Transducer selection & protocol compliance detection algorithm design as a basis for automated point-of-care lung ultrasound screening of tuberculosis.

UNIVERSITY OF TWENTE.

Faculty of Science and Technology
Master Biomedical engineering
Imaging & Invitro diagnostics

Master thesis

J.T. Room

Chair:
Physics of Fluids (POF)

Supervisors:
Prof. Dr. Ir.   C.L. de Korte
Ir.   E.J.S. Swinkels

Board of examiners:
Prof. Dr. Ir.   C.L. de Korte
Prof. Dr.   M. Versluis
Dr.   J. Veltman
Ir.   E.J.S. Swinkels

June 28, 2023

DELFT
IMAGING

Acknowledgements

After a rocky start to my Bachelor's almost seven years ago, here we are. It is a wrap. I thoroughly enjoyed my time in Enschede with friends, study, and work, from which I learned a lot in a relatively short period of time. Time flies when you are having fun. Now I am finishing my Master's thesis and preparing for my first job. But first things first.

I would like to thank Delft Imaging for the opportunity of an external graduation. I learned a lot during my Master's thesis from both academic and business perspectives. I enjoyed working in a friendly and productive environment and I had a great experience working with and talking to you all. In particular, I would like to thank Elze and Frank for their supervision during my thesis. Elze guided me throughout the entire process and was my daily supervisor. Although I was the first student she ever guided during a Master's thesis, she did a great job and helped me a lot by asking the right questions.

I would also like to thank Chris for the flexibility during my graduation and for making it possible to graduate from the University of Twente in a company. Without him, it would not have been possible. I also thank Jeroen for his guidance during my previous internship and for being a part of my graduation committee for the Master's thesis. Also, a big thank you to Michel for being the third member of my committee and for being a good advocate during the beginning of my thesis.

During the quality assurance experiments I was guided by Gert, I would like to thank him for his instructions and especially for the processing of the results. I would like to thank Bram for the time he gave me during his promotion. He provided a very helpful introduction to lung ultrasound when I started and did me a great favor by labeling the test data set.

Last but not least, I would like to thank my family and friends for supporting me throughout my studies and my graduation. In particular, I want to thank my parents and girlfriend for keeping up with me and being there for me even during my worst moments. Thank you.

Abstract

Tuberculosis(TB) is one of the most prevalent diseases in the world. Most TB infections occur in third world countries, where the diagnosis of TB is difficult due to limited access to health-care, funds, and infrastructure. To make early diagnosis possible in these areas, a cost-effective point-of-care device screening tool in combination with a follow-up test would be ideal. A point-of-care ultrasound (POCUS) device for automated TB screening could increase the number of TB cases detected in areas of interest. These cases can then be sent to the hospital to confirm whether the symptoms are TB. For the device, a standardized easy-to-learn scan protocol with an affordable point-of-care ultrasound transducer can be combined with computer-aided detection to eventually exclude the health worker from the screening process.

This thesis takes the first steps toward the design of a cost-effective standardized computer-assisted lung ultrasound (LUS) screening tool for TB. By evaluating different commercial affordable POCUS transducers on their usability and their ability to detect TB-related LUS signs via an assessment matrix, B-line phantom, and quality assurance test. The highest scoring transducer was used for a clinical study on TB detection with LUS using a designed standard scan procedure.

Usable scan lines for diagnosis in LUS are operator dependent. The quality of the recording is influenced by the scan angle toward the pleural line and the proper contact between the skin and the transducer. Therefore, in the second part of this thesis, clinical study data was used to design a shadow peak-based protocol compliance algorithm that is capable of checking whether a scan line is performed correctly. This algorithm was designed to guide an inexperienced user during the scan procedure to ensure proper data collection. The data collected using this protocol compliance protocol can be used to train a deep learning algorithm for computer-aided screening of TB.

The ideal transducer out of the tested transducers for a cost-effective POCUS screening tool for the detection of TB with LUS is the Clarius C3 HD3 multipurpose transducer. The protocol compliance algorithm designed is capable of accurately selecting sufficiently recorded videos of scan lines with a sensitivity of 100% out of a data set containing 30 videos of scan lines. The algorithm made a correct judgment for 87% of the input scan lines with the greatest error in the selection of insufficient scans resulting in a specificity of 60%. The algorithm shows promise, and the physical principle seems to be usable for the selection of sufficient scan lines. However, the algorithm should be re-evaluated using a larger, more representative data set for training and testing to achieve a better judgment of its performance.

The selected transducer and the protocol compliance algorithm design are the first steps toward a cost-effective standardized computer-assisted LUS screening tool for TB.

Contents

List of Figures

List of Tables

Introduction

The World Health Organization estimates that a quarter of the world population in 2022 is infected with tuberculosis(TB) [10]. It is one of the most prevalent diseases in our world. The diagnosis of tuberculosis is currently based on a skin test, molecular test, rapid sputum test, chest x-ray, or a combination of these diagnostic methods [10]. These methods require an advanced clinical setting in which the tests can be performed. However, most TB infections occur in third world countries, where the diagnosis of TB is difficult due to limited access to healthcare, funds, and infrastructure.

To solve this problem and make early diagnosis possible in these countries, a low-end point-of-care device would be ideal. Studies have shown that TB detection with high-end point-of-care ultrasound (POCUS) shows great promise with acceptable sensitivity and specificity [9]. However, there is no standardized POCUS method for automated TB screening and the lack of trained radiographers in the areas of need adds to the complexity of realizing this technique. To overcome these problems, a standardized easy-to-learn scan protocol with an affordable low-end POCUS transducer could be combined with computer-aided detection to eventually exclude the trained radiographer from the diagnostic process.

The first step in the design of this technique is the selection of an ultrasound (US) transducer for the collection of data. Therefore, the first part of this study evaluates different commercial affordable POCUS transducers on their usability and their ability to detect TB-related lung symptoms via an assessment matrix and a phantom experiment. The selected best performing transducer is used for a clinical study in Zambia to collect lung ultrasound (LUS) data. This data will be used in the future to train a deep learning algorithm for a computer-assisted screening tool for automated detection of abnormalities in the lung. To further standardize the POCUS screening method, an easy and standardized LUS scan procedure is necessary. To obtain the best chance of a quick and reliable outcome from an LUS screening procedure, a large area of the lungs must be scanned, but the scan time should be limited. The equilibrium between the two is a 16-step scan line protocol that records multiple important areas of the lung [4]. These areas are scanned with the transducer in the longitudinal position and located on the anterior, lateral, and posterior sides of the thorax

The quality of the recordings of the scan lines from the described protocol for the diagnosis in LUS is operator dependent. Because the quality of the recording is influenced by the scan angle towards the pleural line and the contact between the skin and the transducer [11]. Therefore, the second part of this thesis designs a protocol compliance algorithm that is able to check whether a scan line is performed correctly. The algorithm uses adjusted shadow peak bone segmentation techniques to segment key features in the frames and check their required presence and intensity with respect to the surrounding structures. The algorithm is designed to guide an inexperienced user during the scan procedure to ensure proper data collection.

The selected transducer and the protocol compliance algorithm design are the first steps toward a cost-effective standardized computer-assisted LUS screening tool for the screening of TB.

Theoretical background

2.1 Tuberculosis

This section provides general information on TB and current treatment and diagnosis methods.

2.1.1 Pulmonary Tuberculosis

TB is a bacterial infection that is transferred primarily by air through coughing [10]. The causative agent for TB is Mycobacterium tuberculosis (MTB), and inhalation of this pathogen can have multiple results [12]. The MTB can settle and multiply in alveolar macrophages of the lungs where it survives the adaptive immune system of the host by inhibiting the start of the process that breaks down the MTB. These alveolar macrophages can spread MTB to surrounding lymph nodes. Furthermore, the literature suggests a possible primary infection of the bloodstream from which MTB can migrate to other tissues, therefore becoming extrapulmonary TB without a primary infection of the lungs [13].

However, in most cases, the primary infection starts in the lungs where after the multiplication of MTB in macrophages, a granuloma is created around the MTB infection. The purpose of the granuloma is to create a wall between the bacilli and its surrounding tissue. After the bacilli are contained, the inside of the granuloma dies. In most cases, the MTB is killed in this process. However, in some cases, the bacilli stay alive and are contained within the macrophages, then the infection becomes a latent infection [14]. The infection can remain latent for multiple years without causing serious symptoms.

When the immune system is debilitated, the infection can become active. When the infection becomes active, this can result in rapid multiplication and spread of bacilli in the lungs and cause inflammation. The immune system will start to contain the bacilli by forming granuloma around them and introducing caseous necrosis within the granuloma causing consolidation. This can result in cavities that make it easier for MTB to spread to other parts of the body via the vascular system [14], becoming extra-pulmonary TB.

Active TB causes symptoms in the patient such as weight loss, fever, night sweats, coughing up blood, and chest pains. Active and latent lung TB can cause abnormalities in the lung, multiple of these abnormalities can be seen using POCUS[7]. These abnormalities are described in section 2.3.3.

2.1.2 Extra-pulmonary TB

When the MTB enters the bloodstream, the bacilli can end up in organs, tissues, and lymph nodes throughout the body. The most common places where bacilli end up are lymph nodes, lymphatic TB represents approximately 15% off all extra-pulmonary TB [15]. In addition to lymphatic TB, extra-pulmonary TB can be present in the organs of the abdomen, the central nervous system, the brain, the spine, muscle tissue, and even genital parts. Extra-pulmonary TB infections can have life-threatening and/or functional consequences[16].

2.1.3 Diagnosis

During the active stage of TB symptoms occur, making the diagnosis of TB easier. However, in most cases, TB is still in its latent stage. The latent stage of TB is harder to detect due to the lack of detectable symptoms.

In its latent stage, the best TB detection methods appear to be a skin test and / or a blood test, but the latter of the two is still in an early prototype phase. A positive result of either one of these tests in combination with symptoms can also be used as a diagnosis of active or latent TB[17].

Currently, the most used diagnostic test for TB is a rapid PCR sputum test that looks for the presence of MTB. However, this test lacks specificity, especially for latent TB.

A promising screening method for TB is the chest X-ray in combination with automated detection using deep learning algorithms. This method is able to make a distinction between positive and negative cases of TB, but lacks the ability to differentiate between latent and active TB [17]. The chest x-ray looks for consolidations, cavities, miliary nodules, and effusions, which are definitive of active TB [18].

A combination of the sputum test and chest x-ray can be used as a screening and diagnosis tool for active TB in underdeveloped countries with a lack of infrastructure and access to healthcare. The tests can be combined in a container setting which can be put on the back of a truck. Therefore, it is mobile and can increase the range of the standard hospital setting of TB diagnosis. However, these setups are large, take time to transfer from area to area and are not usable at the bedside of a patient.

2.1.4 Treatment

The treatment for TB depends on the stage of the disease. A latent infection is not contagious and can be treated with a single drug over a long period of time. In the case of active TB, a combination of drugs must be used due to the contagiousness of the active form of TB. This combination of drugs ensures that the patient is no longer infectious to others within a couple of weeks and heals the patient over a long period of time. During the infectious part of the treatment, the patient should be isolated from others to prevent further spread of the disease[10]. Treatment becomes more difficult if a patient suffers from drug-resistant or multidrug-resistant TB. In that case, more severe medications must be used with an increased chance of side effects[10]. Furthermore, drug-resistant TB patients are contagious to others for a longer period of time and are only considered non-contagious after the patient clearly responds to the medication.

To prevent treatment, preventive treatment should be administered soon after primary infection. In addition, a vaccine is available, reducing the risk of infection. The vaccine should be administered at an early age for the best results[10].

2.2 Ultrasound

This section gives an introduction to the working principles behind ultrasound and the different available basic transducers.

2.2.1 Ultrasound basics

US is a non-invasive diagnostic imaging technique that uses high-frequency sound waves to image the interior of the body. The high frequencies that are used in US are generally in the range of $1-10$ MHz and cannot be heard with the human ear. The sound waves travel through the tissue in a wavelike fashion by compressing and expanding the medium. The speed of the wave depends on the tissue through which it travels. Waves that propagate through different tissues can be reflected, absorbed, refracted, or scattered. Each tissue or tissue boundary has its unique properties and therefore reflects, absorbs, refracts, or scatters the wave in a different way. These tissue properties are the basis of US imaging, with scattering being the most prominent[1].

The US waves in diagnostic US are produced by a transducer. The transducer contains multiple piezoelectric elements that expand and shrink by applying an electric field. Therefore, they can emit sound waves into the body. In addition to emitting, these elements can also receive sound waves and transfer the incoming wave into an electric signal that can be processed into an image [1].

At tissue boundaries or at an object, a reflection of an emitted sound wave can occur if the structure is greater than the wavelength of the incident sound wave. The reflection of the sound wave is called an echo and is sent back toward the transducer as shown in figure 2.1. To achieve as much reflection as possible, an angle of 90 degrees is desirable between the transducer and the tissue boundary. If a structure is smaller than the incidence sound wave, scattering will be the most abundant for that structure. Therefore, less reflection takes place and the resulting image will be noisy without much structure.
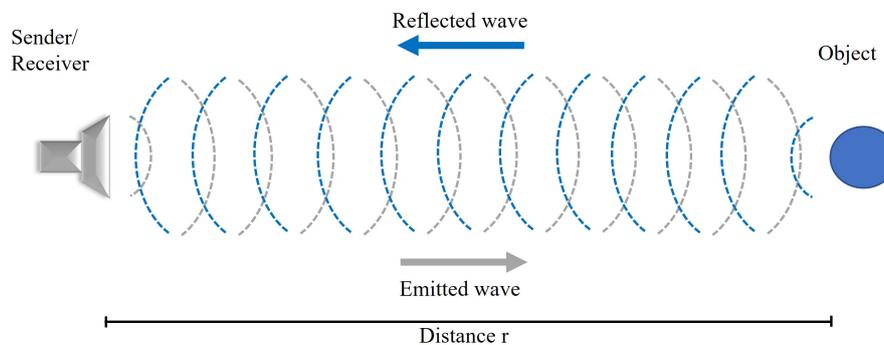
Figure 2.1: Schematic representation of US wave propagation to object, producing a reflected wave propagating back

The time it takes for the echo to return to the transducer indicates the depth (distance r) of the tissue or tissue boundary. Greater distance r means a longer time before the echo returns to the transducer. This relationship between time and distance is used to depict certain structures at certain depths in the US image.

For a single line recording, a high-frequency sound wave is sent into the body by a small number of piezoelectric elements next to one another. The returning high-frequency signals are received by the same group of elements and converted into electric signals. The electric signals of the used piezoelectric elements for this single recording are combined. The envelope or so-called A-mode signal of the combined signals is taken (shown in figure 2.2) and processed into a greyscale line depicting the echoes at a depth corresponding with their location in the body. The brightness of the depicted greyscale line is dependent on the strength/amplitude of the echo at that depth.

By using multiple piezoelectric elements beside one another a transducer array is created that produces several of these A-mode line recordings at the same time. Combining the A-mode lines creates a B-mode image showing anatomical structures in a greyscale image[1].
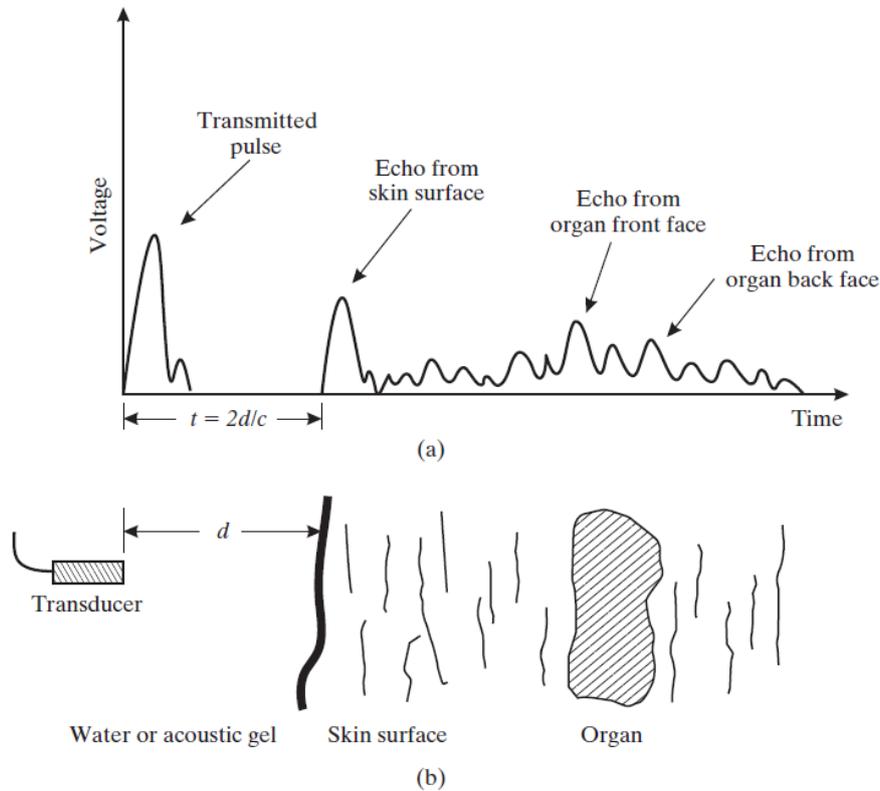
Figure 2.2: (a) A typical A-mode display and (b) its relationship to the anatomy. with t= time, d=distance and c=speed of sound [1]

2.2.2 Ultrasound transducers

There are several types of US transducers, each with its own applications in practice. For this study, the following transducers are considered: linear array, curvilinear array, and micro-convex array. The main difference in hardware between the transducer types is the number of piezoelectric elements that are present in its array and the shape of the array. These influence the area and shape of the produced image. Each type of transducer has its advantages and disadvantages[1].

The linear array has a flat surface that produces an image with a rectangular shape. Most linear transducers have a high frequency range (5-12 MHz) that provides the transducer with a high resolution but a shallow imaging depth of approximately 2-7cm. The Linear transducer is mostly used for imaging of superficial structures[1].

The curvilinear array is commonly used for abdominal scans. This is due to their low-frequency range $(1 - 5MHz)$ which makes tissue penetration around $10cm$ or deeper possible. However, the lower frequency also decreases the resolution of the acquired image compared to the high-frequency linear transducer. Due to the curved shape of the array, a large field of view is accomplished[1].

The micro-convex array has a frequency range of $2-10MHz$, is more curved than the curvilinear array, and has a smaller footprint. Therefore, the transducer has a high resolution in the near field but a low resolution in the far field. furthermore, it has a large field of view due to the large curvature of the transducer [1].

2.3 Lung ultrasound

This section gives an introduction to LUS, the LUS scan protocol used, and the abnormalities present in LUS.

2.3.1 Lung Ultrasound basics

A typical LUS image is depicted in figure 2.3a where the pleural line and the shadows of the ribs can be seen, the combination of these signs is called the bat sign. The bat sign is typical for LUS and gets its name from the shape represented by the lines in figure 2.3b representing the wings of a bat. The shadow under the bone is present because the bone blocks the sound waves from propagating further into the body. The quality of LUS recordings is highly operator dependent, therefore the operator requires training or guidance in advance or during the scan procedure[11].

(a) Example of a typical LUS   (b) Bat sign in typical LUS
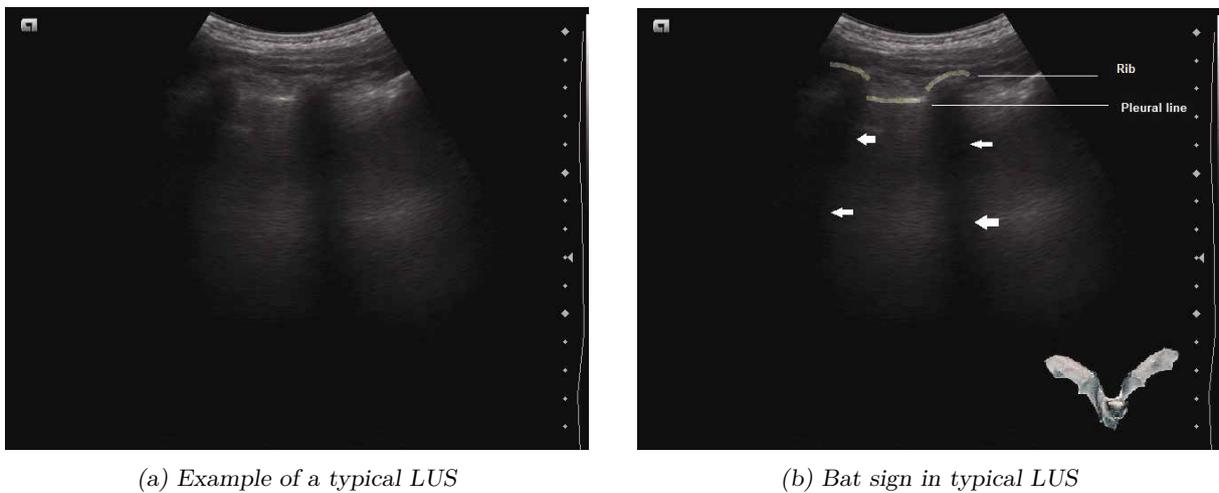
Figure 2.3: Example of (a)normal LUS and the same image with (b) highlighted pleural line and rib shadows representing the bat sign [2]

Historically LUS was not used much due to a lack of anatomical structures posterior to the pleural line and the presence of artifacts [19]. A pioneer in the field of LUS is D. Lichtenstein, he showed that the artifacts could be used to differentiate between healthy and unhealthy lungs. He continued showing that different artifacts are present for different lung diseases [7]. Furthermore, the global COVID-19 lung infection pandemic caused a lot of research in the field of point-of-care LUS [20]. This resulted in a better understanding of the usability of LUS for lung infection detection and its usability for other lung infections or abnormalities.

The lack of anatomical structures in the US image is caused by the air inside the lungs. The tissue-air boundary reflects almost all of the initial sound waves back to the transducer, therefore, blocking most of the sound waves from propagating deeper into the lungs. This means that structures posterior to the pleural line will be visualized with low intensity in the image if they are depicted in the image at all[2].

LUS is an imaging modality that is mainly based on artifacts caused by abnormalities along the pleural line due to the low amount of input waves propagating beyond this line. In addition to artifacts, some structures/tissues just posterior to the pleural line are able to induce hyper-echoic or hypo-echoic spots visible in the US image [3].

Curvilinear transducers are capable of acquiring a good view of the lung due to the deep penetration depth and wide field of view. For more superficial structures or to image children, a linear transducer can be useful due to its high resolution close to the transducer.

The angle between the transducer and the pleural line should be as close as possible to 90° to

obtain a usable scan of the lungs. This induces the most reflection back to the transducer and will therefore show the artifacts and anatomical structures better. If the angle is not around 90°, important artifacts or abnormal structures are not properly imaged[3]. A good indicator to check whether the angle with the pleural line is around 90° is the intensity of the pleural line. When the angle is good, the intensity of the pleural line is much higher than its surrounding structures.

2.3.2 Scan protocol

The lungs are present in a large part of the human thorax. When the lungs are imaged with CT or thorax X-ray, an overview of almost the entire organ is achieved. However, US only visualizes a slice of the part of the lungs where the transducer is placed on the body. To acquire a complete view of the lungs with LUS, multiple scan lines or scan points across the thorax are needed. A scan of the lungs can be performed with the transducer along the direction of the ribs (transverse) or perpendicular (longitudinal) to the ribs, as shown on the top left of figure 2.4.
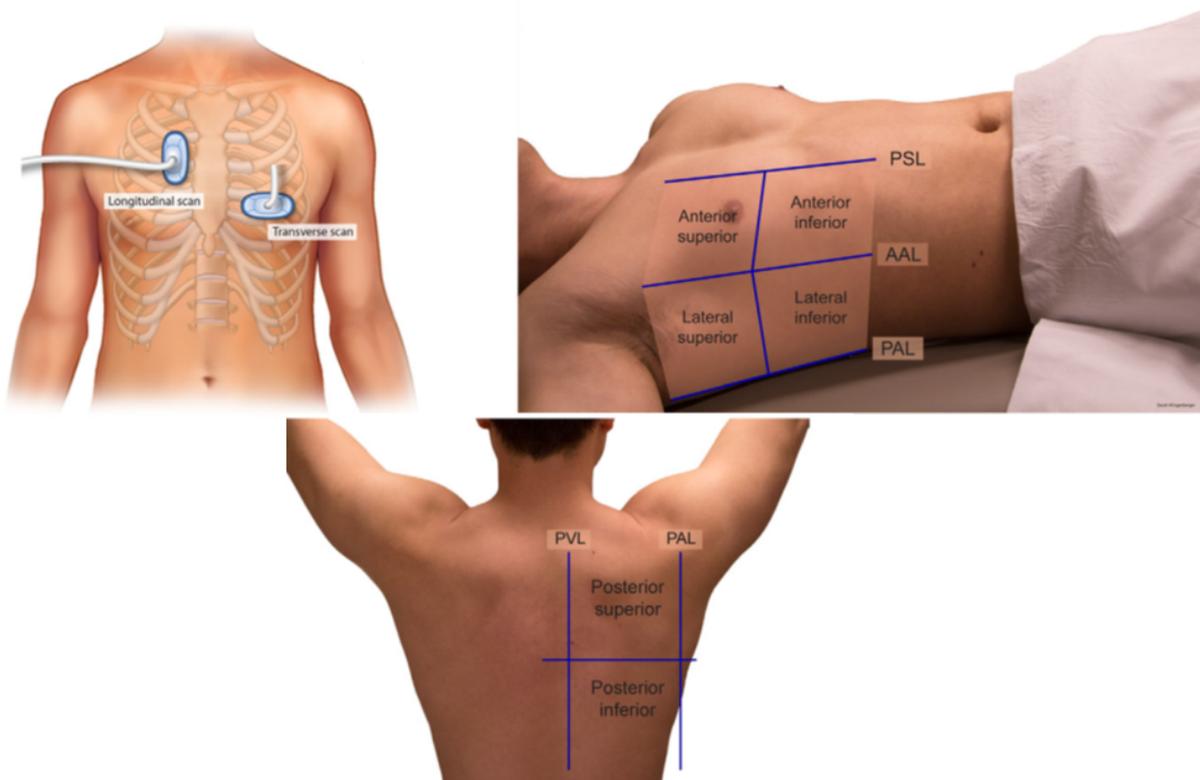
Figure 2.4: Overview of transducer position during US scan of the thorax and the important areas used in LUS [3]

To cover all important parts of the thorax without scanning the entire upper body to save time, several fast scan protocols exist. These protocols divide the thorax into multiple important areas that together give a usable overview of the lung. An example of such a protocol is shown in figure 2.4. Part of each area must be scanned on both sides of the body to make a useful diagnosis from the LUS procedure [3].

For the detection of pulmonary and extra-pulmonary TB, different scan protocols exist in the literature. The two protocols that are widely used are a thoracic and abdominal-focused US protocol based on Focused Assessment with Sonography for HIV-Associated Tuberculosis (FASH) [21] and a 12-step thorax sweep protocol designed to reduce thorax scan time and include the

most important parts of the thorax for TB infection[4].

The focus of this thesis is pulmonary TB. Therefore, the latter of the two protocols is the basis for the used standardized scan protocol designed by M. Fentress [4]. The protocol targets all regions of interest shown in figure 2.4 and consists of 16 scan lines. For each scan line, a 10-second video is recorded while the transducer is slid down from the superior point to the inferior point of the scan line. A schematic overview of the 16-line scan protocol designed by M. Fentress is shown in figure 2.5.
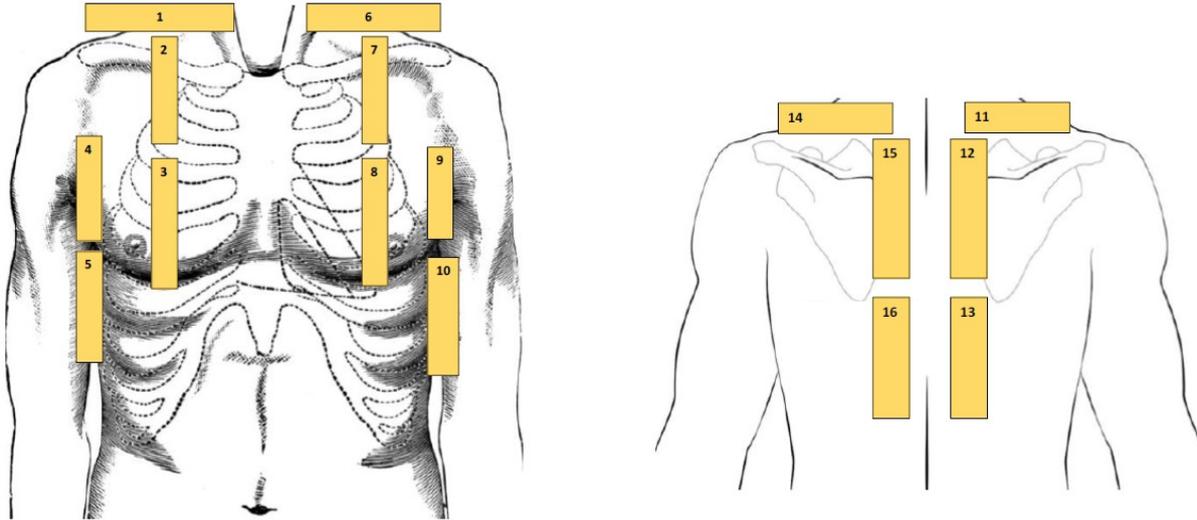
Figure 2.5: The 16 line scan line protocol designed by M. Fentress. [4]

Anterior mid-axillary scan lines 1 and 6 should be performed with the transducer in transverse orientation only, all other anterior and lateral lines should be scanned with the transducer in longitudinal and transverse orientation, but sweeps in transverse orientation are only saved if additional pathology is detected. For the posterior scan lines, the apical sweeps 11 and 14 should again be performed in a transverse orientation and all other posterior lines in the longitudinal orientation. For all scan lines performed, it is important that the angle to the pleura is close to 90 degrees. This will produce a usable scan with high intensity at the pleural line, clear A-lines in healthy patients, or possible artifacts/abnormalities in unhealthy patients.

The 16 lines consist of multiple tissues, at different depths and with slightly different acoustic properties for each patient. Therefore, these regions can be scanned with multiple US transducers with different specifications depending on the desired depth and resolution. Within these regions, several US abnormalities or a combination of abnormalities can indicate a (TB) infection [9]. The most common abnormalities that are taken into account for this study are shown in table 2.1

THORACIC FINDINGS	Ultrasound Finding	Definition
	Small subpleural consolidation (SPC)	Hypoechoic subpleural region less than 10 mm x 10 mm, with distinct borders and trailing comet-tail artefacts
	Consolidation	Subpleural, echo-poor, or tissue-like region >10 mm in depth or length, with or without sonographic air bronchograms
	Cavitation	Consolidation >10 mm in depth or length with hypoechoic central clearing. If color flow is used, no color flow present within the hypoechoic central clearing.
	Pleural effusion	Anechoic collection between the pleural line or diaphragm and the chest wall.
	B-lines	B-1 pattern: 3 or more B-lines in 1 intercostal space.
		B-2 pattern: Diffuse or confluent B-lines in 1 region.
	Irregular pleural line	Pleural line abnormal contour or thickened when imaging performed with probe at 90-degree angle to pleura.

Table 2.1: The abnormalities seen in the US for TB with their definition copied from the following source [9]

2.3.3 TB-related abnormalities and their visual representation

This section presents examples of the representation of pulmonary TB-related abnormalities in an intercostal space on LUS imaging. It is important to note that the listed abnormalities are not only specific to TB, but can also be present for other lung infections such as COVID-19 [3].

A-lines

A-lines are a reverberation artifact caused by the hyper-echoic pleural line and are present in LUS of a healthy lung as shown in figure 2.6b. The A-lines are repetitions of the pleural line at equal intervals posterior to the pleural line. The trapping of sound waves between the pleural line and the transducer surface causes these repetitions. A high percentage of the sound waves emitted by the transducer are reflected at the pleural line as a result of the presence of air below the chest wall. The transducer itself is also a strong reflector, therefore part of the sound waves are reflected back and forward before being received by the transducer elements [5]. A schematic representation of this principle is depicted in figure 2.6a.

(a) Schematic drawing of A-line generation

(b) example of A-lines in LUS
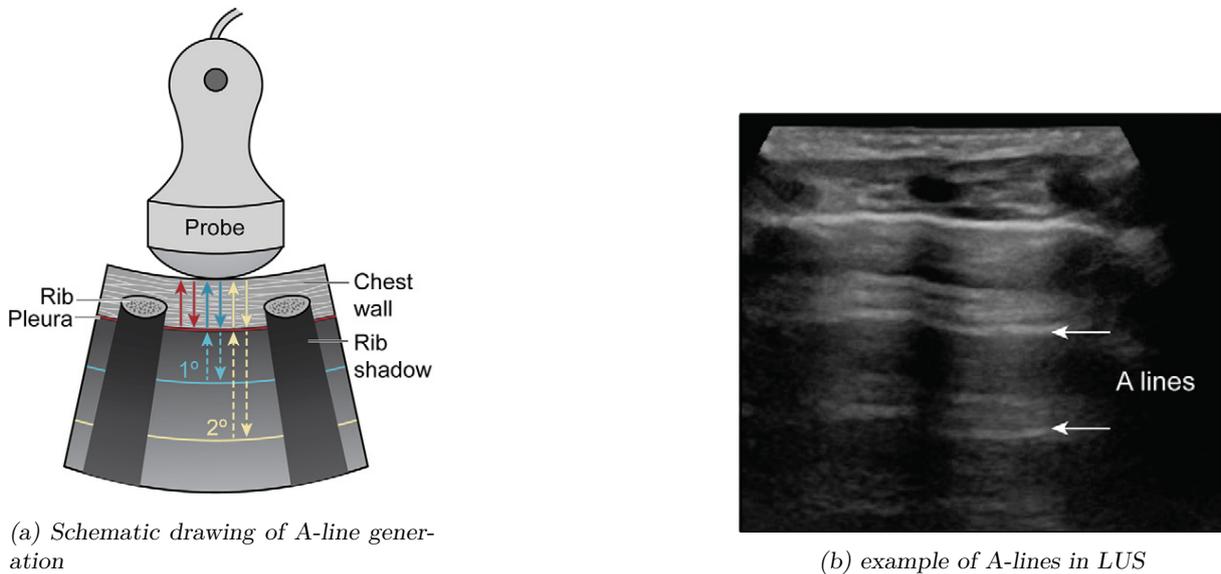
Figure 2.6: A-line artifacts caused by the repeating reflection of sound waves between the pleural line and the transducer [3]

B-Lines

Figure 2.7 shows an example of B-lines. Multiple B-lines (rule of thumb is more than 3 [4]) indicate abnormalities present around the pleural line. B-lines are caused by vertical reverberation in a small area at the pleural line. B-lines appear as hyperechoic, vertical, laser-like lines that move with lung sliding. The lines originate from the pleura and continue without fading to the bottom of the screen as shown in figure 2.7[22][23].

B-lines can signify pulmonary edema caused by TB. However, B-lines are not specific to TB and can indicate multiple different diseases. To obtain more information for the screening of TB, other abnormalities in combination with B-lines are beneficial [24]. Curvilinear transducers seem to improve the visualization of B-lines compared to linear transducers. For the best image quality, a depth between 12-18 cm would be recommended with a gain greater than 50% and the focus at the pleural line[22].

Z-lines can be confused with B-lines because both appear as vertical artifacts originating from

the pleural line. However, Z-lines are short comet tail artifacts that do not reach far down in the image. In general, the Z-lines reach a depth of 2-4cm from the pleural line [25] [26].
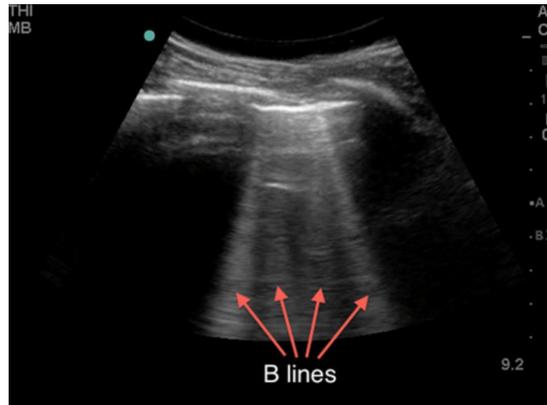
Figure 2.7: Example of multiple B-lines originating from the pleural line [5]

Consolidation

An example of a small subpleural consolidation (SPC) in the lung using US can be seen in figure 2.8a. An SPC is represented by a hypoechoic region smaller than 1x1 cm with a trailing artifact posterior to the consolidation.

An example of bigger consolidation of the lung can be seen in figure 2.8b. It can be recognized by a hypoechoic region, in this region air-bronchograms can be present, which are depicted as white pulsing dots. Consolidation is assumed to be cavitation if the depth is greater than 10mm with a hypoechoic central clearing. A cavity is defined as a gas-filled space within a nodule or a zone containing pulmonary consolidations [27]. Cavities are difficult to distinguish from the background with US imaging [4]. For visualization of small or large consolidations, a good sub-pleural resolution is desired [28].
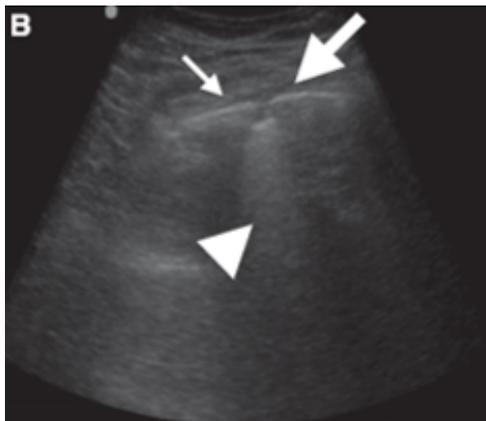
(a) Small subpleural consolidation (SPC) (thick arrow)
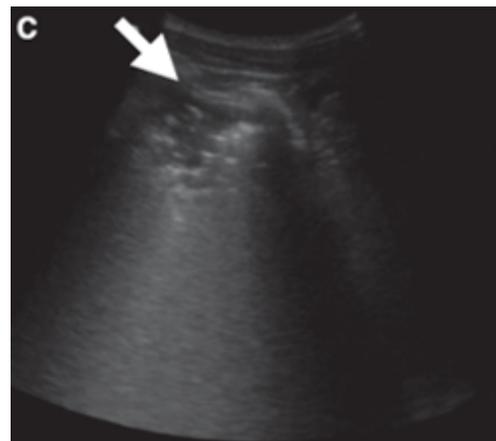
(b) Consolidation (thick arrow)

Figure 2.8: Typical appearance of SPC and consolidation [4]

Pleural effusion

A pleural effusion is the accumulation of fluid in the space between the lung and the pleural line. A pleural effusion can be recognized in figure 2.9 as the anechoic region depicted within the white shape. Pleural effusion can also be shown as black with white strands, black with white septa, or homogeneously echogenic (white)[29]. For the detection of pleural effusion, US is a perfect modality due to the fact that fluids appear distinctive on US images. To detect a pleural effusion, the scan protocol and the position of the patient are important [18]. In addition to a suitable scan procedure, the transducer must be able to image the abdomen.
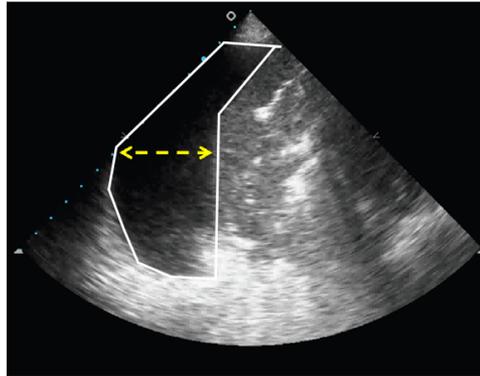
Figure 2.9: Example of pleural effusion depicted within the white shape as an anechoic structure. [6]

Pneumothorax

A pneumothorax, better known as a collapsed lung, is the presence of air in the space between the lung and the pleural line. This symptom can be recognized by the absence of lung sliding during an A-mode scan of an area in the lungs. A normal A-mode scan of the lungs is shown in figure 2.10a as the so-called beach sign. This shows that the lungs under the pleural line move in a normal fashion. An abnormal scan is shown in figure 2.10b here a barcode sign is seen that represents the absence of lung sliding and the possible presence of a pneumothorax[7].

(a)
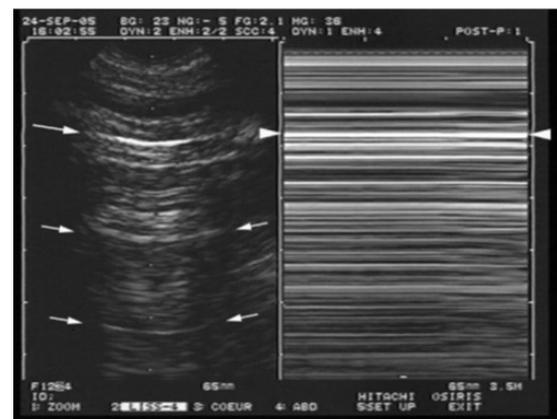
(b)

Figure 2.10: Representation of normal A-mode recording of the lungs in figure 2.10a and abnormal scan of the lungs in figure 2.10b [7]

Irregular pleural line

The high intensity at the pleural line in LUS is caused by the high mismatch in acoustic impedance between the intercostal tissues and the air inside the lung. The pleural line is represented as a hyperechoic line (white) at the border between the intercostal tissue and the air filled lung, an example is shown in 2.7 [30]. Irregularities or thickening of the pleural line can be signs of SPC or pulmonary fibrosis that can result from pulmonary TB [18][31]. For a good analysis of irregularities in the pleural line, the focus of the transducer should be around the pleural line. An irregularity can be a thickening of the pleural line or a large difference in thickness between adjacent areas of the pleural line.

2.4 Computer-aided detection for LUS

The abnormalities described in section 2.3.3 are used to make a distinction between normal and abnormal lung tissues. Traditionally, this is done by a radiologist who is trained to work with the imaging modality and knows how to record and interpret images. However, in recent years, the interpretation of medical images is also done via computer algorithms. These algorithms are able to process large amounts of data and produce results on whether a patient is healthy or non-healthy. For example, it is possible to produce a list of structures found in an image or to automatically segment tumors [32]. By using computer algorithms, a lot of time can be saved and the examiner using the imaging modality does not have to be able to interpret the images. For LUS the first steps have been taken for the automated detection of normal and abnormal structures in images. Research is being done into algorithms that can detect the pleural line, rib shadows, and multiple abnormalities. The first results have been produced over the last years and show promise [26],[30],[33],[34],[35],[36],[37]. It is important to note that most of this research was done on rectangular images resulting from high-end linear transducers.
About computer-aided detection in LUS using low-budget curvilinear POCUS transducers, not much is known. Therefore, knowledge needs to be gained in this field by first designing a standardized scan protocol and checking the quality of the gathered data. To achieve this, a protocol compliance detection algorithm for automated point-of-care LUS TB screening is designed. The data resulting from the compliance algorithm can be used to train a deep-learning algorithm for computer-aided detection of TB using LUS.

2.4.1 Deep Learning

Machine learning or deep learning networks are algorithms that are trained on labeled or unlabeled data sets. After training, these algorithms should be able to analyze similar unlabeled data and assign a label to the data. To make sure that the outcome of a trained network is reliable, it is beneficial to use a large training data set that is representative of the data that needs to be analyzed. To make sure that this is the case, a standard way of data collection for labeled training data and the eventual input data is necessary [38].

Method

3.1 Transducer selection

Seven different US transducers were compared to each other to find the most suitable transducer for the design of a LUS TB screening tool. The list of transducers available for comparison is given below. These transducers were graded on their general and technical properties using an assessment matrix. The selected US transducer will be used for a clinical study and must be able to visualize lung structures and abnormalities in the best way possible.

Because the goal was to design a POCUS device for TB screening in a low-resource setting, the price, usability, robustness, and technical specifications of the transducers were all taken into account. Therefore, the general properties of the tested transducers were also included in this project. Information on different properties was obtained using the information stated by the manufacturers, by taking a survey, and by doing experiments for some more detailed technical properties. The methods used to compare the transducers are explained in more detail in this section.

- Clarius C3 HD3 Curvilinear

- Clarius C7 HD3 micro-convex

- Telemed MicrUs Pro-C60S Curvilinear

- Telemed MicrUs Pro-L40S Linear

- Butterfly iQ+ Multi-purpose

- SonoHealth D2CL Curvilinear & Linear

- MedSingLong MSLPU79 transducer Curvilinear

3.1.1 The assessment matrix

The assessment matrix was used for the final verdict on transducer selection. All the properties on which the transducers were compared were used for the assessment matrix. For each property, the transducers get a score between one and five, one being poor performance and five being excellent performance. Each property was of certain importance for the selection of the transducer. Therefore, a weight was assigned to each property. The weights were assigned by the researcher based on the stated importance of the properties by the company after several conversations and feedback moments. The score of the property of each transducer was multiplied by the weight of the property. The resulting values were summed for all properties of each transducer. The result of the assessment matrix was a total score. The transducer with the highest score was the most ideal for the design of a cost-effective LUS TB screening tool.

3.1.2 Specification stated by manufacturers

Several specifications stated by the manufacturer were used for the assessment matrix. These specifications come from the available website and/or catalog of the transducers. Some examples of these properties are price, robustness, battery life, charging time, and adjustable parameters. These properties were graded with a scoring system. In the case of a value such as a price, a price range was selected that ranged from the lowest price (best = 5) to the highest price (worst = 1) with 5 intervals. If the price of the transducer was in the third interval it would get a score of 3. Some properties were not tied to a number, but were available or not available for a transducer/application. In that case, available gets a score of 5 and not available gets a score of 1. If property information was not available on the website or in catalogs, the specifications were taken from values in the app, or an average score of 3 was used if it could not be determined. In this way, the selection of the transducer was less likely to be influenced by poor manufacturer reporting.

3.1.3 User-friendliness

User-friendliness is important for the design of the screening tool because the goal is to extract the trained radiographer from the screening process. The purpose of the tool is that it is easy to use in a low-resource environment by relatively inexperienced examiners who do not need to interpret the images. The examiners will be trained for a few days and eventually perform the screening procedure themselves. Therefore, the transducer should be easy to use and connect to a phone or tablet.

To grade this property of the transducer, two testers used the transducers for multiple scans of the lungs for a stretched period of time. Both testers had experience using the transducers and knew how to connect and operate them. The testers completed a survey on the transducers and had to grade each determined user-friendliness aspect of the transducers in the survey with a scoring system. The determined aspects on which the transducers were graded are:

- Connection to the phone

- Changeability of imaging parameters

- Positioning of the transducer

- Recording data

- Saving data

- Exporting data

The scoring system with the translation to the scores used to calculate the score for the assessment matrix is shown in table 3.1. The total average score for each transducer was calculated for each tester and the average of the resulting scores of the two testers was used as the input for the assessment matrix.

Label	Meaning	Score
++	Very easy	5
+	Easy	4
+-	Average	3
-	Difficult	2
- -	Very difficult	1

Table 3.1: Scoring system with translation to scores for calculating user friendliness score

3.1.4   B-line phantom experiment

A phantom was created to check the ability of the commercial transducers to image B-lines. These B-lines are associated with TB and other pulmonary infections or diseases [22]. The phantom was constructed based on literature where B-lines were induced in basic LUS phantoms [39] [40].

Phantom design

A schematic drawing of the proposed LUS phantom design is shown in figure 3.1. The layer of meat and the B-line-inducing pockets of fluid have multiple possible options, the bullet points represent the available options. Each part of the phantom should introduce part of what could be seen in a LUS scan of an unhealthy lung section. The meat represents the layers of skin and muscle tissue. The polypropylene sheet introduces a hyper-echoic layer representing the pleural line. The pleural line is hyper-echoic due to the acoustic impedance mismatch at the air and soft tissue interface [3]. The literature suggests that alveoli close to the pleural line filled with fluid cause B-lines [40]. To mimic these fluid-filled alveoli, pockets of fluid were used in the phantom. The phantom described in figure 3.1 gives two options for fluid pockets, these options have been proven in the literature to cause B-lines[39][40].
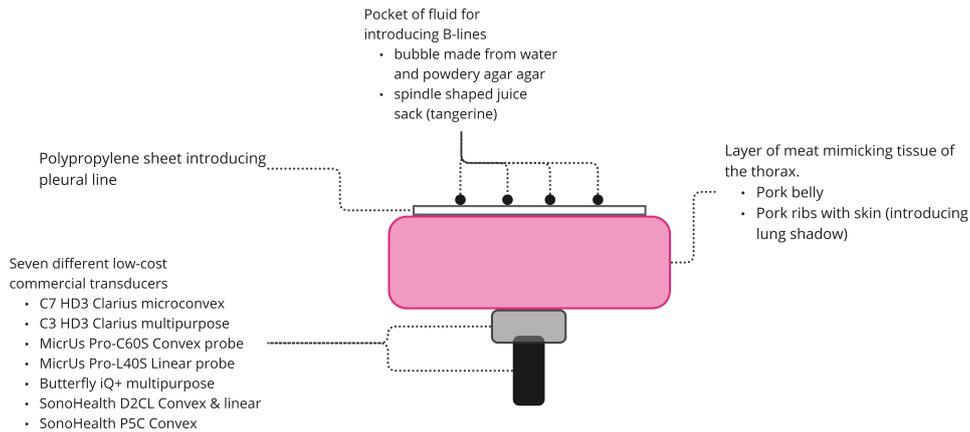
Figure 3.1: Schematic representation of the proposed phantom design

The phantom options were tested with the same US transducer (Telemed MicrUs C60S Micr-curvilinear) with the same settings and visually compared to each other. The first tests were done with pork belly to check whether the spindle-shaped juice sack or the agar bubble worked better for inducing B-lines. The spindle-shaped juice sack phantom and the agar bubble phantom with their frame from the resulting US video are shown in figure 3.2. From the videos and the shown frames, the choice was made to use the agar bubbles because they induced a longer, better recognizable B line, as shown in figure 3.2d.

(a)

(b)

(c)

(d)
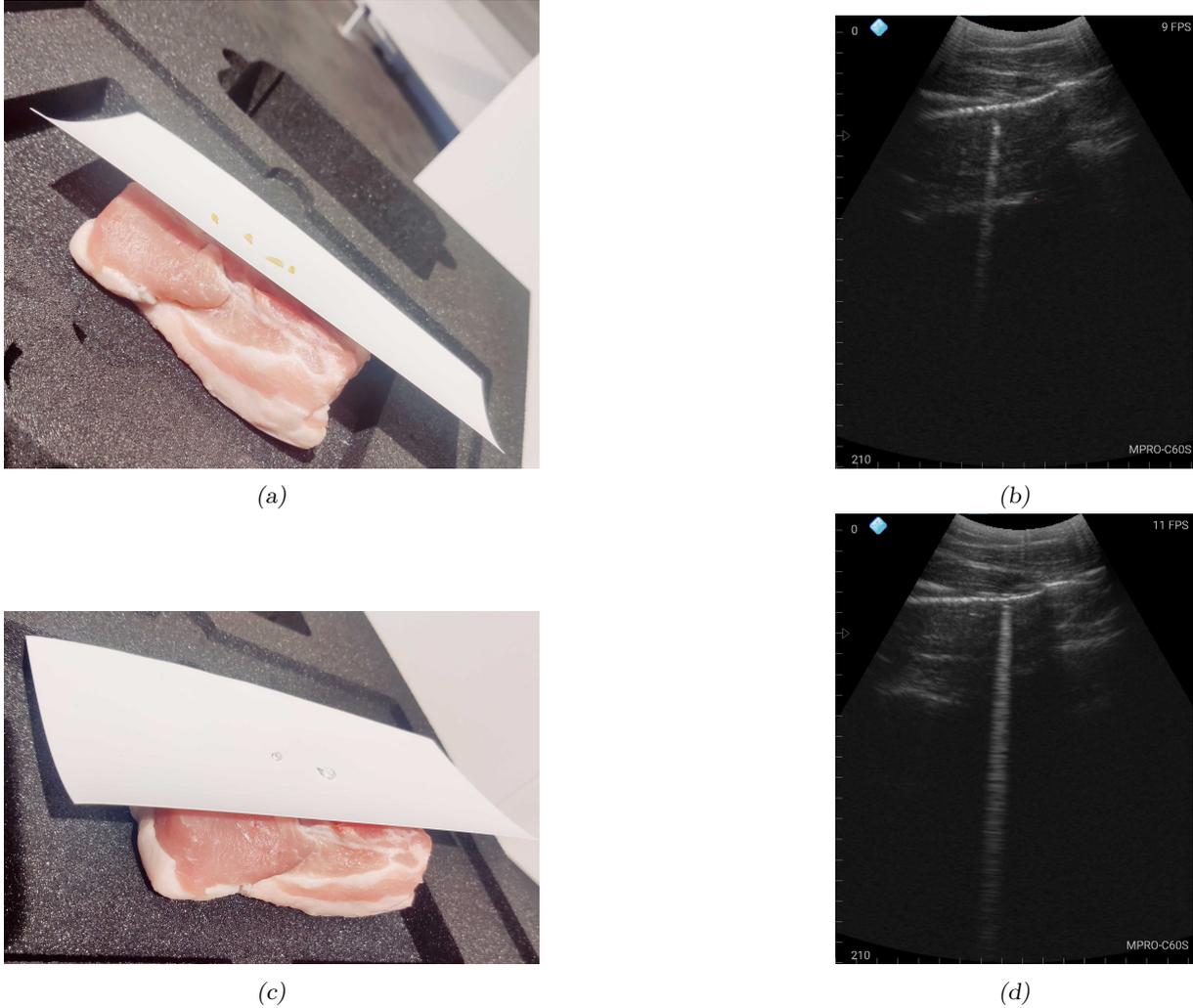
Figure 3.2: comparison of the B-line inducing objects for the pork belly phantom. The spindle-shaped juice sack phantom is shown in image 3.2a with its US representation in 3.2b. The agar bubble phantom is shown in image 3.2c with its US representation in 3.2d

After deciding on the best B-line-inducing object, a test was performed to check whether the pork belly or the pork ribs would be better as a basis for the phantom. A frame of a US recording of a measurement with the pork ribs agar bubble phantom is shown in figure 3.3. For this measurement, more agar bubbles were used to induce more B-lines. The depth and the focus point were adjusted, but the rest of the image settings were equal to the figure 3.2d.
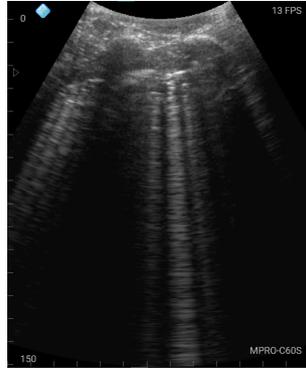
Figure 3.3: US representation of the B-line inducing agar bubbles on the pork rib phantom

The choice was made to continue using the pork ribs for the final B-line experiment because of the presence of the ribs and the resulting rib shadow. This is a better representation of what is seen in the human body shown in figrue 2.3. The last adjustment to the phantom was to exclude the polypropylene sheet because in most cases the connection between parts of the pork ribs and the sheet was poor. This caused an artifact in the image due to the presence of air between the sheet and the ribs. However, the acoustic impedance mismatch at the border between the pork ribs and the air still produced a good representation of the pleural line. The final phantom design for the experiment is shown in figure 3.4.
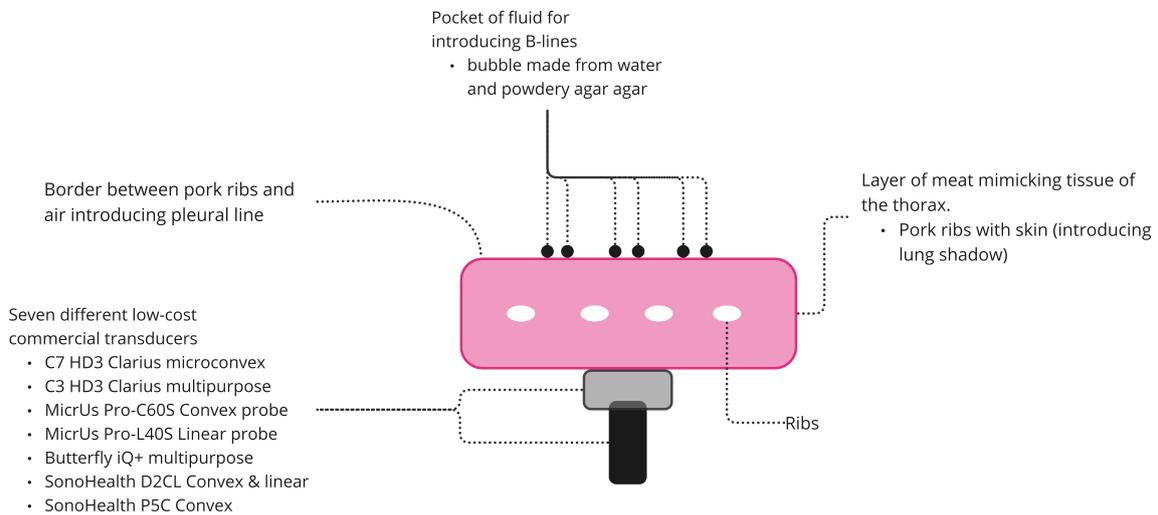
Figure 3.4: Schematic design of the final B-line phantom

Phantom validation

A phantom should mimic reality, therefore, a comparison between what we see in the human body and in the designed phantom was made. All transducers listed in section 3.1 were used to scan the designed phantom (without the B-line inducing agar bubbles) along the same line with the transducer perpendicular to the ribs. Frames of the acquired US B-mode video were compared to frames gathered from data from three different test subjects. These subjects were scant longitudinal on the lateral anterior side of the chest. This is a combination of scan lines 2 and 3 of the protocol shown in figure 2.5. Test subjects declared that they had never been diagnosed with long-term lung infection and therefore were assumed to be healthy.

An example of a comparison between an intercostal space of data acquired from the phantom and the three test subjects is shown in figure 3.5. It can be seen that both the phantom and test subjects have a similar anatomical structure in the form of a white pleural line between two darker rib shadows. The biggest difference between the test persons and the phantom was the thickness of the muscle tissue and the layering of fat between the muscle tissue between the skin and the ribs. This also differs between test subjects. Due to the similarity in anatomical structure, it was assumed that the phantom can be used to simulate the human body for LUS purposes. Therefore, it will be used to perform an experiment to determine the ability of the transducers to detect B-lines.
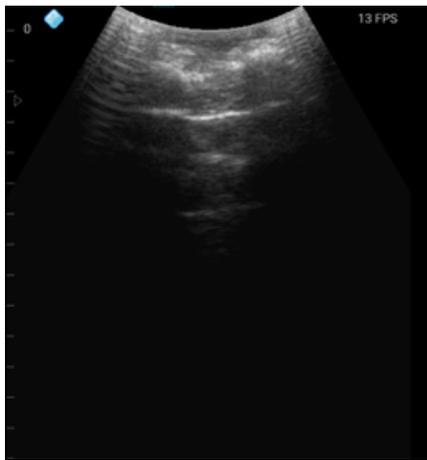
(a) phantom

(b) test subject 1

(c) test subject 2

(d) test subject 3

Figure 3.5: Comparison between the phantom and the three test subjects used for the validation of the phantom for the Telemed MicrUs C60S transducer

After the clinical study in Zambia, the phantom including the agar bubbles was compared again using images acquired from positive patients of B-lines originating from the pleural line shown in figure 3.6. The images show similarity in the structures that were present in the image. However, there was a difference in the brightness of the structures.

(a)   (b)
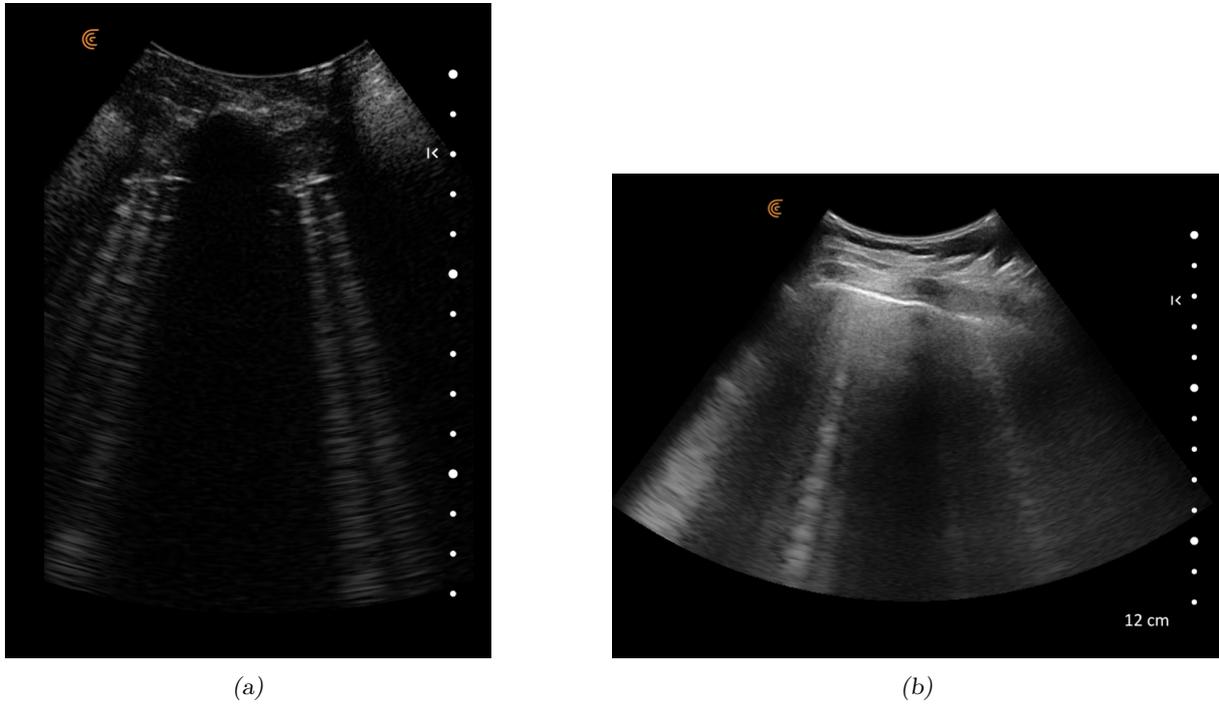
Figure 3.6: Comparison of the finale B-line pork rib phantom with agar bubbles measurement in figure 3.6a and a scanline recording from a patient showing b-lines in figure 3.6b recorded with the same Clarius C3 transducer

Phantom experiment

For the B-line experiment, the phantom described in 3.4 was built and scanned with a standard scan line at the bottom. The phantom was used in two different setups, one experiment with 12 large ($5 - 2mm$) and one with 10 small agar bubbles ($2 - 0.5mm$). The phantom is shown in figure 3.7. The standard scan line was scanned three times with all transducers for both the small and the large agar bubble phantom. In total, six videos per transducer were recorded, and the videos were analyzed by two people who counted the number of B-lines visible in each video. The total number of B-lines counted by each counter for each transducer was calculated and the mean of the two counters was taken. The percentage of detected B-lines was calculated by dividing the mean total of the counted B-lines by the total agar bubbles present on the phantom for the six recordings. The definition of a B-line for counting was as follows:

- The B-line must originate from the area around the pleural line and represent a vertical line.

- A B-line was counted if it could be distinguished from the background and traveled to a depth of at least 2/3 of the maximum depth in the video. The scale bar on the side was used as a reference point.

- In the case that multiple adjacent B lines were present, they were counted as separate B-lines if there was an area of lower intensity in between them.

- B-lines can appear, disappear, and reappear when the transducer is moved along the scan line. Therefore, the horizontal position of the B-line relative to the rib shadow should be watched. A reappearing B-line was not counted twice.

(a)

(b)
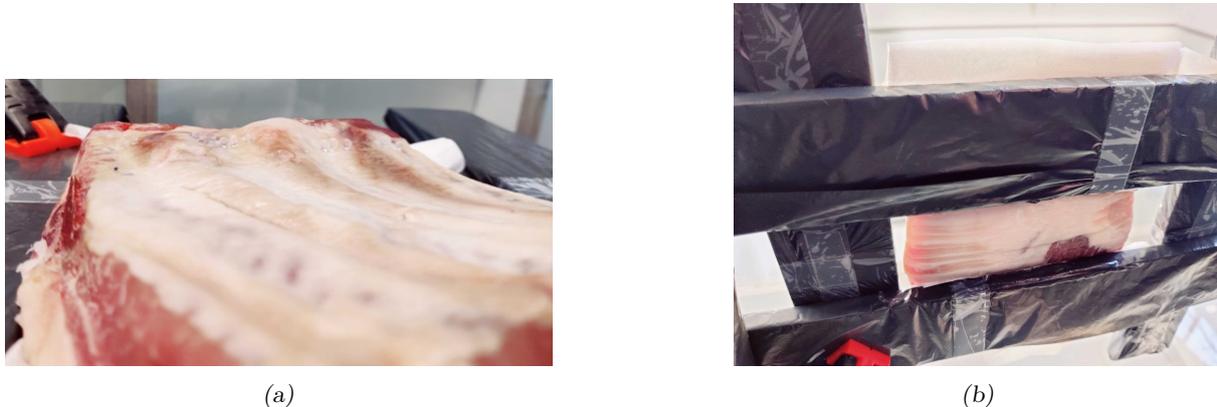
Figure 3.7: Photos of the experimental phantom for the detection of B-lines from above 3.7a and underneath 3.7b

3.1.5 Quality Assurance resolution experiment

The final experiment performed on the transducers is part of a quality assurance (QA) test described in the literature [41]. This test was performed for the three highest-scoring transducers in the assessment matrix that contained the results of the previous experiments. The QA test was used to compare the transducers on their technical properties. This QA4US protocol [41] was used to determine the penetration depth and the subpleural lateral/axial resolution of the transducers. The test was performed at Radboudumc and G. Weijers of Radboudumc was responsible for the processing of the data. The transducer settings were important for this test, a lung preset was used if present for the transducer. This was done because the purpose of the eventual transducer is to perform LUS. If no lung preset was present, a preset with characteristics similar to a lung preset was created. Each part of the QA4US protocol that was used was measured five times to establish an average between the measurements with a standard deviation. The test procedure starts with determining the elevation focus of the transducer with a phantom containing a diagonal slice of hyperechoic material of standard thickness. The focus point was slid to the deepest point and the transducer was shifted over the phantom to take images of the slice at different depths. The thickness of the slice was constant in the phantom, but due to the shape of the US field, the thickness in the image differs. The depth at which the slice is the smallest is the elevation focus of the transducer for the settings used. The transducer's focus was set to this depth to obtain quasi-plane wave conditions. Under these quasi-plane wave conditions, a linear lookup (LUT) table for contrast resolution was created using multiple areas of known dB echo level. This LUT will be used for calibration, a straight (linear) line is desirable. The contrast resolution line will give the relation between the number of gray levels per dB echo level for the transducer for the selected settings.

After creating the LUT, the depth at which the signal was still 6 dB above the background (noise) signal was calculated. This depth can be seen as the penetration depth of the signal for the selected settings. For this measurement, the LUT was used to calculate the depth at which the dB echo level was 6 dB by translating the gray level into the dB echo level. Finally, the axial and lateral resolutions were determined at a depth of about $4cm$. This was done by keeping the focus point at the depth found during the slice thickness test and scanning a vertical line of small elements in a phantom. The element closest to a depth of 4 cm was chosen for analysis and its in-plane point spread function was measured. The gray-level profiles of the in-plane

point spread function in the axial and lateral directions were translated to dB echo level and a parabola was fitted for the upper dB echo level measurements. The full-width half maximum at the $-6dB$ level was taken as the axial and lateral resolution of the transducer for the selected settings at the depth of the wire element.

All technical specifications found for the transducers were translated to a score and an accompanying weight for the assessment matrix. For the penetration depth, a depth above $50mm$ is desirable because anatomical structures can be present a view centimeters posterior to the pleural line. The axial and lateral resolutions are desired to be smaller than $1mm$ because the smallest anatomical structure used in LUS is an SPC and is defined around a size of $1mm$, therefore a resolution below $1mm$ would be perfect to make a distinction between two SPCs. A resolution greater than $1mm$ but below $1.5mm$ is graded as sufficient because this resolution would still be able to show the consolidation. However, the distinction between two neighboring SPCs is harder to make.

3.2   Protocol compliance detection algorithm

This section will elaborate on the written algorithm for protocol compliance. An overview of important steps in the code and their physical basis are given below. The data that will be used to generate results is gathered during a clinical study using the transducer selected in the transducer selection study of this thesis. However, the algorithm design is based on the physical properties of the thorax tissues in US imaging. Therefore, the working principle could be applied to LUS recordings of other US transducers as well.

3.2.1   Data set

The data used to design and test the protocol compliance algorithm was gathered during a clinical study in Zambia. The clinical study performed LUS scans on presumptive TB-positive patients using the Clarius C3 multipurpose US transducer with an abdominal preset, a depth of 11cm, and the focus point at the pleural line. The choice of using the abdominal preset instead of the lung preset was made because the radiologist in Zambia advised the use of this preset due to his experience with LUS. The radiologist trained radiographers for data collection with this preset because in his opinion it was easier to see the position in relation to the real world with this preset. The protocol used for data gathering is described in section 2.3.2.

From the data collected during the clinical study, two data sets were created. Not all data was recorded with the same depth described in the determined preset. Therefore, only recordings with a depth setting between 10-14 cm were used in the data sets. The first data set is used for the design and determination of the threshold for the protocol compliance algorithm. The second data set is used for the performance testing of the designed algorithm.

Both data sets contain 30 scan lines from 30 different patients. The aim was to include a variety of scan lines from the protocol described in section 2.3.2 in each data set. Furthermore, both data sets should contain a combination of sufficient and insufficient recorded scan lines. Therefore, a selection of 60 videos was made from all available data by someone with some experience in LUS. From the list that contains the 60 labeled scan line recordings, the two data sets were selected.

The training data set was labeled by someone with some experience with LUS. The test data set was labeled by the expert Drs. B. Kok. He is an internist at Radboudumc with experience in LUS. Labeling was either "sufficient" or "insufficient" for both the test and training data sets. The data sets do not contain the same recordings.

3.2.2   Pleural threshold algorithm

Section 2.3.2 describes that an angle close to 90 degrees to the pleura is necessary to record a usable scan. If this is not the case, abnormal tissues or artifacts induced by abnormal tissues might be missed. An improper angle to the pleural line can be identified by lack of intensity at the pleural line and bone surface or absence of A-lines in healthy lung tissue. If the angle of the transducer is close to 90 degrees, then high grayscale values with respect to the surrounding structures are expected at the pleural line in the image. This is also the case for the bone surface. Therefore, the grayscale values within the segmented bone or pleural structures are subjected to a threshold to check whether a recording was performed correctly.

In addition to the expected high intensity at the pleural line and the bone surface, there is a drop in the pixel intensity after these structures due to their high reflection index. These pleural line and bone surface properties can be exploited to segment the structures using a shadow peak-based algorithm. The shadow peak algorithm requires a rectangular grid. Therefore, a back scan conversion was used to transform the cone-shaped output image into a rectangular-shaped image.

Backscan conversion

A curvilinear US transducer records an image line by line which results in a rectangular (Cartesian) grid of pixel intensities, this grid is then transformed into a cone-shaped polar coordinate system. This results in the output of a cone-shaped B-mode image that takes the curvature of the US probe into account. The shadow peak method for bone segmentation is based on the attenuation of pixel values in the y direction of a 2D x,y rectangular grid. When the shadow peak method is applied to a cone-shaped curvilinear B-mode image, pixel values outside the cone (which have intensity zero) are part of the analyzed rectangular grid. Furthermore, the angle of the cone-shaped grid will result in a wrongful definition of a column because the column in the rectangular cartesian grid is not the same as the column in the cone-shaped polar grid.

To solve these problems, a backscan conversion needs to be applied to the cone-shaped output image. This will convert the image back to its original rectangular line-by-line grid, which does not take the curvature of the US probe into account. A schematic and visual representation of this transformation is shown in figure 3.8.

For the backscan conversion, the (imaginary) top of the US cone must be calculated by defining the angle of the sides of the cone with respect to the middle line of the cone. The found top (the rotational center) is used to calculate the polar coordinate system within the Cartesian coordinate output image [8]. By extracting the polar coordinates from the Cartesian image, a new rectangular Cartesian grid is created with the transformed US image as shown in figure 3.8.
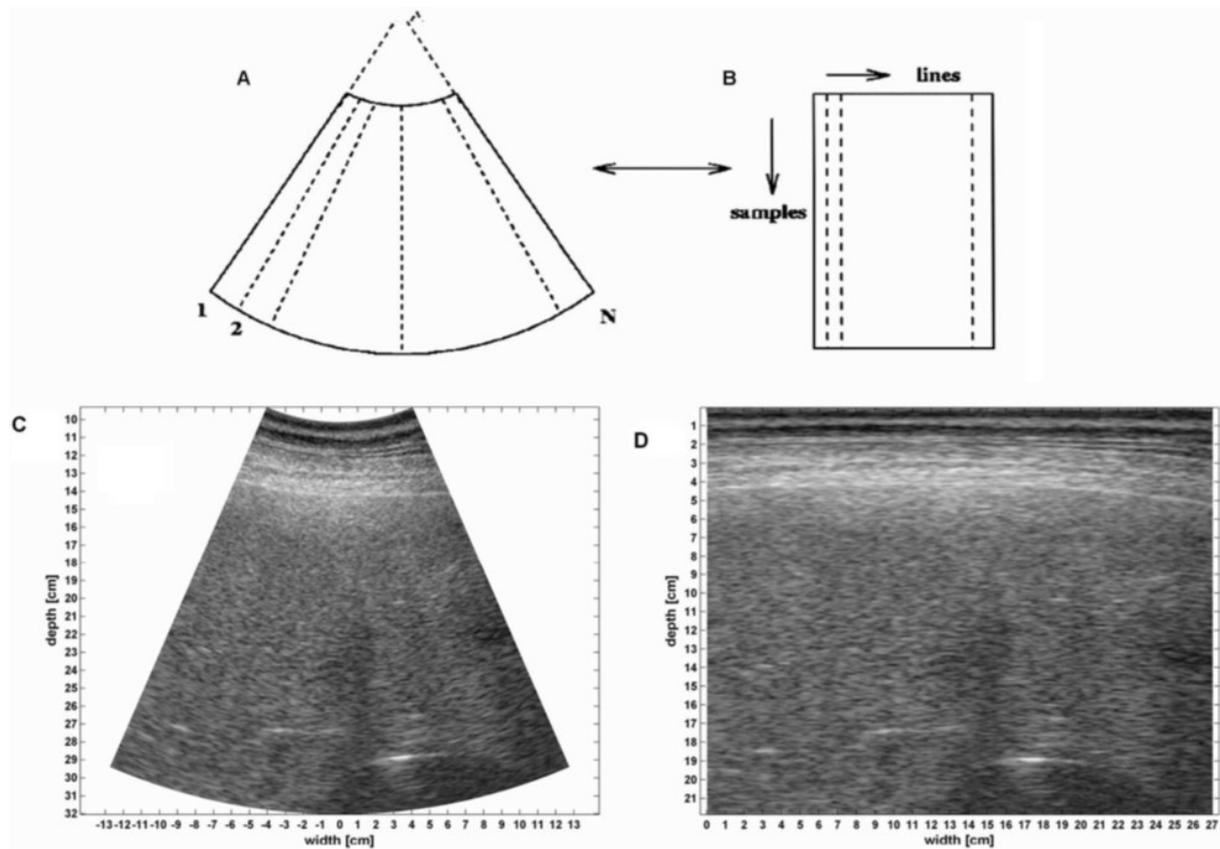
Figure 3.8: Back-scan conversion: (a) scheme of original sector B-mode image; (b) same sector after back-scan conversion; (c) B-mode US image of the liver; (d) same US B-mode image after back-scan conversion [8].

Shadow peak bone segmentation

For bone segmentation in US images, the shadow peak algorithm can be used. This algorithm makes use of the shadows under the bone. The shadow is an area of low intensity posterior to the high-intensity bone surface. By locating the areas under which there is an intensity drop, one can theoretically find the bone in the image. Pixel intensities in a grayscale image are usually between the values 0 and 255. To locate areas with a sudden drop in intensity, a measure of confidence is calculated. This confidence defines the measure in which the pixel represents a shadow. This is repeated for each voxel in the image with equation 3.1 [42].

$$S(x, y, z) = 1 - \sqrt{\frac{\sum_{i=1}^{i=y} I_{x,i,z}}{\sum_{j=1}^{j=Y} I_{x,j,z}}} \tag{3.1}$$

In this equation, $I_{x,y,z}$ represents the intensity in the voxel with coordinates x, y, z representing (column, row, frame). The capital Y is the maximum y coordinate. So, under the root, the cumulative sum of the intensities I in a certain row for a voxel in a certain column and frame is divided by the total sum of intensities in that column of the selected frame. This results in a confidence map for the bone surface. From this confidence map, the peak value for each column is taken to estimate the possible bone position for that row. By looking for connected pixels in the peak confidence voxel map the bone structure in the image can be found.

In addition to the use of this method for bone segmentation, it is expected that it can also be used for pleural line segmentation due to similar physical properties in the US image[37]. Posterior to the pleural line, the US signal will be attenuated. After multiplication with the confidence map, the pleural line and bone structures will be present in the column-based peak intensity image.

Shadow peak based algorithm

The shadow peak bone segmentation algorithm described in the section above is implemented in Python to create a confidence map for the bone and pleural line because they share similar properties. An example of a confidence map with its accompanying US frame is shown in figure 3.9.

(a)   (b)
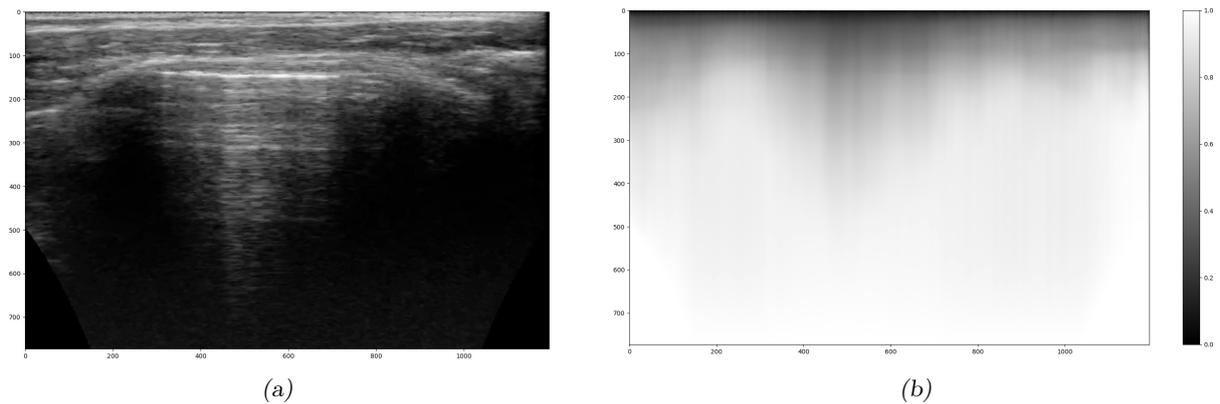
Figure 3.9: (a)Backscan converted US frame with its accompanying (b) shadow peak confidence map

The next step is multiplying the confidence map value by the pixel intensity of the corresponding image pixel. This results in a multiplication map image in which the area anterior to the rib and the pleural line is attenuated while the rib and pleural line areas are enhanced as shown in figure 3.10a. In the resulting image, the peak value for each column is selected and a mask is created with an intensity value of 1 for the peak pixel positions and an intensity of zero for the background. A dilatation on the mask structures is done to connect pixels that are close together so line elements are formed. The resulting mask image is shown in figure 3.10b.

(a)   *(b)*
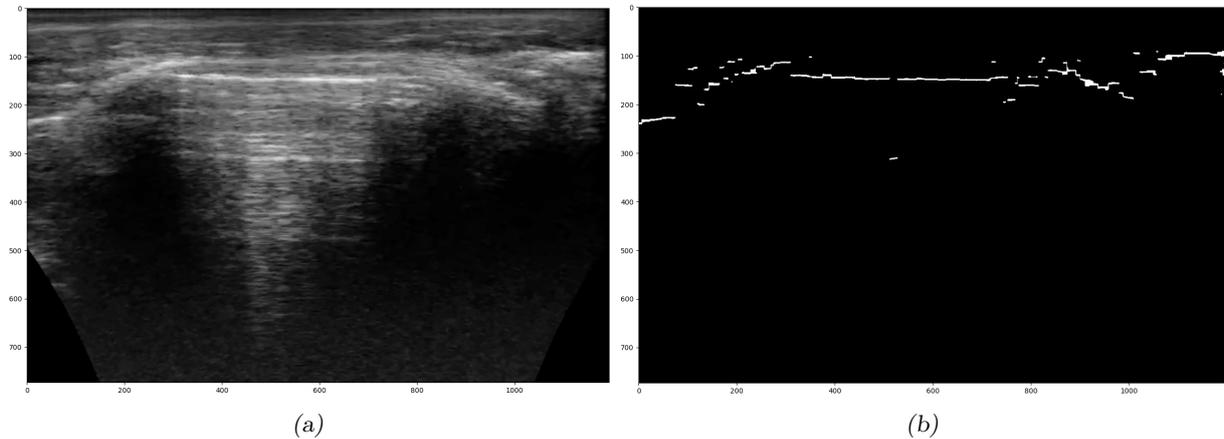
Figure 3.10: (a)The backscan converted frame multiplied by the shadow peak confidence map (multiplication map). (b) The mask resulting from dilating the maximum pixel value positions for each column in image 3.10a

The dilated mask is multiplied by the original backscan converted image and a threshold is applied such that only pixels with high intensity are kept. In this way, pixels that do not represent a high intensity in the multiplication map are compared to the other selected maximum pixels and are removed. This is done to get rid of pixels selected in columns with low intensities, in those columns no pleural line or bone surface is present. The resulting image after applying the threshold is shown in figure 3.11, this image will be used to segment the pleural line and part of the bone surface after additional thresholds are applied.
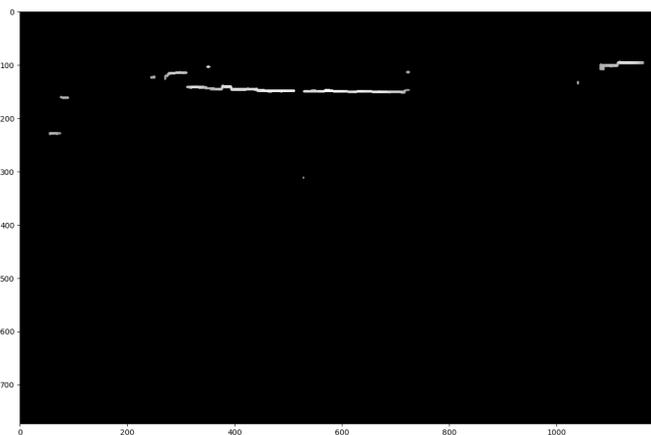
Figure 3.11: Resulting mask structures after thresholding with their pixel intensities from the original input backscan converted image

Segmenting ribs and pleural line

The image resulting from the shadow peak algorithm shown in 3.11 is the input for the final part of the segmentation of the pleural line and the bone surface. The goal is to create separate masks for both structures that contain the pleural line or bone surface if the intensity compared to its surroundings is sufficient. Thus, a distinction is made between a recording with a sufficient angle and an insufficient angle to the pleural line. The processing of frames with the shadow peak algorithm is similar for the creation of bone and pleural line masks. The main difference is the orientation of the structures relative to each other for the selection of the final structures in the mask.

To select the pleural line and ribs, a mean intensity threshold is applied. This threshold is different for each frame and is calculated with equation 3.2. With $meanIntensity$ representing the mean intensity of the upper $\frac{2}{3}$ of the analyzed frame, $maxValueCone$ representing the maximum intensity value in the US cone and $multiplicationFactor$ representing a factor between 0 and 1. To determine the value of the multiplication factor, five scan line recordings of 10 seconds of scan lines 2 and 3 of figure 2.5 are made on a test subject. Each of the five scans was performed at a different angle to the thorax, the scan angles were 50, 70, 90, 110, and 130 degrees. These angles were achieved during scanning using a triangle ruler. The angle between the pleura and the thorax is fixed. So, if the angle between the transducer and the thorax is changed, the angle between the pleura and the transducer changes in equal steps. Three frames were randomly selected from each recording. For the selected frames the $meanIntensity$, $maxValueCone$, and a manually segmented pleural line mean intensity are calculated. The average was taken for these values from the three frames for each scan angle recording. The average values are used as input for equation 3.2. The equation was modeled for multiplication factors between 0 and 1 and plotted together with the found mean intensities of the pleural line for the scan angles in figure 3.12.

$$Thres = meanIntensity + (multiplicationFactor * (maxValueCone - meanIntensity)) \quad (3.2)$$

The goal is to exclude mean pleural line intensities that are not high enough in intensity to be a well-recorded pleural line. During the experiment, it was decided that the mean intensity of the pleural line corresponding to the 90 degree recording was sufficient. The recordings with a 50, 70, 110, and 130 degree angle with respect to the chest did not contain a good representation of the pleural line. Therefore, their intensity should be excluded using the set threshold. A value is excluded if it is below the angle threshold line. To achieve this, the multiplication factor needs to have a value for which the pleural intensity of a 90 degree recording is above the threshold and the pleural intensities for the other angle recordings are below the threshold. Using figure 3.12 and taking a standard deviation for the mean pleural intensity of approximately 5 into account, the multiplication factor chosen for the mean intensity threshold of the algorithm is 0.7. Due to a lack of time, there was no possibility to check the interval between 80 and 100 degrees with a smaller step size.

The mean intensity threshold ensures that only high intensity structures are selected in relation to the mean intensity of the original image. If the angle of the probe is insufficient, no pleural line or bone surface will be selected due to this threshold.
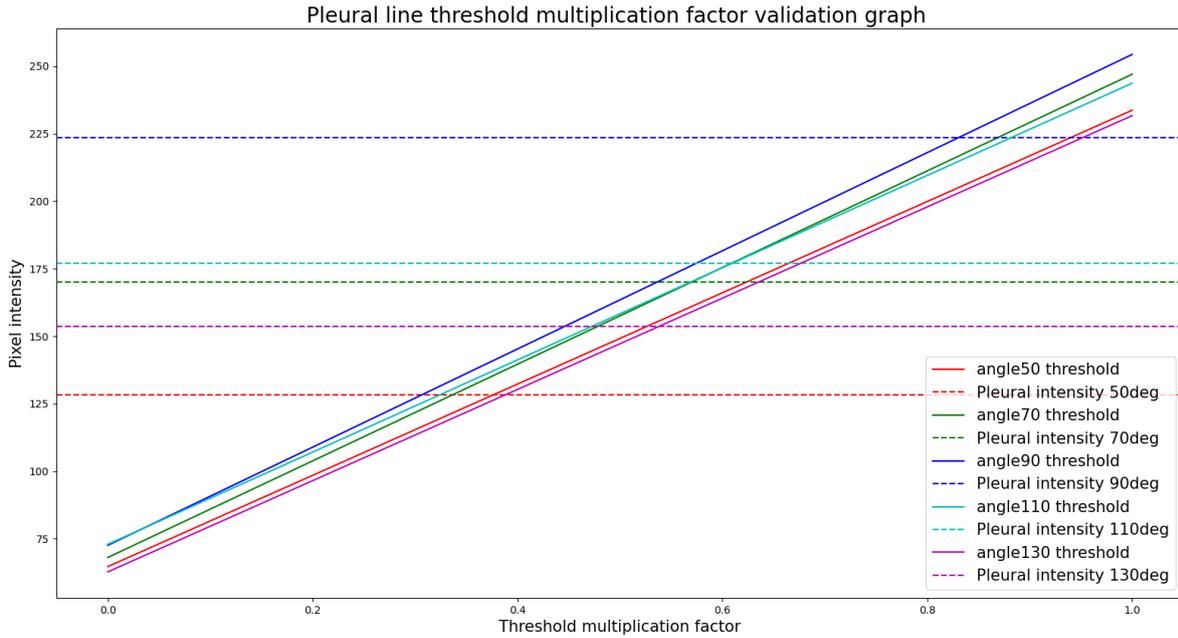
Figure 3.12: Pleural line intensity lines for multiple scan angles with their accompanying modeled thresholds for multiplication factors between 0 and 1 represented in the same color to determine the pixel intensity threshold

If the mean pleural line intensity of the segmented structures is above the set threshold, the bone or pleural lines will be transferred to a new mask. Structures with a mean intensity below the set threshold will be removed. The remaining structures are then analyzed based on their mean y-position related to each other. Bone always has an anterior position with respect to the transducer in comparison with the pleural line therefore the anterior structures are used for a bone mask and the posterior structures are used for the pleural line mask. The resulting masks are shown in figure 3.13a & 3.13b. The mask plotted as an overlay over the back scan converted image is shown in figure 3.14a & 3.14b.
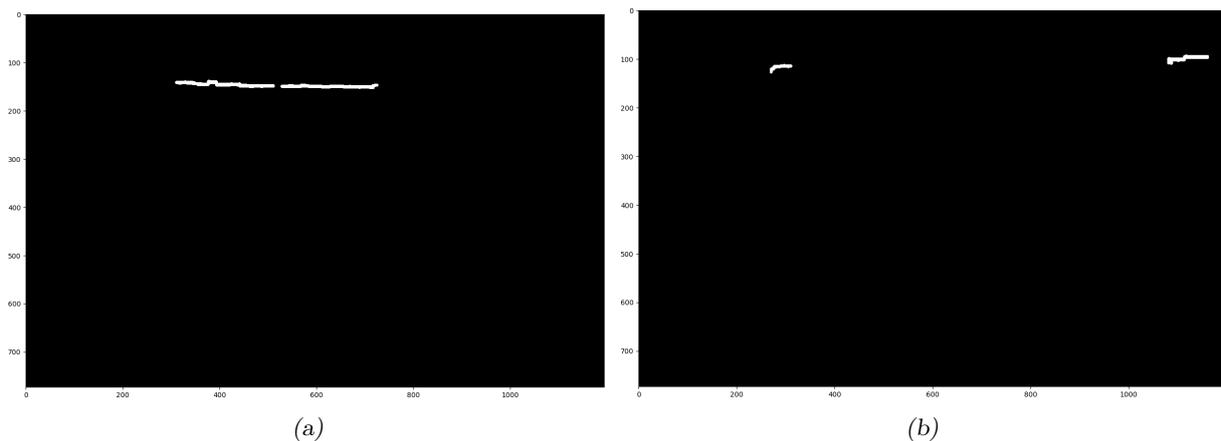
Figure 3.13: (a)Example of the resulting segmentation for the pleural line (b) Example of the resulting segmentation for the bone surface
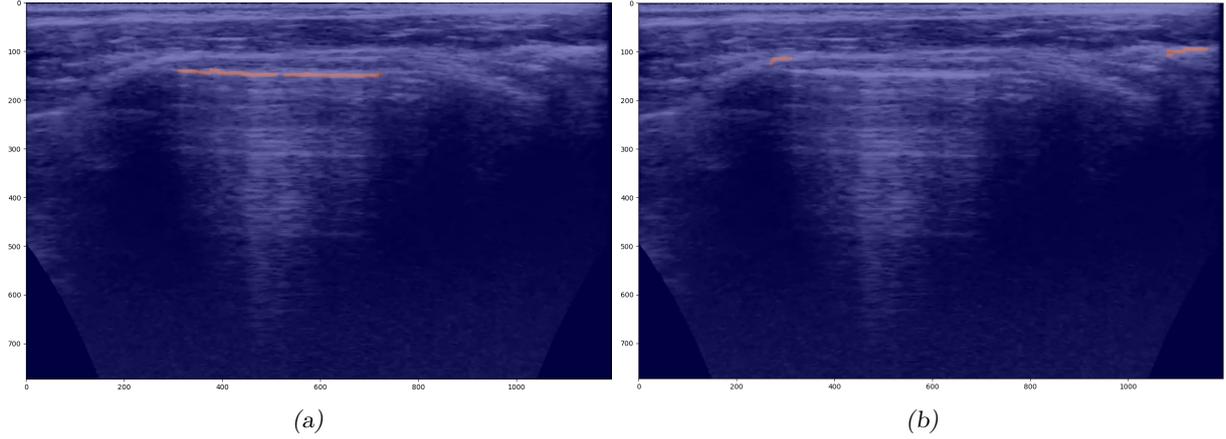
Figure 3.14: (a)Example of the resulting segmentation overlay for the pleural line on the backscan converted US frame (b) Example of the resulting segmentation overlay for the bone surface on the backscan converted US frame

The described process for segmentation is repeated for all frames or a number of the total frames to generate an eventual score to determine if the recording was done correctly. If a structure is present in the resulting pleural segmentation mask, output 1 (sufficient) is given for the frame analyzed. If no structure is present in the segmentation map, an output of 0 (insufficient) is given for the frame analyzed. The average output of the frames analyzed is taken and displayed as a score for the detection of the pleural line in all frames analyzed. This score is subjected to a threshold to give the output 'sufficient' or 'insufficient' recording.

The threshold is defined using the training data set with its labels. The labels are used as the ground truth so the results of the algorithm can be defined as true positive (TP), true negative (TN), false positive (FP), or false negative (FN). These definitions are used to calculate the sensitivity and specificity with equations 3.3 & 3.4.

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3.3)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (3.4)$$

The sensitivity is the true positive rate of the algorithm or the probability of a positive test result if the input of the system is positive. The specificity is the true negative rate of the algorithm or the probability of a negative test result if the input of the system is negative.
The 30 training data set video files are used as input for the algorithm using thresholds between zero and one with a stepsize of 0.01. The ideal threshold is found from a ROC curve constructed by calculating the specificity and sensitivity for each threshold using formulas 3.3 & 3.4. The resulting ROC curve is shown in figure 3.15. For an algorithm to be useful the curve needs to be on the left side of the baseline, this is the case for the curve. The best threshold for the algorithm results in a combination of high specificity and high sensitivity, because that would result in the best ability of the algorithm to distinguish incorrect from correct recorded scanlines. This is the case for the point in the graph closest to the top left corner of the graph because the x-axis is in $1 - Specificity$. It is shown in figure 3.15 that the best threshold for the training data is 0.3. Therefore, this threshold is used in the final algorithm for all scan lines in the protocol. If the resulting pleural line score is above the threshold for an analyzed video the algorithm gives the following output: 'Video is sufficient for analysis. Continue to the next scan line'.
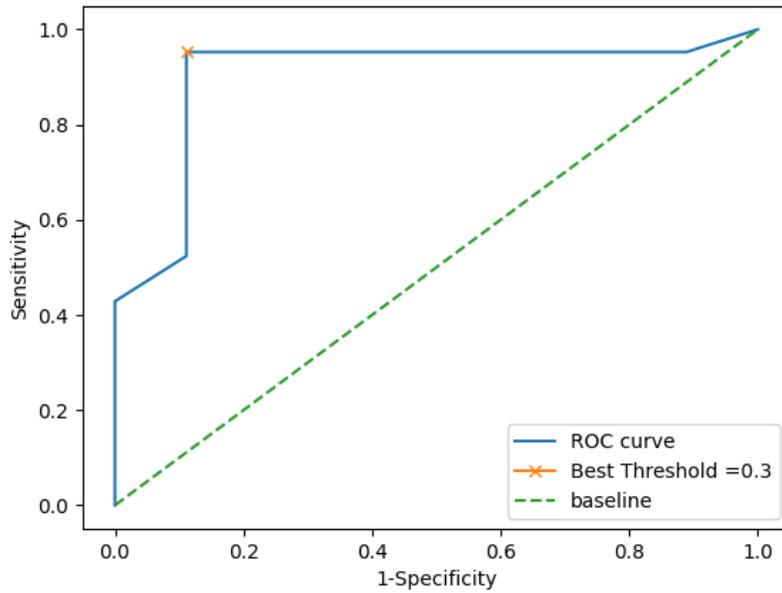
Figure 3.15: ROC curve of the pleural line detection algorithm for the labeled scanlines in the training dataset

3.2.3 Skin contact algorithm

Besides the pleural mask score, the algorithm checks if contact with the skin was sufficient during the recording for the scan lines that are rejected by the pleural threshold algorithm. This is done by checking the intensity values of the upper 40 rows of the frames. In the case of insufficient skin contact, the US image will not be completely filled with pixel intensities because a low amount of signal is returned to the transducer. This principle is already present at a depth of $1cm$ on the right side of the US cone shown in figure 3.16.
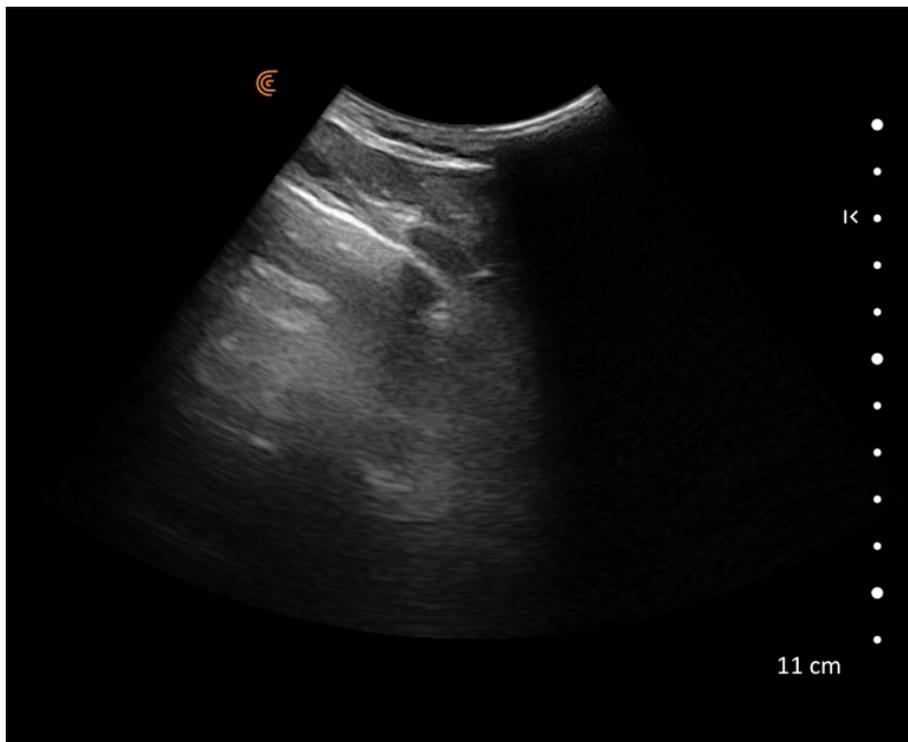
Figure 3.16: Example of US frame recorded with bad skin contact

37

The dark pixel values do not have a higher intensity than 10 in a possible pixel intensity range between 0 and 255. Therefore, the number of pixels with a value lower than 10 is counted in the top 40 rows of the backscan converted US cone. This number is divided by the total number of pixels present in the top 40 rows to calculate the percentage of pixels that are not expected to contain usable data (dead pixels). The resulting percentage is subjected to a threshold for the minimum number of pixels that should contain usable data. This threshold is 34% and is chosen after four scan lines of scan line 2 from figure 2.5 with different skin contact were analyzed. The scan lines were labeled as very good, good, bad, or very bad skin contact by a reader who had knowledge of the LUS recordings. The average percentage of dead pixels in the top 40 rows is calculated for all frames in the four recordings and is shown in table 3.2. Videos 2 and 3 had a percentage below 30% but videos 1 and 4 had a percentage above 37%. A threshold was chosen between the percentages of dead pixels in the good- and bad-labeled videos. The percentage between 29% and 38% resulted in a rounded average of 34%.

Video number	Skin contact label	% of dead pixels
1	very bad	43%
4	bad	38%
3	good	29%
2	very good	3%

Table 3.2: LUS recordings used for determining the percentage of dead pixels threshold

If the percentage of dead pixels is above 34% a 0 is saved in an array, if the percentage of dead pixels is below 34% a 1 is saved, this process is repeated for each frame.

The final skin contact score is calculated by taking the mean of the skin contact array that contains either a 0 or a 1 for the frames in the video. The threshold for the skin contact score is found by analyzing the training data set videos that are defined as insufficient recordings by the pleural threshold algorithm. These videos are labeled by someone with experience in LUS as a video with sufficient or insufficient skin contact. These labels are used as ground truth and with the skin contact scores for the videos the ROC curve shown in figure 3.17 is made to determine the skin contact threshold value.

In figure 3.17 it is shown that the best possible threshold resulting from the training data set is 0.25, this threshold is used for the final algorithm. If the skin contact score is below the threshold of 0.25, the algorithm will give the output: 'Video is NOT sufficient for analysis. Check scan angle and contact with the skin or use more gel. Record scan line (scan line number) again'

If the score for the segmentation of the pleural line is below the set pleural line thresholds but the skin contact score is above its threshold, the algorithm outputs: 'Video is NOT sufficient for analysis. Check scan angle. Record scan line (scan line number) again". The final algorithm includes the skin contact threshold algorithm.
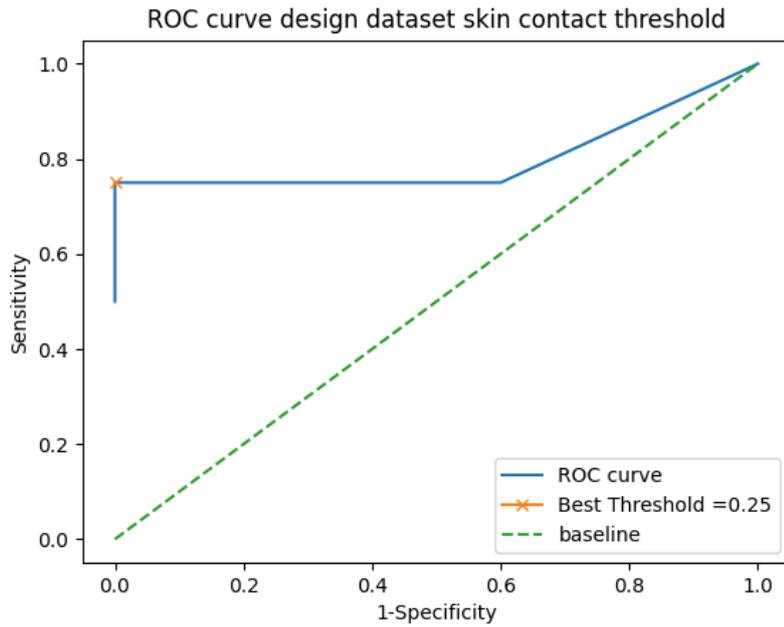
Figure 3.17: ROC curve of the skin contact threshold for the labeled scanlines rejected by the pleural threshold in the training dataset

3.2.4 Overview algorithm thresholds

This section presents the thresholds resulting from the previous sections in table 3.3.

Threshold	Value
Multiplication factor pleural line intensity threshold	0.7
Pleural line score threshold	0.3
% of dead pixels for skin contact	34%
Skin contact score threshold	0.25

Table 3.3: Overview of the thresholds used in the algorithm and their value

3.2.5 Pleural threshold performance test

The performance of the algorithm was tested against expert opinion. Expert opinion is the label assigned by Drs B. Kok to the separate video files in the test dataset. To rate the performance of the algorithm, the expert opinion was used as the ground truth. The selected test data set containing 30 scan lines was annotated by the expert as a 'sufficient' or 'insufficient' LUS recording and the reason for the ruling was added. The expert decided whether the recorded scan line was a sufficient scan for the recorded scan line of the scanned thorax area. Sufficient means that the scan line can be used to look for possible symptoms. Insufficient means that it lacks too much information and needs to be recorded again.

The same test data set was analyzed by the designed protocol compliance algorithm that also gives an output of 'sufficient' or 'insufficient'. The results of the expert labeling and the results found by the algorithm were compared and the expert findings were used as ground truth. Using this ground truth, the algorithm's results were classified as true positive (TP), true negative (TN), false positive (FP) or false negative (FN) so that the accuracy, sensitivity, and specificity of the designed algorithm with the selected thresholds can be calculated using equations 3.3, 3.4, and 3.5. The accuracy is the percentage of correct predictions made by the model.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.5}$$

3.2.6 Skin contact threshold performance test

The results of the pleural threshold performance test that were classified as TN or FN show that the algorithm gives the output of an insufficiently performed scan. For these instances, skin contact was analyzed to give advice, in addition to that the scan line needs to be recorded again. This advice will be 'check contact with the skin or use more gel' in case of a score below the set skin contact threshold. To check the performance of the skin contact threshold, the scan lines that produce a negative result for the pleural threshold were analyzed with the skin contact algorithm. These negative outputs were labeled by someone with experience in LUS as either sufficient contact with the skin or insufficient contact with the skin.

The labels were used as the ground truth so that the algorithm results can be defined as true positive (TP), true negative (TN), false positive (FP) or false negative (FN). These definitions were used to calculate the sensitivity, specificity, and accuracy of the skin contact algorithm with equations 3.3, 3.4, and 3.5.

3.2.7 Speed test

Besides the performance of the algorithms, a speed test was performed for the entire protocol compliance algorithm. For this test, the processing time and pleural line score were recorded for each of the 30 videos in the test data set. This was done for different amounts of the total frames. It was done for all frames (≈ 245) divided by a value from the vector $frameDevide$ [0,2,4,8,16,32]. So, the algorithm was used 6 times for all 30 videos. The average run time for the 30 videos was taken for each value of $frameDevide$ and the pleural line scores were used to calculate the sensitivity, specificity and precision of each value in the $frameDevide$ vector. The final results are displayed for the number of frames analyzed.

Results

4.1 Transducer selection

This section presents the results of the transducer selection experiments and the resulting assessment matrix.

4.1.1 Specifications stated by the manufacturers

The results transferred to the score for the assessment matrix are listed in the appendix with a total summed score for the specifications stated by the manufacturers in appendix A.1. Two examples of property values transferred to the score are shown in table 4.1. Table A.1 shows that four of the seven transducers have a total score of 98 or higher and three transducers score 81 or lower. The transducers listed from highest to lowest scoring are Telemed C60S, Telemed L40S, Clarius C3, Clarius C7, MSLPU79, SonoHealth D2CL, and Butterfly iQ+. An example of a large difference in property between the transducers is shown in table 4.1 for the price. Other big differences between the transducers are the availability of raw data and a software development kit (SDK).

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Telemed MicrUs Pro-L40S Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear & linear	MSLPU79 transducer Curvilinear
Price	€3395 (1)	€3395 (1)	€2300 (3)	€2400 (3)	€2541 (3)	€1600 (5)	€1480 (5)
Battery life	60 min (3)	60 min (3)	Max scan time 120 min (5)	Max scan time 120 min (5)	144 min (5)	90 min (4)	90 min (4)

Table 4.1: Value transferred into assessment matrix score for the properties Price and Battery life

4.1.2 User-friendliness

The results transferred to the score for the assessment matrix are listed in table 4.2 for the user-friendliness surveys taken by two testers. The completed surveys can be seen in Appendices A.3 and A.4. The resulting scores in table 4.2 show that the Clarius C3 transducer scores highest for User-friendliness. This is due to the usability of the app, its integration with the phone storage, and the good contact with the skin that produces a clear image. Close seconds are the Clarius C7, Telemed C60S, and Telemed L40S. These transducers are easy to use, but for the Clarius C7, skin contact is not as good as for the Clarius C3. The differences between the Clarius and Telemed transducers are caused by differences in data storage, recording, and phone connection. The MSLPU79, SonoHealth D2CL, and Butterfly iQ + transducers score at the lower end of the spectrum due to the limited changeability of parameters and the saving / extracting of data.

However, these transducers scored higher on the connection to the phone. The resulting scores from table 4.2 are put into the assessment matrix with a weight of five due to their importance.

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Telemed MicrUs Pro-L40S Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear & linear	MSLPU79 transducer Curvilinear	Weight
User-friendliness	5	4	4	4	1	2	2	**5**

Table 4.2: Results of the user-friendliness surveys transferred to score with the corresponding weight

4.1.3 B-line phantom experiment

The results of the B-line experiment with their transferred score for the assessment matrix are listed below in table 4.3. The results show that the Clarius C7 detects the highest percentage of B-lines in the phantom. In addition to the C7 transducer, the Clarius C3, Telemed C60S, and Butterfly iQ+ transducers also perform at a high percentage of detection above 80%. The SonoHealth D2CL transducer score is decent with a percentage of 76%, which translates to an average score of three for the assessment matrix. The Telemed L40S and MSLPU79 score below average with a score on the assessment matrix of two and a percentage around 65%. The scores on the assessment matrix are based on intervals of 10% between 100% and 50%.

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Telemed MicrUs Pro-L40S Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear & linear	MSLPU79 transducer Curvilinear	Weight
Percentage of total B-lines detected	84% (4)	92% (5)	84% (4)	64% (2)	82% (4)	76% (3)	66% (2)	**5**

Table 4.3: Results of the B-line experiment in the percentage of total B-lines detected and their transferred score with the corresponding weight

4.1.4 Quality assurance resolution experiment

The quality assurance test results for the penetration depth, axial resolution and lateral resolution with their scores and weights from the assessment matrix are shown in table 4.4. The Telemed C60S and Clarius C3 are shown to have a penetration depth below the 50 mm mark for the chosen transducer settings. The axial resolution is the smallest for the Clarius C7 but all values are far below the preferable $1mm$ mark. For the lateral resolution, the Clarius C7 and Telemed C60S have a value above the $1mm$ mark. However, the Clarius C3 has a good lateral resolution of $0.86 \pm 0.13mm$

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Weight
Penetration depth [mm]	84.8 (4)	44.5 (2)	111.4 (5)	**1**
Depth for resolution calc. [mm]	38.0 ± 0.25	36.5 ± 0.13	36.4 ± 0.19	**-**
Axial resolution [mm]	0.40 ± 0.04 (5)	0.34 ± 0.01 (5)	0.51 ± 0.16 (5)	**3**
Lateral resolution [mm]	0.86 ± 0.13 (5)	1.14 ± 0.09 (3)	1.21 ± 0.09 (3)	**3**

Table 4.4: Results of the QA tests for the focus point, axial resolution, and lateral resolution with their transferred score and corresponding weight

4.1.5 Results assessment matrix

The final results of the transducer selection are shown in table 4.5. Technical specifications are only evaluated for the three highest-scoring transducers before the evaluation of the technical specifications. The total score before technical specifications is shown in Appendix A.2. The focus of the final results is on the three transducers selected for technical evaluation. These are the Clarius C3, the Clarius C7 and the Telemed C60S transducers. The transducer with the highest final score is the Clarius C3 that scores 88% of the total possible score. The second highest scoring transducer is the Telemed C60S with also 88% of the total possible score but one point lower for the total score. The lowest scoring transducer is the Clarius C7 with 81% of the total possible score. The greatest differences in the total score between the transducers are introduced by a difference in battery life, charging time, price, cloud storage options, penetration depth and lateral resolution. The accompanying scores for these properties can be found in tables A.1 & A.2. Based on the results from the assessment matrix, the Clarius C3 transducer is used for the data gathering for the protocol compliance algorithm.

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Weight
Total manufacturer spec	102	98	110	-
User-friendliness	5	4	4	5
B-line detection	4	5	4	5
Total before technical specs	147	143	150	-
Penetration depth	4	2	5	2
axial resolution	5	5	5	3
Lateral resolution	5	3	3	3
Total score	185 (88%)	171 (81%)	184 (88%)	210 (100%)

Table 4.5: Final assessment matrix results table containing the scores for the technical properties for the three highest scoring transducers before technical specs, their resulting total score, and the percentage of the total possible score.

4.2 Protocol compliance detection algorithm

The results of the protocol compliance algorithm are described in the following section. The performance of the pleural threshold and skin contact algorithms is tested separately.

4.2.1 Pleural threshold performance test

The process described in method section 3.2.5 is followed to generate the results for the protocol compliance algorithm with the selected thresholds. After analyzing all frames of the 30 test videos and comparing the output with the expert labels, the numbers of TP, TN, FP and FN were found and are shown in the confusion matrix in figure 4.1 From the values in the confusion matrix the following sensitivity, specificity, and sensitivity were found for the designed algorithm.

- $Sensitivity = 100\%$

- $Specificity = 60\%$

- $Accuracy = 87\%$
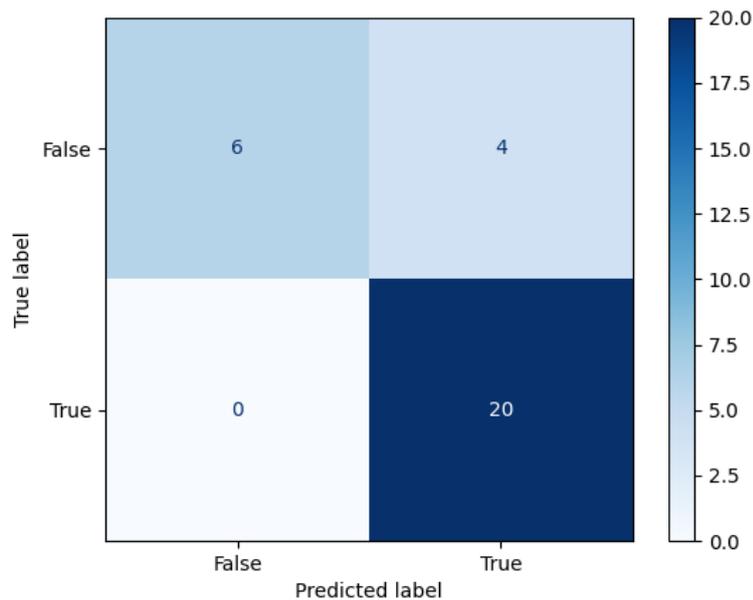
Figure 4.1: Confusion matrix of the results of the pleural threshold algorithm with a threshold of 0.3

4.2.2 Skin contact threshold performance test

The results of the pleural threshold performance test in figure 4.1 show that the algorithm gives six times the output of an insufficiently performed scan. For these instances, the skin contact algorithm will analyze whether the contact with the skin was sufficient. The results of the algorithm compared to the ground truth are shown in figure 4.2. Using the results shown in figure 4.2 the following sensitivity, specificity, and sensitivity were found for the test data set with a skin contact threshold of 0.25.

- $Sensitivity = 100\%$

- $Specificity = 100\%$

- $Accuracy = 100\%$
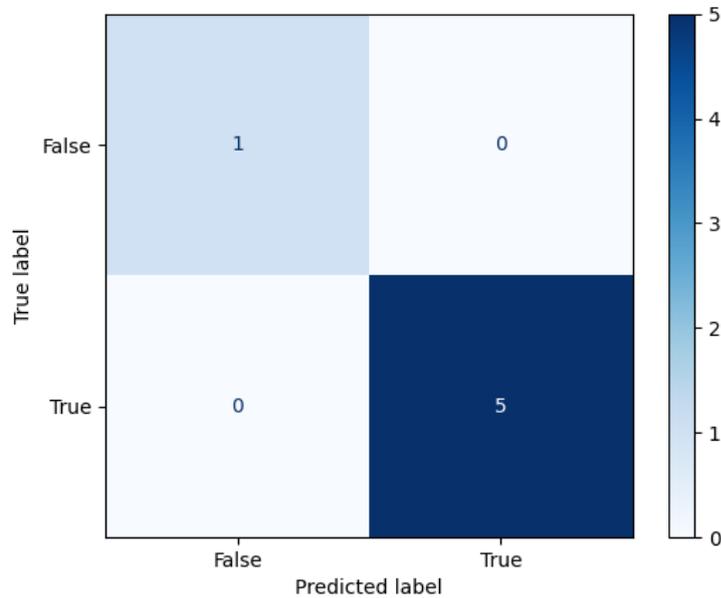
Figure 4.2: Confusion matrix of the results of the skin contact algorithm

4.2.3 Speed test

The resulting data from the speed test of the complete protocol compliance algorithm is shown in figure 4.3 & 4.4. It is shown in figure 4.3 that the specificity increases with the decreasing number of frames analyzed. Sensitivity and accuracy decrease as the number of frames analyzed decreases. For a frameDevide of 4 (total frames analyzed = 61), it is shown that sensitivity, specificity and accuracy remain equal compared to the analysis of all frames for frameDevide=0 (total frames analyzed = 245). In figure 4.4 it is shown that the processing time for frameDevide of 4 is around 3 seconds. This is three times less than the processing time for frameDevide=0.

Figure 4.3: The performance of the pleural line detection expressed in sensitivity, specificity, and accuracy for multiple frameDivide values
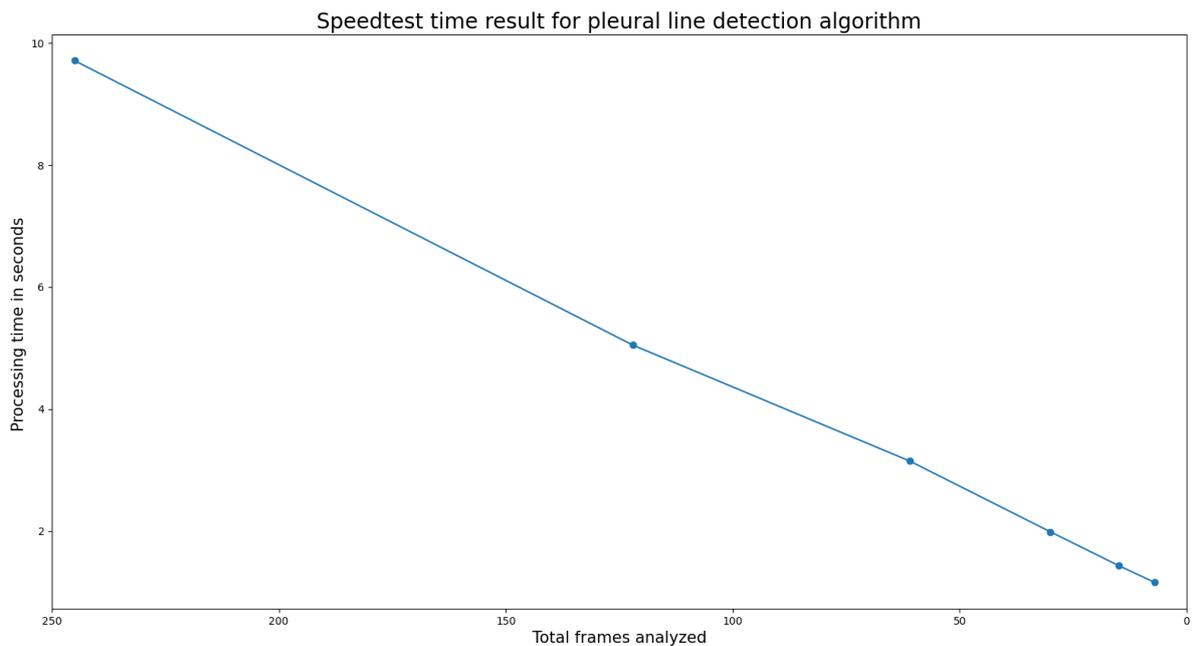
Figure 4.4: The processing time of the protocol compliance algorithm for multiple frameDivide values

Discussion

The discussion is divided into two parts to provide a better overview of the two parts of the project. However, some points of discussion apply to the entire project, which are stated below.

During the thesis, time management was of great importance due to the accessibility of materials for the studies. Transducer selection had to be completed after three months and the selected transducer would be shipped to Zambia to be used in the clinical study. Therefore, there was no access to a transducer to record videos for the design of the algorithm. To solve this problem, the algorithm was first designed using videos recorded with the Telemed C60S transducer and later adapted to Clarius C3 videos. Two months before the thesis deadline, a new Clarius C3 transducer arrived at the company. This transducer could only be used for 2 weeks because it was shipped to Zambia. During the two weeks, the transducer was used to determine and execute the experiments for the threshold determinations in the algorithm. This resulted in a basic and quick method to determine the intensity of the pleural line intensity threshold and the skin contact percentage threshold.

Furthermore, there was no lab present so improvisation was needed to perform experiments like the B-line phantom experiment. Basic materials were purchased to build the phantom in a short period of time. For the comparison between phantom and reality, colleagues were used as test persons after giving their consent.
For the determination of thresholds in the algorithm, only one test person was used due to the limited amount of time. To achieve better thresholds, more data for the threshold determination is needed from multiple test persons.

5.1   Transducer selection

The Clarius C3 transducer scored the highest total score for the assessment matrix as shown in table 4.5 and was therefore used as the probe for the collection of data for the protocol compliance algorithm. The highest total score means that the Clarius C3 is the most suitable candidate for the transducer. The second highest scoring transducer is the Telemed C60S with one point difference. The biggest difference between the two high-scoring probes is the way of connecting to the phone, battery life, lateral resolution, and penetration depth. The final score indicates that there is not a big difference in performance between the two highest-scoring transducers. However, the visual quality of the Clarius C3 recordings seemed to be better.
The commercial probes all include their own presets and a way of post-processing that influences image quality. Certain settings, such as line density, frame rate, and frame averaging, could not be turned off or adjusted for the Clarius C3 transducer. These settings affect image quality. This was seen in the resolution values, but it was not possible to express this difference in other technical property values. This played a role in the selection of the Clarius C3 transducer. It could be beneficial to include a measure of temporal resolution that expresses the resolution over time. However, this is already somewhat included by including the frame rate. Furthermore, a

47

measurement of the contrast-to-noise ratio at the required depth could be included, to obtain another image quality parameter. Including more quality assessment parameters is beneficial for an objective US image quality assessment. These could not be found in the literature and require further research.

In the results section in table 4.4 it is shown that the axial resolution for all transducers is smaller than the required axial resolution. Therefore, all transducers score a five in the assessment matrix for this property. The lateral resolution for all transducers was sufficient, but the Clarius C3 transducer can make a better distinction between small structures at a depth of around $37mm$ than the other transducers. The penetration depth is a dividing property between the transducers. Both the Clarius C3 and Telemed C60 score above average and have a penetration depth that allows for an image with a sufficient signal at a depth of $80mm$. However, the Clarius C7 has a penetration depth of $44.5mm$ this is not deep enough because a penetration depth greater than the depth of the pleural line is needed. In patient data from the clinical study, it was seen that the depth of the pleura during a posterior scan line was around $30mm$. For a good indication of abnormalities, such as the difference between the Z lines and the B lines, a penetration depth of at least $40mm$ posterior to the pleural line is necessary [25]. This results in a penetration depth of $70mm$ that the Clarius C7 does not possess.

If manufacturer properties for the assessment matrix could not be found, an average score of three was used. This was done because a missing property does not indicate that the probe does not have that property, but that it has not been stated by a manufacturer. By applying a score of three, the probe would not be excluded based on a missing property.

The weights assigned to the properties were assigned after conversations and feedback moments with the company for which the transducer was selected. A list of important properties was made, and weights were assigned. The weights for the properties were sent to the company supervisor for feedback and, after a discussion, a final list of weights was assembled. However, it should be stated that this process is prone to be subjective because it is difficult to express importance in numbers. This affects the reliability of the transducer selection method.

During the phantom experiment, the phantom meat was spoiled and not at body temperature. This influences the acoustic properties of the meat and thus the phantom. For further use of the designed phantom, the meat should be at body temperature when scanned. Furthermore, the presets for the probes were not the same because not all frequencies, focus points, and imaging depths could be selected for all probes. Therefore, the setting that resulted in the most ideal visual representation was chosen for each probe and used for all phantom measurements with that probe.
Measurements with the Sonohealth linear were excluded because the recordings stopped before the entire phantom scan line was scanned. Therefore, the score for the curvilinear side of this transducer was used as the score for the Sonohealth transducer.

For the QA test results, it should be stated that the properties found only apply to the settings used for the recording of the data. If the focus point or center frequency is shifted, the results will be different. Furthermore, the recording settings between the probes were not the same because it was not possible to change the frequency for all probes.

5.2 Protocol compliance detection algorithm

The sensitivity for the test data set of the designed pleural threshold algorithm with the threshold 0.3 is 100%, which means that the algorithm is able to select sufficient LUS scans recorded with the used protocol as useful scans for screening. However, a specificity of 60% shows that the algorithm will select more than half of the insufficient scans, but not all, this requires some improvement. The specificity could be improved slightly by changing the threshold, but this would greatly decrease the sensitivity and accuracy. Furthermore, this would not be in line with the threshold determination method using a training data set. The accuracy (87%) of the pleural threshold algorithm with a threshold of 0.3 is a good first result but can be further improved by improving the specificity. The accuracy shows that shadow peak-based segmentation in combination with intensity thresholding is a useful method to check whether a LUS scan line is performed correctly.

The speed test results shown in figure 4.3 show that the algorithm performs the same when $\frac{1}{4}$ of the total frames are used for analysis. This divides the processing time by a factor of $\frac{1}{3}$ as shown in figure 4.4. Furthermore, figure 4.4 shows that there is an almost linear relationship between the amount of frames analyzed and the processing time for the algorithm. The goal is to use the algorithm to guide the examiner during the recording of the scan lines. By reducing the processing time with a $frameDevide$ equal to 4, the examiner gets feedback on his performance with high sensitivity, specificity, and accuracy in a time of about 3 seconds. This is beneficial for the use of the algorithm. Processing time could be reduced even further, but this would affect the overall performance of the algorithm. As a result of a decrease in sensitivity and accuracy.

The sensitivity, specificity, and accuracy of the skin contact algorithm are 100%. This means that the algorithm labeled the data with the same labels as the ground truth and worked perfectly for the test data set. This skin contact algorithm is a great addition to the pleural threshold algorithm as part of the total protocol compliance algorithm. This algorithm provides additional feedback to the examiner if skin contact was the limiting factor in the scan procedure.

In some cases, the bone mask contains white structures that represent other non-target structures in the original image. This affected the accuracy of the algorithm. The problem was solved by using the pleural line mask to identify a correct or incorrect recording and not taking the bone mask into account.
The shadow peak method for the segmentation of the pleural line worked for the purpose of the designed algorithm. However, the results of the shadow peak method did not always segment the pleural line. If the recording contained well-defined A or B lines, the algorithm would also select these structures. This does not affect the ability of the method to determine whether the scan line was performed correctly, because these structures are also signs of a useful recording [3].

In some frames, the algorithm selected structures with high intensity at the borders between different tissues anterior to the pleura. This could be solved by setting a minimum depth for the pleural line in the algorithm. However, this can introduce some difficulty for patients with a large amount of fat tissue. A possible solution to this difficulty could be to implement a minimum depth threshold that depends on the circumference of the thorax. To determine this dependency, more research is required.
Furthermore, scan lines that can contain part of the heart, abdominal organs, or the diaphragm could introduce some high-intensity structures that were sometimes considered pleural structures by the algorithm. A solution to this problem could be the usage of the lung preset of the US probe because lung presets contain minimal image enhancement methods. This hypothesis needs further research. For this research, all thresholds must be defined with a new data set

containing only data recorded with the lung preset.

To improve the thresholds and get a more reliable validation of the designed pleural threshold algorithm larger test and training data sets should be used. The new data sets must be labeled by an LUS expert to achieve a more reliable validation. For the current training data set, labels were assigned by someone with some experience in LUS but not by an expert. The expert should label all data with a ruling on whether the recording is sufficient or insufficient and make a second ruling on whether the skin contact is sufficient or insufficient. The second ruling is necessary for a larger and more reliable data set for the determination of the skin contact threshold and the validation of the skin contact algorithm. Currently, the skin contact threshold was determined with a training data set of nine scan line recordings and the algorithm was validated with a test data set containing six videos. Therefore, a larger data set should be tested to determine the threshold and validate the skin contact algorithm better.

The training and test data sets contained video files with a depth setting between 10-14cm. This choice was made due to the greater presence of dark spaces in images recorded with a greater depth setting and the more abundant presence of white depth dotes in the US cone. These structures introduced difficulty in the backscan conversion and the shadow peak-based segmentation. The second problem can be overcome by a preprocessing step to get rid of these structures and use, for example, nearest-neighbor averaging to fill in the empty spaces in the frame. The algorithm can be used for deeper images that contain more dark spaces because of a lack of signal at greater depth. However, new thresholds need to be determined for these depths. A standard preset is recommended for data collection.

The test data set contained multiple doubt cases in which the expert had difficulty distinguishing between a sufficient or insufficient scan. When in doubt, the scans were labeled as bad scans. This influenced the specificity of the algorithm because the training data set did not contain more than one of these doubt cases. Furthermore, the training data set was not labeled by an expert, but the test data set was. For better agreement between the data sets, the same labeling method should be used for both data sets.

5.3 Recommendations

For the design of a computer-assisted screening tool for LUS, a large data set of TB-positive and TB-negative LUS scans must be collected. To ensure that the data is gathered in a standardized and good way, it is recommended to use a large tablet in combination with the commercial probe. In this way, the user can see where he is imaging within the body. The same preset should be used for all recordings to ensure equal data gathering. However, the focus point should be set to the pleural line for the best image quality for abnormality detection.

For the design of a deep learning algorithm for the detection of TB using LUS, it should be taken into account that the abnormalities or abnormal tissue structures found are not only TB-specific, but can also be signs of other lung diseases [3].

QA tests were only performed for the three highest scoring transducers after all other properties in the assessment matrix were graded. For a more complete overview of all commercial probes tested, it would be beneficial to also perform the QA results for the excluded probes.

The designed algorithm is easy to adapt using the threshold determination methods described in this thesis. The hypothesis is that the algorithm can be used for multiple probes because it is based on the physical properties of LUS. This needs to be investigated by collecting a labeled dataset with the protocol described in figure 2.5 with multiple transducers. It is important that the scans for a transducer are recorded with the same preset because presets change the intensity values in a US frame.

The performance of the algorithm could also be improved by using a more random data set selection representing the normal situation. The current data sets were selected to represent a data set that contains doubt cases and good scans. The reason for this was to see whether the algorithm could distinguish the truly hard cases from each other. However, during the selection of the data sets it was seen that a more representative larger data set would contain a large group of correctly recorded scans, some doubt cases, and a number of really bad scans. The latter was not included in the data sets. Therefore, the validation of the algorithm must be done with a data set that represents the real situation.

In addition to the shadow peaks-based method that was used for segmentation, there exist other line segmentation methods. A radon transform-based algorithm was researched and showed promise in locating the pleural line and B-lines [26]. This location could be used to segment the pleural line. This requires further research. A comparison between the method in this thesis and the radon-based method could be made for better insight.
In this thesis, other possible segmentation methods have not been tested, and no research in the literature on these segmentation methods for LUS was found in addition to some deep learning classification algorithms. For further research, a comparison could be made between some other segmentation methods to achieve the best segmentation. Possible segmentation methods could be Canny edge detection [43], [44] or the Hough line transform [45]. However, these methods are based on line or edge detection, and not on the physical properties of the LUS image. Therefore, these methods need a very strong preprocessing step to be applicable to multiple transducers

The backscan conversion on the image data results in some averaging of the data. It would be better to use the output radio frequency data instead of the video files of the probe and apply the designed algorithm to a rectangular grid constructed directly from the line data. In that case, less pixel data is added or adjusted due to interpolation. However, the choice was made not to use RF data because not all commercial US transducer suppliers give raw RF data as a possible output of the device. By using video files, the algorithm is easier to adapt to other transducers.

Conclusion

The most suitable transducer based on a combination of qualitative and quantitative criteria, of the transducers tested for data collection for the design of a protocol compliance and deep learning algorithm for the detection of TB with LUS is the Clarius C3 HD3 multipurpose transducer. The Clarius C3 transducer has good subpleural lateral and axial resolution and scores sufficient on the usability and manufacturer specifications fronts that are necessary for the use of the transducer by untrained users in rural environments.

The designed protocol compliance algorithm is able to select sufficient recorded videos of scan lines with a sensitivity of 100% from a data set containing 30 videos of scan lines. The algorithm made a correct judgment for 87% of the input scan lines with the greatest error in the selection of insufficient scans, resulting in a specificity of 60%. The performance of the algorithm is equal for processing $\frac{1}{4}$ of the total frames, resulting in a decrease in processing speed by a factor $\frac{1}{3}$. The shadow peak-based algorithm shows promise, but should be reevaluated using a larger and more representative data set for training and testing to achieve a better judgment on its performance.

Bibliography

[1] Jerry L. Prince and Jonathan M. Links, *Medical imaging signals and systems*, Pearson Education, Upper Saddle River, NJ, 2015, second edition edition, (2015).

[2] Rohit Bhoil, Ajay Ahluwalia, Rajesh Chopra, Mukesh Surya, and Sabina Bhoil, Signs and lines in lung ultrasound, *Journal of Ultrasonography*, 8 2021, 21(86):e225.

[3] Thomas J. Marini, Deborah J. Rubens, Yu T. Zhao, Justin Weis, Timothy P. O'connor, William H. Novak, and Katherine A. Kaproth-Joslin, Lung Ultrasound: The Essentials, *Radiology: Cardiothoracic Imaging*, 4 2021, 3(2).

[4] Matthew Fentress, Phillip Ezibon, Akuot Bulabek, Carla Schwanfelder, David Schrift, Sachita Shah, James Tsung, and Adi Nadimpalli, A Lung Ultrasound Scanning Technique for Children and Adults in Low-Resource Settings: Preliminary Experiences in Sub-Saharan Africa, *The American Journal of Tropical Medicine and Hygiene*, 11 2021, 105(5):1148.

[5] A Miller Mbchb and Frca Fficm, Practical approach to lung ultrasound, *BJA Education*, 2 2016, 16(2):39–45.

[6] E. Brogi, L. Gargani, E. Bignami, F. Barbariol, A. Marra, F. Forfori, and L. Vetrugno, Thoracic ultrasound for pleural effusion in the intensive care unit: A narrative review from diagnosis to treatment, *Critical Care*, 12 2017, 21(1).

[7] Daniel A. Lichtenstein and Gilbert A. Mezière, Relevance of lung ultrasound in the diagnosis of acute respiratory failure the BLUE protocol, *Chest*, (2008), 134(1):117–125.

[8] Gert Weijers, Johan M. Thijssen, Alexander Starke, Alois Haudum, Kathrin Herzog, Jürgen Rehage, and Chris L. De Korte, Computer-aided ultrasound diagnosis of hepatic steatosis, *IFMBE Proceedings*, (2008), 22:843–847.

[9] Matthew Fentress, Patricia C. Henwood, Priya Maharaj, Mohammed Mitha, Dilshaad Khan, Philip Caligiuri, Aaron S. Karat, Stephen Olivier, Anita Edwards, Dirhona Ramjit, Nokwanda Ngcobo, Emily B. Wong, and Alison D. Grant, High sensitivity of ultrasound for the diagnosis of tuberculosis in adults in South Africa: A proof-of-concept study, *PLOS Global Public Health*, 10 2022, 2(10):e0000800.

[10] WHO, Global Tuberculosis Report 2022, (2022).

[11] K. C. See, V. Ong, S. H. Wong, R. Leanda, J. Santos, J. Taculod, J. Phua, and C. M. Teoh, Lung ultrasound training: curriculum implementation and learning trajectory among respiratory therapists, *Intensive Care Medicine*, 1 2016, 42(1):63–71.

[12] Erik J. Hasenoehrl, Dannah Rae Sajorda, Linda Berney-Meyer, Samantha Johnson, Jo Ann M. Tufariello, Tobias Fuhrer, Gregory M. Cook, William R. Jacobs, and Michael Berney, Derailing the aspartate pathway of Mycobacterium tuberculosis to eradicate persistent infection, *Nature Communications*, 12 2019, 10(1).

[13] Michelle B. Ryndak and Suman Laal, Mycobacterium tuberculosis Primary Infection and Dissemination: A Critical Role for Alveolar Epithelial Cells, *Frontiers in Cellular and Infection Microbiology*, 8 2019, 9:299.

[14] Smitha J. Sasindran and Jordi B. Torrelles, Mycobacterium Tuberculosis Infection and Inflammation: what is Beneficial for the Host and for the Bacterium?, *Frontiers in Microbiology*, (2011), 2(JAN).

[15] Sara Yukie Rodriguez-Takeuchi, Martin Eduardo Renjifo, and Francisco José Medina, Extrapulmonary tuberculosis: Pathophysiology and imaging findings, *Radiographics*, 11 2019, 39(7):2023–2037.

[16] Delphine Natali, Georges Cloatre, Christian Brosset, Pierre Verdalle, Alain Fauvy, Jean Pierre Massart, Quy Vo Van, Nelly Gerard, Claudia C. Dobler, and Philippe Hovette, What pulmonologists need to know about extrapulmonary tuberculosis, *Breathe*, 12 2020, 16(4):1–18.

[17] Ruvandhi R. Nathavitharana, Alberto L. Garcia-Basteiro, Morten Ruhwald, Frank Cobelens, and Grant Theron, Reimagining the status quo: How close are we to rapid sputum-free tuberculosis diagnostics for all?, *eBioMedicine*, 4 2022, 78:103939.

[18] Ashu Seith Bhalla, Ankur Goyal, Randeep Guleria, and Arun Kumar Gupta, Chest tuberculosis: Radiological review and imaging recommendations, *The Indian Journal of Radiology & Imaging*, 8 2015, 25(3):213.

[19] Giovanni Volpicelli, Point-of-Care Lung Ultrasound, *http://dx.doi.org/10.1024/1661-8157/a001690*, 6 2014, 103(12):711–716.

[20] S. Kulkarni, B. Down, and S. Jha, Point-of-care lung ultrasound in intensive care during the COVID-19 pandemic, *Clinical Radiology*, 9 2020, 75(9):1–710.

[21] Tom Heller, Claudia Wallrauch, Sam Goblirsch, and Enrico Brunetti, Focused assessment with sonography for HIV-associated tuberculosis (FASH): a short protocol and a pictorial review, *Critical Ultrasound Journal*, (2012), 4(1):21.

[22] Nicole M. Duggan, Andrew J. Goldsmith, Ahad Alhassan Al Saud, Irene W.Y. Ma, Hamid Shokoohi, and Andrew S. Liteplo, Optimizing Lung Ultrasound: The Effect of Depth, Gain and Focal Position on Sonographic B-Lines, *Ultrasound in Medicine and Biology*, 8 2022, 48(8):1509–1517.

[23] Francis Chun, Yue Lee, Christian Jenssen, Christoph F Dietrich, and Med Klinik, A common misunderstanding in lung ultrasound: the comet tail artefact, *Med Ultrason*, (2018), 20(3):379–384.

[24] Gino Soldati, Marcello Demi, Andrea Smargiassi, Riccardo Inchingolo, and Libertario Demi, The role of ultrasound lung artifacts in the diagnosis of respiratory diseases, *https://doi.org/10.1080/17476348.2019.1565997*, 2 2019, 13(2):163–172.

[25] Daniel A. Lichtenstein, Gilbert A. Mezière, Jean François Lagoueyte, Philippe Biderman, Ivan Goldstein, and Agnès Gepner, A-lines and B-lines: Lung ultrasound as a bedside tool for predicting pulmonary artery occlusion pressure in the critically ill, *Chest*, 10 2009, 136(4):1014–1020.

[26] Nantheera Anantrasirichai, Wesley Hayes, Marco Allinovi, David Bull, and Alin Achim, Line Detection as an Inverse Problem: Application to Lung Ultrasound Imaging, *IEEE Transactions on Medical Imaging*, 10 2017, 36(10):2045–2056.

[27] L. Beth Gadkowski and Jason E. Stout, Cavitary Pulmonary Disease, *Clinical Microbiology Reviews*, 4 2008, 21(2):305.

[28] M. Fentress, P. C. Henwood, P. Maharaj, M. Mitha, D. Khan, R. Jackpersad, R. Pitcher, A. Redfern, E. Lopez Varela, M. M. van der Zalm, E. B. Wong, M. Palmer, and A. D. Grant, Thoracic ultrasound for TB diagnosis in adults and children, *Public Health Action*, 3 2022, 12(1):3.

[29] Gaetano Rea, Marco Sperandeo, Roberta Lieto, Marialuisa Bocchino, Carla Maria Irene Quarato, Beatrice Feragalli, Tullio Valente, Giulia Scioscia, Ernesto Giuffreda, Maria Pia Foschino Barbaro, and Donato Lacedonia, Chest Imaging in the Diagnosis and Management of Pulmonary Tuberculosis: The Complementary Role of Thoraci Ultrasound, *Frontiers in Medicine*, 12 2021, 8:753821.

[30] Leonardo Carrer, Elena Donini, Daniele Marinelli, Massimo Zanetti, Federico Mento, Elena Torri, Andrea Smargiassi, Riccardo Inchingolo, Gino Soldati, Libertario Demi, Francesca Bovolo, and Lorenzo Bruzzone, Automatic Pleural Line Extraction and COVID-19 Scoring From Lung Ultrasound Data, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 11 2020, 67(11):2207–2217.

[31] M. Sperandeo, P. Filabozzi, A. Varriale, V. Carnevale, M. L. Piattelli, G. Sperandeo, E. Brunetti, and M. Decuzzi, Role of thoracic ultrasound in the assessment of pleural and pulmonary diseases, *Journal of Ultrasound*, 6 2008, 11(2):39.

[32] Zhihua Liu, Lei Tong, Long Chen, Zheheng Jiang, Feixiang Zhou, Qianni Zhang, Xiangrong Zhang, Yaochu Jin, and Huiyu Zhou, Deep learning based brain tumor segmentation: a survey, *Complex and Intelligent Systems*, 2 2023, 9(1):1001–1026.

[33] Wenyu Xing, Guannan Li, Chao He, Qiming Huang, Xulei Cui, Qingli Li, Wenfang Li, Jiangang Chen, and Dean Ta, Automatic detection of A-line in lung ultrasound images using deep learning and image processing, *Medical Physics*, 1 2022.

[34] Marco La Salvia, Gianmarco Secco, Emanuele Torti, Giordana Florimbi, Luca Guido, Paolo Lago, Francesco Salinaro, Stefano Perlini, and Francesco Leporati, Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification, *Computers in Biology and Medicine*, 9 2021, 136:104742.

[35] Hamideh Kerdegari, Phung Tran Huy Nhat, Angela McBride, Reza Razavi, Nguyen Van Hao, Louise Thwaites, Sophie Yacoub, and Alberto Gomez, Automatic Detection of B-lines in Lung Ultrasound Videos From Severe Dengue Patients, *Proceedings - International Symposium on Biomedical Imaging*, 2 2021, 2021-April:989–993.

[36] Sourabh Kulhare, Xinliang Zheng, Courosh Mehanian, Cynthia Gregory, Meihua Zhu, Kenton Gregory, Hua Xie, James McAndrew Jones, and Benjamin Wilson, Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2018), 11042 LNCS:65–73.

[37] R ; Moshavegh, K L Hansen, H ; Moller-Sorensen, M B Nielsen, and J A Jensen, Automatic Detection of B-lines in In-Vivo Lung Ultrasound, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, (2019), 66(2):309–317.

[38] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib, Deep Learning for Medical Image Processing: Overview, Challenges and the Future, In: , (2018), 323–350.

[39] Toru Kameda, Naohisa Kamiyama, and Nobuyuki Taniguchi, The Mechanisms Underlying Vertical Artifacts in Lung Ultrasound and Their Proper Utilization for the Evaluation of Cardiogenic Pulmonary Edema, *Diagnostics 2022, Vol. 12, Page 252*, 1 2022, 12(2):252.

[40] Toru Kameda, Naohisa Kamiyama, Hideo Kobayashi, Yuko Kanayama, and Nobuyuki Taniguchi, Ultrasonic B-Line-Like Artifacts Generated with Simple Experimental Models Provide Clues to Solve Key Issues in B-Lines, *Ultrasound in medicine & biology*, 7 2019, 45(7):1617–1626.

[41] Johan M. Thijssen, Gert Weijers, and Chris L. de Korte, Objective performance testing and quality assurance of medical ultrasound equipment, *Ultrasound in Medicine and Biology*, 3 2007, 33(3):460–471.

[42] Prashant Pandey, Pierre Guy, Antony J Hodgson, and Rafeef Abugharbieh, Fast and automatic bone segmentation and registration of 3D ultrasound to CT for the full pelvic anatomy: a comparative study, *International Journal of Computer Assisted Radiology and Surgery*, (2018), 13:1515–1524.

[43] Hum Yan Chai, Lai Khin Wee, and Eko Supriyanto, Edge Detection in Ultrasound Images Using Speckle Reducing Anisotropic Diffusion in Canny Edge Detector Framework.

[44] Marina Nikolic, Eva Tuba, and Milan Tuba, Edge detection in medical ultrasound images using adjusted Canny edge detection algorithm, *24th Telecommunications Forum, TELFOR 2016*, 1 2017.

[45] Payam S. Rahmdel, Richard Comley, Daming Shi, and Siobhan McElduff, A review of hough transform and line segment detection approaches, *VISAPP 2015 - 10th International Conference on Computer Vision Theory and Applications; VISIGRAPP, Proceedings*, (2015), 1:411–418.

Appendix

A.1 Dutch abstract

Tuberculose (TB) is een van de meest voorkomende ziektes ter wereld. De meeste TB-infecties bevinden zich in derdewereldlanden, waar de diagnose van TB moeilijk is vanwege beperkte toegang tot gezondheidszorg, financiële middelen en infrastructuur. Om vroege diagnose mogelijk te maken in deze gebieden, zou een kosteneffectief, eenvoudig te gebruiken point-of-care (POCUS) apparaat ideaal zijn. Een gestandaardiseerd long ultrasound point-of-care apparaat voor geautomatiseerde TB-screening zou het aantal gedetecteerde TB-gevallen in de betreffende gebieden verhogen. Voor dit apparaat kan een gestandaardiseerd, gemakkelijk aan te leren scanprotocol worden gecombineerd met computerondersteunde detectie, zodat uiteindelijk de getrainde ziekenhuis medewerker kan worden uitgesloten van het diagnostische proces.

Deze scriptie zet de eerste stappen in de richting van het ontwerp van een kosteneffectieve gestandaardiseerde computerondersteunde long ultrasound tool (LUS) voor de screening van TB. Dit wordt gedaan door verschillende commerciële betaalbare POCUS-transducers te evalueren op hun bruikbaarheid en vermogen om TB-gerelateerde longsymptomen te detecteren via een beoordelingsmatrix en een B-lijn fantoomexperiment. De best presterende transducer wordt gebruikt voor een klinische studie naar de detectie van tuberculose met behulp van longechografie en een ontworpen standaard scanprocedure.

Bruikbare scanlijnen voor diagnose in LUS zijn sterk afhankelijk van de gebruiker en worden beïnvloed door de scanhoek ten opzichte van de pleurale lijn en het juiste contact tussen de huid en de transducer. Daarom wordt in het tweede deel van deze scriptie de klinische studiedata gebruikt om een protocolcompliance-algoritme te ontwerpen dat kan controleren of een scanlijn correct wordt uitgevoerd. Dit algoritme is ontworpen om een onervaren gebruiker tijdens de scanprocedure te begeleiden om ervoor te zorgen dat de data op de juiste manier wordt verzameld. De met behulp van dit protocolcompliance-protocol verzamelde data kan worden gebruikt om een deep learning algoritme te trainen voor computerondersteunde screening van tuberculose.

De meest ideale transducer voor een kosteneffectief point-of-care screeningsinstrument voor de detectie van tuberculose met LUS, uit de geteste transducers, is de Clarius C3 HD3 multipurpose transducer. Het ontworpen protocolcompliance-algoritme kan goed opgenomen video's van scanlijnen selecteren met een gevoeligheid van 100% uit een dataset met 30 video's van scanlijnen. Het algoritme maakte een juiste beoordeling voor 87% van de geanalyseerde scanlijnen, met de grootste fout in de selectie van fout opgenomen scans als gevolg van een specificiteit van 60%. Het algoritme geeft veel belovende resultaten maar moet opnieuw worden geëvalueerd en gekalibreerd met behulp van een grotere, representatievere dataset om een beter oordeel te kunnen geven over de prestaties.

De geselecteerde transducer en het ontworpen protocolcompliance-algoritme zijn de eerste stappen naar een kosteneffectieve gestandaardiseerde computerondersteunde screenings LUS-tool voor de screening van TB.

A.2 Specifications stated by manufactures results

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Telemed MicrUs Pro-L40S Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear & linear	MSLPU79 transducer Curvilinear	Weight
Price	1	1	3	3	3	5	5	2
Certification	5	5	5	5	5	1	1	1
Robustness	5	5	5	5	5	3	3	3
Field of view	5	3	5	2	4	3	5	2
Battery life	3	3	5	5	5	4	4	3
Charging time	3	3	5	5	2	5	5	2
Frame rate	5	5	3	3	4	3	3	1
Adjustable parameters	5	5	5	5	1	3	3	1
Preset can be saved	5	5	5	5	1	3	3	2
Local storage	5	5	5	5	1	5	5	1
Cloud storage	5	5	1	1	5	1	1	0
Raw data	5	5	5	5	1	1	1	2
SDK available	5	5	5	5	1	2	3	3
Total manufacturer spec	102	98	110	106	71	74	81	

Table A.1: Results table of the manufacturer specifications containing the scores for the properties for each transducer and the resulting total score computed by multiplying the scores by the weight and taking the sum for all properties per transducer.

A.3 Assessment matrix

Properties	Clarius C3 HD3 multi-purpose	Clarius C7 HD3 micro-convex	Telemed MicrUs Pro-C60S Curvilinear	Telemed MicrUs Pro-L40S Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear & linear	MSLPU79 transducer Curvilinear	Weight
User-friendliness	5	4	4	4	1	2	2	5
B-line detection	4	5	4	2	4	3	2	5
Total before technical specs	147	143	150	136	96	99	101	-
Penetration depth	4	2	5	-	-	-	-	2
axial resolution	5	5	5	-	-	-	-	2
Lateral resolution	5	2	1	-	-	-	-	3
Total score	185 (88%)	171 (81%)	184 (88%)	136	96	99	101	205

Table A.2: Final assessment matrix results table containing the scores for the technical properties for each transducer and the resulting total score computed by multiplying the scores by the weight and adding the sum to the previous total for all transducers.

A.4 User-friendliness survey tester 1

User-friendliness (tester 1)	Clarius C7 HD3 — Micro-Convex	Clarius C3 HD3 — Multi-purpose	MicrUs Pro-C6oS — Curvilinear	MicrUs Pro-L4oS — Linear	Butterfly iQ+ — Multi-purpose	SonoHealth D2CL — Curvilinear linear	MedSingLong probe P5C — Curvilinear
Connection to phone	- Some Android version have difficulties in connecting or take long time to connect. No USB option.	- Some Android version have difficulties in connecting or take long time to connect. No USB option.	+- When using the app, you always need to enable OTG connection (which is not possible on all OS versions or could be easily forgotten by users)	+- When using the app, you always need to enable OTG connection (which I believe is not possible on all OS versions or could be easily forgotten by users)	+ Only USB, direct connection	++ USB and wireless possible, no problems with wireless	++ USB and wireless possible, no problems with wireless
Changeability of imaging parameters	+ Very flexible, adjust and save application, ensure that user does not change all parameters (but gain easily adjusted)	+ Very flexible, adjust and save application, ensure that user does not change all parameters (but gain easily adjusted)	++ Very flexible, possibility to save new preset and hide all settings except for depth, frequency, and focus. Always extra step needed when changing a parameter value.	++ Very flexible, possibility to save new preset and hide all settings except for depth, frequency, and focus. Always extra step needed when changing a parameter value.	+ Only depth, gain, and TGC can manually be changed. Depth and gain are very easily by accidence adjusted. You cannot make your own preset.	- Fixed values for e.g. focus every 2 or 3 cm and deep focus not possible	- Fixed values for e.g. focus every 2 or 3 cm and deep focus not possible
Positioning of the probe	+ Probe size a bit too large in my opinion, but gives good images quite soon.	+ Probe size a bit too large in my opinion, but gives good images quite soon.	+ Good size but cable sometimes anoying.	+ Good size but cable sometimes anoying.	+ Good size but cable sometimes anoying.	++ Very manageable and easy to change between linear / convex sides. However, small scan surface.	+ Very manageable
Recording data	++	++	++	++	++	+ Acquiring image and video easy, only start video immediately when unfreezing which might be confusing.	Acquiring image and video easy, only start video immediately when unfreezing which might be confusing.
Saving data	++ You can choose between DICOM or mp4 instead of downloading them both. RF data possible, otherwise sides of pyramid are cropped without connection to DICOM server.	++ You can choose between DICOM or mp4 instead of downloading them both. RF data possible, otherwise sides of pyramid are cropped without connection to DICOM server.	++ Saves image immediately to storage when pressing save button. However, we already have code to save the RF data and convert to B-mode.	++ Saves image immediately to storage when pressing save button. However, we already have code to save the RF data and convert to B-mode.	+ Only saving to cloud, not to local memory. Crops sides of pyramid	+ dcm and mp4. RF not possible now.	+ dcm and mp4. RF not possible now.
Exporting data	++ You can export all images of one patient at once.	++ You can export all images of one patient at once.	+ Done at the same time as saving?	+ Done at the same time as saving?	- The need to download one by one. Exporting to local memory of device without using the cloud is not possible.	+ extra safety with extra step (save image/video + Export), but takes more time since you need to save every video/image separately. Saving videos takes quite some time.	+ extra safety with extra step (save image/video + Export), but takes more time since you need to save every video/image separately. Saving videos takes quite some time.
Recapitulatory user-friendliness	+	+	+	+	-	-+	-+

Table A.3: Filled in user-friendliness survey of tester 1

A.5 User-friendliness survey tester 2

User-friendliness (tester 2)	Clarius C7 HD3 Micro-Convex	Clarius C3 HD3 Multi-purpose	MicrUs Pro-C6oS Curvilinear	MicrUs Pro-L4oS Linear	Butterfly iQ+ Multi-purpose	SonoHealth D2CL Curvilinear linear	MedSingLong prob P5C Curvilinear
Connection to phone	Connection without wire is preferred however the time to connect is longer for older android versions +	Connection without wire is preferable however the time to connect is longer for older android versions +	OTG storage needs to be turned on with every new connection. connection with wire is okay but does restrain the movement -+	OTG storage needs to be turned on with every new connection. connection with wire is okay but does restrain the movement -+	Connection with wire, goes quick but restrains movement. +	Wireless connection with phone is easy to make ++	Wireless connection with pho is easy to make ++
Changeability of imaging parameters	Settings are easy to adjust and the presets can be saved. ++	Settings are easy to adjust and the presets can be saved. ++	Settings can all be adjusted. however no sliders or fancy options are present for the adjustment. Therefore it takes some time +	Settings can all be adjusted. however no sliders or fancy options are present for the adjustment. Therefore it takes some time +	Only basic settings can be adjusted. not much room to find good settings. -	For the focus only standard positions exist. not all settings can be changed. buttons are small on phone -	For the focus only standard positions exist. not all settin can be changed. buttons ar small on phone -
Positioning of the probe	The contact with the skin can be a bit hard due to the shape of the probe. Furthermore it is quite big -+	The probe has good contact with the skin without to much pressure. However it's quite big +	The contact with the skin is good. wire can restrain the movement +	The contact with the skin is good. wire can restrain the movement +	Easy to use and good contact with scan. However the scan surface is small. -+	Contact with skin is good but on both sides of the probe the scan surface is limited -+	Contact with skin is good bu the probe feels like it is easy break +
Recording data	Recording images and videos is easy and quick ++	Recording images and videos is easy and quick ++	Recording images and videos is easy and quick ++	Recording images and videos is easy and quick ++	Recording images and videos is easy and quick ++	Recording is difficult to understand, the recording sta as soon as you unfreeze and continues to a set number of frames or until freezin -+	Recording is difficult to understand, the recording sta as soon as you unfreeze an continues to a set number c frames or until freezin -+
Saving data	Saving the data is easy and can be done in multiple formats. A patient file can be made ++	Saving the data is easy and can be done in multiple formats. A patient file can be made ++	Saving the data is easy and can be done in multiple formats. However it's hard to add patient information and store headers in the standard application. +	Saving the data is easy and can be done in multiple formats. However it's hard to add patient information and store headers in the standard application. +	Saving data is only possible in the cloud and not directly on the phone. It is possible to add patient information to the data easily. +	Saving the images is possible for multiple recordings but has to happen in two steps. loading the save is time consuming. Furthermore adding patient data to the data is harder and seems to difficult. Multiple formats are possible +	Saving the images is possible multiple recordings but has happen in two steps. loading save is time consuming. Furthermore adding patien data to the data is harder ar seems to difficult. Multiple formats are possible +
Exporting data	Easy, data can be extracted directly from the phone or from the clarius cloud. ++	Easy, data can be extracted directly from the phone or from the clarius cloud. ++	Easy, data can be extracted directly from the phone ++	Easy, data can be extracted directly from the phone ++	Data extraction is very time consuming because each image or video needs to be downloaded separately. Data can only be extracted from cloud. -	Easy, data can be extracted directly from the phone ++	Easy, data can be extractec directly from the phone ++
Recapitulatory user-friendliness	+	++	+	+	-	-+	-+

Table A.4: Filled in user-friendliness survey of tester 2