UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

DEPARTAMENTO DE ESTRUCTURA DE LA MATERIA, FÍSICA TÉRMICA Y
ELECTRÓNICA

TRABAJO DE FIN DE GRADO

Código de TFG: ETE20

Imagen por Ultrasonidos

Ultrasound Imaging

Supervisor/es: Joaquín López Herraiz

Ignacio Ortega Alonso

Grado en Física

Curso académico 2020-2021

Convocatoria de junio

B-line detection in lung ultrasound

Abstract:

Ultrasound imaging is a widely used technique for medical diagnosis. It is based on the emission of high-frequency sound waves and the reception of their returning echoes. Measuring time delays and changes in pressure amplitude of the received waves, the scanned object's structure and acoustical properties can be inferred. Despite their acoustical properties being unsuited for ultrasound imaging, lungs can also be examined with this technique. Certain artifacts known as B-lines which may appear in lung ultrasound are characteristic of pneumonia and similar pathologies such as COVID-19. Therefore, it has now become relevant to find methods able of detecting these artifacts. In this work, using a collection of assessed lung scans a machine learning classifier algorithm was trained to detect B-lines following two approaches. One of them involves extracting specific features from the images that are representative of B-lines, while the other takes as an input the whole image to be classified. The second approach showed better results, while the results of the first one could be improved with changes in the scoring system and better quality data.

Contents

1 Introduction

1.1 Motivation and context

1.1.1 The broad topic: Medical Physics and Artificial Intelligence

Medical Physics consists of every application of physical theories to medicine. Among them are all of the medical imaging techniques that are carried out in radiology units of hospitals in order to diagnose diseases, such as X-rays, computed tomography (CT), magnetic resonance imaging (MRI) or ultrasound (US), as well as positron emission tomography (PET), which belongs to the area of nuclear medicine. Furthermore, there are therapeutical applications too such as radiotherapy, which allows to treat tumors with the aid of ionizing radiation.

Artificial Intelligence (AI) is an umbrella term that englobes all the models and algorithms designed by humans in order to solve certain problems. In recent years, AI and in particular deep learning - a subcategory of the earlier - have shown great performance in several areas, medical physics being one of them [16]. For instance, it has been possible to detect skin cancer [8] and pneumonia from chest X-rays [14] with great accuracy thanks to AI. There is such an interest in these techniques that the number of published articles on these topics in medical physics journals has grown exponentially (see figure 1).
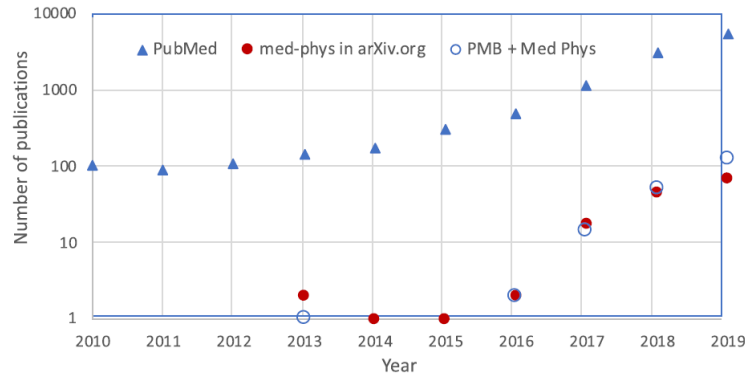

Figure 1: Number of published papers on IA y DL [16]

1.1.2 The specific topic: Lung ultrasound and Covid-19

Recently, lung ultrasound (LUS) has become increasingly relevant due to the current situation of Covid-19 pandemic. Despite CT being the most established imaging technique for pneumonia monitoring and assessment, lung ultrasound advantages make it a promising tool to face this health crisis. First, ultrasound does not use ionizing radiation, which makes it a suitable technique for pregnant women and children. In addition, it can be used in emergency and intensive care units without the need to take patients to the radiology unit, thus reducing the risk of infection and the workload for healthcare workers [1]. Lastly, diagnosis of Covid-19 with lung ultrasound has been proven to be comparable in accuracy to the CT standard [12].

Similarly to other fields of medical physics, lung ultrasound too has the potential for promising applications of AI. In particular, regarding lung ultrasound it is of great importance to correctly identify certain artifacts known as B-lines (see section 1.2.4), whose presence is linked to different diseases, Covid-19 being one of them. An algorithm able to detect such artifacts would have multiple advantages: it would aid in the training of less experienced physicians, it could assist doctors in the diagnosis and it would speed up patient inspection times. Several of this type of algorithms

have already been accomplished, both with the help of neural network classifiers [18] [4] [7] [15] and with more traditional algorithmic approaches [5] [10].

1.2 Physics of ultrasound imaging

1.2.1 Physical principles of ultrasound imaging

Medical ultrasound allows to obtain images of different body tissues thanks to the controlled emission of high-frequency sound waves and the reception of their reflections. Ultrasound probes are equipped with piezoelectric transducers that vibrate when excited with high-voltage electric pulses. This vibration is then transmitted through the body as a pressure wave whose behaviour is influenced by the mechanical properties of the tissues. A fraction of the energy emitted by the probe is reflected on tissue interfaces and received by the transducers. The transducers then vibrate and produce an electric signal that is processed in order to create an image of the examined region.

Sound waves can be understood as periodic variations of pressure on a propagation medium. The rate at which the pressure varies (frequency f) and the spatial distance between neighbouring points with equal pressure (wavelength λ) are related through the speed of sound $c = \lambda f$. The speed of sound c depends on the elastic properties of the propagation medium, and it can be expressed as a function of bulk modulus B and density ρ by:

$$c = \sqrt{\frac{B}{\rho}}$$

In soft tissue, sound travels at an average speed of $1540 \mathrm{m\,s^{-1}}$, but it can be as slow as $600 \mathrm{m\,s^{-1}}$ in the lungs, or as fast as $3200 \mathrm{m\,s^{-1}}$ in bones (see table 1).

Another property of a medium which is related to the speed of sound is acoustic impedance, defined as:

$$Z = \rho c$$

Differences in acoustic impedances of adjacent tissues are responsible for the reflection of sound waves. More exactly, the ratio between the pressure amplitudes of the reflected (P_r) and incident (P_i) wave can be expressed as

$$\frac{P_r}{P_i} = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

where Z_1, Z_2 are the acoustic impedances of the proximal and distal tissues respectively. An intensity reflection coefficient can also be defined as:

$$R_I = \frac{I_r}{I_i} = \left(\frac{P_r}{P_i}\right)^2$$

making use of the fact that intensity is proportional to the squared power of pressure amplitude. For some interfaces like fat-muscle this coefficient is very small, meaning that most of the energy carried by the wave keeps propagating. For others, such as muscle-air almost all the energy is reflected.

When the size of the objects responsible for reflection is similar or smaller than the wavelength, sound is reflected in all directions or scattered. The amount of scattered energy depends on acoustic impedance differences, but also on the texture of the tissue and the wavelength.

Scattering, together with tissue absorption, are responsible for the attenuation or loss of energy of the propagating waves. Different tissues have different attenuation coefficients. They are typically measured in units of dB/cm and are proportional to frequency. In other words, waves with longer wavelength can penetrate deeper into the body. [6]

4

	$\rho(kg \cdot m^{-3})$	$c(m \cdot s^{-1})$	$Z(kg \cdot m^{-2} \cdot s^{-1})$
Air	1.2	330	400
Lung	300	600	$1.8 \cdot 10^5$
Fat	924	1450	$1.3 \cdot 10^6$
Soft tissue	1050	1540	$1.6 \cdot 10^6$
Muscle	1068	1600	$1.7 \cdot 10^6$
Bone	1650	3200	$5.3 \cdot 10^6$

Table 1: Typical values of density, speed of sound and acoustical impedance [2]

1.2.2 Transducers, beam properties

Ultrasound waves are produced by transducers made of piezoelectric materials. Because of the way electric dipoles are arranged in this kind of materials, they vibrate when a voltage difference is applied to their surfaces. Similarly, when an external force makes them contract or expand, a difference in voltage appears between their surfaces. This property allows them to operate both as transmitters and receivers of sound waves. An important property of a transducer is its Q factor, which is defined as:

$$Q = \frac{f_0}{\Delta f}$$

where f_0 is the resonant frequency of the transducer and Δf is its bandwith, that is, the range of frequencies around f_0 where energy is also emitted. These frequencies are introduced by dampening the vibration of the piezoelectric and thus making the emitted pulse shorter. Transducers that are lightly damped show a narrow bandwith (high Q factor) and a long spatial pulse length, whereas heavily damped transducers have broad bandwiths (low Q factor) and short spatial pulse length (figure 2).

(a) Low Q, broad bandwith, short pulse
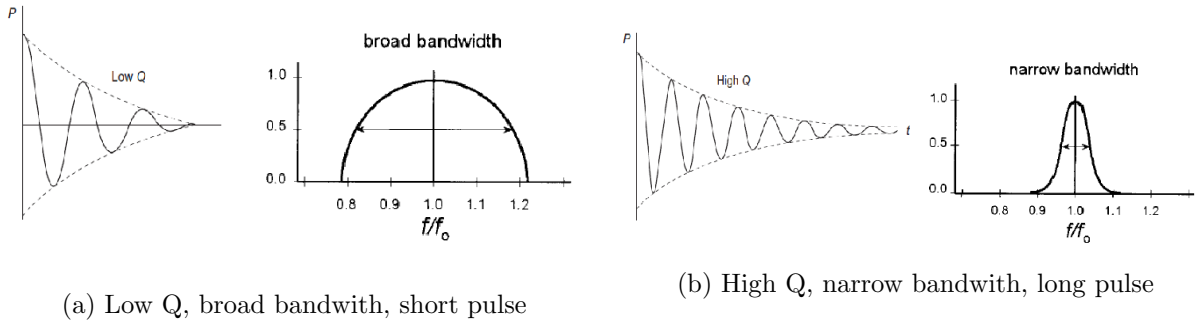
(b) High Q, narrow bandwith, long pulse

Figure 2: Low Q and high Q: spectrum and pulse comparison [6] [2]

Several piezoelectric elements are arranged in transducer arrays and operate together to produce an ultrasound beam. The shape of the beam is the result of interference between every individual emitter. It is characterized by two regions: a near field where the beam converges and a far field where it diverges. The length of the near field (or in other words, the focal distance) is related to the effective diameter of the transducer and its frequency by:

$$l \propto \frac{d^2}{\lambda}$$

That is, wider transducers and higher frequencies produce longer focal distance.

The ability to resolve objects in the acquired image is determined by a series of parameters of the beam and transducers. In the axial direction (direction of the beam), resolution is limited by

(a) A single pulse is emitted against interfaces A and B.

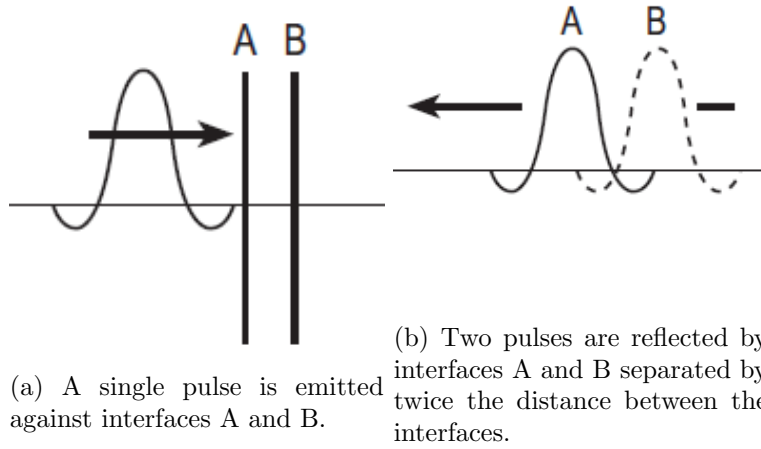(b) Two pulses are reflected by interfaces A and B separated by twice the distance between the interfaces.

Figure 3: Spatial resolution along axial direction is determined by half the spatial pulse length (SPL) [2]

the spatial pulse length (SPL). Two returning echoes can only be resolved if the distance between them is greater than the SPL. Therefore, for two reflectors to be distinguished they cannot be separated by a distance shorter than half the SPL (figure 3). Shorter SPL can be achieved with shorter wavelength or fewer cycles per pulse, i.e. lower Q factor. Typical axial resolution values are of $0.5mm$. Because SPL is constant along the beam, axial resolution is independent of depth. In the perpendicular direction resolution is determined by the diameter of the beam. For this reason lateral resolution is best at the focal zone and it decreases in the near and far fields. Some transducer arrays can produce beams of variable focal distance, creating an extended focal zone where lateral resolution is optimal.

1.2.3   Image formation

In order to generate a two-dimensional ultrasound image several steps have to be followed. First transducers are set to 'transmit mode' to produce sound pulses using a high voltage spike. After the pulse is emitted, they are set to 'receive mode' in order to capture vibrations of returning echoes. Echoes received earlier correspond to reflections taking place close to the transducer, while later echoes are produced by deeper reflectors. The returning echoes are very weak and can only induce small voltages in the transducers that need to be amplified. In addition, later echoes are significantly weaker than earlier ones due to attenuation experienced by the sound wave through-out its round-trip. To compensate for this an additional time-dependent amplification known as time gain compensation (TGC) is introduced. The outcome of this process is a signal containing information of the scanned tissues along the beam direction, where amplitude is proportional to acoustic impedance differences and time corresponds to depth. Assuming an average speed of sound of $1540m/s$ along the beam path, time and depth are related through

$$D = \frac{ct}{2}$$

where the 2 factor accounts for the round-trip distance. After a certain time interval, transducers are set to emit the next pulse and the whole process is repeated several times per second with different focus parameters to achieve a two-dimensional image.

1.2.4 Lung Ultrasound and artifacts

As with any other imaging device, ultrasound cannot reproduce anatomy with unlimited accuracy. Artifacts are deviations from the actual aspect of the scanned objects due to the intrinsic working principles of the device. Lungs are particularly prone to exhibit artifacts when probed with ultrasound, but valuable information can still be obtained from its examination when the origin of these artifacts is well understood.

A typical lung ultrasound image shows the following features. First, a very bright and approximately horizontal line can be seen in the upper half of the picture, corresponding to the pleura. The pleura is a thin membrane inside the chest cavity where lungs are contained. When probing perpendicular to the ribs, these can be also seen as dark regions above the pleura projecting shadows downwards. Due to the high difference of acoustic impedance between the soft tissue above the pleura and the air-filled lung below it, almost all the energy carried by sound waves is reflected, making observation of deeper tissues impossible. A common artifact that can be observed in the region below the pleural line is A-lines (figure 4.a-b). A-lines appear like equally distanced bright lines parallel to the pleural line and they can be explained as a reverberation artifact. Due to the high reflectivity of the pleura, echoes can bounce back and forth several times between it and the transducer surface. Because of their delay with respect to the emitted pulse, these echoes are interpreted as multiple objects located further away from the transducer [3].

Another reverberation artifact that can be commonly observed in lung ultrasound is B-lines (figure 4.c). Whereas A-lines are usually present in healthy lungs, B-lines are correlated with interstitial syndrome, a kind of disease where inflammation or acummulation of fluid makes thin tissues within the lung called interstitium to thicken. This can be observed for example in pneumonia patients. While the exact origin of this artifact is not clearly understood, it is thought that it is originated from reverberation between closely located high reflective surfaces, like air-liquid. Because of the proximity between reflectors the resulting echoes are not distinct, but rather appear as a bright and diffuse vertical line stretching from the pleural line downwards. During the current pandemic, lung ultrasound has proved to be a reliable method for diagnostic and monitoring of Covid-19 pneumonia through the assessment of B-lines and similar phenomena.
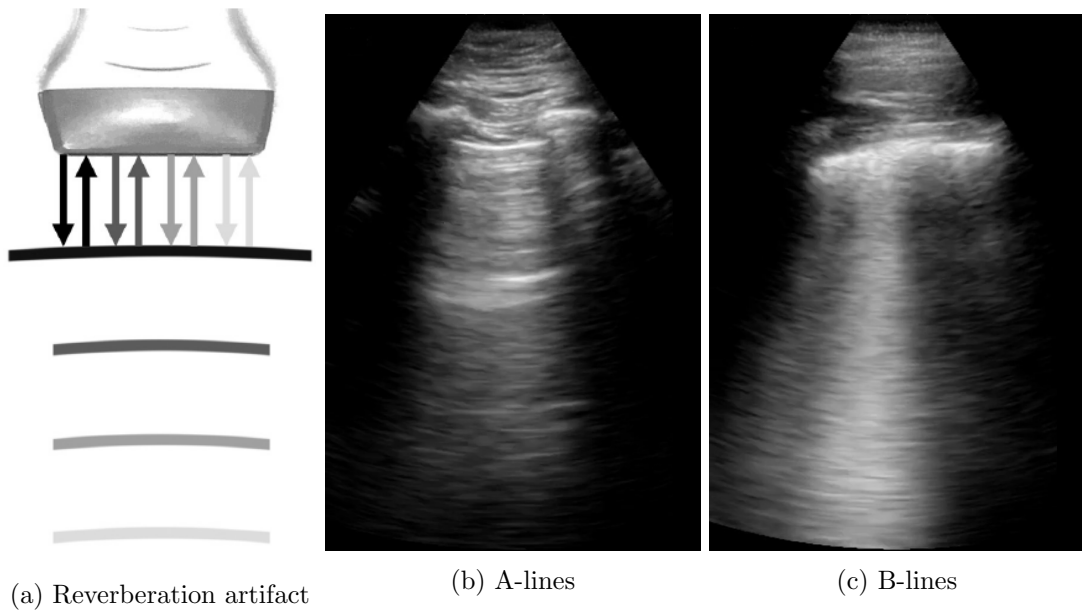

(a) Reverberation artifact   (b) A-lines   (c) B-lines

Figure 4: Common ultrasound artifacts found in lungs

2 Objectives

The present work explores a particular topic from different perspectives and with different goals.

Firstly, it focuses on understanding ultrasound imaging from a theoretical point of view. It aims to identify the physical principles that rule its operation, as well as explaining the image artifacts that are characteristic of this technique. In particular, those artifacts present in lung ultrasound are explored.

Secondly, the current state of research on the topic in question is explored in the scientific literature. Besides considering the broader context of medical physics to which the topic belongs to, emphasis is placed on recent applications of artificial intelligence to lung ultrasound.

Lastly, a practical approach has also been taken on the topic under discussion in order to become familiarised with a programming language and other tools. An analysis of a collection of lung ultrasound videos has been carried out, and its results have been interpreted. The goal has been to design a machine learning model capable of detecting B-lines after being trained on a set of labelled images. This has been done taking two different approaches: first, by extracting those features of the image that are more likely to be correlated with B-line appearance; then, by feeding all the information in the images to the model.

Apart from that, this work also mentions the results of a collaboration with Ultracov project, with the aim to have a first experience in research.

3 Methods

3.1 Software tools

Different tools have been used throughout the project. Python programming language has been chosen to perform the analysis on the videos due to its popularity among the scientific community and the wide offer of libraries available. These are collections of pre-written and documented code designed for different purposes: numpy for matrix and numerical calculations, matplotlib for graphs and visualizations, openCV for image analysis and scikit-learn for machine learning algorithms, among others.

Once we have all the necessary libraries to develope our algorithms, we need an application to edit our code, test it and debug it: this is the role of an integrated development environment (IDE). Both Sypder and PyCharm have been used.

Finally, it is also useful to have some way of organising our code and handling updates. To this purpose, version control applications such as GitHub are of great help. They allow developers to save their source code online and keep track of any changes or updates. It is specially useful when collaborating with other developers.

3.2 Data set

In order to train our model we have made use of a collection of 240 ultrasound videos of lungs. The videos correspond to a single patient throughout a 40 day period of regular examinations monitoring the evolution of Covid-19 pneumonia. They were taken and labelled according to B-lines presence by an experienced physician (Dr. Yale Tung Chen [17] - Hospital Universitario Puerta de Hierro) as part of the ULTRACOV project. All videos were taken with a Butterfly iQ probe set to 13 cm of scan depth. Most of the videos correspond to regions of the chest where lung is the only visible organ, but on some of them other organs such as liver or spleen show up. An image resolution of 1080×632 pixels and framerate of 21 fps is constant across the whole database. Video durations vary widely, ranging from 1 second to 20 seconds of length, with an average of 6 seconds per video.
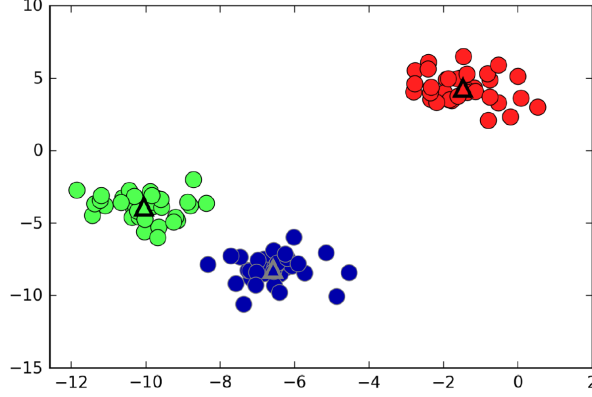
Figure 5: Representation of a K-Means algorithm for a dataset of two features: each cluster (red, green, blue) is represented by its centroid (triangles) [11]

Each video was labelled depending on whether B-lines were present or not at any instant throughout the video. In other words, videos labelled as 'no B-lines' do not exhibit them on any of their frames whereas videos labelled as 'B-lines' do, at least on a fraction of their frames but not necessarily in all of them. This is due to the natural motion of the lungs while breathing, as well as motion of the probe and changes in direction of observation. Because of this, it is necessary to separate videos in smaller sections and isolate those corresponding to B-lines presence. A total number of 85 videos out of 240 are labelled as 'B-lines', which adds up to a 35% frequency of occurrence.

The frames making up each video were extracted on separate directories and cropped to remove unnecessary details for the later analysis. As a result, a collection of gray-scaled images sized 905×632 pixels and 8 bits of depth (i.e., brightness values ranging from 0 to 255) was obtained.

In order to facilitate analysis we need to transform the original set of videos into a set of images. To achieve this, a K-means clustering algorithm implemented in the Scikit-learn toolkit was used [13].

Clustering algorithms are a kind of unsupervised learning algorithms whose aim is to subdivide a dataset into separate groups or clusters in such a way that data points belonging to the same cluster are as close as possible to each other. K-means is one of the simplest of these algorithms. When applied to a two-dimensional dataset the result looks like shown in figure 5 [11]. Each cluster is represented by a cluster center or 'centroid', which corresponds to the average of the data points belonging to that cluster. Each data point is assigned to the cluster whose centroid is the closest. Different configurations are computed one after another until the total sum of distances within clusters is minimized, also known as 'inertia' parameter:

$$I = \sum_{j}^{N} \sum_{x_i \in C_j} ||\vec{x_i} - \vec{\mu_j}||^2$$

where N is the total number of clusters, $\vec{\mu_j}$ is the 'centroid' of the j-th cluster C_j and $\vec{x_i}$ are data points.

When working with images, instead of two dimensions each data point $\vec{x_i}$ belongs to an n-dimensional space with n being equal to the number of pixels in the image ($905 \times 632 = 571960$ in our case). For every video in our original collection we have calculated three clusters and selected the closest frame to each cluster centroid. On those videos labelled as 'B-lines' only the frames that do show B-lines have been manually selected. As a result, we obtained a set of 638 labelled images,
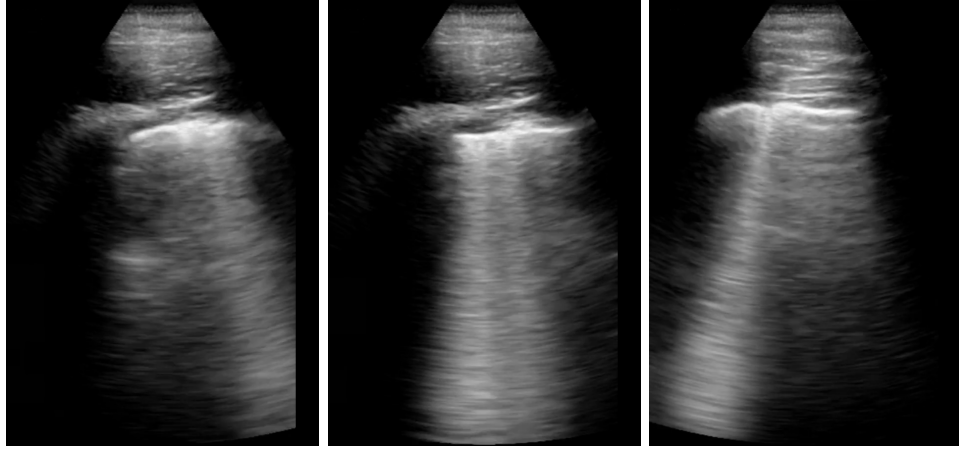
Figure 6: Selected frames representing a video

173 of which correspond to B-lines and the remaining 465 showing no B-lines (27% positives). An example of selected frames for a video through this method can be seen in figure 6.

3.3 Feature based classification

On this method we try obtain a representation of every image based on a few features. These features have been intentionally selected to capture the presence or absence of B-lines on the images, and they require a previous processing step.

3.3.1 Polar format of ultrasound images

Before performing any analysis on the samples this approach requires us to further simplify the data set. To achieve this, we perform a change of coordinates on the ultrasound images in order to obtain images in a rectangular shape equivalent to what the raw data obtained by the ultrasound device would look like. First, we manually measured the aperture angle on the images. Next, we divided the circular sector into several regions of equal radial and angular coordinates (figure 7.a). Finally, we took the mean value of brightness on each region and assigned it to the reshaped image (figure 7.b). By doing so, we obtained a rectangular-shaped image of N-rows by M-columns. In particular, we have chosen $N = M = 50$ (figure 7.c).

This reshaping of the images also facilitates further calculations, as we have greatly reduced the size of the images. It also helps identify B-lines more easily, as they should now appear as vertical lines whereas originally they could take diagonal shapes.

3.3.2 Selected B-line features

We now describe a scoring system used to evaluate the likeliness of an image to show B-lines. Such a scoring system must take into account known features of B-line pattern: bright, uniform vertical shapes extending from the pleural line to the lower margin of the image. Images exhibiting B-lines should generate high scores on this scale, whereas those who show none should produce lower scores. The final score of a video is taken as the maximum score of all of its three centroids. All scorings are performed on the images in polar format. In this work five features have been taken into account:

- **Column mean brightness**: average intensity is calculated for every column on the image. The highest value obtained is chosen as the outcome. All averages are calculated excluding

(a) Sector-shaped image

(b) Square-shaped image

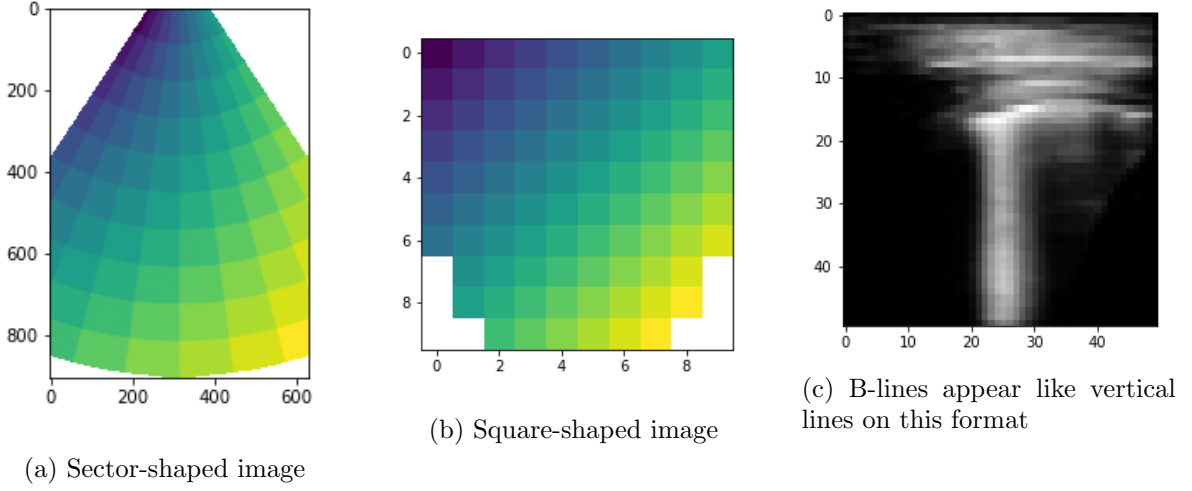(c) B-lines appear like vertical lines on this format

Figure 7: Reshaping of images into polar coordinates

the region above the pleural line, which is estimated to be located at the peak of intensity in that column.

- **Bottom quarter mean brightness**: average intensity is calculated for every column on the quarter section that is further away from the transducer. This should highlight the fact that B-lines extend down to the lowest margin of the image, while other artifacts fade away faster.

- **Bottom value**: intensity values at the lowest point are taken for every column. The maximum value among them is selected.

- **Center of mass**: vertical center of mass coordinates are calculated for every column using pixel brightness as weights. The highest value (lowest center of mass) is selected.

- **Length times mean**: for every column its length excluding the region above the pleura (above peak brightness coordinate) is multiplied by its mean brightness. The highest value among them is selected.

Once scored, every video should yield a set of three values corresponding to each scoring scale. All values are in the range $[0, 1]$ due to the normalization chosen. It is important to note that other scoring systems could be used. The one we present has been chosen because of its simplicity and easy interpretation.

3.4 Image based classification

On this other method, we feed the same algorithm with the raw data of the image. That means, our training data has now as many features as pixels in the image, with the intensity of each pixel being the value of the feature. To speed up calculations, we have resized all images to a square 200×200 shape with no previous processing or change of format.

3.5 Classification algorithm: logistic regression

If our scoring system correlates well with the presence of B-lines, we should be able to train a machine learning algorithm that is able to predict the presence or absence of B-lines in a video
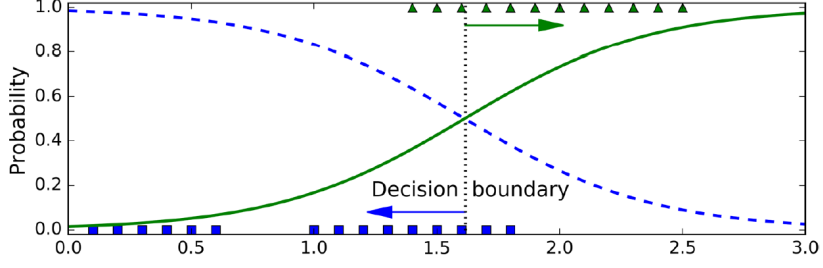
Figure 8: Example of a logistic function fit to a set of data. Values above the decision boundary are assigned to class 1. [9]
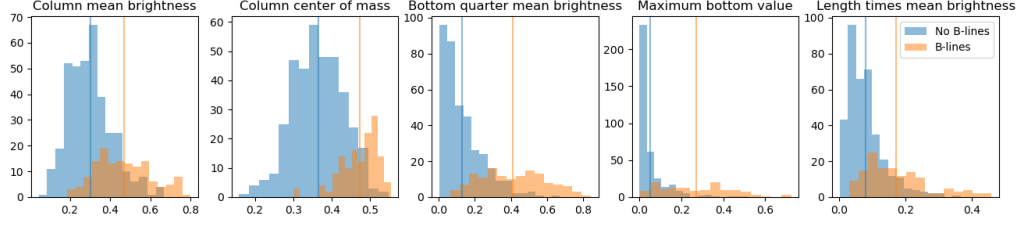

Figure 9: Overview of feature scores

based on its score. We have chosen a logistic regression implementation in Scikit-learn toolkit [13]. Logistic regression is a kind of supervised learning algorithm that assigns a data point to a class where it is most probable to belong to. It takes advantage of logistic functions, also known as sigmoids, by taking a continuous variable as an input and yielding a probability value between 0 and 1 as output.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Probabilities below or above 0.5 correspond to one or the other class. The value of the input variable corresponding to a probability of 0.5 is known as the decision boundary (figure 8). In this case, three values per sample are fed to the algorithm as input, corresponding to each scoring scale, and two different values make up the output, corresponding to the probabilities of presence or absence of B-lines.

4   Results

Before training the logistic regression algorithm on any of both methods, our data set has been split into two different sets. The first one is the training set, i.e. the collection of labelled images or its scores that will be fed to the algorithm. The second one is the test set, on which the algorithm's prediction will be evaluated. Ratios of 0.8 and 0.2 out of the total dataset have been chosen respectively for the training and test sets. In other words, the algorithm is trained on 510 out of the total 638 samples, and it is tested on the remaining 128 samples.

Scores of all videos in the training set have been calculated and its distributions have been represented in figure 9. For each feature the score distributions for positive and negative B-line cases can be seen, and their mean values are indicated by vertical lines. Mean values for B-line cases are higher than those of non-B-line ones for each feature, which indicates that the selected features are indeed correlated with the presence or absence of B-lines on the images. However,
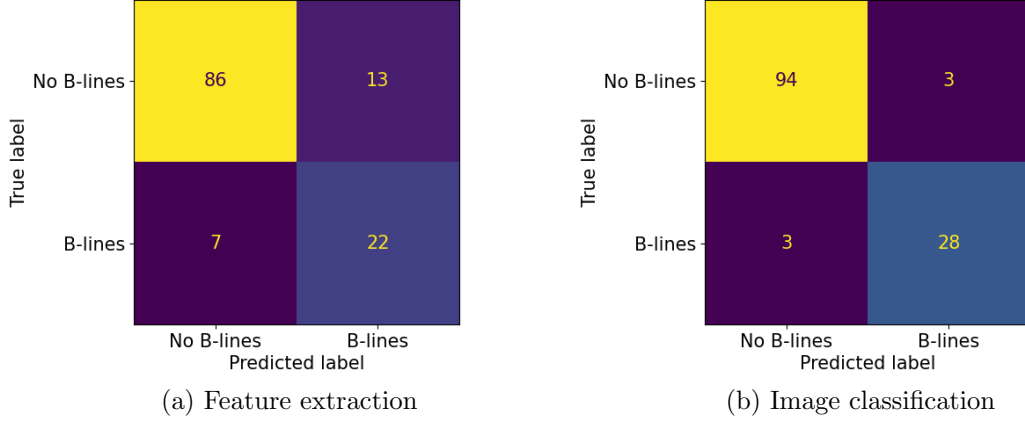
(a) Feature extraction   (b) Image classification

Figure 10: Confusion matrices of both classifiers

there is some overlap between the values scored by postive and negative cases. In other words, in this representation there is not a clear boundary between each class of data points.

A comparison between the algorithm's prediction and the true values is summarized in figure 10 for both methods in what is known as a 'confusion matrix'. Elements in the diagonal of the matrix are correct predictions (true positives or true negatives), while elements out of the diagonal are wrong predictions (false positives or false negatives).

We now make use of different statistical measures to assess how good the classification is. First, accuracy (ACC) is defined as the ratio of correct predictions (both true positives and true negatives) to the total number of cases:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity or 'true positive rate' (TPR) is defined as the ratio of correctly identified positives to the total number of real positives (true positives and false negatives):

$$TPR = \frac{TP}{TP + FN}$$

Conversely, specificity or 'true negative rate' (TNR) is the ratio of correctly identified negatives to the total number of real negatives (true negatives and false positives):

$$TNR = \frac{TN}{TN + FP}$$

Correct predictions can also be normalized against the number of predictions made for each class. In the case of positive predictions we have 'positive predictive value' (PPV), which is defined as the ratio of true positives to all predicted positives (true positives and false positives):

$$PPV = \frac{TP}{TP + FP}$$

On the other hand, 'negative predictive vaue' (NPV) measures the ratio of true negatives to all predicted negatives (true negatives and false negatives):

$$NPV = \frac{TN}{TN + FN}$$

Performance of both methods on all the previously mentioned statistical measures has been summarised in table 2.

	Accuracy	True positive rate	True negative rate	Positive predictive value	Negative predictive value
Features	84,4%	75,9%	86,9%	62,9%	92,5%
Image	95,3%	90,3%	96,9%	90,3%	96,9%

Table 2: Comparison of results: feature extraction and image classification

5 Discussion

Of both methods, the image classification is the one that shows better results across all statistical measures. However, the feature extraction method has some advantages, as it provides a better insight of what are the characteristics that make both classes distinguishable.

The poorer performance of the feature extraction method can be explained by the score distribution of the data set. Due to the overlap in scores between images of different classes, the classifier algorithm cannot determine a clear boundary between classes.

The reason behind the overlap of scores is in the scoring system itself, and it could be overcome by selecting more appropriate features. For example, the 'quarter mean' scale may not always be indicative of B-lines. When organs such as the spleen appear on a video, the brightness on the lower half of the image increases (see figure 11.a) even if there are no B-lines. This could cause a false positive. There is also the case of images that do not show B-lines but have a high average brightness, thus inducing false positives as well (figure 11.b). Therefore, changes on the scoring system could improve the accuracy of the method.

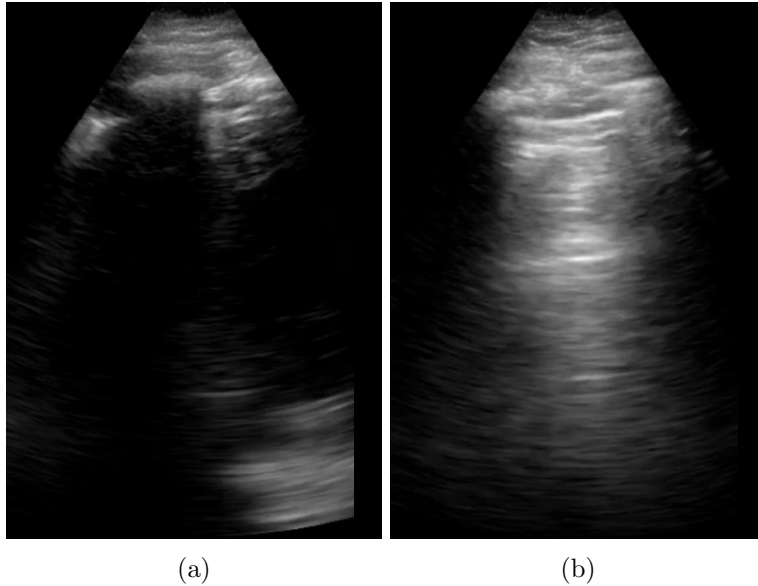(a)   (b)

Figure 11: Potential false positives: (a) presence of organs such as the spleen may be mistaken for B-lines, (b) A-lines may have high average brightness.

A similar approach has been taken by Brattain et al. [5]. In their study, pleural lines were removed from the images before evaluating the presence of B-lines, which might facilitate uniformity across the dataset. The duration of the videos was also shorter, with a constant 60 frames per video. Another approach by van Sloun [18] used deep learning techniques, achieving accuracy of up to 90%.

5.1 ULTRACOV project

As part of this work a collaboration with Ultracov project has also been made. Ultracov project aims to develop an ultrasound device specific for Covid-19 detection. To this purpose, different AI algorithms are being developed. As our contribution to the project, we have gathered these tools and integrated them in a graphical user interface (GUI) using Python and the PySimpleGUI library. The GUI also allows to display videos and manually assign labels to them based on the anomalies that they show, in order to calculate a score of lung involvement.
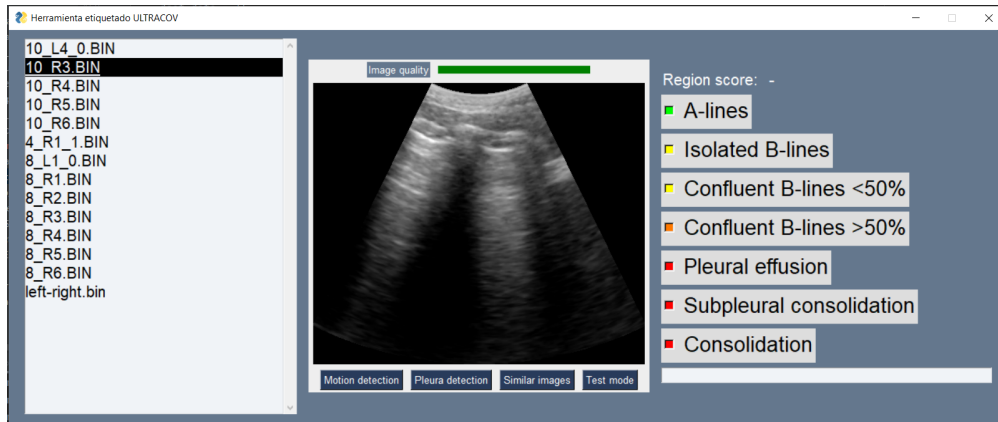

Figure 12: Graphical user interface including analysis tools for Ultracov project.

The following tools have been included:

- Motion detection: it provides a graph of the lateral displacement in pixels of the video through an optical flow algorithm.

- Pleura detection: it shows the predicted location of the pleural line on the current video using a pre-trained neural network.

- Similar images: it displays a selection of the most similar images to the current one from a dataset.

- Image quality: it analyses whether the probe is properly placed based on the pleura detection tool (if no pleura can be seen the quality is not good).

References

[1] Marco Allinovi, Alberto Parise, Martina Giacalone, Andrea Amerio, Marco Delsante, Anna Odone, Andrea Franci, Fabrizio Gigliotti, Silvia Amadasi, Davide Delmonte, et al. Lung ultrasound may support diagnosis and monitoring of covid-19 pneumonia. *Ultrasound in Medicine and Biology*, 46(11):2908–2917, 2020.

[2] Penelope J Allisy-Roberts and Jerry Williams. *Farr's physics for medical imaging*. Elsevier Health Sciences, 2007.

[3] Michael Baad, Zheng Feng Lu, Ingrid Reiser, and David Paushter. Clinical significance of us artifacts. *Radiographics*, 2017.

[4] Jannis Born, Gabriel Brändle, Manuel Cossio, Marion Disdier, Julie Goulet, Jérémie Roulin, and Nina Wiedemann. Pocovid-net: automatic detection of covid-19 from a new lung ultrasound imaging dataset (pocus). *arXiv preprint arXiv:2004.12084*, 2020.

[5] Laura J Brattain, Brian A Telfer, Andrew S Liteplo, and Vicki E Noble. Automated b-line scoring on thoracic sonography. *Journal of Ultrasound in Medicine*, 2013.

[6] Jerrold T Bushberg and John M Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.

[7] B Cristiana, T Grzegorz, K Seungsoo, M Katelyn, L Rachel, M Shaw, RL McNamara, R Balasundar, and CL Moore. Automated lung ultrasound b-line assessment using a deep learning algorithm. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[8] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[9] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[10] Ramin Moshavegh, Kristoffer Lindskov Hansen, Hasse Møller Sørensen, Martin Christian Hemmsen, Caroline Ewertsen, Michael Bachmann Nielsen, and Jørgen Arendt Jensen. Novel automatic detection of pleura and b-lines (comet-tail artifacts) on in vivo lung ultrasound scans. In *Medical Imaging 2016: Ultrasonic Imaging and Tomography*, volume 9790, page 97900K. International Society for Optics and Photonics, 2016.

[11] Andreas C Müller and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.", 2016.

[12] S Ottaviani, M Franc, E Ebstein, L Demaria, C Lheure, MP Debray, A Khalil, B Crestani, R Borie, and P Dieudé. Lung ultrasonography in patients with covid-19: comparison with ct. *Clinical radiology*, 75(11):877–e1, 2020.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011.

[14] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[15] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, et al. Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging*, 39(8):2676–2687, 2020.

[16] Chenyang Shen, Dan Nguyen, Zhiguo Zhou, Steve B Jiang, Bin Dong, and Xun Jia. An introduction to deep learning in medical physics: advantages, potential, and challenges. *Physics in Medicine & Biology*, 65(5):05TR01, 2020.

[17] Yale Tung-Chen. Lung ultrasound in the monitoring of covid-19 infection. *Clinical Medicine*, 20(4):e62, 2020.

[18] Ruud JG van Sloun and Libertario Demi. Localizing b-lines in lung ultrasonography by weakly supervised deep learning, in-vivo results. *IEEE journal of biomedical and health informatics*, 2019.