UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

Máster Universitario en *Deep Learning*
for Audio and Video Signal Processing

MASTER THESIS

Deep learning analysis of vessel reduction images after EVAR

Javier Riera del Moral
Advisor: Pablo Carballeira López

September 2021

Deep learning analysis of vessel reduction images after EVAR

AUTOR: Javier Riera del Moral
TUTOR: Pablo Carballeira López

Dept. Tecnología electrónica y de las comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
September 2021

Resumen (castellano)

Un aneurisma aórtico es una dilatación de la aorta, la mayor arteria que suministra sangre al organismo. Los aneurismas más frecuentes son los de la aorta abdominal (AAA). Los AAA tienden a crecer y a romperse, con el consiguiente elevado riesgo de muerte por hemorragia interna. El tratamiento quirúrgico más empleado hoy en día es la colocación de una endoprótesis aórtica, EVAR (de las siglas en inglés, EndoVascular Aneurysm Repair). La EVAR es un dispositivo autoexpandible que se coloca por dentro de la arteria enferma para excluir el aneurisma de la circulación, reduciendo la presión que soporta y eliminando el riesgo de rotura del AAA. El mejor signo pronóstico de este tratamiento es la reducción del tamaño del AAA a lo largo del tiempo, una vez producida la despresurización. Llamamos saco aneurismático al AAA correctamente tratado con EVAR que queda excluido de la circulación. Sin embargo, el EVAR no está exento de complicaciones, las más frecuentes son las llamadas fugas, que son pequeñas entradas de sangre en el AAA tratado que condicionan la presurización del saco aneurismático y por tanto la reaparición del riesgo de rotura y hemorragia, que siempre se acompaña de una ausencia de reducción del tamaño del saco. Hay distintos tipos de fugas dependiendo del lugar por donde entra la sangre en el saco. Su detección temprana es fundamental para planificar un tratamiento adecuado a tiempo. La prueba diagnóstica actual para realizar el seguimiento de los EVAR es el CT (Tomografía Computarizada). Las imágenes obtenidas con esta técnica se estudian por los médicos en busca de manchas de contraste dentro del AAA tratado que indican la presencia de fugas. Estos picos de contraste pueden ser evidentes en algunos casos, pero difíciles de ver en otros, especialmente considerando el gran volumen de imágenes por cada CT. El modelo propuesto en este estudio consiste en una red de detección, basada en RetinaNet, para localizar el saco en las imágenes del CT y eliminar el ruido circundante. Después utilizar un modelo de clasificación binaria basada en redes convolucionales, tanto 2D como 3D, para analizar las imágenes y realizar una predicción de la evolución del tamaño del aneurisma, lo que permitiría a los médicos realizar una vigilancia focalizada en los pacientes con más riesgo de presentar fugas.

Abstract (English)

An aortic aneurysm is an enlargement of the aorta, the largest artery supplying blood to the body. The most common aneurysms are abdominal aortic aneurysms (AAA). AAAs tend to grow and rupture, resulting in a high risk of death from internal bleeding. The most commonly used surgical treatment today is the placement of an aortic stent graft, EVAR (EndoVascular Aneurysm Repair). EVAR is a self-expanding device that is placed inside the diseased artery to exclude the aneurysm from circulation, reducing the pressure on the aneurysm and eliminating the risk of AAA rupture. The best prognostic sign of this treatment is the reduction in size of the AAA over time, once depressurization has occurred. An AAA that is correctly treated with EVAR and excluded from circulation is called an aneurysmal sac. However, EVAR is not free of complications, the most frequent are the so-called leaks, which are small inflows of blood into the treated AAA that condition the pressurization of the aneurysmal sac and therefore the reappearance of the risk of rupture and bleeding, which is always accompanied by a lack of reduction in the size of the sac. There are different types of leaks depending on where blood enters the sac. Early detection is essential to plan appropriate treatment in time. The current diagnostic test to follow up EVARs is CT (Computed Tomography). The images obtained with this technique are studied by physicians looking for contrast spots within the treated AAA that

indicate the presence of leaks. These contrast peaks may be evident in some cases, but difficult to see in others, especially considering the large volume of images per CT scan. The model proposed in this study consists of a detection network, based on RetinaNet, to localize the sac in the CT images and remove the surrounding noise. Then using a binary classification model based on convolutional networks, both 2D and 3D, to analyze the images and make a prediction of the evolution of the aneurysm size, which would allow physicians to perform targeted surveillance on patients at higher risk of leaking.

Palabras clave (castellano)

Aneurisma de aorta, endoprótesis aórtica, fuga, saco aneurismático, tomografía computarizada (CT), aprendizaje automático (ML), aprendizaje profundo (DL).

Keywords (inglés)

Aortic aneurysm, Endovascular aortic repair, EVAR, endoleak, leak, aneurysm sac, computerized tomography (CT), machine learning (ML), deep learning (DL).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Motivation

Abdominal aortic aneurism (AAA) is a life-threatening disease consisting in a dilatation of the abdominal aorta exceeding the normal vessel diameter by 50% (around 3cm of diameter is considered an AAA). The AAA is characterized by a progressive expansion in size, although the progression of this growth is variable, remaining stable for some patients and other growing rapidly. The continuous expansion of the aneurysm size may end on a rupture, a weaken of the wall in the aneurysm sac that leads to internal bleeding. Aneurysms are usually asymptomatic until its rupture, which is often lethal with a mortality rate of 85 to 90% [1][2]. The objective is to identify and treat aneurysms before they rupture.

The actual procedure to identify if a patient has an AAA is through an axial CT-scan. CT scans are large volumetric grayscale images that show the anatomical structures of the body region explored, including the path of the aorta. CT scans are x-ray based, although these scans output a 3D volume of images instead of the 2D "projectional x-rays". The result encodes the radiodensity of millions of points (approximately a volume of 512x512x1000 pixels) and measures them in Hounsfield units, where lower values, for example air, are shown in black, and higher values, like bones, are shown in white [3][4]. Intermediate tissue´s values are shown in Table 1-1.

Table 1-1-1: Hounsfield scale table [4]

TISSUE	HU
Bone	+1000
Liver	40 to 60
Blood	40
Kidney	30
Muscle	10 to 40
Water	0
Whiter mater	-20 to -30
Grey mater	-37 to -45
Fat	-50 to -100
Air	-1000

Actual treatment relies on open or endovascular repair. The decision to treat depends on the evaluation of the risk of AAA growth and rupture, which can be difficult to assess in practice [1]. This decision is not only made based on the characteristics of the AAA but also the patient's operative risk and longevity [2].

Open repairs consist of an abdominal or flank incision. The aneurysm sac is opened and an interposition of a synthetic graft is sutured to both the healthy edges of the aorta. Endovascular repair is a less invasive approach that involves the intraluminal introduction of a covered stent graft through the femoral and iliac arteries; the stent graft acts as a sleeve that passes through the aneurysm sac landing in the healthy aorta above the aneurysm and in the iliac arteries below the aneurysm [1].

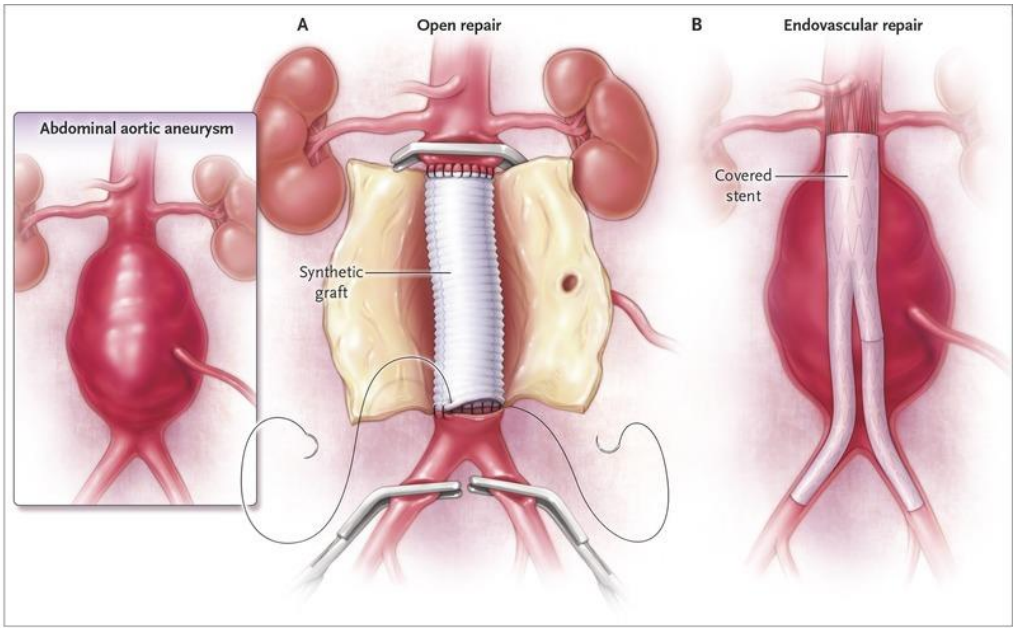

Figure 1-1: Abdominal Aortic Aneurysm repair techniques

For this project we will be focusing on post endovascular repair patients. In some series, as much of 20 to 30% of the patients with an EVAR may require a second intervention during the following years after the procedure, most of them also endovascular [1]. This are usually related to the development of endoleaks, wich are small blood flows inside the aneurysm sac that could lead to pressure raising that may led to continued aneurysm growth and subsecuent increase the risk of rupture [1][5].

These endoleaks can be detected in the CT scans as white peaks of contrast into the aortic sac. Several factors can reduce the capacity of the technique to detect some of these endoleaks, such as patient´s supine position, delay of the time between contrast is injected

and CT images obtained, precise moment of the cardiac beating, etc, can make this contrast can be difficult to detect for the human eye [6][7][8].

The main motivation for this master's thesis is to program a computer aided diagnosis system (CAD) focusing on the post-operative evolution for patients with an endovascular repair using state-of-the-art deep learning technology in image processing. The future utilization for this application can help post-EVAR patients' surveillance, detecting flaws of the stent through the CT output images, helping surgeons to detect and diagnose the problem and plan an adequate treatment, increasing the life expectancy of the patient.

This thesis also seeks to remark the advantages of deep learning technology on the medical field, especially in diagnosis, and introduce these systems into our hospitals in Madrid [9][10].

1.2 Objectives

The main objective of this master's thesis is to develop a CAD system to help predict the evolution of the aneurysm size after an EVAR using CT images and deep learning technology.

There are three secondary objectives. The first objective is to create a dataset of CT-scans of post-EVAR treated patients from "La Paz" university hospital. Secondly, to detect endoleaks as the principal prognostic factor for the aneurysm sac degrowth. Finally, I will evaluate the advantages and limitations of Deep learning on medical imaging and the development of a CAD system.

1.3 Structure of the report

This report has the following chapters:

- **Related work.** An explanation of the state-of-the-art and the last advances on image processing for EVAR CT-scans.
- **Design.** Evolution of the design of the model and post-processing.
- **Development.** Explanation of the process of creating the dataset and training the different models evaluated.
- **Integration and experimental results.** Presentation of the implementation of the models and validation results obtained.
- **Conclusions and future work.** Ideas for improvements and for future research on the topic.

2 Related work

In this section will be discussed the state of the art used in the design of the model proposed in this thesis along with the latest work on deep learning for medical imaging, focusing on CT scans and AAA related work.

2.1 Computer-Aided Diagnosis systems in medicine

Computer-aided diagnosis (CAD) in medicine is the result of a large amount of effort expended in the interface of medicine and computer science. Some CAD systems in medicine try to emulate the diagnostic decision-making process of medical experts. These systems may process data that can be complicated and/or massive in size and are capable of infer knowledge from data, improving future diagnosis, and therefore, their performance over time based on their successes and failures [10].

This CAD systems can contribute immensely to the decision making in clinics and hospitals but also must meet some objectives to be successful. These objectives are managing large volumes of clinical data, objective and quantitative judgments and effectiveness and efficiency.

2.1.1 Managing large volumes of data

To provide accurate clinical diagnosis to a patient, medical professionals frequently must analyze various types of clinical data. Usually consist of the patient's clinical information, such laboratory test results, physical symptoms, imaging techniques and other findings. Other data to consider can be a patient´s medical history with his/her past medication records and the history of past diseases, social status, diet, smoking history, exercise habits, etc.

Clinical data keep becoming more sophisticated, complicated and increases dramatically in size and becomes too difficult for medical professionals to understand the whole spectrum of a patient´s conditions for them to diagnose the problem in time. Is this why CAD systems needs the capability to process and analyze the vast volumes of clinical data making use of optimized algorithms [10].

2.1.2 Objective and quantitative judgments

Traditional human-based diagnostic approaches mainly depend on the judgments of the healthcare professionals, which sometimes can be subjective [10]. These judgments can be affected by many reasons, from the inexperience of some new healthcare professionals to fatigue and distractions caused by overwork or night shifts. Moreover, human´s eye might be limited to find out pathological data that are already detected by radiological images. In our particular area of interest, it´s possible to have CT scans with small amounts of contrast within the aneurysm sac which may be missed by a human eye and may be an early but relevant endoleak. Under such circumstances, human errors are an unavoidable reality. Even under normal conditions, it is very difficult to quantify patients' information and accurately diagnose.

CAD systems in the other hand do not depend on a single healthcare professional´s analysis or skills, but in the knowledge obtained from large amounts of data, being capable of making more objective diagnosis consistently [10] and have the capabilities of finding said small amounts of contrast to make a more accurate early diagnosis.

2.1.3 Effectiveness and efficiency

CAD systems can be cost-effective, especially if a disease is detected in the early stages and can be treated before its evolution progresses into a more complex stage when treatment is more expensive and perhaps not as effective.

In the particular case of treated AAA, a close surveillance is mandatory to detect leaks and misplacements of the device that may condition the prognosis of the patient. Early detection of this complications clearly would facilitate their treatment with less technical requirements and fewer risks for the patients [11]. In addition, CAD may improve not only the early detection of diseases but increase the workflow of a diagnostic procedure [12].

Deep learning technology has been of a large importance in CAD systems state of the art because its capability to meet these objectives.

2.2 Deep Learning approaches for AAA classification

Although the task in hands is an image classification, the differences between typical datasets like ImageNet [13] and CT data volumes effects on the model proposed.

CT scans are groups of N images with size 512x512, N being the number of slices and variable for each patient. The idea behind the models studied is to extract the useful information about the aneurysm to reduce the amount of "noise" for our task. Then uses this information and evaluates them using a classifier.

The state of the art in machine learning algorithms for classification analysis of abdominal aortic aneurysms is divided into three different approaches.

First, S. Mohammadi et al [14] proposed an automatic algorithm that detects the aorta region along other regions of the abdomen (abdominal inside region, body borders and bones) from the CT scan using a patch extraction. This patch extraction is done by scanning through the CT slice with a window of 64x64 pixels. Then uses a CNN (table 2-2) to classify each patch into its corresponding label.

Table 2-1: CNN layers and parameters [14]

Layers	Training parameters
Data	Dim: 64 × 64 × 3 Batch size: 100 Data augmentation: mirror true
Conv 1	Output Num: 20 Kernel size: 5 Stride: 1 Pad:0
ReLU 1	Type: ReLU
Conv 2	Output Num: 30 Kernel size: 5 Stride: 1 Pad:0
ReLU 2	Type: ReLU
Pooling 1	Type: Max Pooling Kernel size: 2 Stride: 2
Conv 3	Output Num: 40 Kernel size: 7 Stride: 1 Pad:0
ReLU 3	Type: ReLU
FC 1	Output Num: 96 Type: InnerProducts
ReLU 4	Type: ReLU
FC 2	Output Num: 4 Type: InnerProducts
ReLU 5	Type: ReLU
Drop out 1	Drop out Ratio: 0.2
Accuracy	–
Loss	Type: SoftMax with loss

The second stage of the method proposed in this paper consist in using the Hough circles algorithm, detecting the shape of the aorta, defining its borders, and measuring its diameter in pixels for later conversion to millimeters. Depending on the output of the algorithm, this algorithm prints a risk evaluation message.

The results obtained by this method reaches high sensitivity, precision and accuracy in detecting the aorta border and measuring its diameter (98.41, 98.33, and 98.41% respectively).

Although this paper focuses on predicting the risk of developing an AAA instead of a post-surgery evaluation, the methods used for detecting the aorta and measuring the diameter could be of use in our application.

The second approach proposed uses an evaluation of the texture as features for the classification of the AAA post-EVAR. This approach classifies the CT scan in 2 labels, favorable and unfavorable evolution.

Texture analysis can be categorized into tree methods: statistical methods, model-based methods, and structural methods. Focusing on the statistical texture methods and specifically on spatial domain statistical techniques, we can adopt three grey-level matrices to extract texture features: the grey-level co-occurrence matrix (GLCM), the grey-level run length matrix (GLRLM) and the grey-level difference method (GLDM) [5][15].

With these methods we can extract features as the following, shown in Table 2-2.

Table 2-2: Detailed features in three greyscale matrices. Grey-Level Co-occurrence Matrix=GLCM. Grey-Level [15]

GLCM matrix	GLDM matrix	GLRLM matrix
Energy	Contrast	Short-run emphasis
Correlation	Angular second moment	Long-run emphasis
Inertia	Entropy	Grey level nonuniformity
Entropy	Mean	Run percentage
Inverse difference moment	Inverse difference moment	Run length nonuniformity
Sum average		Low-grey-level run emphasis High-grey-level run emphasis
Sum variance		
Sum entropy		
Difference average		
Difference entropy		
Two information measures of correlation		

Once extracted the features, the authors use a three-layer backpropagation neural network with a nonlinear sigmoid function as activation for each neuron. The network is trained to give 0.9 output value if the evolution is favorable or 0.1 if is unfavorable [5].

The results obtained with this approach shows are shown in table 2-3.

Table 2-3: AUC results of different texture methods for AAA evolution

Texture method	AUC_mean [5]	AUC_mean [15]

GLCM	0.977	0.901
GLDM	0.960	0.831
GLRLM	0.851	0.861

The third approach focuses on using deep learning models for both localization of the aorta in the CT volume, and classification of possible endoleaks on each slide. The localization scheme uses RetinaNet [16] to localize the aorta with a bounding box for each slide of the CT scan. Then the bounding box is extracted and used as a separate image, reshaped, and classified by a ResNet50 network [17]. The paper describing this approach shows positive results in classification, giving an 0.94±0.03 ROC AUC with data augmentation [18]. Although, to obtain these results, uses a big dataset for training and testing, reaching a final amount of 760 CT volumes and a total of 239,935 slices [18].

A later improvement of the algorithm [19] helps not only to find endoleaks but also to measure the aneurysm maximum diameter, AAA volume and endoleak volume using a 3D U-net [20] for a 3D segmentation of the aneurysm.

Convolutional neural networks are of great importance in image processing systems, including the detection and classification of the aorta region in AAA prediction systems. Although, in the presented approaches focuses on 2D CNNs, other areas in medical imaging uses 3D CNNs for a better understanding of the third dimensional space, and the relation of consecutive slices in the data volumes.

Some of these studies use 3D convolutions in the feature extraction. This scheme focuses on these three-dimensional spatial relations. The main problems with this approach are the computational power needed for a fast training and the input dimensions, which could change between patients. In this paper [21] the authors reshape the input scan to a 32x32x32 volume using the linear interpolation method, but the volumes sizes treated in this master thesis are in the order of 512x512xN, N being between 200-1600, and a reshape of this magnitude can represent a great loss of information.

Another approach is mixing 2D and 3D convolutions in a network as proposed by R. L. Draelos et al in their CT-net [22]. Extracting features using a pretrained ResNet18 with

ImageNet on to multiple stacks of 3 slices treated as 3 channel images. The features extracted are then concatenated and processed with several 3D convolutional layers. Finally, a classifier using three fully connected layers to obtain a final prediction. The pipeline of this model is represented in figure 2-1

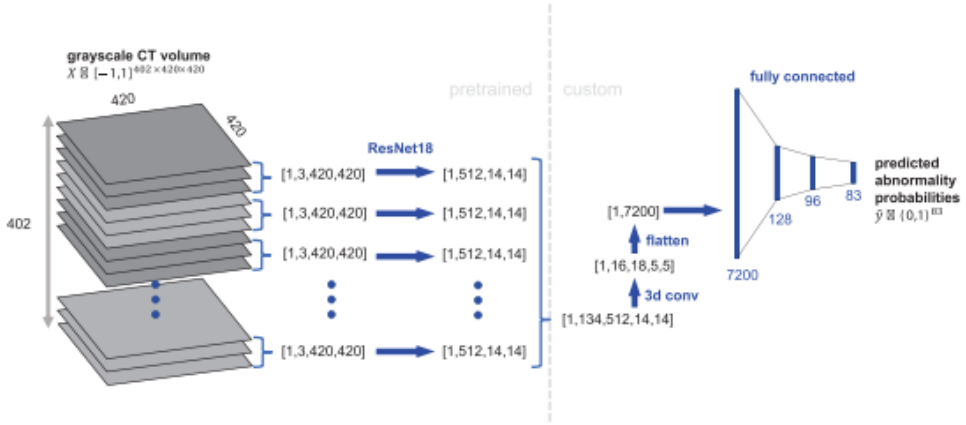

Figure 2-1: CT-Net volume classification architecture [22]

3 Design

The Design decided for the architecture of the model follows the idea behind S. Hahn et al's work [18][19]. Dividing the task into 2 different architectures, the first one focusing on localizing the aorta section from each slide of the CT volume where an aneurysm is present. The second one focuses on the classification task, using different feature extraction networks, from multiple ResNet architectures to a custom 3D architecture.

3.1 Localization Task

The idea behind this task is to reduce the input size used in training. Because the dimensions of each patients CT volume, training will be slow and ineffective, due the large number of slices and the existence of slices that don't contain information about the EVAR or the aneurism. The procedure planned is to train a RetinaNet model to localize and set a squared bounding box around the aorta region. Then, this detection is going to be used in the classification task as the input image.

The objective on this task is not obtaining perfect results in localization but a clear region where the aorta is present for later use. The main intention for this stage is not obtaining a precise bounding box but ensuring that all the slices where the aorta is present is localized with a bounding box, especially if the AAA is visible too.

To ensure the correct functioning of the localization module, there is also a post-processing phase. In this phase there are three modifications to the direct output of the RetinaNet module: Padding the results, interpolation between slices and removing isolated predictions.

Padding increases the bounding box by 10 pixels in each direction, ensuring that the aorta is fully included in the bounding box. This is effectively adding noisy pixels to the detection, but because of the typical localization of the endoleaks, detecting the walls of the aneurysm is of grate importance for the later feature extraction.

Interpolating between detections from consecutive slices is necessary to reduce the number of false negatives. To include slices where should be a detection of the aneurysm, the system decides the values of the corners of the bounding box by averaging the positions of previous and posterior bounding boxes. Because of the possible existence of false

positives, we should not interpolate slices without bounding boxes in a range from the current.

To reduce the number of false detections in regions far from the aneurysm, if the detection is isolated from other detections of the volume, it is not considered and removed from the output.

3.2 Classification Task

Classification is the main objective of the system. Deciding if the patient should require a second intervention. In this system´s area, the plan is to train a CNN that inputs each detection from the localization task and passes through the model deciding if the growth is favorable or unfavorable.

The first approach is to follow S. Hahn et al's papers [18][19] and use a ResNet model to classify each of the slices of the CT volume, and then using a voting mechanism to decide the final case predictions. However, their approach had individual labels for each slice of the CT scan representing the existence of leaks in said slice, while our dataset count with one label for the complete data volume. Further explain the characteristics of the dataset in chapter 4.1.

Another approach is to use 3D CNNs for all the CT volume, obtaining a final case prediction directly from the network. This brings the before mentioned problem of variable sized volumes present in our dataset. To solve this problem, the input of the network uses a fixed size in the three dimensions, using a window to select which slices are selected for each iteration, and use a voting algorithm to return a final and unique value to the complete CT volume.

The quantity of non-redundant data available for training a CNN is a concern when using data volumes of medical images, being similar between consecutive slices and between different patients' scans. Therefore, is important to use data augmentation techniques to increase the variability of the training data, being an instrumental data-processing step to achieve a more generalizable and accurate performance [23]. The data augmentation techniques planned for the classification task are catalogued as basic augmentation, performing to the input a series of small random translations, random rotations and random flips in the horizontal and/or vertical axis.

4 Development

4.1 Creating the dataset

The data used in this project was gathered from the hospital "La Paz" university hospital, with permission from the ethics committee. The data was downloaded directly from the hospital system, one CT scan at a time. The final dataset gathered consisted of 88 non-labeled CT volumes from 44 different patients. These CT volumes were taken from patients operated with an EVAR from 2010 to 2012. For each patient could obtain a CT scan from seven days after the surgery and a second scan performed from six to twelve months later.

This dataset needs labels for both the localization and the classification task.

For localization, a vascular surgeon created a ground truth bounding box around the AAA and the aorta region surrounding the aneurysm for 25 patients (49 CT volumes) using a labeling open-source software called LabelImg [24] and a modification called Video Frame Mode [25] to improve labeling efficiency.

In the case of classification, obtaining labels is complicated. For some patients who had a clear contrast peak in their CT scan, we can obtain their label from the case reports filed at the time of the CT scan obtention. Although, for the other patients, in other to know if their EVAR is leaking we analyze their evolution measuring the diameter of both CT scans in possession and write the corresponding label based on the evolution of the AAA size. Six patients were discarded from this stage because lack of information needed to label their condition, ending with 38 CT scans for the classification step.

4.2 Localization

The CT volumes of the dataset are saved on disk as a number of .DICOM files, one for each slice, sorted in a folder and named after their corresponding position on the volume.

To read these files we make use of the dcmread function implemented in the pydicom library [26] for python. Once read the image corresponding to the slice, we start a pre-processing stage to normalize the image to the range [0,1].

To train the model we make use of the 49 CT labeled volumes.

Training for this model was based on the intersection over union evaluation between the bounding boxes presented in each iteration and the corresponding ground truth. RetinaNet not only finds the bounding box around the object but also returns a confidence related score of the type of object detected. After a result is given by the model, we filter the output to detections with a confidence level over a threshold of 0.9.

Once trained the model, we develop a script that returns a list with the directories of the .DICOM files that present a detection and the position of the 4 corners of the bounding box and uses this list for the post-processing phase described in a previous chapter. The post-processing output is a data frame containing the directories and the position of the final detections, which will be used for the classification task.

The training of this model for twenty epochs lasted eight hours using a personal laptop with a NVIDIA GeForce RTX 2060 graphics card.

4.3 Classification

For classification two different approaches were developed. A ResNet network, both 18 and 50 layers were attempted, and a custom 3D network.

The first step in developing the classification scheme is to pre-process the output of the localization algorithm to adapt to the input of our network.

For ResNet architectures we needed to extract from each slice of the multiple volumes the directory to the DICOM file and the position of the corresponding bounding box. Then extract the detected region and reshape it into a 64x64 image.

For the 3D CNN architecture, we also need to group these images in packs of a fixed number of slices, all from the same CT scan. The output are multiple 64x64x32 data volumes. For both 2D and 3D CNNs, data augmentation techniques are applied as specified in the design.

The custom 3D architecture consists of three custom 3D ResNet blocks. These blocks are made of two 3D convolutional layers, two 3D batch normalization layers and two ReLU activation layer. Organization of the layers in the model is shown in figure 4-1.

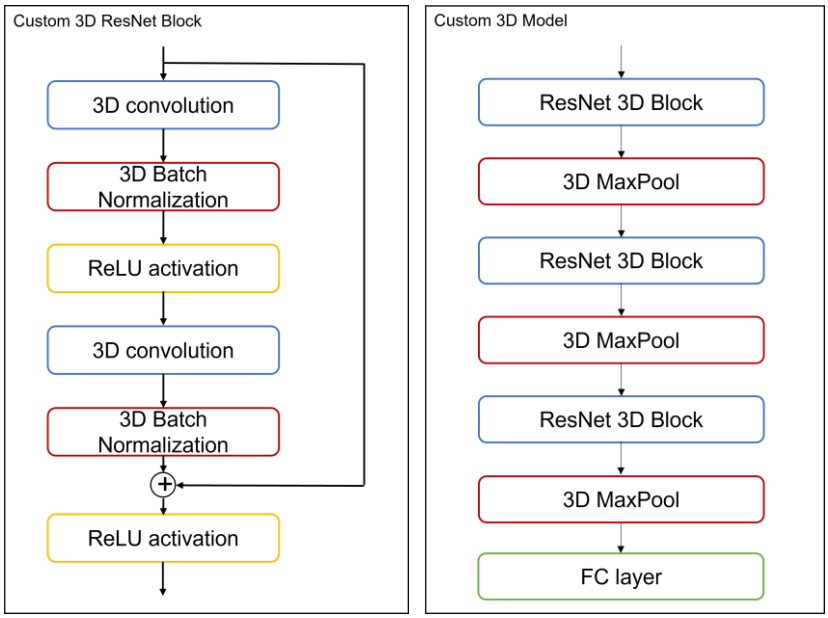

Figure 4-1: Organization of layers in the 3D custom model

To implement the voting system, we decided to use a probability pondered voting system with the output returned by the network. First, we analyze the output of the model by obtaining the mean probability for each label, favorable or unfavorable evolution. This means that if a small number of slices present a peak in contrast, alerting the existence of an endoleaks, and the model defines that slice with high probability of an unfavorable evolution the voting system will consider this as the correct output even if most of the slices are classified as the opposite label. We also developed a majority voting system for comparison.

Training each model for 25 epochs lasted fourteen hours each using the same GPU. The dataset used in classification counted with 28 CT scans for training and 10 for testing.

All the algorithms are implemented in python, using the pydicom library [26] for reading thee .DICOM files and an extensive use of the pytorch library [27].

5 Integration and experimental results

5.1 Localization

Localization task, as mentioned before, has the objective of selecting a rough estimation of the localization of the aorta. The exact position of the bounding boxes is not as important as the precision, selecting the minimum possible false negatives and detecting all the true positives.

Running the algorithm used for localization and evaluating its output we obtain a precision value of 82.015% with 2695 True positives and 591 False positives. This means that the system still has a considerably high number of false detections even with the post-processing applied. This affects future utilization of this data, making more complex the training for classification.

In figure 5-1 and figure 5-2 we can see the evolution of the precision and the loss through training. Qualitative results are closer to the expectations, represented in figure 5-2.

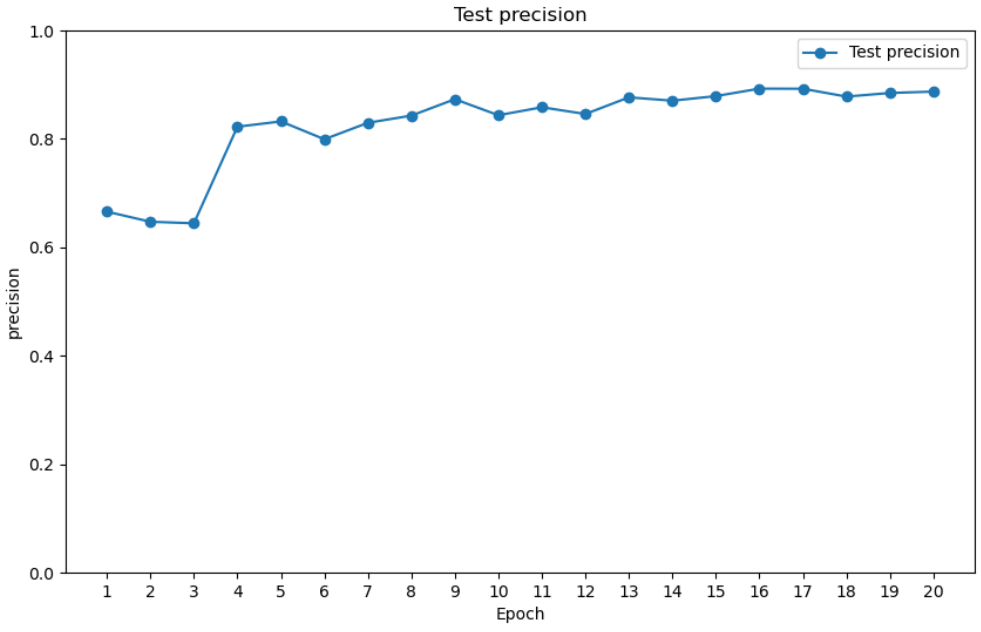

Figure 5-1: Localization test precision

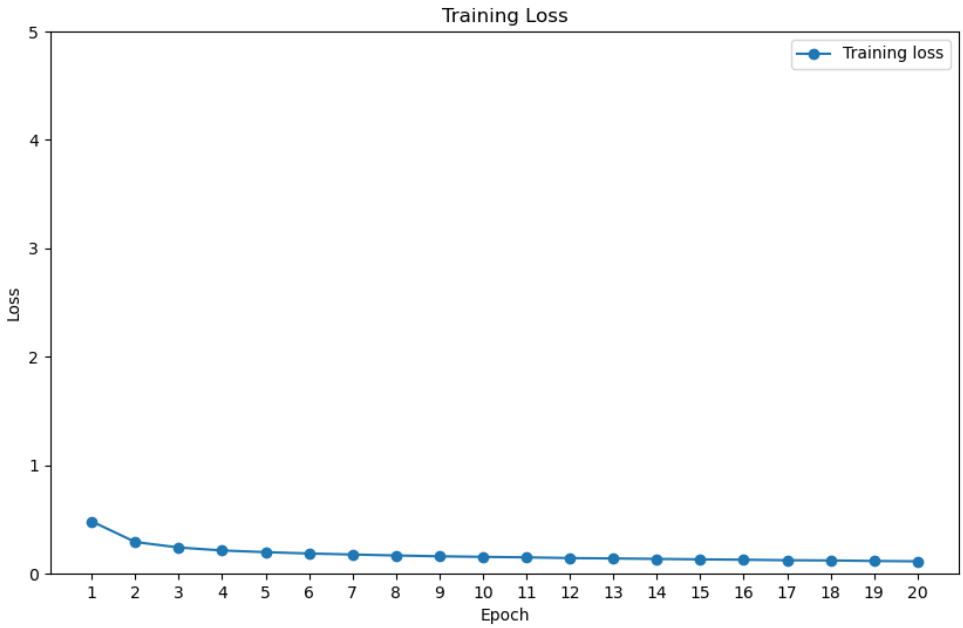

Figure 5-2: Localization training loss

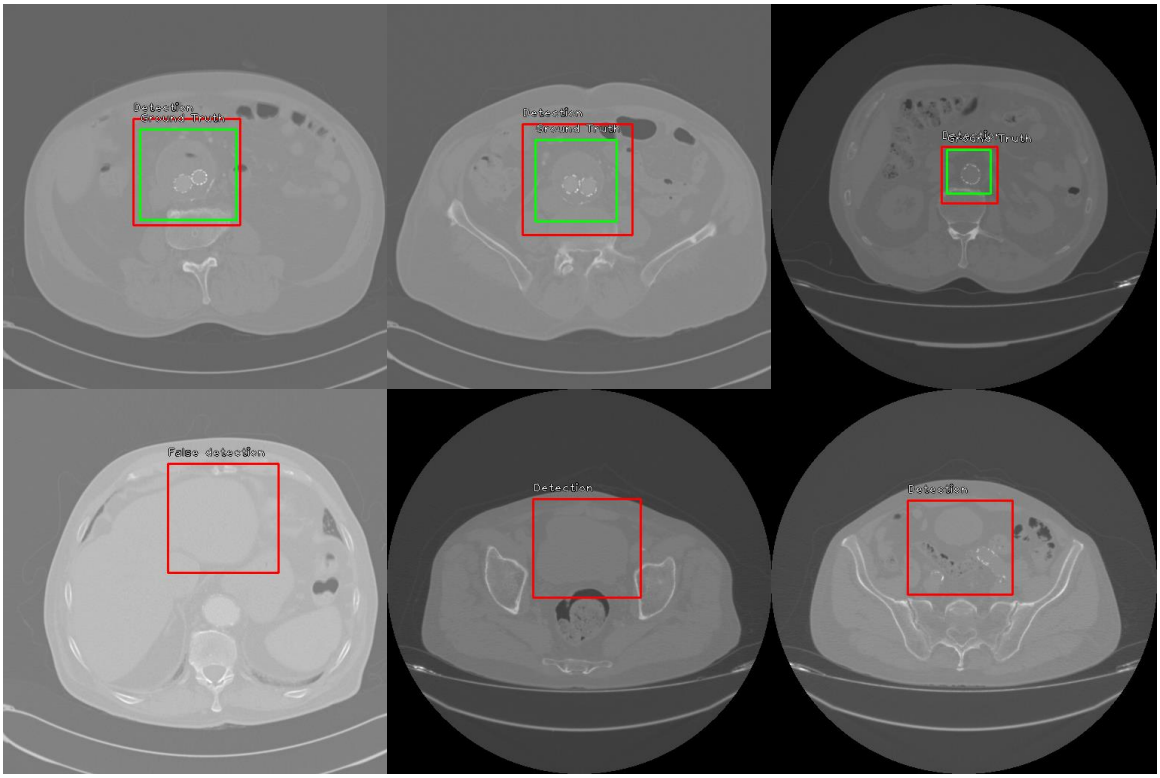

Figure 5-3:In red final detections made by the system, in green the ground truth. Top row true positives, bottom row false positive

5.2 Classification

The training of the classifiers was affected by the small number of available CT scans for training and the false detections made in localization, dragging the error to this step. Training graphs show the effect of overfitting in training for both architectures.

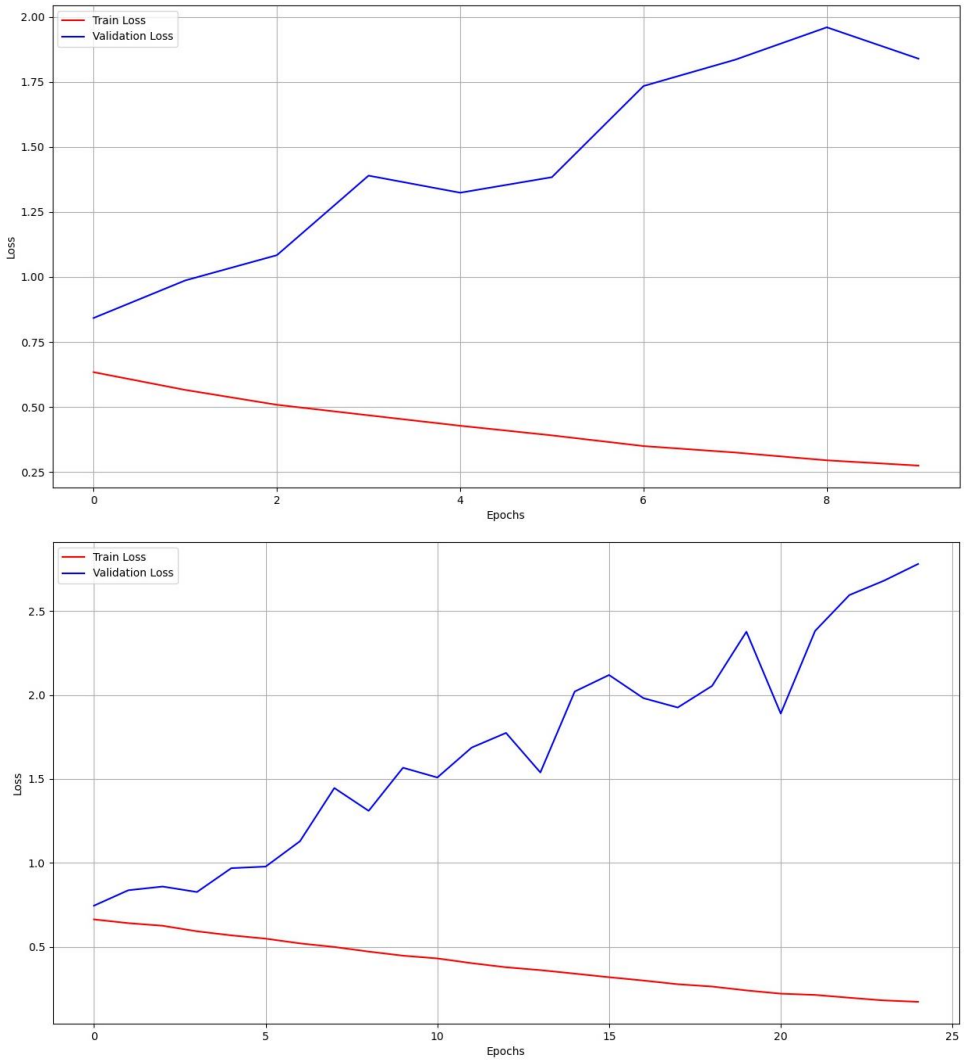

Figure 5-4: ResNet training losses. Up ResNet18, down ResNet50

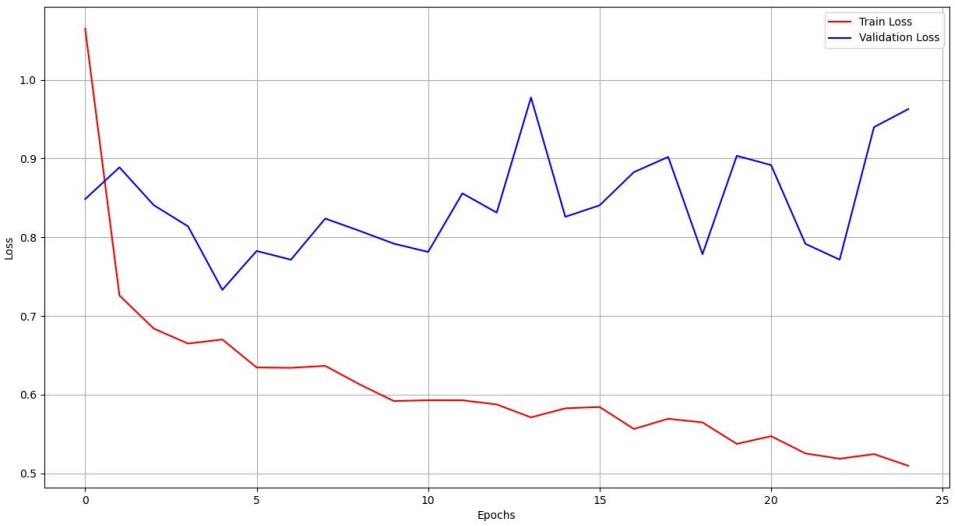

Figure 5-5: Custom 3D CNN training losses

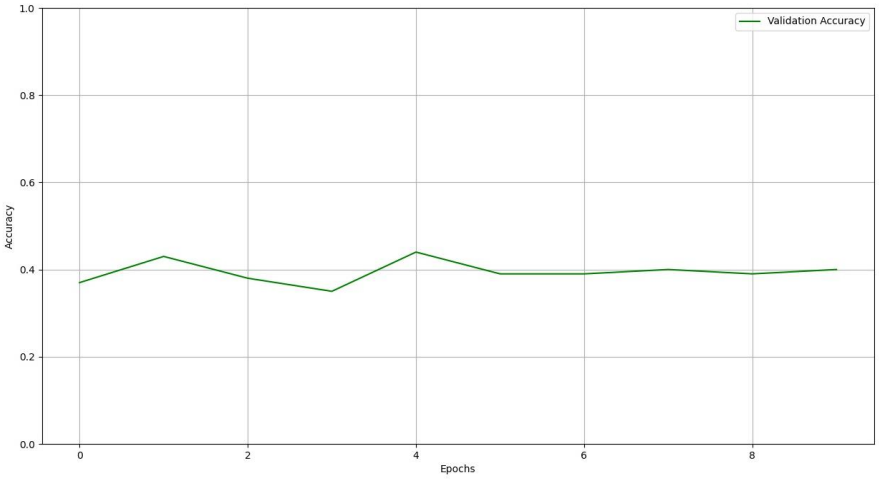

Figure 5-6: Validation accuracy for ResNet18

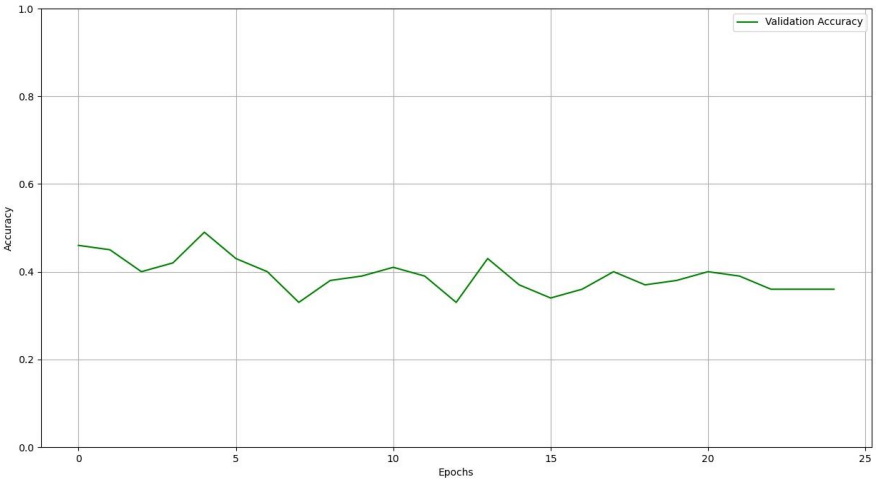

Figure 5-7: Validation accuracy for ResNet50

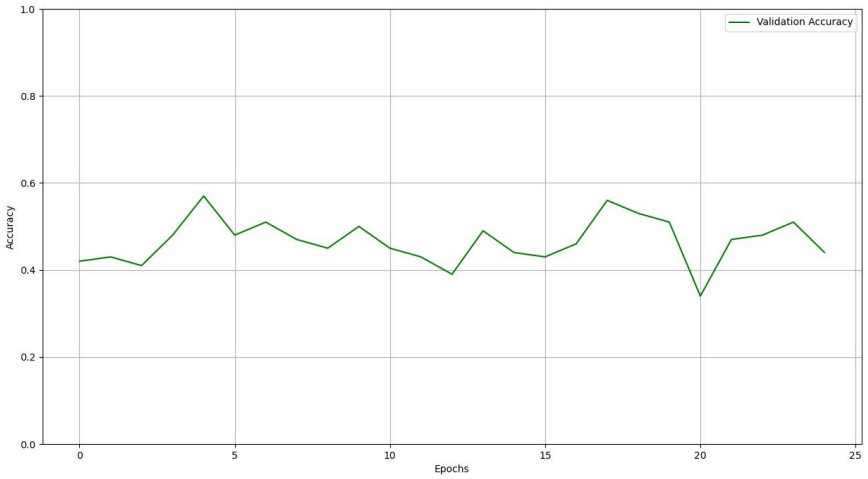

Figure 5-8: Validation accuracy for Custom 3D CNN

The classifier is affected by overfitting, obtaining unvaluable results. Both models learn to return the same output for each iteration, either that all the CT scan belongs to a favorable evolution or the opposite. This explains why the accuracy graph is surrounding the 50% value.

6 Conclusions and future work

6.1 Conclusions

This Master thesis attempted to develop a CAD system to help detect endoleaks in post-operative EVAR patients, helping the medical professionals to plan an early correction of the stent and reducing the risk of rupture.

The system was based in a two-step pipeline. The first step consisted in a localization DL model using RetinaNet, detecting the aorta in each slice of the CT volume, especially the Aneurysm. The second step consist of a classification model, using both a ResNet architecture and a 3D custom CNN model. This step is focused in extracting the features of the slices and predict between a favorable or unfavorable evolution for the patient.

Although the idea of using computer algorithms to help medical experts to aid their diagnosis is becoming an important research field in the world, this concept is not extended in hospitals in Madrid. The use of CAD systems in hospitals is increasing due their capability to process large amounts of data, present objective results and presenting a high effectiveness.

The system show a large overfitting between the model and the data used in training, obtaining a poor performance when used with a validation set. This overfitting can be led by many things, starting by a small dataset available for training, a limited performance in the localization step and the high complexity of the input data.

This study was neither able to develop a system capable of predicting the evolution of an AAA, nor the contrast peaks from the CT scan, signaling the existence of an endoleak.

From the secondary objectives, the creation and labeling of the dataset was a high time-consuming task and was completed successfully for a small number of patients (44) considering the task in hand. The system is not ready yet to detect with sufficient accuracy endoleaks. The limitation of this system is the need of larger amount of data from patients to be trained enough to be useful in the medical day practice.

6.2 Future work

Future research on this topic should start with a better-defined dataset, with more accurate labels and more patients to train with. With more data we could train better the

models designed, starting from the localization task, focusing on reducing the number of false positives returned.

Classification models used are not optimal for the task considering the limitations of the dataset and labels. Researching on new models and investigating their capabilities to extract the correct features could be helpful to the classification task.

The next area of work is the use of overfitting reduction techniques, as more complex data augmentation, pretrain the models with similar datasets or using self-supervised learning.

Bibliography

[1] K. C. Kent, "Abdominal Aortic Aneurysms," *N. Engl. J. Med.,* vol. 371, *(22),* pp. 2101-2108, 2014. Available: https://doi.org/10.1056/NEJMcp1401430.

[2] J. Raffort et al, "Artificial intelligence in abdominal aortic aneurysm," Journal of Vascular Surgery, vol. 72, (1), pp. 321-333. e1, 2020.

[3] R. L. Draelos, "Automatic Interpretation of Chest CT Scans with Machine Learning," Glass Box Medicine, 2020

[4] Cuong et al, "Porosity Estimation from High Resolution CT SAN Images of Rock Samples by Using Housfield Unit". Open Journal of Geology., vol.8, (10), pp. 1019-1026. 2018. Available: https://doi.org/10.4236/ojg.2018.810061.

[5] G. García et al, "Evaluation of Texture for Classification of Abdominal Aortic Aneurysm After Endovascular Repair," J. Digital Imaging, vol. 25, (3), pp. 369-376, 2012. Available: https://doi.org/10.1007/s10278-011-9417-7.

[6] O'Donnell, T. F. et al, "Select early type IA endoleaks after endovascular aneurysm repair will resolve without secondary intervention," Journal of vascular surgery, vol. 67, (1), pp. 119-125. 2018.

[7] Deery, S. E. et al, "Aneurysm sac expansion is independently associated with late mortality in patients treated with endovascular aneurysm repair," Journal of vascular surgery, vol. 67, (1), pp. 157-164. 2018.

[8] O'Donnell, T. F. et al, "Aneurysm sac failure to regress after endovascular aneurysm repair is associated with lower long-term survival," Journal of vascular surgery, vol. 69, (2), pp. 414-422. 2019.

[9] B. Rengarajan et al, "A Comparative Classification Analysis of Abdominal Aortic Aneurysms by Machine Learning Algorithms," Ann. Biomed. Eng., vol. 48, (4), pp. 1419-1429, 2020. Available: https://doi.org/10.1007/s10439-020-02461-9.

[10] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," Expert Syst. Appl., vol. 138, pp. 112821, 2019. Available: https://doi.org/10.1016/j.eswa.2019.112821.

[11] Rand, T. et al, "Quality improvement guidelines for imaging detection and treatment of endoleaks following endovascular aneurysm repair (EVAR)," Cardiovascular and interventional radiology, vol. 36, (1), pp. 35-45. 2013.

[12] C. D. Lehman et al, "Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection," JAMA Intern Med, vol. 175, (11), pp. 1828-1837, 2015. Available: https://doi.org/10.1001/jamainternmed.2015.5231.

[13] Deng J et al, "Imagenet: A large-scale hierarchical image database," 2009 IEEE conference on computer vision and pattern recognition. p. 248-255. 2009.

[14] S. Mohammadi et al, "Automatic Segmentation, Detection, and Diagnosis of Abdominal Aortic Aneurysm (AAA) Using Convolutional Neural Networks and Hough Circles Algorithm," Cardiovascular Engineering and Technology, vol. 10, (3), pp. 490-499, 2019. Available: https://doi.org/10.1007/s13239-019-00421-6.

[15] N. Ding et al, "CT texture analysis predicts abdominal aortic aneurysm post-endovascular aortic aneurysm repair progression," Scientific Reports, vol. 10, (1), pp. 12268, 2020. Available: https://www.ncbi.nlm.nih.gov/pubmed/32703988.

[16] T. Lin et al, "Focal loss for dense object detection," in Proceedings of the IEEE International Conference on Computer Vision, 2017.

[17] K. He et al, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[18] S. Hahn et al, "Deep learning for recognition of endoleak after endovascular abdominal aortic aneurysm repair," 2019, Available: https://ieeexplore.ieee.org/document/8759187.
[19] S. Hahn et al, "Machine deep learning accurately detects endoleak after endovascular abdominal aortic aneurysm repair," JVS-Vascular Science, vol. 1, pp. 5-12, 2020. Available: http://dx.doi.org/10.1016/j.jvssci.2019.12.003
[20] Ronneberger et al. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. 2015.
[21] T. T. Ho et al, "A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects," Scientific Reports, vol. 11, (1), pp. 34, 2021. Available: https://doi.org/10.1038/s41598-020-79336-5.
[22] R. L. Draelos et al, "Machine-learning-based multiple abnormality prediction with large-scale chest computed tomography volumes," Medical Image Analysis, vol. 67, pp. 101857, 2021. Available: http://dx.doi.org/10.1016/j.media.2020.101857.
[23] Abdollahi B. et al, "Data Augmentation in Training Deep Learning Models for Medical Image Analysis," Deep Learners and Deep Learner Descriptors for Medical Applications. Intelligent Systems Reference Library, vol 186. 2020. https://doi.org/10.1007/978-3-030-42750-4_6.
[24] Tzutalin, "labelimg," https://github.com/tzutalin/labelImg, 2015.
[25] jochengietzen, "labelimg" https://github.com/jochengietzen/labelImg, 2019.
[26] Mason, D. L., et al, "pydicom: An open source DICOM library," https://github.com/pydicom/pydicom
[27] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; pp. 8024-8035. 2019.
[28] C. Chen et al, "Deep Learning for Cardiac Image Segmentation: A Review," Front Cardiovasc Med, vol. 7, pp. 25, 2020. Available: https://doi.org/10.3389/fcvm.2020.00025.
[29] L. Mesin, "Biomedical Image Processing and Classification," Electronics-Switz, vol. 10, (1), pp. 66, 2021. https://doi.org/10.3390/electronics10010066.
[30] Chun Hei Michael Chan, "Pytorch: Step by Step implementation 3D Convolution Neural Network," Towards data science, 2020.
[31] J. A. Fries et al, "Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences," Nat Commun, vol. 10, (1), 2019. Available: https://doi.org/10.1038/s41467-019-11012-3.

Glossary

AAA	Abdominal Aortic Aneurysm
CT	Computerized Tomography
EVAR	EndoVascular Aneurysm Repair

Appendix

A Ethnics committee permission

Ref: 57/368728.9/20

Hospital Universitario La Paz
Hospital Carlos III
Hospital Cantoblanco

INFORME DEL COMITE DE ÉTICA DE LA INVESTIGACIÓN

Dª Almudena Castro Conde, Presidenta del Comité de Ética de la Investigación con medicamentos del Hospital Universitario La Paz

CERTIFICA

Que este Comité ha evaluado el Trabajo de Fin de Master del alumno del Grado de Ingeniería de Telecomunicaciones y del Master de inteligencia Artificial de la Universidad Autónoma de Madrid, Javier Riera del Moral, titulado: "**USO DE TECNOLOGÍAS DE APRENDIZAJE PROFUNDO PARA ANÁLISIS DE LA REDUCCIÓN DEL SACO ANEURISMÁTICO TRAS EVAR**", Protocolo **HULP-VAS-CT-IA-2020**, versión 1 de 27 noviembre 2020; código HULP: **PI-4559**,

Y considera que,

- El procedimiento de recogida de datos mantiene la confidencialidad de los datos de carácter personal y se cumplen los requisitos éticos y legales exigibles para este tipo de estudios.
- La capacidad de la investigadora y los objetivos del proyecto son apropiados para llevar a cabo el estudio.

Y que este Comité acepta que dicho estudio sea llevado a cabo por Javier Riera del Moral con la supervisión del Dr. Luis Riera del Moral del Servicio de Angiología y Cirugía Vascular del Hospital Universitario La Paz, como tutor responsable del Trabajo de Fin de Master

Lo que firmo en Madrid a 29 de diciembre de 2020

Firmado digitalmente por: FERNANDEZ DE UZQUIANO MARIA EMMA
Fecha: 2020.12.30 08:37

Firmado:
Dª Almudena Castro Conde
Presidenta del CEIm

P.O. Dª Emma Fernández de Uzquiano
Secretaria Técnica del CEIm