Division of Biomedical Engineering

Department of Human Biology

University of Cape Town

Feature detection in ultrasound images for computer aided diagnosis of Hodgkin's Lymphoma

Submitted to the University of Cape Town in fulfilment of the academic requirements for the degree of MSc in Biomedical Engineering by full dissertation.

Tareen Dawood (DWDTAR001)

Supervisors: A/Prof Tinashe Mutsvangwa[1] and A/Prof Estelle Verburgh[2]

Date: 10 September 2021

[1] Department of Biomedical Engineering, UCT, SA
[2] Clinical Haematology, UCT, SA

Declaration

I, **Tareen Dawood**, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the university to reproduce for research either the whole or any portion of the contents in any manner whatsoever.

Signature: ….. Signed by candidate …… Date: ……10 September 2021……

Abstract

The varying clinical presentation of Hodgkin's lymphoma (HL) poses a diagnostic challenge in South Africa, as the clinical picture of this lymphoma overlaps with prevalent comorbidities such as tuberculosis (TB) and the Human Immuno-Deficiency Virus (HIV). HIV infection additionally increases the risk of developing HL. These factors motivate for the need to investigate the role of imaging modalities in the diagnostic pathway of HL.

The goal of this project was to develop and evaluate an automated framework for improving diagnostic imaging interpretability of ultrasound for HL diagnosis in a HIV TB endemic environment. To achieve this, a precise abdominal ultrasound protocol was developed with clinical guidance. The specific frames in the protocol were used to detect several image biomarkers of clinical interest: splenic enlargement (splenomegaly), splenic lesions, splenic microabscesses, abdominal lymph node enlargement, ascites, and effusions (pleural and pericardial). The developed protocol provided a novel guideline to identify an abnormality from the available ultrasound images. A secondary outcome of the protocol was the development of a prospective guide to image Hodgkin's lymphoma patients using ultrasound, however further testing and evaluation is required to validate its use.

Image processing techniques were then applied to identified frames, and geometrical and textural features extracted, to develop an automated abnormality characterisation framework. A total of 36 features were extracted and used to characterise each abnormality. Thereafter, an automated algorithm was used to characterise and classify Hodgkin's lymphoma. A support vector machine model was built, with two experiments performed to evaluate the model. The model achieved a maximum training accuracy of 83%, similar in performance to support vector machine classification models used in medical applications. Noticeably the classification accuracy increased favourably when specific abnormalities were assessed: an enlarged spleen, splenic micro abscesses, ascites, pleural effusions, and pericardial effusions. This may indicate that these specific abnormalities are sufficient to differentiate patients with and without Hodgkin's lymphoma but understanding the reasoning for the decision taken by the system requires further investigation.

In this study we show how image processing and automated classification techniques when applied to ultrasound images, have the potential to improve the differential diagnostic pathway of HL. Further evaluation using a larger dataset is planned, to validate and implement these findings in a strained healthcare setting.

Acknowledgements

I would like to express my sincere gratitude to my primary supervisor A/Prof Tinashe Mutsvangwa. He navigated, supported, and encouraged me throughout the project with immense patience. I appreciate his consistent effort to ensure I was always taking time to create scientific work of a high standard. I would also like to express my gratitude to Dr Estelle Verburgh for her unwavering support, given the demanding clinical environment she currently works in. The guidance, consistent and timeous knowledge from both my immediate supervisors has contributed to the outcomes of my work. Their ability to do all this whilst having many other commitments, whilst we navigated uncertain times the past year during COVID, will always be valued and highly appreciated. In addition to my supervisors, I must express my gratitude to Dr Adeola for his initial and continuous encouragement that led to the start of this project, leading to a new collaboration with Dr Estelle and our Biomedical Engineering division at UCT.

A sincere thank you to the two radiologists who helped me in between meetings, patients scans and even on holidays, Drs Innocent Ncube and Qonita Said Hartley. They both provided me with immense guidance even when under pressure with their long hours and strained environments. Their dedication to improving and serving those around us is admirable and I hope they know how dearly I appreciated their time.

I would also like to thank all my fellow researchers in the Medical Image Inferencing & Distributed Diagnostics (Mi2D2) Group: Ms X Thusini, Ms Y Karanja, Ms B Malila, Mr Jean R Fouefack, Mr E Kamuhire, Ms C Namayega, Mr N Tegang, Mr F Fehr and Mr J Fan for being a supportive and kind team to work with. We shared our experiences and made lifelong friendships and I wish them well in their academic careers and journey ahead.

I would also like to thank Dr Simba, Ms Lillian, and Ms Jenna who reside in the Hematology Department and relentlessly helped me understand, analyse, and extract information to make the key aspects of my analysis validated and verified.

Lastly and most importantly I would like to thank my husband, family, and close friends for all their love and care. I chose a new path after falling ill a few years ago and left my full-time occupation. There have been many challenges and the idea of going back to study was daunting, but their unwavering support has been so generously received. This thesis is dedicated to all of you for always being my pillar of support, strength and keeping my laughter alive on the days I struggled but never spoke about.

Table of Contents

Declaration i

Abstract ii

Acknowledgements ii

Table of Contents iv

List of Figures viii

List of Tables x

List of Abbreviations xi

1. Introduction 1

1.1 Aims 2

1.2 Project overview 3

1.3 Scope and Limitations 4

1.4 Ethical considerations 4

1.5 Dissertation overview 4

2. Literature Review 6

2.1 Characteristics and diagnostic approach of HL 6

2.2 Role of ultrasound imaging in HL 7

2.3 Computer vision 13

2.4 Computer vision and medical imaging applications 13

2.5 Ultrasound image acquisition 13

2.6 Ultrasound image processing 14

2.7 Ultrasound segmentation 14

2.8 Ultrasound feature detection 15

2.9 Machine learning in ultrasound imaging classification 17

2.10 Summary of literature review 18

3. Theoretical considerations for ultrasound image processing 20

3.1 Ground truth image data 20

3.2 Medical image/ultrasound image 21

3.3 Wavelet transform 21

3.3.1 Discrete wavelet transform 22

3.3.2 Two-dimensional discrete wavelet transform (2D DWT) 23

3.3.3 Wavelet thresholding methods 24

3.4 Feature extraction in medical images/ultrasound 26

3.4.1 Textural features 26

3.4.2 Gabor Filter 26

3.4.3 GLCM 27
3.4.4 Textural feature extraction 28
3.4.5 Geometric feature extraction 30
3.5 Classification using medical/ultrasound images 30
3.6 Evaluation metrics 34
3.6.1 Ground truth image similarity 34
3.6.2 Image quality metrics 34
3.6.3 Root mean square error 35
3.6.4 Peak signal to noise ratio 35
3.6.5 Machine learning model evaluation metrics 35
3.7 Feature evaluation metrics 38
3.7.1 Gabor parameters 38
3.7.2 Normalisation levels 39
4. Project data pre-processing and tools 40
4.1 Data overview 40
4.2 Data extraction 40
4.3 HL cohort 40
4.3.1 HL cohort for objective 1 41
4.3.2 HL cohort for objective 2,3 and 4 41
4.3.3 RADLAC cohort for objective 2, 3 and 4 41
4.4 Hardware and software tools 42
5. Development of a prospective abdominal ultrasound imaging protocol 44
5.1 Overview 44
5.2 Methods: new ultrasound imaging protocol and frame sequence 44
5.3 Findings: new ultrasound imaging protocol and frame sequence 45
5.4 Methods: identifying geometrical and textural descriptors 46
5.5 Findings: identifying geometrical and textural descriptors 46
5.6 Conclusion 48
6. Ground truth feature development 49
6.1 Data preparation 49
6.2 Abnormality identification, capture and segmentation 51
6.3 Pre-processing 53
6.4 Denoising and normalisation 55
6.4.1 Methods 55
6.4.2 Results of denoising and normalisation 55

6.5 Geometrical feature extraction 57
6.5.1 Methods 57
6.5.2 Results of geometrical feature extraction 57
6.6 Textural feature extraction 58
6.6.1 Methods and findings 58
6.7 Discussion 60
6.8 Conclusion 61
7. Automated abnormality characterisation framework 62
7.1 Overview 62
7.2 Methodology 62
7.2.1 Input files 62
7.2.2 Automated descriptions 62
7.2.3 Automated extraction 63
7.3 Results 64
7.4 Conclusion 65
8. Evaluation of an automated feature classification model for Hodgkin's lymphoma 66
8.1 Support vector machine model development 66
8.2 Experiment overview 67
8.3 Experiment one 68
8.3.1 Experiment one: Training, testing, and evaluating the support vector machine classifier 69
8.3.2 Experiment one: Support vector machine results 70
8.3.3 Experiment one: Discussion 70
8.4 Experiment 2 71
8.4.1 Experiment two: Training, testing, evaluation and results the support vector machine classifier 71
8.4.2 Discussion 72
8.5 Conclusion 73
9. Conclusion 75
9.1 Summary of findings 75
9.1.1 Development of an ultrasound imaging protocol 75
9.1.2 Ground truth feature development 76
9.1.3 Automated abnormality characterisation framework 77
9.1.4 Evaluation of an automated classification model 77
9.2 Limitations and recommendations for future work 78

9.3 Overall conclusions and contribution of the project 80

Appendix A i

Table A.1: A single consolidated view of the protocol and all descriptors. i

Table A.2: A snippet of the drop-down list (spleen only shown) utilised by the radiologists to capture abnormalities. ii

Table A.3: Radiologists 1's findings iii

Table A.4: Radiologists 2's findings iii

Table A.5: A final consolidated table of mutual abnormalities between radiologists utilised for the automated frameworks iv

References 1-14

List of Figures

Figure 2.1: Classical concepts in the pathway to diagnosis and treatment of HL.

Figure 2.2: Multiple focal lesions (depicted by white arrows) illustrates a high degree of splenic involvement in an HL patient.

Figure 2.3: A series of steps in a computer vision medical imaging application.

Figure 3.1: Different types of wavelets used to decompose a signal.

Figure 3.2 Illustration of different levels of decomposition when a 2-D DWT is applied to an image. On the left an image with a single level decomposition creating four sub-bands. The image on the right has two levels of wavelet decomposition, creating four new sub-bands to further decompose the LL1 sub-band.

Figure 3.3 Graphical representation of hard (left) and soft (right) thresholding.

Figure 3.4 A visual representation to indicate the four directions and single pixel distance that can and has been used to build a GLCM.

Figure 3.5 A visual illustration of two hyper-planes chosen, one with a small margin (left) versus a larger margin (right). The image on the right illustrates a better choice for a hyperplane to discriminate between classes to develop a model that generalises well on unseen data.

Figure 3.6 Illustrates the support vectors and identified optimal canonical hyperplane, using training data x for linearly separable data.

Figure 3.7 A quadratic curve (solid line) versus a linear separation line (dotted line). The data points in red would be mis-classified if a linear separation model were used.

Figure 3.8: Left: Overlapping datasets of diseased or non-diseases ultrasound images indicating additional metrics to use for classifier evaluation. Right: ROC curve (Hallinan, 2014).

Figure 4.1: Steps taken to create the HL cohort.

Figure 4.2: Steps taken to create the RADLAC cohort.

Figure 6.1 The naming conventions and folders used to store and save all images, findings, and outcomes from objective 2.

Figure 6.2 Original frame analysed after cropping and removing patient details.

Figure 6.3 Frame illustrating just the ROI or the enlarged spleen after the *get_final_mask* function is applied to the frame.

Figure 6.4 Radiologist 1 (top left) segmentation and radiologist 2 segmentation (bottom left) with the final single objective segmentation illustrated at the bottom, with the SSIM metric values included.

Figure 6.5: Frame 1 for patient 4 from the HL cohort after the *denoise* function is applied with an estimated noise level of 0.02 versus 0.08 for the image on the right. At a noise estimate of σ = 0.02, noise is removed while not over smoothing the image or leave a very granular appearance when compared to the image on the right (σ = 0.08). In addition, PSNR is slightly better to for the 0.02 noise estimate.

Figure 6.6: The image on the left illustrating the texture identified with the Gabor filter at λ = 10 compared to λ = 100 for the image on the right when all other parameters are kept the same.

List of Tables

Table 2.1: Table summarising the review of abdominal biomarkers.

Table 3.1 Common kernel functions

Table 3.2: Automated classification model outcome versus actual gold standard of diagnosis (Parikh et al., 2008).

Table 4.1: Data characteristics of the HL cohort.

Table 4.2: Data characteristics of the RADLAC cohort.

Table 5.1 New ultrasound imaging protocol and frame sequence developed for the seven abnormalities of interest.

Table 5.2 Geometrical and textural descriptors corresponding to each frame in the new protocol to identify each abnormality.

Table 6.1: Results obtained from evaluating two wavelet families and 3 different wavelet levels, accompanied by the PSNR and RMSE values.

Table 6.2 Evaluating the difference between ground truth abnormality size of the spleen versus the calculated value using the *geometric_size* function.

Table 6.3 Six statistical textural measures extracted using the *greycoprops* function.

Table 6.4 The variable names of all features extracted as calculated in the code with the corresponding number of features extracted.

Table 7.1: Table indicating each frame and the corresponding description.

Table 7.2: The snippet of the data frame generated using the large algorithmic framework for HL Patient 4.

Table 8.1: Tabulated description of each test performed within every experiment.

Table 8.2: Demographic and clinical characteristics of the cohort used in experiment 1.

Table 8.3 The variable values used for each test.

Table 8.4 Final overall accuracy when training the model with LOO cross validation principle.

Table 8.5: Demographic and clinical characteristics of the cohort used in experiment 2.

Table 8.6 Final overall accuracy evaluated with the LOO cross validation principle.

List of Abbreviations

CT	Computed tomography
CV	Computer vision
DWT	Discrete wavelet transform
EPTB	Extra pulmonary tuberculosis
HIV	Human immuno-deficiency virus
HL	Hodgkin's lymphoma
IPS	International prognostic score
ML	Machine learning
NN	Neural network
RADLAC	Rapid access diagnostic lymphadenopathy clinic
ROI	Region of interest
SVM	Support vector machine
STAPLE	Simultaneous truth and performance level estimation
TB	Tuberculosis

1. Introduction

Hodgkin's lymphoma (HL) is a highly curable cancer. It is a common cancer in young people and the incidence is growing among young people in Human Immunodeficiency Virus (HIV) endemic regions (Patel et al., 2011; Verburgh and Antel 2019). There are many obstacles to earlier diagnosis of HL in an HIV endemic environment. Imprecise diagnostic pathways, barriers to healthcare access and the overlapping clinical presentation of HL with extra pulmonary tuberculosis (EPTB) results in diagnostic confusion. Incorrect empiric treatment for TB is often administered resulting in a missed opportunity to treat HL at an early and more favourable stage (Patel et al., 2011; Verburgh and Antel 2019). Consequently, there is a need to identify patterns of presentations, and gain crucial insights into image biomarkers, that represent HL in HIV and TB endemic regions. Leveraging knowledge gained from image biomarkers has the potential to improve the diagnostic pathway, and thereby improve patient outcome (Antel et al., 2019).

Ultrasound is the common first point of care imaging modality for patients with a differential diagnosis of HL. In developing countries this initial ultrasound may be the only opportunity for radiologists to detect potential abnormal biomarkers indicating HL. The early identification of imaging biomarkers from abdominal ultrasound determines the extent of splenic and abdominal abnormalities and if detected early may lead to improved prognosis of a patient (Saboo et al, 2012). While more sensitive imaging modalities like computed tomography (CT) enhanced by positron emission tomography (PET) exist for this purpose, they are often more costly to procure, operate and maintain; a cost which is passed on to the patient (Brattain et al., 2019; Griesel et al., 2019). Regardless of the benefits of being non-invasive and cheap, interpretation of ultrasound is subjective and dependent on a sonographer, and or radiologist training and experience. Thus, inadequate interpretation by healthcare personnel may contribute to observational oversights, leading to missed diagnostic opportunities (Walczyk and Walas 2013). The issue is more acute in developing countries where healthcare personnel are often working in strained and under resourced healthcare environments (Walczyk and Walas, 2013).

Automated ultrasound image analysis and machine learning algorithms have the potential to improve HL diagnosis and provide a useful adjunct for sonographers and radiologists in developing countries (Waite et al., 2016; Brattain et al., 2019). This could lead to new standards for characterising abdominal abnormalities which present across numerous disease states. Furthermore, the adjunct could aid clinicians in discovering and synthesising patterns in HL disease progression (Gunčar et al., 2018; Brattain et al., 2019). It may influence the International

Prognostic Score (IPS), a risk stratification tool, to include guidelines for prognosis within a HIV and TB endemic area (Barta et al., 2014). However, to date there has been no report on the use of computer-aided approaches for HL ultrasound-based assessment of the abdomen and spleen (Brattain et al., 2018). Interestingly, automated classification of lymphoma subtypes using pathological images has been reported, but such images may not be available earlier on in the diagnostic pathway (Orlov et al., 2010; Schmitz et al., 2012; Bai et al., 2019).

Thus, considering that a low-cost ultrasound imaging procedure is routinely included when a patient first presents with HL symptoms, it is worth exploring the use of an automated imaging tool to enhance the current diagnostic pathway. Such a tool may enable non-subjective computational characterisations of specific abdominal biomarkers directly from ultrasound images.

1.1 Aims

The first aim of the research project was to develop an algorithmic framework for automatic identification and characterisation of diagnostic abnormalities associated with HL, TB, and HIV, using abdominal ultrasound images. Several ultrasound imaging biomarkers of interest were investigated: splenic enlargement (splenomegaly), splenic lesions, splenic microabscesses, abdominal lymph node enlargement, ascites, and effusions (both pleural and pericardial). Ultrasound images were extracted from two separate cohorts of patients from the Groote Schuur Hospital Haematology patient registry. Cohort 1, consisting of patients with known diagnosis of HL, and cohort 2, consisting of patients referred to the rapid access diagnostic lymphadenopathy clinic (RADLAC) with a differential diagnosis of TB and lymphoma (the validation cohort). **The second aim of the project was to investigate ultrasound abnormality patterns between the HL and RADLAC cohorts using an automated classification framework.**

To achieve both these aims the following four objectives were formulated:

1. Develop a precise prospective imaging protocol by analysing retrospective abdominal ultrasound images and reports of HL patients.
2. Develop geometric and textural mathematical ground truth descriptors for seven abdominal abnormalities.
3. Develop an automated algorithmic abnormality characterisation framework that leverages the identified imaging protocol in objective 1 and the descriptors in objective 2.
4. Evaluation of an automated feature classification model for HL using the automated algorithmic abnormality characterisation framework from objective 3.

1.2 Project overview

The primary aim of the project was to develop the automated algorithmic framework to extract features and characterise abnormalities from abdominal ultrasound images. Machine learning was then employed to identify abnormality patterns between HL confirmed patients and those without HL. Six steps were implemented to achieve this:

The first step, implemented with clinical guidance, was a manual identification of two patient cohorts that met the criteria for inclusion. Using an available patient registry, the first cohort of known HL patients was identified (HL cohort), and available ultrasound images extracted. Only patients who had a confirmed diagnosis of HL close to a diagnostic ultrasound were included. A second cohort of RADLAC patients was subsequently derived from the RADLAC clinic for patients presenting with peripheral lymphadenopathy of unknown etiology, in order to obtain a diagnosis of either HL, or other lymphoma, or TB, or other cancer/infection. Patients who were diagnosed with confirmed HL in the RADLAC clinic were removed to create a validation cohort known as the RADLAC cohort. Thereafter, all available ultrasound images were extracted for this cohort.

The second step was to identify ultrasound frames that enhanced the observation for each abdominal biomarker. This was performed with the guidance of two radiologists. The outcome of this step was a set of common frames across the patient cohort. A secondary outcome was the development of a new imaging sequence or protocol for prospective ultrasound imaging for HL.

The third step was the development of an abnormality detection table, leveraging the identified set of frames from the second step. The visual geometric and textural attributes associated with the biomarkers were manually derived with two radiologists and all data captured into one consolidated table.

The fourth step was the development of ground truth descriptors for the seven abdominal abnormalities of interest. The table developed in the third step was used as a guideline to capture the findings from two radiologists. Each radiologist was blind to the diagnosis of a patient and to each other's work. The radiologists identified the frame of an observed abnormality corresponding to the developed protocol. Each abnormality was then manually segmented and used to develop the ground truth descriptors.

The fifth step employed image processing and feature extraction techniques, first to improve the image quality and ensure features that were extracted were characterising the abnormalities of interest. Secondly to extract geometrical and textural features that can characterise the

abnormalities. Evaluation metrics were used to quantify and measure the performance of methods implemented.

The final step used the extracted features from step 5 to train a machine learning algorithm to classify features representing both cohorts identified for this project. Testing and predictive classification accuracy were the two metrics used to measure the performance of the developed algorithm.

1.3 Scope and Limitations

The research project applied computer vision and machine learning techniques on available ultrasound images. However, five limitations were identified: 1) this was a retrospective study and the researcher had no control on acquisition parameters, such as multiple scanner settings and model types; 2) there may have been an absence of ultrasound imaging biomarkers to highlight the clinical presentation of HL for some patients, because not all HL patients will have disease sites in the abdomen; 3) specific ultrasound frames required to observe an abnormality could have been absent or were not saved when the ultrasound scan was performed; 4) there can be multiple views for every frame identified using the developed imaging protocol but radiologists may only pick one of such frames and lastly 6) development of ground truth descriptors is highly dependent on clinical guidance and assistance.

1.4 Ethical considerations

The Human Research Ethics Committee (HREC) at the University of Cape Town (UCT), (reference number HREC REF: 459/2020) granted ethical approval for the extraction of information from an available patient registry, to generate the two patient cohorts. The patient registry is under the curatorship of Associate Professor Estelle Verburgh, a clinical haematologist at UCT, and co-investigator in the project.

1.5 Dissertation overview

The dissertation is divided into eight chapters. Chapter 2 presents the literature review, highlighting the current gaps in research. Chapter 3 provides a review of the theoretical concepts that underpin the technical methods used in the work. Chapter 4 presents the research methodology, highlighting the data and tools used in the research project to implement the aims and objectives. Chapter 5 presents the first objective on the creation of the ultrasound imaging protocol from a database of patient images. The visual descriptors that are used for the several abnormalities of interest are also presented. Chapter 6 presents the second objective which is the development of geometric and textural mathematical ground truth descriptors. Chapter 7 presents

the third objective on the development of an automated abnormality detection framework. Chapter 8 presents the final objective, to evaluate the performance of an automated model to learn features to characterise and classify HL. Evaluation metrics used to analyse the performance of the model using two experiments. Chapter 9 presents a final discussion of all findings in the context of similar work identified in the literature. Conclusions that could be drawn from the observed findings together with recommendations for future work are also discussed.

2. Literature Review

This chapter presents a review of the clinical characteristics of, and diagnostic approach for, Hodgkin's lymphoma (HL). In addition, computer vision techniques for medical imaging applications are presented. The chapter ends with a summary of the gaps in the current HL diagnostic approach within a tuberculosis (TB) and Human Immunodeficiency Virus (HIV) endemic region. The gaps identified highlight the potential of computer vision techniques to improve the HL diagnostic pathway using image biomarkers.

2.1 Characteristics and diagnostic approach of HL

Lymphoma is a malignancy characterised by the abnormal clonal proliferation of lymphocytes leading to tumour formation in potentially all tissues of the body (Swerdlow et al., 2016). Although lymphocytes are originally produced in the bone marrow and thymus, they circulate in blood via the lymph nodes and spleen to reach the whole body. Lymphocytes are found in all sites in the body, but especially find their home in the lymph nodes and the spleen, which explains why malignant lymphoma transformation would usually occur primarily in either lymph nodes or spleen.

Hodgkin's lymphoma (HL) is a B-cell lymphoma with a high cure rate, if detected early (Pileri et al., 2002). The development of HL at various locations in the body requires the use of biopsies and medical imaging to understand the extent of disease progression. Consequently, the diagnostic pathway for HL is not always clear because of complex clinical expression and presentation (Sathiya & Muthuchelian, 2009; Verburgh & Antel, 2019). The clinical presentation becomes even more challenging in South Africa where HL overlaps with prevalent comorbidities like TB and HIV (Antel et al., 2019; Verburgh and Antel, 2019).

The diagnostic pathway for HL should follow the classical model used in healthcare systems comprising of three stages: 1) the patient experiences symptoms of HL and the clinician performs a physical exam; 2) followed by diagnostic tests incorporating a biopsy of a suspected cancer site in concert with imaging to localise all cancer sites; and 3) treatment is given based on extent of cancer sites. The three stages of an ideal and classical diagnostic pathway are illustrated in Figure 2.1 (Sathiya & Muthuchelian, 2009; Verburgh & Antel, 2019).

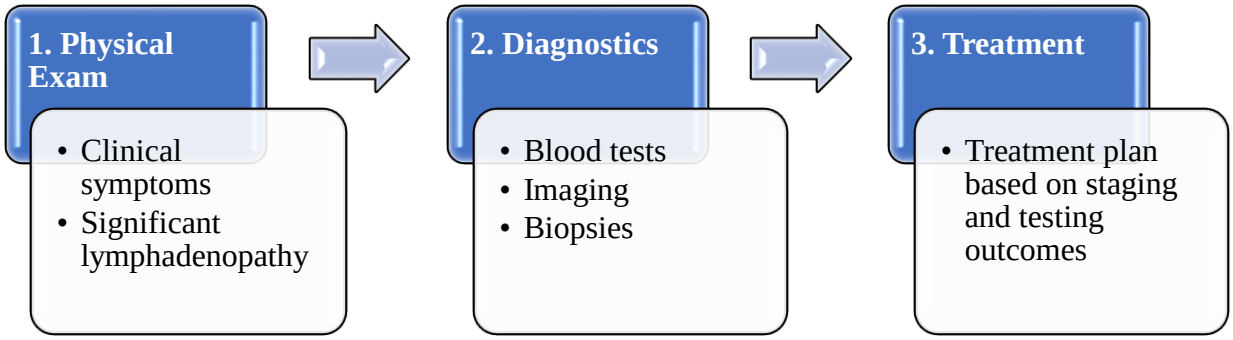

Figure 2.1: A classical and ideal pathway for diagnosis and treatment of HL.

In South Africa, however, this classical diagnostic pathway for HL is derailed by the inordinate focus on seeking a positive TB diagnosis. In the first-place health care workers are unaware of lymphoma as a commonly occurring diagnosis, and secondly, there are overlapping symptomatology and signs between TB and HL. Thus, there is an inherent bias towards seeking a positive TB diagnosis in a TB endemic region such as South Africa. Additionally, inadequate access to sensitive imaging technologies such as computed tomography (CT) compounded by overburdened clinical personnel continually exposed to TB diagnoses, reduces the ability to identify abnormalities that may indicate HL. Furthermore, the abnormalities associated with HL overlaps with those found in patients with TB with the result that radiologists within TB endemic regions may be biased towards a TB diagnosis when interpreting ultrasound images. This contributes towards a deviation from the classical pathway depicted in Figure 2.1 above and delays a potential HL diagnosis. Antel et al (2019) investigated the pathway to lymphoma diagnosis in South Africa and quantified different time intervals contributing to deviations from the typical diagnostic pathway. The authors concluded that a significant contributor to diagnostic delay for HL in South Africa is the *healthcare practitioner interval* or time taken from the first healthcare visit to a diagnostic biopsy, with a median value of 7 weeks; much higher than the acceptable standard of 4 weeks (Antel et al., 2019). Various factors were found to impact the interval, such as the inability of physicians to identify the clinical features that may suggest lymphoma, and the use of inadequate sampling methods for enlarged lymph nodes. Additionally, the suspicion of TB and delay in access to diagnostic biopsies further contributed to the overall diagnostic delay for HL, leading to HL patients being diagnosed in an advanced stage of disease.

2.2 Role of ultrasound imaging in HL

When a patient is first admitted into a hospital, an abdominal ultrasound and X-ray of the chest are frequently used to aid HL diagnosis (Kebede & Getaneh, 2015). In the ultrasound case, a trained sonographer completes the task by following standard imaging protocols to ensure abnormalities

are detected and annotated. A final report is generated to summarise normal or abnormal biomarkers, using appropriate taxonomy, that may guide diagnosis. However, a sonographer may not always consult an experienced radiologist during the process. Additionally, in strained healthcare environments, protocols are often not adhered to. This can lead to a missed opportunity for earlier detection of abnormalities (Geijer & Geijer, 2018).

Bruno et al (2015) reviewed the errors associated with diagnostic radiology and offered potential strategies radiologists should explore to improve their performance. Two interpretation errors were highlighted: *perceptual errors,* where biomarkers are not detected and *cognitive errors*, where biomarkers are detected but their relevance may not be understood. The suggestion was to reduce errors by implementing standardisation of image capture. The authors further suggested using tools that can aid in consistent detection of errors to prevent the potential missed opportunity to detect image relevant markers. Additionally, radiologists should continuously update and expand their knowledge of relevant visual biomarkers and their role across a wide scope of differential diagnostic pathways. These suggestions aim to promote structured and consistent clinical reporting to reduce potential interpretation errors and performance by radiologists (Bruno, Walker, & Abujudeh, 2015).

Splenic and abdominal involvement, which occurs in one third of all HL patients, is a non-specific prognostic finding that may overlap with other comorbidities like HIV and extra pulmonary tuberculosis (EPTB) (Caremani et al., 2013). It is commonly understood that CT provides more accurate assessment of lymph nodes than ultrasound (Białek and Jakubowski, 2017). However, regardless of reduced efficacy compared to CT, ultrasound is more commonly used for assessing lymph nodes because it is readily available and accessible. The nature of the image quality of ultrasound may affect the interpretation of abnormal lymph nodes resulting in incorrect differential disease associations. Additionally, abdominal structures may only indicate partial views of abnormalities further contributing to missed opportunities. Ideally, any abnormal ultrasound finding must be immediately clarified by CT, but in resource constrained countries like South Africa, ultrasound may be the only imaging modality that is readily at hand, and therefore the most important diagnostic imaging tool by which the patient can be assessed for the differential diagnosis of lymphoma, or more specifically, HL. Correctly interpreting the distinctive differentiating patterns found on abdominal ultrasound may improve the diagnostic pathway to differentiating HL from other diseases such as TB in South Africa (Verburgh and Antel, 2019).

Yang et al (1999) used CT and found a potential differentiating pattern between EPTB and HL, based on the anatomical prevalence of enlarged lymph nodes (lymphadenopathy). Lymphadenopathy appeared to occur in the retroperitoneal cavity, with smooth (homogenous) texture for HL patients. In EPTB patients, lymphadenopathy occurred within the mesenteric region of the abdomen (Kebede & Getaneh, 2015; Zhang et al., 2015). Manzella et al (2013) investigated abdominal features for subtypes of lymphoma using CT and confirmed the presence of abnormal lymph nodes in the mesentery as a rare finding in HL patients. Recently Griesel et al (2019), described the combination of three clinical biomarkers in ultrasound that may be more sensitive for EPTB diagnosis in HIV patients. These include lymph node length greater than 1 cm; splenic hypoechoic lesions (dark masses on ultrasound indicating solid components); and effusions due to excess build-up of fluid in either the pericardial or pleural cavity. Charting the extent of lymph node involvement is the cornerstone of staging lymphomatous malignancies and affects the overall prognosis for a patient (Ganeshalingam, 2009). The Lugano classification model is the current standard used in clinical practise for staging lymphoma and is used as a guideline to determine the optimal treatment and response outcomes for patients (Ragheb et al., 2020).

Two image biomarkers that may be used in the HL diagnostic pathway are the presence of an enlarged spleen (splenomegaly), and splenic focal lesions (see Figure 2.2). The presence of splenic lesions can alter the treatment and prognosis of HL patients (Saboo et al., 2012). Although important biomarkers for lymphoma, the radiographic presence of splenic lesions is not specific, in that lesions may be seen in other haematological malignancies, as well as numerous other disease states. Saboo et al (2012) used CT to identify imaging patterns of splenic involvement. The study reported that splenomegaly should not be a biomarker used in isolation, as normal sized spleens may have disease infiltration. The volume of the spleen was discussed as a more sensitive measure to assess the size of the spleen; however, a height measurement of 13 cm is a preferred and common measurement used to identify an enlarged spleen using ultrasound.

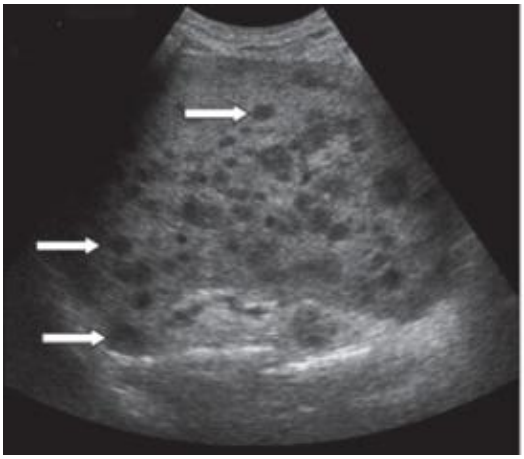

Figure 2.2: Multiple focal lesions (depicted by white arrows) illustrate a high degree of splenic involvement in an HL patient.

Berzaczy et al (2019) reported on a study to evaluate the efficacy of the threshold value of 13 cm recommended for classifying a spleen as enlarged. The study was performed on lymphoma patients in the absence of any infectious diseases. The authors highlighted the 13 cm cut-off as a sensitive measurement to use, only if the measurement is taken vertically, along the craniocaudal diameter. Although the study noted that splenic enlargement does not always imply a patient has lymphoma, it did suggest the marker should be used in conjunction with other radiomic markers.

Schafer et al (2019) performed a systematic review to understand the prevalence of splenic microabscesses as a biomarker for EPTB in HIV patients. The study noted that in HIV patients with confirmed TB, more than half presented with splenic microabscesses. The study also concluded that ultrasound imaging is a useful adjunct for early identification of splenic microabscesses in low resource settings. Leite et al (2007) used magnetic resonance imaging (MRI) to understand the clinical presentation of lymphoma involving extra nodal structures in the abdomen. It was found that extra nodal involvement for all types of lymphoma is common, making specific presentation and characteristics for subtypes difficult. An interesting observation for lymphoma patients is that they may often be immunocompromised, making them more susceptible to fungal or bacterial microabscesses.

Bhatt el al (2015), reported pleural effusions to be a rare finding in HL, although still affecting the overall prognosis for a patient. Conversely, patients with EPTB and HIV have a higher prevalence of pleural effusions, a potential differential to use when examining ultrasound scans (Heller et al., 2012). However, the prevalence of pleural effusions presenting without comorbidities should be a prompt to consider a differential diagnosis of a malignancy (Heller et al., 2012). Reuter et al (2005) studied the prevalence of pericardial effusions in a cohort of South African TB patients. The paper

identified about 70% of HIV patients coinfected with TB presenting with pericardial effusions. In HL, pericardial effusions have not been extensively investigated, but a study conducted by Marks et al (2018) on a cohort of children with HL, indicated that the presence of the biomarker correlated with poor patient prognosis.

Ascites are similar to pleural effusions but are located within the peritoneal cavity (the abdominal lining) and are less prevalent in lymphomas (Liu, Wang, & Yang, 2019). Sinkala et al (2009) reported that ascites were prevalent in HIV patients with TB if accompanied by para-aortic lymphadenopathy. A shortcoming of both studies was the identification and observation of only one patient with lymphoma, potentially indicating a lack of in-depth studies to understand the full role of ascites in HIV, TB, and HL. However, both studies did indicate that ascites and effusions had a negative impact on the prognosis of lymphoma, suggesting the importance of detecting these biomarkers.

The review of various image biomarkers from literature highlighted two shortcomings; biomarkers tend to be evaluated in isolation and used very small patient datasets. Moreover, biomarkers are evaluated within either the sphere of infectious diseases (TB), or in the sphere of malignant diseases (lymphoma), but seldom in datasets containing both disease states. This indicates the uncertainty and current difficulty in identifying exactly which abdominal biomarkers are independently associated with HIV/TB involvement of the abdomen, or identical to, or distinct from the features found in abdominal HL. Tools that extract all details at once may be beneficial to improve diagnostic interpretability, to highlight and identify patterns across HL, TB, and HIV. The findings highlighted in this section are summarised in Table 2.1.

Table 2.1: Table summarising the review of abdominal biomarkers.

Biomarker	Study	Findings
Lymph node	Yang et al (1999) Ganeshalingam (2009) Manzella et al (2013) Kebede and Getaneh (2015) Griesel et al (2019)	i. Measured across short axis (transverse) length ii. Abnormal node: >=1 cm across disease states and hypoechoic[3] iii. Prevalent specific marker in EPTB and HL iv. Diffuse nodules >=1 cm in HL and EPTB v. Retroperitoneal lymph nodes associated with HL vi. Mesenteric lymph nodes associated with EPTB
Splenic lesions	Saboo et al (2012) Griesel et al (2019)	i. Presence indicates abnormality ii. Focal lesions (single/multiple) may be seen in HIV patients coinfected with TB
Splenic enlargement (splenomegaly)	Berzaczy et al (2019)	i. >=13 cm (craniocaudal height) indicates enlargement ii. Enlargement is likely to be because of HL not EPTB if patient is HIV negative
Splenic microabscesses	Leite et al (2007) Schafer et al (2019)	i. Presence indicates abnormality ii. TB or HL (unclear)
Pleural effusion	Heller et al (2012) Bhatt el al (2015)	i. Presence indicates abnormality ii. Prevalent in TB/HIV iii. A rare finding in HL
Pericardial effusion	Reuter et al (2005) Marks et al (2018)	i. Presence indicates abnormality ii. A rare prevalence in HL patients without TB iii. A higher prevalence in TB patients
Ascites	Sinkala et al (2009) Liu et al (2019)	i. Presence indicates abnormality ii. More prevalent in EPTB and HIV patients iii. Rare complication in lymphomas

[3] Appears darker than surrounding area in an ultrasound image.

2.3 Computer vision

Computer vision is a field of science that aims to mimic the human visionary system and extract detailed information from static and sequential images (Forsyth, 2003). Computer vision covers a range of techniques for various applications, such as automated segmentation of images, pattern recognition, or object tracking (Vial et al., 2018). Computer vision techniques aim to analyse the images to extract measurable representations (features) for an application of interest.

2.4 Computer vision and medical imaging applications

Computer vision may aid the development of an automated image processing adjunct for sonographers, radiologists, and clinicians, to automatically extract and represent visual markers of interest and potentially reduce interpretation times (van Timmeren et al, 2020). The first step in medical image processing is the application of techniques to improve the quality of the acquired images. Automated or manual segmentation is then applied, to identify biomarkers of interest. Feature detection algorithms are developed to extract measurable features for each biomarker using sophisticated computational algorithms. A computer may then be taught to learn and classify the developed features. Lastly, evaluation and validation of algorithms are performed to test the derived features (Forsyth et al., 2003; Szeliski, 2010; Timmeren et al., 2020). Figure 2.3 illustrates the steps described (Huang, Zhang, & Li, 2018).

Figure 2.3: A series of steps in a computer vision medical imaging application.

Radiomics is an approach used in medical imaging to extract data *difficult for the human eye to perceive.* (Vial et al, 2018). A review of radiomics by van Timmeren et al (2020) discusses the usability of a radiomics workflow to enhance clinical decision making. The authors outline similar steps to a typical computer vision application illustrated in Figure 2.3 with some additional constraints on the feature selection step. Notably, analysis techniques to select features for the development of feature classification models to exclude non-reproducible features. The applicable points raised in the review are discussed in the relevant sections to follow.

2.5 Ultrasound image acquisition

In resource constrained countries, ultrasound is used as a first point of care diagnostic imaging tool. It performs real-time, low-cost image acquisition without exposing a patient to harmful

ionizing radiation (Brattain et al., 2018; Stewart, 2020). Ultrasound images are generated using high frequency sound waves and may be retrieved from multiple scanners, often without adherence to standard imaging protocols, resulting in variable acquisition parameters (van Timmeren et al., 2020). This creates noisy and low contrast images that may adversely impact interpretation and clinical findings (Brattain et al., 2018).

2.6 Ultrasound image processing

Image processing is a fundamental step used to enhance an image before any detailed information or features are extracted. The images are usually retrospectively analysed with non-standardised protocols and scanner parameters. This creates illumination differences, reducing image quality and limiting further analysis (van Timmeren et al., 2020). Image processing techniques such as denoising and normalisation are often used to enhance the quality of an ultrasound image to further process images (Canuma, 2018; Poudel et al., 2018, Kociołek et al., 2020).

Normalisation techniques aim to mitigate the effect of illumination differences by reducing the intensity range of an image. The review by van Timmeren et al (2020) highlighted the importance of normalising images to reduce intensity variation. One of the easiest methods is to limit the region of interest to a fixed range of intensities. This may create reproducible features when using multiple images for analysis. However, the review suggested the optimal identification of the intensity range may be difficult to choose.

Speckle noise presents as the characteristic granular appearance associated with ultrasound images. The presence of speckle noise reduces the quality of an image, limiting interpretability and further extraction of fine detail. The management of speckle noise is well researched with the K-distribution, Rice distribution, Rayleigh and spatial filters reported to mitigate its effect (Noble, 2009; Gupta et al., 2018). However, research suggests the wavelet transform to be more effective in preserving features within the high frequency bands, whilst keeping the edge and structure of the object intact (Deka et al., 2013). The wavelet is a logarithmic transformation technique in both frequency and time domain, that is used to preserve information within the low and high frequency bands for ultrasound images (Mikhailovich and Tannenbaum, 2006; Kaur et al., 2018).

2.7 Ultrasound segmentation

Delineation of objects through algorithmic segmentation techniques creates boundaries to isolate only a particular region of interest (ROI) (Noble, 2009). Edge (contour) and region-based computer-based approaches are the two most widely used segmentation algorithms to calculate shape, area, and volume measurements for objects of interest (Linguraru et al., 2010).

Manual segmentation is time consuming and subject to intra and inter-operability errors (Withey and Koles, 2007). To mitigate these errors the simultaneous truth and performance level estimation (STAPLE) algorithm is used, to create a probabilistic estimate and objective ground truth. Multiple expert manual segmentations are fed into the STAPLE algorithm to develop one combined delineated object from which computer-based features can be derived (Warfield et al., 2004). However, the algorithm underperforms in cases where only two expert segmentations are available, negatively influencing the optimal segmentation estimate (Van Leemput and Sabuncu, 2014). In the absence of multiple expert clinicians and large datasets, manual segmentations and annotations may be beneficial for better representations of biomarkers (Krig et al., 2014; Forghani et al., 2019).

Automated segmentation using machine learning algorithms has improved consistency of segmentations to reduce errors encountered and time taken when performing the task manually. However, challenges exist when these algorithms are used to segment ultrasound images compared to the more expensive imaging modalities such as CT (Brattain et al., 2018). Noticeably, imbalanced datasets, variable data quality, limited availability of shape priors for ultrasound imaging markers and inconsistent annotations. Moreover, the experience of a sonographer or radiologist may affect the consistency of ultrasound image acquisition. This may produce inadequate views, restricting standardised representations (Brattain et al., 2018). As image acquisition and quality of images produced from ultrasound scanners improves, automated segmentation may be the preferred method to generate consistent, robust delineation of objects or biomarkers of interest (Brattain et al., 2018).

Despite the range and advancement of techniques for ultrasound segmentation, literature highlighting the application of these techniques for the spleen and abdominal abnormalities are scarce (Brattain et al., 2018). Automated segmentation of ultrasound images lacks the availability of shape priors and access to large well annotated datasets, limiting the use of machine learning automated segmentation algorithms. Manual segmentation may be a preferred method to initiate standardised representations of biomarkers to develop robust ultrasound shape priors, usable in automated algorithms (Saini et al., 2010; Brattain et al., 2018; Zheng et al., 2018).

2.8 Ultrasound feature detection

Medical imaging classification models are built using handcrafted features, clinically driven by radiologists' insights. User developed algorithms may be used to detect features of interest or alternatively automatically detected using machine learning algorithms. To create a gold standard

or ground truth more than one skilled radiologist performs the task of characterising features for a particular application, reducing subjectivity and bias (Brattain et al., 2019).

Afshar et al, 2019 reviewed the challenges and opportunities that exist between a user designed detection algorithm versus a deep learning-based approach that automatically detects features directly from large high dimensional medical images. The authors discuss the opportunities and pitfalls of extracting features with clinical guidance (*handcrafted features*) or deep learning approach (*deep features*) with the latter given more focus because for its ability to extract information that may not be detectable by a human. This approach has shown to have the potential to provide generalisable models for medical imaging applications. However, the lack of interpretability of features detected using deep learning approaches advocate for handcrafting features with clinical guidance. A noticeable drawback of deep learning-based feature extraction methods are the need for large datasets which may or may not be annotated and consistent. Deep learning approaches currently lack the ability to explain why particular features are used to infer predictions for image classification models. Moreover, they cannot relate the features detected with clinically relevant concepts limiting the interpretation of a deep learning computer-based model for feature extraction.

Geometric and textural analysis is typically used by radiologists to identify and characterise visual biomarkers (Afshar et al., 2019). Typically, size and shape features are used to quantify geometrical properties of a biomarker (Afshar et al., 2019). You et al (2014) suggested the use of a contour to describe the shape of an object for modality classification. The classification model achieved an accuracy of 74% when using a contour feature compared to other types of individual descriptors such as colour and texture. van den Heuvel et al (2018) investigated the development of an automated tool to measure fetal heads from ultrasound for monitoring the growth of a fetus. The tool first identified the fetal skull and then extracted features to identify the elliptical shape characteristic of a fetal head. This was fed to a classification model to test extracted features. Using a large dataset compared to other studies, they achieved results comparable to experienced sonographers' manual segmentations of the fetal skull.

Textural analysis is not a new technique and has been used to identify lesions in image-based classification models (Haralick et al., 1973; Garra et al., 1993). The models evaluated in literature, use textural descriptors to classify breast and liver lesions in ultrasound, but are not exclusively applied to HL diagnosis (Brattain et al., 2019). Subramanya et al. (2015) developed a liver disease classification model with results achieving an average classification accuracy of ∼86%. Xu et al

(2019) also used the texture extraction for automated liver disease classification model, but characterised texture using the gray-level-occurrence matrix (GLCM). The model achieved an accuracy of ∼76%. Gómez-Flores and Ruiz-Ortega (2016) explored a fully automated approach using texture to segment and identify breast lesions. A Gabor filter bank was used to identify texture patterns from ultrasound and results were evaluated against studies using similar approaches. The outcomes were accurate segmentations of different types of breast lesions present with little computation time.

Literature has highlighted the benefit of implementing the combination of a Gabor filter and the GLCM to improve the accuracy of texture-based classification models (Tou et al., 2007; Lahmiri, 2013). Furthermore, the results seen in literature highlight the ability of classification models to learn from textural features. However, models for classification of splenic abnormalities seen in HL, TB or HIV patients have not been exclusively researched (Xu et al., 2019; Brattain et al., 2019).

In summary, the literature highlights promising results for automated detection systems in a clinical setting. Automation has been achieved through using a combination of geometric and textural descriptors to represent the visual biomarkers typically used by clinicians and radiologists (Brattain et al., 2019). Contour detection has been the tool of choice to develop shape and size descriptors. The application of gray level co-occurrence matrices have shown promise for medical classification applications (Humeau-Heurtier, 2019). Notably the success of breast tumor and liver lesion texture extraction using GLCM proves the viability of the technique. However, none of the above approaches are currently used to improve the diagnostic pathway for HL.

2.9 Machine learning in ultrasound imaging classification

Machine learning approaches aim to teach a computer to perform certain tasks by learning from experience. Supervised machine learning is frequently applied to ultrasound to build computer aided diagnostic tools or classification and segmentation models (Foster et al., 2014; Erickson et al., 2017; Brattain et al., 2018). Deep learning differs from traditional machine learning as it can handle large amounts of high-dimensional data. Classification systems using deep learning algorithms utilise the neural network (NN) algorithm for feature learning (Huang et al., 2018). However, reviewing existing literature there are a vast number of ultrasound imaging classification models built but few studies use large ultrasound repositories for deep learning models and rarely are studies specific to the abdominal region (Brattain et al., 2018). Interestingly a large majority of research implements supervised machine learning approaches. The approach achieves

consistent and high accuracy prediction performance when learning with smaller feature sets (Huang et al., 2018; Brattain et al., 2018).

Brattain et al (2018) provides a comprehensive review on machine learning approaches for ultrasound applications. The authors focus on detection and classification of lesions in the breast and liver. One challenge identified by the authors was imbalanced datasets, as ultrasound is often a first point of care imaging modality; not always used for identifying markers of disease. Due to inherent subjectivity in interpretation of ultrasound images, there also exists a scarcity of large, well annotated data. The authors conclude that improving current image acquisition methods and image quality control standards may facilitate increased use of machine learning for ultrasound applications.

Support vector machine (SVM) learning algorithms have been proposed in literature to learn and detect features from medical images to minimise operator dependency and improve diagnostic accuracy, which are common findings associated with ultrasound evaluation. The SVM algorithm has been widely investigated for ultrasound imaging-based classification models. Yu et al (2015) used the SVM to identify the optimal point of entry for needle insertion, based on ultrasound images of the spines of pregnant women requiring epidurals. Nascimento et al (2016) developed an SVM machine learning algorithm able to detect breast lesions with a mean accuracy of ~88%. Sjogren et al (2016) characterised free fluid in the abdomen using an SVM classifier and achieved a 95% sensitivity rate. Across all studies a high classification rate was achieved indicating the utility of automated machine learning based classification models using ultrasound. Thus, there is an opportunity to leverage similar approaches to enhance ultrasound based differential diagnosis.

2.10 Summary of literature review

Techniques from the field of computer vision have aided research and development of computer aided diagnostic systems. Typically, visual descriptors from medical images are transformed into measurable features to provide robust and unbiased image interpretation. However, these techniques have not been explored to improve the differential diagnosis of HL (Brattain et al., 2018; Verburgh and Antel, 2019). The exploration of HL biomarkers in the abdomen may provide a means to improve differential diagnosis using automated models.

While state of the art deep learning models exist to extract features from large datasets for classification tasks, they require the availability of well annotated datasets. Large datasets of ultrasound images are not currently available for HL patients as machine learning based approaches are widely applied to more prevalent global diseases (Brattain et al., 2018). In addition,

to broaden the diagnostic accuracy of deep learning models, suitable datasets should contain sufficient diversity in disease states. Machine learning approaches may therefore be promising, but most of the literature has thus far focussed on breast and liver tissue. The creation of a platform for repeatable and robust biomarker identification, may enhance the differential diagnostic imaging pathway in HL, particularly in HIV and TB endemic South Africa.

3. Theoretical considerations for ultrasound image processing

This chapter describes the theoretical concepts relevant to the research methodology followed in this project. First the concept of ground truth image data is introduced, followed by an overview of medical image data with focus on ultrasound imaging. Image processing techniques relevant for noisy ultrasound images are then outlined. These include methods to denoise data, normalise data, and finally those used to preserve features. Extraction of features from ultrasound using mathematical and statistical techniques follows thereafter. Finally, the chapter concludes with an explanation on support vector machines and how they are used for automated classification applications using ultrasound images.

3.1 Ground truth image data

Ground truth data is an essential component used in the development of automated imaging applications (Krige, 2014). In computer vision tasks, ground truth data is generated by an expert. Robustness of ground truth data is extremely important in medical applications. For imaging analysis applications, medical experts use their clinical knowledge to annotate and generate a set of ground truth images. The ground truth is used to develop and evaluate automated medical imaging applications (Krig, 2014). A popular algorithm for estimation of ground truth is the simultaneous truth and performance level estimation (STAPLE) algorithm. The algorithm aims to provide a consensus for multiple clinical experts' manual segmentations, thereby creating a probabilistic ground truth estimate for downstream processing tasks (Warfield et al., 2004). However, a less computationally efficient method of summing binary masks over the area of combined segmentation overlap, normalised by the number of clinical experts involved, may be used to generate a ground truth segmentation, (Deeley et al., 2011).

Time consuming manual segmentation remains the current gold standard for producing robust automated image classification models for ultrasound. Typically, this requires large sets of well annotated images. However, maintaining the required quality of manually derived ground truth data and standardising the task particularly for ultrasound imaging remains challenging. Ground truth data of insufficient quality impacts the performance of automated models developed against such data (Strohm et al., 2020). Despite the potential shortcomings described above, manually derived ground truth data extraction is often necessary in exploratory work. As such, research is ongoing to improve standards in developing ground truth image data (Kohli et al., 2017).

3.2 Medical image/ultrasound image

Ultrasound imaging is non-invasive and low-cost but is inherently noisy, producing low quality images that may impact diagnostic accuracy. This necessitates the application of pre-processing techniques to ultrasound images prior to use in clinical applications (Gupta et al., 2018). Intensity normalisation is a pre-processing technique used to mitigate intensity variation and limit the effects of uneven and varying illumination often encountered in ultrasound imaging. The effects may be more pronounced depending on scanner settings during image acquisition (Om et al., 2012) and the intensity normalization functions to create more robust textural features, which can improve downstream accuracy in automated ultrasound imaging applications (Kociołek et al., 2020).

3.3 Wavelet transform

The wavelet transform is used to analyse a signal (image). It is a multiresolution technique that aids preservation of textural features that may reside within the high frequency band of ultrasound images (Prabusankarlal et al, 2017). Wavelet transforms have been shown to keep the edge and structure of the object intact while removing speckle noise associated with ultrasound images (Mikhailovich and Tannenbaum, 2006; Kaur et al., 2018).

A wavelet transform converts the signal (image) through a series of translations and dilations in the time and frequency domains. Approximation and detail coefficients are extracted to represent the signal, corresponding to low and high frequencies, respectively. Wavelet packets are used to apply further levels of signal decomposition to improve resolution and separation of noise from the true signal (Jin et al., 2005). Wavelets are chosen to meet a particular requirement whilst using the least amount of processing time to transform the signal.

There are different types of wavelets or wavelet families, each one exhibiting a different characteristic. Common types are the Daubechies, Coiflet, Haar and Symlet (Choi and Jeong, 2020). The Daubechies wavelet is used because of its ability to represent polynomial behaviour. Conversely, the Haar wavelet is a piecewise approximation which can be more computationally efficient. Coiflet and Symlet wavelets are similar to the Daubechies, but generate smoother signals (Jin et al., 2005, Stolojescu-Crisan et al., 2010). An important factor for wavelet choice is the vanishing moment, which affects the smoothing attribute of a wavelet. A higher moment is often favoured as it creates a large and sparse representation of the high frequency bands of a signal, where noise is often found, particularly for ultrasound images. Ultrasound image quality degrades in the presence of speckle noise (present in the high frequency bands) and may corrupt the true signal. Therefore, methods to efficiently ‘spread’ the signal can aid removal of noisy components

to retain only the true signal of the image. Figure 3.1 provides a visual representation of the different wavelets described (Stolojescu-Crisan et al., 2010). The Haar wavelet's piecewise approximation is clear compared to the Symlet and Coiflet which exhibit smoother signal approximations.

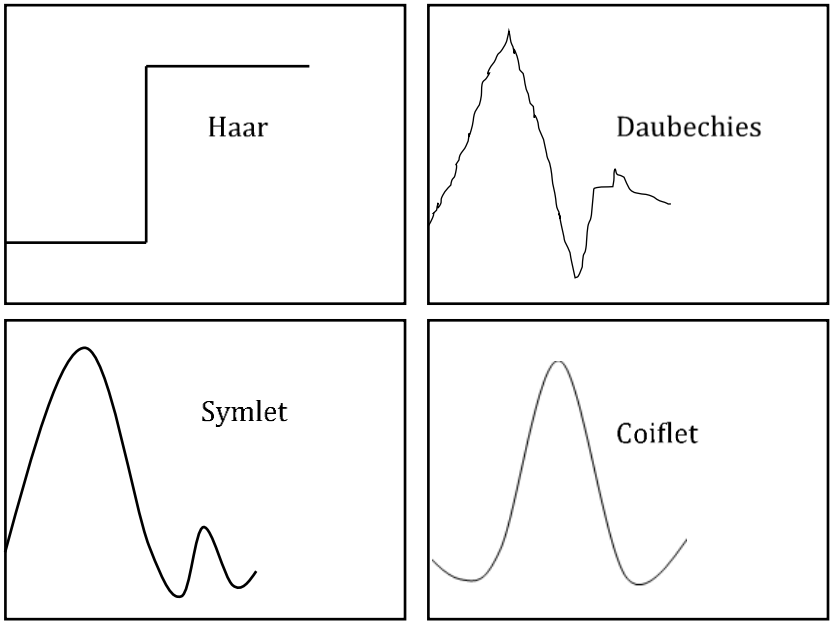

Figure 3.1: Different types of wavelets used to decompose a signal.

3.3.1 Discrete wavelet transform

The discrete wavelet transform (DWT) is used in ultrasound image processing as a preferred method to retain both spatial information and textural properties whilst reducing speckle noise (Graps, 1995; Kaur et al., 2018).

The *mother* wavelet characterises the basic wavelet shape (Φ) and is defined by:

$$\boldsymbol{\Phi_{s,l}(x) = 2^{-\frac{s}{2}}\,\Phi(2^{-s}x - l)} \tag{3.1}$$

Where the *s* denotes the scale (width) and *l* the dilation (position). A scaling equation is then applied to the mother wavelet for different resolutions as defined by:

$$\boldsymbol{W(x) = \sum_{k=-1}^{N-2} (-1)^k\, c_{k+1}\Phi\,(2x + k)} \tag{3.2}$$

The scaling function $W(x)$ now represents the mother wavelet Φ in terms of wavelet coefficients c_k. These wavelet coefficients are constrained by:

$$\sum_{k=0}^{N-1} c_k = 2 \text{ and } \sum_{k=0}^{N-1} c_k c_{k+2l} = 2\delta_{l,0} \tag{3.3}$$

where δ is the delta function or generalised function used to model an impulse and constrain the wavelet.

The wavelet coefficients are chosen to meet an application specific requirement and are stored in a transformation matrix. When the matrix is applied to the original signal, a new signal representation is generated with higher resolution. The coefficients act as iterative filters, which are constrained within a user-specified threshold, creating a smoothing effect to reduce noise whilst retaining detailed information. The coefficients are effective in identifying which parts of the signal have high and low signal to noise ratio (SNR). The coefficients representing low SNR values (meaning more noise) are removed by applying thresholds and then reconstructed using the inverse DWT to ensure the true signal is isolated from the noise (Yadav et al., 2015). The outcome, after the thresholds are applied, is a smoother image reconstructed from wavelet coefficients; or a removal of small details representative of noise (Graps, 1995; Kaur et al., 2018).

3.3.2 Two-dimensional discrete wavelet transform (2D DWT)

When applied for image denoising, the two-dimensional DWT decomposition uses a one-dimensional DWT along the rows (horizontal filter) of an image, and then a one-dimensional DWT along the columns (vertical filter). Four sub-bands are produced, *low-low (LL1), low-high (LH1), high-low (HL1) and high-high (HH1)* (Yadav et al., 2015). The LL1 sub-band represents the coarse level coefficients, while the LH1, HL1 and HH1 sub-bands representing the fine scale coefficients. Depending on the application, the LL may be further reduced, to create more levels of decomposition. Figure 3.2 illustrates a single level (image on the left) versus a two-dimensional (2D) wavelet (image on the right) sub-band decomposition (Anutam and Rajni et al., 2014; Yadav et al., 2015). Each new level of decomposition introduced will identify four new sub-bands to represent the coarse or LL coefficients.

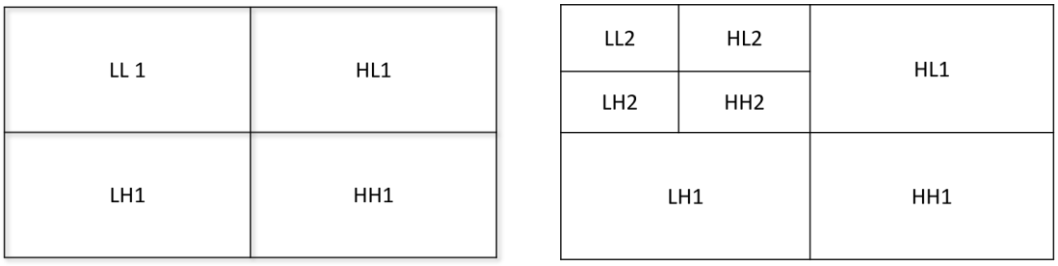

Figure 3.2 Illustration of different levels of decomposition when a two-dimensional DWT is applied to an image. On the left an image with a single level decomposition creating four sub-bands. The image on the right has two levels of wavelet decomposition, creating four new sub-bands to further decompose the LL1 sub-band.

3.3.3 Wavelet thresholding methods

Thresholding wavelet coefficients is an effective method for identifying the true signal hidden within noisy data (Graps, 1995). The technique is applied after the two-dimensional DWT is applied to an image. The coefficients that are not part of the true signal are removed or modified based on a user-selected threshold value. The ideal threshold value is chosen by evaluating the outcomes from three steps, 1) estimating the noise variance in an image; 2) implementing a threshold and removing unwanted ranges of coefficients based on the estimated noise; 3) reconstructing the image with the inverse DWT and evaluating the new signal representation. These steps are iteratively completed until the reconstructed image visually resembles the true signal and the noise has been sufficiently removed (Anutam and Rajni et al., 2014). Soft and hard thresholding techniques are used to denoise an image. An ideal threshold is found by identifying an optimal peak signal to noise ratio (PSNR). Equation (3.4) represents hard thresholding and equation (3.5), soft thresholding. Figure 3.3 is the graphical representation of the two thresholding equations (Anutam and Rajni et al., 2014).

$$\boldsymbol{D}\,(\boldsymbol{U},\boldsymbol{\lambda}) = \boldsymbol{U}\ \boldsymbol{for\ all}\ |\boldsymbol{U}| >, \boldsymbol{\lambda} \tag{3.4}$$

$$\boldsymbol{D}\,(\boldsymbol{U},\boldsymbol{\lambda}) = \boldsymbol{sgn}(\boldsymbol{U}) * \mathbf{max}\,(\mathbf{0}, |\boldsymbol{U}| - \boldsymbol{\lambda}) \tag{3.5}$$

where D is the representation of wavelet coefficients, U the estimated wavelet coefficients after decomposition and λ is the threshold value.

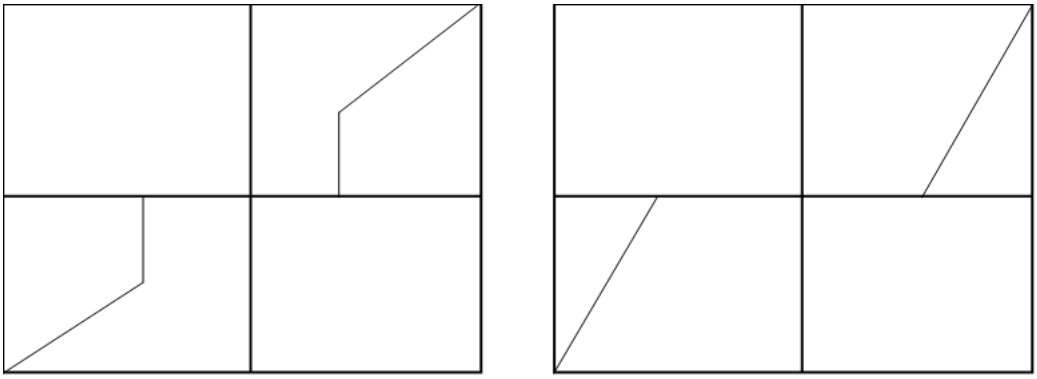

Figure 3.3 Graphical representation of hard (left) and soft (right) thresholding.

Hard thresholding is discontinuous and creates abrupt or sharp changes when reconstructing the denoised signal. Soft thresholding is therefore preferred when reducing significant unwanted noise in imaging applications whilst preserving and retaining the visual representation (Chang et al., 2000). Thresholding is performed only on the coefficients that represent the detailed sub-bands. The two frameworks used to model thresholds are BayesShrink and VisuShrink (Hedaoo and Godbol, 2011).

BayesShrink assumes a Gaussian distribution and is defined by the threshold equation below (Anutam and Rajni et al., 2014):

$$\boldsymbol{t_B = \sigma^2 / \sigma_s} \tag{3.6}$$

where t_B is the Bayes threshold, σ^2 is the noise variance, and σ_s the signal variance without noise.

Equation (3.6) is transformed to solve for σ_s using the definition of additive noise, represented by the equation below:

$$\boldsymbol{\sigma_s = \sqrt{\max(\sigma_w{}^2 - \sigma^2, 0)}} \tag{3.7}$$

where $\sigma_w{}^2$ represents the additive noise variance.

The VisuShrink threshold is defined by the equation below:

$$\boldsymbol{t_V = \sigma\sqrt{2\log N}} \tag{3.8}$$

where t_V is the Visu threshold, N is the size of the original image and σ is the noise variance.

Studies have shown that the wavelet transform is more effective in preserving features within the high frequency bands, whilst keeping the edge and structure of the object intact which is critical for noisy ultrasound images (Deka et al., 2013). The wavelet technique transformation in both frequency and time domain is therefore an effective tool to preserve information within the low

and high frequency bands for ultrasound images (Mikhailovich and Tannenbaum, 2006; Kaur et al., 2018).

3.4 Feature extraction in medical images/ultrasound

3.4.1 Textural features

Texture is an essential visual descriptor used in radiological interpretation of medical images. However, identification of textural features varies with a clinician's experience and visual capability (Afshar et al., 2019). In biomedical imaging applications texture classification models use textural features extracted automatically from an image. The features are developed using mathematical descriptors creating measurable quantities that describe the spatial interaction and relationship of pixels (Humeau-Heurtier, 2019).

The Gabor filter and gray level co-occurrence matrix (GLCM) are techniques used to extract first and second order statistical texture features in multiple directions from ultrasound (Xu et al., 2019). First order statistical measures identify individual pixel value characteristics, in contrast to second order measures which analyse the relationship between pixels. Tou et al (2007) highlighted the benefit of implementing a combination of both a Gabor filter and GLCM, to extract accurate textural features for texture classification models. The study highlighted that the Gabor filter can be used to extract textural measures at a specific frequency and has been found to be more effective when supplemented with the GLCM to incorporate texture analysis by developing features to represent the spatial relationships of actual pixels values. Xu et al (2019) demonstrated the use of incorporating textural features from ultrasound using the GLCM and Gabor for a breast lesion diagnosis system. The study used three-dimensional ultrasound images to highlight the improved accuracy of the systems once these features were included into the model.

3.4.2 Gabor Filter

The Gabor filter aims to model the way in which humans perceive texture (Gómez-Flores and Ruiz-Ortega, 2016). Human vision identifies distinct spatial arrangements to develop localised patterns that can be correlated with one another to describe texture (Lahmiri and Boukadoum, 2013). A similar effect can be achieved by using multiple Gabor filters (banks) to generate statistical measures for texture to identify specific content by varying frequencies, scales, and orientations of an image. A 2D Gabor filter is a combination of a sinusoidal signal modulated by a Gaussian wave represented by equation (3.9). The filter has both a real and imaginary component (Bianconi and Fernández, 2007; Lahmiri and Boukadoum, 2013).

$$\boldsymbol{\psi(x,y) = \frac{F}{2\gamma\eta}\, e^{-\mathrm{F}^2\left[\frac{x'}{\gamma}+\frac{y'}{\eta}\right]} e^{i2\pi F x'}} \tag{3.9}$$

with $x' = x\,cos\theta + ysin\theta\ and\ y' = -xsin\theta + ycos\theta$.

where F is the central frequency of the filter, θ the orientation of the filter, γ, η are smoothing parameters. The variables x, y represents the images pixel position and collectively the values of F, θ, γ, η will determine how an image is analysed in the spatial and frequency domains (Bianconi, and Fernández, 2007).

The 2D Gabor filter is flexible in representing patterns of texture. Varying the parameters values F, θ, γ, η creates different Gabor filters corresponding to the regions of interest needing to be analysed for a specific application (Lahmiri, 2013). However, Bianconi and Fernández (2007) note that increasing frequency and orientations may not have a significant effect on texture classification compared to smoothing parameters.

The use of Gabor filters for extracting textural features in ultrasound images has resulted in improved performance of automated classification models, with high performance still achieved with noisy medical images such as ultrasound (Tamilselvi, 2010; Liu et al., 2014).

3.4.3 GLCM

The GLCM is developed by creating a *nxn* matrix based on the number (*n*) of gray levels present in an image. The matrix is populated by counting the number of occurrences of pixel values with the same intensity value at a certain angle and distance apart (Humeau-Heurtier, 2019).

The GLCM is defined by (Xu et al., 2019):

$$\boldsymbol{C_{d,\theta}(i,j) = \sum_{x=1}^{N}\sum_{y=1}^{M} \begin{cases} 1, if\ l(x,y) = i, I(x+\Delta x, y+\Delta y) = j \\ 0, otherwise \end{cases}} \tag{3.10}$$

where $C_{d,\theta}(i,j)$ represents the number of occurrences for a pair of gray levels i and j, at distance d and angular distance θ , apart. I represents the intensity values in the x^{th} row and y^{th} column of an image.

Equation (3.10) may not always produce a value if an occurrence of a particular combination does not exist; this may create poor approximation. To mitigate this phenomenon, the gray level intensity values are normalised, reducing the size of the GLCM. (Liu et al., 2014). A probabilistic

function defined by equation (3.11) is then applied once the intensity levels are normalised to create a new GLCM representation given by equation (3.12) (Xu et al., 2019).

$$\boldsymbol{P_{i,j}(i,j) = \frac{C\,(i,j)}{\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} C\,(i,j)}} \tag{3.11}$$

$$\boldsymbol{G = \begin{bmatrix} P(0,0) & \cdots & P(0,l-1) \\ \vdots & \ddots & \vdots \\ P(l-1,0) & \cdots & P(l-1,l-1) \end{bmatrix}} \tag{3.12}$$

A set of common θ values used are $0°, 45°, 90°, 135°$ with a distance $d = 1$ pixel apart, to develop a spatial distribution representation comparing the relationship of two neighbouring pixels. Figure 3.4 illustrates the use of four θ directions; in each case a separate GLCM matrix is created for analysis (Xu et al., 2019).

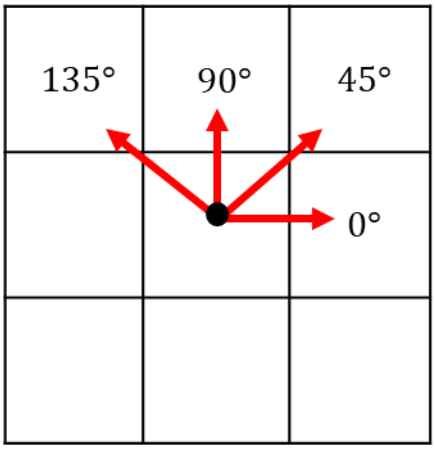

Figure 3.4 A visual representation to indicate the four directions and single pixel distance used to build a GLCM.

The GLCM has been used for automated ultrasound imaging applications with reported increase in accuracy for segmenting breast lesions (Gómez-Flores and Ruiz-Ortega, 2016). Its ability to discern texture differences makes it a useful tool for ultrasound applications where texture discrimination is important for classifying potential disease states.

3.4.4 Textural feature extraction

Harlick proposed a now widely used protocol for extracting textural properties from an image using fourteen measures. Each textural measure represents a condensed representation of a computed matrix such as the GLCM (Harlick et al., 1973). A recent study highlighted that the use of just five of the fourteen measures may be sufficient to represent and characterise texture in images in order to build texture classification models (Humeau-Heurtier, 2019). Seven of the most widely used Harlick features, and their associated equations are now presented.

The first textural feature is **Contrast,** which measures the intensity variations between pixels in an image at d distance, and at θ, angle apart. It is calculated using:

$$\boldsymbol{Contrast} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1}|i-j|^2 P_{d,\theta}(i,j) \tag{3.13}$$

where i and j represents a pair of gray levels at a distance d and angle θ apart (Humeau-Heurtier, 2019).

Dissimilarity is a textural feature used to describe the uniformity of pixels in an image identifying the linearity of intensity variations and calculated using:

$$\boldsymbol{Dissimilarity} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1} P_{d,\theta}(i,j) \tag{3.14}$$

Energy is a textural feature are used to describe the uniformity of pixels in an image and calculated using:

$$\boldsymbol{Energy} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1} P_{d,\theta}(i,j) \tag{3.15}$$

Entropy is a textural feature measuring of disorder of an image. If an image is not uniform, then the calculated values of the GLCM would be small and produce a large entropy value indicating disorder or non-uniformity. The feature is calculated using:

$$\boldsymbol{Entropy} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1} P_{d,\theta}(i,j).\log P_{d,\theta}(i,j) \tag{3.16}$$

Correlation describes linear dependency by identifying how a single central pixel relates to surrounding pixels. Correlation is calculated using:

$$\boldsymbol{Correlation} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1} P_{d,\theta}(i,j)\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \tag{3.17}$$

Homogeneity is a textural feature used to describe the distribution of pixel intensities. A high value indicates an absence of variation, and little to no variation in intensity of pixels. The feature is calculated using:

$$\boldsymbol{Homogeneity} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\frac{1}{1+(i-j)^2} P_{d,\theta}(i,j) \tag{3.18}$$

Angular second moment (ASM) is a textural feature used to measure the local uniformity in an image and calculated using:

$$\boldsymbol{ASM} = \sum_{i=0}^{l-1}\sum_{j=0}^{l-1} P_{d,\theta}(i,j).\log P_{d,\theta}(i,j) \tag{3.19}$$

The textural features and methods described in this section to extract features from images, have been used to build automated ultrasound classification models. Xu et al (2019) demonstrated the ability of an ultrasound classification model to differentiate and identify liver disease. The study evaluated textural feature as inputs into the model extracted from a GLCM and obtained performance accuracies of 90%. Interestingly, the application only required a few textural measures or features to provide a high classification accuracy.

3.4.5 Geometric feature extraction

Shape and intensity descriptors are geometric properties used to represent a region of interest (ROI) (Afshar et al., 2019). The descriptors are derived from measures of the distribution of pixel intensities. Contour detection is used to estimate the shape and size of an object and has been used to extract automated measurements from 2D ultrasound images (You et al., 2014). Once a contour is found using morphological operations, the pixels that lie within the contour are counted. The pixel values are converted to an actual physical measurement for an object of interest using a pixel spacing ratio, usually provided as meta data on medical images (Thomas et al., 1999; Calvo-Lobo et al., 2018).

3.5 Classification using medical/ultrasound images

The support vector machine (SVM) uses statistical theory to develop a model that learns how to separate binary or multi-class data with an optimal hyper-plane. In the case of non-linear data, a kernel function is used to map inputs into a high dimensional space, to try and a find a linear separation of classes. Multiple kernels may be combined to improve accuracy and separation of data for classification models (Dioşan et al., 2012). A supervised SVM model is developed by learning from labelled inputs and outputs. Data is split into training and test sets (Dioşan et al., 2012). In a binary classification model, training data is modelled by the equation (Kecman, 2005):

$$(\boldsymbol{x_1},\boldsymbol{y_1}),(\boldsymbol{x_2},\boldsymbol{y_2}),\ldots..(\boldsymbol{x_l},\boldsymbol{y_l}),\boldsymbol{x} \in \boldsymbol{R^n},\boldsymbol{y} \in \{-\boldsymbol{1},+\boldsymbol{1}\} \tag{3.20}$$

where $y \in \{-1,+1\}$ represents the binary class output labels for input data $x \in R^n$.

In a 2D space (n=2), the *decision boundary* or hyperplane to partition two classes is defined by (Kecman, 2005):

$$\boldsymbol{w_1 x_1 + w_2 x_2 + b = 0} \tag{3.21}$$

The aim is to identify which hyperplane optimally separates the data classes (Kecman, 2005). To develop a model that generalises well to unseen data, the hyperplane should have a large margin (*M*) separating the two classes with minimal training error (Kecman, 2005). Figure 3.5 illustrates this concept with the figure on the left having a small margin and the figure on the right, a large margin. The larger margin will be better at discriminating between classes when processing new data (Kecman, 2005).

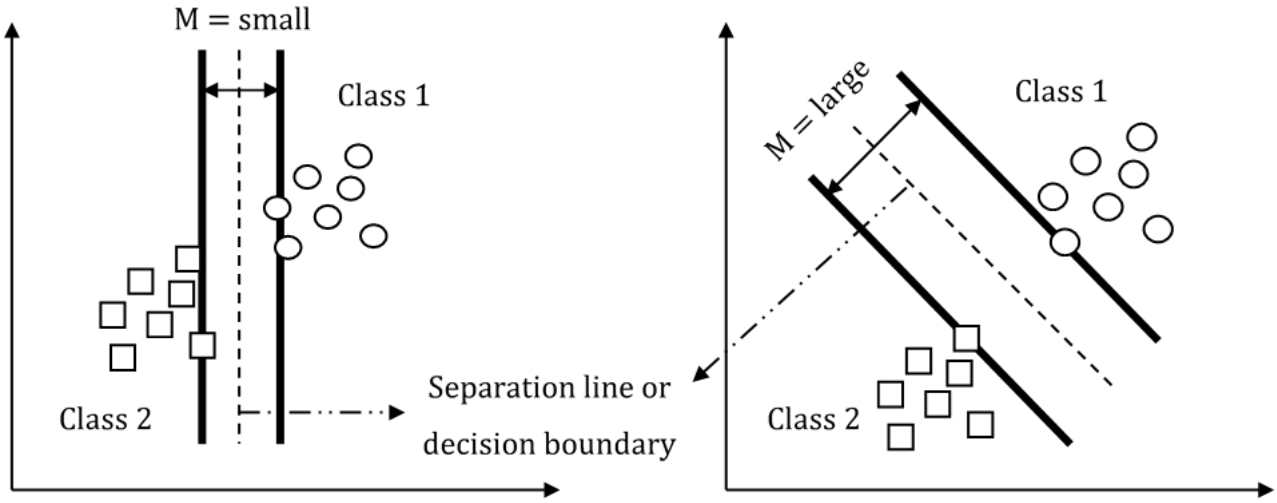

Figure 3.5 A visual illustration of two hyper-planes chosen, one with a small margin (left) versus a larger margin (right). The image on the right illustrates a better choice for a hyperplane to discriminate between classes to develop a model that generalises well on unseen data.

The training data is used to develop learning parameters, optimised to manage the trade-off between maximising margins for accurate classification and high error rates (misclassification) (Prochazka et al., 2019). The parameters *w* and b, are used to represent the *decision function* or *discriminant,* (Kecman, 2005) and defined by:

$$d(x, \mathbf{w}, b) = \mathbf{w}^T x + b = \sum_{i=1}^{n} w_i x_i + b \tag{3.22}$$

where $\mathbf{w} = [w_1, w_2 \ldots . w_n]$ and b is the bias. When $d(x, \mathbf{w}, b) = 0$, this represents the separation boundary (dotted line in Figure 3.5).

Identifying the optimal decision function to discriminate between two classes is a challenging task. Each training set produces a support vector to aid the optimal margin between classes, however an optimal boundary to maximise the margin between classes needs to be found. A *canonical hyperplane* is therefore used to represent the optimal boundary between support vectors to ensure a maximal margin is obtained for binary class separation (Kecman, 2005). The canonical hyperplane is defined by equation (3.23) below, for every training set x ∈ X.

$$\mathbf{min}\,|\mathbf{w}^{T}\boldsymbol{x_i} + \boldsymbol{b}| = \mathbf{1} \tag{3.23}$$

Using equation (3.22) and the constraints for support vectors represented by the canonical hyperplane equation (3.23), a new representation of the decision function is defined by:

$$\boldsymbol{D}(\boldsymbol{x}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{l}} \boldsymbol{w_{0i}x_i} + \boldsymbol{b_o} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{l}} \boldsymbol{y_i\alpha_i x^T}_{\boldsymbol{i}}\mathbf{x} + \boldsymbol{b_o} \tag{3.24}$$

Figure 3.6 below illustrates the concept of the support vectors and the canonical hyperplane defined and constrained by equations (3.23) and (3.24) respectively (Kecman, 2005).

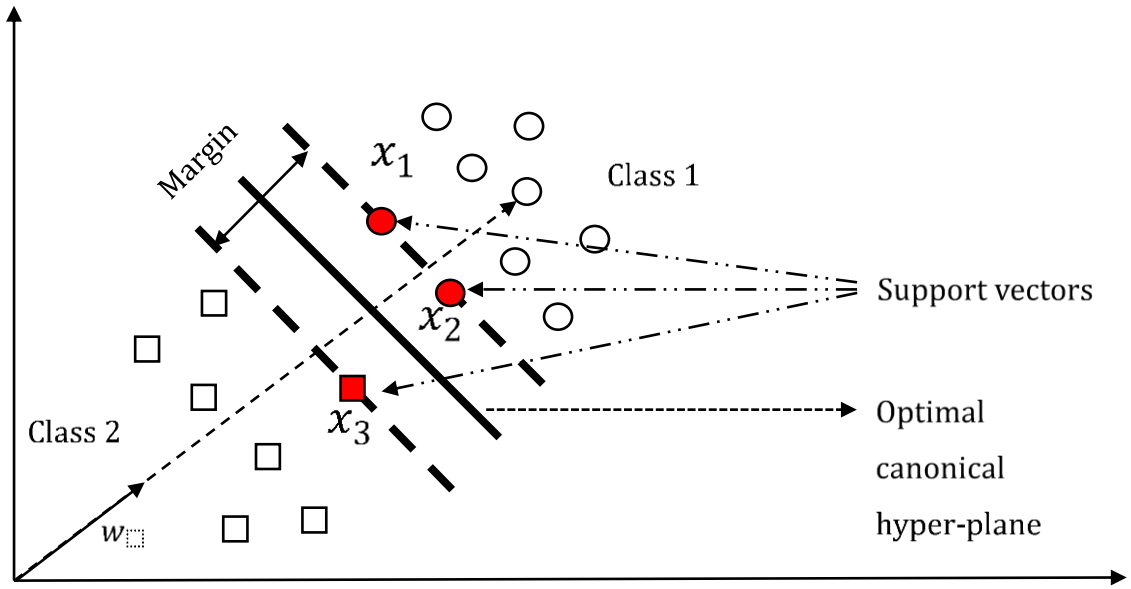

Figure 3.6 Illustrates the support vectors and identified optimal canonical hyperplane, using training data x for linearly separable data.

In feature classification applications with non-linear data, the input feature space is transformed into an *n*-dimensional feature space using kernel functions (K); typically radial basis or polynomial kernels (Subramanya et al. 2015). Table 3.1 summarises the popular kernel functions and the type of classifier.

Table 3.1: Common kernel functions

Kernel functions	Type of classifier
$\mathbf{K}(\mathbf{x}, \boldsymbol{x_i}) = (\mathbf{x}^{\mathbf{T}}\boldsymbol{x_i})$	Linear, dot, kernel
$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_i}) = [(\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x_i}) + \mathbf{1}]^{\boldsymbol{d}}$	Complete polynomial of degree *d*
$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_i}) = \boldsymbol{e}^{\mathbf{0.5}[(x-x_i)^T \Sigma^{-1}(x-x_i)]}$	Gaussian radial basis function

Instead of a canonical decision function to identify a hyperplane for linearly separable data points, a quadratic curve is used to separate non-linear data (Kecman, 2005). Figure (3.7) below illustrates the concept of a quadratic curve to separate non-linear data. If a linear curve were used, a larger set of data points (indicated in red) would be mis-classified. This highlights the need to use a quadratic curve to separate non-linear data effectively.

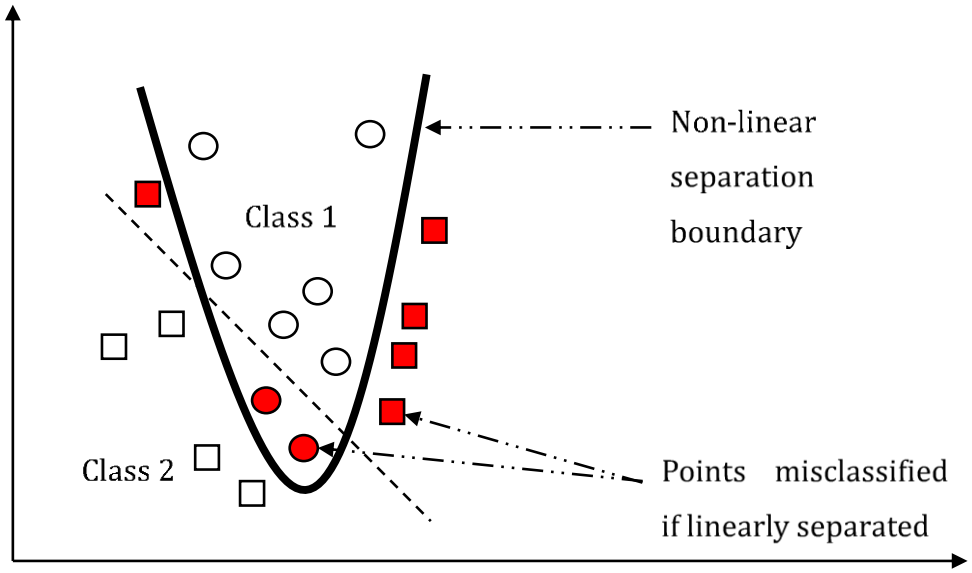

Figure 3.7: A quadratic curve (solid line) versus a linear separation line (dotted line). The data points in red would be mis-classified if a linear separation model were used.

The quadratic curve decision surface is given by:

$$\boldsymbol{D}(\boldsymbol{x}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{l}} \boldsymbol{y_i \alpha_i K}(\boldsymbol{x}, \boldsymbol{x_i}) + \boldsymbol{b} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{l}} \boldsymbol{v_i K}(\boldsymbol{x}, \boldsymbol{x_i}) + \boldsymbol{b} \qquad (3.25)$$

With: $C \geq \alpha_i \geq 0,\ with\ \ i = 1, l \quad and\ \ \sum_{i=1}^{l} y_i \alpha_i = 0$

where K is the kernel function, C the penalty parameter and α, the scaling values used to represent features in a *n*-dimensional space (Kecman, 2005).

The penalty parameter C is model specific and is the upper bound scaling value for α used to constrain equation (3.25). The parameter limits the effect of training data points that fall on the incorrect side of the decision surface and chosen during model development to optimise the decision surface (Kecman, 2005; Nalepa and Kawulok 2019).

3.6 Evaluation metrics

This section discusses the metrics used to assess the performance of image processing, feature extraction and machine learning techniques.

3.6.1 Ground truth image similarity

In automated segmentation tasks the dice similarity coefficient (DSC) is used to measure the intersection of spatial overlap when compared against the ground truth segmentation (Yeghiazaryan and Voiculescu, 2018). The DSC range is between 0 and 1, with 1 indicating a complete overlap. The dice similarity coefficient is calculated by (Deely et al., 2011):

$$\boldsymbol{Dice\ similarity\ coefficient}\ (\boldsymbol{A},\boldsymbol{B}) = \frac{\boldsymbol{2(A \cap B)}}{\boldsymbol{A+B}} \tag{3.26}$$

where A and B represents the two segmentations that are being compared and ∩ the intersection where both segmentation regions overlap.

The structural similarity index measure (SSIM) is an objective measure used to quantify the differences of image quality in terms of contrast and structure between two images. The SSIM range is between 0 and 1, with 1 indicating identical images and 0 no structural similarity. The SSIM is defined by (Wang et al., 2004):

$$\boldsymbol{SSIM(x,y)} = \frac{\boldsymbol{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}}{\boldsymbol{(\mu_x^2+\mu_y^2 + C_1)(\sigma_x^2+\sigma_y^2 + C_2)}} \tag{3.27}$$

where x and y are the two segmentations, μ, σ, C, are luminance, contrast, and structure, which are parameters used to calculate the structural similarity.

3.6.2 Image quality metrics

Ultrasound images often produce low quality images due to the prevalence of speckle noise and contrast variance (Gupta et al., 2018). Denoising and normalisation techniques can be used to enhance the quality of the ultrasound images (Canuma, 2018; Poudel et al., 2018, Kociołek et al., 2020). To evaluate these methods, two statistical measures are used: the root mean square error

and PSNR ratio. These quantitative measures provide sufficient assurance that ultrasound image quality has improved (Loizou et al., 2006; Damodaran, 2009).

3.6.3 Root mean square error

The root mean square error (RMSE) measure is used to quantify the quality of an image after denoising techniques are applied (Loizou et al., 2006). A low RMSE indicates the effectiveness of denoising filters (Damodaran, 2010). The error is defined by:

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{MN}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{M}} \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} \left(\frac{\boldsymbol{g_{i,j}} - \boldsymbol{f_{i,j}}}{\boldsymbol{lpg_{i,j}}} \right)^{\mathbf{2}}} \tag{3.28}$$

where M and N represent the size of an image and particular pixel pairs are represented as *i* and *j*, respectively. The variable $lpg_{i,j}$ represents the filtered image, after denoising and $g_{i,j}$ is the original image.

3.6.4 Peak signal to noise ratio

The PSNR is a measure used to identify the power of the signal related to the power of noise present in an image (Damodaran, 2010). The ratio is defined in decibels by:

$$\boldsymbol{Peak\ signal\ to\ noise\ ratio} = \mathbf{10}\ \mathbf{log_{10}} \frac{\boldsymbol{s}}{\boldsymbol{RMSE}} \tag{3.29}$$

where s represents the maximum intensity value from the original image before denoising.

This ratio determines how well an image is transformed after applying denoising techniques. The aim is to find an optimal value to ensure the signal is separated from noise, which corrupts the image (Damodaran, 2010).

The RMSE and PSNR are therefore two metrics used in conjunction with a visual inspection of an image to determine the quality of an image after denoising and normalisation techniques are applied. This is a useful technique for noisy medical images such as ultrasound (Loizou et al., 2006; Damodaran, 2010).

3.6.5 Machine learning model evaluation metrics

The correctness and robustness of a machine learning algorithm are two common tests to evaluate performance. Robustness is a measure of how resilient a classifier is. Perturbations are applied to the classifier whilst monitoring performance (Zhang and Sejdić, 2019). Correctness determines the model's performance by measuring the accuracy of positive prediction using unseen test data.

Machine learning classifiers score efficiency based on four measures, true positive rate, true negative rate, false positive rate, and false negative rate. The measures are developed by outcomes produced from the machine learning model (either true or false) against the actual outcomes or ground truth (positive or negative) (Parikh et al., 2008). Table 3.2 describes the four measures when applied to an automated classification model. The example presented refers to diseased (positive) or non-diseased populations (negative).

Table 3.2: Automated classification model outcome versus actual gold standard of diagnosis (Parikh et al., 2008).

	Disease: Positive	**Disease: Negative**
True	True Positive	True Negative
False	False Positive	False Negative

In clinical applications there is trade-off among the four performance measures. The sensitivity and specificity tests are used to assess an automated model's accuracy versus the actual diagnostic outcomes (Danjuma, 2015). Equations (3.28) to (3.30) are used to calculate the accuracy, sensitivity, and specificity, respectively (Danjuma, 2015).

$$\boldsymbol{Accuracy}\ (\%) = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}}\ \boldsymbol{100} \tag{3.30}$$

$$\boldsymbol{Sensitivity}\ (\%) = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}\ \boldsymbol{100} \tag{3.31}$$

$$\boldsymbol{Specificity}\ (\%) = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}}\ \boldsymbol{100} \tag{3.32}$$

The accuracy yields a value within a range of 0 and 1, with 1 as a definite positive prediction. However, if a value of 0.5 is produced it is also considered a positive score. For a medical application this might not be appropriate and fails to account for false positives even though the classification accuracy meets a numeric passable criterion of 50%. Performance evaluation for machine learning models are therefore application specific and aim to find the optimal trade-off for true positive and true negative rates (Danjuma, 2015).

The receiving operating characteristic (ROC) curve is another metric used to depict the trade-off between true positive and false positive rates using a threshold value. The performance measure of the ROC curve is analysed using the area under curve (AUC) method. When this value is 1, the learned model is perfectly accurate classifier (Zhang and Sejdić, 2019). The optimal threshold

value for a classifier is chosen based on its ability to maximise the true positives (Zhu et al., 2010). The sensitivity and specificity metrics of a classifier aid predictive evaluation and are generated from the true positive rates and false positive rates, respectively. Figure 3.8 provides an example of the datasets that are generated and the ROC curve (Hajian-Tilaki, 2013).

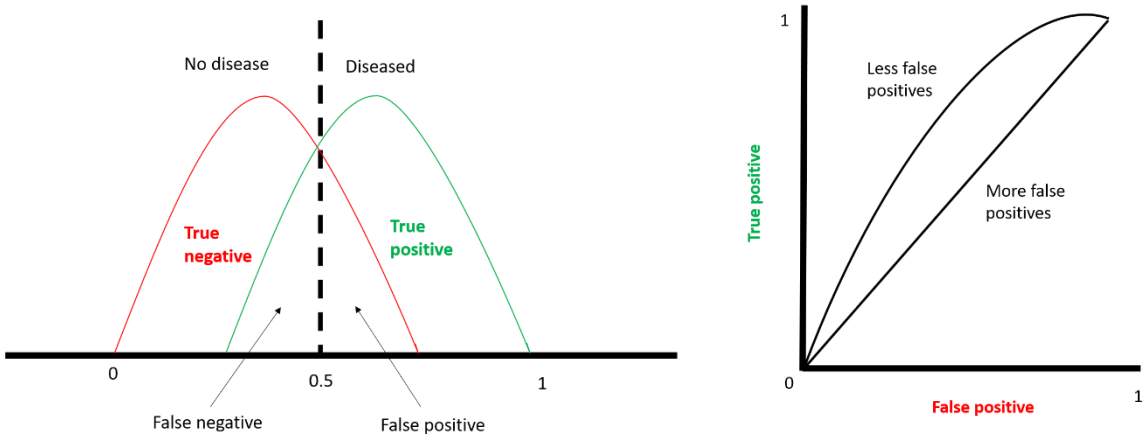

Figure 3.8: Left: Overlapping datasets of diseased or non-diseases ultrasound images indicating additional metrics to use for classifier evaluation. Right: ROC curve (Hallinan, 2014).

The validation of a machine learning model is crucial in the medical environment as a false positive result may misdiagnosis a patient resulting in the administration of incorrect treatment. The use of the ROC technique versus trained specialists can provide an analysis regarding the performance of a trained machine learning classifier (Zhu et al., 2010). However, the predictive accuracy for small, imbalanced datasets, may result in erroneous outcomes, either overoptimistic or remarkably poor performance. Cross-validation methods aim to reduce the effect by training and testing the model on different folds of data. The training and test sets *must cross-over successively such that each data point has a chance* to be tested (Danjuma, 2015). The *k*-fold cross validation method is used to help reduce bias for training, test, and hold-out data choices. First the data is randomly shuffled, the parameter *k* is chosen and represents how many groups or folds the entire data sample is split into (Danjuma, 2015). In each *k-fold* a portion of data is held for testing, and the rest used for training the model. The model is then trained, evaluating the accuracy of the model using the test set. The process is repeated for every fold, saving the score but without retaining each model. Finally, an average accuracy score is calculated across all folds, to obtain one accuracy measure to summarise the performance of the model (Danjuma, 2015). The k-fold cross validation methodology is often used to prevent an overoptimistic representation of performance for machine learning algorithms (Danjuma, 2015).

Equation (3.33) is computed to provide the final accuracy for the model:

$$\boldsymbol{Cross\ validation\ accuracy} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{A_i} \tag{3.33}$$

where *k* is the number of folds chosen, and *A* is the accuracy of the model based on each fold.

Support vector machine models used for computer aided diagnostic systems for ultrasound have used the evaluation metrics presented. Liu et al (2014) used the AUC and cross fold validation techniques whilst varying the parameters gamma and *C* of the support vector machine. An optimal combination of these parameters achieved an accuracy of 90%. A pilot study by Sjogren et al (2016) evaluated the performance and feasibility of developing an SVM to detect abdominal free fluid. The evaluation techniques of sensitivity and specificity provided confidence to clinicians on the use of automated ultrasound classification.

3.7 Feature evaluation metrics

Features extracted from images may vary depending on the number of orientations and scaling parameters used (van Timmeren et al, 2020). Classification models are at risk of model overfitting in cases when used on small datasets are coupled with multiple extracted features. To overcome this and create a generisable model without overfitting, the feature selection technique may be used (van Timmeren et al, 2020). Feature reduction methods provide a guideline to identify the optimal features which are reproducible and relevant. Manual descriptors used to identify abnormalities may create the most reproducible visual features, when developed with more than one clinical expert. This may reduce the number of features early on, ensuring only the most informative features are included. Test and retest methods are an alternate iterative process, used to exclude and or include different features fed into a classification model (van Timmeren et al, 2020). In each iteration the classification accuracy is calculated, using equation 3.31, to measure the robustness and relatability of feature choice. Multiple test and retest iterations are performed to determine which combination of features provides optimal performance, ensuring only relevant features are used.

3.7.1 Gabor parameters

The Gabor filter presented in the theoretical section identified four parameters that need to be chosen F, θ, γ, η. The determination of an optimal set of parameters may be time consuming. However, for small sets of extracted textural features, the accuracy of a classification model can be used to evaluate different combinations of features (Tou et al., 2007).

3.7.2 Normalisation levels

Normalisation methods reduce the effect of uneven sensitivity from data, by reducing the range of intensity values. The range chosen influences the outcomes for textural classification models (Kociołek et al., 2020). The optimal choice of normalisation levels may therefore be evaluated by optimising the accuracy of a classification model as the range of gray levels are varied (Tou et al., 2007).

In this chapter the theoretical considerations used in the project were discussed to highlight the various techniques and methodologies involved in developing an automated classification application using ultrasound images. Ground truth data acquisition and image pre-processing techniques were identified as an essential first step to remove noise inherently present in ultrasound images. Thereafter feature extraction techniques were discussed, with GLCM and the Gabor filters identified to enhance textural and geometrical feature identification in order to characterise abnormal and normal features in medical images. The widely implemented SVM algorithm was explained in detail, followed by an overview of feature and model evaluation metrics to complete the analysis of techniques used to guide the research methodology in this project. Collectively the chapter aids the understanding of the methodology required to produce an automated tool that can be used to improve the differential diagnostic pathway of HL.

4. Project data pre-processing and tools

4.1 Data overview

Preliminary analysis was performed on the available Haematology patient registry, to create two distinct cohorts of Hodgkin's lymphoma (HL) and rapid access diagnostic lymphadenopathy clinic (RADLAC) patients. The two cohorts were identified with guidance from trained clinicians.

4.2 Data extraction

The HL and RADLAC patient populations static ultrasound images were extracted from picture archiving and communication system (PACS), in the digital imaging and communications in medicine (DICOM) format. The ultrasound imaging had been performed in previous studies using multiple scanners: Phillips, Toshiba TUS-A400, Toshiba Xario, Toshiba TUS-X100 and Toshiba TUS-X200. The ultrasound scanner settings were diverse and, in some instances, not captured.

The ultrasound image set for every patient correlated to an instance in time. Homogenised datasets from different time periods were not included as this would not represent the scenario radiologists would use when interpreting ultrasound scans.

Data was maintained on electronic platforms, only available to the researcher and principal project supervisors. Data was stored on Figshare, a repository specifically designed for storing clinical information used in research projects (Figshare, 2012).

4.3 HL cohort

The HL cohort was a retrospective inclusion of a well-defined set of patients with HL (a disease state commonly presenting with peripheral lymphadenopathy), and the US images retrospectively reviewed. Each patient had at least one of the seven imaging biomarkers noted either on the ultrasound scan or in the accompanying radiological report (where available) and a confirmed HL diagnosis. A total of 35 patients met the criteria for inclusion. Figure 4.1 illustrates the steps taken to examine, analyse and collate the HL cohort.

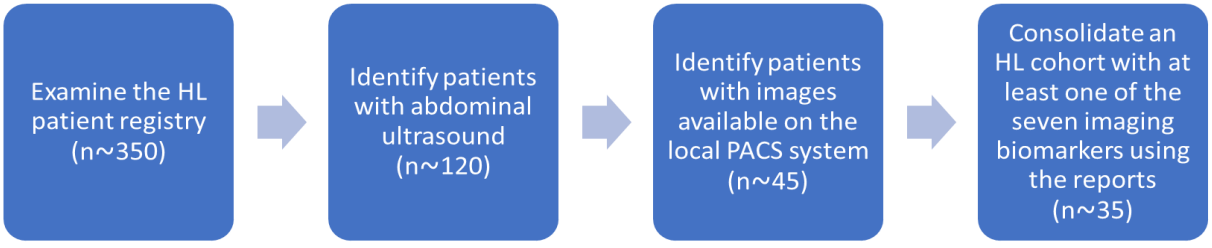

Figure 4.1: Steps taken to identify the HL cohort.

4.3.1 HL cohort for objective 1

The data requirements for objective 1 (**the development of a prospective abdominal ultrasound imaging protocol**) was to identify a small subset of the HL patient population. The image data for the identified subset was used to develop a precise ultrasound imaging protocol from the retrospective image data. This was performed with assistance from two radiologists and insight from a clinical haematologist. Descriptors and standard taxonomy for seven biomarkers (abnormalities) were formulated to create a well annotated table, describing the developed protocol. A subset of 5 HL patients were used to create the protocol for Objective 1.

4.3.2 HL cohort for objective 2,3 and 4

A subset of 23 HL patients were identified from the larger 35 HL patient population and excluded the patients used in objective 1. All 23 HL patients had diagnostic ultrasound scans performed close to the HL diagnosis date. This improved the likelihood that any noticeable ultrasound abnormalities could be associated with a positive HL diagnosis. In at least one of the ultrasound images one abnormality of interest needed to be identified based on the review of the ultrasound scan reports together with the review of annotated ultrasound images, where available. The review was performed by the researcher with guidance from the radiologists involved with the project. The data characteristics for the HL cohort is presented in Table 4.1.

Table 4.1: Data characteristics of the HL cohort.

	HL patients (n = 23)
Male	16
Female	7
Average Age	39
HIV positive	19
HIV negative	4
TB treatment	8
TB organism found	0
Proven TB and HIV positive	0
Proven TB only	0

4.3.3 RADLAC cohort for objective 2, 3 and 4

The RADLAC cohort was identified using patient data from a prospective biopsy clinic to evaluate patients with peripheral lymphadenopathy of unknown aetiology, in order to obtain a diagnosis of either HL, or other lymphoma, or TB, or other cancer/infection. The US images were retrospectively reviewed and a cohort of 24 patients were identified, with confirmed HL patients excluded. Together with the HL cohort, the identified RADLAC cohort was used to develop the

abnormality detection framework for objective 3, and the automated classification model for objective 4. Figure 4.2 below illustrates the steps taken to examine, analyse and collate the RADLAC cohort.

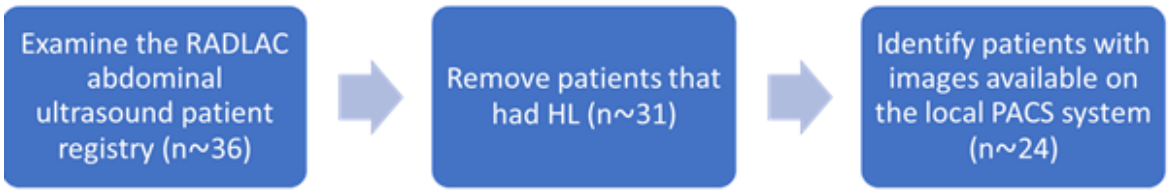

Figure 4.2: Steps taken to create the RADLAC cohort.

The data characteristics for the RADLAC cohort is presented in Table 4.2.

Table 4.2: Data characteristics of the RADLAC cohort.

	RADLAC patients (n = 24)
Male	8
Female	16
Average Age	43
HIV positive	16
HIV negative	8
TB treatment	7
TB organism found	5
Proven TB and HIV positive	4
Proven TB only	1

4.4 Hardware and software tools

The research project was implemented on a 64-bit Dell laptop with an Intel ® i7-855OU CPU @ 1.8 GHz, integrated graphics card and 8 GB of RAM.

The project was designed and developed using a variety of software packages: 1) MicroDicom to view ultrasound images (https://www.microdicom.com/); 2) ImageJ, an open source software tool for image processing, specifically used in this project for segmentation of imaging biomarkers (https://ImageJ.Net/) and 3) Anaconda, an open source platform used to distribute packages for Python and R programming languages for computer science and machine learning applications (https://www.anaconda.com/). The Jupyter notebook web-based application was chosen to implement the algorithmic framework in Python code.

A variety of packages were downloaded to provide a seamless implementation of the framework and used with the notebook, notably; Scikit-learn (https://scikit-learn.org/stable), OpenCV (https://opencv.org/), Scikit-image (https://scikit-image.org/), NumPy (http://www.numpy.org), Pandas (https://pandas.pydata.org) and Matplotlib (https://matplotlib.org).

The Google Chrome Remote Desktop application (https://remotedesktop.google.com/support/) was installed on both the researcher's and radiologist's computers. This allowed the radiologist to have access to the researcher's computer to use the necessary tools required to identify, capture, and segment the abnormalities.

5. Development of a prospective abdominal ultrasound imaging protocol

5.1 Overview

The primary aim of the research project was to develop and validate an algorithmic framework for automated feature detection using abdominal ultrasound images. The clinical motivation was to explore the use of computer aided diagnosis to detect abdominal abnormalities of interest earlier on in the Hodgkin's lymphoma (HL) diagnostic pathway. To achieve this, a precise abdominal ultrasound protocol was required and therefore developed. The protocol provided an outline of specific frames required to detect seven abnormalities of interest for this project. Identifying the required frames ensured subsequent image processing techniques applied to ultrasound images would extract only geometrical and textural features to characterise the seven abnormalities (image biomarkers) of interest. A secondary outcome of the protocol was a prospective guide to image HL patients using ultrasound. This may enhance the ability to capture abnormalities (clinical biomarkers) earlier on in the HL diagnostic pathway. This chapter therefore presents and describes how this was achieved.

5.2 Methods: new ultrasound imaging protocol and frame sequence

Two radiologists and a clinical haematologist were consulted during the development of the ultrasound protocol. The radiologists reviewed non-specific sequences of ultrasound frames available from 5 HL patients, as described in Chapter 4. Each patients' set of images were assessed to determine, through an iterative and consultative process, the most suitable frame that enhanced the visibility of each of the seven abnormalities of interest. A pre-labelled EXCEL table was developed to capture the frame view and corresponding descriptions as each frame was chosen to constitute a new protocol.

In the first iteration of the process, the frames that consistently contained each of the abnormalities were identified and captured. However, not every abnormality was clearly identified due to the quality of the retrospective images and inconsistent frames or views of the 5 HL patient image set. The radiologists performed a second iteration, benchmarking their first round of analysis with typical abdominal ultrasound protocols. The frames were updated to ensure a more robust identification of the abnormalities. Notably, the orientation of probe and anatomical landmarks were confirmed by both radiologists to create a comprehensive set of frames for the new protocol.

In the final iteration the clinician assessed the developed protocol to ensure the HL imaging biomarker detection requirements were met. Minor refinements were made with detailed descriptions of the anatomical landmarks required to identify each of the frames. The final clear

and concise protocol developed and agreed to by the radiologists and the clinician comprised of eight specific frames. These specific frames were identified as a robust sequence that would enhance consistent identification of the several abnormalities of interest within the scope of this project.

5.3 Findings: new ultrasound imaging protocol and frame sequence

Splenomegaly and splenic lesions were two of the abdominal abnormalities identified using one frame. Splenic microabscesses was categorised as frame 2 and the only frame in the new protocol captured using a linear probe. The lymph nodes were separated into three anatomical regions: epigastric, retroperitoneal, and mesenteric each identified by frames 3, 4 and 5, respectively. Ascites, pericardial and pleural effusions required three different frames for identification, represented by frames 6, 7 and 8 in the new protocol. Table 5.1 presents the details developed for each frame and summarises the findings for the new ultrasound imaging protocol developed for this project. The table provides specific orientations of the ultrasound probes to locate the ideal position to capture the frame and is supplemented with anatomical landmarks and descriptions that should be identified before each frame is captured.

Table 5.1: New ultrasound imaging protocol and frame sequence developed for the seven abnormalities of interest.

Frame number	Orientation of probe and view required	Anatomical landmarks and descriptions to capture abnormality
1	Longitudinal probe position, along long axis of the spleen. The view of spleen should be through the hilum.	*Locate in sequential order:* i. Diaphragm: with the superior margin to lower pole of spleen visible. ii. Hilum: The fat around hilum is bright and the vessels to spleen must be visible. iii. Spleen which is bean shaped. The texture must be visually assessed by radiologist to identify if lesions are present or not. iv. Left kidney is a landmark used over and above points 1-3 in terms of relative position to the probe.
2	Longitudinal probe position along long axis of the spleen.	Spleen is typically bean shaped and once identified, its texture should be visually assessed by radiologist to

	A linear high frequency probe must be used and zoomed in to detect abscesses.	identify if microabscesses are present or not.
3	Longitudinal probe position, placed on epigastric region located around the pancreas	*Locate in sequential order:* i. Liver ii. Porta hepatis iii. Pancreas iv. Stomach: Sometimes difficult to see using ultrasound.
4	Longitudinal probe position, placed on retroperitoneal region, located around aorta and above and below renal veins.	Identify the relative location of the inferior vena cava and the abdominal aorta.
5	Longitudinal probe position placed on mesentery located around the bowel area.	Identify the branches of the superior mesenteric artery and vein.
6	Longitudinal probe position, placed on transhepatic region relative to the right kidney (most sensitive area to identify abnormality).	*Locate in sequential order:* i. Morison's pouch: Between right kidney and liver. ii. Pelvis: Behind bladder
7	Longitudinal probe position, placed on epigastric area angled to see heart from under the diaphragm.	Identify heart and look around for excess fluid.
8	Longitudinal probe position, placed along the sides of the ribs.	*Locate in sequential order:* i. Diaphragm ii. Lungs iii. Spleen (left effusion) iv. Liver (right effusion)

5.4 Methods: identifying geometrical and textural descriptors

Table 5.1 presented the new ultrasound protocol describing the orientation of the probe and specific landmarks required to enhance the detection of each of the abnormalities of interest. The next step performed by both radiologists was to identify the morphological and textural descriptors routinely utilised to describe and assess every abnormality. The descriptors were captured alongside each frame in the new protocol and reviewed with the clinician working on the project.

5.5 Findings: identifying geometrical and textural descriptors

The abnormalities were characterised by either the same or different textural and geometrical descriptors. Frames 1 and 2 share the same textural descriptors to describe splenic lesions and splenic microabscess as homogenous or heterogenous but have different geometrical descriptors. Frames 3 to 8 are described by textural descriptors categorised as either hyperechoic (dark area in

ultrasound) or hypoechoic (lighter area in ultrasound). However, their geometric descriptors varied across each frame and corresponding abnormality.

Splenic lesions were categorised as either greater than 10 mm or less than or equal to 10 mm, when assessing frame 1. A splenic lesion categorised with a diameter less than 5 mm was defined as a splenic microabscess when analysing frame 2. The categorisation of a lesion with a diameter between 5 mm and 10 mm was noted as a potential gray area between radiologists as the final categorisation depends on the texture, frequency, and presence of the lesions to classify them as a splenic lesion or splenic microabscess. The shape descriptors captured for splenic lesions and microabscesses were identified as either round or irregular for frames 1 and 2. The lymph node shape descriptors were categorised as normal, round, irregular or ovoid for frames 3,4 and 5, respectively. The geometrical and textural descriptors for each frame in the new protocol are provided in Table 5.2 below.

Table 5.2 Geometrical and textural descriptors corresponding to each frame in the new protocol to identify each abnormality.

Frame number	Geometric descriptors	Textural descriptors
1	i. Spleen size >=130 mm ii. Lesion size >10 mm or <10 mm iii. Lesion shape (Round/Irregular)	Homogenous (smooth) /heterogenous (coarse/uneven)
2	i. Microabscess <=5 mm ii. Microabscess shape (Round/Irregular)	Homogenous (smooth) /heterogenous (coarse/uneven)
3	i. Lymph node >=10 mm ii. Shape (Normal, Round, Irregular, Ovoid)	Hypoechoic/hyperechoic (dark area/lighter area)
4	i. Lymph node >=10 mm ii. Shape (Normal, Round, Irregular, Ovoid)	Hypoechoic/hyperechoic
5	i. Lymph node >=10 mm ii. Shape (Normal, Round, Irregular, Ovoid)	Hypoechoic/hyperechoic
6	i. Sliver of fluid (size <=5 mm) ii. Free fluid (size >5 mm)	Hypoechoic/hyperechoic
7	i. Sliver of fluid (size <=5 mm) ii. Free fluid (size >5 mm)	Hypoechoic/hyperechoic
8	i. Sliver of fluid (size <=5 mm) ii. Free fluid (size >5 mm)	Hypoechoic/hyperechoic

Table 5.1 and 5.2 were combined to create one large table of the developed protocol, descriptions of each frame and the geometrical and textural features identified for each frame. The new table

was reviewed by the radiologists and refined to separate the geometric descriptors into shape and size categories. The textural descriptors were categorised as echotexture and three new categories (status, frequency, and location) were included to provide supplementary fields to assess each frame. An additional column was introduced for radiologists to capture commentary and insights on potential patterns of abnormalities they believe suggests a differential outcome or diagnosis for each patient. The new table is presented in Table A.1, Appendix A.

5.6 Conclusion

Ultrasound imaging protocols are a set of standard procedures radiologists utilise when examining a patient. However, these protocols are not always adhered to and more often so in strained healthcare environments (Geijer and Geijer 2018). This may potentially lead to missed opportunities to identify HL biomarkers from abdominal ultrasound (Antel et al., 2019; Verburgh and Antel, 2019). In an HIV endemic environment, these HL biomarkers may be falsely attributed to be synonymous with extrapulmonary TB. (Caremani et al., 2013).

The developed protocol in this project is a novel guideline that aids the development of an automated feature characterisation algorithm, that can now be applied to common abdominal ultrasound findings and be validated for the full range of common differential diagnoses. In addition, it has potential to contribute to the design of an imaging protocol that can be used in prospective studies. However, further testing and evaluation is required by trained radiologists through an ethically approved study to use the developed protocol for ultrasound imaging of suspected HL patients. A successful outcome of such an exercise would result in consistent detection of abnormalities that may not be routinely captured using current ultrasound imaging guidelines in the HL diagnostic pathway.

To conclude, a set of distinctive frames for the seven abnormalities that often present across HL, TB and HIV patients were successfully identified with clinical guidance. This provided a precise protocol to facilitate the development of ground truth descriptors which is described in the following chapter.

6. Ground truth feature development

This chapter describes how objective 2, *developing geometric and textural mathematical ground truth descriptors for seven abdominal abnormalities* was achieved. The chapter begins with the methods used to prepare the data followed by the steps taken to capture and segment abnormalities. Next, the methods used for denoising, and intensity normalisation of the ultrasound images are presented. Subsequently, the methodology used for textural feature development is outlined followed by a description of the geometric feature development process. Finally, the results and outcomes of the objective are presented, and conclusions drawn.

6.1 Data preparation

The ultrasound images for the identified 23 Hodgkin's lymphoma (HL) and 24 rapid access diagnostic lymphadenopathy clinic (RADLAC) patient cohorts were extracted from the picture archiving and communication system (PACS) with the patient folder number kept as key identifier in a password protected file, stored on the researcher's computer. Prior to being assessed, each patient's images were anonymised; a first anonymisation removing the reference to the specific patient folder number; and a second anonymisation randomising patient files to blind any patient information from the radiologists and reduce bias. The researcher kept a master file with the original data and patient folder numbers and the new allocated anonymised names. All data was saved to the Figshare repository, which was updated regularly.

The patients were first categorised numerically in the format *Patient x HL and Patient x RADLAC*, where *x* represented a number from 1 to 23 and 1 to 24 for the HL and RADLAC cohorts, respectively. Two sub folders were created in each patient's folder, one for the original PACS ultrasound images and reports (where available) and one for all the combined segmentation outcomes. Two separate folders for each radiologist were created: *'Radiologist 1' and 'Radiologist 2'*. In each radiologist's folder, a file for every patient was created, saved in an alternating order of HL and RADLAC patients; and stored as *'Patient x'* where *x* represents a numerical value. Every *'Patient x'* had a saved mapping to the first round of numerical categorisations for the HL and RADLAC cohorts, which were only available to the researcher.

Each radiologist's folder was divided into three subfolders: *'New Sequence,' 'PACS Original Images and Report' and 'Segmented.'* The original ultrasound scans were copied into the *'PACS Original Images and Report'* folder for each patient and these were the ultrasound images shown to each radiologist for assessment. The two radiologists involved in the project, who assessed the ultrasound images, were therefore blind to the actual patient folder number, the final patient

diagnoses, and each other's work. This ensured that the researcher was the only observer aware of the mappings to either the HL or RADLAC groups. Figure 6.1 illustrates the different folder and subfolders used to anonymise and store images, findings, and outcomes.

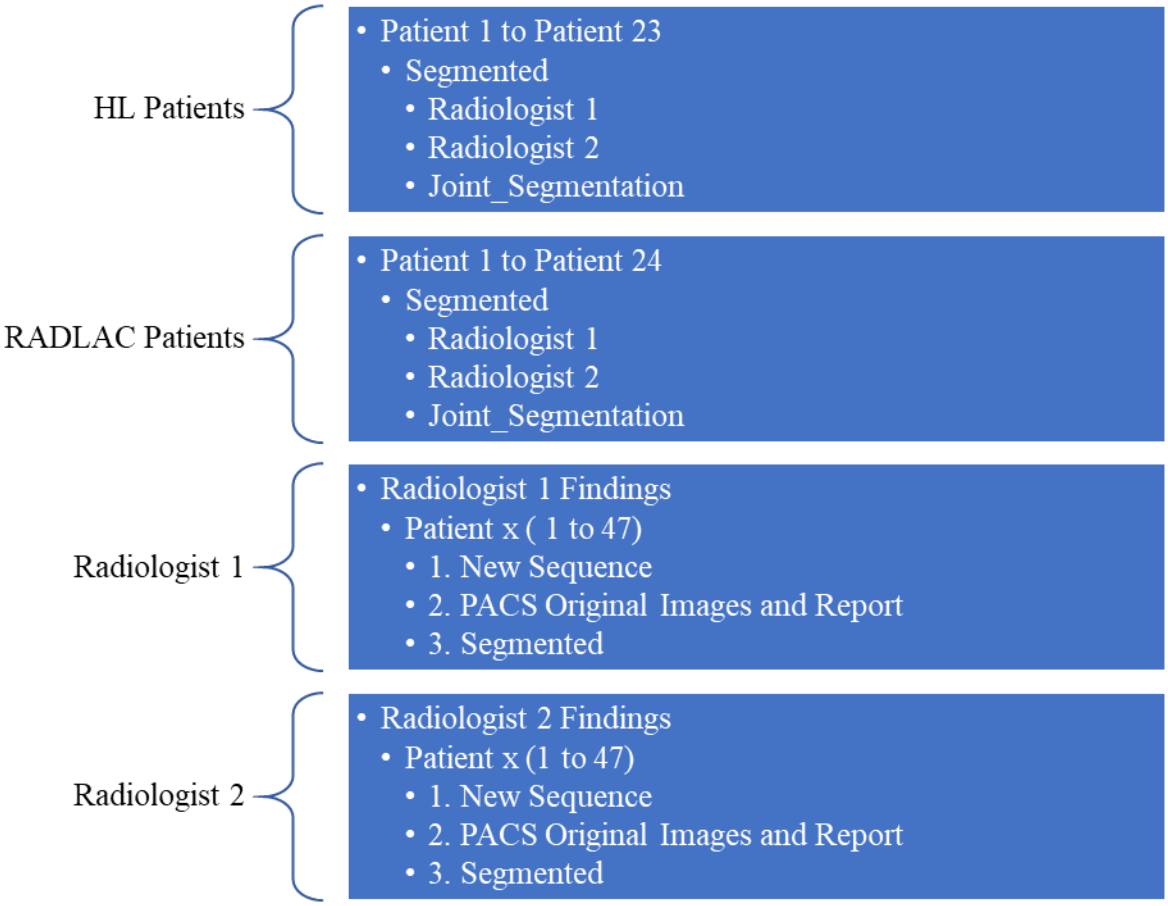

Figure 6.1: The naming conventions and folders used to store and save all images, findings, and outcomes from objective 2.

The Google Chrome remote desktop application allowed the radiologists to access the researcher's computer to assess the ultrasound images. Due to time constraints and availability of radiologists only 10 HL patients and 5 RADLAC patients with confirmed tuberculosis (TB) were analysed by both radiologists. However, radiologist 2 analysed an additional 3 RADLAC patients who did not have confirmed TB. All 3 cohort subsets had a large percentage of Human Immunodeficiency Virus (HIV) patients (over 80%) which may contribute to distinct radiological findings. These smaller patient subsets were used in the algorithm development and experiments presented in this research project in subsequent sections.

6.2 Abnormality identification, capture and segmentation

Table A.1 in Appendix A, was used as a template to create a pre-labelled EXCEL table for radiologists to assess the ultrasound images. Drop down lists were created to facilitate easy capture of abnormalities present and the descriptors developed and identified from objective 1. Each radiologist had their own pre-labelled EXCEL sheets saved as *'Radiologist 1 Findings' and 'Radiologist 2 Findings'*. All patient's images were examined, identified abnormalities captured, and appended to a new row in the EXCEL sheet. The column for radiologists to suggest a suspected disease state (seen in Table A.1, Appendix A) was populated using the same approach as current radiological reports. A snippet of the drop down pre-labelled table, specifically illustrating spleen capture, can be found in Table A.2, Appendix A.

Every ultrasound image was assessed independently by each radiologist and saved in their respective folders using the pre-labelled EXCEL table described. The identification and capture of abnormalities was performed in three key steps: 1) first the radiologists had to assess all ultrasound images available for a patient and identify any abnormalities; 2) the accompanying textural and geometric visual descriptors were captured and lastly 3) all abnormalities were manually segmented by each radiologist and saved.

The original PACS images were shown to each radiologist using the MicroDicom viewer (https://www.microdicom.com/). When the radiologist identified an abnormality, the original image number and new frame number defined by the developed imaging protocol were captured. The type of abnormality identified determined which geometric and textural descriptors were captured. During the process, rules were developed to standardise capture for both radiologists and limit variation in abnormality identification. Firstly, since the frame number in the new protocol could have multiple views for a specific frame, each radiologist was instructed to capture the frame that best fit the descriptions developed in Table 5.1. Secondly, if there was any uncertainty whether an abnormality was present, the radiologist did not capture any frames. Thirdly, if there was more than one occurrence of an abnormality of interest, the frame with the best visual representation of that abnormality was captured. The final findings across the 10 HL and 5 RADLAC patients assessed by both radiologists indicated all 10 HL having at least one abnormality present. However, only 3 RADLAC patients with confirmed TB had at least one abnormality present.

The radiological findings common to both radiologists were included for analysis to derive a single objective ground truth and develop the algorithmic frameworks. A total of 25 abnormalities were identified by both radiologists for the 10 HL and 3 TB patients from the RADLAC cohort. Frame

1 identified either or both splenomegaly and splenic lesions, therefore only 22 unique frames were captured to represent the 25 abnormalities across both cohorts. The additional 3 RADLAC patients with no confirmed TB, and only assessed by radiologist 2, had 5 unique frames to represent 5 abnormalities. A summary of findings are tabulated and presented for radiologist 1 and 2 in Tables A.3 and A.4 in Appendix A, respectively. Additionally, a final table presented as Table A.5 in Appendix A, was developed to clearly highlight and retain only the same abnormalities and frames chosen by both radiologists.

The abnormalities identified by both radiologists for HL and TB patients highlighted differences between cohorts even though only a small group were analysed. Hodgkin's lymphoma patients had a larger prevalence of splenomegaly (6 out of 10 patients,) compared to 1 of the 3 TB patients. Splenic microabscesses were present in HL patients but not encountered as frequently as splenic lesions. Abnormal epigastric lymph nodes were found more often in TB patients (2 out 3 times) compared to just one occurrence across all 10 HL patients. Retroperitoneal lymph nodes were found in just one HL patient's image data and in none of the TB patients. Enlarged lymph nodes in the mesentery were not seen across any of the patients and this potentially indicates the difficulty of identifying these anatomical regions using ultrasound. Ascites were only identified and prevalent in images of HL patients who were HIV positive. Only one pericardial effusion was identified across all patients for an HL HIV patient. The pleural effusions were seen more frequently in the images of HL patients (2 out of 10) however this could be due to the lower number of TB patients presenting with abdominal abnormalities.

The researcher used Tables A.3 and A.4 to save the new frame sequence for every patient in the *'New Sequence'* subfolder in tag image file format (.tif) which supports a lossless compression and is supported by the software utilised in the project (https://ImageJ.Net/). The images were stored as *'Frame x'* where *x* is a number from 1 to 8, corresponding to the developed protocol sequence. Using the Google Chrome remote desktop application, the radiologist accessed the researcher's computer and used the ImageJ tool to segment each abnormality. The segmented images were saved as binary masks in .tif file format in the *'Segmented'* folder as *'Frame x'* with *x* being the specific frame number corresponding to sequence identified in the developed protocol. The continuous use of the naming convention *'Frame x'* was implemented to prevent any potential errors in allocating incorrect frames through the analysis and implementation of the objectives.

6.3 Pre-processing

Before image enhancement and noise reduction techniques were applied, the two separate manual segmentations developed by each radiologist were combined to form one objective ground truth segmentation through a series of steps:

1. Using the Jupyter notebook the original ultrasound image and each segmentation from radiologist 1 and 2 were cropped to 530 x 420-pixels and converted to grayscale using the *cropcenter* function with OpenCV and the developed *join_segmentations* function.
2. Using the two segmented masks and the original frame, only the region of interest (ROI) or segmented abnormality was extracted. A bitwise joining of the two images ensured only the area where both segmentations overlapped was retained using the *join_segmentations* function. The *get_final_mask* function was developed within *join_segmentations* and used to crop the ground truth segmentation to its contours. The abnormality could now be limited to its contours to ensure the features that were extracted were only from the abnormality of interest and not the entire ultrasound frame. The developed function used the built in *findContours* function, extracting the contour points. Evaluation of the *get_final_mask* function was first visually inspected using the original frame and the segmented frame to confirm that only the abnormality of interest was extracted. Figure 6.2 illustrates the process with frame 1 from an HL patient, with an enlarged spleen. The frame includes other structures around the spleen compared to Figure 6.3 illustrating just the spleen after the *get_final_mask* function is applied to the frame.

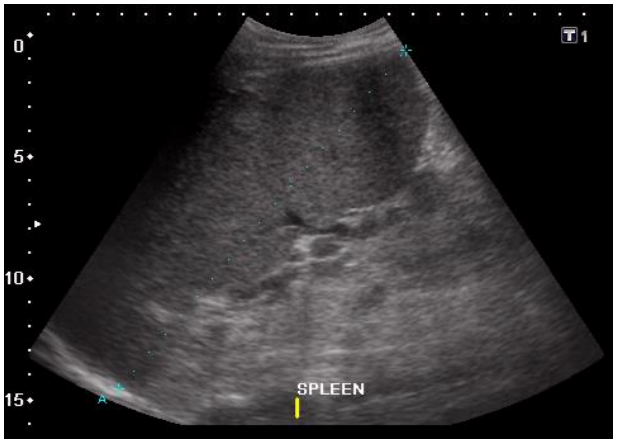

Figure 6.2 Original frame analysed after cropping and removing patient details.

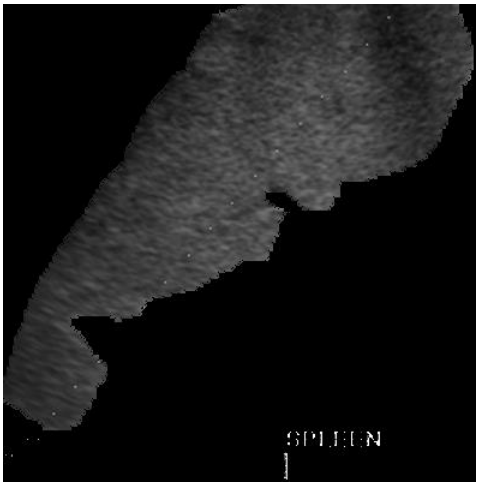

Figure 6.3 Frame illustrating just the ROI or the enlarged spleen after the *get_final_mask* function is applied to the frame.

3. The structural similarity index measure (SSIM) evaluation metric was used to identify the difference in segmentations between two radiologists. The two segmentations were compared against one another and against the derived single objective segmentation. Figure 6.4 indicates the two different segmentations and the final derived single objective segmentation for the abnormality captured in frame 1 for patient 4 from the HL cohort. The SSIM output was 0. 894 between the two segmentations and once joined and compared to the original segmentation the SSIM had a value of 0.981. The SSIM range is between 0 and 1 with 1 indicating identical images and 0 no structural similarity.

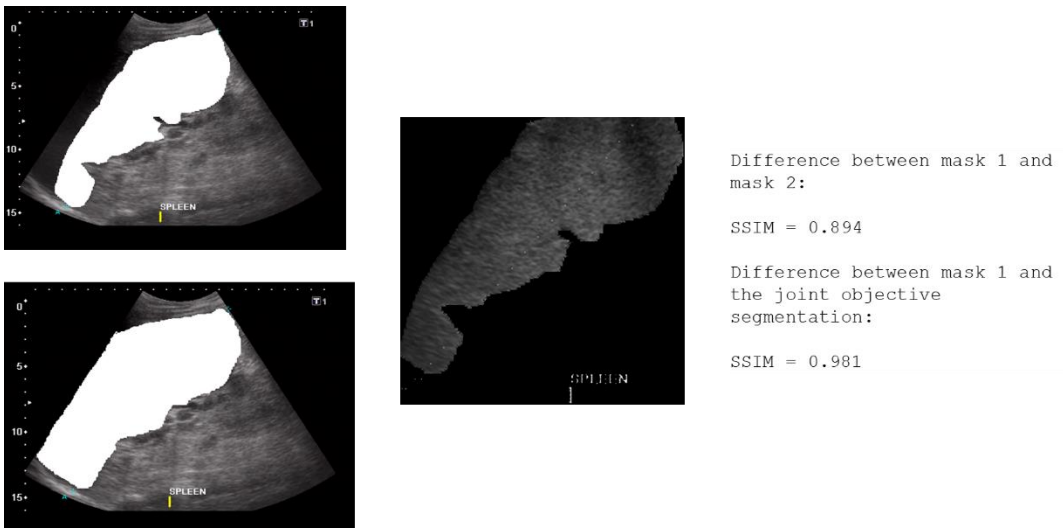

Figure 6.4 Radiologist 1 (top left) segmentation and radiologist 2 segmentation (bottom left) with the final single objective segmentation illustrated at the bottom, with the SSIM metric values included.

The images acquired using the *join segmentations* function were saved directly into every patient's folder in the *'Joint_Segmentation'* subfolder as per the file structure illustrated in Figure 6.1. Each image was saved as '*Frame x'* where *x* is a number from 1 to 8, corresponding to the developed protocol sequence. This ensured the stored images had consistent naming conventions and it was easy to determine what abnormalities a patient had based on the frame number.

6.4 Denoising and normalisation

Every frame saved in the *'Joint_Segmentation'* folder for each patient was passed through two steps to denoise and normalise every image before the ground truth features were extracted:

1. Application of a two-dimensional (2D) discrete wavelet transform (DWT) using the *denoise* function.
2. Normalisation of every frame/image to account for varying illumination settings across all ultrasound images using the *normalise_image* function.

6.4.1 Methods

An iterative process was performed to identify an optimal set of parameter combinations to effectively denoise all images. The parameters utilised were the noise estimate level, different wavelet types, wavelet levels and the final denoised image, to visually inspect any change.

Literature highlighted the use of 128 gray levels as a minimum sufficient intensity range to normalise images for textural feature classification models, using the Gabor transform and GLCM-based features (Kociołek et al., 2020). This was implemented using the *normalise_image* function. The effect of the gray level normalisation was evaluated using the SVM algorithm developed and discussed in chapter 8.

6.4.2 Results of denoising and normalisation

Table 6.1 illustrates the combination of parameters analysed for noise estimate levels, wavelet family and wavelet level choice accompanied by the evaluation metrics peak signal to noise ratio (PSNR) and root mean square error (RMSE). There were only marginal differences in improving PSNR and RMSE when varying these parameters. In contrast Figure 6.5 illustrates a more pronounced difference in reconstructed images after denoising when comparing two different estimated noise values. The image on the left has a smoother representation after denoising the image than the one the right. Considering the combination of factors evaluated a noise estimate level ($\sigma = 0.02$) was chosen with the Daubechies wavelet and a five-level decomposition for this project. This parameter value wase chosen as it produced an overall better-quality reconstructed

image when both a soft threshold (VisuShrink) and the '*bior'* wavelet family of 5 levels were used to denoise the image.

Table 6.1 Results obtained from evaluating two wavelet families and 3 different wavelet levels, accompanied by the PSNR and RMSE values.

Wavelet_Name	Sigmas	Level	RMSE_compared_to_original_image	PSNR_compared_to_original_image (dB)
bior2.8	0.01	2	57.392284	12.954
bior2.8	0.02	2	57.392818	12.954
bior2.8	0.05	2	57.394196	12.953
bior2.8	0.08	2	57.395638	12.953
bior2.8	0.01	5	57.392780	12.954
bior2.8	0.02	5	57.393696	12.954
bior2.8	0.05	5	57.396011	12.953
bior2.8	0.08	5	57.398529	12.953
bior2.8	0.01	6	57.392929	12.954
bior2.8	0.02	6	57.394001	12.953
bior2.8	0.05	6	57.396908	12.953
bior2.8	0.08	6	57.399426	12.953
db5	0.01	2	57.392303	12.954
db5	0.02	2	57.392815	12.954
db5	0.05	2	57.394093	12.953
db5	0.08	2	57.395340	12.953
db5	0.01	5	57.392776	12.954
db5	0.02	5	57.393669	12.954
db5	0.05	5	57.396244	12.953
db5	0.08	5	57.398338	12.953
db5	0.01	6	57.392929	12.954
db5	0.02	6	57.393944	12.953
db5	0.05	6	57.396828	12.953
db5	0.08	6	57.399586	12.95

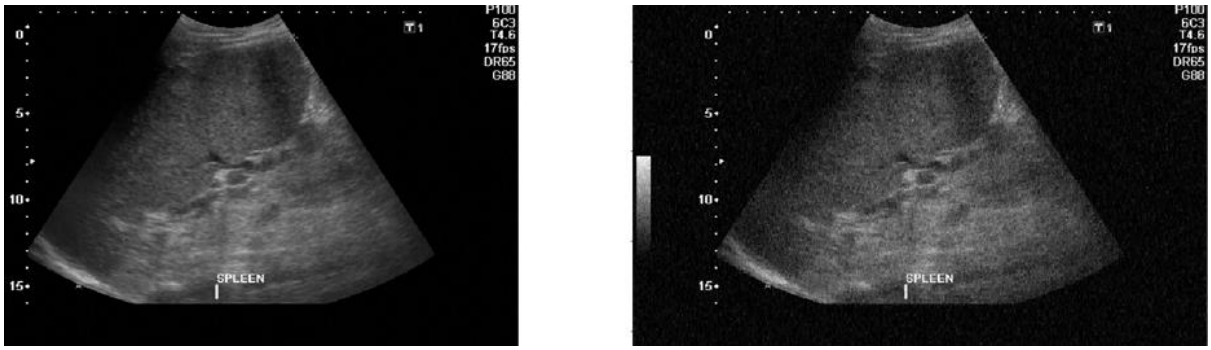

Figure 6.5 Frame 1 for patient 4 from the HL cohort after the *denoise* function is applied with an estimated noise level of 0.02 versus 0.08 for the image on the right. At a noise estimate of $\sigma = 0.02$, noise is removed while not over smoothing the image or leave a very granular appearance when compared to the image on the right ($\sigma = 0.08$). In addition, PSNR is slightly better to for the 0.02 noise estimate.

6.5 Geometrical feature extraction

6.5.1 Methods

Geometric features were extracted directly from each frame in the '*Joint_Segmentation'* folder for every patient, before denoising or normalisation was applied to the images to keep the structure intact when measuring size/assessing shape.

1. Two first order features were extracted to identify variance and standard deviation of pixels using the *geometric_properties* function and the SciPy package.
2. The shape of each abnormality was calculated using the heuristic ratio, to identify how 'round' an object was and determine shape characteristics for abnormal spleens/lymph nodes.
3. Size was used as a metric to analyse the spleen, lymph node, ascites, and effusions. Size is an essential descriptor used by radiologists to categorise these anatomical structures and determine if they breach a specified threshold value to classify them as abnormal. The spleen threshold for enlargement was a craniocaudal diameter greater or equal to 13 cm and for a lymph node, lengths greater than 1 cm (Berzaczy et al., 2019; Griesel et al., 2019). The size of an effusion considered as a sliver versus free fluid had a cut off at 5 mm. A *geometric_size* function was developed using OpenCV to calculate *the Abnormality Size cm* and *Heuristic Size* (ratio).

The typical DICOM header metadata termed *pixel spacing* was not activated on the ultrasound machines and therefore this information was missing in the saved images. The *'Physical Delta X'* and *'Physical Delta Y'* metadata was available but differed across patient image metadata as different scanners were used. Given the above constraints, an average '*Physical Delta'* value of 0.039 was used to represent all scanner values. The average value was used to convert the number of pixels along the longest diameter to a physical distance in cm. This conversion was developed as standard function called *geometric_size* and used for all abnormality size calculations. The automated calculation of spleen size using the *geometric_size* function was evaluated against the ground truth spleen size identified by the radiologists to determine the performance of the function.

6.5.2 Results of geometrical feature extraction

Only patients that presented with an enlarged spleen, corresponding to Frame 1 from the developed protocol were part of the evaluation as the abnormality was more prevalent in the cohorts analysed. Table 6.2 presents the ground truth measurement versus the automated measurement. It is clearly

noticeable that the results obtained differ and only in some instances the ground truth and automated measurements are slightly similar. The larger the physical delta distance value was from the average used in the *geometric_size* function a larger noticeable difference was observed between ground truth and automated measurements.

Table 6.2 Evaluating the difference between ground truth abnormality size of the spleen versus the calculated value using the *geometric_size* function.

Patients (Frame 1) Enlarged spleens	Ground Truth Measurement (cm)	Automated Measurement (cm)
HL Patient 4	18.25	18.27
HL Patient 24	13.01	14.66
HL Patient 25	13.43	14.36
HL Patient 28	13.54	16.76
HL Patient 32	15.49	11.82
HL Patient 35	20.85	20.67
RADLAC Patient 3	14.35	15.81
RADLAC Patient 7	13.92	22.26

6.6 Textural feature extraction

6.6.1 Methods and findings

Textural features were extracted from the denoised and normalised images using the following steps and parameters:

1. Each frame in the '*Joint_Segmentation'* folder of every patient was passed through the *gabor_image* function. The function used the *getGaborKernel* in OpenCV to create a 10x10 kernel (filter) with a spread of $\sigma=50$. Tou et al (2007) reported using four directions $(0°, 45°, 90°, 135°)$ to develop GLCM's with a spatial distance of 1 pixel apart, resulting in a texture classification model accuracy of ~ 80%. The same four directions were used to create the four Gabor kernels for this project.
2. The frequency/wavelength is an important factor a Gabor filter uses to identify textural patterns (Bianconi, and Fernández, 2007). The value of $\lambda = 10$ was selected after evaluating two extreme frequencies. At a wavelength value of $\lambda = 100$, texture was not preserved when analysed visually. This is evident in the illustration in Figure 6.6 below, on the left is the 10x10 kernel at $\lambda = 10$ with the accompanying convolved image. On the right a 10x10 kernel with wavelength $\lambda = 100$ and its accompanying convolved image indicating textural degradation.

Using these parameters, the ground truth segmentations were convolved with the 10x10 kernel with λ = 10 to generate four Gabor images for each of the four directions.

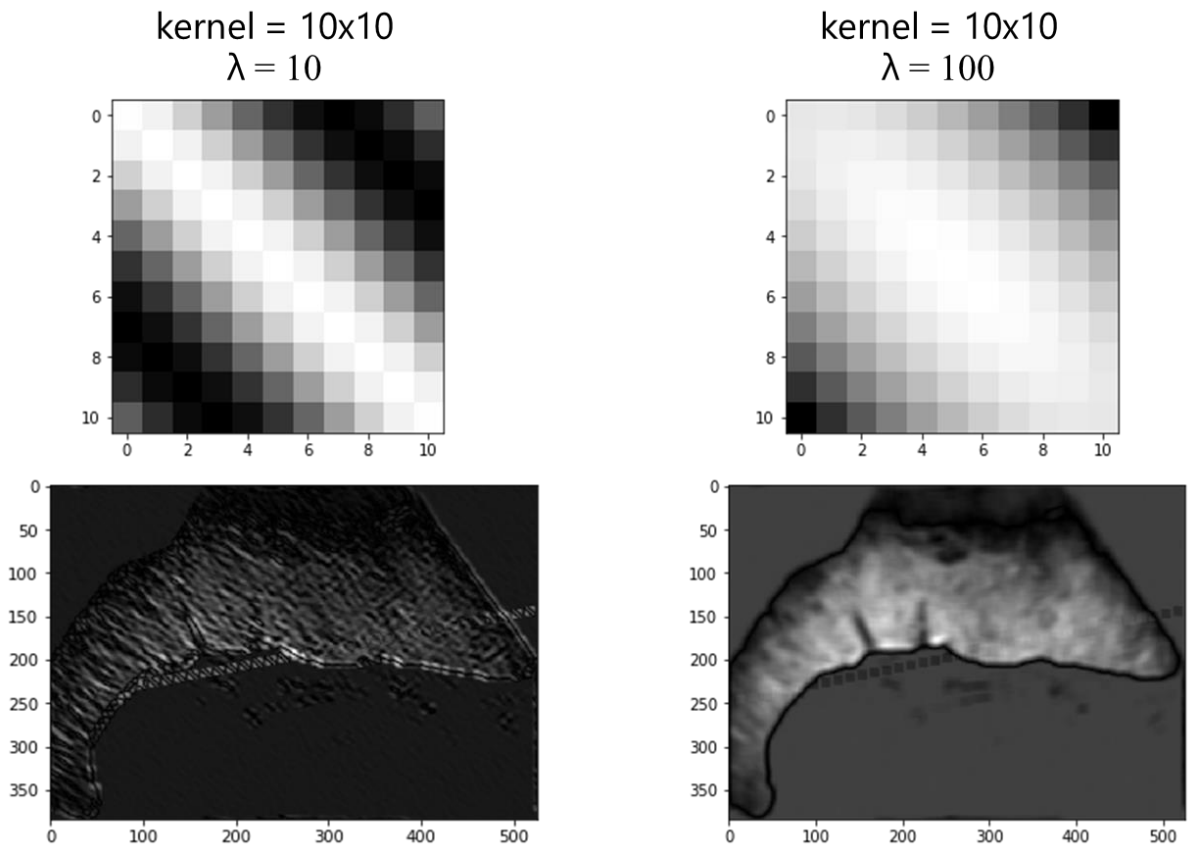

Figure 6.6: The image on the left illustrating the texture identified with the Gabor filter at λ = 10 compared to λ = 100 for the image on the right when all other parameters are kept the same.

3. The Gabor and the GLCM have shown to improve the accuracy of texture-based classification models (Tou et al., 2007; Lahmiri, 2013). Therefore, two additional statistical measures for texture were extracted through the developed *gabor_feature* function from the Gabor image in each direction to enhance the number of textural features extracted:
    a. Gabor mean
    b. Gabor variance
4. The *glcm_properties* function was developed to create a gray level co-occurrence matrix (GLCM) for every Gabor image and extract six statistical measures using:
    a. The *greycomatrix* function and *skimage* algorithm from the sci-kit package to develop the GLCM.
    b. The *greycoprops* function then extracted six statistical textural measures (features) from each of the GLCM's. The measures were chosen as they have previously provided

optimal classification outcomes from ultrasound. The measures and descriptions are presented in Table 6.3 below (Tou et al., 2007; Xu et al., 2019).

Table 6.3: Six statistical textural measures extracted using the *greycoprops* function.

Textural Measure	Description of measure
Contrast	Measures detailed changes of texture.
Dissimilarity	Identifies dissimilar patterns of texture.
Homogeneity	Describes the uniformity of texture by identifying how similar the pixel values are.
Energy	Measures the homogeneity of the image.
Correlation	Measures the variation in pixels in an image to understand texture changes.
Angular second moment (ASM)	Describes uniformity by looking at the distribution of pixel values.

Therefore, the Gabor and GLCM collectively extracted 8 features for each of the four directions to develop 32 textural features to characterise the abnormalities. Table 6.4 below provides a summary of all extracted geometrical and textural features.

Table 6.4 The variable names of all features extracted as calculated in the code with the corresponding number of features extracted.

	Variable name	Number of features
Geometric	1. Abnormality Size cm	1
	2. Heuristic Size	1
	3. Variance of pixel values	1
	4. Standard deviation of pixel values	1
Textural	1. Contrast	4
	2. Dissimilarity	4
	3. Homogeneity	4
	4. Energy	4
	5. Correlation	4
	6. ASM	4
	7. Gabor Mean	4
	8. Gabor Variance	4
Total		**36**

6.7 Discussion

This chapter provided the methodology and techniques used to generate the objective ground truth by only using frames that were identified by both radiologists. The identified abnormalities differed between the HL and RADLAC cohorts. A total of 3 TB confirmed RADLAC patients had abdominal abnormalities present from the 5 included for analysis. All 10 HL patients presented

with at least one abnormality which may create a slight bias for HL abnormality pattern identification. In total 25 abnormalities were found across the 10 HL and 3 TB patients and were used to segment abnormalities and extract features to develop measurable ground truth descriptors.

The first step to develop ground truth descriptors was to denoise and normalise images. The evaluation metrics PSNR and RMSE were used with the reconstructed denoised images to ensure noise removal was adequate. A final choice of a Daubechies wavelet comprised of five levels and a noise estimate of $\sigma=0.02$ produced the best visually reconstructed image and evaluation metrics.

Textural and geometric descriptors were developed to extract a total of 36 geometrical and textural features presented in Table 6.4. However, due to the use of an average conversion value to calculate the size of an abnormality a noticeable difference in ground truth and automated measurements was observed. The metadata for *pixel spacing* would have therefore provided better results as it would be read in directly from the image into the developed function. The consistency of results may be improved in prospective images acquired for similar applications when the *pixel spacing* metadata is available.

6.8 Conclusion

Geometrical and textural descriptors are used by radiologists to identify and characterise visual biomarkers (Afshar et al., 2019). These descriptors are used to inform the development of automated techniques to extract features directly from medical images.

This chapter presented the methodology and techniques utilised to develop objective ground truth descriptors. Functions were developed to extract geometrical and textural properties to mathematically represent potential markers of HL using ultrasound images. These steps are a necessary in order to develop a completely automated algorithm to characterise frames or ultrasound scans implemented in the chapter to follow.

7. Automated abnormality characterisation framework

7.1 Overview

This chapter focuses on the *development of an automated algorithmic abnormality characterisation framework*. The chapter begins with the methods used to develop the framework starting with the mandatory input fields required to read in each frame, if available, using the protocol developed and discussed in Chapter 5. Next the descriptions for each frame are presented followed by a detailed explanation of the *get_Frame_features* function which was developed to automatically extract any of the 36 features from every available frame. The extracted features were measurable descriptors to represent each abnormality. The chapter concludes with the final results and outcomes achieved.

7.2 Methodology

7.2.1 Input files

The automated framework developed was built to require two mandatory manual inputs; 1) *Patient_Name x*, where *x* is the number assigned to a patient within each particular cohort, Hodgkin's lymphoma (HL)/ rapid access diagnostic lymphadenopathy clinic (RADLAC). This follows the standardised format implemented and used to save every HL or RADLAC patient's set of images as detailed in Chapter 5; 2) The second mandatory field is the current directory of the patient files containing the final joint segmentation.

7.2.2 Automated descriptions

Once the two mandatory fields were provided the algorithm searched the *'Joint_Segmentation'* folder for each patient to retrieve the ground truth segmentations, if available. The segmentations were saved as frames using the naming convention *'Frame x'* where *x* is the number between 1 and 8, corresponding to the identified protocol developed in Chapter 5 and a particular abnormality. If a frame was found, a matching description to identify the abnormality would accompany the frame. Table 7.1 provides each frames description as implemented in the algorithm.

Table 7.1: Table indicating each frame and the corresponding descriptions

Frame	Description of each frame
1	'Spleen Enlarged and/ lesions present'
2	'Splenic Lesions/Microabs present'
3	'Epigastric LN present'
4	'Retroperitoneal LN present'
5	'Mesenteric LN present'
6	'Ascites present'
7	'Pericardial effusions present'
8	'Pleural effusions present'

Key: LN – lymph node

7.2.3 Automated extraction

Every frame or ground truth segmentation was read into the framework using the PIL package. The *get_Frame_features* function was then called to perform the following steps on each frame and append the output automatically to a two-dimensional tabular data frame:

a) The frame name was stored in the first column of the data frame under the heading *'Input Image'*. This ensured the frame identified for a patient can be both easily identified, and the features are extracted and appended to the correct frame.

b) Each frame was denoised and normalised (to 128 gray levels) using the developed *denoise* and *normalise_image* functions developed in Chapter 6.

c) A Gabor image for each of the four directions chosen $(0°, 45°, 90°, 135°)$ was created and two Gabor features extracted using the *gabor_image* and *gabor_feature* functions, respectively. These functions were developed and discussed in Chapter 6. A total of 8 textural features could be extracted using these two functions and saved under the headings: *Gabor Mean Input and Gabor Variance Input.*

d) A GLCM matrix was created for every Gabor image, and six statistical textural features using the *glcm_properties* function developed and discussed in Chapter 6 were extracted. Six features from four different Gabor images provided 24 textural features, with each one appended to the data frame under the headings: *Contrast Input x, Dissimilarity Input x, Homogeneity Input x, Energy Input x, Correlation Input x, and Asm Input x*. The *x* represents a number from 1 to 4 as each of the four Gabor images had six features extracted. Therefore, a total of 32 textural features can be extracted from steps c and d.

e) The *geometric_properties* function developed and described in Chapter 6 was used to extract two geometrical features stored under the headings *Variance of pixels* and *Std Deviation of pixels* in the data frame.

f) Lastly the size of the abnormality and the heuristic ratio using the *geometric_size* function developed in chapter 6 were appended to the data frame as *Abnormality Size cm and Heuristic Size*. Therefore, adding these two features a total of 4 geometrical features could be extracted from the frame.

7.3 Results

The search function for every patient's available ground truth segmentation frame was evaluated. The developed framework would indicate the Patient number and respective cohort in addition to the frame found in the folder. An example of a successful search is illustrated below for Patient 35 from the HL cohort who only had Frame 1 identifying an abnormality.

```
Patient 35 HL has the following abnormalities:

Frame 1 = Spleen Enlarged and/ lesions present
```

The final output from the developed algorithmic framework is a data frame of 36 features (4 geometrical and 32 textural) to characterise each abnormal image if present in a patient's folder. Evaluation at each step in the automated framework was performed to sort and check for any potential errors in both format and the value of the output variables. A random extraction of features was performed using the standalone functions developed in Chapter 6 versus the entire algorithmic framework. This provided a validation step to ensure values extracted in the automated framework were the same as the outputs from the standalone single functions. Additionally, for each of the 10 HL patients and each of the 6 RADLAC patients (3 TB confirmed and 3 other) the feature values obtained from the individual functions aligned to the values obtained when extracting the same features using the developed automated algorithm.

A snippet of the final data frame generated after passing through the *get_Frame_features* function is presented in Table 7.2 for Patient 4 from the HL cohort. The table illustrates how each patient's data frame stored the identified frame name and extracted features. In the table Patient 4 from the HL cohort presents with three abnormalities identified as frames 1,3 and 6, with only a snippet of all features shown that would append to the data frame.

Table 7.2: The snippet of the data frame generated using the large algorithmic framework for HL Patient 4.

Input Image	Contrast Input1	Dissimilarity Input1	Homogeneity Input1	Asm Input1	Variance of pixels	Std Deviation of pixels	Abnormality Size cm	Heuristic Size
Frame 1.tif	17.95	2.04	34.28	0.1	14.25	3.77	18.27	0.34
Frame 3.tif	76.69	4.03	102.15	0.07	21.52	4.64	2.64	0.53
Frame 6.tif	81.7	2.23	153.49	0.19	8.62	2.93	7.28	0.33

7.4 Conclusion

The automated extraction of textural and geometrical features from ultrasound images has been used to develop classification models to aid earlier diagnosis, with classification accuracy comparable to radiologists and clinicians (Brattain et al., 2019). However, automated feature extraction techniques has not been applied within the HL diagnostic pathway using abdominal ultrasound.

The developed algorithmic framework in this project could successfully extract all 36 features for each frame to characterise abnormalities directly from abdominal ultrasound images. The textural and geometrical features represent clinical biomarkers that are used to determine the extent of HL disease progression and guide treatment. The automated framework may therefore aid earlier detection of markers for the HL when implemented with an automated classification system, which was developed and described in the following chapter.

8. Evaluation of an automated feature classification model for Hodgkin's lymphoma

This chapter describes the development and evaluation of an automated framework for characterising and classifying Hodgkin's lymphoma (HL) using ultrasound images. The working hypothesis was that abdominal biomarkers in ultrasound images could differentiate HL from TB, as two prevalent disease states in a HIV endemic environment like South Africa. The automated model was developed using textural and geometrical features extracted with the algorithmic framework built in Chapter 7. To develop the automated classification framework, first, a support vector machine (SVM) model was built for differentiating HL patients from rapid access lymphadenopathy clinic (RADLAC) patients who may or may not have had TB. Next, two experiments were performed to evaluate the performance of the model. In each experiment a different dataset was used to determine the testing and predictive classification accuracy of the model. A total of fourteen tests were performed within each experiment to understand how different parameters altered accuracy metrics. The chapter concludes with a discussion on findings and recommendations for future improvements to the developed model.

8.1 Support vector machine model development

The SVM model was built using the *svm* function from the Scikit-learn package (https://scikit-learn.org/stable). The function requires three inputs or parameters: *kernel, gamma,* and *C*. These three parameters are described in Chapter 3 and require fine tuning to optimise classification accuracy. The initialisation of the gamma and *C* values to 0.5 and 1 was guided by the literature using previously developed ultrasound classification models (Liu et al., 2014; Yu et al., 2015). Subsequently, the values were varied to evaluate their impact on model classification accuracy performance. Although literature indicates that the radial bias kernel provides optimal classification performance compared to the polynomial and linear kernels, they were all evaluated within each experiment. (Lahmiri and Boukadoum 2013; Liu and Xu 2014).

Lastly, the dataset used for both experiments required input images mapped to labelled outputs, to train and test the classification model. Patient image data was assigned a binary label outcome of 1 if it was of an HL patient and 0 for RADLAC and or TB patients. Due to the small and imbalanced dataset there was a risk that predictive capacity of the developed model would either overfit or perform poorly and therefore, the cross-validation method was used while training the SVM model. Additionally, the test and training dataset sizes chosen to evaluate the model were meant to maximise the training data available. However, this may have created a bias towards

classification performance because there was a larger HL dataset compared to other disease states, within each experiment.

8.2 Experiment overview

The researcher created a folder for each of the two experiments and every folder contained a new EXCEL table to capture all the results. The first experiment was saved in a folder named *'10 HL versus 3 TB'.* The second experiment was named and saved as *'10 HL versus 6 RADLAC'*.

Each of the folders contained the HL and RADLAC patients ground truth segmentations developed in Chapter 6. Both experiments used the 10 HL and 3 TB patients assessed by both radiologists. However, the second experiment included the RADLAC dataset that only radiologist 2 assessed. In both experiments the method of test and retest described and discussed in section 3.64 was used.

In each experiment fourteen tests were performed and each one saved as folders named *'Test x'* where x is a value from 1 to 14. This ensured the work and findings for experiments were structured and repeatable. Each test evaluated the effects of different hyper parameters on classification performance. The training and validation accuracy metrics were used to evaluate the performance as parameters were changed. The accuracy is measured on a scale from 0 to 1, with 1 as the highest accuracy achievable.

The first test was designed to identify how changes of three parameters to the SVM model would impact classification performance. Next with clinical guidance the effects on classification performance using abdominal features focused on the spleen and lymph node abnormalities was evaluated. Then the Gabor filter frequency and kernel size were varied in test 4 and 5. Next textural features were at random included or excluded in tests 6,7,8,9 and 10 to evaluate the effect on classification performance (Brattain et al., 2019). The test cases were then expanded to include additional tests for other combinations of features and presented as tests 11,12,13 and 14. In each test the variables for C, gamma and test size were varied to prevent bias in model choice. When evaluating the performance of the model in each test, only the most accurate classifier with associated variables were reported. Table 8.1 below provides a clear description of each test as described.

Table 8.1 Tabulated description of each test performed within every experiment.

Test Case	Description of variables used to evaluate performance of support vector machine
Test 1	This test was just to evaluate the C, gamma, and test size effects on performance of the SVM
Test 2	Only frames 1,2,3,4 and 5 were used
Test 3	Only gray level used for intensity normalisation was reduced
Test 4	Only the Gabor filter frequency was reduced
Test 5	Only the Gabor kernel size was increased
Test 6	Only the textural features were used
Test 7	The contrast textural feature in four directions was not included
Test 8	The contrast and ASM textural features in four directions were not included
Test 9	The contrast, homogeneity and ASM textural features in four directions were not included
Test 10	The dissimilarity textural feature in four directions was not included
Test 11	Only frames 1 and 2 were used
Test 12	Only frames 1,2 and 3 were used
Test 13	Only frames 1,2,6,7 and 8 were used
Test 14	Only frames 1,2,4 and 5 were used

8.3 Experiment one

The automated abnormality framework described in Chapter 7 was used to extract 36 features for every frame/abnormality present for each patient. Every feature extracted was appended and stored in a two-dimensional (2D) tabular data structure called a data frame. Two data frames were created for both experiments, one for the HL patient cohort (n=10) and one for the TB patient cohort from the RADLAC group (n=3). Every test performed within the first experiment was stored in a data frame in the corresponding *'Test x'* folder. In each test the SVM had a binary outcome as either an HL or TB patients assigned as 1 and 0, respectively. Table 8.2 provides the detailed characteristics of the dataset used for experiment 1.

Table 8.2: Demographic and clinical characteristics of the cohort used in experiment 1.

	HL patients (n = 10)	RADLAC – (TB confirmed) (n = 3)
Male	7	0
Female	3	3
Average Age	41	34
HIV positive	8	3
HIV negative	2	0
TB empiric treatment	6	0
TB proven treatment	0	2
TB organism found	0	3
TB and HIV positive	0	2
TB only	0	1

8.3.1 Experiment one: Training, testing, and evaluating the support vector machine classifier.

A data frame for each cohort was developed. The two data frames were joined to create one final data frame called *'Final_Dataset'* The data frame was spilt into training and test sets and labelled as either X for inputs or Y for the outputs. The Y labels for patients were either a 1 for an HL patient or 0 for a TB patient. The X or inputs were all the features extracted for each patient. A summary and description of the tests performed is provided in Table 8.3. For each test, the same SVM variables were changed together with a specific feature or model hyper parameter explicitly described in Table 8.3 below.

Table 8.3: The variable values used for each test.

Test Case	**SVM Variables**				
	Gamma	C	Test Size	Kernel	Specific Variable value
Test 1	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Only SVM variables
Tet 2	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Frames changed
Test 3	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Gray levels = 64
Test 4	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Gabor frequency = 1/100
Test 5	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Gabor kernel size = 20 x 20
Test 6	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	32 features
Test 7	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	32 features
Test 8	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	28 features
Test 9	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	24 features
Test 10	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	32 features
Test 11	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Frames changed
Test 12	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Frames changed
Test 13	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Frames changed
Test 14	0.5 and 1	1 or 100	0.1 or 0.2	RBF/poly/ linear	Frames changed

8.3.2 Experiment one: Support vector machine results

The HL and RADLAC groups were imbalanced as 10 HL patients had at least one abnormality but only 3 of the 5 RADLAC patients had at least one abnormality, with all 3 having confirmed TB. The k-fold cross validation leave one out (LOO) method as described in section 3.6.3 was used to mitigate or reduce the overfitting of data in every test when evaluating the SVM model.

Upon varying the parameters, the gamma value was changed from 0.5 to 10, however no noticeable changes in SVM performance was observed. When the Gabor frequency hyper parameter was increased, the simulation time was quite noticeably longer, potentially due to the specifications of the laptop used. Table 8.4 presents the highest optimal combination of accuracies achieved in each test. The training accuracy and standard deviation (SD) was captured together with the validation (prediction) accuracy for each test performed.

Table 8.4 Final overall accuracy when training the model with LOO cross validation principle.

Test Case	Highest training accuracy (SD)	Highest validation accuracy
Test 1	0.76 (0.42)	0.8
Test 2	0.77 (0.42)	0.50
Test 3	0.79 (0.41)	0.67
Test 4	0.76 (0.42)	0.8
Test 5	0.76 (0.42)	0.8
Test 6	0.76 (0.42)	0.8
Test 7	0.76 (0.42)	0.8
Test 8	0.76 (0.42)	0.8
Test 9	0.76 (0.42)	0.8
Test 10	0.76 (0.42)	0.8
Test 11	0.78 (0.42)	1
Test 12	0.69 (0.46)	0.5
Test 13	0.81 (0.39)	1
Test 14	0.80 (0.40)	1

8.3.3 Experiment one: Discussion

Test 13 achieved the highest performance using the features from frames 1,2,6,7 and 8. The frames corresponded to an enlarged spleen, splenic micro abscesses, ascites, pleural effusions, and pericardial effusions. A training accuracy of 81% and validation accuracy of 100% was achieved to differentiate between an HL and TB patient using LOO cross validation. Further investigation into the test set of the highest performing model showed that there were only HL patients in the test set. This may indicate a potential bias in the model towards HL classification, attributed to a larger HL patient population used to train the model and not enough data to evaluate the model.

Interestingly the change in Gabor filter size, Gabor frequency and reducing the number of features did not change the accuracy when compared to reducing the number of abnormalities used to differentiate the two classes of patients. The number of features used to train the model did not change the performance of the classification model but a noticeable drop in performance occurred when only frames 1, 2 and 3 were used to train the model. In conclusion, due to the small sample sizes further analysis must be completed to improve and further validate the accuracy to prove the hypothesis of using automated feature detection to improve HL diagnosis.

8.4 Experiment 2

The data frames developed for experiment one was used in experiment 2 but an additional data frame was created for the 3 patients from the RADLAC group only assesed by radiologist 2. Using the abnormality framework developed in objective 3, a total of 36 features were extracted for every frame/abnormality present for each patient. As each patient frame was passed through the framework the extracted features were appended and stored in data frames specific to each test and saved in the corresponding *'Test x'* folder. All RADLAC patients with TB or no TB were assigned a binary label of 0 and HL patients a value of 1. Table 8.5 provides the detailed characteristics of the dataset used for experiment 2.

Table 8.5: Demographic and clinical characteristics of the cohort used in experiment 2.

	HL patients (n = 10)	**RADLAC – (TB confirmed) (n = 3)**	**RADLAC (non-TB) (n=3) Test Set**
Male	7	0	1
Female	3	3	2
Average Age	41	34	41
HIV positive	8	2	3
HIV negative	2	1	0
TB empiric treatment	6	0	1
TB proven treatment	0	2	0
TB organism found	0	3	0
TB and HIV positive	0	2	0
TB only	0	1	0

8.4.1 Experiment two: Training, testing, evaluation and results the support vector machine classifier.

The same methodology used for experiment 1 was used in experiment 2 except the 10 HL and 3 TB patients were all used to train the classification model and the testing was performed on the 3 RADLAC patients whom only radiologist 2 assessed.

Similar to experiment one due to the small dataset size, the HL and RADLAC groups were imbalanced. The k-fold cross validation leave one out (LOO) method as described in section 3.6.3 was used to mitigate or reduce the overfitting of data in every test when evaluating the SVM model.

Table 8.6 presents the highest optimal combination of accuracies achieved in each test. The training accuracy and SD was captured together with the validation (prediction) accuracy for each test performed.

Table 8.6: Final overall accuracy evaluated with the LOO cross validation principle.

Test Case	Highest training accuracy (SD)	Highest validation accuracy
Test 1	0.59 (0.49)	0.6
Tet 2	0.6 (0.49)	0.6
Test 3	0.77(0.42)	0.2
Test 4	0.50 (0.5)	0.6
Test 5	0.77 (0.42)	0
Test 6	0.86 (0.34)	0
Test 7	0.64 (0.48)	0.4
Test 8	0.64 (048)	0.4
Test 9	0.64 (0.48)	0.4
Test 10	0.64 (0.48)	0.6
Test 11	0.80 (0.40)	1
Test 12	0.73 (0.44)	1
Test 13	0.83 (0.37)	1
Test 14	0.83 (0.37)	1

8.4.2 Discussion

Tests 13 and 14 achieved the highest performance. In both tests only a subset of frames were used to differentiate an HL patient from a RADLAC patient. Further insights into the highest achievable performance of 100% in these tests needs to be understood by using a larger dataset to ensure the model is not trained with a bias towards an HL patient as there were many more HL patients with an abnormality present when compared to a TB/RADLAC patient.

Similar to experiment 1 the accuracy did not change when the Gabor frequency, Gabor kernel size or if the features were reduced. The significant increase in both training and validation accuracy changed when specific abnormalities were used to differentiate the two classes of patients. The accuracy across the tests in experiment 2 indicated a lower validation accuracy which may indicate a less biased system as experiment 1 where the model had a larger feature set to learn from for the HL population. However, further analysis is required to validate and prove the hypothesis of using automated feature detection to improve HL diagnosis as the validation accuracy is limited to a small group of patients.

8.5 Conclusion

The SVM is the most widely applied model for supervised classification and typically yields a mean accuracy level of 80% or more (Nascimento et al., 2016). The models developed in this project achieved 81% and 83% training accuracies. However, the reasoning for why the model produced these optimal results requires further testing using larger datasets as there is a bias towards HL patients due to the uneven dataset sizes of the HL and RADLAC patients used in the experiments.

The optimal combination of abnormalities producing the highest accuracies in both experiments for test 13 correlated to abnormalities perceived to cause some ambiguity. An enlarged spleen, splenic micro abscesses, ascites, pleural effusions, and pericardial effusions influenced the training accuracy achieving 81% and 83% in each experiment and comparable to a reported mean accuracy benchmark of 80% for SVM classification models (Nascimento et al., 2016). This may indicate that these specific abnormalities are sufficient to differentiate HL patients from patients without HL but understanding the reasoning for the decision taken by the system requires further investigation.

Nascimento et al (2016) provided evidence of the SVM approach as more advantageous when compared to alternate learning methods using a reduced feature set. The SVM approach under-performed with reduced prediction accuracy when a large feature set was used, indicating a potential disadvantage of an SVM classification model to differentiate and characterise patients with large feature sets. Future work should consider including a larger patient cohort with a larger prevalence of abnormalities of interest. Additionally, a prospective study using the developed protocol may enhance the ability to capture a larger number of abnormalities of interest to further evaluate the classification model.

The developed model was trained to learn textural and geometrical features from ultrasound images to classify and differentiate HL using several abnormalities of interest. However, the training and prediction accuracies obtained in each test achieved are biased and produce optimistic predictive performance. Considering the small dataset utilised further evaluation is required to validate the use of SVM models to improve the HL diagnostic pathway.

9. Conclusion

The primary aim of the research project was to develop and validate an automated feature detection framework to identify biomarkers of disease progression for Hodgkin's lymphoma (HL) within a tuberculosis (TB) and Human Immunodeficiency Virus (HIV) endemic region, using abdominal ultrasound images. The research was clinically motivated by the need to identify biomarkers for HL earlier on in the diagnostic pathway to prevent misdiagnosis of TB in HIV patients, who all have similar clinical representations to HL.

9.1 Summary of findings

9.1.1 Development of an ultrasound imaging protocol

A cohort of HL and rapid lymphadenopathy (RADLAC) patients were identified and selected for the project. Seven abdominal abnormalities were identified using a set of ultrasound images from the identified HL cohort. Splenomegaly, splenic lesions, splenic microabscesses, lymph node enlargement in three anatomical regions (epigastric, mesenteric, and retroperitoneal), ascites, pericardial effusion, and lastly pleural effusion. These specific abnormalities are known to be potential markers for disease but are often missed due to technical faults such as incorrect radiological protocol adherence. Additionally, these specific biomarkers may be misinterpreted due to a biased association of their prevalence with a differential diagnosis of TB favoured over HIV. Interestingly, current research evaluates the impact of these abnormalities in isolation and within the spheres of either TB, HIV or HL (Heller et al., 2012; Bhatt el al., 2015; Liu, Wang et al. 2019). Therefore, in order to build an automated algorithmic framework, a precise protocol was developed to identify specific ultrasound frames to capture the abnormalities of interest for this project.

The developed protocol has a set of eight distinct frames and provides a novel guideline to aid abnormality detection of HL patients within TB and HIV endemic regions. Literature does not describe specific imaging protocols used for HL patients in these regions and focus is placed on TB and HIV only (Heller et al., 2012). Additionally, current guidelines do not explicitly cater to prevalent overlapping disease states in HIV and TB endemic regions. Developing reproducible and consistent studies is therefore a challenge. The developed protocol aims to mitigate the problem and aid future studies by providing a coherent image acquisition guideline for HL within TB and HIV endemic regions. Furthermore, it provides a concise approach to further validate the hypothesis that abnormality patterns seen on ultrasound can differentiate TB, HIV or HL earlier on in the diagnostic pathway.

Two radiologists and a clinician guided the development of the protocol and chose frames based on their experience and clinical knowledge. However there exists a potential bias when identifying the optimal frame for abnormality capture. A prospective study is therefore required to validate the developed protocol for first point of care in ultrasound imaging examinations. Validation of the developed protocol may encourage clinicians to trust the use of a cheaper and more available imaging tool to enhance the ability to consistently identify specific biomarkers which overlap across TB, HIV and HL and prevent missed HL diagnostic opportunities. Moreover, the groundwork to reduce variation in image capture when performing the retrospective analysis provides a consistent approach that can be used in future work to enhance the ability to validate and benchmark research findings against one another.

9.1.2 Ground truth feature development

Feature development using ultrasound has been used to characterise texture predominantly for breast lesion classification models (Brattain et al, 2018). In this study multiple abnormalities are characterised with a set of textural and geometrical properties that can be extracted for several abnormalities of interest. This approach provides a more insightful and comprehensive view for clinicians, identifying various combinations of biomarkers that may be representative of overlapping comorbidities and malignancies such as TB, HIV and HL, respectively.

Each of the seven abnormalities were captured using the developed protocol as a guideline. Each abnormality was segmented with assistance from the two radiologists to create ground truth segmentations. Image processing techniques were used to reduce noise and the effects of varying illumination settings due to the different types of ultrasound scanners used. Functions were developed to extract geometrical and textural properties to characterise every ground truth segmentation or abnormality with guidance from two radiologists and a clinical haematologist.

Image processing techniques such as normalisation and denoising aim to enhance the quality of images before features are extracted (Canuma, 2018; Poudel et al., 2018, Kociołek et al., 2020). Various image processing techniques are often performed retrospectively and limited by the inability to control scanner settings utilised during acquisition. This results in a potential drawback of incorrectly characterising a clinically relevant feature such as texture, impacting downstream analysis (Kociołek et al., 2020). Normalising to a smaller intensity range is standard practise to reduce the effect of varying illumination differences due to multiple scanner settings and was used in this project. However, the ideal method would be a prospective study to set specific scanner

settings and reduce a potential uncertainty of correct textural characterisation, notably important when characterising ground truth features.

9.1.3 Automated abnormality characterisation framework

The functions developed to extract geometrical and textural features were used to build an automated algorithmic abnormality characterisation framework. The framework required two mandatory fields to specify; first, the specific drive or path containing patients' set of segmented frames, and secondly, a specific patient number for the required features to be extracted. The algorithm would then automatically extract features and append them to a data frame for every available frame (abnormality) present in every patient's folder. A total of 36 features could be successfully extracted to characterise each of the abnormalities.

A drawback of a single generisable framework is the prevention of modular modification to feature extraction functions. The risk here is that a single parameter change within the functions impacts all abnormalities and subsequently, extracted features. Furthermore, a technical difficulty which was unavoidable was the average value used for the geometric size calculation, which produced a large deviation from ground truth measurements. The difference was attributed to the imaging acquisition settings which were not controllable as this was a retrospective study. The lack of consistent settings and incorrect set up of scanners significantly contributed to inaccuracies in measurements calculated automatically as a mean value was used for the pixel conversion ratio. However, this may be mitigated in future work by switching on the correct settings on the scanner and calibrating during image capture to ensure pixel settings are the same and retrieved directly from an images metadata.

9.1.4 Evaluation of an automated classification model

The literature review identified a gap within HL research in TB and HIV endemic countries highlighting the lack of research on image feature extraction techniques to build HL classification models. The developed support vector machine (SVM) classification model built in this project aimed to contribute to the identified research gap. An SVM model was trained to learn textural and geometrical features from ultrasound images to classify and differentiate HL. The classification model achieved a training accuracy similar to the mean benchmark of 80% obtained in SVM classification models (Nascimento et al., 2016). Interestingly, the training accuracies achieved in the project indicated specific abnormalities were sufficient to differentiate HL patients from patients without HL. However, understanding the reasoning for the decision taken by the system requires further investigation.

A drawback of the small patient population used in the final evaluation of the classification model was the large and overfitted validation accuracies obtained. The near perfect accuracies were attributed to the large difference in HL versus RADLAC patient population sizes, producing a classification bias towards the HL cohort. However, this was the first study to evaluate the potential of automated approaches for HL diagnosis within a TB and HIV endemic region. The lack of research invested into low resource health care environments with large TB and HIV endemic regions, prevented direct comparisons of classification results obtained for the HL cohort. As such, the work developed in this study provides an initial proof of principle for utilising automated image analysis tools to diagnose HL. Although limited in scope and data, the presented algorithmic framework may provide a good starting point towards HL automated feature extraction and automated model development. This in turn may provide and create more accessible and consistent diagnosis for HL within the African healthcare context.

9.2 Limitations and recommendations for future work

The study was a retrospective analysis using available ultrasound images acquired from multiple scanners with varying settings, which limited the number of images that met the criteria for analysis, and reduced the total cohort analysed within study. A prospective study that includes the developed protocol for image capture and multimodal data to incorporate all types of imaging could potentially improve and provide a more realistic clinical presentation of HL. Additionally, it may reduce bias and achieve more realistic accuracy metrics for an automated HL classification model using a larger patient cohort.

A combination of 36 textural and geometrical features were successfully extracted to characterise each of the seven abnormalities. However, the abnormality size measurement calculated using the automated abnormality framework had a noticeable difference when compared to the ground truth measurements. This was attributed to the average conversion value used across all images irrespective of the individual scanner settings due to the lack of the *pixel spacing* meta data available on the scans. A direct extraction of the metadata for each scan would improve the automated size measurement and limit the variation in the calculate sizes when compared to ground truth measurements.

The preliminary results observed in the study indicate the potential use of automated techniques to aid the diagnostic pathway earlier on to differentiate between overlapping diseases states such as HL, TB, and HIV. Interestingly, observing the results, the similarities of abdominal ultrasound abnormalities between EPTB and HL patients highlights why there is confusion between these

disease states when scans are interpreted. This indicates that if EPTB is suspected, HL could be an equally valid diagnostic outcome. However, a much larger cohort should be used to maximise the evaluation of the automated classification model to develop associations of features that may represent a more distinct differentiation amongst patient cohorts.

The SVM technique used to develop the automated algorithm is part of the growing area of research in the field of artificial intelligence. Various techniques are actively researched for medical applications with recent interest tending towards the development of more reliable and interpretable automated models. The SVM is an established method and more interpretable compared to newer more complex deep learning algorithms, such as neural networks, which use many layers of computations to extract finer levels of detail. These deep learning algorithms often require larger sets of well-defined data, that are well annotated and adhere to clear and well-defined protocols for image capture. To leverage these deep learning models a much larger well curated dataset would be required for our application. However, these complex models may not be as interpretable as established techniques like SVM. Furthermore, for prospective studies, it should be noted that acquiring a larger set of images to improve prediction accuracies should not be favoured over poorly labelled images, which contributes to a lower predictive performance of supervised classification models (Nalepa & Kawulok, 2019). A hybrid approach for retrospective analysis could be used by incorporating available sensitive modalities like computed tomography (CT) to complement ultrasound images and enhance the quality of outcomes when training a classification model. This approach could be investigated to enhance an automated classification model's ability to encapsulate the entire clinical representation of patient cohorts from medical images and improve the accuracy of automated classification predictions (Nalepa and Kawulok 2019).

Lastly this project was a first step towards identifying the techniques and protocols required to create an automated system to improve the differential diagnostic pathway of HL. The development of an end-to-end solution in a healthcare setting was out of scope for this project but would require further evaluation and validation of the algorithm, using additional data. Furthermore, testing and validation of the developed protocol should be performed to ensure it is robust and can consistently assist radiologists to acquire the required frames for a differential diagnosis of HL. Once these validation steps are complete, further enhancements and development is required, namely a user-friendly application for radiologists. In the long term the application and protocol could be developed together and reside within the ultrasound machine itself. This

would provide the ability to scan and display the potential diagnosis in real time to the radiologist to prevent any missed opportunity for an HL diagnosis.

9.3 Overall conclusions and contribution of the project

The primary aim to develop an algorithmic framework to automatically extract textural and geometrical features from a specific sequence of frames, was achieved. Development and evaluation was performed using ultrasound images from the identified HL and RADLAC patient cohorts. Thereafter an automated classification model was built to learn the features extracted to characterise and identify HL. However, the limitations to assess the entire available cohort and the lack of consistent frame capture contributed to overfit validation accuracies. Additionally, the imbalanced datasets skewed classification, overstating the predictive performance with a bias towards HL patient identification.

Despite these challenges, the small dataset analysed by both radiologists indicates the project is a successful preliminary analysis. However, further evaluation is required to understand and interpret the effects of each feature in the model within a larger cohort of HL patients living in a TB and HIV endemic environment.

Appendix A

Table A.1: A single consolidated view of the protocol and all descriptors.

Category	Spleen	Splenic Lesion	Splenic microabscess	Lymph node Epigastric	Lymph node Retroperitoneal	Lymph node Mesenteric	Ascites	Pericardial Effusion	Pleural Effusion	Radiologist/s findings
Frame Number of new protocol/ sequence	**Frame 1**		**Frame 2**	**Frame 3**	**Frame 4**	**Frame 5**	**Frame 6**	**Frame 7**	**Frame 8**	
Status (overall)	Enlarged	Visible but need to see Frame 2 of protocol to determine if it is microabscessess or is it a lesion. Frame 2 view not required. No other view available and differential on lesion/microabscess to be made on Frame 1.	Present	Normal or Abnormal	Normal or Abnormal	Normal or Abnormal	Present	Present	Present	Radiologist/s commentary on what the observed abnormalities may indicate
Shape	N/A	Round or Irregular	Round or Irregular	Normal (kidney-shaped) or Irregular or Round or Ovoid	Normal (kidney-shaped) or Irregular or Round or Ovoid	Normal (kidney-shaped) or Irregular or Round or Ovoid	N/A	N/A	N/A	
Size	>=130 mm	>10 mm	>5mm	>10 mm	>10 mm	>10 mm	Sliver of fluid	Sliver of fluid	Sliver of fluid	
	<130 mm	<=10 mm	<=5mm	<=10 mm	<=10 mm	<=10 mm	Free fluid	Free fluid	Free fluid	
Frequency	N/A	Single or Innumerable	Single or Innumerable or Multiple	Single or Multiple	Single or Multiple	Single or Multiple	N/A	N/A	N/A	
EchoTexture	Homogenous or Heterogeneous	Homogenous or Heterogeneous	Hypoechoic or Hyperechoic	Hypoechoic or Hyperechoic	Hypoechoic or Hyperechoic	Hypoechoic or Hyperechoic	N/A	N/A	N/A	
Location	N/A	N/A	N/A	Epigastric	Retroperitoneal	Mesentery	Pelvic fluid or Intra-abdominal	N/A	N/A	
Additional Information										
Description of Frame view for new protocol and sequence	Longitudinal view along long axis of the spleen		Longitudinal view along long axis of the spleen BUT images acquired with linear high frequencty probe and zoomed in to see abscesses.	View the epigastric region lcoated around the pancreas	View the retroperitoneal region, located around aorta and above and below renal veins.	View of the mesentery , located around the bowel area.	View of the adominal/pelvic region. Specifically the transhepatic longitudinal section wrt right kidney (most sensitive area to identify abnormality).	View the epigastric area angled to see heart from under the diaphragm	A longitudinal view along the sides of the ribs	N/A
Detailed descriptions for locations of lymph nodes, based on the 3 anatomical categories				Epigastric lymph nodes are located in upper abdomen and coeliac region 1. Porta hepatic/peri portal/ peri pancreatic/ peri splenic (could be acessory spleen and should be verified).	Retroperitoneal lymph nodes are located in three regions: 1. Superior mesenteric lymph node/s by the superior mesenteric artery 2. Iiliac chain node/s 3. Para aortic lymph node/s	Mesenteric region has the Iliac fossa node/s				
Landmarks for segmenting abnormalities (in order of precedence)	1. Diaphgram (superior margin to lower pole of spleen) 2. Hilum - fat is bright and the vessels to spleen must be seen 3. The orientation is through hilum (important) 4. Typically a bit bean shaped 5. Left kidney is a landmark used over and above points 1-3 as it is relative to the position of the probe	N/A - Frame 1 - based on finding the spleen and measured on texture	Based on finding the spleen and measured on texture using a linear probe	1. Liver 2. Porta hepatis 3. Pancreas 4. Stomach (diffcult to see sometimes on ultrasound)	Identify the relative location of the inferior vena cava and the abdominal aorta	Identify the branches of the superior mesenteric artery and vein	1. Morissons pouch - between right kidney and liver 2. Pelvis - behind bladder	Identitfy heart and look around for excess fluid	1. Diapghram 2. Lungs 3. Spleen (left effusion) 4. Liver (right effusion)	

Table A.2: A snippet of the drop-down list (spleen only shown) utilised by the radiologists to capture abnormalities.

Patient Number	**Spleen**				
Descriptions	Original US Image Number	Image number in sequence	Status	Size as per report else resized value	Echotexture

Table A.3: Radiologists 1's findings

Patient (HL/RADLAC)	Enlarged Spleen >130mm	Splenic Lesion	Splenic microabs	Lymph nodes Epigastric	Lymph nodes Retroperitoneal	Lymph nodes Mesenteric	Ascites	Pericardial Effusion	Pleural Effusion	Suggestive indication or differential diagnosis
4HL	1			1			1			Lymphoma
6HL		1							1	Lymphoma
10HL					1				1	Lymphoma
19HL		1					1			Cant say
24HL	1							1		Lymphoma
25HL	1	1					1			Lymphoma
28HL	1		1							Lymphoma
30HL			1				1		1	TB
32HL	1			1						Lymphoma
35HL	1	1		1						Lymphoma
1 RADLAC										Cant say
2 RADLAC			1	1						TB
3 RADLAC	1	1								Lymphoma
6 RADLAC										Cant say
17 RADLAC				1					1	TB
Total	7	5	3	5	1	0	4	1	4	

Table A.4: Radiologists 2's findings

Patient (HL/RADLAC)	Enlarged Spleen >130mm	Splenic Lesion	Splenic microabs	Lymph nodes Epigastric	Lymph nodes Retroperitoneal	Lymph nodes Mesenteric	Ascites	Pericardial Effusion	Pleural Effusion	Suggestive indication or differential diagnosis
4HL	1			1	1		1			TB
6HL		1							1	TB
10HL					1					Lymphoma
19HL							1			Cant say
24HL	1							1		TB
25HL	1	1					1			Lymphoma
28HL	1		1							TB
30HL			1				1		1	Lymphoma
32HL	1									Lymphoma
35HL	1	1		1						Lymphoma
1 RADLAC										TB
2 RADLAC			1	1						TB
3 RADLAC	1	1					1			TB
6 RADLAC										TB
17 RADLAC				1					1	TB
7 RADLAC	1			1						Lymphoma
8 RADLAC					1				1	Cant say
10 RADLAC			1							Cant say
Total	8	4	4	5	3	0	5	1	4	

Table A.5: A final consolidated table of mutual abnormalities between radiologists utilised for the automated frameworks

Patient (HL/RADLAC)	Enlarged Spleen >130mm	Splenic Lesion	Splenic microabs	Lymph nodes Epigastric	Lymph nodes Retroperitoneal	Lymph nodes Mesenteric	Ascites	Pericardial Effusion	Pleural Effusion
4HL	1			1			1		
6HL		1							1
10HL					1				
19HL							1		
24HL	1							2	
25HL	1	1					1		
28HL	1		1						
30HL			1				1		1
32HL	1								
35HL	1	1		1					
1 RADLAC									
2 RADLAC			1	1					
3 RADLAC	1	1							
6 RADLAC									
17 RADLAC				1					1
Mismatched frames				1			1		
Total	7	4	3	3	1	0	3	1	3
Total abnormalities	**25**								

References

Afshar, P., Mohammadi, A., Plataniotis, K. N., Oikonomou, A., & Benali, H. (2019). From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. IEEE Signal Processing Magazine, 36(4), 132-160. doi:10.1109/MSP.2019.2900993

Antel, K., Levetan, C., Mohamed, Z., Louw, V. J., Oosthuizen, J., Maartens, G., & Verburgh, E. (2019). The determinants and impact of diagnostic delay in lymphoma in a TB and HIV endemic setting. BMC Cancer, 19(1), 384. doi:10.1186/s12885-019-5586-4

Antel, K., & Verburgh, E. (2019). Lymphadenopathy in a tuberculosis-endemic area: Diagnostic pitfalls and suggested approach (Vol. 109).

Anutam, & Rajni, R. (2014). Performance Analysis of Image Denoising with Wavelet Thresholding Methods for Different Levels of Decomposition. The International journal of Multimedia & Its Applications, 6, 35-46. doi:10.5121/ijma.2014.6303

Bai, J., Jiang, H., Li, S., & Ma, X. (2019). NHL Pathological Image Classification Based on Hierarchical Local Information and GoogLeNet-Based Representations. BioMed research international, 2019, 1065652-1065652. doi:10.1155/2019/1065652

Barta, S. K., Xue, X., Wang, D., Lee, J. Y., Kaplan, L. D., Ribera, J.-M., Sparano, J. A. (2014). A new prognostic score for AIDS-related lymphomas in the rituximab-era. Haematologica, 99(11), 1731-1737. doi:10.3324/haematol.2014.111112

Berzaczy, D., Haug, A. R., Raderer, M., Kiesewetter, B., Berzaczy, G., Weber, M., & Mayerhoefer, M. E. (2019). Is there a reliable size cut-off for splenic involvement in lymphoma? A [18F] FDG-PET controlled study. PLOS ONE, 14(3), e0213551. doi: 10.1371/journal.pone.0213551

Bhatt, N., Brandt, G., Niebrugge, D., & Khan, A. U. (2015). Prognostic Value of Pleural Effusion in Hodgkin's Lymphoma Patients. Blood, 126(23), 5001-5001. doi:10.1182/blood.V126.23.5001.5001

Białek, E. J., & Jakubowski, W. (2017). Mistakes in ultrasound diagnosis of superficial lymph nodes. Journal of ultrasonography, 17(68), 59-65. doi:10.15557/JoU.2017.0008

Bianconi, F., & Fernández, A. (2007). Evaluation of the effects of Gabor filter parameters on texture classification. Pattern Recognition, 40, 3325-3335. doi: 10.1016/j.patcog.2007.04.023

Brattain, L. J., Telfer, B. A., Dhyani, M., Grajo, J. R., & Samir, A. E. (2018). Machine learning for medical ultrasound: status, methods, and future opportunities. Abdominal Radiology, 43(4), 786-799. doi:10.1007/s00261-018-1517-0

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. Radiographics: a review publication of the Radiological Society of North America, Inc, 35(6), 1668-1676. doi:10.1148/rg.2015150023

Calvo-Lobo, C., Useros-Olmo, A. I., Almazán-Polo, J., Martín-Sevilla, M., Romero-Morales, C., Sanz-Corbalán, I., López-López, D. (2018). Quantitative Ultrasound Imaging Pixel Analysis of the Intrinsic Plantar Muscle Tissue between Hemiparesis and Contralateral Feet in Post-Stroke Patients. International journal of environmental research and public health, 15(11), 2519. doi:10.3390/ijerph15112519

Canuma. (2018). Image Pre-processing. Retrieved from https://towardsdatascience.com/image-pre-processing-c1aec0be3edf

Caremani, M., Occhini, U., Caremani, A., Tacconi, D., Lapini, L., Accorsi, A., & Mazzarelli, C. (2013). Focal splenic lesions: US findings. Journal of ultrasound, 16(2), 65-74. doi:10.1007/s40477-013-0014-0

Chang, S. G., Bin, Y., & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. IEEE Transactions on Image Processing, 9(9), 1532-1546. doi:10.1109/83.862633

Chen, M.-J., Huang, M.-J., Chang, W.-H., Wang, T.-E., Wang, H.-Y., Chu, C.-H., Shih, S.-C. (2005). Ultrasonography of splenic abnormalities. World journal of gastroenterology, 11(26), 4061-4066. doi:10.3748/wjg. v11.i26.4061

Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., Chen, C.-M. (2016). Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. Scientific reports, 6, 24454-24454. doi:10.1038/srep24454

Choi, H., & Jeong, J. (2020). Despeckling Algorithm for Removing Speckle Noise from Ultrasound Images. Symmetry, 12, 1-26. doi:10.3390/sym12060938

Ciurte, A., Bresson, X., Cuisenaire, O., Houhou, N., Nedevschi, S., Thiran, J.-P., & Cuadra, M. B. (2014). Semi-Supervised Segmentation of Ultrasound Images Based on Patch Representation and Continuous Min Cut. PLOS ONE, 9(7), e100972. doi: 10.1371/journal.pone.0100972

Cuccaro, A., Bartolomei, F., Cupelli, E., Galli, E., Giachelia, M., & Hohaus, S. (2014). Prognostic Factors in Hodgkin Lymphoma. Mediterranean journal of hematology and infectious diseases, 6, e2014053. doi:10.4084/MJHID.2014.053

Damodaran, N. (2009). Implementation of Wavelet Filters for Speckle Noise Reduction in Ultrasound Medical Images: A Comparative Study.

Danjuma, K. J. (2015). Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients. ArXiv, abs/1504.04646.

Deeley, M. A., Chen, A., Datteri, R., Noble, J. H., Cmelak, A. J., Donnelly, E. F., Dawant, B. M. (2011). Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. Physics in medicine and biology study, 56(14), 4557-4577. doi:10.1088/0031-9155/56/14/021

Deka, B., & Bora, P. K. (2013). Wavelet-based Despeckling of Medical Ultrasound Images. IETE Journal of Research, 59(2), 97-108. doi:10.4103/0377-2063.113026

Dioşan, L., Rogozan, A., & Pecuchet, J.-P. (2012). Improving classification performance of Support Vector Machine by genetically optimising kernel shape and hyper-parameters. Applied Intelligence, 36(2), 280-294. doi:10.1007/s10489-010-0260-1

Ehimwenma, O., & Tagbo, M. T. (2011). Determination of normal dimension of the spleen by ultrasound in an endemic tropical environment. Nigerian medical journal: journal of the Nigeria Medical Association, 52(3), 198-203. doi:10.4103/0300-1652.86141

Eitrich, T., & Lang, B. (2006). Efficient optimization of support vector machine learning parameters for unbalanced datasets. Journal of Computational and Applied Mathematics, 196, 425-436. doi: 10.1016/j.cam.2005.09.009

Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine Learning for Medical Imaging. Radiographics: a review publication of the Radiological Society of North America, Inc, 37(2), 505-515. doi:10.1148/rg.2017160130

Evans, D. (2013). Ten years on ART – where to now? (Vol. 103).

Figshare. (2012). Figshare.

Forghani, R., Savadjiev, P., Chatterjee, A., Muthukrishnan, N., Reinhold, C., & Forghani, B. (2019). Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Computational and structural biotechnology journal, 17, 995-1008. doi: 10.1016/j.csbj.2019.07.001

Forsyth, e. a. (2003). Computer Vision: Modern Approach. Retrieved from https://www.academia.edu/38213969/Computer_Vision_A_Modern_Approach_2nd_Edition

Foster, K. R., Koprowski, R., & Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomedical engineering online, 13, 94-94. doi:10.1186/1475-925X-13-94

Gaiolla, R. D. (2017). Hodgkin's lymphoma in developing countries: can we go further? Revista brasileira de hematologia e hemoterapia, 39(4), 299-300. doi: 10.1016/j.bjhh.2017.08.004

Ganeshalingam, S., & Koh, D.-M. (2009). Nodal staging. Cancer imaging: the official publication of the International Cancer Imaging Society, 9(1), 104-111. doi:10.1102/1470-7330.2009.0017

Garra, B. S., Krasner, B. H., Horii, S. C., Ascher, S., Mun, S. K., & Zeman, R. K. (1993). Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrasound Imaging, 15(4), 267-285. doi:10.1177/016173469301500401

Geijer, H., & Geijer, M. (2018). Added value of double reading in diagnostic radiology, a systematic review. Insights into Imaging, 9(3), 287-301. doi:10.1007/s13244-018-0599-0

Gómez-Flores, W., & Ruiz-Ortega, B. A. (2016). New Fully Automated Method for Segmentation of Breast Lesions on Ultrasound Based on Texture Analysis. Ultrasound in Medicine & Biology, 42(7), 1637-1650. doi: https://doi.org/10.1016/j.ultrasmedbio.2016.02.016

Graps, A. (1995). An Introduction to Wavelets. IEEE Computer. Sci. Eng., 2(2), 50–61. doi:10.1109/99.388960

Griesel, R., Cohen, K., Mendelson, M., & Maartens, G. (2019). Abdominal Ultrasound for the Diagnosis of Tuberculosis Among Human Immunodeficiency Virus-Positive Inpatients with World Health Organization Danger Signs. Open forum infectious diseases, 6(4), ofz094-ofz094. doi:10.1093/ofid/ofz094

Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M., & Notar, M. (2018). An application of machine learning to haematological diagnosis. Scientific reports, 8(1), 411-411. doi:10.1038/s41598-017-18564-8

Gupta, M., Taneja, H., & Chand, L. (2018). Performance Enhancement and Analysis of Filters in Ultrasound Image Denoising. Procedia Computer Science, 132, 643-652. doi: https://doi.org/10.1016/j.procs.2018.05.063

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian journal of internal medicine, 4(2), 627-635. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24009950

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/

Hallinan, J. (2014, 2014). Assessing and Comparing Classifier Performance with ROC Curves.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610-621. doi:10.1109/TSMC.1973.4309314

Hedaoo, P., & Godbole, S. (2011). Wavelet Thresholding Approach for Image Denoising. International Journal of Network Security & Its Applications, 3. doi:10.5121/ijnsa.2011.3402

Heller, T., Wallrauch, C., Goblirsch, S., & Brunetti, E. (2012). Focused assessment with sonography for HIV-associated tuberculosis (FASH): a short protocol and a pictorial review. Crit Ultrasound J, 4(1), 21. doi:10.1186/2036-7902-4-21

Huang, Q., Zhang, F., & Li, X. (2018). Machine Learning in Ultrasound Computer-Aided Diagnostic Systems: A Survey. BioMed research international, 2018, 10. doi:10.1155/2018/5137904

Humeau-Heurtier, A. (2019). Texture Feature Extraction Methods: A Survey. IEEE Access, PP, 1-1. doi:10.1109/ACCESS.2018.2890743

Iberoamerican Congress, o. P. R. P. n. (2018). Progress in pattern recognition, image analysis, computer vision and applications. In M. e. al (Ed.), Progress in pattern recognition, image analysis, computer vision and applications. Switzerland: Springer.

Illanes, A., Esmaeili, N., Poudel, P., Balakrishnan, S., & Friebe, M. (2019). Parametrical modelling for texture characterization—A novel approach applied to ultrasound thyroid segmentation. PLOS ONE, 14(1), e0211215. doi: 10.1371/journal.pone.0211215

Jabarulla, M. Y., & Lee, H.-N. (2018). Speckle Reduction on Ultrasound Liver Images Based on a Sparse Representation over a Learned Dictionary. Applied Sciences, 8(6), 903. Retrieved from https://www.mdpi.com/2076-3417/8/6/903

Jiang, Y., Xie, W., Hu, K., Sun, J., Zhu, X., & Huang, H. (2013). An aggressive form of non-Hodgkin's lymphoma with pleural and abdominal chylous effusions: A case report and review of the literature. Oncology letters, 6(4), 1120-1122. doi:10.3892/ol.2013.1501

Jin, Y., Angelini, E., & Laine, A. (2005). Wavelets in Medical Image Processing: Denoising, Segmentation, and Registration. In J. S. Suri, D. L. Wilson, & S. Laxminarayan (Eds.), Handbook of Biomedical Image Analysis: Volume I: Segmentation Models Part A (pp. 305-358). Boston, MA: Springer US.

Kaur, P., Singh, G., & Kaur, P. (2018). A Review of Denoising Medical Images Using Machine Learning Approaches. Current medical imaging reviews, 14(5), 675-685. doi:10.2174/1573405613666170428154156

Kebede, A. A., & Getaneh, F. B. (2015). PATTERNS OF ULTRASOUND FINDINGS IN ABDOMINAL LYMPHOMA PATIENTS AT TIKUR ANBESSA SPECIALIZED HOSPITAL, ADDIS ABABA, ETHIOPIA. Ethiopian medical journal, 53(4), 199-207. Retrieved from http://europepmc.org/abstract/MED/27182586

Kecman, V. (2005). Support Vector Machines – An Introduction. In (Vol. 177, pp. 605-605).

Kennedy-Nasser, A. A., Hanley, P., & Bollard, C. M. (2011). Hodgkin disease and the role of the immune system. Pediatric hematology and oncology, 28(3), 176-186. doi:10.3109/08880018.2011.557261

Kociołek, M., Strzelecki, M., & Obuchowicz, R. (2020). Does image normalization and intensity resolution impact texture classification? Computerized Medical Imaging and Graphics, 81, 101716. doi: https://doi.org/10.1016/j.compmedimag.2020.101716

Kohli, M. D., Summers, R. M., & Geis, J. R. (2017). Medical Image Data and Datasets in the Era of Machine Learning—Whitepaper from the 2016 C-MIMI Meeting Dataset Session. Journal of digital imaging, 30(4), 392-399. doi:10.1007/s10278-017-9976-3

Krig, S. (2014). Ground Truth Data, Content, Metrics, and Analysis. In S. Krig (Ed.), Computer Vision Metrics: Survey, Taxonomy, and Analysis (pp. 283-311). Berkeley, CA: Apress.

Lahmiri, S., & Boukadoum, M. (2013). Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Features Extraction from Biomedical Images. Journal of Medical Engineering, 2013, 13. doi:10.1155/2013/104684

Leite, N. P., Kased, N., Hanna, R. F., Brown, M. A., Pereira, J. M., Cunha, R., & Sirlin, C. B. (2007). Cross-sectional imaging of extranodal involvement in abdominopelvic lymphoproliferative malignancies. Radiographics: a review publication of the Radiological Society of North America, Inc, 27(6), 1613-1634. doi:10.1148/rg.276065170

Linguraru, M. G., Sandberg, J. K., Li, Z., Shah, F., & Summers, R. M. (2010). Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. Med Phys, 37(2), 771-783. doi:10.1118/1.3284530

Liu, E.-S., Wang, J.-S., & Yang, W. C. (2019). Peritoneal lymphoma with ascites mimicking portal hypertensive ascites: A case report. Medicine, 98(8), e14583-e14583. doi:10.1097/MD.0000000000014583

Liu, H., Tan, T., van Zelst, J., Mann, R., Karssemeijer, N., & Platel, B. (2014). Incorporating texture features in a computer-aided breast lesion diagnosis system for automated three-dimensional breast ultrasound. Journal of medical imaging (Bellingham, Wash.), 1(2), 024501-024501. doi: 10.1117/1.JMI.1.2.024501

Liu, Z., & Xu, H. (2014). Kernel Parameter Selection for Support Vector Machine Classification. Journal of Algorithms & Computational Technology, 8, 163-178. doi:10.1260/1748-3018.8.2.163

Loizou, C. P., Pattichis, C. S., Pantziaris, M., Tyllis, T., & Nicolaides, A. (2006). Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering. Medical and Biological Engineering and Computing, 44(5), 414. doi:10.1007/s11517-006-0045-1

Manzella, A., Borba-Filho, P., D'Ippolito, G., & Farias, M. (2013). Abdominal Manifestations of Lymphoma: Spectrum of Imaging Features. ISRN Radiology, 2013. doi:10.5402/2013/483069

Marks, L. J., McCarten, K. M., Pei, Q., Friedman, D. L., Schwartz, C. L., & Kelly, K. M. (2018). Pericardial effusion in Hodgkin lymphoma: a report from the Children's Oncology Group AHOD0031 protocol. Blood, 132(11), 1208-1211. doi:10.1182/blood-2018-02-834465

Michailovich, O., & Tannenbaum, A. (2006). Despeckling of Medical Ultrasound Images. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 53, 64-78. doi:10.1109/TUFFC.2006.1588392

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review. Artificial Intelligence Review, 52(2), 857-900. doi:10.1007/s10462-017-9611-1

Nascimento, C. D. L., Silva, S. D. d. S., Silva, T. A. d., Pereira, W. C. d. A., Costa, M. G. F., & Costa Filho, C. F. F. (2016). Breast tumor classification in ultrasound images using support vector machines and neural networks. Research on Biomedical Engineering, 32, 283-292. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2446-47402016000300283&nrm=iso

Noble, J. A. (2009). Ultrasound image segmentation and tissue characterization. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 224(2), 307-316. doi:10.1243/09544119JEIM604

Noble, J. A., & Boukerroui, D. (2006). Ultrasound image segmentation: a survey. IEEE Transactions on Medical Imaging, 25(8), 987-1010. doi:10.1109/TMI.2006.877092

Om, Y., Benes, R., & Riha, K. (2012). Suitable Image Intensity Normalization for Arterial Visualization. International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems, 1. doi:10.11601/ijates. v1i2-3.37

Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S., & Goldberg, I. G. (2010). Automatic classification of lymphoma images with transform-based global features. IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society, 14(4), 1003-1013. doi:10.1109/TITB.2010.2050695

Ortiz, S., Chiu, T., & Fox, M. (2012). Ultrasound image enhancement: A review. Biomed. Signal Process. Control., 7, 419-428.

Parekh, V. S., & Jacobs, M. A. (2019). Deep learning and radiomics in precision medicine. Expert review of precision medicine and drug development, 4(2), 59-72. doi:10.1080/23808993.2019.1585805

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity, and predictive values. Indian journal of ophthalmology, 56(1), 45-50. doi:10.4103/0301-4738.37595

Patel. (2012). Hodgkin’s Lymphoma and Human Immunodeficiency Virus Infection. Open access peer-reviewed chapter (Immunodeficiency).

Patel, M., Philip, V., & Fazel, F. (2011). Human Immunodeficiency Virus Infection and Hodgkin Lymphoma in South Africa: An Emerging Problem. Advances in Hematology, 2011. doi:10.1155/2011/578163

Pileri, S. A., Ascani, S., Leoncini, L., Sabattini, E., Zinzani, P. L., Piccaluga, P. P., Stein, H. (2002). Hodgkin's lymphoma: the pathologist's viewpoint. Journal of Clinical Pathology, 55(3), 162. Retrieved from http://jcp.bmj.com/content/55/3/162.abstract

Pontet, J., Yic, C., Díaz-Gómez, J. L., Rodriguez, P., Sviridenko, I., Méndez, D., Cancela, M. (2019). Impact of an ultrasound-driven diagnostic protocol at early intensive-care stay: a randomized-controlled trial. The Ultrasound Journal, 11(1), 24. doi:10.1186/s13089-019-0139-2

Poudel, P., Illanes, A., Sheet, D., & Friebe, M. (2018). Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches. Journal of Healthcare Engineering, 2018, 13. doi:10.1155/2018/8087624

Prabusankarlal, K. M., Thirumoorthy, P., & Manavalan, R. (2017). Classification of breast masses in ultrasound images using self-adaptive differential evolution extreme learning machine and rough set feature selection. Journal of medical imaging (Bellingham, Wash.), 4(2), 024507-024507. doi: 10.1117/1.JMI.4.2.024507

Prochazka, A., Gulati, S., Holinka, S., & Smutek, D. (2019). Classification of Thyroid Nodules in Ultrasound Images Using Direction-Independent Features Extracted by Two-Threshold Binary Decomposition. Technology in cancer research & treatment, 18, 1533033819830748-1533033819830748. doi:10.1177/1533033819830748

Pupale. (2018). Support Vector Machines (SVM) — An Overview. Retrieved from https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

Re, A., Cattaneo, C., & Rossi, G. (2019). Hiv and Lymphoma: from Epidemiology to Clinical Management. Mediterranean journal of hematology and infectious diseases, 11(1), e2019004-e2019004. doi:10.4084/MJHID.2019.004

Reuter, H., Burgess, L., & Doubell, A. F. (2005). Epidemiology of pericardial effusions at a large academic hospital in South Africa. Epidemiology and infection, 133, 393-399. doi:10.1017/S0950268804003577

Saboo, S. S., Krajewski, K. M., O'Regan, K. N., Giardino, A., Brown, J. R., Ramaiya, N., & Jagannathan, J. P. (2012). Spleen in haematological malignancies: spectrum of imaging findings. The British journal of radiology, 85(1009), 81-92. doi:10.1259/bjr/31542964

Saini, K., Dewal, M. L., & Rohit, M. (2010). Ultrasound Imaging and Image Segmentation in the area of

Ultrasound: A Review. International Journal International Journalof Advanced Science and Technology Advanced Science and Technology Advanced Science and Technology.

Sathiya, M., & Muthuchelian, K. (2009). Significance of Immunologic Markers in the Diagnosis of Lymphoma. Acad J Cancer Res, 2.

Schafer, J. M., Welwarth, J., Novack, V., Balk, D., Beals, T., Naraghi, L., Hoffmann, B. (2019). Detection of splenic microabscesses with ultrasound as a marker for extrapulmonary tuberculosis in patients with HIV: A systematic review. In 2019 %9 Clinical ultrasound; Ultrasound; HIV; Tuberculosis; Extrapulmonary tuberculosis; Systematic review %! Detection of splenic microabscesses with ultrasound as a marker for extrapulmonary tuberculosis in patients with HIV: A systematic review (Vol. 109).

Schmitz, e. a. (2012). Automated Image Analysis of Hodgkin lymphoma. Retrieved from https://arxiv.org/abs/1209.3189

Sharma, N., Ray, A., Sharma, S., Shukla, K., Pradhan, S., & Aggarwal, L. (2008). Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network. Journal of medical physics / Association of Medical Physicists of India, 33, 119-126. doi:10.4103/0971-6203.42763

Sinkala, E., Gray, S., Zulu, I., Mudenda, V., Zimba, L., Vermund, S. H., Kelly, P. (2009). Clinical and ultrasonographic features of abdominal tuberculosis in HIV positive adults in Zambia. BMC Infectious Diseases, 9(1), 44. doi:10.1186/1471-2334-9-44

Sjogren, A., Leo, M., Feldman, J., & Gwin, J. (2016). Image Segmentation and Machine Learning for Detection of Abdominal Free Fluid in Focused Assessment with Sonography for Trauma Examinations: A Pilot Study. Journal of Ultrasound in Medicine, 35. doi:10.7863/ultra.15.11017

Spina, M., Carbone, A., Gloghini, A., Serraino, D., Berretta, M., & Tirelli, U. (2011). Hodgkin's Disease in Patients with HIV Infection. Advances in Hematology, 2011, 402682. doi:10.1155/2011/402682 Allen, P. B., & Gordon, L. I. (2017). Frontline Therapy for Classical

Hodgkin Lymphoma by Stage and Prognostic Factors. Clinical Medicine Insights. Oncology, 11, 1179554917731072-1179554917731072. doi:10.1177/1179554917731072

Stewart, K., Navarro, S., Kambala, S., Tan, G., Poondla, R., Lederman, S., Lavy, C. (2020). Trends in Ultrasound Use in Low- and Middle-Income Countries: A Systematic Review. International Journal of MCH and AIDS, 9, 103-120. doi:10.21106/ijma.294

Stolojescu-Crisan, C., Răilean, I., Moga, S., & Isar, A. (2010). Comparison of wavelet families with application to WiMAX traffic forecasting.

Strohm, H., Rothlübbers, S., Eickel, K., & Günther, M. (2020). Deep learning-based reconstruction of ultrasound images from raw channel data. International Journal of Computer Assisted Radiology and Surgery, 15(9), 1487-1490. doi:10.1007/s11548-020-02197-w

Subramanya, M. B., Kumar, V., Mukherjee, S., & Saini, M. (2015). SVM-Based CAC System for B-Mode Kidney Ultrasound Images. Journal of digital imaging, 28(4), 448-458. doi:10.1007/s10278-014-9754-4

Swerdlow, S. H., Campo, E., Pileri, S. A., Harris, N. L., Stein, H., Siebert, R., Jaffe, E. S. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood, 127(20), 2375-2390. doi:10.1182/blood-2016-01-643569

Syed, F. F., & Mayosi, B. M. (2007). A Modern Approach to Tuberculous Pericarditis. Progress in Cardiovascular Diseases, 50(3), 218-236. doi: https://doi.org/10.1016/j.pcad.2007.03.002

Szeliski. (2010). Computer Vision: Algorithms and Applications.

Teomete, U., Tulum, G., Ergin, T., Cuce, F., Koksal, M., Dandin, O., & Osman, O. (2018). Automated computer-aided diagnosis of splenic lesions due to abdominal trauma. Hippokratia, 22(2), 80-85. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/31217680

Thomas, J. G., Jeanty, P., Peters, R. A., 2nd, & Parrish, E. A., Jr. (1991). Automatic measurements of fetal long bones. A feasibility study. J Ultrasound Med, 10(7), 381-385. doi:10.7863/jum.1991.10.7.381

Tou, J. Y., Tay, Y. H., & Lau, P. Y. (2007). Gabor Filters and Grey-level Co-occurrence Matrices in Texture Classification.

Tseng, H. H., Wei, L., Cui, S., Luo, Y., Ten Haken, R. K., & El Naqa, I. (2018). Machine Learning and Imaging Informatics in Oncology. Oncology. doi:10.1159/000493575

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. PLOS ONE, 14(11), e0224365-e0224365. doi: 10.1371/journal.pone.0224365

van den Heuvel, T. L. A., de Bruijn, D., de Korte, C. L., & Ginneken, B. v. (2018). Automated measurement of fetal head circumference using 2D ultrasound images. PLOS ONE, 13(8), e0200412. doi: 10.1371/journal.pone.0200412.

van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging— "how-to" guide and critical reflection. Insights into Imaging, 11(1), 91. doi:10.1186/s13244-020-00887-2

van Zelst, J. C. M., Tan, T., Clauser, P., Domingo, A., Dorrius, M. D., Drieling, D., . . . Mann, R. M. (2018). Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. European radiology, 28(7), 2996-3006. doi:10.1007/s00330-017-5280-3

Vancauwenberghe, T., Snoeckx, A., Vanbeckevoort, D., Dymarkowski, S., & Vanhoenacker, F. M. (2015). Imaging of the spleen: what the clinician needs to know. Singapore medical journal, 56(3), 133-144. doi:10.11622/smedj.2015040

Verburgh, E., & Antel, K. (2019). Approach to lymphoma diagnosis and management in South Africa (Vol. 109).

Vial, A., Stirling, D., Field, M., Ros, M., Ritz, C., Carolan, M., Miller, A. A. (2018). The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. Translational Cancer Research, 7(3), 803-816. Retrieved from http://tcr.amegroups.com/article/view/21823

Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., & Reede, D. (2016). Interpretive Error in Radiology. American Journal of Roentgenology, 208(4), 739-749. doi:10.2214/AJR.16.16963

Walczyk, J., & Walas, M. K. (2013). Errors made in the ultrasound diagnostics of the spleen. Journal of ultrasonography, 13(52), 65-72. doi:10.15557/JoU.2013.0005

Wang, C.-C. J., Silverberg, M. J., & Abrams, D. I. (2014). Non-AIDS-Defining Malignancies in the HIV-Infected Population. Current infectious disease reports, 16(6), 406-406. doi:10.1007/s11908-014-0406-0

Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging, 23(7), 903-921. doi:10.1109/TMI.2004.828354

WHO. (2014). HODGKIN LYMPHOMA (ADULT). Review of Cancer Medicines on the WHO List of Essential Medicines. Retrieved from: https://www.who.int/selection_medicines/committees/expert/20/applications/HodgkinLymphoma_Adult.pdf?ua=1

Withey, D. J., & Koles, Z. J. (2007, 12-14 Oct. 2007). Medical Image Segmentation: Methods and Software. Paper presented at the 2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging.

Xu, S. S.-D., Chang, C.-C., Su, C.-T., & Phu, P. (2019). Classification of Liver Diseases Based on Ultrasound Image Texture Features. Applied Sciences, 9, 342. doi:10.3390/app9020342

Yadav, A., r.roy, kumar, a., kumar, c., & Dhakad, S. (2015). De-noising of Ultrasound Image using Discrete Wavelet Transform by Symlet Wavelet and Filters.

Yang, Y., Su, Z., & Sun, L. (2010). Medical image enhancement algorithm based on wavelet transform. Electronics Letters, 46, 120-121. doi:10.1049/el.2010.2063

Yang, Z. G., Min, P. Q., Sone, S., He, Z. Y., Liao, Z. Y., Zhou, X. P., . . . Silverman, P. M. (1999). Tuberculosis versus lymphomas in the abdominal lymph nodes: evaluation with contrast-enhanced CT. American Journal of Roentgenology, 172(3), 619-623. doi:10.2214/ajr.172.3.10063847

Yeghiazaryan, V., & Voiculescu, I. (2018). Family of boundary overlap metrics for the evaluation of medical image segmentation. Journal of medical imaging (Bellingham, Wash.), 5(1), 015006-015006. doi: 10.1117/1.JMI.5.1.015006

You, D., Antani, S., Demner-Fushman, D., & Thoma, G. (2014). A Contour-based Shape Descriptor for Biomedical Image Classification and Retrieval (Vol. 9021).

Yu, R.-S., Zhang, W.-M., & Liu, Y.-Q. (2006). CT diagnosis of 52 patients with lymphoma in abdominal lymph nodes. World journal of gastroenterology, 12(48), 7869-7873. doi:10.3748/wjg.v12.i48.7869

Yushkevich, P., & Gerig, G. (2017). ITK-SNAP: An Intractive Medical Image Segmentation Tool to Meet the Need for Expert-Guided Segmentation of Complex Medical Images. IEEE Pulse, 8, 54-57. doi:10.1109/MPUL.2017.2701493

Zhang, G., Yang, Z.-g., Yao, J., Deng, W., Zhang, S., Xu, H.-y., & Long, Q.-h. (2015). Differentiation between tuberculosis and leukemia in abdominal and pelvic lymph nodes: evaluation with contrast-enhanced multidetector computed tomography. Clinics, 70, 162-168. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-59322015000300162&nrm=iso

Zhang, J., & Cheng, Y. (2020). Despeckling Method for Medical Images Based on Wavelet and Trilateral Filter. In J. Zhang & Y. Cheng (Eds.), Despeckling Methods for Medical Ultrasound Images (pp. 103-122). Singapore: Springer Singapore.

Zhang, L., & Rusinkiewicz, S. (2018). Learning to Detect Features in Texture Images.

Zhang, Z., & Sejdić, E. (2019). Radiological images and machine learning: Trends, perspectives, and prospects. Computers in Biology and Medicine, 108, 354-370. doi: https://doi.org/10.1016/j.compbiomed.2019.02.017

Zheng, Q., Warner, S., Tasian, G., & Fan, Y. (2018). A Dynamic Graph Cuts Method with Integrated Multiple Feature Maps for Segmenting Kidneys in 2D Ultrasound Images. Academic radiology, 25(9), 1136-1145. doi: 10.1016/j.acra.2018.01.004

Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., & Zhang, J. (2010). Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. Journal of digital imaging, 23(1), 51-65. doi:10.1007/s10278-009-9185-9

Zulkarnain, N. Z., & Meziane, F. (2019). Ultrasound reports standardisation using rhetorical structure theory and domain ontology. Journal of Biomedical Informatics: X, 1, 100003. doi: https://doi.org/10.1016/j.yjbinx.2019.100003