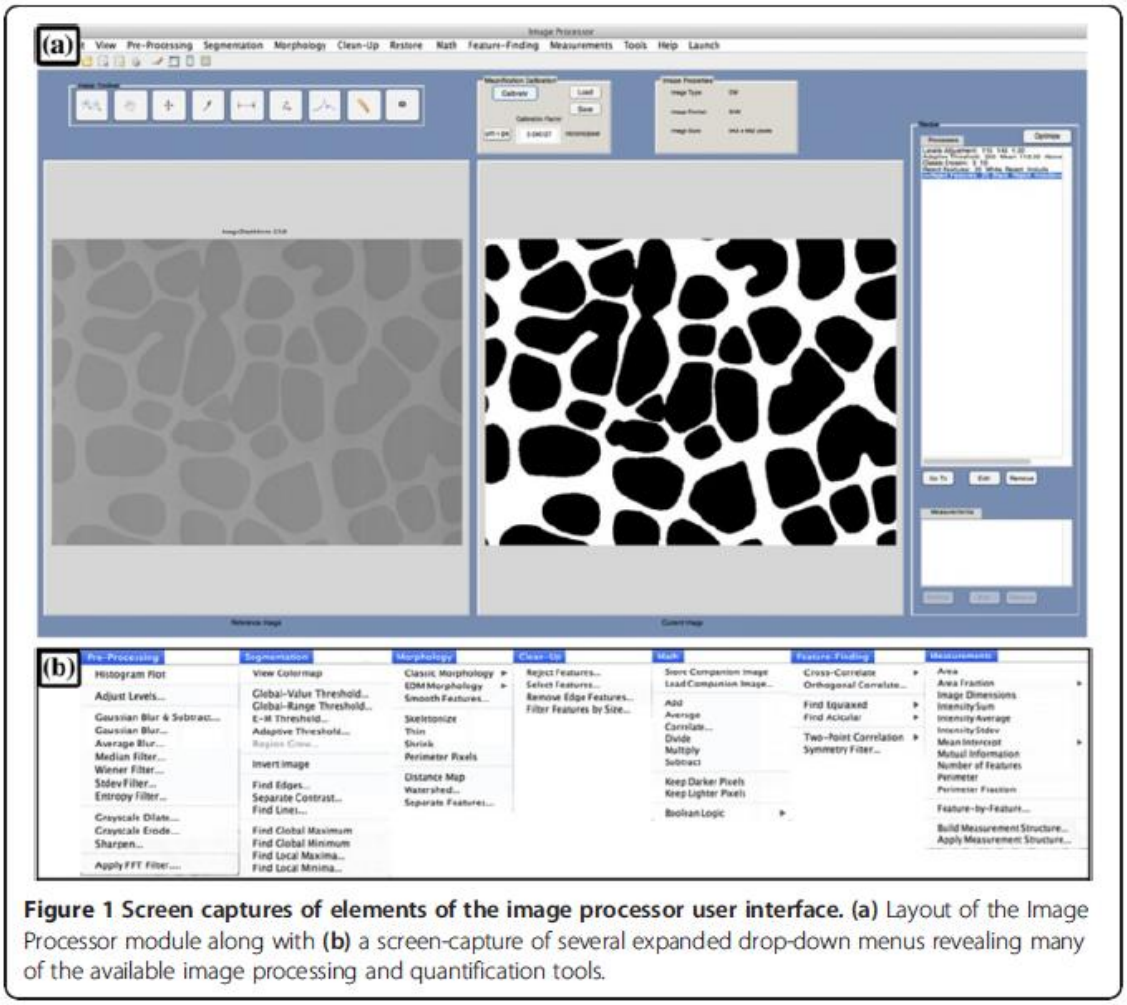


Figure 1 Screen captures of elements of the image processor user interface. (a) Layout of the Image Processor module along with **(b)** a screen-capture of several expanded drop-down menus revealing many of the available image processing and quantification tools.

Batch processor interface of MIPAR™ software

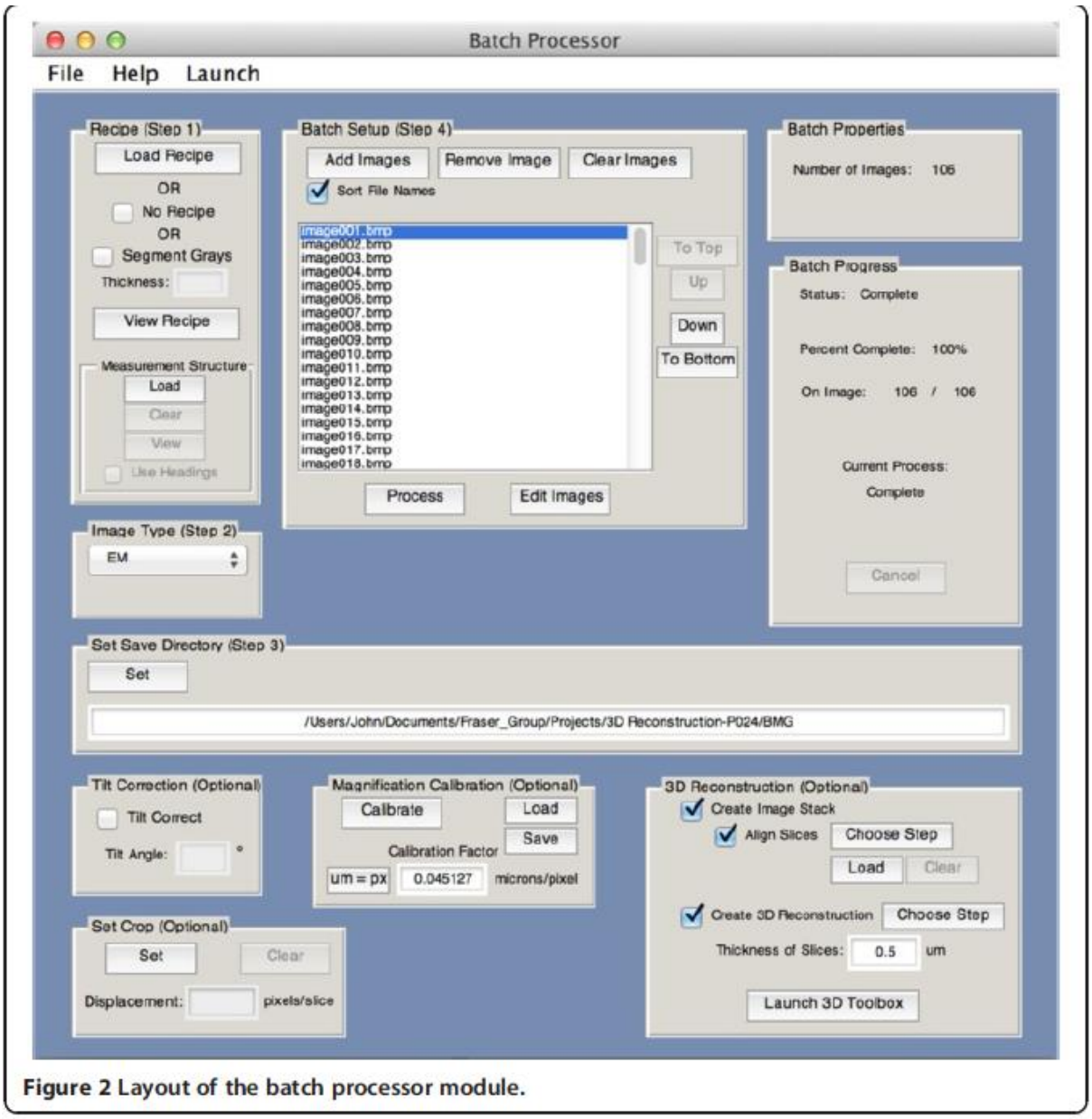


Figure 2 Layout of the batch processor module.

3D toolbox interface of MIPAR™ software

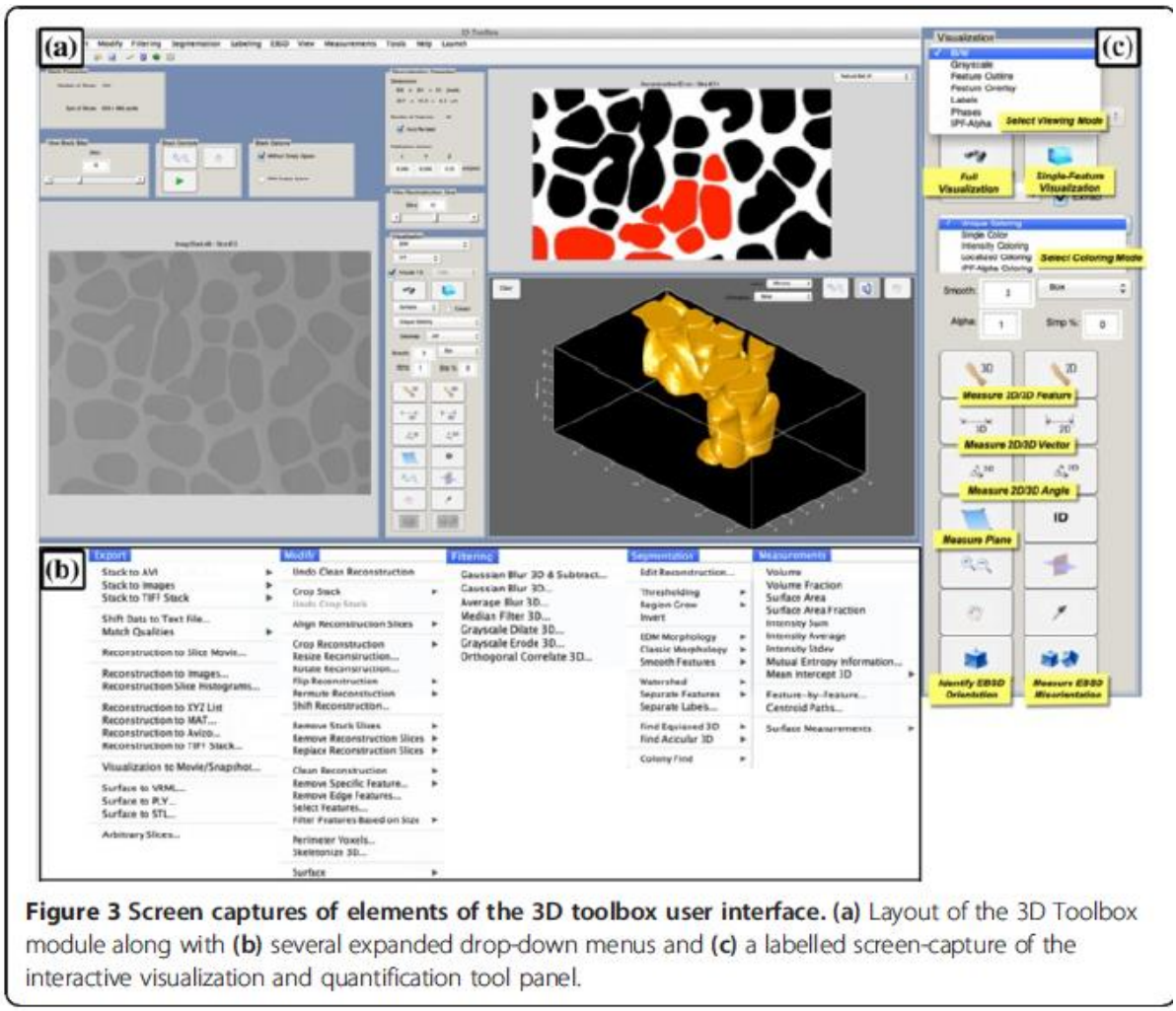


Figure 3 Screen captures of elements of the 3D toolbox user interface. (a) Layout of the 3D Toolbox module along with **(b)** several expanded drop-down menus and **(c)** a labelled screen-capture of the interactive visualization and quantification tool panel.

Feature set application for MIPAR™ software

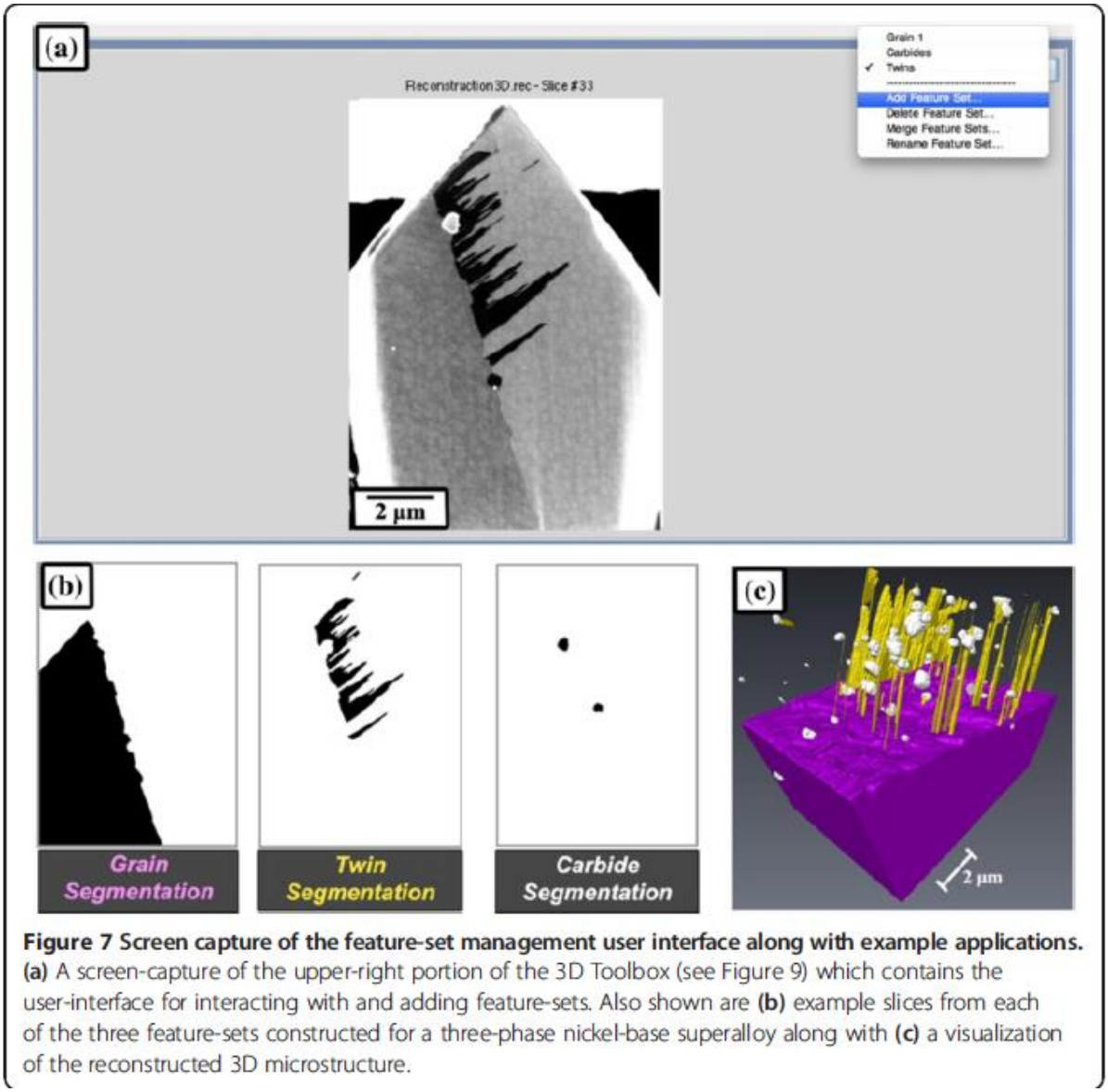


Figure 7 Screen capture of the feature-set management user interface along with example applications. **(a)** A screen-capture of the upper-right portion of the 3D Toolbox (see Figure 9) which contains the user-interface for interacting with and adding feature-sets. Also shown are **(b)** example slices from each of the three feature-sets constructed for a three-phase nickel-base superalloy along with **(c)** a visualization of the reconstructed 3D microstructure.

Annexure F: Statistician support letter

BIOSTATISTICS UNIT

27 January 2020

LETTER OF STATISTICAL SUPPORT

This letter confirms that **K. Malherbe** from the **Department of, Faculty of Health Sciences of the University of** discussed her project: **"A DIAGNOSTIC ALGORITHM FOR ACCURATE MORPHOLOGIC IDENTIFICATION OF BREAST CARCINOMA ON ULTRASOUND"** with me. I confirm that I will assist with the statistical analysis of the study data.

Data analysis

The descriptive statistics mean, median, standard deviation and inter-quartile range will be used to describe the continuous variables such as age and time to event. Frequencies and proportions will be used to describe the categorical variables. Clustering techniques such as K-means and hierarchical clustering will be considered to cluster the bivariate intensity means for breast and surrounding tissue. A Gaussian mixture model will also be considered to model groups, with three groups chosen to reflect the 3 groups of cancer types. Cohen's kappa will be calculated to as measure of agreement for comparing methods in classifying cancers as malignant or benign. Percent of agreement will be reported. McNemar's test will be used to test for agreement. Multinomial logistic regression will be used to predict the probability of type of cancer regressing the specialist's classification on that of the algorithm's classification. Tests will be evaluated at 5% level of significance. All analysis will be done using STATA 15.

Sample size

The researcher will use a database of images to train the deep learning model to classify images and do the segmentation between the breast and surrounding tissue. The learned model will be used to classify the 1100 images (retro- and prospective). Of these 200 – 250 will be sampled to for the purpose of measuring agreement between the algorithm and a specialist who also classifies it.

Name: C Janse van Rensburg
Biostatistics Unit
MRC Pretoria
012 339 8529
Charl.JansevanRensburg@mrc.ac.za

MEDICAL RESEARCH COUNCIL
Biostatistics Unit
Private Bag X385
Pretoria
0001
Tel: 012 339 8523 / Fax: 012 339 8582

Annexure G: Non-disclosure agreement Parklane Radiology Department Letter of intent and clinical advisory support from Helen Joseph Hospital and Milpark Breast Centre of Excellence of Prof. Carol-Ann Benn

LETTER OF INTENT

This letter serves as a request for clinical advisory board members to assist and provide relevant imaging data during the research, development and prototype testing of SMART algorithm software.

The software is currently under consideration by the Technology Innovation Agency (TIA) as well as the Innovation Hub of South Africa for funding and incubation of the product.

SMART will be used in the public and private radiology and breast imaging sectors. The purpose of the software is to promote second reader for a radiologist by means of segmentation as well as identification of breast cancer by means of ultrasound imaging. Furthermore, the software will also allow a patient user interface, which will allow access to Breast Cancer Support Pretoria NPO who serves nationwide breast cancer survivors in the form of breast cancer education and support. The implementation of satellite groups are also promoted and funded by the NPO which strives to promote job creation in the form of volunteer.

The software will be a pioneer of deep machine learning and artificial intelligence in South Africa. It will be open software, able to interact with various platforms of ultrasound units currently used in practice.

We are striving to improve the clinical validity of this software development by allowing clinical specialist in the field of breast imaging, surgery and oncology to become part of the research process to improve the clinical validity thereof. Your participation would require telephonic/ email interaction and clinical advice of certain disease processes as applicable to your area of expertise.

If you would like to be partake in the clinical advisory team for this initiative, as well as provide data for the data training process of the software to the amount of 40 000 images, please send back this letter signed and dated as below to info@breastcancersupport.co.za or katmandukathryn@gmail.com.
Thank you for considering to be a part of this exciting venture.
(Kathryn Malherbe, Director/Founder of Diagnostic Imaging Solutions)

Prof Carol Ann Benn
Specialist Surgeon and Head of the Helen Joseph and Netcare Breast Care Centre

I, Peter Schaub would like to provide clinical assistance and advice during the research development process of SMART as part of Tech innovation Agency initiative.

Kyara Bergstrom
Head of Research and Complementary health Netcare Breast Care Centre of Excellence
+27 82 887 8268
+27 11 480 5673
Kyara.Bergstrom@netcare.co.za

19/12/19

DIAGNOSTIC IMAGING SOLUTIONS

Registered name: Melherbe Imaging Inc

Document Title: Mutual non-disclosure agreement
Document code: DIS001.NDA001
Version: 1.1
Date: 9/7/2019

1. Preamble.

The Parties as identified in Annexure A attached hereto are engaged in discussions relating to their potential collaboration in the Field as likewise described herein; and as they desire and required to disclose Confidential Information to one another; and have agreed to do so subject to the terms and conditions as set out in this agreement.

2. Definitions.

2.1. The following words and/or phrases, when used in this agreement, shall have the following meanings:

2.1.1. "Confidential Information" shall mean all scientific, technical, business, financial, past, present or future research, development, business activities, products, services and technical knowledge or marketing information, whether inside or outside the Field, which one party (the "Disclosing Party") discloses to the other party (the "Receiving Party") in connection with the discussions, and which has been identified in writing as confidential or is of such a nature (or has been disclosed in such a way) that it should be obvious to the Receiving Party that it constitutes Confidential Information (without limiting the generality of the aforegoing). "Confidential Information" shall include any information that falls within the definition of Personal Information (as defined in the Protection of Personal Information Act ("POPI");

2.1.2. "Disclosing Party" shall mean the Party disclosing Confidential Information under this agreement;

2.2. **"Disclosing Purpose"** shall mean, as pertains to any particular joint opportunity(ies) in the Field, the discussions held or to be held between the Parties regarding their possible collaboration and future working relationship with regards to any such opportunity(ies);

2.3. "Effective Date" shall mean the date of the commencement of this agreement as indicated in Annexure "A";

2.4. **"Notice"** shall mean a written document addressed by one Party to the other and either delivered by hand, sent per registered post, e-mail or sms/text to the addresses as indicated in Annexure "A";

2.5. **"Receiving Party"** shall mean the Party receiving Confidential Information under this agreement.

3. Obligation of Confidentiality.

3.1. The Receiving Party undertakes and agrees:

3.1.1. to use the Disclosing Party's Confidential Information only to give effect to the Disclosing Purpose;

3.1.2. to hold in strict confidence and not to publish or disclose to any unauthorised third parties any of the Confidential Information of the Disclosing Party without the prior written consent of the Disclosing Party;

3.1.3. to use the same degree of care (and in any event not less than reasonable care) to safeguard the confidentiality of the Disclosing Party's Confidential Information that it uses to protect its own information of like kind;

3.1.4. to limit any disclosure of such Confidential Information only to those of its employees and professional advisors who have a specific need –to- know to access such Confidential Information and either entered into a written agreement which imposes, or are otherwise bound by the same restrictions as those imposed upon it by virtue of this agreement;

3.1.5. treat any Personal Information in a manner compliant with the POPI Act once in force;

3.1.6. not to disclose or reveal to any third party, whomsoever, either the fact that discussions or negotiations are taking, or have taken, place between the Parties, the content of any such discussions, or other facts relating to the Disclosing Purpose;

3.1.7. on termination of this agreement, to act with the Disclosing Party's Confidential Information in accordance with a Notice delivered to it by the Disclosing Party, and if no such Notice is delivered to the Receiver, to destroy the Disclosing Party's Confidential Information in a similar manner to which it would destroy its own Confidential Information.

4. Exclusions.

4.1. The Receiving Party recognises that this agreement is not intended to restrict use or disclosure of any portion of the Disclosing Party's Confidential Information which:

4.1.1. is as of the Effective Date, or later, made known to the public or otherwise enters the public domain through no default by the Receiving Party of its obligations under this Agreement;

4.1.2. it can show was in its possession prior to the initial disclosure by the Disclosing Party as evidenced by written documents to the like;

4.1.3. is rightfully received by it from a third party owing no obligation of confidentiality to the Disclosing Party;

4.1.4. is independently developed by the Receiving Party by a person(s) who did not have access to the Confidential Information of the Disclosing Party;

4.1.5. is disclosed by the Receiving Party after receipt of written permission from the Disclosing Party; or

4.1.6. it is required or required by subpoena, court order, or similar process to disclose, provided that, in such an event, it will provide the Disclosing Party with prompt written notice of such requirement so that the latter may seek an appropriate protective order and/or waive the Receiving Party's compliance with the provisions of this agreement.

5. Ownership and protection of information.

5.1. The Disclosing Party shall retain ownership of all its Confidential Information as disclosed hereunder.

5.2. Nothing contained in this agreement or in any disclosures made hereunder shall create or imply or be construed as to grant to the Receiving Party any license or other rights in or to the Confidential Information and/or any trade secrets, property rights, patents thereto, or act as a waiver of any rights that the Disclosing Party may have to prevent infringement or misappropriation of any patents, patent applications, trademarks, copyright, trade secrets, know-how or other intellectual property rights owned or controlled by the Disclosing Party as at the Effective Date.

5.3. The Disclosing Party discloses the Confidential Information "as is" and accordingly no disclosure thereof by it hereunder shall constitute any representation, warranty, assurance, guarantee or endorsement by such Disclosing Party with respect to infringement of patents or other rights of third parties, nor is any warranty or representation as to the accuracy, completeness, or technical or scientific quality of any of the Disclosing Party's Confidential Information provided hereunder. (For the avoidance of doubt it is stated expressly that the Disclosing Party neither makes nor have made, any representation or warranty as to the merchantability or fitness for a particular purpose of any Confidential Information disclosed hereunder).

6. Term of obligation.

The Parties' obligations concerning non-disclosure of Confidential Information contained in the above clauses shall commence on the Effective Date and shall continue for five (5) years from the date of each disclosure, unless otherwise agreed between the parties in writing, which after such obligations shall then with terminate.

7. No Violation.

Each party represents that its compliance with the provisions of this agreement will not violate any duty which such party may have towards any third party, including obligations concerning the provision of services to others, confidentiality of information and assignment of inventions, ideas, patents or copyright.

8. Breach.

It is acknowledged that the breach of this agreement by the Receiving Party would cause the Disclosing Party irreparable injury not compensable in monetary damages alone. Accordingly, in the event of a breach, or a threat of a breach, the Disclosing Party, in addition to its other remedies, is entitled to a restraining order, preliminary injunction or similar relief so as to specifically enforce the terms of this agreement or prevent, cure or reduce the adverse effects of the breach.

9. Domicilium citandi et executandi.

The Parties hereto respectively choose as their domicilium citandi et executandi for all purposes of, and in connection with this agreement, the physical addresses and contact details stated in the attached annexure A.

10. Notices.

Any Notice to be given hereunder shall be given in writing and may be given either personally or may be sent by post, email or facsimile and addressed to the relevant party at its domicilium citandi et executandi address as chosen in Annexure "A". Any notice given by post shall be deemed to have been served on the expiry of 7 (seven) working days after same is posted by registered delivery post or air mail. Any notice delivered personally or sent by telefax or email be deemed to have been served at the time of delivery or sending.

11. Governing law and Jurisdiction.

This agreement will be governed and construed by the laws of the Republic of South Africa and the Parties hereby submit to the exclusive jurisdiction of the South African courts to hear any dispute arising therefrom which the Parties are unable to settle amicably.

DocuSign Envelope ID: 2D581B9A-2D59-4994-857A-CD2F48148644

LETTER OF INTENT

DIAGNOSTIC IMAGING

Attention: Prof Carol Ann Benn
Milpark Breast Centre
of Excellence

This letter serves as a request for clinical advisory board members to assist and provide relevant imaging data during the research, development and prototype testing of SMART algorithm software.

The software is currently under consideration by the Technology Innovation Agency (TIA) as well as the Innovation Hub of South Africa for funding and incubation of the product.

SMART will be used in the public and private radiology and breast imaging sectors. The purpose of the software is to promote second reader for a radiologist by means of segmentation as well as identification of breast cancer. Furthermore, the software will also allow a patient user interface, which will allow access to Breast Cancer Support Pretoria NPO who serves nationwide breast cancer survivors in the form of breast cancer education and support.

The software will be a pioneer of deep machine learning and artificial intelligence in South Africa. It will be open software, able to interact with various platforms of ultrasound units currently used in practice.

We are striving to improve the clinical validity of this software development by allowing clinical specialist in the field of breast imaging, surgery and oncology to become part of the research process to improve the clinical validity thereof. Your participation would require telephonic/ email interaction and clinical advice of certain disease processes as applicable to your area of expertise.

If you would like to be partake in the clinical advisory team for this initiative, as well as provide data for the data training process of the software to the amount of 40 000 images, please send back this letter signed and dated as below to info@breastcancersupport.co.za or katmandukathryn@gmail.com.
Thank you for considering to be a part of this exciting venture.
(Kathryn Malherbe, Director/Founder of Diagnostic Imaging Solutions)

Signed: [signature]
Date: 4/12/19
Place: Milpark

Prof Carol Ann Benn
Specialist Surgeon and Head of the Helen Joseph and Netcare Breast Care Centre

by signing this document I am providing clinical assistance, ultrasound images of minimum 40 000 and advice during the research development process of SMART as part of the Technology Innovation Agency Initiative.

DocuSigned by:
Signed: Kyara Bergstrom
Date: [illegible]
Place: Milpark

Kyara Bergstrom 02-Dec-19 | 07:09:08 SAST
Head of Research and Complementary health Netcare Breast Care Centre of Excellence

by signing this document I am providing clinical assistance, ultrasound images of minimum 40 000 and advice during the research development process of SMART as part of the Technology Innovation Agency Initiative.

Contact information for above applicants/ clinical advisors

+27 82 887 6268
+27 11 480 5673
Kyara.Bergstrom@netcare.co.za

DS
KM.

Annexure H: Ethics committee letters of approval

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

Institution: The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

29 September 2020

Approval Certificate
New Application

Ethics Reference No.: 327/2020
Title: A diagnostic algorithm for accurate detection of breast carcinoma on ultrasound: Development of novel MIPAR imaging software algorithm

Dear Mrs K Malherbe

The **New Application** as supported by documents received between 2020-07-09 and 2020-09-16 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2020-09-16 as resolved by its quorate meeting.

Please note the following about your ethics approval:

- Ethics Approval is valid for 1 year and needs to be renewed annually by 2021-09-29.
- Please remember to use your protocol number (327/2020) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers
MBChB MMed (Int) MPharmMed PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag x323
Gezina 0031, South Africa
Tel +27 (0)12 356 3084
Email: deepeka.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo

Annexure I: Funding and commercialisation of PhD research

THE INNOVATION HUB
for Gauteng City Region
A subsidiary of the Gauteng Growth and Development Agency

21 November 2019

Malherbe Imaging
Pretoria
Gauteng
0001

Dear Malherbe Imaging

Congratulations on your selection by the adjudication panel of the Gauteng Accelerator Programme (GAP) Medical 2019, as the **Third (3rd)-prize winner,** for your innovation: **AI software for Breast Carcinoma Detection through Ultrasound Scans.**

GAP Medical is part of The Innovation Hub Management Company (**"TIHMC"**)'s annual suite of innovation competitions aimed at closing the gap between scientific research and commercialisation in Gauteng Province, based on the Gauteng Innovation and Knowledge Economy Strategy (2012) (**"GIKES"**). GIKES seeks to use competitions as a way of identifying innovations that would enhance the competitiveness of Gauteng based companies, improve the efficiency of the public sector, as well as contribute towards the improvement of quality of life of all citizens.

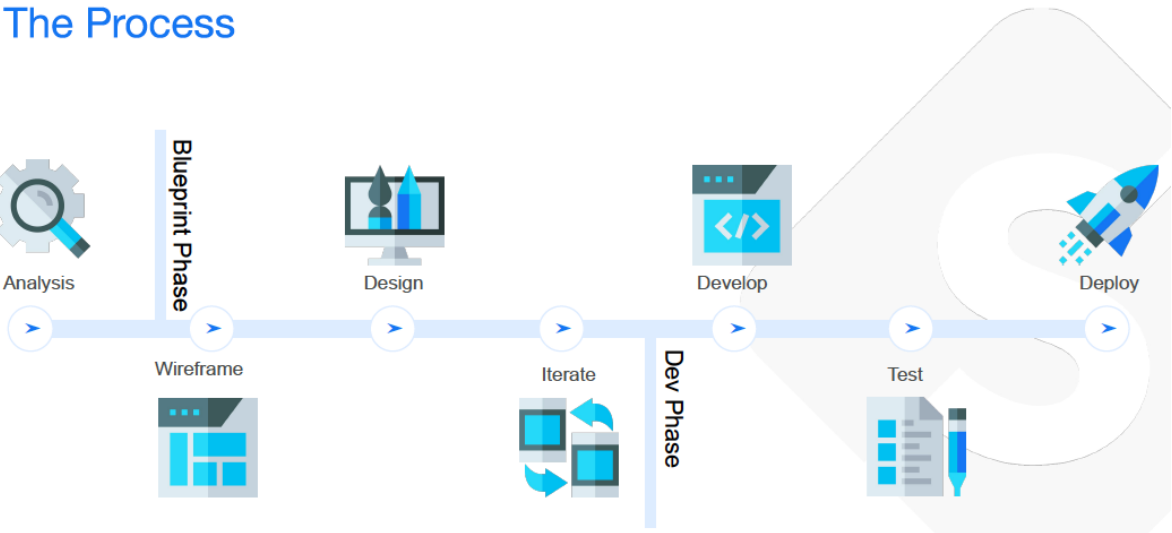
In terms of the competition rules, you qualify for the total seed fund of **R 100 000-00**.

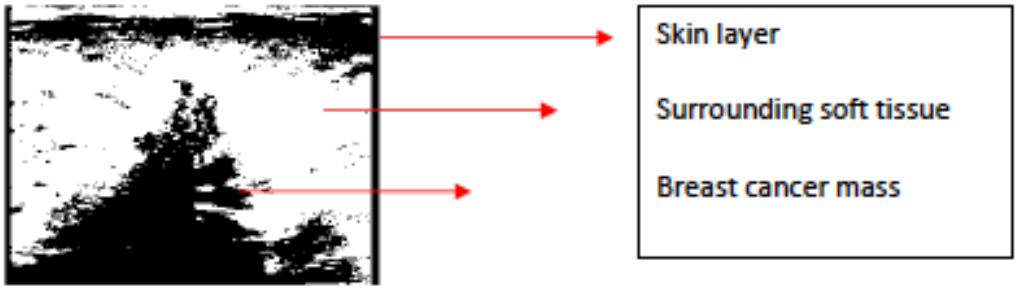
The seed funding will be allocated to you as follows:

- R20 000-00 cash prize to be used at your discretion to cover for expenses you incurred during the competition
- R80 000-00 as seed funding tied to developmental milestones defined in your agreement with the BioPark@Gauteng.

A meeting invitation to discuss the milestones and payments will be relayed to you shortly. Your incubation contract, needs and milestones will be discussed at the meeting.

The seed funding, as well as the business support services to be provided, will have no implications whatsoever on the intellectual property rights relating to your innovation, which will vest with you and/or institute.


The Process
Analysis
Blueprint Phase
Wireframe
Design
Iterate
Develop
Dev Phase
Test
Deploy


Skin layer
Surrounding soft tissue
Breast cancer mass

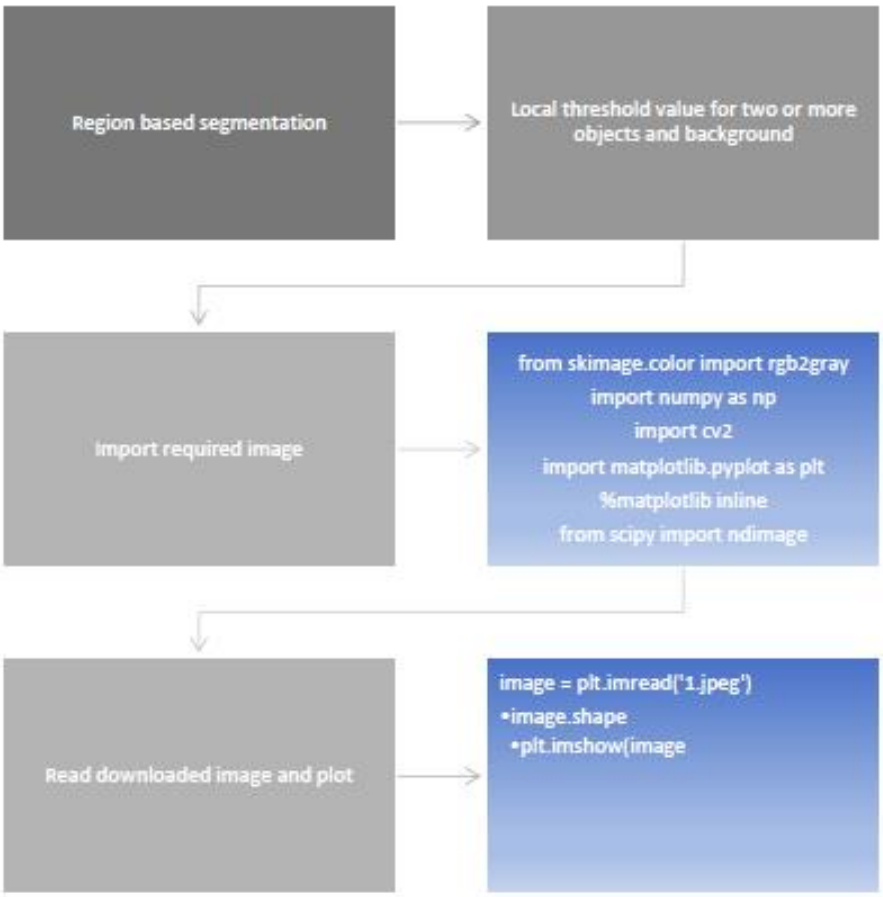
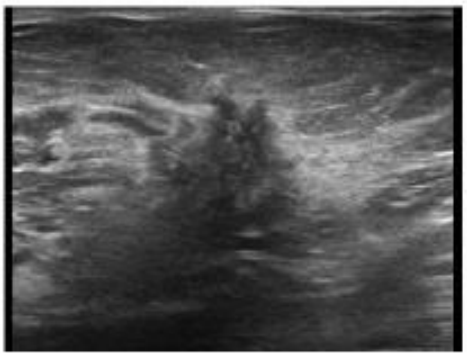
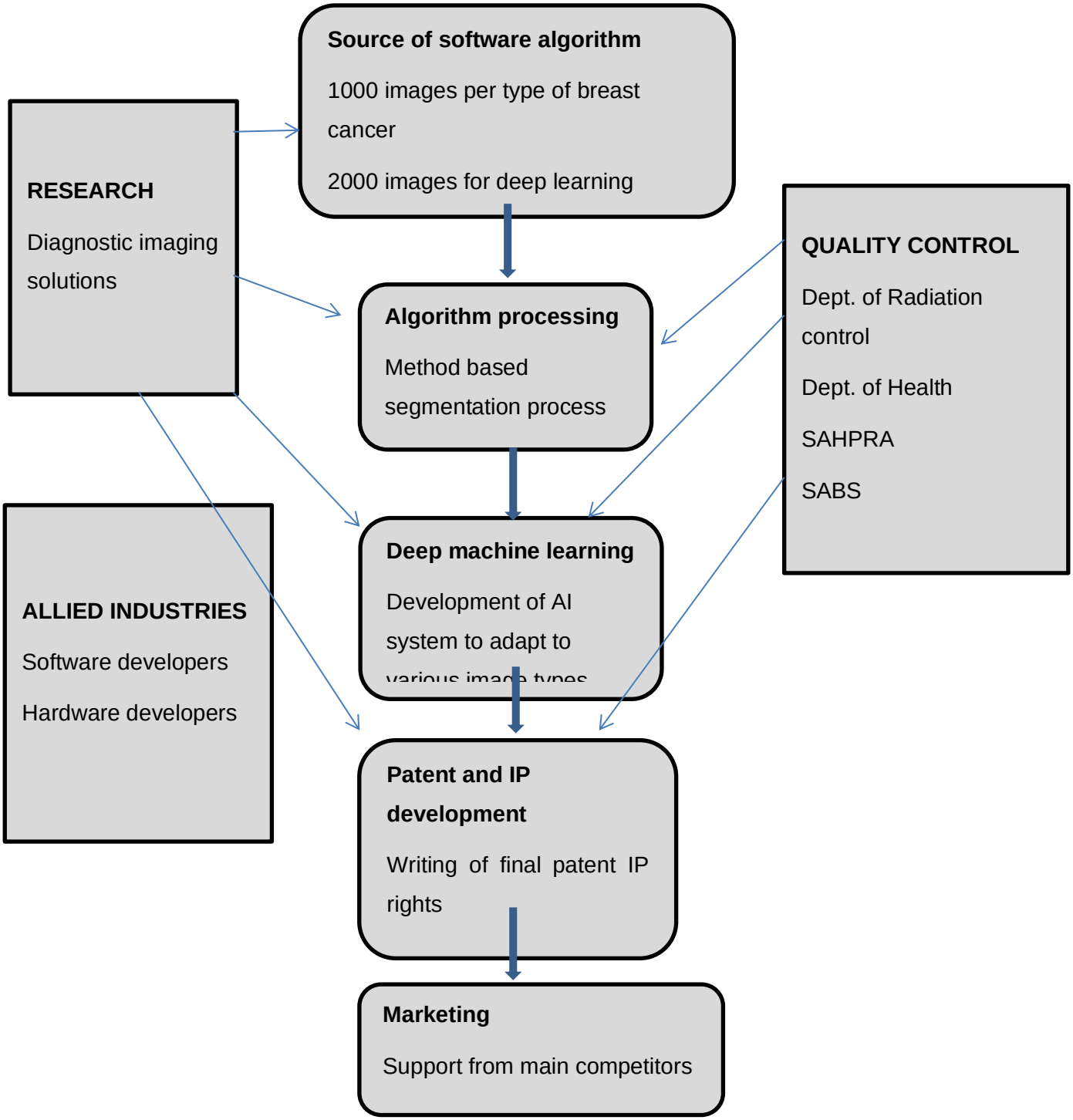


Image may require crop function if imported from PACS database.



Commercialisation of algorithm for in practice use

This is the flowchart planning of commercialisation of software product over a three to five year term.



Annexure J: Published articles related to current thesis research output

ELSEVIER

The American Journal of Pathology

Certificate of publication for the article titled:

"Tumor Microenvironment and the Role of Artificial Intelligence in Breast Cancer Detection and Prognosis"

Authored by:

Kathryn Malherbe

Published in:

Volume 191, Issue 8 (2021), Pages 1364-1373

Serial number: PR-283278-68F97708B3B5

Annexure K: Proof of language editing of final thesis

LET'S EDIT

EDITING CERTIFICATE

25 August 2021

TO WHOM IT MAY CONCERN

DECLARATION: Editing of PhD Thesis

This is to certify that the PhD in Clinical Anatomy thesis entitled **"A DIAGNOSTIC ALGORITHM FOR ACCURATE IDENTIFICATION OF BREAST CARCINOMA ON ULTRASOUND DEVELOPMENT OF A NOVEL IMAGING SOFTWARE ALGORITHM"** submitted by **Kathryn Malherbe** was edited for English language, grammar, punctuation, and spelling by the undersigned. Editing also included addressing the layout and formatting of the document.

The editor will not be held accountable for any later additions or changes to the document that were not edited by the editor, nor if the client rejects/ignores any of the changes, suggestions or queries, which he/she is free to do. The editor can also not be held responsible for errors in the content of the document or whether or not the client passes or fails. It is the client's responsibility to review the edited document before submitting it for evaluation.

Name of Editor: Shirley Wilson

Qualification: Bachelor of Arts (in Education)

Signature:

Let's Edit is a Level 1 EME B-BBEE Contributor (Procurement Recognition Level = 135%)

Address: **570 Fehrsen Street, Brooklyn Bridge Office Park, Brooklyn, Pretoria, 0181**
Tel No.: **012 433 6584**, Fax No.: **086 267 2164** and Email Address: **editor@letsedit.co.za**