**Main artery Pulmonary Embolism detection,   
using traditional computer vision techniques**



Master of Science (MS) in Data Science  
Module: ITC6009A1 – Machine Vision in Data Science  
Instructor: Dr. Drakakis Georgios

Students’ names & IDs:

Alkiviadis Kariotis 241735

Konstantinos Tzoras 312632

Christos Karatsiolis 270254

Dimitra Chatzopoulou 270580

Term: Fall Term 2023  
Type: Group Term Project - Report

Submission Date: Tuesday 19th of December

Words:

Table of Contents

[Abstract 3](#_Toc153651832)

[Keywords 3](#_Toc153651833)

[Introduction 4](#_Toc153651834)

[Materials & Methods 6](#_Toc153651835)

[Reading & creating slices 6](#_Toc153651836)

[Segmenting the slices 6](#_Toc153651837)

[Cropping the chosen slices 7](#_Toc153651838)

[Isolating the pulmonary artery and detecting the embolus 7](#_Toc153651839)

[Results & Discussion 8](#_Toc153651840)

[Conclusion and Future Work 16](#_Toc153651841)

[References 16](#_Toc153651842)

[Appendices 17](#_Toc153651843)

[Appendix 1 17](#_Toc153651844)

[Radiologist multiple Interviews Summary 17](#_Toc153651845)

[Appendix 2 21](#_Toc153651846)

[Appendix 3 21](#_Toc153651847)

# Abstract

200-250 Words

Pulmonary embolism (PE) and more specifically, a large PE can lead to severe life-threatening problems or even mortality. Over 100.000 deaths have been reported between the last 10 years, due to this disease, from which the 10-30% of people died within the first month of diagnosis. Thus, there is a necessity for further investigations and improvements of the early detection of this disease.   
In order to diagnose the PE, the patients should take a computed tomography pulmonary angiography (CTPA). This examination consists of multiple scans which are pictures of the blood vessels that run from the heart to the pulmonary arteries.

The aim of this project is to diagnose if a patient is positive to pulmonary embolism, by detecting the presence of the embolism from computed tomographic pulmonary angiographies dataset. The scope is to provide rapid automated detection of pulmonary embolisms of patients so as to avoid human errors and reduce their effort. The dataset contains 91 patients with multiple scans per CTPA.   
Computer vision techniques are utilized in order to approach this goal since there is a pattern of the identification of PE through the computed pulmonary angiographies. Specifically, the diagnosis is based on sufficient contrast enhancement. However, the diverse range of clinical representations from people that are not ail to death diagnosis can be very challenging. Thus, aiming to address these challenges, various methods of computer vision were utilized such as histogram equalization, blur, canny edge detector, thresholding from which some of them brought better results than others.

## Keywords

Embolism, main pulmonary artery, detection, optimal parameters,

# Introduction

700-800 words Literature Review

SHOULD MENTION HERE THE SCOPE OF OUR PROJECT:  
Detection of embolisms located inside the main pulmonary artery only and not in the smaller branches of it, located in the area of the lungs.

Computer Vision is a part of artificial intelligence, which is used for training computers to comprehend and interpret the visual world. This field includes image processing techniques which aim to assist machines to recognize and detect objects from digital images. These advanced techniques enable the automation of tasks; thus, human effort can be minimized, as well as multiple fields of research and applications can benefit from increased efficiency and accuracy.  
 This development can be applied to various industries and fields such as medicine, agriculture, healthcare. The following project aims to apply this knowledge for the detection of Pulmonary Embolism (PE).  
Pulmonary Embolism is a blood clot that blocks the pulmonary artery in the lungs. To be more precise this blood clot comes from a deep vain in the leg and afterwards transfer itself to the lungs where it gets trapped in a smaller lung artery. Pulmonary Embolism can be life threatening without treatment and cause various problems such as cardiac arrest and arrhythmia. The occurrence and fatality of this disease have been increased between the last 10 years, with more than 100.000 deaths from which the 10-30% died within the first month of diagnosis. Thus, with that rate of mortality and the creation of future health issues it is necessary to investigate and provide ways that will help to address this illness. One improvement which is also the scope of the project, is to provide rapid automated detection of pulmonary embolisms in patients so as to avoid human errors and reduce their effort.   
The dataset that is used in the project consists of 91 computed topographies of various patients. The files are in the format of nrrd, thus, Simple ITK library will be used since it’s well known for medical image data processing, as well as nrrd library which is used to read and write NRRD files. Each CTPA consists of numerous scans since it takes pictures of the blood vessels that go from the heart to the lungs. The arteries can be detected through these scans as bright and white. In contrast, the blockages or blood clots are inside or at the outline of an artery and are presented darker and with higher contrast (grey color). That is the pattern that will be used.   
Initially, each CTPA consisted of hundreds of images from which only a few of them display the two arteries clearly and from which is accurate if the patient is positive to PE or not. In this part, the first challenge was to identify and keep only the images that were containing the central artery in a clear way. Thus, the first filtering was completed by applying a specific threshold to the intensities of pixels from the grayscale images. However, that wasn’t enough, thus, by tuning the parameters of trackbars, it was managed to detect the circles in the image. In this way, we filtered all the images that were containing a circle meaning that they were having a central artery.  
Afterwards, it was crucial by using the scans that were found and saved from the above steps, to manage to detect the embolisms more accurate as possible. Aiming to do so, multiple image processing techniques as well as multiple different parameters were examined and applied again in order to end up with the more accurate result. For instance, gaussian blur, crop of an image, application of binary threshold (also adaptive and band thresholds were tested), connected components, erosion, and dilation. The final step was to save the images that contained the central artery with colored embolism. During these processes, we were facing many challenges since there was a diverse range of clinical representations from people that are not ail to death diagnosis. Moreover, a lot of parts outside of the main artery were misleading thus, there could be applied many improvements that will be profoundly analyzed during this project.

Jupyter Notebook will be the environment that will be used in order to run the desired processes and the files that contains the codes will be saved in “. ipynd” format.

# Materials & Methods

## Reading & creating slices

Initiated our project by reading the NRRD files (Appendix 2), which consisted of 91 patients along with their tomography pulmonary angiograms positive and negative for pulmonary embolism. For each patient, a different number of slices/topographies have been extracted, and each one had a completely non-identical number of slices/frames (Appendix 1, Question 2).

As a next step, we converted the image to grayscale and all the following methods were used sequentially after this operation. This operation took place to deal with a single-color channel and at the same time, reduce the level of complexity and computational requirements.

Following our interview with the radiologist, we inspected various window levels and window widths (medical terms used on the tomographies to change pixel intensities) to get rid of the branches of the main pulmonary artery and have “clearer” frames of the main pulmonary artery. In terms of machine vision techniques, the above-mentioned operation was performed to apply two zero thresholds to set the appropriate upper and lower gray levels ("Radiopaedia," n.d.).

After applying the medical masks, we mirrored the slices to be used in the same way as they are inspected and diagnosed in the medical industry. For each patient, we stored all the processed slices in png format and saved them to a respectively named folder.

## Segmenting the slices

To capture dynamically, between all the patients of the dataset, the specific slices in each patient that the main pulmonary artery is depicted, we applied the Hough Circles cv2 method. This method was used to identify the circles shapes inside the images, which contained the ascending thoracic aorta. It uses the mathematical equation ((x−xcenter)2+(y−ycenter)2=r2) to capture the center of the circle. Hough Circles technique had several parameters (6) which can be depicted below as follows:

* **Input image:** The processed image incorporated inside the function.
* **cv2.HOUGH\_GRADIENT:** This parameter aims to utilize bothedge detection and gradient information to identify the circles in each image.
* **dp:** The value of 1 was used, which provides a good balance between accuracy and speed (computational time). Even though processing time was not an issue in our scope, after trial and error we observed that the specific parameter value was achieving the desired result.
* **minDist:** Set a value of 100, which worked in our tryouts managing the prevention of overlapping circles and at the same time, not miss detecting the desired circle.
* **param1:** A value of 30 was used, which provided a reasonable balance between sensitivity and accuracy, resulting in identifying the proper circle.
* **param2:** To specify the size of circles to be detected, we used a value of 35.
* **minRadius & maxRadius:** Set values of 20 and 35 respectively, to capture all the desired circles between the aforementioned range of radius/diameter.

As we iteratively looped through all slices of a specific patient, in which we were able to detect a circle, were stored the slice number in a list. As a result, we created the main list which contained all the frames, where the circle was successfully detected. In this list, the first and the last slices were not depicting the main pulmonary aorta, so we decided to use the center of the list and look 15 slices above and below that number. As we discuss in the conclusion/results section, a different way could have been used.

## Cropping the chosen slices

Firstly, we created two copies of the chosen slices. After observing the dataset, we came up to the conclusion that all the main pulmonary arteries of patients, located in a specific window of our images. Based on this consideration, we processed by cropping the image to avoid at the same time miss detection of objects that are out of our concern (descending thoracic aorta, thoracic spine, thoracic ribs etc.).

By initializing a start and end point at 0.25 and 0.75 respectively, and by identifying the correct coordinates (upper left, upper right, lower left, lower right) we correctly created the rectangle, which was the only thing kept from the image, and turned every pixel located outside the rectangle to completely black.

## Isolating the pulmonary artery and detecting the embolus

Specific steps were follower to correctly spot the embolus, located inside the main pulmonary artery as follows:

1. **Binary Threshold:** Set parameter of 115 and kept all the values above this threshold.
2. **CCA (Connected Component Analysis):** Performed this method and reserved the second largest component (which was composed of the main artery, the ascending thoracic aorta and the superior vena cava), because the largest one was the background.
3. **Kernel:** For erosion and dilation purposes, we defined a static variable of (3,3) kernel, to use multiple times throughout our methods. We reached the conclusion of selecting this size of kernel after multiple trials and errors.
4. **Erosion:** Performed two rounds of erosion to separate the main pulmonary artery from the ascending thoracic aorta.
5. **CCA:** Kept the second largest component, which was only depicting the main pulmonary artery.
6. **Cropping:** Since the embolisms in patients were detected at the bottom half of the main artery, we proceeded by dynamically cropping the component horizontally. The upper half was turned black, and we proceeded with the lower half that was in our concern.
7. **Dilation:** Performed six rounds of dilation to dilate and make sure that we include the embolus.
8. **Masking:** From the bottom half dilated component, we masked/mapped it back to the original slice.
9. **Blurring:** On the original masked part of the image, we applied a Gaussian blur with an internal applied kernel of (5,5).
10. **Band Thresholding:** Used a band thresholding of the blurred image, with values between 90 and 120, to isolate the pixel range that the embolus had.
11. **Erosion:** One round of erosion was performed to eliminate the noise.
12. **CCA:** Kept the second largest component, which was the embolus.
13. **Dilation:** Two rounds of dilation were applied to make the detection of embolus more visible.
14. **Masking:** We mapped and colored (red) the detected pixels of the embolus to the untouched initial image that we imported.

# Results & Discussion

By reading and extracting the images in grayscale, from the NRRD files, the original images that we successfully retrieved were looking like the following one:

An x-ray of a chest

Description automatically generated

Figure : Original not processed slice 341/Patient 006

Following the interviews performed with the radiologists and the overall feedback and guidelines that they provided us; we began by applying the suggested window and width level. The levels used in their industry were translated into our specific thresholds, which resulted in identifying the relevant parameters and interpreting their values in the context of our scope. As a result, we proceeded with the specific parameters of THRESHOLD\_TOZERO that were mentioned in “Materials and Methods” section. This process was implemented to eliminate the medical views, which aim to identify different sicknesses, and highlight the main pulmonary aorta and the embolisms included inside it. After performing all mentioned modifications, the result was the following:

A close-up of a ct scan

Description automatically generated

Figure : Processed slice 341/Patient 006

Afterwards, we have aligned and settled all our modifications, in relation with all the radiologists’ instructions, we proceeded with creating a loop to generate png images (with the applied medical masks) for all tomographies that each patient had inside our dataset.

The primary goal was to be able to capture all the slices containing the main pulmonary artery dynamically. At our baby steps, we started investigating a single slice, which was picked manually and contained the main pulmonary artery. Initially, we plotted the histogram (normal and equalized) and started to search for similarities between different slices of the same patient. This seemed to be working in the context of a single patient. When we observed different patients, we found no connection with the pattern observed to the previous one. Based on this failure, we decided to proceed with a band thresholding aiming to find if all the embolisms (in all patients) located in the same value range of pixels. Even though this method seemed to be successful for some patients because it retrieved the desired slices of each one, it didn’t work on a broader scale.

At this dead end, we figured out a similarity between the slices that contained the main pulmonary artery, which was the circle located at its upper right corner. We reached out to the radiologists again, who confirmed to us that this was called ascending thoracic aorta, and it is something that is obligatory to always be present in every patient. So, we tried to figure out a way to detect shapes in a circle format, indicating the use of the respected cv2 method. The method is called Hough Circles and by using trackbars we were able to spot the sweet combination of different parameters that would fulfill our purpose. Initially the method was working as follows:

A close-up of an x-ray

Description automatically generated

Figure : Non-parameterized Hough Circles method of slice 341/Patient 006

Subsequently, we fined tuned all six different parameters of our method, managing to successfully spot the ascending thoracic aorta as follows:

A close-up of a ct scan

Description automatically generated

Figure : Fined tuned Hough Circles method of slice 341/Patient 006

Having made possible to acquire all the slices from every patient that contained the main pulmonary artery, we proceeded with the steps of its isolation. In every step of the process that follows, multiple trials and errors tryouts attempts took place. Through histogram inspection, we performed a binary thresholding managing to eliminate various parts of the images that were not in our interest. The most-right image of *Figure5*, depicts this result:

A graph of blue lines

Description automatically generated

Figure : Binary thresholding through inspection of Histogram keeping values above 115 and turning everything else to black

By obtaining the thresholded image, we observed that on different patients we had a large variation of components sizes. Sometimes the descending thoracic aorta and spine were detected as single components in combination with a smaller main pulmonary artery. Thus, we decided to crop the image into a specific rectangle to avoid the aforesaid false targeting. The cropping can be shown in Figure6:

A close-up of a ct scan

Description automatically generated

Figure : Original and Binary thresholding image before and after cropping

On the cropped image, we executed a connected components analysis and kept the second largest component which was the one of our interest (largest component was the background). To separate the multiple components detected, we implemented two rounds of erosion with a kernel mask of (3,3). The eroded result can be depicted in the third image of Figure7:

A white logo on a black background

Description automatically generated

Figure : Performed CCA to keep the components containing the main pulmonary artery and two rounds of Erosion

Conducted a connected component analysis on the eroded image and kept the second largest component which was only the main pulmonary artery (largest component was the background). The result can be shown in the second image of Figure8:

A black and white image of a person

Description automatically generated

Figure : Performed CCA to obtain only the main pulmonary artery

Since the embolisms were detected at the half bottom of the main pulmonary artery, we performed a dynamic cropping process, taking into account the height of the component, finding the middle value pixels and cropping it accordingly. Furthermore, with the same kernel (3,3), we dilated the remaining component six more times. Through the dilation process we ensured that we would definitely incorporate the embolism to our next step. Then, we masked the dilated image back to the original image values. With the same order that this process is analyzed, it can be respectively shown in the following Figure9:

A black and white logo

Description automatically generated

Figure : Cropped main artery component, dilated and masked

Having our masked half lower cropped main pulmonary artery, we tried may and different binary thresholds parameters combinations, as to locate the embolisms. After several tries, we managed to achieve great performance for a single patient but not in others, bringing us eventually to the conclusion that we must blur our image and then give it a try. So, we applied a Gaussian Blur with a mask of (5,5), on the masked pixels on the original image. Now, we were able to find the sweet parameters of the band thresholding technique, that was able to work for numerous patients and bring us back parts of the outline of the main pulmonary artery and most importantly the embolisms. The second image is the blurred one and the third image is the band threshold, shown in Figure10:

A white logo on a black background

Description automatically generated

Figure : Blurring the masked image and applying a band thresholding

Then, we performed just one round of erosion (with the small kernel of (3,3)), which completely discarded some noise but most importantly the lines of the outline of the main pulmonary artery, leaving us with only the embolus. In the second image of Figure11, there are some different objects left so a connected component analysis came handy, by keeping the second largest one which is the embolus. After keeping the aforementioned component, we performed with the small mask again, two rounds of dilation so as to be able to spot the embolus more intense. All of the process with the order mentioned in this paragraph can be seen in Figure11:

A black and white image of a planet

Description automatically generated

Figure : From band threshold, we performed erosion, then CCA to get the embolus and the dilation to spot it better

Finally, we reached the stage where we turned the embolus detected into RGB version, colored it in red color, and mask it back to the untouched original slice of the patient that we examined through the loop process that is taking place into our code (Figure12). Through this part of the code snippet a new folder is created, which for the selected images through the Hough Circles methods, creates a new folder and stores inside of it all the image after performing the whole aforementioned machine vision techniques and processes, mentioned in this particular section of the project.

A close-up of a ct scan

Description automatically generated

Figure : Original not processed slice 341/Patient 006, with the embolus detection colored in red

# Conclusion and Future Work

Drakakis Instructions

Conclusion and Future Work: approximately 700-800 words. This contains a very

brief summary of the results, leading to the decision as to whether or not the

goal was fulfilled and why. It is common to reiterate a major shortcoming and

provide realistic future work, i.e. use a different method and why, try to include

more data from source X, apply the same methodology to a different problem

(transferability of your work is very useful), or the like.

From all the steps throughout our analysis, we reached a point where we were able to dynamically target specific slices of CT scans of every patient, locate the one’s containing the main pulmonary artery and specifying the embolus exact location. So, the project goals was reached as we managed to classify negative patients with above 90% accuracy, and in the positive patients misclassified (meaning that we didn’t correctly detect/spot the embolus) around 30% of the slices that examined. In simple words, at positive patients from the 31 slices that we examined we identified 70% correctly the embolisms. In the case of negative patients, we identified embolisms outside the main pulmonary artery or on the outline of it, as depicted in Figure13:

A close-up of a ct scan

Description automatically generated

Figure : Two slices from random negative patients in which we located the embolisms just on the edge of the outline of the main pulmonary artery

This brings up a modification that should have been made. The optimal solution would be, to use a percentage of the received slices when we looked up above and below of the median slice of the list containing the main pulmonary artery images detected. Based on our method, by retrieving the median value and 15 slices above and below, we resulted in specific patient cases getting completely wrong slices. This occurred because some patients had around 70 slices (in comparison with 90% of our dataset which contained around 500 slices), which was due to the fact of emergency situations (Appendix 1, Question 4). One of the main reasons of wrong classifications was the fact that there was a diverse representation of every component of the blood cycle, from patient to patient there were several differences in their organ’s sizes and positions as well as that the computed tomography pulmonary angiographies come from various equipment (Appendix 1). One more future modification that could be done is to try to detect the outline of the main pulmonary artery. By being able to perform this kind of operation, we could fix our misclassifications of the negative patients, by using the coordinates of the contour and for each detection made on it or outside of it just ignore it and don’t map it on the original image. Then in the context of our project scope, by being able to detect and exclude for the predictions the exact outline, the results on the positive patients will perform drastically. From the 30% that we wrongly identified the embolisms, in all of the case one of the band thresholds that is being used through the process, detects a part on the outline of the main pulmonary artery, because it ranges in the same pixel value range. So, this could result in providing better numbers of embolisms detections and also open up the way for further machine vision techniques and methods applications. Another interesting future approach that could be applied towards our project scope, are deep learning techniques that could assist in classifying if the patient was positive or not to pulmonary embolism. In this view, we would need to have extensive assistance from radiologists, who would provide us a labeled dataset in the way that we showed them to. The labels would annotation on the CT scans/slices of the patients that would mark the exact place where the embolisms are detected, in an appropriate way. In the same manner but with a simpler way of labeling procedure, a set of negative patients CT scans will be necessary. From that point, we could continue with applying deep learning techniques, ranging from simple Neural Networks to even Convolution Neural Networks, which will provide with much better results towards detecting and identifying the exact spot of the embolisms in the positive patients (Soffer et al., 2021). Concluding, the steps and the methods that we applied throughout our entire project, if slightly and accordingly modified, could be easily transferred to various other applications. Especially in the medical area, ,in Cardiac images or videos to detect mis functionalities or blockages in coronary arteries, on Cancer detection for finding breast cancers or brain tumors, in pet scans for various oncological issues, in Orthopedics towards the diagnose of bone fractures or even Dermatology for detecting a melanoma or other skin conditions.

# References

Drakakis Instructions

References: Approximately 10-12 should suffice in this case. Choose any style you

wish but be consistent

**WE HAVE GONE WITH APA STYLE. THE BULLETS ARE WRONG BUT DON’T DELETE THEM UNTIL YOU HAVE THE FULL LIST OF SOURCES THAT WE WILL USE (easy sorting purposes). ASK WHEN REACHING THAT POINT**

<https://www.svhlunghealth.com.au/procedures/imaging/ctpa-ct-pulmonary-angiogram>

* Radiopaedia. (n.d.). Windowing (CT). Radiopaedia. <https://radiopaedia.org/articles/windowing-ct>. Accessed on December 16, 2023.
* Soffer, S., Klang, E., Shimon, O., Barash, Y., Cahan, N., Greenspana, H., & Konen, E. (2021). Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Scientific reports*, *11*(1), 15814.

# 

# Appendices

## Appendix 1

### Radiologist multiple Interviews Summary

Question 1 *:Can you explain to us, the depicted components, and their use in the following CT scans?*

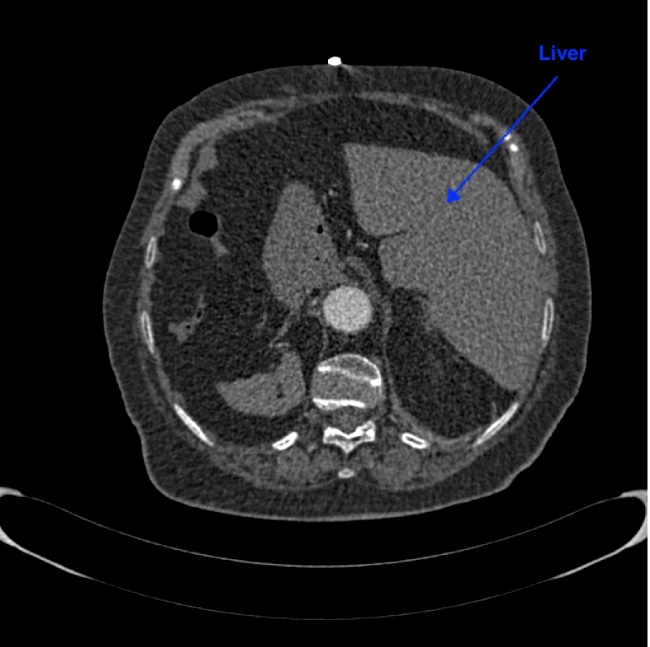
In the Figure … we can see the Liver depicted, which is responsible for the production of bile which helps towards the breakdown of fats in the small intestine during digestions. 

Figure :CT slice 145 of patient 006

In the Figure \_\_ we can see the Heart depicted, which is the center of the circulation system and pubs around the body as it beats. In simple words pushes the highly oxygen blood and pumps it back to the body.

An x-ray of a chest

Description automatically generated

Figure : CT slice 238 of patient 006

In the Figure \_\_ we can the left atrium, which takes the blood from the pulmonary veins, which is the only case where veins are filled with arteria blood, brought from the lungs completely cleaned.

An x-ray of a chest

Description automatically generated

Figure : CT slice 278 of patient 006

In the Figure \_\_ we can see the aortic arch, which is the middle transformation from the ascending thoracic aorta to the descending one and is the one the leaves from the left ventricle of the heart in order to get the blood to the tissues and organs. During this process it splitted into various branches.

An x-ray of a chest

Description automatically generated

Figure :CT slice 413 of patient 006

In the Figure \_\_ we can have a depiction of the various components that are in a slice which contains the main pulmonary artery (the main Λ shaped element at the center). Some “general” or basic elements are the thoracic ribs which are connected to the thoracic spine and at to the sternum. The darker/black part are the lungs which they contain smaller tissues or artery branches. Below the main pulmonary artery is the descending thoracic aorta, which role is mentioned in the following question. To the upper right of the main pulmonary artery is the ascending thoracic aorta and the smaller white reversed “triangle” below it is the superior vena cava or great vein trunk, which is a vein who drains the vein blood and returns the vein blood to the right atrium of the heart from the brain, the neck, and the upper limbs. Finally, is the main pulmonary artery with the embolism depicted at it’s right lower branch (the grayish dark sport)§.

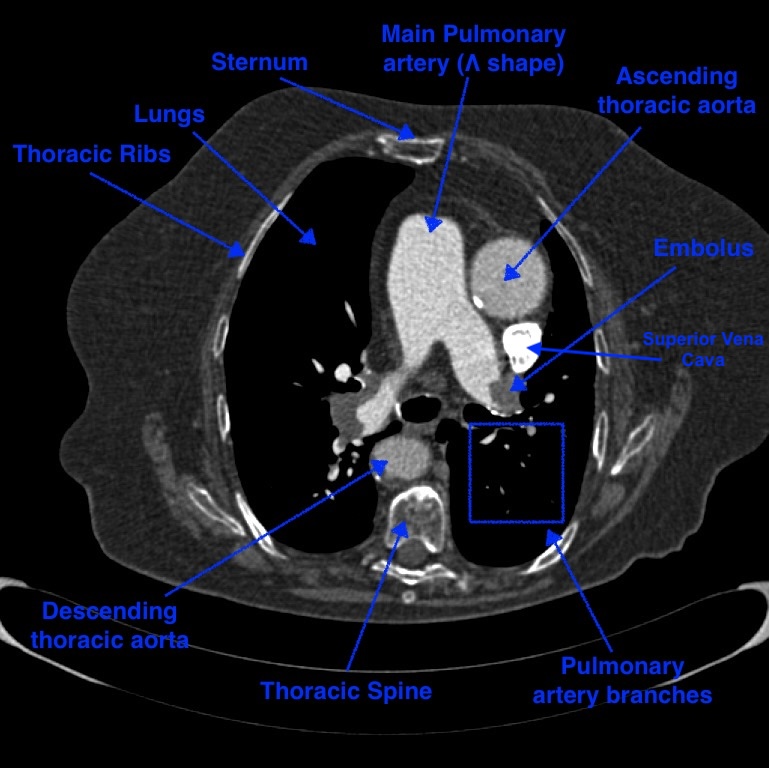


Figure :CT slice 341 of patient 006

Question 2 *:Can you give us a brief overview of the whole context in order to provide us the necessary knowledge to proceed with our project?*

Some general information about the blood cycle. Arterial blood begins with the aorta from the left ventricle of the heart and delivers oxygen and nutrients to all organs and tissues. Then it removes carbon dioxide and other waste products from the organs and tissues to the right atrium. From the right atrium the blood goes to the right ventricle and then with the pulmonary article goes to the two lungs. In lungs the blood is filled with oxygen and get ridded of the carbon dioxide. Afterwards, it becomes arterial blood again because it has been oxygened again, and it return to the right atrium with pulmonary veins. From there to the right ventricle to the aorta and the whole process in loop again.

In the process of the blood transition from the main pulmonary artery to the lungs, there embolisms can be spotted. Embolisms are blood clots (πήγματα). Those travels through vessels where it can cause a blockage. In simple words they are blocking the arteries, not letting the blood flow through the artery to the lungs, stopping the entire blood cycle.

Throughout our interview with the team, they have built for us an UI interface in which they demonstrated the basic techniques that we were using to inspect the CT scans. With our guidance we ended up with the specific window and width levels which in our industry understanding are parameters for change the brightness and the contrast of the CT scans. There are many windows in the medical industry, such as lung window, bone window, brain window and many more, but the one tailored to the project was the soft tissue one, which highlighted the arteries and veins of the CT scans.

Question 3 *: Why are there different number of CT scans for different patients in our dataset?*

The reason of the different number of CT scans slices can be due the gender of the patient, the weight of the patient and also the height of the patient. The reason why a small percentage of your patients is having much less than around 500 hundred slices (which is the normal) is due to the fact that maybe the urgent that was examined, was in a rush to get to the surgery so we just needed a few scans to serve as guidelines for the surgeon to operate.

Question 4 *: Why are there different coloring and pixel variations from patient to patient?*

This happens because the dataset does not consist of patients, who have examined in the same Computed tomography system. Each system has as default values for the CT scans different values of brightness and contrast, so that means that different masking is needed from one to another. After inspecting with the team approximately 10-15 patients from their dataset, we came up with some level and width windows that could serve their purpose for almost all of them.

## Appendix 2

NRRD or Nearly Raw Raster Data files, are the files that are stored in this specific format. Those files find application in the medical industry such as CT or MRI scans, and they are used for visualizing the medical image data. This type of file makes handy because of the various support for numerous data types and meta-information, providing the possibility for processing and analysis.

## Appendix 3

|  |  |
| --- | --- |
| **Terminology** | **Explanation** |
| Medical masking | In our industry translated as a specific combination of parameters of threshold to zero |
| Positive Patient | Patient who is “sick”, meaning who has an embolus |
| Negative Patient | Patient who is” healthy” and does not have an embolus |
|  |  |
|  |  |
|  |  |
|  |  |