

Article

Development of a Collaborative Robotic Platform for Autonomous Auscultation

Daniel Lopes ^{1,†} , Luís Coelho ^{1,2,†}  and Manuel F. Silva ^{1,3,*,†} 

¹ Instituto Superior de Engenharia do Porto, Instituto Politécnico do Porto, Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal

² Center for Engineering Innovation and Industrial Technology, Instituto Superior de Engenharia do Porto, Instituto Politécnico do Porto, 4249-015 Porto, Portugal

³ INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

* Correspondence: mss@isep.ipp.pt; Tel.: +351-22-83405-00

† These authors contributed equally to this work.

Abstract: Listening to internal body sounds, or auscultation, is one of the most popular diagnostic techniques in medicine. In addition to being simple, non-invasive, and low-cost, the information it offers, in real time, is essential for clinical decision-making. This process, usually done by a doctor in the presence of the patient, currently presents three challenges: procedure duration, participants' safety, and the patient's privacy. In this article we tackle these by proposing a new autonomous robotic auscultation system. With the patient prepared for the examination, a 3D computer vision sub-system is able to identify the auscultation points and translate them into spatial coordinates. The robotic arm is then responsible for taking the stethoscope surface into contact with the patient's skin surface at the various auscultation points. The proposed solution was evaluated to perform a simulated pulmonary auscultation in six patients (with distinct height, weight, and skin color). The obtained results showed that the vision subsystem was able to correctly identify 100% of the auscultation points, with uncontrolled lighting conditions, and the positioning subsystem was able to accurately position the gripper on the corresponding positions on the human body. Patients reported no discomfort during auscultation using the described automated procedure.

Keywords: autonomous auscultation; MobileNetV2-UNet; healthcare; computer vision; auscultation points; medical robotics; vision robotics



Citation: Lopes, D.; Coelho, L.; Silva, M.F. Development of a Collaborative Robotic Platform for Autonomous Auscultation. *Appl. Sci.* **2023**, *13*, 1604. <https://doi.org/10.3390/app13031604>

Academic Editor: Yutaka Ishibashi

Received: 30 November 2022

Revised: 19 January 2023

Accepted: 21 January 2023

Published: 27 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

First introduced into factories in the 1960s [1], robotic systems have gained a foothold in the manufacturing world, particularly in the automotive industry. More recently, collaborative robots created to work side by side with humans are transforming the world of traditional robotics and production facilities [2,3]. In the medical sector, the introduction of robotics and autonomous technologies is contributing to face the increasing number of patients per health professional and maintaining the existing level of quality in health services. In this context, automation allows, among other possibilities, professionals to carry out easier and safer diagnostic activities, not necessarily implying the presence of, or direct contact between, a doctor and his patient [4–7].

Auscultation has been an integral component of clinical examination since the 19th century, and is a highly cost-effective screening method for the detection of abnormal clinical symptoms [8]. Auscultation will continue to be essential in modern times, as cardiopulmonary disease is one of the largest underlying and direct causes of death worldwide, with a major influence on the quality of life and healthcare expenditure [9]. Additionally, this is a viable diagnostic technique for patients with COVID-19, and can be used as a follow-up tool for non-critical patients [10,11]. Remote auscultation is advantageous in terms of infection prevention, as conventional auscultation requires clinicians to make physical contact with

patients [12,13]. The main challenge in auscultation is that its effectiveness depends on the hearing abilities and expertise of physicians [14].

Electronic stethoscopes have recently emerged as interesting solutions for remote auscultation [15–18]. They allow the visualization of heart and lung sonograms during auscultation, making it simpler to distinguish between various types of heart and lung sounds, as well as a computer-assisted diagnosis in coronary artery disease [19]. Artificial intelligence (AI) is being applied to this equipment to decrease the subjectivity of auscultation [20,21].

Although technology has increased the effectiveness of auscultation, the practice still involves physicians, or healthcare personnel, placing portable stethoscopes on the patient's body, requiring their physical presence. Additionally, during auscultation, they must place the stethoscope with adequate contact forces against the patient's skin, i.e., to ensure patient safety and to remove extraneous noise. Given the shortage of healthcare professionals, particularly in rural areas, where medical resources are scarce [19,22], the creation of a robotic platform that can perform autonomous auscultation with high accuracy and repeatability, without relying on physician intervention, would be useful.

Through a literature search, it was only possible to find two studies already conducted in this area. In the work developed by Tsumura et al. [23], the goal was to develop a robotic auscultation platform, able to estimate auscultation positions and safely place the stethoscope in the estimated position. The estimation method is based on registration with a light detection and ranging (LiDAR) sensor. To autonomously perform the auscultation, the robotic system recognizes the auscultation positions based on the shape of the whole body and places the stethoscope in the determined auscultation positions, considering the positional relationship between the body and the stethoscope. The work developed by Zhu et al. [24], demonstrates, to the best of our knowledge, the first robotic system capable of performing autonomous auscultation of the heart and lungs, based on the TRINA robot. The goal of the robot is to provide diagnostically useful sounds, with high quality, for a physician to hear, rather than to perform the diagnosis itself. First, the system captures a scan of the patient's 3D point cloud, registers a model of the human body, estimates the location of key anatomical landmarks, and produces an initial map of high-quality auscultation locations. It then adopts informative path planning, using auditory feedback to adaptively search the region of interest for high-quality auscultation. The auditory feedback is based on sound quality estimators trained on a database of cardiac and pulmonary stethoscope recordings. To determine the optimal auscultation site, a Bayesian optimization (BO) problem was formulated where the unknown sound quality field was estimated as a semi-parametric residual Gaussian process model with a prior map that depends on a latent translation offset and sound quality scaling parameters. However, this proposed system configuration was based on remote control and did not focus on autonomous auscultation. From these studies, and our literature search, it can be concluded that there is still little work on platforms for autonomous auscultation. It can also be concluded that the robots used in these studies are collaborative robotic arms, due to the possibility of safety interaction with humans. Allied to the robotic arm there is always a sensor/camera to visualize the patient and the location to where the robotic arm should move.

Given the above ideas, this paper presents a robotic system that allows the autonomous placement of a stethoscope based on external body information, while satisfying patient safety in terms of stethoscope surface contact with the body. This platform performs the autonomous auscultation on patients' back, since it is easier to acquire volunteers to provide images of the back than of the chest to create the dataset, differentiating itself from the work performed by Tsumura et al. [23]. The auscultation points shown in Figure 1 were used in the present work, bearing in mind the shoulder blades positions not to auscultate on top of them, because they make auscultation difficult (block the sound transmission). These points are obtained through the knowledge of human anatomy and from reference

points whose location is estimated through a red, green, blue depth (RGB-D) vision system supported on AI, which are the main novelties in this work.

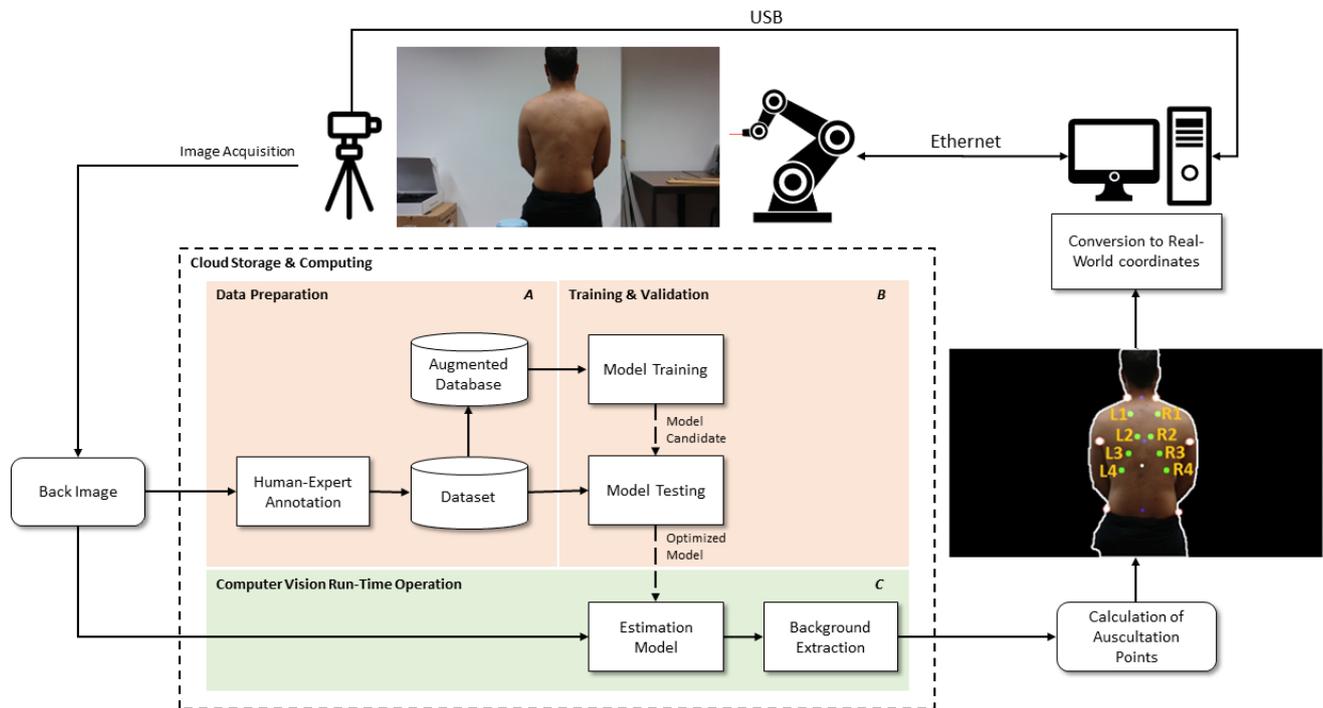


Figure 1. General architecture of the proposed system, showing the most important functional blocks (software), the involved materials (hardware), and their respective flow connections.

2. Materials and Methods

2.1. Materials

The equipment used in the construction of this project is shown in Figure 1, where the general system architecture can be seen.

Concerning the collaborative robotic manipulator, we have used the UR3e device, manufactured by Universal Robots. It weighs 11 kg, has a payload of 3 kg (enough to grab and move a stethoscope), 360° rotation in all joints of the arm, and an infinite rotation in the end joint. Applications of the UR3e robot cover manufacturing industries, from medical devices to circuit boards and electronic components [25,26].

For image acquisition, we have used a RGB-D camera produced by OAK-D. It has three on-board cameras that implement stereo and red, green, and blue (RGB) image acquisition. This camera uses the USB-C cable for power and to send data to a host. This camera is a popular option to develop applications in a wide variety of areas, namely in education, neonatal care, assistive technologies for the disabled, augmented reality/virtual reality (AR/VR), warehouse inspection using drones, robotics, agriculture, sports analytics, retail, and even advertising [27].

Finally, a personal computer was used to interface the UR3e robotic controller with the OAK-D camera and to run the related computer vision algorithms. The supporting hardware was composed of an Intel Core i5-8300H 2.30 GHz with 8 Gb memory and a NVidia GTX 1050 video card. The communication between the computer and the robot controller was established over a wired connection through the Ethernet protocol. The computer and camera communication was established through a universal serial bus (USB) type C cable.

2.2. Methods

Figure 1 shows the system overall architecture. It is possible to verify that there is, first, a data preparation phase (A), consisting of the creation of the dataset and then the

model training (B). Next, autonomous real-time auscultation is performed (C), where first the offset between the camera and robot alignment is obtained. Next, the image of the person's back is obtained from the camera, and this is inserted into the trained model and a model to perform background extraction from the image.

Then, the auscultation points are calculated through the obtained landmarks and their respective 3D spatial coordinates. Finally, through the calculated distances between the 3D spatial coordinates, the robot reaches the auscultation points on the patient's back. The person is without a shirt during this whole process, since in this way the person's back has less detail, making it easier to estimate the key points and to auscultate the desired points.

2.2.1. Model Training

First of all, it was necessary to train a prediction model to predict the landmarks present on the backs of the individuals. U-Net was used for this purpose. It is an architecture developed by Olaf Ronneberger et al., for biomedical image segmentation, in 2015 at the University of Freiburg, Germany [28], and is one of the most widely used approaches currently in any semantic segmentation task. It is a fully convolutional neural network that is designed to learn from little training data. This is an improvement over the existing fully convolutional networks (FCN) developed by Jonathan Long et al., in 2014 [29]. U-Net is a U-shaped network architecture composed of two parts: an encoder network and a decoder network. This model merges high-level features and low-level features through a skip connection between the two parts, which improves the ability to segment details in the image. In this work, the preformed classification network MobileNetV2 [30] was used as the encoder network, forming the MobileNetV2-UNet. The decoder network comprises four deconvolution layers and three inverted residual blocks, so that the input and output dimensions of the MobileNetV2-UNet are identical. The architecture is shown in Figure 2.

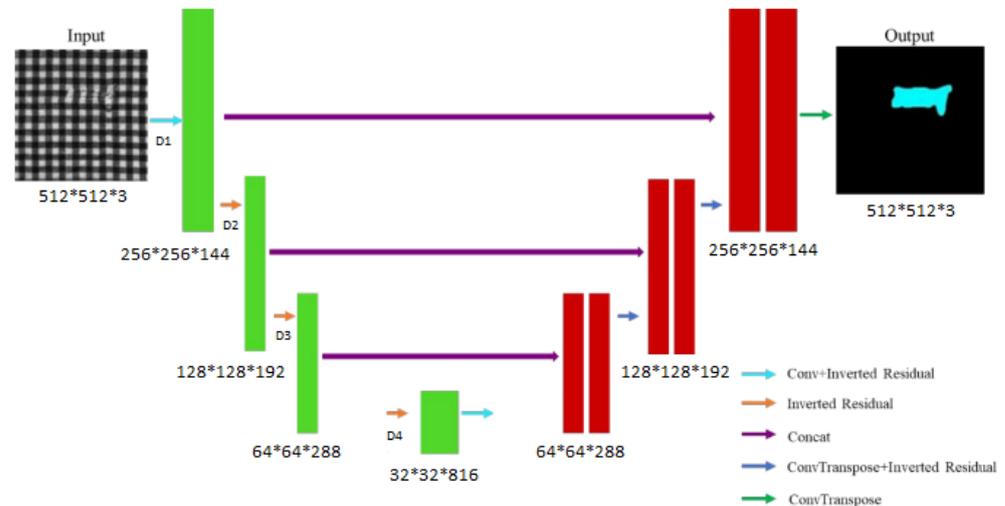


Figure 2. MobileNetV2-UNet architecture (based on [31]).

Some advantages of this network structure are:

- The model has fewer parameters, requiring less computational effort during training and eventually less data;
- Using a pre-trained encoder helps the model converge much faster compared to the non-pre-trained model;
- A pre-trained encoder helps the model achieve high performance compared to a non-pre-trained model.

Once the network was formed, it was necessary to create a dataset composed of a training set and a validation set. These were separated in folders with the back images and their respective masks, both with a size of $512 \times 512 \times 3$ pixels. In Figure 3, it is possible

to observe a small example of a back image with its respective mask, that is, the reference points. The mask was developed manually by an experienced professional and followed current medical recommendations and standards. Since the initial dataset with the back images was made up of only 172 images for training, the data augmentation techniques described in Figure 4 were used. These techniques are useful for improving the performance and results of the machine learning model by forming new and different examples to train datasets. If the dataset of a machine learning model is rich and sufficient, the model performs better and more accurately. Furthermore, for machine learning models, data collection and labeling can be tiresome and costly processes. Transformations on datasets, using data augmentation techniques, allow these operational costs to be reduced [32,33]. In the end, the created dataset had 6536 back images in the training set, where 34 different female and male persons were present, and 24 back images, from different people, in the validation set. In Figure 5 it is possible to observe the distribution of volunteers in terms of gender at the creation of the dataset.

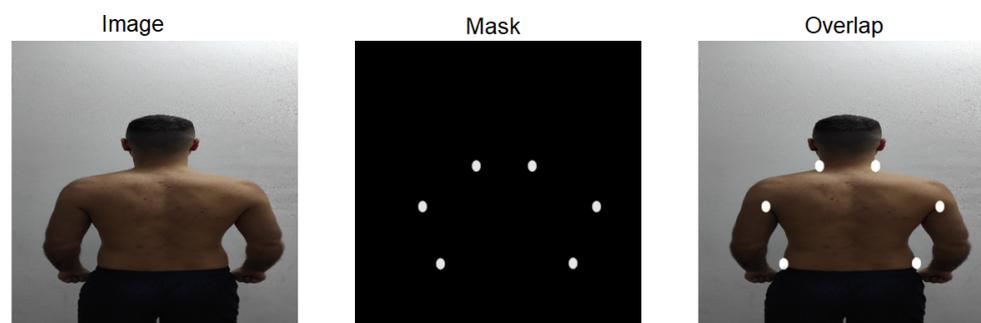


Figure 3. Example of an image pair with the reference points, for training. The image's aspect ratio was adjusted to fit the machine learning model input.

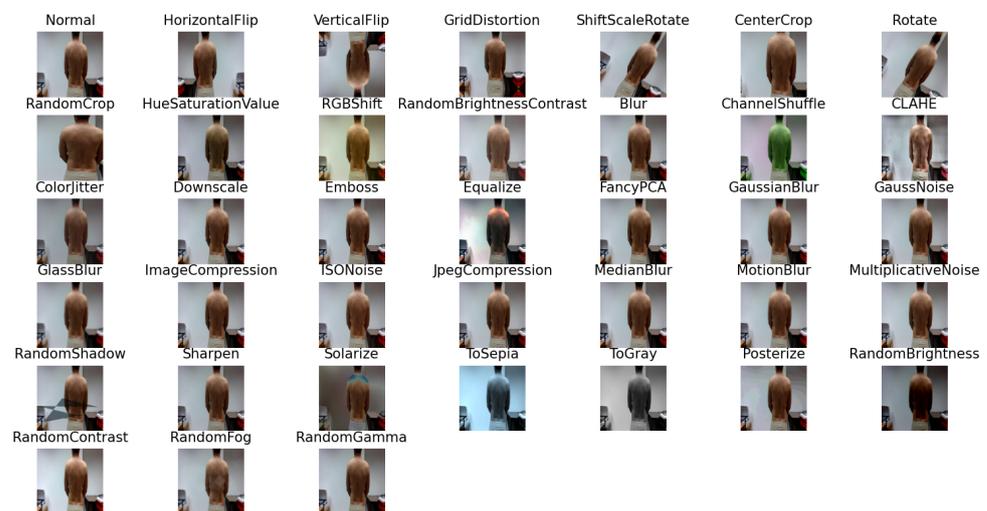


Figure 4. Techniques used for data augmentation.

The Adam optimizer [34] was used to compile the model, with a learning rate of 10^{-4} ; the dice loss [35] was used as the loss function during training and validation; and the dice similarity coefficient (DSC), intersection over union (IoU), recall, and precision metrics were used to analyze the performance of the model over the course of training. To evaluate the obtained results, the metrics accuracy, F1-score, Jaccard, recall, and precision were used because they provide a good performance insight and are widely used by the scientific community.



Figure 5. People present at the dataset creation.

The model was trained over 9 epochs, with a batch size of 1, using the computer GPU. Once the model training was completed, we moved on to the development of the autonomous auscultation algorithm.

2.2.2. Calculating the Offset of the Camera's Alignment with the Robotic Arm

The camera and robot arm must be aligned so that the center of the image obtained from the camera corresponds to the center of the robot's gripper, and the spatial auscultation point (calculated later) corresponds to the distance traveled by the robot arm. Since the camera was not fixed in our setup (to keep the system flexible to the environment conditions), any slightest touch to the camera support causes the camera to move. In this way, the offset of the camera's alignment with the robotic arm is always recalculated.

For this purpose, a socket was created, with the host being the internet protocol (IP) of the robotic arm and port 30002 being used. Via this socket, the robotic arm was moved (with a speed of 2 m/s and an acceleration of 2 m/s²) to the following joint positions, using the `movej()` [36] function: base = 180°; shoulder = −180°; elbow = 90°; wrist 1 = −90°; wrist 2 = 90°; wrist 3 = 245°;

The camera was then placed on the table, so that the red point in the left side of Figure 6A, which corresponds to the image coming from the color camera, approximately coincides with the green point in the center of the green circle, since this green point approximately coincides with the middle of the gripper on the front of the robotic arm. Markings were placed on the table for the placement of the camera support, so that every time there was an idea of where to place it, so that the camera would be approximately centered with the robot [27].

Once the color camera image was taken, it was converted to a hue, saturation, value (HSV) format. The red point is the center of the image. The circles present in the image were determined using the `HoughCircles()` function, given the parameters entered into it. The green point corresponds to the center of the green circle found in the intervals hue [75:109], saturation [63:238] and value [0:255].

To obtain a more precise value of the 2D coordinates of the green point, the values obtained after 250 instances were averaged.

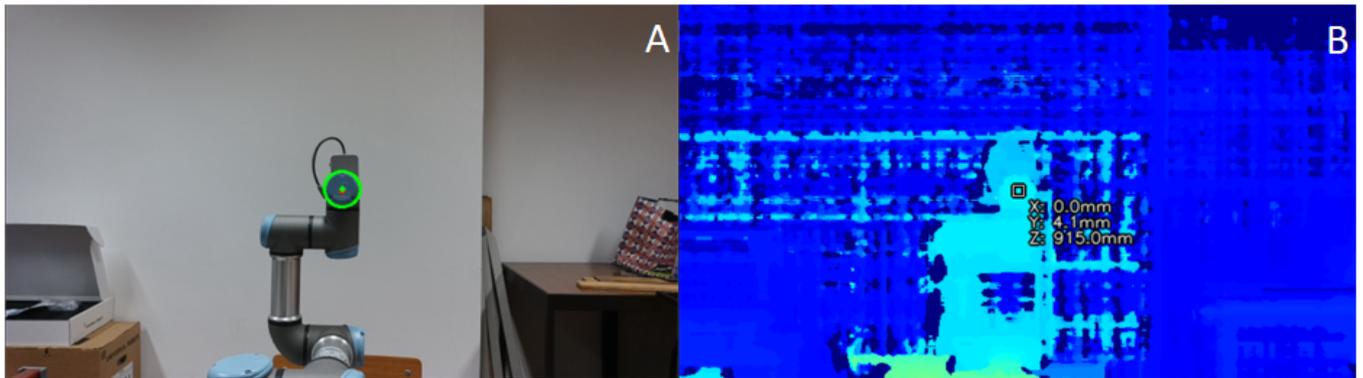


Figure 6. Calculation of the alignment offset between the robotic arm with the camera. On the left (A), photographic image frame, with the middle of the gripper represented by a green dot (center of the green circle). On the right (B), depth image frame with the middle of the gripper (green dot) represented by a square, with related spatial coordinates, near bottom right corner vertex.

Using the 2D coordinates of the green point, their 3D spatial coordinates were calculated, as illustrated on the right side of Figure 6B. These coordinates (x , y , z) are calculated using the `hostSpatialCalc` algorithm available at the Luxonis GitHub repository [37]. This algorithm calculates the spatial coordinates of a region-of-interest (ROI) or point-of-interest (POI) based on the depth image, averaging the depth values in the ROI, and removing values outside the range. To use a POI in this algorithm, it is necessary to create a ROI through this POI that is as small as possible. From this, the function `calc_spatial()` of the algorithm `HostSpatialCalc` was then used, and the x and y coordinates of the POI that is to be calculated were inserted. The value of these coordinates is the offset, because if this green point is in the center of the image, the spatial coordinates x and y will be 0. To obtain a more precise value of the offset, an average of the values obtained after 250 instances was also performed.

2.2.3. Prediction of Reference Points

Once the offset of the camera alignment with the robotic arm was calculated, the reference points were predicted. Using the previously created socket, the robotic arm was moved (with a velocity of 2 m/s and an acceleration of 2 m/s²) to the following joint positions using the `movej()` function [36]: base = 90°; shoulder = 0°; elbow = 0°; wrist 1 = −90°; wrist 2 = 90°; wrist 3:240°;

This movement of the robotic arm has been performed so that the camera can take a frame of the person's back without the arm being in front of them. The person should be approximately in the center of the image.

Taking the frame from the person's back, as illustrated on the left side of Figure 7, it is then resized to 512 × 512 × 3 pixels, since this is the input of the model. Then, the prediction model for image segmentation in humans, already trained and available on the Internet, was used. This model predicts a mask that corresponds to the outline of the person, and with this, it is possible to perform the extraction of the image background.

The mask predicted by this model (contour of the person), presented on the right side of Figure 7, will be used later to calculate the center of this same contour. Next, the keypoints, shown in Figure 8, were predicted using the previously trained model.

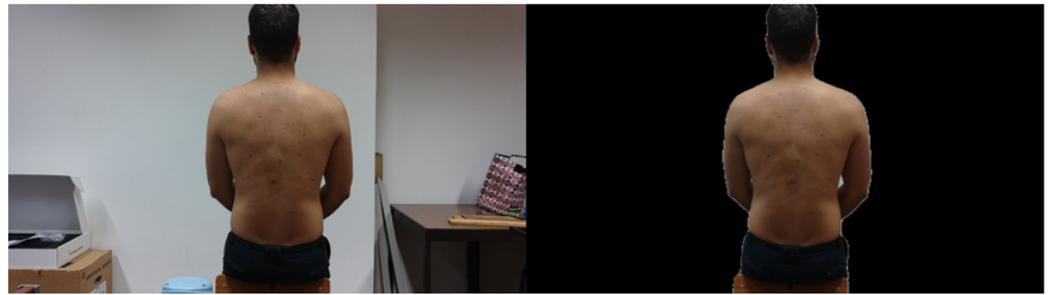


Figure 7. Frame of the person's back with and without background, respectively.

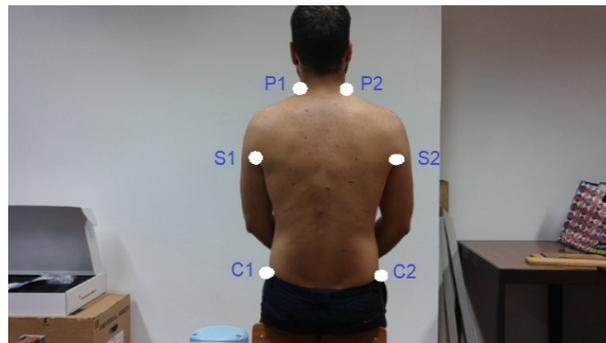


Figure 8. Keypoints on the person's back, represented with white circles. Labels attributed according to: P for neck, S for armpit, C for waist, 1 for left side and 2 for right side.

2.2.4. Calculation of the Points to Be Auscultated

Once the key points on the person's back were obtained, the points to be auscultated are calculated. For this, we used the previewed image with the key points, the mask with the body contour, and the frame of the person's back.

First, the unwanted points that were erroneously predicted in the image with the key points (in other words, the noise) were removed. Morphological transformations were used for this, namely image opening. Image opening is the name given to an erosion followed by dilation transformation. These transformations were performed using the `morphologyEx()` function.

Next, the 2D coordinates of the reference points (key points) of the person's back were obtained using the `SimpleBlobDetector()` function with the following parameters: `filterByColor = True`; `blobColor = 255`; `minArea = 0.1`; `minCircularity = 0.01`; `minConvexity = 0.01`; `minThreshold = 1`; `minInertiaRatio = 0.01`.

In addition, the center of the largest contour present in the mask with the contour of the person was also obtained. This center should correspond to the center of the person's contour.

The key points are stored in different variables, since the `SimpleBlobDetector()` function stores all these points in one variable. They are also separated into right side or left side, considering the previously calculated center. These points are then also sorted considering the value of the y coordinate.

After the keypoints are all sorted and stored, the following computations are performed, depending on the key points which are found (see Figure 9):

- In the axilla area ($S1$ and $S2$):
 - is computed the midpoint between points $S1$ and $S2$ (S_{mid}), and the horizontal length of the back through the x coordinate of these points (difference between these two points) ($dist_x$);
- In the neck area ($P1$, $P2$):
 - is computed the midpoint between points $P1$ and $P2$ (P_{mid});

- In the waist area (C1 and C2):
 - is computed the midpoint between points C1 and C2 (C_mid);
- In the neck and waist area (P1, P2, C1 and C2):
 - is computed the vertical length of the back ($dist_y$) by difference of the y coordinate of the intermediate points calculated earlier (P_mid , C_mid);
- Only at the neck and waist on the left side (P1 and C1):
 - is computed the vertical length of the back using the y coordinate of these points P1 and C1 (the difference between these two points) ($dist_y$);
- Only at the neck on the right side and at the waist on the left side (P2 and C1):
 - is computed the vertical length of the back through the y coordinate of these points P2 and C1 (difference between these two points) ($dist_y$);
- Only at the neck on the left side and at the waist on the right side (P1 and C2):
 - is computed the vertical length of the back through the y coordinate of these points P1 and C2 (difference between these two points) ($dist_y$);
- Only at the neck and waist on the right side (P2 and C2):
 - is computed the vertical length of the back through the y coordinate of these points P2 and C2 (difference between these two points) ($dist_y$);

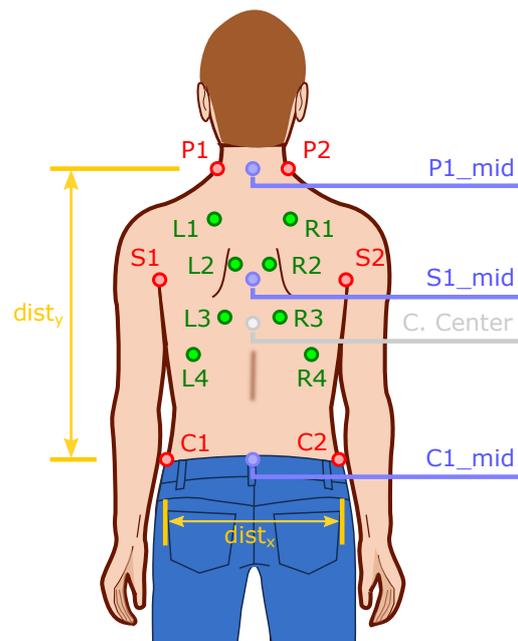


Figure 9. Dorsal mapping of auscultation points (green), reference points (red), and distances.

The algorithm is designed in such a way that it will only proceed to the next step when it has found at least four reference points, two of which must be the points S1 and S2, because that is the only way it can calculate the horizontal length of the back. It must find a point at the neck (P1 or P2) and another one at the bottom of the back (C1 or C2), to be able to calculate the vertical length of the back. The calculation of the auscultation points' 2D coordinates, represented in Figure 9 in green, are based on the measurements of the length of the back (vertical and horizontal). If these are successfully obtained, and depending on the key points found, the following equations apply:

$$L1_x = S_mid_x - (dist_x/2) * 0.30 \quad L1_y = Ref_point_y + dist_y * 0.15 \quad (1)$$

$$R1_x = S_mid_x + (dist_x/2) * 0.30 \quad R1_y = Ref_point_y + dist_y * 0.15 \quad (2)$$

$$L2_x = S_mid_x - (dist_x/2) * 0.15 \quad L2_y = Ref_point_y + dist_y * 0.35 \quad (3)$$

$$R2_x = S_mid_x + (dist_x/2) * 0.15 \quad R2_y = Ref_point_y + dist_y * 0.35 \quad (4)$$

$$L3_x = S_mid_x - (dist_x/2) * 0.35 \quad L3_y = Ref_point_y + dist_y * 0.50 \quad (5)$$

$$R3_x = S_mid_x + (dist_x/2) * 0.35 \quad R3_y = Ref_point_y + dist_y * 0.50 \quad (6)$$

$$L4_x = S_mid_x - (dist_x/2) * 0.50 \quad L4_y = Ref_point_y + dist_y * 0.65 \quad (7)$$

$$R4_x = S_mid_x + (dist_x/2) * 0.50 \quad R4_y = Ref_point_y + dist_y * 0.65 \quad (8)$$

The variable Ref_point_y corresponds to whatever point is found in the neck area, i.e., P_mid_y if both are found, or $P1_y, P2_y$, depending on which one is found. This algorithm is detailed in the flowcharts shown in Figures 10 and 11.



Figure 10. First part of the algorithm for calculating auscultation points' 2D coordinates.

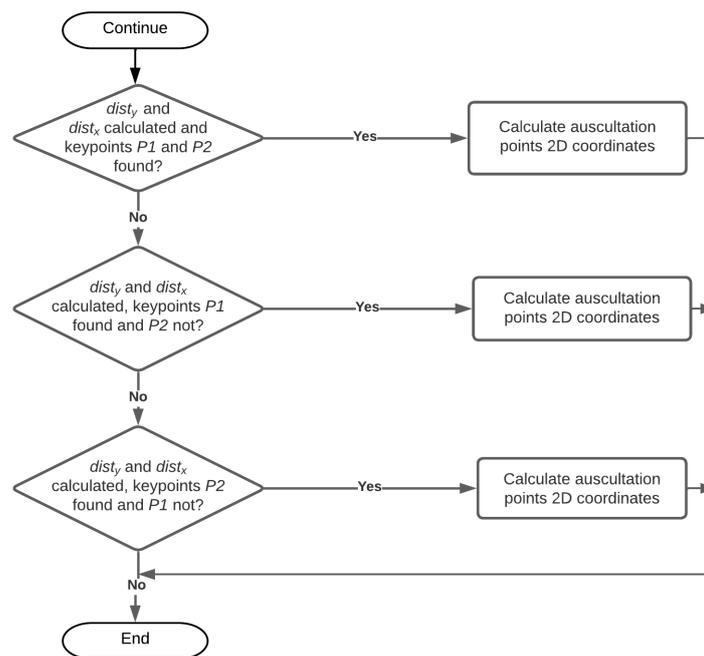


Figure 11. Second part of the algorithm for calculating auscultation points’ 2D coordinates.

2.2.5. Calculation of the Auscultation Points 3D Spatial Coordinates

Having determined the auscultation points’ 2D coordinates, we proceeded to calculate the spatial coordinates of these points. This calculation was performed using the same algorithm mentioned above: the function `calc_spatial()` was used where its input parameters were the depth image and the x and y coordinates of each auscultation point. This function returns the 3D spatial location of each point, calculated earlier. Figure 12 (left) shows a frame of the spatial coordinates of the points to be auscultated.

To obtain more accurate values, during 250 instances when there are values available for each point, the algorithm accumulates their values in the same variables and at the end makes an average. These final values are also converted to meters.

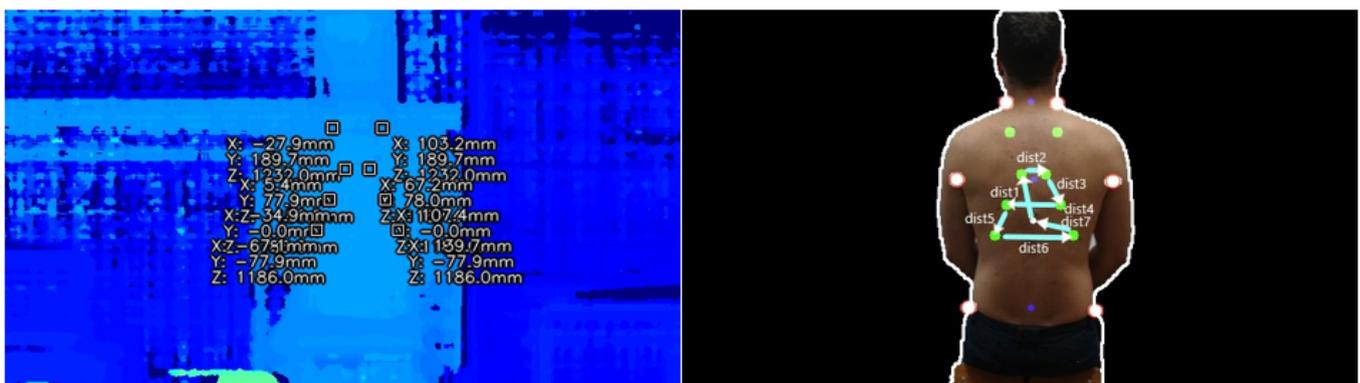


Figure 12. 3D spatial coordinates of the points to be auscultated and the path that should be taken by the robotic arm, respectively.

2.2.6. Performing the Auscultation

Once the spatial coordinates of the points to be auscultated are obtained, the robotic arm is moved to the same joint positions mentioned previously, when the offset was calculated, and with the same speed and acceleration.

After, the distance between the auscultation points is calculated for the robotic arm to perform continuous auscultation without having to return to the starting position. This

calculation started at auscultation point $L2$, because the UR3e robotic arm does not have enough range to reach points $L1$ and $R1$.

To calculate these distances, first, the y coordinate value of the two points at each level was averaged:

$$y1 = (eL1_y + eR1_y) / 2 \quad (9)$$

$$y2 = (eL2_y + eR2_y) / 2 \quad (10)$$

$$y3 = (eL3_y + eR3_y) / 2 \quad (11)$$

$$y4 = (eL4_y + eR4_y) / 2 \quad (12)$$

This average was calculated because the auscultation points on the right and left sides must be at the same y coordinate to always auscultate at the same place, but on the opposite side of the trunk.

Next, the distances are calculated as follows:

$$dist1_x = eL2_x - robot_offset_x \quad dist1_y = y2 - robot_offset_y \quad (13)$$

$$dist2_x = eR2_x - eL2_x \quad dist2_y = 0 \quad (14)$$

$$dist3_x = eR3_x - eR2_x \quad dist3_y = y3 - y2 \quad (15)$$

$$dist4_x = eL3_x - eR3_x \quad dist4_y = 0 \quad (16)$$

$$dist5_x = eL4_x - eL3_x \quad dist5_y = y4 - y3 \quad (17)$$

$$dist6_x = eR4_x - eL4_x \quad dist6_y = 0 \quad (18)$$

$$dist7_x = -eR4_x \quad dist7_y = -y4 \quad (19)$$

where $robot_offset_x$ and $robot_offset_y$ corresponds to the values of the x and y coordinates, of the offset value calculated earlier.

Variables that contain an e at the beginning correspond to the spatial coordinate of the respective point. The last distance corresponds to bringing the robot back to the default position. Figure 12(right) shows the path the robot arm will follow.

To move the robotic arm, taking these distances into account, the function `move()` was used [36]. These movements of the robotic arm between points are performed at a speed of 0.2 m/s and an acceleration of 0.2 m/s².

When the robot reaches an auscultation point, it moves towards the person's back (at a speed of 0.2 m/s and acceleration of 0.2 m/s²), while the force it applies to the tip of the gripper is monitored along the x axis represented in Figure 13. Whenever this force exceeds the force initially received (when the robotic arm reached the auscultation point) by 3 N, it stops for 10 s (simulates auscultation). Next, the robotic arm moves away from the person's back and continues to the next point.

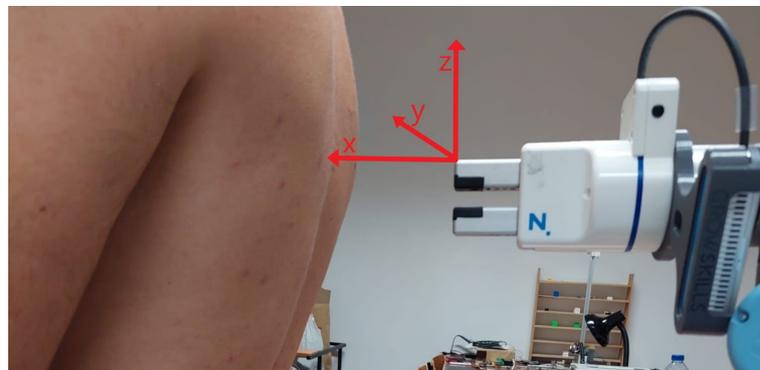


Figure 13. Robotic arm referential.

To obtain the force applied to the tip of the gripper, the real-time data exchange (RTDE) interface is used. This interface provides a way to collect information from the robotic arm, allowing synchronizing variables that can be chosen by the client to be written and

read, which must be contained in the actual synchronization's instruction packet. Each instruction has a unique ID. The RTDE control package setup outputs returns the values of the variables, in the same order in which they were requested. The RTDE was initialized on port 30004 (the host was the IP of the robotic arm) and with a frequency of 125 Hz. Figure 14 presents the flowchart of this algorithm.

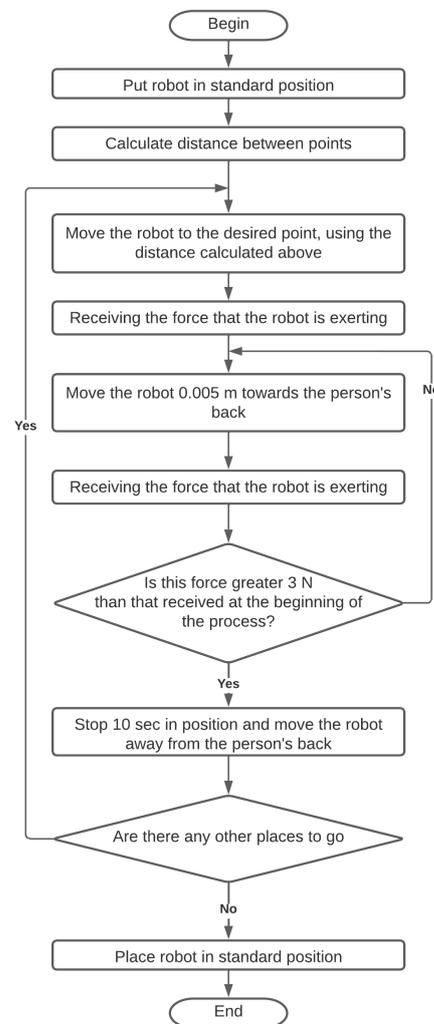


Figure 14. Control algorithm for the robotic arm.

3. Results

To test the platform developed for autonomous auscultation, a simulation of pulmonary auscultation was performed. The patient, during the whole process, was standing against the back of a chair for better stability. These simulations were performed on six patients. This section will be divided into the vision subsystem (responsible for training the prediction model and identifying the auscultation points) and positioning subsystem (responsible for taking the stethoscope surface into contact with the patient's skin surface at the various auscultation points).

3.1. Vision Subsystem

Through Figure 15 it is possible to see the evolution of the model learning over the 9 epochs, and that it obtained the best results in epoch 2, since then the model started over-fitting.

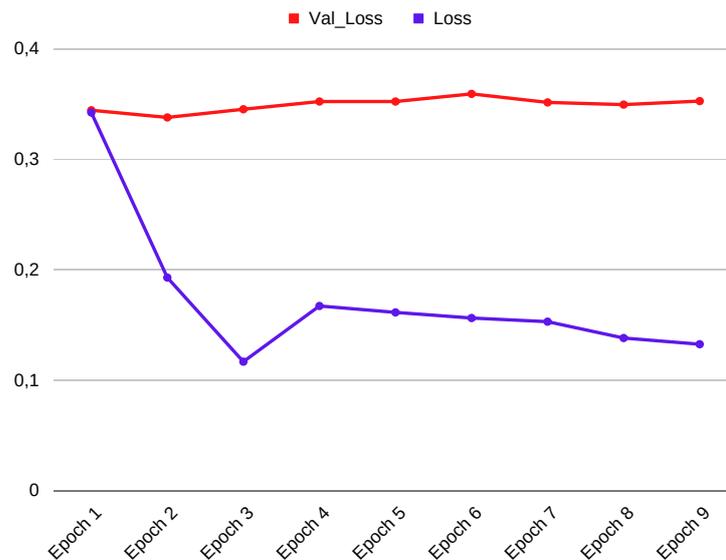


Figure 15. Training evolution over 9 epochs.

From Table 1, it is possible to verify that the model from epoch 2 has achieved good training results, with values of 0.9973 for accuracy, 0.805 for precision, and a loss function value of 0.162 in training. Although the value of the loss function in the validation is 0.338, the model only fails to predict a key point in 2 of the 24 images used for validation, and these 24 images were of different people. These results are very acceptable, since the keypoints do not need to be predicted with high precision and Jaccard index, it is only necessary to know the approximate location of these.

Table 1. Performance metrics resulting from model training.

Val_loss	Loss	Accuracy	F1-Score	Jaccard	Recall	Precision
0.338	0.162	0.9973	0.713	0.557	0.654	0.805

To perform the real-time auscultation, first, the calculation of the camera alignment offset with the robotic arm, shown in Figure 6, was performed. This calculation was successfully performed, and the 2D values of the offset and the 3D spatial coordinates shown in the figure on the right side were obtained. This offset is a matter of a few millimeters, but it makes a difference in placing the gripper on the person's back in the correct place.

Next, the frames of the six patients' backs used in this test were obtained (see Figure 16), as well as their respective extractions from the image background (see Figure 17) and the key points were predicted using the trained prediction model, as illustrated in Figure 17 (white points).

With this information, the auscultation points' 2D coordinates were successfully calculated (as shown in green in Figure 17) using the key points, also successfully predicted. The model successfully predicted 35 out of 36 desired key points (as shown in Figure 17), leading to an accuracy of 97.2%. This missed key point did not influence the performance of the algorithm, since it is built to handle these issues. The respective 3D spatial coordinates of the auscultation points were also calculated, and are presented in Figure 18.

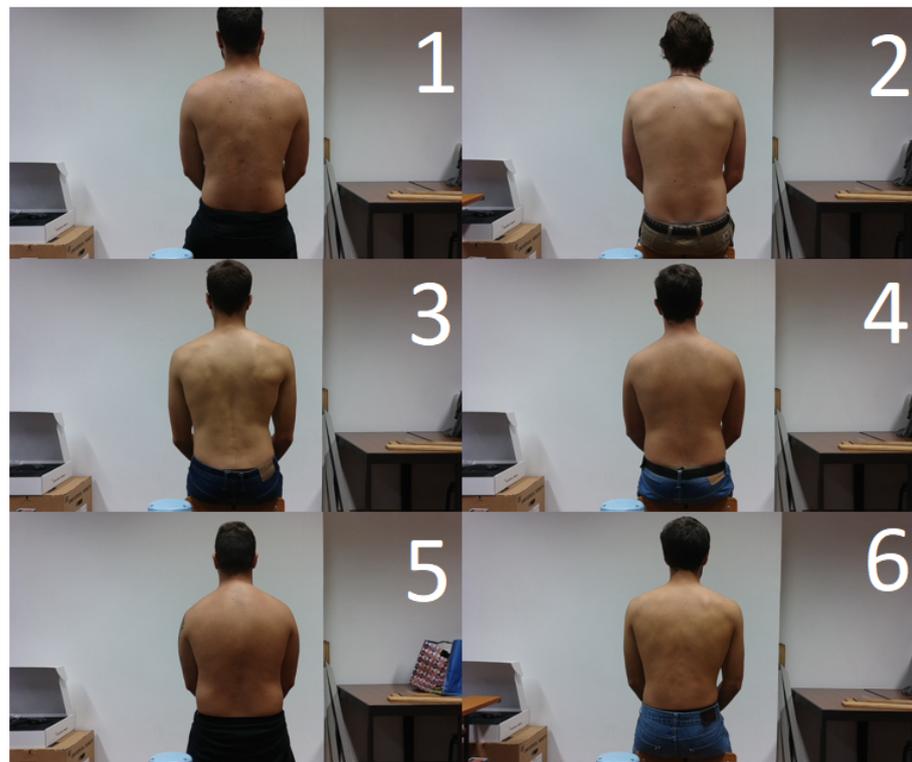


Figure 16. Photographic back images for six distinct patients, prepared for auscultation keypoints prediction.

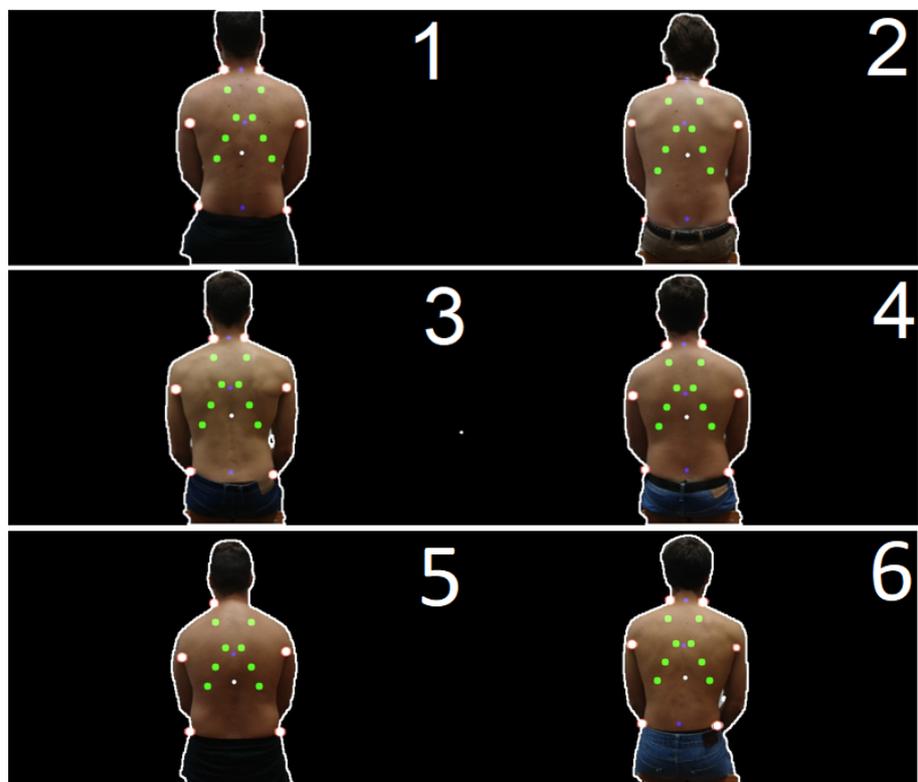


Figure 17. Auscultation points (represented by green circles) calculated from morpho-anatomical clues (represented by white circles, estimated), for six distinct patients. (After processing subfigures from Figure 16).

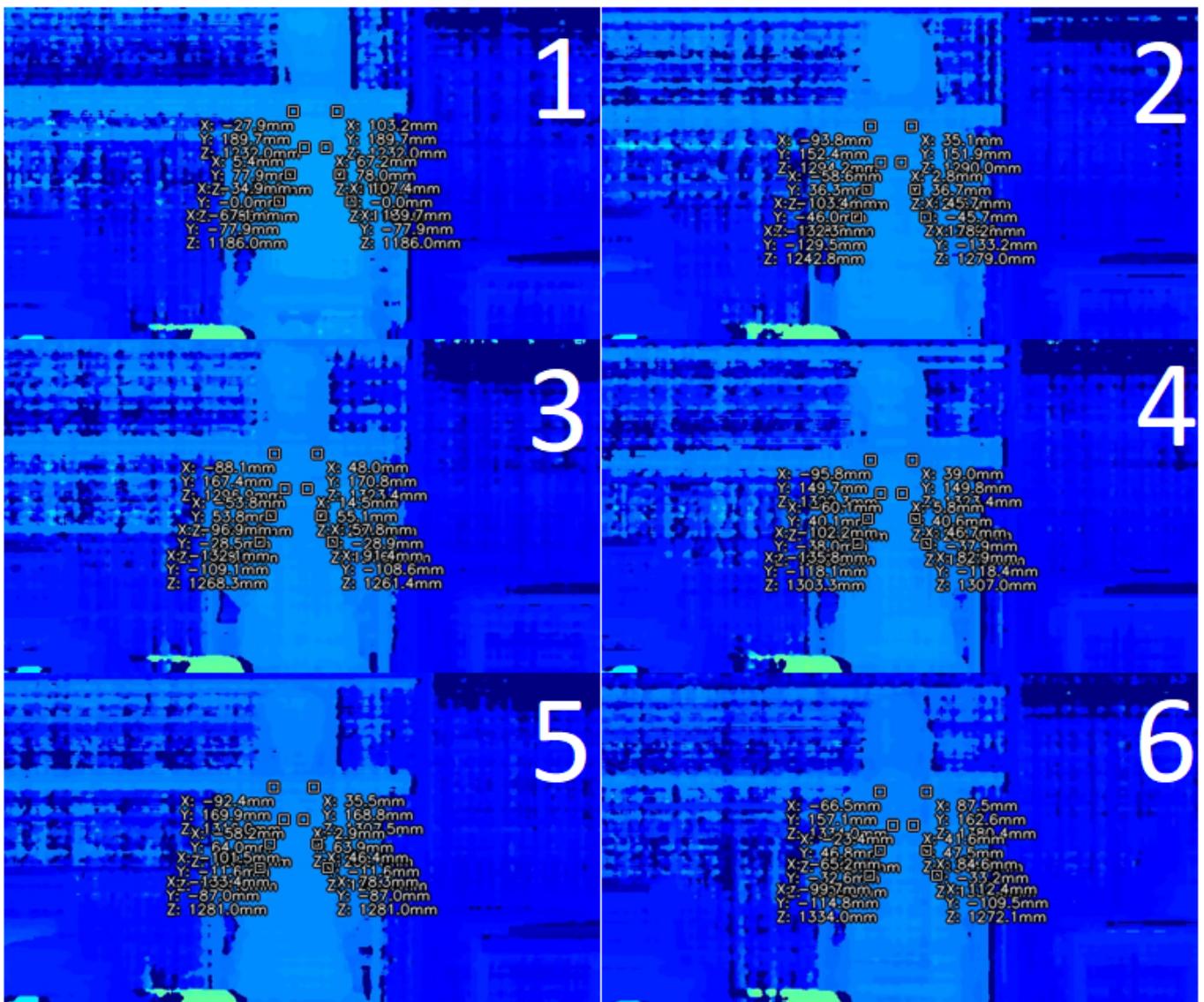


Figure 18. Depth image showing the torso plane and background plane for six distinct patients. Auscultation points are represented by squares with the related 3D spatial coordinates on the patients' back.

3.2. Positioning Subsystem

Finally, the gripper of the robotic arm was placed at the calculated auscultation points of four patients, as shown, for example, in Figure 19.

The robotic arm successfully positioned the gripper on the desired auscultation points in patients 1, 3, 5, and 6, except for points *L1* and *R1*. These were not entered into the distance calculation because the UR3e robot does not have enough reach to get there.

As can be seen from Figure 20, the points auscultated by the robotic arm are similar to the 2D points calculated earlier by the algorithm, leading to the conclusion that the robotic arm has a good accuracy when placing the stethoscope. These may have some divergences due to the individual not being able to be 100% still during the entire auscultation process.

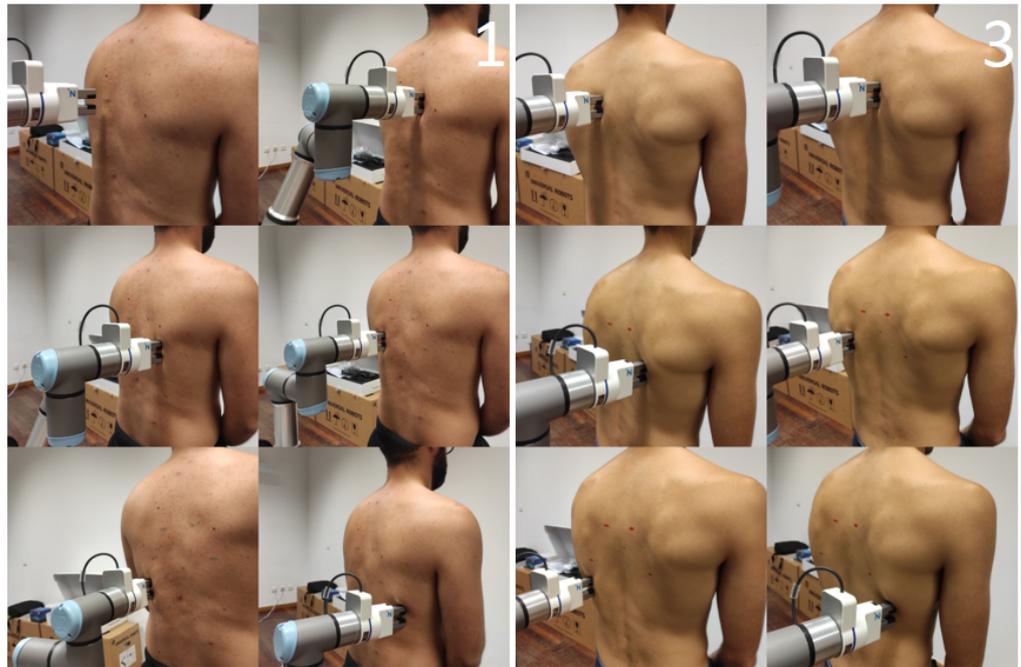


Figure 19. Robotic arm at the auscultation points of individuals 1 (left images) and 3 (right images).



Figure 20. Points auscultated (red and green) in patients 1, 3, 5, 6 by the developed platform.

4. Discussion

The developed platform was tested on six individuals, with distinct height, weight, and skin color. Through these tests, it was possible to conclude that the model trained through the dataset was robust enough to predict the key points in different individuals. The model also did not prove to be sensitive to variations in light, since during the tests there was no control of the lighting conditions. Thus, it is possible to say that the results obtained

in terms of localization were quite good, being sufficient to perform the auscultation task, which does not require precise positioning.

The robotic arm was able to reach the desired auscultation points (6 of 8, due to the limited reach of the arm) successfully, depending on the individual's physical structure, namely his back length. The reason for this is that in patients 2 and 4 the robotic arm was unable to go to points $L4$ and $R4$, because when descending from point $L3$ to point $L4$, the elbow joint hit the robotic arm's support table, as can be seen in Figure 21.



Figure 21. Collision of the robotic arm elbow joint with the bench where it is installed, when trying to reach the point $L4$ of patients 2 and 4.

This was due to these patients being shorter, because by the calculation of the 3D spatial coordinates, as shown in Figure 18, these patients have the lowest y coordinate value at points $L4$ and $R4$ among all. From the depth image (Figure 18) it can be assumed that the robotic arm can successfully reach all desired auscultation points in all patients that have a y coordinate, at points $L4$ and $R4$, above -114.0 mm. The remaining auscultation points in patients 2 and 4 were successfully reached and are shown in Figure 22. The results are very similar to the previous four patients, in the first four auscultation points.



Figure 22. Points auscultated (red) in patients 2 and 4.

On the other hand, there is the possibility of the person moving; the system is not designed to reorient itself whenever the person moves, but this can be minimized if the person becomes seated and as stable as possible, as is also required in other medical

examinations. The results so far are preliminary, and all the individuals tested were pleased with the robot's operation and did not feel any discomfort.

When compared with the work by Zhu et al. [24], the system presented on this paper uses a RGB-D camera to capture an image of the patients back, while the system from Zhu et al. uses a RGB-D camera to capture a 3D point cloud scan of the patient, and processes this point cloud to determine the auscultation points. Regarding practical experiments, Zhu et al. only performed experiments with four subjects, and no information is presented regarding the percentage of detected auscultation points [24]. Regarding the work by Tsumura et al. [23], its goal was to develop a robotic auscultation platform for estimation of the landing positions to hear the sounds of four cardiac valves, but there is no information of tests having been conducted on live subjects (it has only been tested on a mannequin). Furthermore, this work uses a LiDAR camera to acquire the 3D contour of the body surface as point cloud data.

5. Conclusions

Auscultation is an essential process in clinical activity, but its practice has been recently targeted by some challenges. First, the safety of both participants, doctor and patient, is at risk in the presence of airborne diseases (e.g., COVID-19, but also the common cold, chickenpox, and measles, among many others). On the other hand, the auscultation task, which requires the direct contact of the stethoscope with the patient's body, may involve some preparation time, especially in people with physical limitations, such as the elderly. This may represent a barrier for the healthcare professional. In addition, computer systems already exist to automatically classify heart and breath sound abnormalities with a very small error rate [38,39], paving the way to assisted diagnosis. Hence, the health professional can be released for more differentiated tasks. Finally, there are privacy concerns, as the patient is not always comfortable with the physical exposure that the procedure requires. This article describes a new system that can automatically perform the auscultation process. A state-of-the-art review in the field was first presented, covering both vision systems and robotics, in the medical context or for auscultation. The automation of the auscultation process is still a relatively unexplored area. After providing an overview of the system, all details were thoroughly described. Preliminary results have shown that the robotic platform allows the estimation of auscultation positions for listening to the sounds of four heart valves, based on reference information from the body anatomy, and simulates the placement of a stethoscope while maintaining contact force. The developed robotic platform has the potential to address the critical issue of an increasing demand for healthcare services while offering total safety for the users. This technology can further increase the efficiency of screening for abnormal clinical signs with similar resources, making it also useful as a telemedicine tool.

As future work, this platform should undergo some changes. For better accuracy in placing the stethoscope at the auscultation points, the patient should be more stable, in other words, the patient should be in a sitting position, as this is also the best position to obtain a precise estimation of points and better auscultation results. The robotic arm will also have to be replaced by another one with a larger reach (since the one used had a limited reach for our purposes, hindering its movements), and to have a larger work area, it should not be mounted on a table, but on a specific platform that maximizes its functioning. The camera should also be placed in a fixed position where it cannot be moved with the slightest touch. To complete the purpose of this platform, a digital wireless stethoscope should be integrated with the gripper.

Author Contributions: Conceptualization, M.F.S. and L.C.; methodology, D.L.; writing—original draft preparation, D.L.; writing—review and editing, L.C. and M.F.S. All authors have read and agreed to the published version of the manuscript.

Funding: Partially supported by FCT-UIDB/04730/2020 and FCT-LA/P/0063/2020 projects.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is not publicly available, but the authors can provide limited access for research purpose in the condition of clear authorship attribution.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Groover, M.P. *Automation*; Encyclopedia Britannica: Chicago, IL, USA, 2020.
2. El Zaatari, S.; Marei, M.; Li, W.; Usman, Z. Cobot programming for collaborative industrial tasks: An overview. *Robot. Auton. Syst.* **2019**, *116*, 162–180. [CrossRef]
3. Aaltonen, I.; Salmi, T. Experiences and expectations of collaborative robots in industry and academia: Barriers and development needs. *Procedia Manuf.* **2019**, *38*, 1151–1158.
4. Zayas-Cabán, T.; Haque, S.N.; Kemper, N. Identifying Opportunities for Workflow Automation in Health Care: Lessons Learned from Other Industries. *Appl. Clin. Inform.* **2021**, *12*, 686–697. [CrossRef]
5. Dupont, P.E.; Nelson, B.J.; Goldfarb, M.; Hannaford, B.; Menciassi, A.; O'Malley, M.K.; Simaan, N.; Valdastrì, P.; Yang, G.Z. A decade retrospective of medical robotics research from 2010 to 2020. *Sci. Robot.* **2021**, *6*, eabi8017. [CrossRef]
6. Stumpo, V.; Staartjes, V.; Klukowska, A.; Kafai Golahmadi, A.; Gadjradj, P.; Schröder, M.; Veeravagu, A.; Stienen, M.; Serra, C.; Regli, L. Global adoption of robotic technology into neurosurgical practice and research. *Neurosurg. Rev.* **2021**, *44*, 1–13. [CrossRef]
7. Athanasiou, A.; Xygonakis, I.; Pandria, N.; Kartsidis, P.; Arfaras, G.; Kavazidi, K.R.; Foroglou, N.; Astaras, A.; Bamidis, P. Towards Rehabilitation Robotics: Off-The-Shelf BCI Control of Anthropomorphic Robotic Arms. *BioMed Res. Int.* **2017**, *2017*, 5708937. [CrossRef] [PubMed]
8. Montinari, M.R.; Minelli, S. The first 200 years of cardiac auscultation and future perspectives. *J. Multidiscip. Healthc.* **2019**, *12*, 183. [CrossRef] [PubMed]
9. Ferlay, J.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Pineros, M. *Global Cancer Observatory: Cancer Today*; Technical Report; International Agency for Research on Cancer: Lyon, France, 2020.
10. Alyafei, K.; Ahmed, R.; Abir, F.F.; Chowdhury, M.E.; Naji, K.K. A comprehensive review of COVID-19 detection techniques: From laboratory systems to wearable devices. *Comput. Biol. Med.* **2022**, *149*, 106070. [CrossRef] [PubMed]
11. Huang, Y.; Meng, S.; Zhang, Y.; Wu, S.; Zhang, Y.; Zhang, Y.; Ye, Y.; Wei, Q.; Zhao, N.; Jiang, J.; et al. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods. *medRxiv* **2020**. [CrossRef]
12. Hirosawa, T.; Harada, Y.; Ikenoya, K.; Kakimoto, S.; Aizawa, Y.; Shimizu, T. The Utility of Real-Time Remote Auscultation Using a Bluetooth-Connected Electronic Stethoscope: Open-Label Randomized Controlled Pilot Trial. *JMIR mHealth uHealth* **2020**, *9*, e23109. [CrossRef]
13. WHO. *Chronic Obstructive Pulmonary Disease (COPD)—World Health Organization*; Technical Report; World Health Organization: Geneva, Switzerland, 2022.
14. Sarkar, M.; Madabhavi, I.V.; Niranjana, N.; Dogra, M. Auscultation of the respiratory system. *Ann. Thorac. Med.* **2015**, *10*, 158–168. [CrossRef]
15. Rennoll, V.; McLane, I.; Emmanouilidou, D.; West, J.; Elhilali, M. Electronic Stethoscope Filtering Mimics the Perceived Sound Characteristics of Acoustic Stethoscope. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1542–1549. [CrossRef] [PubMed]
16. Nowak, L.; Nowak, K. Sound differences between electronic and acoustic stethoscopes. *BioMedical Eng. OnLine* **2018**, *17*, 104. [CrossRef]
17. Kalinauskienė, E.; Razvadauskas, H.; Morse, D.; Maxey, G.; Naudžiūnas, A. A Comparison of Electronic and Traditional Stethoscopes in the Heart Auscultation of Obese Patients. *Medicina* **2019**, *55*, 94. [CrossRef] [PubMed]
18. Leng, S.; Tan, R.S.; Chai, K.T.C.; Wang, C.; Ghista, D.N.; Zhong, L. The electronic stethoscope. *BioMedical Eng. OnLine* **2015**, *14*, 66. [CrossRef]
19. OECD. *Health Statistics*; Technical Report; OECD: Paris, France, 2022.
20. Ma, Y.; Xu, X.; Yu, Q.; Zhang, Y.; Li, Y.; Zhao, J.; Wang, G. LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using bi-ResNet Deep Learning Algorithm. In Proceedings of the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17–19 October 2019; pp. 1–4. [CrossRef]
21. Kim, Y.; Hyon, Y.; Lee, S.; Woo, S.D.; Ha, T.; Chung, C. The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulm. Med.* **2022**, *22*, 119. [CrossRef] [PubMed]
22. Liu, J.X.; Goryakin, Y.; Maeda, A.; Bruckner, T.; Scheffler, R. Global Health Workforce Labor Market Projections for 2030. *Hum. Resour. Health* **2017**, *15*, 11. [CrossRef]
23. Tsumura, R.; Koseki, Y.; Nitta, N.; Yoshinaka, K. Towards fully automated robotic platform for remote auscultation. *Int. J. Med. Robot. Comput. Assist. Surg.* **2022**, *19*, e2461. [CrossRef]
24. Zhu, Y.; Smith, A.; Hauser, K. Automated Heart and Lung Auscultation in Robotic Physical Examinations. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4204–4211. [CrossRef]
25. UR3e Technical Specifications. Available online: <https://www.universal-robots.com/media/1807464/ur3e-rgb-fact-sheet-landscape-a4.pdf> (accessed on 18 January 2023).

26. UR3e Collaborative Robot Arm That Automates Almost Anything. Available online: <https://www.universal-robots.com/products/ur3-robot/> (accessed on 18 January 2023).
27. OAK-D—DepthAI Hardware Documentation 1.0.0 Documentation. Available online: <https://docs.luxonis.com/projects/hardware/en/latest/pages/BW1098OAK.html> (accessed on 18 January 2023).
28. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*; Springer: Cham, Switzerland, 2015. [CrossRef]
29. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015. [CrossRef]
30. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
31. Jing, J.; Wang, Z.; Rättsch, M.; Zhang, H. Mobile-Unet: An efficient convolutional neural network for fabric defect detection. *Text. Res. J.* **2020**, *92*, 004051752092860. [CrossRef]
32. Perez, L.; Wang, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv* **2017**, arXiv:1712.04621.
33. Yang, S.; Xiao, W.; Zhang, M.; Guo, S.; Zhao, J.; Shen, F. Image Data Augmentation for Deep Learning: A Survey. *arXiv* **2022**, arXiv:2204.08610.
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 240–248. [CrossRef]
36. Robots, U. The URScript Programming Language, 2019. Available online: <https://www.universal-robots.com/download/manuals-e-series/user/ur3e/513/user-manual-ur3e-e-series-sw-513-portuguese-pt/> (accessed on 18 January 2023).
37. GitHub—Depthai-Experiments/Gen2-Calc-Spatial-on-Host at Master-Luxonis/Depthai-Experiments. Available online: <https://github.com/luxonis/depthai-experiments/tree/master/gen2-calc-spatial-on-host#calculate-spatial-coordinates-on-the-host> (accessed on 30 July 2022).
38. Chen, D.; Xuan, W.; Gu, Y.; Liu, F.; Chen, J.; Xia, S.; Jin, H.; Dong, S.; Luo, J. Automatic Classification of Normal–Abnormal Heart Sounds Using Convolution Neural Network and Long-Short Term Memory. *Electronics* **2022**, *11*, 1246. [CrossRef]
39. Rocha, B.M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.P. Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem? *Sensors* **2020**, *21*, 57. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.