N° Ordre...../FT/UMBB/2023

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE M'HAMED BOUGARA-BOUMERDES

Faculté Technologie

Thèse de Doctorat

Présentée par :

Amel Laidi

En vue de l'obtention du diplôme de **DOCTORAT - LMD** en **:**

Filière : Génie Biomédical

Spécialité : Instrumentation Biomédicale

Segmentation et Analyse des Séquences Cardiaques par les techniques d'apprentissage en profondeur (Deep Learning)

Devant le jury composé de :

Mr	Rahmoune	Fayçal	Prof.	Univ. Boumerdes	Président
Mr	Ammar	Mohammed	Prof.	Univ. Boumerdes	Directeur
Mr	El Habib Daho	Mostafa	MCA	Univ. Tlemcen	Co- Directeur
Mr	Harrar	Khaled	MCA	Univ. Boumerdes	Examinateur
Mr	Lazouni	Mohamed Elamine	MCA	Univ. Tlemcen	Examinateur
Mr	Omari	Tahar	MCA	Univ. Boumerdes	Examinateur

Année Universitaire 2022/2023
Soutenue le 16 Mai 2023

N° Ordre...../FT/UMBB/2023

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

M'HAMED BOUGARA UNIVERSITY - BOUMERDES

Faculty of Technology

PhD Dissertation

Presented by :

Amel Laidi

In order to obtain the degree of **PhD - LMD** in **:**

Field: Biomedical Engineering

Speciality: Biomedical Instrumentation

CARDIAC STRUCTURES SEGMENTATION AND ANALYSIS USING DEEP LEARNING TECHNIQUES

Defense committee members :

Mr	Rahmoune	Fayçal	Prof.	Univ. Boumerdes	President
Mr	Ammar	Mohammed	Prof.	Univ. Boumerdes	Supervisor
Mr	El Habib Daho	Mostafa	MCA	Univ. Tlemcen	Co- Supervisor
Mr	Harrar	Khaled	MCA	Univ. Boumerdes	Examinator
Mr	Lazouni	Mohamed Elamine	MCA	Univ. Tlemcen	Examinator
Mr	Omari	Tahar	MCA	Univ. Boumerdes	Examinator

Academic Year 2022/2023

Defended on May 16th, 2023

Acknowledgements

Words cannot express my gratitude to my supervisor **Prof. Mohammed Ammar**, for his invaluable patience and feedback, through all stages of research.

I am extremely grateful to my co-advisor **Dr. Mostafa El Habib Daho,** without whom this manuscript would not be a coherent final product. And to my co-author **Dr Mahmoudi Saïd** for his contribution in all joined publications.

I offer my appreciation to the president of the defense committee **Prof. Fayçal Rahmoune**, and the members of the defense committee **Dr. Khaled Harrar** , **Dr. Mohamed Elamine Lazouni** , and **Dr Tahar Omari** for kindly taking the time to review and discuss my work.

I offer my thanks to all my fellow lab mates, especially Khadidja Messaoudene and Aicha Benyoucef, for their constant counseling and feedback and moral support.

I would like to thank my parents, my brothers, and my friends for being there for me every step of the way, always pushing me forward, always believing in me.

Lastly, I would like to thank everyone who had offered any kind of help, big or small, and made this thesis possible.

Dedication

PhD has a reputation for being a dark path. The first beam of light is the supervisor, they can make or break your experience. In that I was most fortunate, Mr. Ammar was the ultimate mentor, and for him I will be forever grateful.

For my parents and my brothers Kamel and Rahim, who always pushed me and believed in me even when I doubted myself…

For my friends, my second family, who were there through the breakdowns and failures, telling me I am better than I believed I was, listening to my complains, easing my fears, and rationing me into impossible situations…

They say, "When the snow falls and the white winds blow, the lone wolf dies but the pack survives".

For everyone who was my pack…

I dedicate this work...

Abstract

In this thesis we work on two of the most urgent research problems surrounding the use of deep learning in cardiac applications: the shortage of labeled data, and the lack of trustworthy Artificial Intelligence (AI) models.

We first tested the impact of using synthetic data as a solution for dataset imbalance on the task of atherosclerosis screening from Coronary CT Angiography images. We recorded an improvement in sensitivity from 60.8% to 89.0%, an unprecedented performance.

For the second part of the thesis, we focused on heart structure identification using a novel AI model that mimics doctor reasoning; creating a mind map that describes every step the doctor uses when manually segmenting and identifying the heart structures, then replacing each step with a corresponding machine learning algorithm, making the model trustworthy.

Keywords: Deep Learning; Cardiac Segmentation; Atherosclerosis; Classification; Transfer Learning.

Résumé

Dans cette thèse, on travaille sur deux grandes problématiques dans l'utilisation de l'apprentissage approfondi dans les applications cardiaques: le manque de données étiquetées, et le manque d'explicabilité des modèles d'intelligence artificielle (IA).

On a d'abord étudié l'impact d'utiliser des données synthétiques pour gérer le déséquilibre d'un ensemble de données sur la tâche de dépistage de l'athérosclérose à partir d'images d'angiographie coronarienne. On a noté une amélioration de la sensibilité de 60,8% à 89,0%, une performance sans précédent dans la littérature.

Dans la deuxième partie, on travaille sur l'identification des structures cardiaques à l'aide d'un nouveau modèle d'IA qui imite le raisonnement des médecins. On a suivi un organigramme qui décrit chaque étape fait par le médecin lors de la segmentation et de l'identification manuelles des structures cardiaques, en remplaçant chaque étape par un algorithme d'apprentissage automatique approprié.

Mots Clé : Apprentissage Approfondi (Deep Learning) ; Segmentation Cardiaque ; Athérosclérose ; Classification ; Apprentissage par Transfer.

ملخّص

في هذه الأطروحة نناقش مشكلتين يتمّ مواجهتهما عند استخدام التعلّم العميق في التّطبيقات الطّبية: نقص البيانات المصنَّفة ، والإفتقار إلى القابلية للتفسير في نماذج الذكاء الاصطناعي **(AI)**

اختبرنا أولاً تأثير إضافة بيانات مُستحدثة غير حقيقيّة لتعويض نقص البيانات في مهمة تصنيف حالات تصلّب الشرايين من صور تصوير الأوعية التاجية المقطعية .**(CCTA)** سجّلنا تحسناً في حساسية التصنيف من **60.8**٪ إلى **89.0**٪، وهو أداء مستحسن غير مسبوق.

في الجزء الثاني من الأطروحة، ركّزنا على التعرف على بنية القلب باستخدام نموذج ذكاء اصطناعي جديد يقلد تفكير الطبيب. عبر انشاء خطة تصف كل خطوة يستخدمها الطبيب عند تأدية المهمة يدويًا، ثم استبدال كل خطوة بخوارزمية تعلم آلي مناسبة.

الكلمات المفتاحية : التعلم العميق ؛ تصلب الشرايين ؛ التصنيف .

Table of Contents

List of Figures

List of Tables

General Introduction

Cardiovascular diseases are a group of diseases that attacks the heart, the coronary arteries, and the blood vessels. They are considered the number one cause of death globally. Over 17 million people die from them every year, which is on the rise, according to the world health organization [1]. The diagnosis and follow-up of such disease are made by professional clinicians, with the help of different medical imaging techniques, that includes Echocardiography, Computer Tomography, Angiography, cardiac MRI, etc,

Even though the advances in cardiac imaging have made an immense improvement in the follow-up of cardiac diseases, they also provide large numbers of images that take significant time to be analyzed, even with an expert's eye. This can be a problem due to the significant lack of medical professionals; especially after the COVID-19 pandemic, the medical industry has a big drop in the number of professionals. Therefore, automated solutions have been proposed and adapted in the field.

The rise of deep learning has revolutionized every field, and that does include healthcare. Cardiac applications have benefited a lot from deep learning advances. The literature is full of ground-breaking results brought by the use of deep learning, and Artificial Intelligence (AI) in general, in cardiac image segmentation, classification, and reconstruction. However, in order to take these achievements from academia to industry and benefit from them properly, a few problems need to be answered:

- Deep learning is known for its big appetite for data. However, annotated medical images are not widely available, which has always been a problem. Is it possible to find a solution? By performing innovative data augmentation techniques, can we perform cardiac disease screening successfully without needing more labeled data?

- If cardiac applications are to be used by non-experts, there would be no one to supervise the final result provided by the computer. In that case, deep learning techniques must be completely trustworthy.

Our research focuses on these two problems limiting the proper use of deep learning for cardiac applications: lack of annotated data and trustworthy AI.

Our contribution to the first problem is in proving that is possible to use generative models such as GAN to generate new synthetic data and use it to balance imbalanced datasets. While for the second problem, we came up with a trustworthy AI workflow that helps with the task of structure segmentation from Cardiac MRI images.

This thesis is outlined as follows:

In chapter 1, we provide the medical context of the research for a better understanding of the terminologies used further down the thesis. Then, we describe the heart anatomy and the blood vessel system, a general description of the most common cardiovascular diseases, and finally, and then we give the existing imaging modalities used for cardiac imaging. This chapter shows the complicated aspects of cardiovascular diseases and the numerous techniques available for cardiac imaging, all providing an increasingly overwhelming amount of data.

In chapter 2, we introduce the fundamentals of machine learning and deep learning, where it all started, and how it turned into a powerful AI tool. Then in section 2.2, we talk about the state-of-the-art deep learning models in cardiac image segmentation and classification to give an idea of the existing research related to our study and how it inspires our research questions.

In chapter 3, we work on atherosclerosis screening from Coronary Computed Tomography Angiography (CCTA) images using transfer learning; we test different types of pre-trained models, with different combinations of hyperparameters. Next, we study the influence of data imbalance on classification; we propose using Generative Adversarial Networks (GANs) for data augmentation to deal with the data imbalance and improve the model's performance. Parts of this chapter were published as in [2].

Chapter 4 proposes a new approach for a trustworthy model for cardiac segmentation and identification. We attempted to replicate the steps taken by a human expert from the

moment they see the image to the moment they decide on its segmentation, replacing each step with a machine learning technique. This chapter was partially published in [3].

We finish up with a general conclusion, and research perspectives.

Chapter 1

Medical Context

1 Medical Context

Given the nature of our research, the remainder of the thesis will involve a lot of medical terms. To make them easy to understand, we present a medical context. This chapter includes general information of the heart position and anatomy, the vascular system, a description of most common cardiovascular diseases and current heart imaging modalities.

1.1 Heart Anatomy

1.1.1 Presentation, Position and Dimensions of the Heart

The heart, a hollow muscular organ of conical shape, is the most important muscle in the body. It essentially provides the blood circulation system with indispensable pressure. Despite its power, the heart is quite small in size, approximately the size of a human fist. It is 12 cm in length, 9 cm in width at its widest point, and 6 cm thick. It weighs only around 300 grams. [4]

The heart is vertically divided by the septum into the right heart and the left heart, each half is in turn divided into two chambers, atrium and ventricle which makes the heart composed of four chambers: right atrium (RA), left atrium (LA), right ventricle (RV); and left ventricle (LV). It is located between the lungs, near the center of the rib cage, with about two thirds of its mass lying on the left of the body's midline, as shown in Figure 1-1. The heart rests on the diaphragm in a space called the mediastinum, surrounded by a fluid-filled sac called the pericardium.

The heart muscle is what produces the electrical signals that cause the heart to contract, pumping blood throughout the whole body. The combination of the heart and circulatory system is what is referred to as the cardiovascular system.

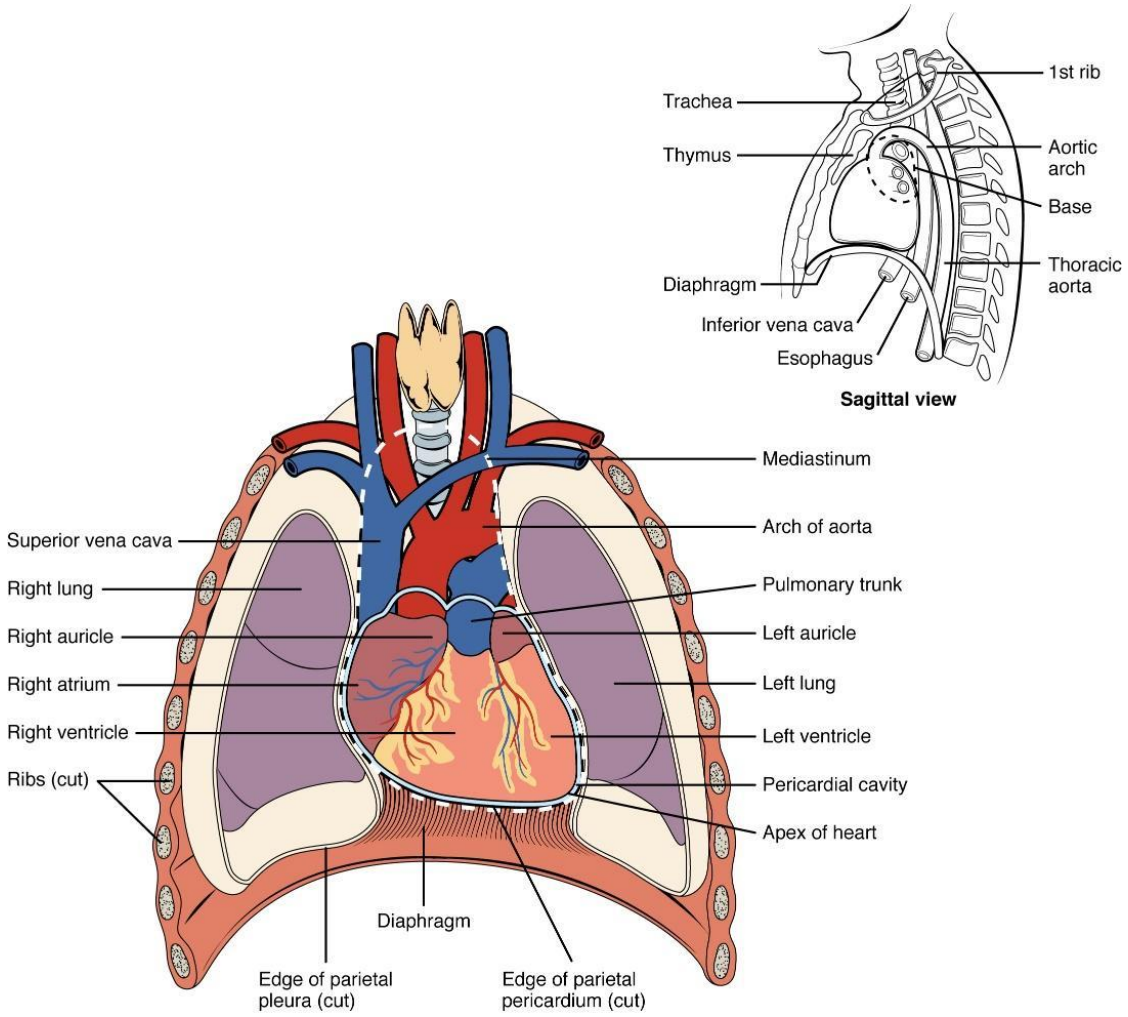

Figure 1-1 Position of the heart in the thorax [5]

1.1.2 The Pericardium and Heart Wall

The heart is surrounded and held in place by the pericardium (peri: around), a three-layered sac that surrounds and protects the heart. Its main role is to keep the heart in position while allowing it enough space for vigorous and rapid contact movements.

The wall of the heart consists of three layers (Figure 1-2):

- Epicardium (outer layer): constitutes the wall's thin, transparent outer layer.

- Myocardium (middle layer): forms a major part of the heart wall and is responsible for its pumping action. It is the cardiac muscle tissue, it consists of involuntary, striated, and grooved fibers. which is the cardiac muscle tissue,
- The endocardium (inner layer).

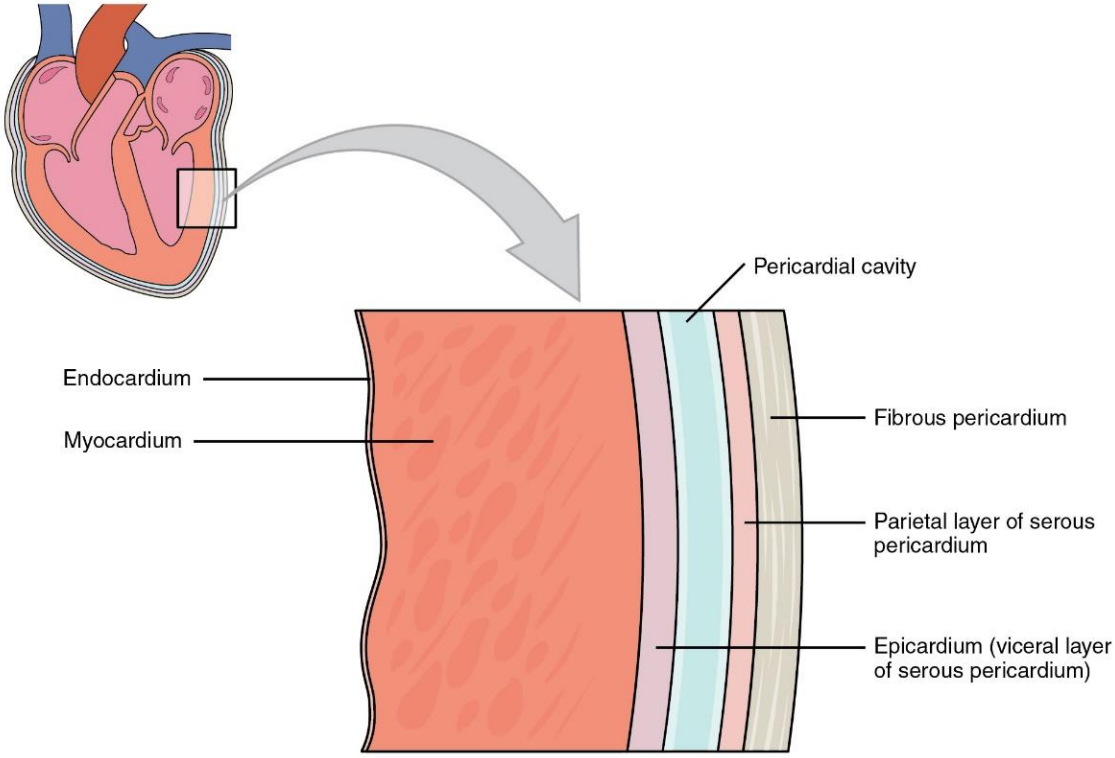

Figure 1-2 Different layers of the heart muscle [5]

1.1.3 Blood Circulation in The Heart

Although we are not consciously aware of the heart's activity most of the time, it has an incredible working capacity. Even when the body is at rest, the heart pumps thirty times its own weight every minute, and about 5 liters of blood are sent to the lungs and the same amount to the rest of the body. Since we do not spend all of our time at rest, the heart pumps more vigorously during periods of activity and its actual output is much greater.

The cardiovascular system provides the "pump" that circulates blood. As the blood circulates through the body's tissues, it supplies nutrients and oxygen to the interstitial fluid

and then to the cells. At the same time, it collects waste, carbon dioxide and heat. The heart's pumping cycle happens as follows:

The right atrium receives blood coming back from the cells, deoxygenated, through three veins: the superior vena cava which brings blood from body organs situated above the heart; the inferior vena cava which brings blood from body organs situated below the diaphragm, while the coronary sinus drains blood from the mesh vessels supplying the heart walls. The blood then moves from the right atrium to the right ventricle, which pumps it to the lungs from the pulmonary trunk.

The pulmonary trunk splits into two arteries, the right pulmonary artery carries blood to the right lung, while the left pulmonary artery carries blood to the left lung. When it reaches the lung, the blood is stripped of carbon dioxide and supplied with oxygen. This oxygenated blood is carried back to the heart through the four pulmonary veins right into the left atrium.

From the left atrium, the blood goes into the left ventricle, which pushes it into the ascending aorta. From there blood flows into the coronary arteries, which carry it to the heart, the aortic arch, the thoracic aorta, and the abdominal aorta. Finally, the aorta and its branches carry the blood into the systemic circulation.

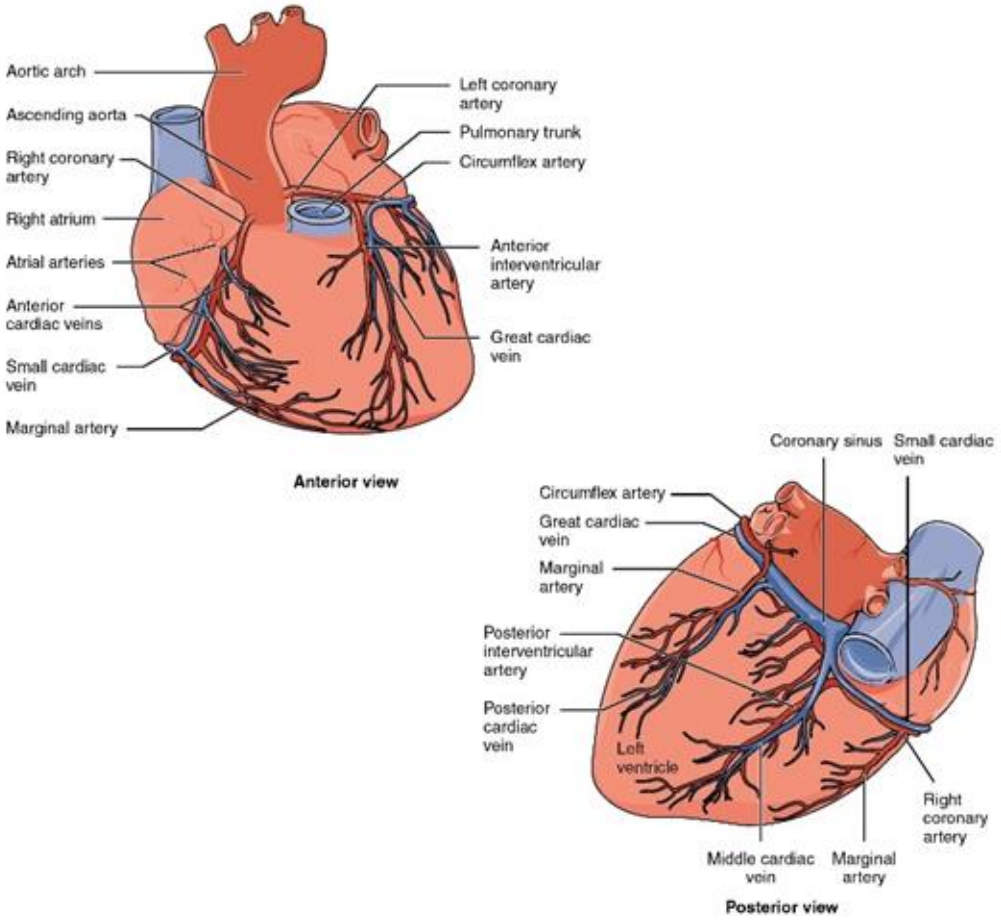

Figure 1-3 Anterior and posterior views of the Coronary Circulation [5]

1.1.4 The Valves of The Heart

When a chamber of the heart contracts, it expels some blood into a ventricle or out of the heart through an artery. The heart has valves, which prevent the blood from flowing back.

The contraction and relaxation of the heart causes an alteration in pressure that eventually triggers the valves to open and close.

- Atrioventricular valves : The atrioventricular valves are located between the atria and the ventricles; The right atrioventricular valve, located between the right atrium and the right ventricle, is also called the tricuspid valve, because it is composed of three cusps (points). The left atrioventricular valve is located between the left

atrium and the left ventricle. It is also called a bicuspid valve because it has two cusps.

When an atrioventricular valve is open, the pointed ends of the valve project into the ventricle.

- Sigmoid Valves (semi-lunar valves): are positioned between the arteries and the heart; The pulmonary sigmoid valve is located at the gap where the pulmonary trunk leaves the right ventricle. The aortic sigmoid valve is located in the gap between the left ventricle and the aorta.

1.1.5 Blood Supply to The Heart

The wall of the heart has its own network of blood vessels. Coronary circulation is the blood movement through the mesh of vessels running through the myocardium, it is called “coronary” after the crown-shaped arrangement of the heart's blood vessels (corona: crown).

1.1.6 Cardiac Conduction

Cardiac conduction refers to the speed at which electrical impulses travel through the heart. This is achieved by both conductive cells that generate and transmit impulses and contractile cells (muscles) that contract in response to electrical impulses. They can also transmit and sometimes generate impulses.

The cardiac conduction system includes the spread of electrical activity from the sinoatrial node (SA) to the atrioventricular node (AV), down the bundle of His and through the Purkinje fibers. As the electrical activity moves through the system, it triggers muscle contraction in the surrounding myocardial tissue, starting with the atria then the ventricles.

1.1.7 Cardiac Cycle

The cardiac cycle is the series of actions taking place at each heartbeat. It has two phases:

- Diastole: The heart's ventricles are relaxed, while the blood flows into the heart.
- Systole: The ventricles contract expulsing blood to the arteries.

1.2 The Blood Vessels

Blood vessels are the circulatory part of the **cardiovascular** system. They are complex networks of hollow tubes that carry blood away from the heart to the body's tissues and back to the heart. There are two types of blood vessels:

1.2.1 Arteries

Arteries are the vessels transporting blood from the heart toward the tissues. The large elastic arteries come out the heart and split into muscular arteries of medium caliber, branching out into different body areas.

These medium-sized arteries then divide into smaller arteries, which branch into even smaller arteries called arterioles. As the arterioles enter a tissue, they branch into countless microscopic vessels called capillaries. Through the walls of the capillaries, substances are exchanged between the blood and the body's tissues.

Before leaving the tissue, groups of capillaries come together to form small veins called venules. They, in turn, join together to form progressively larger vessels called veins.

- Aorta: the origin and provider of most major arteries. It is the body’s largest artery. It is responsible for pumping oxygenated blood from the heart to the rest of the body.
- Brachiocephalic artery: carries oxygenated blood from the aorta to the head, neck, and arms of the body.
- Carotid arteries: provide oxygenated blood to the head and neck regions of the body.
- Common Iliac Arteries: transports oxygenated blood from the abdominal aorta to the legs and feet.
- Coronary arteries: transports oxygenated and nutrient-filled blood to the heart muscle.
- Pulmonary artery: carries deoxygenated blood from the right ventricle to the lungs.
- Subclavian arteries: maintain proper blood circulation of the arms.

1.2.2 Veins

Veins are the vessels that carry blood from the tissues to the heart.

- Brachiocephalic veins: Two large veins that merge to create the superior vena cava, which carries deoxygenated blood from the head, neck, and arms to the heart.
- Common iliac veins: Veins that join to form the inferior vena cava, , which carries deoxygenated blood from the lower body to the heart.
- Pulmonary veins: transport oxygenated blood from the lungs to the heart.
- Cava veins: consisting of the superior and inferior vena cava, transport deoxygenated blood from various body parts to the heart.

1.3 Cardiovascular Diseases

Cardiovascular diseases (CVD) are the number one cause of mortality around the world, according to the World Health Organization (WHO) [1]. CVD include several cardiac and vascular diseases, the most common of which are heart rheumatism, high blood pressure, and coronary artery diseases. In addition, anatomical deformations of the heart or blood vessels, Diphtheria, or lack of nutrition can cause other cardiac diseases.

Cardiovascular diseases refer to a group of disorders affecting the heart and blood vessels, which include:

1.3.1 Coronary Heart Disease:

Coronary Heart Disease (CHD) is a significant contributor to death and impaired quality of life in developed nations. CHD affects the blood vessels that provide blood and nutrients to the heart muscle, making it one of the leading causes of cardiovascular disease. [6].

It is due to plaque, a combination of fatty deposits including cholesterol, connective tissue, white blood cells, and some smooth muscle cells.
Plaque build-up within the artery wall causes the arteries to narrow and decrease flexibility, obstructing blood flow. This condition is called **atherosclerosis**, a hardening of the arteries that involves plaque accumulation.

The occlusion in the coronary blood vessels causes a restriction of blood flow to the tissues, called ischemia, which in turn causes the cells to receive insufficient amounts of oxygen, leading to hypoxia.

A common symptom of coronary artery disease is pain radiating from the chest, but some patients remain asymptomatic. If untreated, coronary artery disease can lead to stroke or a heart attack. Plaque rupture is also the most common cause of coronary thrombosis. [7]

Early detection of atherosclerosis enables timely intervention. Angioplasty is a procedure where blockages are widened using a balloon. A catheter with an expandable tip is inserted through a superficial vessel, typically in the leg, and guided to the blockage site. The balloon is then inflated, compressing plaque and opening the vessel for improved blood flow. After deflation, the balloon is removed, and a stent made of a specialized mesh is often placed at the blockage site to reinforce the weakened walls. Stent insertion has been a common practice in cardiology for over 40 years. [5]

1.3.2 **Cerebrovascular Disease**

Affecting the blood vessels that supply the brain, cerebrovascular diseases include stroke, transient ischemic attack (TIA), aneurysm, and vascular malformation. They can develop from other vascular diseases like atherosclerosis, thrombosis, and embolic arterial blood clots. When not treated in time, cerebrovascular diseases can prevent the brain cells from getting enough oxygen, resulting in eventual brain damage. [8]

1.3.3 Peripheral Arterial Disease

Peripheral Artery Disease (PAD) is a condition that affects the blood vessels supplying the arms and legs and is characterized by stenosis or blockage in the aorta or limb arteries. Other contributing factors to PAD include thrombosis, embolism, vasculitis, fibromuscular dysplasia, entrapment, cystic adventitial disease, and injury. [9]

1.3.4 Rheumatic Heart Disease

Rheumatic Heart Disease is a condition affecting the heart muscle and valves, caused by streptococcal bacteria and resulting from a previous case of rheumatic fever. This disease can lead to long-term damage to the heart, potentially leading to heart valve problems.

1.3.5 Congenital Heart Defects

One of the most common birth defects, Congenital Heart Defects (CHDs) are abnormalities in the structure of the heart that are present at birth and are caused by genetic or environmental factors. These defects can range from minor issues, such as small holes in the heart, to more severe conditions like missing or poorly formed heart chambers.

1.3.6 Deep Vein Thrombosis and Pulmonary Embolism:

Obstruction of the leg veins by a blood clot, which can break free and migrate to the heart or lungs. Deep vein thrombosis (DVT) is a serious disorder. It is caused by a blood clot developing in the deep veins, most commonly in the lower extremities. When part of the clot breaks off and travels to the lungs, it causes a pulmonary embolism. Venous thromboembolism (VTE) refers to DVT, PE, or both.

Heart attacks and strokes are usually acute events primarily due to an artery blockage preventing blood from reaching the heart or brain. Their most common cause is the build-up of fatty deposits on the inner walls of the blood vessels supplying these organs, called atherosclerosis. Strokes can also result from bleeding from a cerebral blood vessel or clots.

The presence of several associated risk factors, such as smoking, poor diet and obesity, sedentary lifestyle, harmful use of alcohol, hypertension, diabetes, and hyperlipidemia, usually causes most cardiovascular diseases.

1.4 Heart Imaging Modalities

The ability to accurately assess heart functionality depends on monitoring parameters such as heart chambers' volumes, cardiac output (CO), and ejection fraction (EF) non-

invasively, all of which can be calculated from cardiac images. The ability to image the heart invasively using different modalities caused a revolution in cardiovascular medicine, allowing a better understanding and early diagnosis of CVDs, leading to eventual longer life expectancy.

Modern cardiovascular imaging techniques involve Computed tomography (CT), magnetic resonance (MR), single-photon emission computed tomography (SPECT), and ultrasound (US). Each technique offers the ability to capture multiple images during the heart's complete cycle, but each one is more suited to certain tasks compared to others.

1.4.1 Echocardiography

A cardiac ultrasound scan, also known as echocardiography, is a medical diagnostic tool that uses high-frequency sound waves to produce real-time images of soft tissues and organs. The sound waves are reflected by the boundaries between tissues of differing densities, based on the impedance difference.

An echocardiography is similar to a standard ultrasound, although the hardware and software are optimized for evaluation of cardiac structure and function. It is broadly used in the clinic because it is non-invasive, portable, and affordable.

Early echocardiography machines featured a single “M-mode” ultrasound displayed overtime on a moving paper sheet. Modern echocardiographs, however, use phased array transducers that discharge sequences of ultrasound, which are reflected and then sensed by the receiving elements. A “scan converter” generates images using information about the timing and magnitude of the reflected ultrasound. This operation repeatedly occurs in almost real-time, generating in-motion images with frame rates from thirty frames per second up to 100 frames per second. Different structures are indicated by varying shades of gray, with liquids appearing black, solids such as calcifications appearing white, tissues like the myocardium appearing gray, and muscle exhibiting a unique speckled pattern. Although two-dimensional echocardiography has replaced M-mode echocardiography in

most cases, M-mode is still used due to its high temporal resolution and accuracy for linear measurements. [10]

1.4.2 Coronary Computed Tomography Angiography (CCTA)

An exam that examines the blood vessels supplying the heart using an iodine-based contrast material and a CT scan to detect any narrowing of the arteries.

Imaging of the coronary arteries by CT is a difficult task due to the small size of their lumen and the movement of the heart and respiratory system. To overcome the issue of respiratory motion, the image is taken during breath-holding. To address the cardiac motion, the heart rate can be decreased by administering appropriate medication. Imaging the whole-heart volume is timed with the administration of intravenous iodinated in a weight-appropriate dose. Image acquisition is ECG-triggered, which makes it adapted to the cardiac cycle.

Prospective ECG triggering is used to minimize the patient's radiation exposure. The x-ray beam is only activated during a specific part of the cardiac cycle with the least movement.

1.4.3 Nuclear Medicine Imaging

Both Single photon emission computed tomography (SPECT) and Positron emission tomography (PET) scans are non-invasive nuclear imaging tests that use radioactive tracers (called radionuclides) to produce images of healthy and damaged heart muscles. They are dedicated to the diagnosis of coronary artery disease and damage due to a heart attack.

A radioactive tracer is injected into the bloodstream. The tracers are usually natural molecules such as glucose or water, which are branded with a small amount of radioactive material. The radioactive tracer produces a gamma ray, which is then detected by a gamma detector and used to produce a series of images of the heart from all different directions and angles.

PET myocardial perfusion imaging is a better option compared to SPECT due to its higher diagnostic accuracy and lower radiation exposure from using shorter-lived radiotracers. Although more expensive, PET imaging is faster than SPECT. Nowadays, PET and SPECT

scanners are combined with CT scanners, where CT is primarily used for positioning the patient in the field of view and correcting radiotracer distribution in soft tissues (attenuation correction). CT can also provide diagnostic data like coronary artery calcium score and CT coronary angiography. [11]

1.5 Conclusion

Medical imaging plays a key role in assessing and treating cardiovascular disease. But with the variety of available modalities, it is important to choose the appropriate test for the disease. Moreover, the interpretation of images and extraction of the most useful and relevant information remains a challenge for many reasons. First, medical professionals are a valuable resource that the world increasingly lacks [12]. Besides, the difference in machine settings and acquisition can make a remarkable difference between tests even for the same subject, which makes it difficult to maintain a stable evaluation. Finally, the increasing number of images generated by each test requires considerable time, even from a trained professional, to assess every piece of information; even then, some information can be overlooked, affecting the quality of the diagnosis.

The advances in cardiac imaging techniques may have revolutionized cardiovascular medicine, but if it is not backed with automatic evaluation systems, their potential can be eventually wasted.

Chapter 2

Deep Learning for Cardiac Classification and Segmentation

2 Deep Learning for Cardiac Classification and Segmentation

2.1 Fundamentals of Deep Learning

2.1.1 Introduction

Since the first notion that computers can “think” and solve complex problems, it has been the ultimate goal to make computers solve complex problems, not by following algorithms and rules, but in a smart, intuitive way like humans do. But to do so, we needed to build a new form of program that imitates a human’s problem-solving skill.

The human learning process is complicated, and it mostly happens as we are children, We learn the most complicated tasks early, from walking and talking to critical thinking. The learning process happens in stages and through different processes. It is almost impossible to grasp how brains learn, but it is certain that it has everything to do with going through different experiences every day and dealing with a lot of information.

In a way, that is exactly how it translated into computer learning. It has to do with a lot of incoming information, data, and how the machine processes and learns from it. This was the beginning of deep learning, machine learning, and artificial intelligence as we know it.

The emergence of deep learning has changed the face of programming. Being exceptionally good at finding intricate structures in data makes it particularly useful. It meant having machines that think like humans but faster and more efficiently.

2.1.2 What Is Machine Learning?

Since the existence of early computers, they have been used to perform complex computations and operations. The process is done following an algorithm. An algorithm is a set of commands created by a user, conducted in a specific order by the computer to perform a particular task.

The algorithm is performed on a set of inputs, to reach a desired output. For instance, in Figure 2-1, the algorithm is performed on a dataset containing images of both normal brain images and images containing tumors, and it turns each image with a caption of either “Normal” or ‘Tumor”.

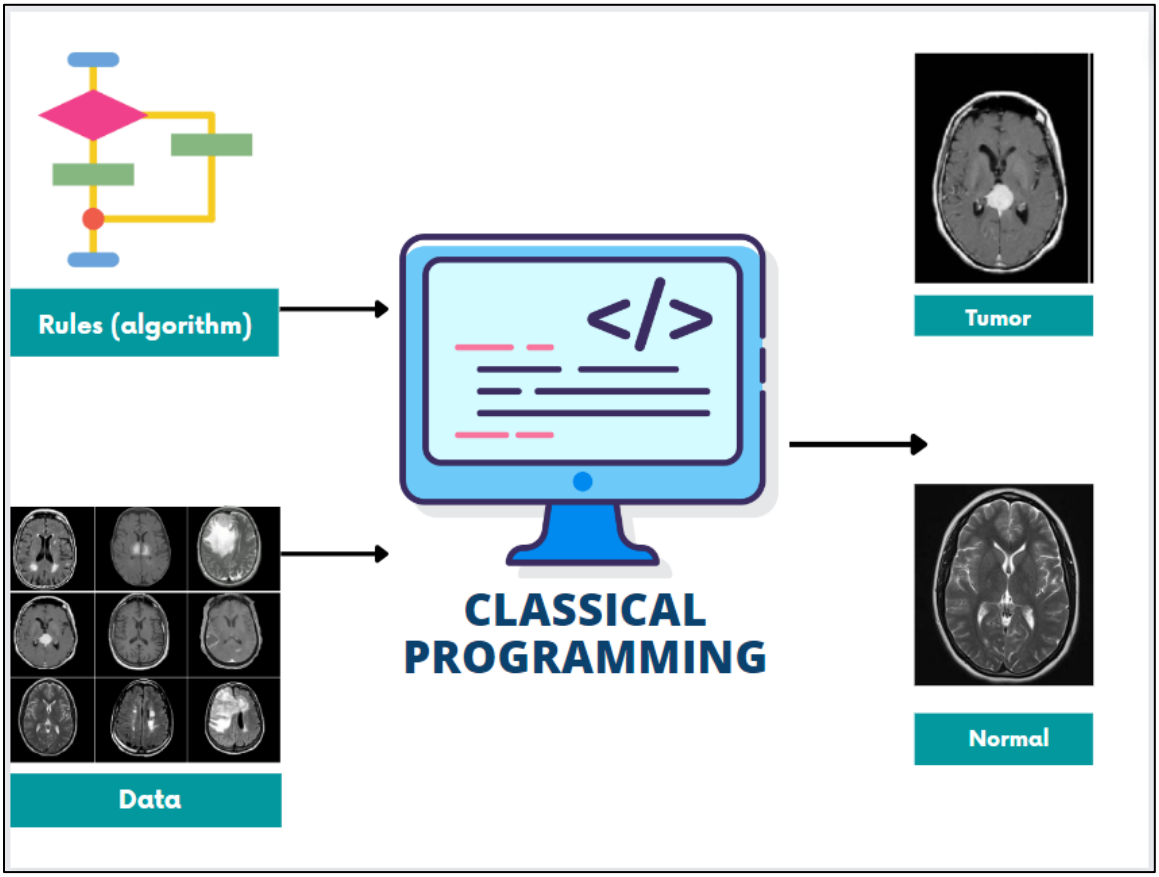

Figure 2-1 An illustration of classical programming

Machine Learning is the science of understanding data and making sense of it. It is a set of methods that allow computers to learn from the data itself to make and improve predictions. Unlike “normal programming”, where all instructions must be explicitly given to the computer, machine learning relies on “indirect programming” where learned parameters map inputs to predictions. This can be a set of weights for a linear model or a neural network. For example, in Figure 2-2 the machine learning model takes labeled brain

MRI images of both tumor and normal cases and learns to tell them apart. The result is a model that can classify unseen data as either normal or tumor.

In other words, a machine-learning model takes input data and finds some hidden pattern to give as output, these patterns are “learned” from repeated exposure to similar sets of inputs and outputs.

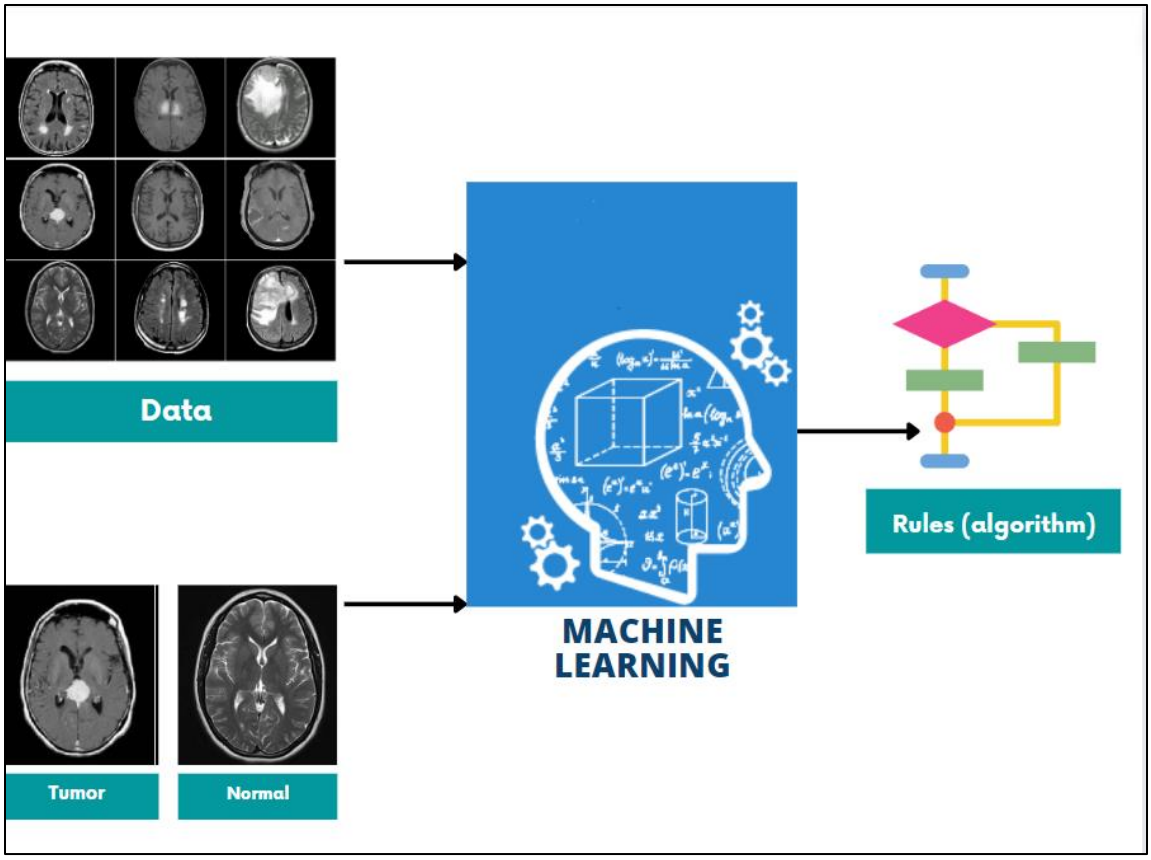

Figure 2-2 An illustration of machine learning

2.1.3 From Machine Learning to Deep Learning

Deep learning is a sub-division of machine learning based on artificial neural networks. It is the perfect imitation of the way humans learn, as it relies on learning from examples and trial and error.

We would like to propose that deep learning is quite similar to the scientific method. The scientific method is, by definition, “the process of objectively establishing facts through testing and experimentation”. The process mainly includes observing, establishing a hypothesis, predicting, experimenting and finally analyzing outcomes.

Similarly, a neural network starts with the assumption that the input is related to the output, that there exists a mathematical formula, simple or complicated and that tells of a correlation between the input and the output.

The first experiment is usually done with random initializations of weights. The results of the first few iterations are usually far from what is desired. Another iteration would be attempted, taking into consideration the error from the previous iteration, and optimizing it. More iterations are attempted until the model reaches the desired result, a minimal error between the desired output and that generated by the model.

This was the simplest, most relatable definition of deep learning, But, of course many new concepts were introduced along the way:

2.1.3.1 Neural Networks

The principle of Deep learning algorithms lies in building a network of connected basic computational elements. These elements are small computational units called “artificial neurons”, which are often referred to simply as neurons. The design of the artificial neuron was modeled after the human neuron, which forms the brain and central nervous system and plays a vital role in our cognitive abilities.

2.1.3.2 Artificial Neurons

The concept of artificial neurons used in machine learning is inspired by real neurons and their interconnections and decision-making process. The first version of an artificial neuron is the “perceptron”. It was first presented in a paper by McCulloch and Pitts in 1943 [13]. A perceptron takes several binary inputs, x1, x2…, and produces a single binary output, as illustrated in Figure 2-3.

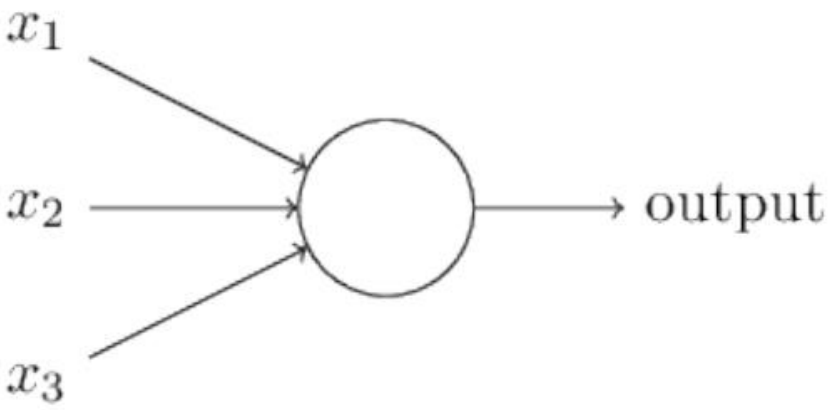

Figure 2-3 The building block of a perceptron

More accurately, the inputs are "weighted" as shown in *Figure 2-4*, and the weights attached to every input are gradually altered until the network reaches the correct output.

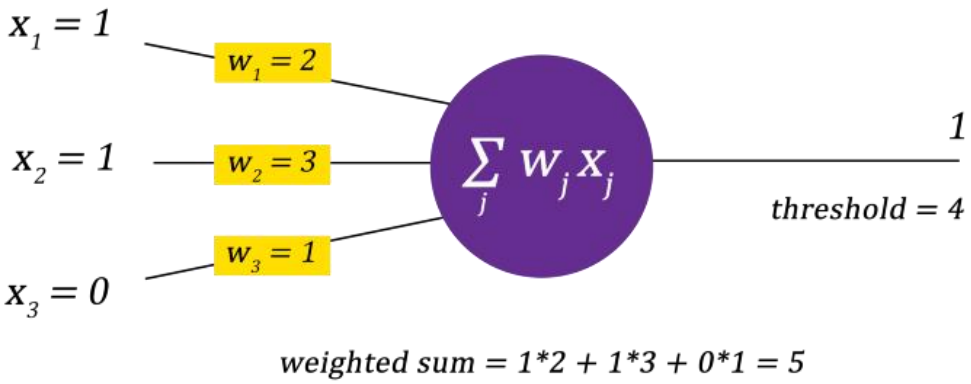

Figure 2-4 An illustration of how weights affect the outcome of a perceptron.

Stacking neurons in a series of layers has proven to be the most efficient way to organize neural networks. Neurons of the same layers are separated from each other, They get their input from a previous layer and feed their output to the next layer.

The layer stacking method allows a hierarchical analysis of data, where early layers analyze simple features, and later layers analyze more complicated ones. Taking the example of computer vision, early layers are more of a simple shape detector. They only detect lines and curves, whereas later layers detect more complicated shapes and textures until the model can analyze a full complex image.

2.1.3.3 Layers

All the learning occurs in the layers. A layer is a series of neurons stacked at a specific depth within a neural network. Layers can be of different types, as shown in *Figure 2-5*: Input layer, hidden layers , and output layer.

The input layer contains the input data, where each variable is a ‘node’.

The hidden layers are the heart of neural networks. They are designed to learn specific aspects of the data by minimizing a cost function. A hidden layer means simply it is neither an input nor an output layer.

Up to now, a neural network’s output of a layer is an input of the next, Such a structure is called a **feed-forward neural network**, and loops are not much sensical. However, there are **recurrent neural networks** that have neurons that fire for a limited time, which makes loops possible.

The output layer usually represents the output of classification problems. It can have one to more neurons, depending on the expected outcome of the problem.

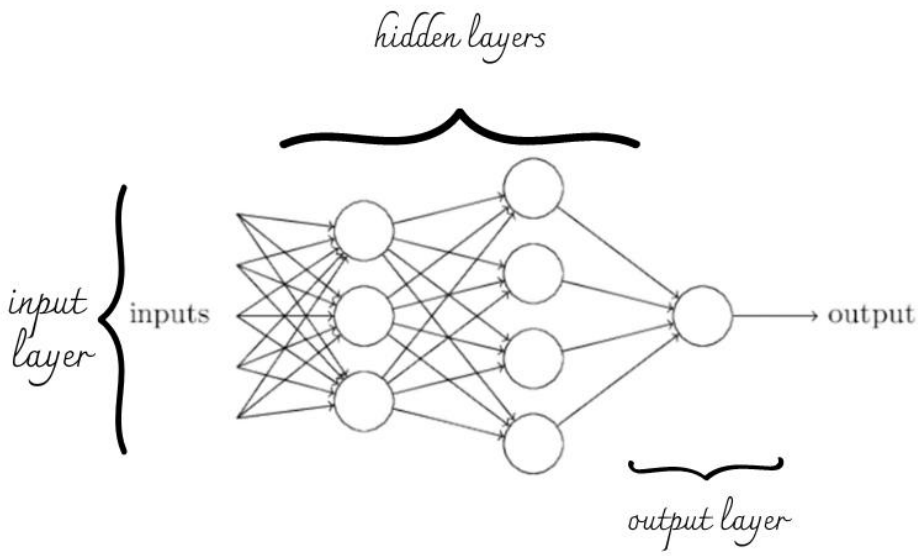

Figure 2-5 The different types of layers in a neural network

2.1.3.4 Loss Function

It is how we monitor the learning process and try to improve it. Loss functions map a number of parameter values for the network onto a scalar value that compares the desired output to the predicted one. Then it measures how efficiently those parameters achieve the intended task.

2.1.3.5 Optimization Algorithms

There are usually too many different parameters to tweak in a neural network (weights and biases) before it performs as it is expected to. It is common to consider the parameters' universe a multidimensional universe, which needs to be searched to find the best parameters' combination for optimal performance. The search process is done through an optimization process, a mathematical error function that needs to be minimized. This function is minimized using one of many popular optimization algorithms [14]:

- Gradient Descent [15]

- Stochastic Gradient Descent (SGD) [16]
- Mini Batch Stochastic Gradient Descent (MB-SGD) [17]
- SGD with momentum [18]
- Root Mean Square Propagation (RMSProp) [19]
- Adam [20]

2.1.4 Popular Architectures:

The success of neural networks leads to more research to make them more suited to different tasks, resulting in different architectures adapted to different research fields. We will mention a few that were more used in computer vision in general and medical applications in particular.

2.1.4.1 Convolutional Neural Networks (CNNs)

The Convolutional neural network (CNN) was introduced in 1995 [21].

CNNs are regular neural networks that have at least one convolutional layer. A convolution is the act of sliding a filter all over the input, which mainly helps consider the neighborhood of each pixel, which is particularly useful when dealing withs images. Additionally, considering single patches of the input at a time can make spotting features easier than looking at the whole input. Convolutional neural networks have been particularly superior in performance with image, speech, and audio signal inputs. Their structure involves three main layers: Convolutional layers, Pooling layers, and fully connected (FC) layers.

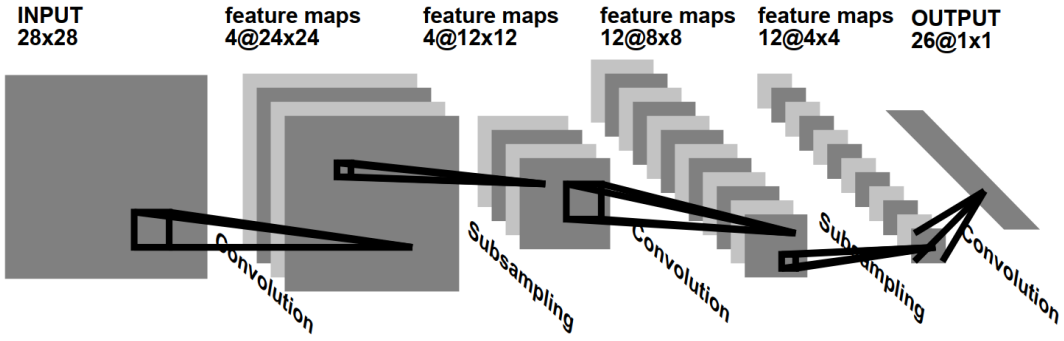

Figure 2-6 The Convolutional Neural Network as described in the original paper [21]

Convolutional Layer

Most of the computing in a CNN happens on the convolutional layers. There is a convolution operation between the input data and the kernels (filters), which results in a matrix referred to as "feature map", which mainly highlights the location and strength of a feature in the input.

Pooling Layer

Pooling layers are responsible for reducing the spatial size of the layer and the number of parameters in the input, which is why they are sometimes referred to as down-sampling layers. The pooling layer sweeps a filter through the image, but instead of doing a convolution operation, the kernel performs an aggregation function on its field of impact and fills an output array in the process. This is useful for spotting dominant features which are invariant to rotation and position. The main types of pooling that exist are:

- **Max pooling:** the most popular type of pooling. It selects the pixel with the maximum value and sends it to the output array, then the filter slides into another region of the input.

- **Average pooling:** in this type of pooling, the output array is filled with the average of the region in contact with the filter, before it slides to the next one and does the same.

The down-sampling that happen in the pooling layers comes with a big loss of information but is considered insignificant. Pooling layers are the strength of CNNs. Not only do they reduce the complexity of the module, but they also help improve efficiency, and limit the risk of overfitting.

Fully Connected Layer

As its name suggests, a fully connected layer has each neuron of the output layer connected to every neuron from the layer before. It is used for classification based on previously extracted features.

The success of CNNs in different tasks leads to more improvements, and variations over the years, including: AlexNet [22], VGG Net [23], GoogleNet [24], and ResNet [25].

2.1.4.2 Fully convolutional Networks

Fully convolutional networks were first developed by Long [26], They are CNN with no fully connected layers in them. The last fully connected layer was replaced with a fully convolutional layer. This change allows the network to have a better pixel-wise prediction.

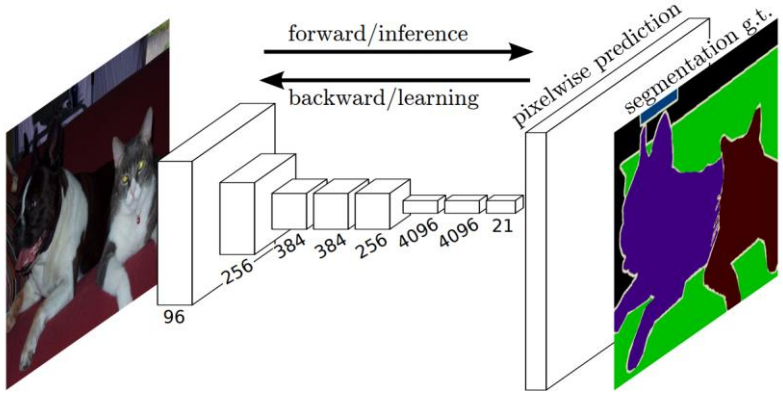

Figure 2-7 The Fully Convolutional Network as described in the original paper [26]

2.1.4.3 U-Net

U-Net is particularly popular in medical image segmentation problems, proposed by Ronneberger [27]. The main architectural novelty in U-Net is the combination of an equal amount of up-sampling and down-sampling layers, leading to a symmetrical shape resembling a "U" (Figure 2-8), connecting them with skip connections between opposing convolution and deconvolution layers. This step aims to concatenate features learned in the contracting path to the expanding paths. U-Net models can process entire images in one forward pass, directly producing a segmentation map. This allows them to consider the full context of the image, unlike patch-based CNNs. A 1x1 convolution is used in the final layer to assign each 64-component feature vector to the desired number of classes. The network has 23 convolutional layers in total.

This architecture was introduced along with data augmentation, decreasing the amount of annotated data needed to train, a feature that made the network very successful in biomedical applications.

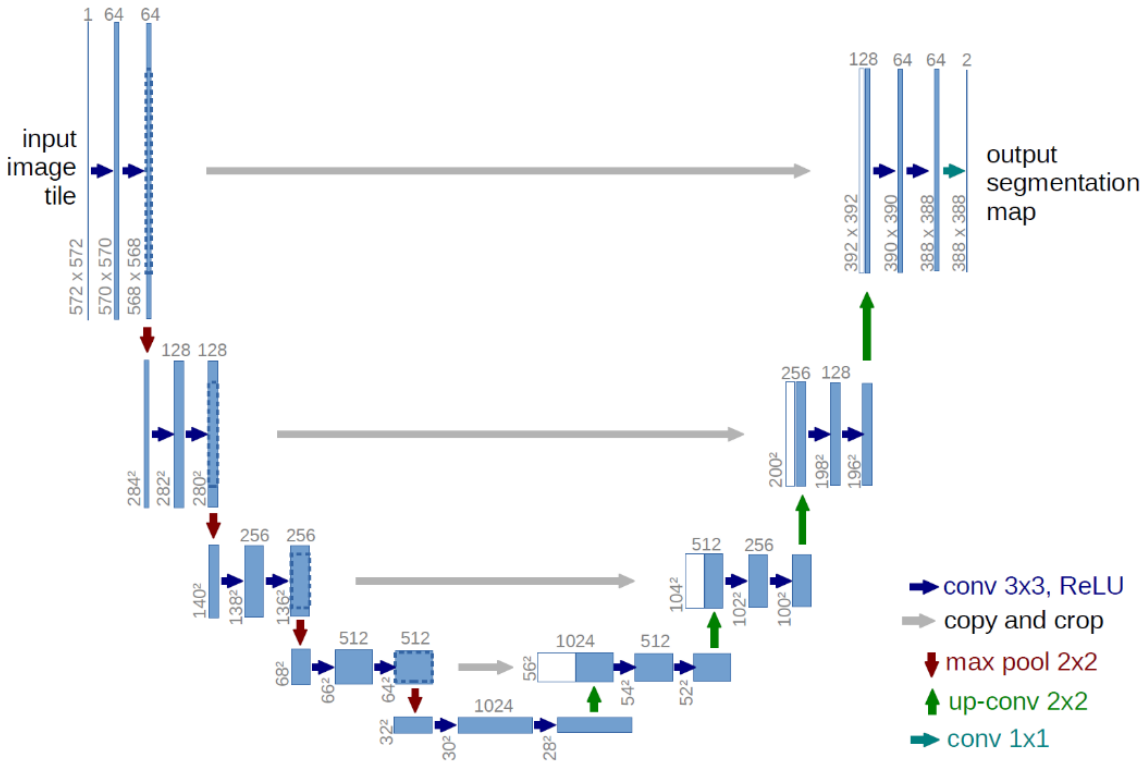

Figure 2-8 The U-Net architecture as described in the original paper [27]

2.1.4.4 Residual Networks

Residual Networks (ResNets) [18] came as answer to an emerging problem in very deep networks, where weights became increasingly large or small after too many operations. ResNets add residual blocks that pact as short-cut connections between the input (x) and the output of a layer (F(x)).

The residual block sums the non-linear transformations performed by the layer and its input. It is reformulated as H(x) = F(x) + x (Figure 2-9).

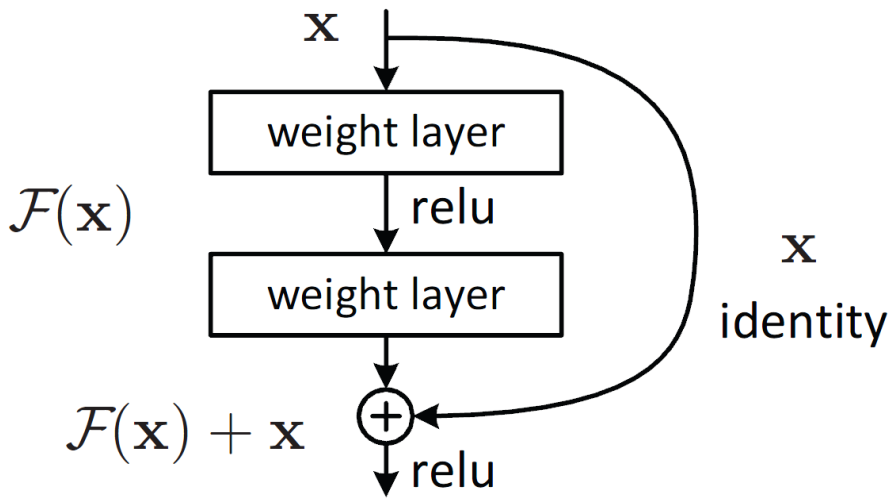

Figure 2-9 The residual block structure

2.1.4.5 Attention-Based Models

The use of attention mechanisms in deep learning models started with the publication of a paper by Vaswani et al [28]. Attention-based models are a type of deep learning architecture that use attention mechanisms to selectively focus on different parts of an input sequence or image. This allows the model to dynamically allocate more computational resources to the most relevant parts of the input, while ignoring the less relevant parts. Attention-based models are commonly used in natural language processing (NLP) tasks, such as machine translation, question answering, and text summarization, where the model needs to selectively attend to different words in a sentence. They are also used in computer vision tasks, such as image captioning and visual question answering, where the model needs to selectively attend to different regions in an image. In the medical field, attention-based models have been used for tasks such as image segmentation, lesion classification, and diagnosis prediction.

2.1.4.6 Visual Transformers

Visual Transformers [29] are a type of deep learning model based on the Transformer architecture, originally proposed for natural language processing tasks. The Transformer

architecture is a type of neural network that uses self-attention mechanisms to capture dependencies between elements in an input sequence. In visual transformers, this architecture has been adapted to process visual data, such as images and videos.

Visual Transformers have been used for various computer vision tasks, including image classification, object detection, and segmentation. They have shown to be effective in capturing global contextual information and handling long-range dependencies, making them well suited for tasks that require a comprehensive understanding of the visual content in an image. Additionally, Visual Transformers can be trained on large-scale datasets, allowing them to learn from vast amounts of data, leading to improved performance on complex visual recognition tasks.

2.1.5 Challenges in deep learning

Thanks to the quick advancement in computational powers and the increasing availability of data, deep learning has taken off in the last few years in all fields. However, when it comes to medical applications, too many challenges are emerging lately, slowing the full impact it could have had.

- Need for data:

As previously mentioned, the availability of data, especially labeled one, has strengthened deep learning in the last few years. Unfortunately, this is not the case with medical applications; there is a big shortage of labeled data, not only because of the lack of resources to begin with, but because of confidentiality concerns. As we will see in the next chapter, one promising solution is data augmentation, either by traditional methods or data synthesis.

- Explainability and fairness:

Even though deep learning algorithms have reached human-level accuracy on many tasks, they still are considered black boxes and cannot be backtracked to explain each prediction made. This makes them hard to trust, especially in sensitive domains like healthcare.

2.2 Deep Learning for Cardiac Image Segmentation

Automatic segmentation of heart chambers is a large field of research. The complex shape of the heart and the existing variety of imaging modalities provide many tasks and challenges corresponding to different segmentation methodologies.

The literature is quite rich and detailed when it comes to cardiac segmentation, most review papers focused on MRI being the best-performing imaging modality. For instance, the authors in [30] reviewed techniques for segmenting different heart chambers in MRI, while in [31], the authors focused on segmentation methods on short-axis cardiac MR images. [32] and [33] reviewed only the right ventricle segmentation from cardiac MRI. Some papers took an interest in segmentation techniques from echocardiography images [[34], [35], cine MR segmentation [31], or Tagged MRI [[36]. Frangi et al. [37] classified cardiac modeling techniques into three classes: surface models, volume models, and deformable models.

Other reviews were more generalized. For instance, in [38], the reviewed 3D cardiac modeling techniques were based on different modalities, angiography, cardiac US, isotope imaging, cardiac CT, and MRI. Similarly, the authors in [39] grouped heart chambers and whole heart segmentation techniques by modalities.

With the emergence of deep learning, reviews focused on its application in cardiology [40], [41], [42].

The study of different review papers offers an overview of the various techniques and applications of segmentation in cardiac imaging.

2.2.1 Cardiac Magnetic Resonance Imaging CMRI

Magnetic resonance imaging is the golden technique for viewing the heart's structure and function. The sharp contrast between different tissues results in good-quality images and easy diagnosis.

Initial segmentation techniques employed basic methods such as thresholding as a preliminary step, followed by a region-expansion technique [43], [44]. Some other techniques relied on Classification-based segmentation [45] or deformable contour [46], [47],

2.2.1.1 Left Ventricle LV

The left ventricle (LV) plays a crucial role in the cardiovascular system as it pumps oxygenated blood from the heart through the aortic valve to be circulated throughout the whole body. With such a key role in blood circulation, the left ventricle has been the most investigated chamber in cardiac segmentation.

The left ventricle has relatively thick myocardial tissues that provide enough pressure for blood circulation. As a result, LV parameters can be abnormal in many cardiovascular diseases, such as hypertension or after myocardial infarction.

The success of Convolutional Neural Networks in medical applications applied quickly to LV segmentation; Romaguera et al. [48] trained a five-layer FCN model using whole cardiac MRI on the MICCAI 2009 SUNNYBROOK Left Ventricle Segmentation challenge dataset [49] without cropping the images while using stochastic gradient descent (SGD) and RMSprop optimizer; They reported better results with SGD. Emad et al. [50] used a 6-layer CNN with a pyramid of scales-based localization, which allowed the algorithm to consider varied sizes of the heart over different images, increasing the system's performance. Finally, Molaei et al. [51] reported an improvement in both specificity and sensitivity in LV segmentation when using Gabor filter initialization; Gabor filters were chosen due to their structure invariant properties, which the authors hoped the DCNN would inherit, resulting in a stronger system and more accurate segmentation. In

addition, Gabor filters act as a preprocessing step that enhances tissue contrast and blurred edges.

Some researchers have investigated hybrid methods, combining traditional image processing techniques with deep learning architectures. Ngo et al. [52] created high-quality segmentation by utilizing a distance regularized level-set method, in combination with a pipeline of Deep Belief Networks (DBNs) to both identify and segment the LV. Luo et al. [53] trained a three-layered CNN to predict the LV segmentation, after employing a LV atlas method to accurately localize the left ventricle. Avendi et al. [54] implemented a three-step process to segment the LV. It starts with using a CNN for LV chamber detection, then using a stacked autoencoder to predict the shape of the LV and finally, completing the segmentation with level-set refinement.

Ngo et al. [43] created high-quality segmentation by utilizing a distance regularized level-set method, in combination with a pipeline of Deep Belief Networks (DBNs) to both identify and segment the LV. Avendi et al. [44] employed a three-stage approach, starting with a CNN to identify the LV chamber, and then using a stacked autoencoder to infer the shape of the LV, and lastly, applying level-set refinement. Luo et al. [45] used an LV atlas mapping method for precise localization, and then trained a three-layer CNN to predict the LV.

Yang et al. [55] created an end-to-end deep fusion network that uses deep learning for label fusion and feature extraction, Learned features are then used to define a similarity measure for MRI atlas selection. They used these learned features to define a similarity measure for selecting MRI atlases. They evaluated their method against other techniques such as majority voting, patch-based label fusion, multi-atlas patch match, and SVM with augmented features and found that their method had superior accuracy. In a subsequent study [56], they used a regression CNN to locate the left ventricle (LV) and then used a U-net-based architecture to segment it within the relevant region of interest.

Rohé et al. [57] developed a multi-atlas framework with a registration method trained with convolutional neural networks, The registration module employs SVF-Net [58] , which is

an encoder-decoder network that replaces the optimization of energy criteria with a deterministic prediction of the parameters from training images.

A recent technique resorted to breaking down the segmentation problem into subtasks by applying neural networks in a multi-stage pipeline. Tan et al. [59] used a CNN for LV endocardium location and another CNN to determine the endocardial radius. They used MICCAI 2011 Left Ventricle Segmentation STACOM [60] and MICCAI 2009 Sunnybrook dataset [49] for training and evaluation, respectively, the model achieved results comparable to the state of the art. While Liao et al. [61] combined a detector with a neural network classifier and used them to detect the region of interest containing the LV. They segmented the LV using a "hyper columns" FCN. The 2D segmentations were used to estimate the LV volume. The model was end-to-end trained alternately on LV segmentation and volume estimation.

Wolterink et al. [62] employed a convolutional neural network that increases the level of dilatation with every convolutional layer, ensuring a larger receptive field with less trainable parameters. While Yang et al. [56] used a 3D fully connected network, with residual connections instead of concatenation, in order to compensate for the volume size imbalance usually encountered with 3D series. They investigated different loss functions, They proposed a Multi-class Dice Similarity Coefficient (mDSC) based loss function to re-weight the training for all classes.

The authors in [63] performed end-to-end LV segmentation combining feature detection with a capsule network. A capsule is a group of neurons that output a vector [64], The length of the vector refers to the probability of the existence of the object, while its direction refers to the instance parameters [65], [64]. A capsule network preserves spatial information; it proved useful in replacing pooling layers, which would significantly decrease the number of parameters while maintaining the same amount of information.

2.2.1.2 Right ventricle RV

Right ventricle (RV) segmentation had long been neglected, considering its role less important than that of the left ventricle, but recent research suggests that any failure of the left ventricle causes an overload of the right ventricle and changes its dynamic, making the RV function a vital biomarker of the progress of all heart diseases.

But even after knowing its importance, RV segmentation remains more challenging due to the complex geometry of the right ventricle; it has a variable and crescent structure, thin indistinct myocardial walls, and non-uniform boundaries. In addition, the literature on RV segmentation remains poor.

Early segmentation algorithms used end-to-end FCNs to segment the LV and RV simultaneously. One of the first papers to apply a plain FCN to CMRI was Tran et al. [66]. They trained a plain four-layer end-to-end FCN model on two different datasets; MICCAI 2009 Sunnybrook dataset [49] for LV and MICCAI 2011 Left Ventricle Segmentation STACOM [60] for RV. The model performed better than state-of-the-art methods on both tasks.

Later works focused on optimizing the network structure to enhance the feature learning capacity for segmentation. Khened et al. [67] developed a model inspired by DenseNets and included Inception modules [24] for a consistent performance when segmenting images with varying anatomical shapes. The model performs parallel convolutions with different kernel sizes, and later combine them in a similar fashion to an inception module.

In [68], the authors created a self-training semi-supervised learning method in which a segmentation network for LV/RV and myocardium was trained on labeled data to estimate labels, then updated the network with both the available true image-label pairs and the estimated labels for the unlabeled images, while the final segmentation was refined using a Conditional Random Field (CRF).

Avendi et al. [69] stack a regular CNN with autoencoders. The model first detects the RV chamber, then segments it.

Archontis et al. [70] use a two-pathway CNN; one pathway captures the finest details from original images, whereas the other learn the higher-level features from down-sampled images.

Other studies use either Recurrent Neural Networks or multi-slice networks (2.5D nets) to enhance the segmentations with spatial information extracted from neighboring slices ([71], [72], [73], [74]).

The authors in [71] start with determining the region of interest (ROI) using ROI-net, a variant of U-Net that takes one MRI image as input and predicts pixel-wise probabilities providing a heart/background segmentation. The ROI image is then segmented using LV-net for left ventricle segmentation or LVRV-net for bi-ventricular segmentation. It performs slice segmentation of a given slice S(i), taking S(i-1), the adjacent slice above, and the segmentation mask as contextual input. In [72], the authors proposed a recurrent U-Net that combines left ventricle detection and segmentation into a single end-to-end architecture, thus simplifying the segmentation pipeline. The network trains on a stack of 2D images. Previously segmented images are fed to the recurrent unit to serve as context for segmenting the current image. The proposed model outperformed known architectures such as FCN and recurrent deep belief networks.

Patravali et al. [73] tested a 2D and 3D U-Net, trained with different loss functions. The model that reported the highest Dice loss was a 2D U-Net designed to take a stack of 3 image slices as input channels.

In [74], Du et al. use an advanced end-to-end encoder-decoder network that takes dilated modules as an encoder and D-Fire dilated modules as a decoder. The model performs Bi-Ventricle segmentation from the pixel level view (Cardiac-DeepIED). The ED is integrated with a convolutional LSTM in one innovative end-to-end network architecture. The encoder and the Convolutional LSTM can capture semantic information (pixel-based information). At the same time, the pooling layers reduce the size of feature maps, resulting in images that are small in size but rich in semantic information. The decoder up-samples the image representation to recover original dimension, then generates a feature map

featuring the image's label. The resulting feature maps are finally fed to a fully connected layer. Kong et al. [75] used a temporal regression network to accurately identify end-diastole and end-systole frames. They combined a 2D CNN to code the spatial information of the cardiac sequence with a long short-term memory (LSTM) to decode the temporal information.

Isensee et al. [76] used a combination of 2D and 3D U-Nets to perform both segmentation of the LV/RV contours and LV myocardium, and disease classification.

Adding skip connections to the models has also shown good results. Lieman et al. [77] created "FastVentricle", a new FCN architecture with integrated skip connections, adding more speed and memory efficiency. The model was inspired by ENet [78]. In [79], the authors introduce a variant of U-net, where they replaced the double convolution layers with a single convolution layer at each depth, which reduces the memory burden and speeds-up the training. The new model, called V-net, is used for segmentation of LV/RV endocardium and epicardium . The method over-performed human experts, particularly for RV segmentation.

2.2.1.3 Left Atrium LA

Using deep learning algorithms for left atrium segmentation can be challenging due to the small presence of the atrial structures compared to the background. To overcome this issue, Vesal et al. [80] started by cropping using fixed coordinates to extract the input images from the center of the image, which allows for a better representation of the LA features on a smaller ROI. They used dilated convolutions instead of regular convolutions in the lowest layer of the encoder branch of a 3D U-Net, for the segmentation of the LA straight from Gadolinium Enhanced MRIs (GE-MRI). Following a similar principle, other researchers ([81], [82], [83]), used a multi-CNN method for atrial segmentation. They concatenated two networks: a CNN to localize and crop out the Left Atrium, and another network to segment the LA from the small image patches.

A recurrent issue that decreases segmentation performance when training is the big variance of the LA anatomical structures. He et al. [84] avoided false classification by using image context information. They incorporated multi-scale pooling in a pyramid module to associate contextual features for a more accurate classification. Similarly, Zhao et al. [85] incorporate object and image context with pyramidal pooling in their neural network new called PSPNet.

Inspired by [84] and the PSPNet [85], Bian et al. [86] extract different scale features with pyramid pooling, incorporated in a multi-scale 2D CNN. The network is robust against the different shapes and forms usually encountered in clinical cases.

Vigneault et al. [87] and Bai et al. [88] applied 2D FCNs to directly segment the LA and RA from standard 2D long-axis images. They also demonstrated that their networks can be trained to also segment ventricles from 2D short-axis images without changes to the network structure. Likewise, Xiong et al. [89] , Preetha et al. [90], and Chen et al. [91] have used 2D FCNs to segment the atrium from 3D LGE (Late Gadolinium Enhancement) images slice-by-slice, optimizing the network structure for improved feature learning. In addition, 3D networks ([82], [92], [93], [80], [81]) and multi-view FCN ([94], [95]) have been investigated to capture 3D information from 3D LGE images for more accurate atrium segmentation. In particular, Xia et al. [82] proposed a two-stage segmentation framework that uses a 3D U-Net to roughly locate the atrial center from down-sampled images, followed by another 3D U-Net to accurately segment the atrium in the full-resolution portions of the original images. This approach is both efficient and accurate.

2.2.2 Computed Tomography CT

Computed tomography of the heart, also known as cardiac CT, is a diagnostic imaging test used to get detailed images of the heart and coronary arteries. This test is commonly performed to evaluate the presence and severity of coronary artery disease (CAD), and to detect and assess atherosclerosis by analyzing plaque and stenosis.

Early research in the field of automatic atherosclerosis screening in cardiac CT focused on improving the analysis of the images obtained from the test. This involved recognizing and segmenting the cardiac arteries for better analysis [96], [97], and also segmenting stenosis [98] and plaques [99], which are the primary indicators of atherosclerosis.

2.2.2.1 Coronary Artery Segmentation

Coronary CT angiography (CCTA) is considered the golden modality for coronary artery disease diagnosis. Segmentation of the coronary artery has versatile uses in CCTA; it provides quantitative information on coronary artery stenosis, 3D reconstruction, and cardiac dynamics assessment.

The authors in [100] proposed a growing method that starts with determining an initial seed which is the meeting point of the ascending aorta and the coronary arteries. Then a convolutional network helps a growing algorithm search for the existence of coronary arteries in the neighboring blocks, all happening iteratively.

[101] proposed a 3D attention FCN method to perform end-to-end coronary artery mapping straight from CCTA. The FCN was associated with a deep attention strategy to highlight semantic features, improving the segmentation accuracy.

Multiple studies investigated vessel segmentation from CTA ([102], [103]), which is the first step towards the diagnosis of atherosclerosis and the detecting and quantifying of stenosis ([104], [105]).

2.2.2.2 Coronary Artery Calcium and Plaque Segmentation

One particularly specific feature of atherosclerosis is Coronary artery calcium (CAC), Coronary artery calcification is a collection of calcium in the heart's coronary arteries which, makes CAC scoring an easy and efficient means of cardiovascular disease assessment [106]. Early papers focused on learning and analyzing spatial information such as dimensions, texture, looks, and overall position of a lesion to distinguish CAC from other similar candidates like aortic calcifications. Position features related to anatomical landmarks were the most efficient, as proven by many studies ([99], [107], [108], [109]).

Shahzad et al. [110] and Wolterink et al. [111] performed calcium scoring for each individual vessel on pre-located major coronary arteries. While Išgum et al. [112] estimated the location of the whole coronary artery tree, Yang et al. [113] extracted coronary artery centerlines in CCTA images to provide location features. Hong et al. [98] were the first to evaluate clinically relevant parameters from deep learning-based coronary artery segmentation and CNNs to segment CTA lumen and calcified plaque.

Most deep learning-based methods aim to predict dense segmentation probability maps using an end-to-end CNN segmentation ([114], [115], [96], [69]). Recent methods, however, resorted to classifying each voxel separately. For example, Wolterink et al. [116] proposed to first identify candidate voxels in CCTA with a CNN, and then further discriminate among identified candidates with a second CNN. Lessmann et al. [117] used a similar approach to locate calcifications with two CNNs, one for identification of labeling of potential calcification based on location, and the other for true calcification's identification from pre-selected candidates. In contrast, using end-to-end regression, Cano-Espinosa et al. [118] and de Vos et al. [119] achieved fast automatic calcium scoring in less than a second.

On the other hand, few works were interested in non-calcified plaque (NCP) and mixed-calcified plaque (MCP) segmentation and quantification [105]. Most machine learning-based methods extracted descriptors of the vessel wall ([120], [121], [122], [123]), then ran them through a linear classifier or SVM to determine the presence of non-calcified plaque. [124]

2.3 Deep Learning for Cardiac Disease Classification

Disease classification straight from MRI or CT is NOT quite common. Instead, the process usually involves a feature extraction step or segmentation.

Zreik et al. [125] identified patients with coronary artery stenoses from the LV myocardium of rest CT. They used a multi-scale CNN to first segment the LV myocardium and then

applied an unsupervised Convolutional Autoencoder (AE) to encode it. Finally, they used a Support Vector Machine (SVM) classifier to perform the final classification based on the encoded and clustered data.

One of the first papers to investigate the possibility of automatic atherosclerosis screening straight from CCTA [126] proposed a deep-learning-based system to classify coronary arteries as normal or abnormal and display the likelihood of atherosclerosis in each coronary artery or branch. They start by identifying and extracting the coronary arteries from the CCTA images through the use of a deformed mean shape model. They then straighten the vessels in a multi-planar reformatting (MPR) to produce a longitudinal view of the coronary arteries. The MPR volumes were classified using a 3D CNN for atherosclerosis screening, associated with a gradient-based class activation map (Grad-CAM) [109], to highlight image regions that influenced the CNN's decision-making process. The model was evaluated using five-fold cross-validation and achieved an accuracy of 90.9%, a positive predictive value of 58.8%, and a sensitivity of 68%.

The authors in [127] employed an automated technique for plaque classification as either normal, calcified, or non-calcified. The algorithm uses a Gabor transform to extract seven features from CTA images: energy, and Kapur, Max, Rényi, Shannon, Vajda, and Yager entropies. They obtained 89.09% accuracy, 91.70% positive predictive value, 91.83% sensitivity, and 83.70% specificity. The results were obtained using all computed features without any feature reduction, using a probabilistic neural network.

White et al. [128] achieved 95% Negative Predictive Value for the task of atherosclerosis screening from CCTA, allowing for a safe discard of atherosclerosis.

2.4 Conclusion

The emergence of deep learning had an important impact on cardiology. It allowed for a full exploration of the big data generated by modern imaging techniques.

Automatizing the assessment of cardiac structure and function is the eventual aim of applying deep learning. As we have seen in this chapter, many breakthroughs have been

accomplished, but there is still room for improvement in existing applications and innovation in new ones.

Cardiac segmentation is a broad field of research. With the rapid rise of learning techniques, choosing the right model for the right task can be a challenge, and it can depend on the following:

- The imaging modality, as with every modality, has its own characteristics and specifics.
- The targeted structure. Different heart chambers have different textures and shapes, and it is important to adapt the segmentation method to match the task's needs.
- Available data, as it is challenging to adopt a deep learning method when working with limited datasets, although it is important to note that most recent research found a way around this, as transfer learning and data augmentation allowed more flexibility.

Deep learning is a powerful tool that has been heavily studied in cardiac research in recent years. However, the challenges and issues related to it need to be addressed properly to make it accepted and commonly used in real-life cardiac applications.

Chapter 3

Coronary Artery Disease, Atherosclerosis Screening

3 Coronary Artery Disease, Atherosclerosis Screening

3.1 Introduction

The coronary arteries form a mesh surrounding the heart muscle. They are solely responsible for delivering blood to the myocardium, hence providing nutrients and oxygen to every blood cell. Any malfunction of the coronary arteries can lead to ischemia, angina, decreased performance, and/or infarction. Consequently, the drop in coronary blood flow below required metabolic needs leads to ischemic myocardium, affecting everything from the heart's electrical activity to its pumping ability. This can lead to a heart attack and possibly death.

Coronary artery disease is one of the most common types of cardiovascular diseases. It is a result of the gradual narrowing of the lumen of the coronary arteries due to atherosclerosis, a buildup of plaque in artery walls leading to blockage or narrowing, making it the primary cause of heart disease.

Atherosclerosis is a condition in which cholesterol builds up in the walls of the arteries. This process begins in childhood, with the formation of fatty streaks within the arteries. As people age, these streaks can worsen, becoming scarred and calcified, which can lead to narrowed or blocked vessels. The effects of the narrowing or blockage depend on which vessels are affected. Affected cerebral vessels can cause strokes, affected coronary vessels can cause angina and heart attacks, while affected renal vessels can cause renal failure, and affected peripheral arteries can cause limb ischemia. [129]

There are several risk factors that can increase the chances of individuals developing atherosclerosis including among others : age, male gender, and raised plasma cholesterol.

The go-to modality for evaluating patients with coronary artery diseases is Coronary CT angiography (CCTA). It is a non-invasive imaging technique that permits a thorough description of coronary artery plaque and grading of coronary artery stenosis. However, like any other imaging modality, CCTA generates a substantial number of images, of which

the examination process is still to date performed manually by an expert or semi-automatically by first segmenting the lumen and the arterial wall and then defining the presence of plaque or stenosis, which can be a time-consuming process and can only be performed by a human expert. Therefore, the need for an automatic system proves more urgent than ever for better and faster management of emergencies.

Automatic atherosclerosis screening remains a relatively novel field of research. Early studies focused on segmenting and classifying plaque scores. [130]

The issue of immediate atherosclerosis screening was first addressed by Candemir [126]. They began with a deformed mean shape model that uses coronary ostia and cardiac chambers as anchor points to locate and extract the coronary arteries from CCTA images. The initial estimation of the centerline was then refined through region-of-interest masks. Next, the vessels are defined as the volumes around the centerline, and the surrounding area of the vessel. Eventually, they produce a longitudinal view of straightened MPR volumes. They then move to a pre-processing step to increase the performance of their eventual CNN. Considering the arteries, they can have varying lengths. They cannot be introduced into a CNN until they are resized, but to avoid any loss of information, they add empty frames at the extremities of shorter arterial volumes. Finally, they augment the data size by rotating MPR volumes between 0° and 360° around the vessel centerlines.

The final data was then processed with a 3D CNN with five-fold cross-validation , resulting in an Accuracy of 90.9%. The positive predictive value (PPV) was 58.8%, Sensitivity was 68.9%, Specificity was 93.6%, and negative predictive value (NPV) was 96.1%. Finally, they used Grad-cam [131] for a weakly supervised abnormality localization to better understand the results. The algorithm produced a map emphasizing the significance of each pixel on the image, providing visible indicators to the most active regions, which hopefully would refer to the presence of atherosclerosis in diseased cases.

Another recent study uses a 3D CNN for coronary artery analysis. The network first extracts image characteristics from singular sections through the MPR volume of the artery,

then processes the features with a recurrent neural network [105]. The proposed system can identify plaque, determining its type and stenosis along with its significance.

We explored a different approach for atherosclerosis screening from CCTA images. We started by applying Transfer learning to evaluate pretrained models on our dataset. Transfer learning has proven to be very efficient, not only for optimizing the time of experimentation but also redressing the shortage of data required for training a new model solely dedicated to the task.

In this chapter, we provide an in-detail description of the proposed method. We start by describing the used data, briefly describing transfer learning, and the models we chose to evaluate. After deciding on the best model, we will discuss the problem of "data imbalance" and present the proposed solution.

3.2 Data

The dataset used in this chapter is an open-source dataset containing Coronary CT Angiography images designated for atherosclerosis screening.

It was gathered and arranged by a research group at Ohio State University Wexner Medical Center [132]. It is a collection of Mosaic Projection View (MPV) Coronary artery images from 500 patients. Each set is composed of 18 views of a straightened coronary artery, stacked vertically.

The Mosaic Projection Views were created from a 3D to 2D projection. First, every volume is rotated, creating unique ray-traced (RT) projections every 10° (0–180°). The projections are then combined by averaging the overlapping intensities, straightened, and arranged in vertical series of 18 images as shown in Figure 3-1.

The dataset was repartitioned as 300 for training, 100 images for testing, and 100 for validation. The validation dataset consists of one randomly selected artery per normal case (50 images) and one diseased case (50 images). The training images were augmented 6-fold to create 2 364 more images to balance the dataset and re-enforce training. The

augmentation was not performed on the 2 304 normal images in the data set, the whole validation dataset, or the whole testing dataset.

We used different data augmentation techniques to increase the network's performance further: Random reflection horizontally and vertically with 50% probability, random rotation between -90 and 90 degrees, and random scaling between 1 and 4.

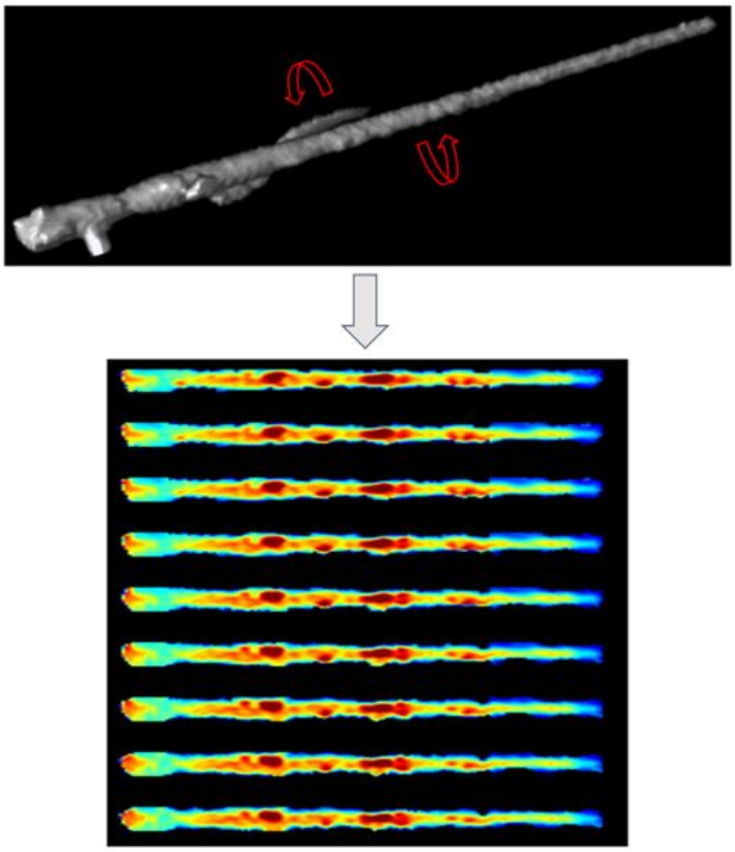

Figure 3-1 Mosaic Projection Views of the coronary arteries as described in the original paper [132]

3.3 Transfer Learning

Atherosclerosis screening is a relatively new task and has not been thoroughly studied. Therefore, we decided against building a new CNN for the task. Transfer learning can easily achieve the same performance, if not better, considering the small dataset.

Transfer learning is one of the most powerful concepts in machine learning. It takes knowledge that a network learns training on big data, then applies it to another task by retaining the weights of the initial layers and only training the final layers with new data for better fine-tuning. It is typically used in one of three cases :

- A shortage in annotated data to train a model from scratch, which is almost always the case in medical applications. Starting with an already trained model with basic knowledge of shapes and patterns can put the little data that is available to better use. It is used only for fine-tuning and perfecting the model's performance.
- A limited time limit. It is never an easy task to build a brand-new model and train it from scratch. There are too many parameters to play around before landing on the perfect architecture for both the data and the task at hand. Testing with transfer learning can at least stir the research towards the most appropriate models.
- Limited computing resources, especially when data isn't a lack in data. Handling bigger datasets and complicated tasks requires powerful computers that aren't always available to researchers.

All the known algorithms are trained on big datasets such as Imagenet [133], Ms coco [134], and Pascal [135]; they are ready to be used on chosen data following one of the following methods:

- Freeze all the weights, remove the last classification layer, and replace it with a classification layer appropriate to the target task.
- Calculate a function from the input to the last layer and use it as a shallow network.
- In some extreme cases, the weights of the trained network can be used as an initialization for training instead of random initialization.

Transfer learning not only answers the lack of data issue but can also add flexibility to the model. For example, no matter how extensive and varied the dataset is. It is inevitable to encounter new pathological cases which have anatomical properties different than any image on the dataset. Such images can puzzle the system. However, models trained on regular images have a more flexible nature and can recognize more details in the image to help them decide better. Therefore, a pre-trained model is theoretically more robust and generalizes better.

3.3.1 Finding The Right Model

Like end-to-end machine learning projects, not every architecture is suitable for all kinds of images. Therefore, it is crucial to find the suitable model for the data. For this purpose, we experimented on several architectures. The choice was not random; the models were picked mindfully depending on previously known advantages of each model.

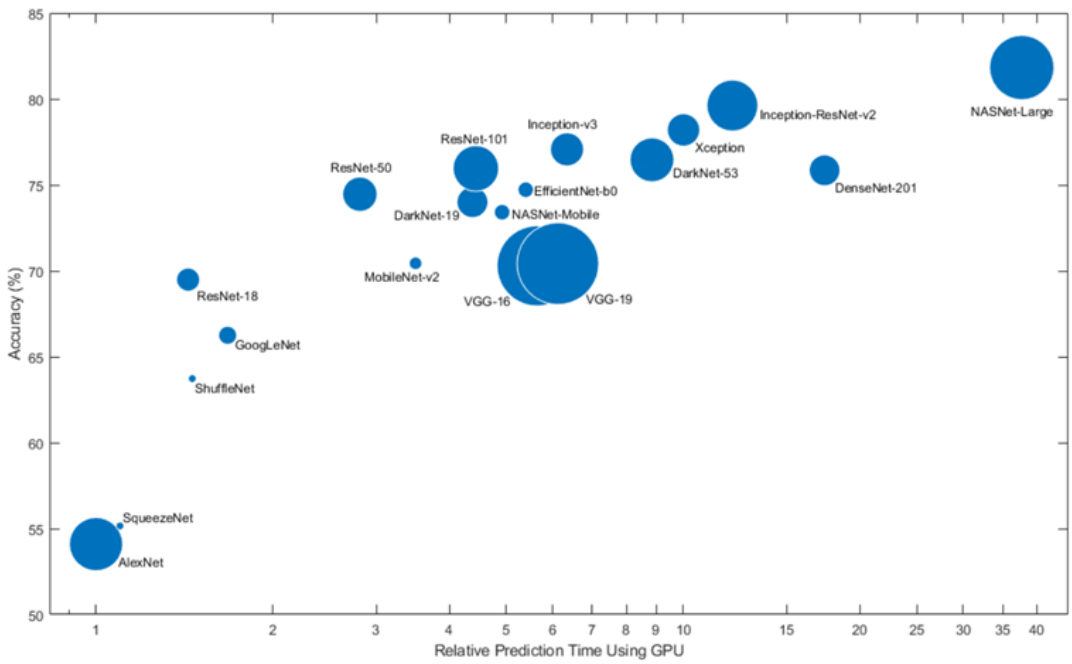

Figure 3-2 Performance of different known pretrained models on Imagenet [136]

Figure 3-2 shows the performance of a group of the most known and powerful architectures trained on the Imagenet dataset. In addition, it shows a prediction of the accuracy of each model and the relative predicted time to reach it by training with a GPU.

ImageNet is a famous dataset containing more than 14 million annotated images following to the WordNet hierarchy. The dataset has been used in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) since 2010 . This dataset spans 1000 object classes and contains 1,281,167 training images, 50,000 validation images and 100,000 test images. [137] The database is publicly available at http://www.image-net.org.

The personal computer we used for the entirety of this thesis is an Intel I5 8th generation, with 8 Go of RAM, an integrated GPU (NVIDIA GEFORCE 940MX). When choosing the models to experiment with , we were restricted by the limitations of the machine, so we chose models that respected a set of criteria:

- CNNs have proven very efficient for computer vision problems. Therefore, we tested convolutional neural network-based models .
- Not too deep: considering the relatively small size of the dataset on hand, we had to be careful not to overfit the models prematurely.
- Fast to train, but with relatively good accuracy. It was important to pick models that are fast to train but that had good performances.

After considering the conditions above, we settled on 6 models for testing: AlexNet, VGG Net, GoogleNet, ResNet, Inception, and Inception-ResNet. Of course, every model is different in structure and function.

3.3.1.1 AlexNet

AlexNet was the best-performing architecture in the 2012 Large Scale Visual Recognition Challenge (ILSVRC). It has one of the most straightforward architectures, with about 60 million parameters and only 8 hidden layers. It is composed of 5 convolution layers followed by three fully connected layers and pooling layers [22]. It was first trained on a restricted

subset of the ImageNet data, provided by the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC).

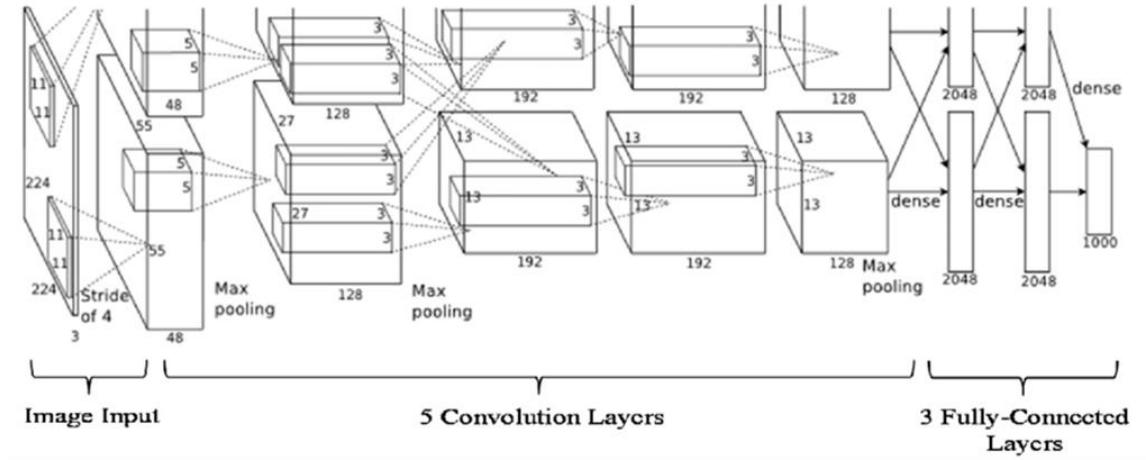

Figure 3-3 AlexNet basic architecture as cited in the original paper [22]

The training was partitioned across two different GPUs. Due to the lack of on-chip memory on the GPU, the authors used (an NVIDIA GeForce GTX 580).

The network consists of 7 layers, with the first 5 being convolutional (including some with max-pooling) and the last 2 being fully connected. The final output layer is a 1,000-unit softmax layer, used to classify the 1,000 different image classes. A diagram of the network can be found in Figure 3-3 as featured in the corresponding paper [128]. Note that many layers are split into two parts, corresponding to the two GPUs.

The input layer contains 224×224×3 neurons, representing the RGB values for a 224 × 224 image. All images had to be rescaled, so the shorter side had a length of 256. They would then be cropped out into a 256 × 256 area in the center of the rescaled image.

The first hidden layer is a convolutional layer with a max-pooling step. The layer has 96 feature maps in total which are divided into two groups of 48 each, with half being processed

on one GPU and the other half on the other GPU. The max-pooling in this and the following layers are performed in 3x3 regions, with a 2-pixel overlap between them.

The second hidden layer is also a convolutional layer, with a max-pooling step. It contains 256 feature maps, distributed evenly across both GPUs. The following three layers, the third, fourth, and fifth, are also convolutional layers but do not include max pooling. Their specific parameters are: the third layer has 384 feature maps, with 3x3 receptive fields and 256 input channels, the fourth layer has 384 feature maps, with 3x3 receptive fields and 192 input channels and the fifth layer has 256 feature maps, with 3x3 receptive fields and 192 input channels. It's worth noting that the third layer necessitates some inter-GPU communication for access to all 256 input channels, as depicted in the figure.

The final two hidden layers in the architecture are fully connected, each containing 4,096 neurons. The last layer, the output layer, comprises a 1,000-unit softmax layer.

To accelerate the training process, the network employed rectified linear units. However, with approximately 60 million parameters, it was prone to overfitting, despite the large training set. To counter this issue, the training set was expanded by randomly cropping and horizontally reflecting 224x224 sub-images from the original 256x256 images, which were then used as inputs for the network. Additionally, overfitting was mitigated by applying a variation of L2 regularization and dropout. Finally, the network was trained using a momentum-based mini-batch stochastic gradient descent method.

The model’s simple structure makes it the fastest to learn. Therefore, it is best suited to choose the initial preferences for fast run-through parameters.

Table 3-1 Results of the tests using AlexNet

Mini-Batch size	**Number of epochs**	**Initial learning rate**	**Dropout**	**Accuracy**	**Training time**
100	5	1e-4	/	83.71	88min 21sec
100	8	1e-4	/	90.18	142min 0 sec
64	20	1e-4	/	82.28	439min 18sec

64	20	1e-4	0.5	89.50	140 min 18 sec

The table above reports the results of some of the tests we ran using AlexNet; two main results are to be concluded:

- Training on bigger mini-batch sizes has better results
- Dropout improves accuracy and reduces training time significantly

3.3.1.2 VGG Net

The VGG or VGG-19 network was created by Simonyan and Zisserman from the Visual Geometry Group (VGG) at the University of Oxford in 2014 [23,23]. VGG Net managed to take second place in the 2014 ILSVR Challenge. The model has a uniform architecture with 19 stacked hidden convolutional layers and contains 138 million parameters, which makes it hard to handle and slow to train. However, we still chose to experiment with it because it is a strong model for feature extraction.

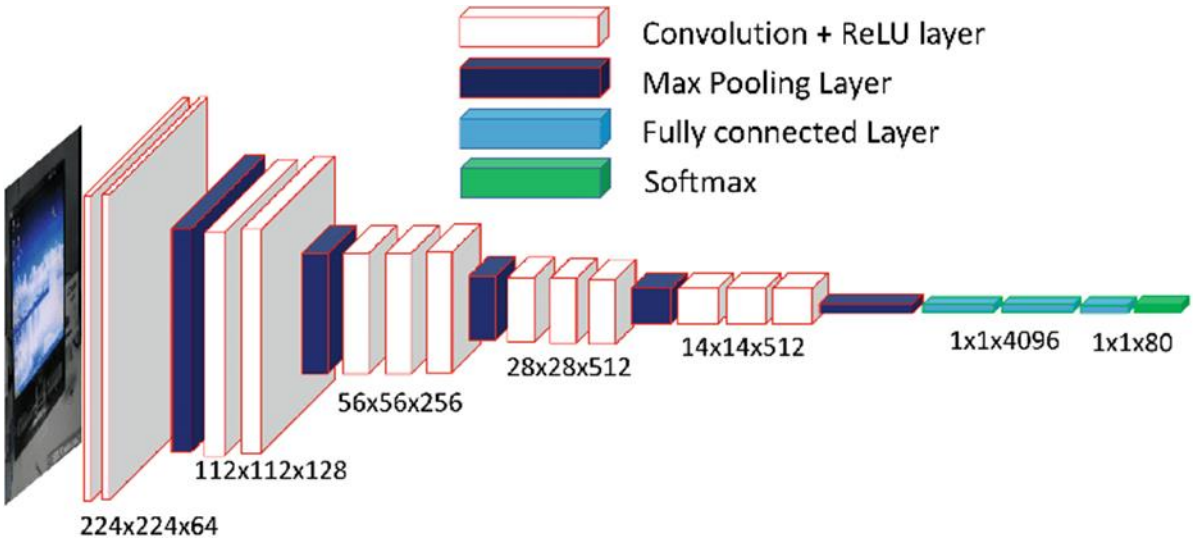

Figure 3-4 The VGG Net architecture as described in the original paper [23,23]

The convolutional layers are more efficient in representing images for classification purposes. After the convolution process, each image is transformed into a vector of 4096 features, which is a significant reduction from its initial representation as a vector of 154,587 features.

The two fully connected layers and the softmax layer are similar to a multilayer perceptron and can be replaced by other classifiers such as Random Forests or Support Vector Machines. However, these layers are crucial for the training phase of the neural network.

Table 3-2 Results of the tests using VGG Net

Mini-Batch size	**Number of epochs**	**Initial learning rate**	**Accuracy**	**Training time**
128	3	1e-4	91.02	939min 33sec
128	4	1e-4	84.89	1001min 55sec

As shown in the table above, we trained the model for a short three and then four epochs. Still, the training time was significantly long. Furthermore, training the longer model caused the system to crash due to many generated parameters and limited computing power.

3.3.1.3 GoogleNet

It was the first version of the Inception network, developed by the Google team, and won the 2014 ILSVRC. It is a relatively deep model with 22 hidden layers but has a reduced number of parameters (4 million). This is due to the use of 1x1 convolutions, which reduces the size of the matrices [24]

With the advances and complications of the task at hand, AlexNet started falling short, and they cost too much computational power once you started getting deeper.

In 2015, Google tried to build deeper models without sacrificing cost. They came up with a special convolutional layer called the inception layer. It is basically a network-in-network

system (Figure 3-5); Each inception module contains a few convolution layers in parallel that go through a dimensionality reduction step through 1 × 1 convolutional layers, and there is one max pooling layer. These layers get concatenated before being passed on to another inception or a regular module. The 1 × 1 convolutions reduce the dimensionality. This gets passed on to the more expensive 5 × 5 and 3 × 3 layers.

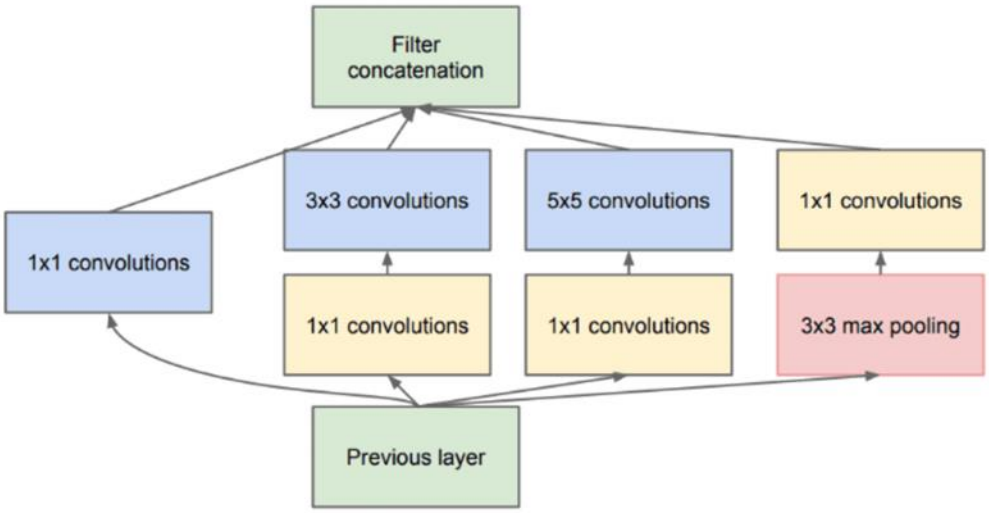

Figure 3-5 The building block of GoogleNet

The network is 22 layers , without the pooling layers. With such a deep network, there arises a problem of vanishing gradients. A vanishing gradient is a problem of depth wherein the errors are not strong enough to produce strong gradients to move the weights in any direction. To avoid this, in between, layers are branched off into one fully connected softmax layer. These classifier layers aim, not to perform better in label accuracies, but to add more errors so that some gradients produce additional discriminative features; they act as regularizers.

GoogleNet has traditional convolution layers at the beginning followed by inception modules. The traditional convolution layers produce activations that can be easily clustered

using the 1 × 1 layers. In addition, the fully connected layers were trained with a 0.7 dropout unlike the 0.5 of AlexNet and a 0.9 momentum.

Having a relatively short training time compared to its high performance, GoogleNet was an obvious choice for our experimentation. We only ran one test with a mini-batch size of 64 and an initial training rate of 10^{-4}. After training for 10 epochs, the model rendered a poor accuracy of 79.09 %.

3.3.1.4 ResNet

Ever since the introduction and success of the AlexNet, most winning models resorted to adding more hidden layers and going deeper to learn better. Unfortunately, although it prevents the model from overfitting the dataset, this method results in one of two problems, vanishing and exploding gradients. This happens when while training a very deep network, the derivatives or the slopes can sometimes get either extremely big or extremely small, even exponentially small, and this makes training difficult.

Proposed by He et al. [25], residual networks or ResNets are a novel architecture, which creates new branches in the architecture, whereas one branch (an identity block) the information is forwarded without changes, and in the other (the convolutional block) it is processed as would be in a typical layer. The unprocessed data or the “residual” is added to the original signal going through the network unaltered. This split in the network means one branch simply propagates the gradient without altering it. Deep residual networks are built by stacking these blocks together allowing them to learn with strong gradients passing through.

The concept of residual blocks is illustrated in Figure 3-6. The "jump connection" as it is called, is in the heart of the residual blocks. It keeps X nonintact to later add it to the outcome. The skip connection causes the output to vary from the traditional approach where the input 'X' is multiplied by the layer weights and then a bias term is added.

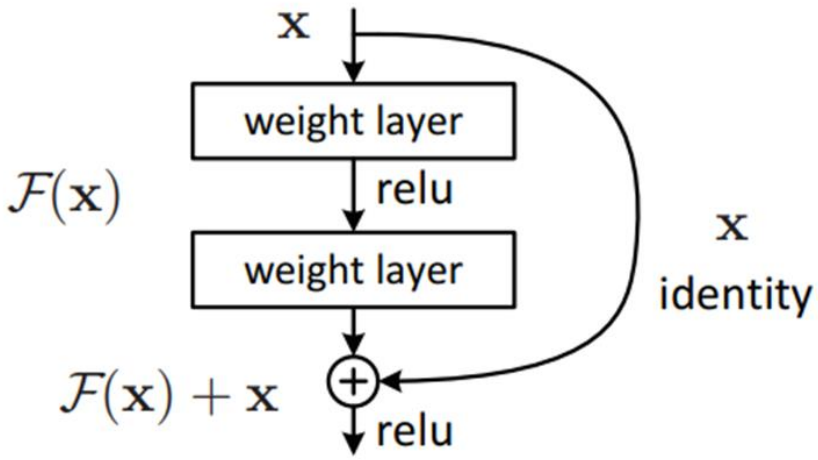

Figure 3-6 The main structure of a residual block

It has been established that ResNets are easy to optimize, and their performance is directly proportional to the increase in their depth [25]. That is why we used two different versions of the network: ResNet-50 which has 50 layers, and ResNet-101 which has 101 layers.

Table 3-3 Results of the tests using ResNet-50 and ResNet-101

	Mini-Batch size	Number of epochs	Initial learning rate	Dropout	Accuracy	Training time
ResNet 50	128	4	1e-4	/	88.08	381min 56 sec
	256	10	1e-4	/	91.02	1099min 55 sec
ResNet-101	64	6	1e-4	/	90.43	310min 18 sec
	32	10	1e-5	/	91.10	329min 54 sec
	32	25	1e-5	0.2	91.94	1471 min 6 sec
	64	10	1e-5	0.1	86.15	2135 min 14 sec

32	10	1e-4	0.3	94.71	1332 min 52 sec
64	20	1e-4	0.5	94.29	
64	20	1e-3	0.5	89.50	140 min 18 sec
64	20	1e-4	0.5	95.21	2999 min 17 sec
128	25	1e-5	0.5	91.18	3083 min 11 sec

3.3.1.5 Inception Net

A common challenge with deep convolutional neural networks is that the number of feature maps tends to grow as the network's depth increases. This can significantly raise the number of parameters and computation time when larger filter sizes, such as 5x5 and 7x7, are used.

A 1x1 convolutional layer is designed to address the problem of dimensionality reduction. This layer, also known as a feature map pooling or projection layer, performs channel-wise pooling, which reduces the number of feature maps while preserving important features. It can also be used for a direct one-to-one projection of feature maps to combine features across channels or to increase the number of feature maps, as in traditional pooling layers.

The Google team had developed new and improved versions of GoogleNet. There was Inception v2 and Inception v3 [138], Inception v4 and Inception-ResNet [139]. Each version offers an improvement over the last one. The original GoogleNet had 9 Inception modules.

The Inception net relies on a bunch of concepts and modules that make it what it is. Mostly the **1 x 1 Convolutions**.

Inception Net v3 was a revision of the original inception network, introducing new features such as the RMSProp Optimizer, Factorized 7x7 convolutions, and BatchNorm in the Auxiliary Classifiers. Additionally, Label Smoothing was added as a regularizing

component in the loss formula to prevent the network from becoming too confident in a class and to prevent overfitting.

3.3.1.6 Inception-ResNet

Inception Net v3 uses residual connections to combine the output of the convolution operation in the inception module with the input. To maintain consistency in dimensions, 1x1 convolutions are applied after the original convolutions, adjusting the depth to match the input (since the depth increases after the initial convolution).

The slowest network of them all. We chose it to study the utility of using complicated networks, and whether it would make much difference to the final result. A powerful and complex architecture that is probably over-qualified for our data. But we decided to test it to stay out of the speculative.

Table 3-4 Results of the tests using Inception-ResNet

Mini-Batch size	**Number of epochs**	**Initial learning rate**	**Dropout**	**Accuracy**	**Training time**
32	6	1e-4	0.35	87.24	1644 min 29 sec
16	10	1e-4	0.4	90.93	2391 min 30 sec

As mentioned before, the Inception-ResNet is rather a deep and complicated network, and it tends to overfit rapidly. Relatively large mini-batch sizes wouldn't allow the computer to memorize the weights during training. We had to work with small batch sizes of 16 and 32. We trained for 10 and 6 epochs, respectively.

3.3.2 Fine-Tuning Hyperparameters

When there is enough data, we can use simple algorithms with little-to-no tweaking. Nevertheless, when the dataset is small, it is particularly important to carefully tune the architecture and the hyperparameters, which can make a significant difference.

Applied deep learning is a very empirical process. Many parameters must be tried to find what works best for the problem. One healthy habit in machine learning is having a clear-eyed understanding of what to tune in order to achieve a certain effect. This is what is called "orthogonalization".

The right model needs to be finely adapted to the data in hand. For this purpose, the main metrics to tweak are :

- **Minibatch size**: a "minibatch" is a subset section from the training dataset built by shuffling and partitioning the dataset. Its size can vary from one image to the total number of examples in the training dataset. Common choices are 32, 64, and 128 elements per minibatch. A bigger batch size computes faster, which is more efficient, while a smaller batch size converges faster and generalizes better. However, it is often reported that when increasing the batch size for a problem, there exists a threshold after, which is a deterioration in the quality of the model.

  The perfect balance is a batch size that significantly speeds up training without sacrificing model accuracy. In our case, we tried 32, 64, and 128. And eventually settled on a batch size of 64.

- **Learning rate**: the hyperparameter that controls how quickly model learns. It determines the amount of error that is applied to the weights of the layers during each update. A higher learning rate allows for faster learning but may not converge to the optimal weights. Lower learning rates provide more optimal learning, but require more time and computational resources. We experimented with different learning rates, including 10^{-3}, 10^{-4} and 10^{-5}. We found that using the lowest learning

rate 10^{-5} resulted in a significant increase in training time, so we decided to use 10^{-4} for a balance of better performance and faster training.

- **Number of training epochs**: how many times the model goes through the dataset can impact its performance. Training for too many epochs can lead to overfitting, while not training for enough can result in underfitting. Determining the optimal number of epochs can be difficult, so it is advisable to monitor the learning curve during training.

- As shown in Figure 3-7, reaching 25 epochs improves training performance without causing overfitting. On the other hand, using only 3 epochs may not be sufficient as it only reaches 91% accuracy.

- **Dropout :** Dropout is a method of regularization that temporarily removes certain neurons from the network during training, according to a pre-set probability, preventing them from participating in forward and backward propagation. This dynamic alteration of the model's architecture leads to more robust training. However, using dropout throughout the entire training process can make it difficult to evaluate the model's performance, as dropout should be disabled during testing.
- In our experiment, we first trained the model without dropout, using a fixed network structure. Then, we tested different levels of dropout (10%, 20%, 30%, and 50%) and found that the highest performance was achieved with a 50% dropout rate.

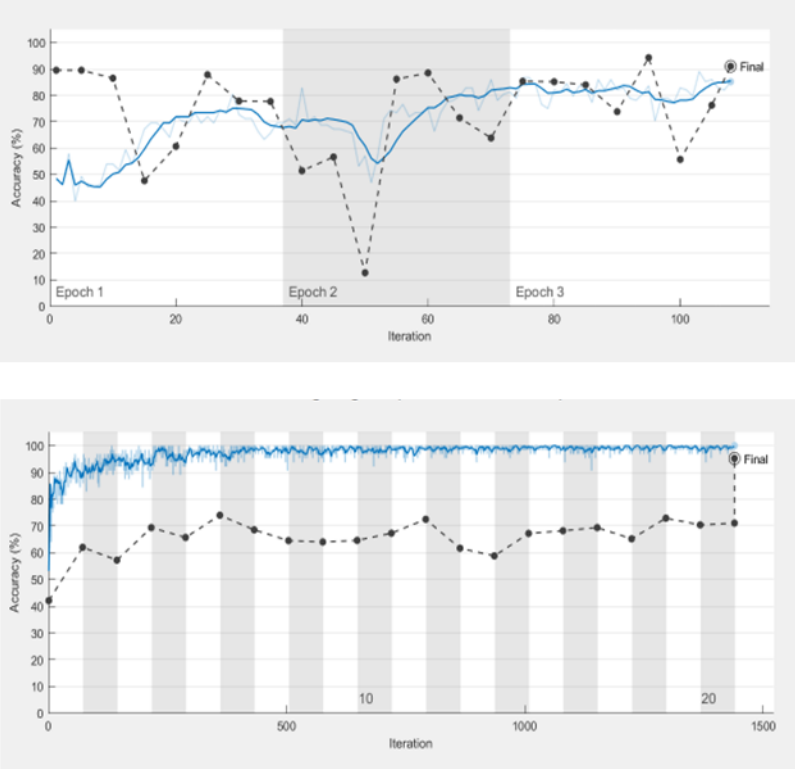

Figure 3-7 A basic illustration for the impact of number of epochs on the stability of the model. The figure on top shows the model after training for 3 epochs. The figure on the bottom shows the model after training for 20 epochs.

3.3.3 Performance Evaluation

It is not always possible to find a model that satisfies all the requirements of a task; To compare the performance of different models, it is useful to set up satisficing as well as optimizing metrics. For instance, let us take accuracy and running time as examples. We usually aim to maximize accuracy but in an acceptable time frame.

In medical applications, every case matters. Therefore, it is important to understand when a model works and when it does not. Even though accuracy can be a good metric to initially evaluate performance, in medicine, other metrics are to be considered, sensitivity, specificity, and prevalence. Sometimes there are different expectations for metrics;

maximizing accuracy is an optimizing metric, while minimizing running time is a satisfying metric.

Accuracy is the probability of the prediction being correct. It considers the classification of each case.

Sensitivity , also known as true positive rate, is the likelihood that the model will correctly identify a patient as having the disease when they actually do have the disease. It measures the proportion of positive cases that are correctly classified as positive.

$$\textbf{Sensitivity} = \frac{\mathbf{TP}}{\mathbf{TP+FN}} \qquad (3\text{-}1)$$

Specificity , also known as true negative rate, is the likelihood that the model will correctly identify a patient as healthy when they are actually healthy. It measures the proportion of negative cases that are correctly classified as negative.

$$\textbf{Specificity} = \frac{\mathbf{TN}}{\mathbf{TN+FP}} \qquad (3\text{-}2)$$

The proportion of patients with a positive diagnosis in a given population is referred to as **prevalence.**

Accuracy can be expressed in terms of sensitivity, specificity, and prevalence, which allows us to view it as a combination of sensitivity and specificity with specific weightings. The expression is stated in equation (3-3).

Accuracy = Sensitivity * prevalence + Specificity * (1 – prevalence) (3-3)

In the clinic, practicians using artificial intelligence models are usually interested in knowing what the probability of the patient having the disease if they are classified as positive, which is referred to as the positive predictive value (PPV) of the model. Similarly, they may be interested in the probability that a patient is healthy when the model's prediction is negative, known as the negative predictive value (NPV) of the model.

Confusion matrix assembles the forementioned metrics in a table to better evaluate the classifier or model.

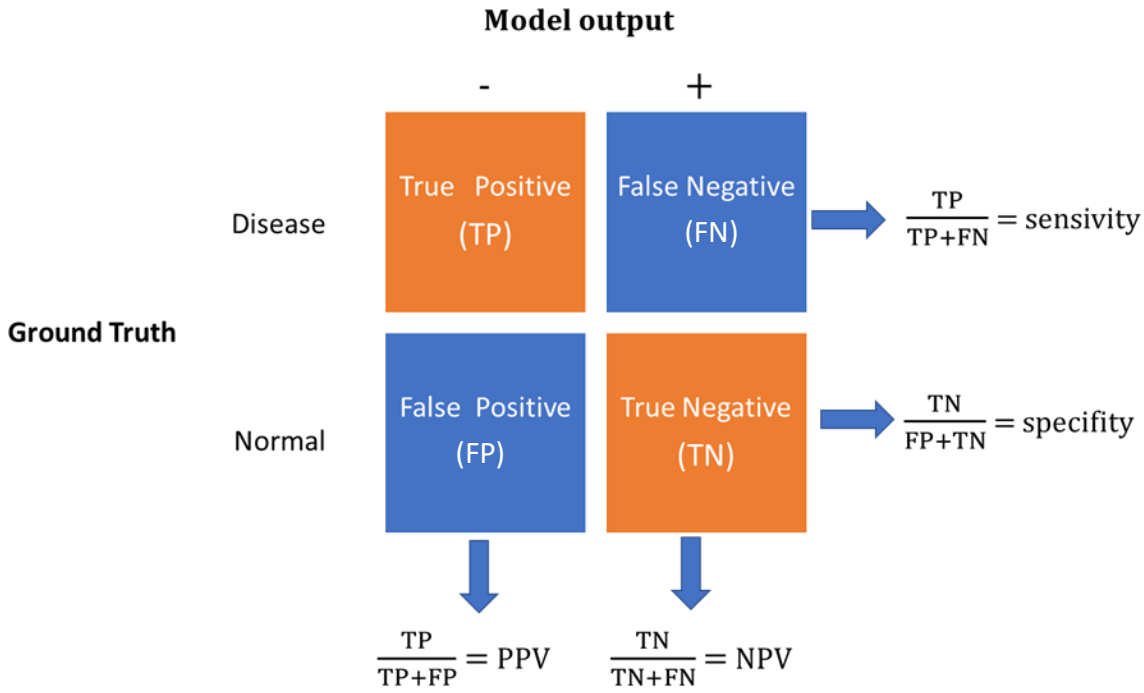

Figure 3-8 A detailed description of the confusion matrix

True Positive [TP]: The model's prediction is “Positive” and matches the actual outcome.

True Negative [TN]: The model's prediction is “Negative” and matches the actual outcome.

False Positive [FP]: The model's prediction is “Positive” but contradicts the actual outcome.

False Negative [FN]: The model's prediction is “Negative” but contradicts the actual outcome.

3.3.4 Results and Discussion

The model that performed best on our data was a Residual Network of 101 layers. Therefore, we considered a mini-batch size of 64, not too big, not too small.

We used a learning rate of 10^{-4} which allowed the model to learn fast enough without missing local minima.

We used 0.5 dropouts, which made the model more universal.

The model trained for 20 epochs which made it perform well without overfitting.

The model took about 50 hours to train. It achieved 95.2% accuracy. 99.25% specificity, 60.8% sensitivity, 90.48% positive predictive value, and 95.6% negative predictive value.

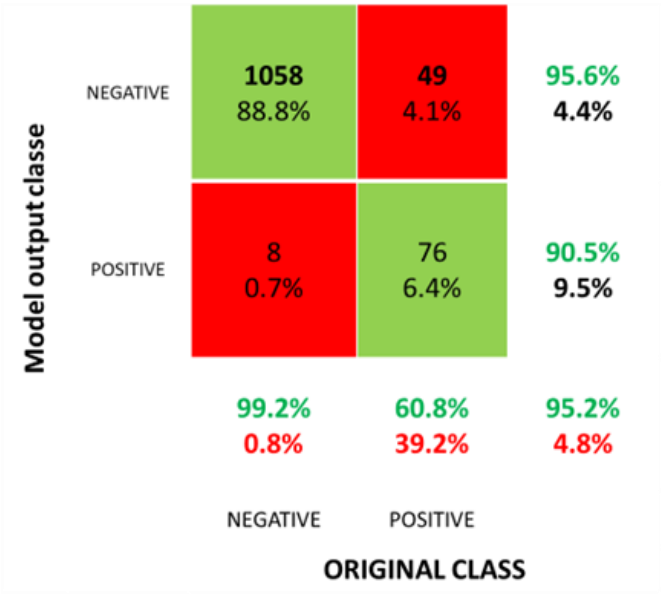

Figure 3-9 The final model's confusion matrix

Transfer learning is one of the most powerful new concepts in deep learning. It allows better data recognition, even when the training data is not big enough. However, similar to regular learning, finding the right architecture and tweaking it finely enough to suit the data

at hand is essential. After testing different variations of the models mentioned above which went on for days, we finally settled on ResNet.

Testing different architectures is not enough. Once we settle on a pre-trained model, it is important to fine-tune it again to find the perfect hyperparameters.

After thorough testing and fine-tuning we settled on using a ResNet-101, with a minibatch size of 64, a learning rate of $10e^{-4}$, and a 50% dropout. We trained the model for 20 epochs.

This model achieved 95.21% accuracy, 99.25% specificity, 60.8% sensitivity, 90.48% positive predictive value, and 95.6% negative predictive value.

It is essential to point out that the model provides a high NPV, enough to confirm negative cases and clear out much-needed beds in the emergency room [128]. Eliminating patients with chest pain from emergency rooms with a high level of certainty (narrow margin for error) is an accomplishment. This means that the model can be implemented as it is for clinical use. Every other metric is relatively high and comparable to human performance, except for sensitivity. So **WHY** is the sensitivity low? Is it acceptable to work with the model as it is?

Sensitivity is, by definition, the probability that positive cases are correctly classified. Conversely, poor sensitivity means a poor classification of positive cases, which can be due to the lack of positive images in the dataset.

Under the assumption that the lack of positive cases in the dataset is the reason behind the low sensitivity, we proceed in the next section to solve that problem.

3.4 Data Imbalance

The imbalance related to medical datasets is referred to as “intrinsic”, which arises from factors such as limited time, storage, or other restrictions on the dataset or data analysis. Intrinsic imbalance is a direct result of the inherent nature of the data space. It reflects the prevalence or the frequency of disease in the real world, where healthy cases are much

more present than pathological cases in the population, even when under the same health risks and triggers, and presenting the same symptomology. In medical applications, misclassification can be a problem of severe consequences. Classifying positive cases as normal leads to discharging sick patients, which puts their lives at risk.

Using one evaluation metric, accuracy in our case, to assess the network is not a safe practice with imbalanced data. Specificity, sensitivity, positive predictive value, and negative predictive value are all important. Furthermore, classifiers and models tend to ignore small classes while learning better from the large ones.

3.4.1 Data Augmentation Techniques

In their paper "Learning from Imbalanced Data" [140], the authors discussed the issue in detail. They proposed several data augmentation solutions. That we have considered but were not enough.

3.4.1.1 Random Oversampling and Undersampling

Random oversampling follows a simple rule where a random selection of examples E from the minority class S_{min} is replicated and then added to the class. Consequently, the number of examples in the dataset S is increased by E, and the data distribution is adjusted accordingly [141].

On the other, random undersampling, removes data from the original dataset following the same principle. Randomly selected examples E from the majority class S_{maj} are removed, decreasing the size of the original dataset S by E examples [141].

As simple and efficient as the method seems, it has its limitations. In the case of undersampling, removing examples from S_{maj} can lead to removing important examples that lead the model to correctly classify, which means reducing the model's performance on the majority class. Oversampling, on the other hand, would expose the model to a set of replicated data, making it learn from the same images, leading to overfitting. Although this may increase the training accuracy, the model would perform worse on unseen data [142].

3.4.1.2 Informed Undersampling

As cited in [140], many methods have been developed to make undersampling less random and avoid cutting off important data. The two models have been thoroughly tested and have shown the best results in informed undersampling [143] : the "Easy Ensemble" and "Balance Cascade" algorithms. "Easy Ensemble" samples multiple subsets from the S_{maj} , the majority class, and develops several classifiers based on the combination of each sus=best with the minority class.

To summarize, all the methods mentioned above, and many others, were developed to make undersampling less random and avoid cutting off essential data.

Undersampling has shown promising results, but it is not as interesting to use in deep learning. Deep networks need more data, and there is not enough of it. In our example, the network can identify negative cases despite the data imbalance, which is useful in discharging risk-free patients. Such a feature is extremely useful, and it should not be sacrificed for the sake of improving the model's performance. The aim here is to improve the performance of the minority class (positive atherosclerosis) without loss in the performance of the majority class (negative atherosclerosis). Therefore, any form of undersampling is unacceptable.

3.4.1.3 Data Augmentation

One commonly used way to deal with insufficient or/and imbalanced data is the so-called "*data augmentation*". As its name suggests, the method uses different techniques to increase the number of images on the dataset. The solutions can vary from simple image processing methods such as rotating, cropping, zooming, and histogram-based methods to Generative Adversarial Networks.

Simple Image transformations like rotation, zooming, and increasing the dataset size through data augmentation and image synthesis, can make the model more resilient and less susceptible to adversarial attacks [144].

There are methods for image augmentation to create image diversity that have been widely used for data generation in CNNs. It not only provides new data, but also makes the network more robust and less vulnerable in case of adversarial attacks. Transformations include rotation, translation, zooming in and out, and sometimes simply replicating the images.

These techniques are not applicable in our case because of the nature of the images. As explained in the section, the dataset is a collection of 18 projections of coronary arteries, straightened and stacked vertically. Therefore, any physical change to the images can eventually alter the data. It would also provide images quite different from the original ones and cannot be replicated in real life.

3.4.1.4 Data Synthesis

Data synthesis in the medical dictionary refers to "a method that uses statistical techniques to combine results from different studies and obtain a quantitative estimate of the overall effect of a particular intervention or variable on a defined outcome—i.e., it is a statistical process for pooling data from many clinical trials to glean a clear answer". The data may vary from fake patient records to fake medical imaging.

As its name suggests, the techniques aim to synthesize new data. When acquiring more data became challenging, researchers resorted to using new algorithms for generating new images. The new data has the same distribution as the original data and can fill even an expert's eye.

3.5 Generative Adversarial Networks (GAN)

3.5.1 Intuition Behind GANs

Discriminative models learn to predict the class label given an input, such as classifiers. Generative models, in contrast, learn the distribution of the data they are trained on and can generate new samples from that distribution.

Generative Adversarial Networks (GANs) are a powerful type of generative model that can create highly realistic objects that are difficult to distinguish from real ones.

Data Augmentation Using GANs was first introduced in a paper in 2019 [145]. There are two models behind a GAN, the **generator** and the **discriminator**. They fight against each other, each trying to outperform the other, resulting in one of the models to being so good that it generates realistic images.

Generative models try to learn how to make a realistic representation of some class; meanwhile, discriminative models distinguish between different classes.

In summary, GANs learn to produce realistic examples without seeing real images, and the discriminator models learn to distinguish between real and fake. They improve over time by receiving more realistic images at each round from the generator and real images, all jumbled up in a pile. They try to develop a keener and keener eye as these images improve.

3.5.2 Discriminator

"The discriminator in GANs is a classifier that takes in d-dimensional inputs and produces a single output in the range of (0,1), representing the probability that the input example is real. A value of 1 means the input is real, while a value of 0 means it is synthetic."

The output probabilities from the discriminator are the ones that help the generator learn to produce better-looking examples over time.

3.5.3 Generator

It is a model used to generate examples from random noise, then improve them over time.

The generator samples from a random latent distribution and generates an image. The second network is a discriminator that tries to predict if an input image was generated by the generator network or was sampled from a dataset.

Suppose the data were sampled from the real world. In that case, we may also train a softmax layer with the final layer of the discriminator's representations to learn the features in a discriminative way.

3.5.4 Workflow of a GAN

As mentioned above, ANs consist of two neural networks, a generator (G(z)) and a discriminator (D(x)), that work in opposition to each other. The generator's goal is to create realistic images that can trick the discriminator into thinking they are real. On the other hand, the discriminator's objective is to accurately identify which images are real and which are generated. The generator and discriminator are trained to optimize different cost functions, with the generator trying to minimize it and the discriminator trying to maximize it. A mathematical representation of GANs was presented in [146] as follows:

Determining the fixed distribution of data, Pdata(x), can be challenging. Therefore, it is common to assume that Pdata(x) follows a Gaussian mixture distribution and use the maximum likelihood method. However, this can be difficult to calculate when dealing with complex models, which can limit the overall performance. As an alternative, the distribution Pg(x) can be calculated using artificial neural networks (ANNs).

The generator, an ANN with parameter θg, takes a random variable z from a given distribution and maps it to a pseudo-sample distribution using the ANN. The generated data is represented as G(z) and its distribution as Pg(z). With the parameter θg, various complex distributions can be generated from a simple input distribution. The goal of the generator is to minimize the difference between Pg(x), the generated images, and Pdata(x), the input image distribution

$$G \quad = argmin_G\left(P_g, P_{data}\right) \qquad (3\text{-}4)$$

As the precise forms of the distributions are not known, it is not possible to calculate the difference directly. To address this, another neural network, referred to as the discriminator, is created to learn the difference between the two distributions. The original GAN [147] employed a binary classifier [148] with the discriminator's parameters denoted

as θ_d, which outputs a value of 1 for a real sample x and 0 for any other input. The loss was measured using binary cross-entropy, a commonly used function in binary classification.

$$Loss = -\big(ylog(\hat{y}) + (1-y)log(1-\hat{y})\big) \qquad (3\text{-}5)$$

y : the label of a given sample

$\hat{y}$: the probability that the model's prediction for a sample is a positive example.

If the model correctly classifies a sample, $\hat{y}$ is set to 1; otherwise, it is set to 0. Positive cases are assigned to P_{data} and negative cases are assigned to P_g.

Basically, GANs are structured such that the Discriminator aims to minimize its output (D, G), while the Generator works to increase the Discriminator's output and thus to decrease its loss. To express this in a formula:

$$V(G,D) = E_{xPdata}[logD(x)] + E_{xPg}\big[log\big(1-D(x)\big)\big] \qquad (3\text{-}6)$$

Where:

P(z) represents the distribution of the generator

x is a sample from the distribution $P_{data}(x)$

z is a sample from the distribution P(z)

D(x) is the Discriminator neural network

G(z) is the Generator neural network

The generator G aims to maximize the output of the discriminator D when presented with a generated (fake) sample, in order to trick the discriminator. Meanwhile, the discriminator attempts to differentiate real data from generated samples. Therefore, the

discriminator aims to maximize V(G, D) while the generator aims to minimize it, creating a minimax relationship.

$$min_G max_D V(G,D) = min_G max_D E_{xPdata}[logD(x)] + E_{zPz}\left[log\left(1 - D\left(G(z)\right)\right)\right] \quad (3\text{-}7)$$

The training process of a GAN involves alternating updates to the generator and discriminator parameters, such as:

- The Discriminator is trained while the Generator remains inactive. During this phase, the network only goes through forward propagation and no backpropagation occurs.

  The Discriminator is trained on real data for n epochs to test if it can correctly identify them as real. It is also trained on fake generated data from the Generator to see if it can correctly identify them as fake.". (Figure 3-10)

- The Generator is trained while the Discriminator remains inactive.

  After the Discriminator has been trained on the generated fake data, the Generator can use these predictions to improve its performance by generating data that is more likely to fool the Discriminator. (Figure 3-11)

The process of alternately training the Generator and Discriminator is repeated until Pdata(x) approaches Pg(x) .

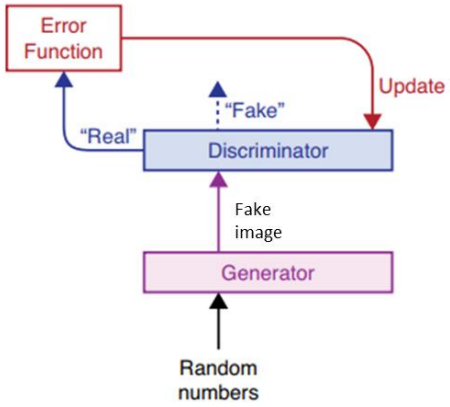

Figure 3-10 The GAN behavior if the discriminator labels the fake image as real, the discriminator is updated to better spot the fakes.

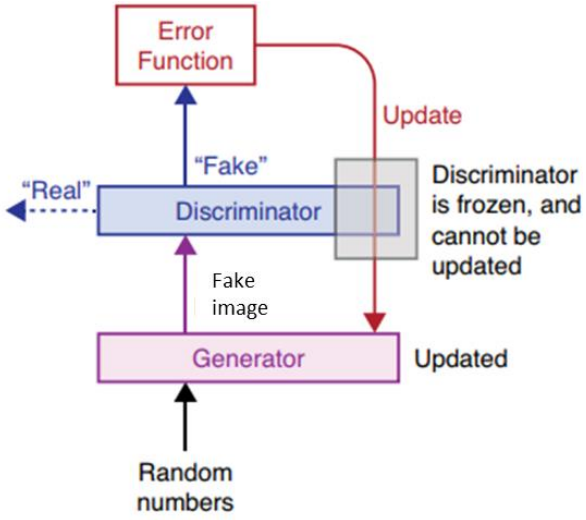

Figure 3-11 The GAN behavior during a fake image generation. The generator receives random numbers, which produces a fake image. If the discriminator labels it as fake, the generator is updated. But the discriminator is untouched.

GANs have become increasingly popular in image processing and computer vision as a method of data augmentation. This technique is particularly valuable in medical applications, where obtaining labeled medical data can be difficult.

GANs have been extensively utilized in the field of cardiology, such as creating realistic cardiac images [149], producing synthetic electrocardiography signals [150], and imitating electronic medical histories [151].

3.5.5 Model Training

We trained a GAN model using our dataset to generate new data. Our training process involved using a mini-batch size of 128 for 50 epochs until the model reached convergence, as illustrated in Figure 3-12. The Adam optimization algorithm was used with a learning rate of $2*10e^{-4}$, a gradient decay factor of 0.5, and a squared gradient decay factor of 0.999.

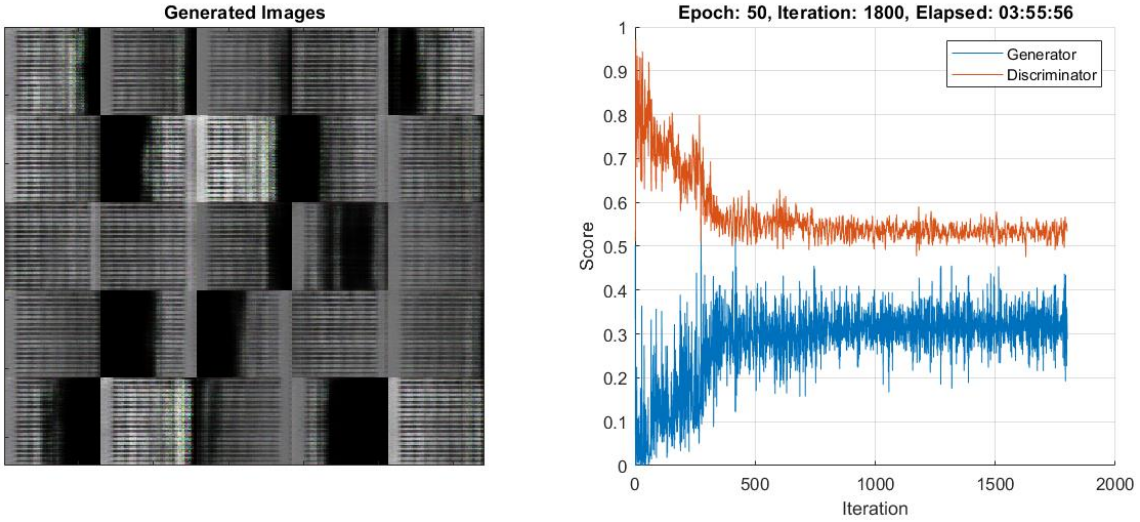

Figure 3-12 The advancement of the generator and discriminator throughout the training process of the GAN [3]

Examples of the created images can be seen in Figure 3-13, which showcases eight authentic positive images from the original dataset in the first row, eight generated positive images in the second row, eight authentic negative images from the original dataset in the third row, and eight generated negative images in the fourth row.

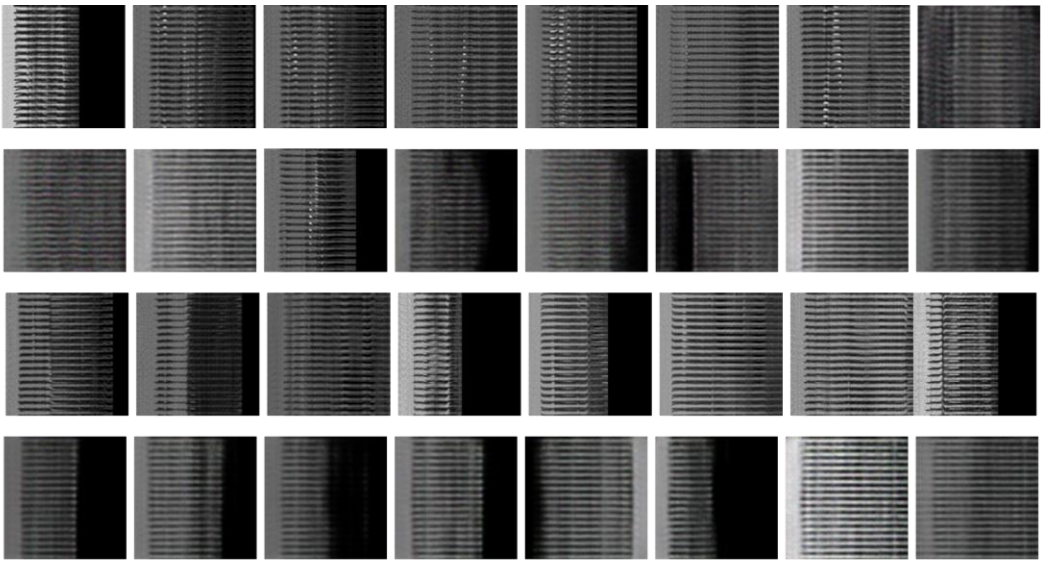

Figure 3-13 Samples from the training dataset. Row 1, original positive images from the dataset; Row 2, generated positive images; Row 3, negative images from the dataset, Row 4, generated negative images.

As outlined previously in section (3.3) , We used a ResNet-101 model on the original dataset [132]. That initial experiment yielded an accuracy of 95.2%, sensitivity of 60.8%, specificity of 99.26%, positive predictive value of 90.48%, and negative predictive value of 95.57%.

Using the images generated from the GAN, we worked on improving the performance of the same ResNet model, The results can be found in Table 3-5, which describe the outcome of the following tests:

In the first experiment, we added 100 generated images each to the positive and negative training folders, resulting in a slight drop in accuracy (94.0%) and an improvement in sensitivity (75.2%). Despite the small decrease in accuracy, the model's Positive Predictive Value dropped from 90.5% to 60.6%.

In the second experiment, we added 300 generated positive images to the training folder, which resulted in a significant drop in accuracy (86.9%) and sensitivity (61.6%).

Additionally, the Positive Predictive Value decreased from 90.5% in the first model to 43.4%

The third test involved relocating the 300 generated images from the previous test into the testing dataset to balance it out. The results showed an improvement in accuracy, sensitivity, and Positive Predictive Value, which were 88.2%, 75.6%, and 94.2%, respectively. This suggests that using generated data to balance the testing dataset is a better approach.

The fourth experiment involved replacing 1000 real positive images from the training dataset with 1000 generated images, while moving the real images into the testing dataset.

The processed increased the number of positive images in the test set, balancing it out. The results of this test showed an improvement in accuracy, sensitivity, and Positive Predictive Value, which were 88.2%, 75.6%, and 94.2%, respectively. This suggests that the best use of the generated data is to balance the testing dataset.

All results are summed up in Table 3-5. Our highest recorded outcomes were 93% accuracy, slightly lower than the accuracy of the original experiment, but with a sensitivity of 89% and a Positive Predictive Value of 97.1%

Table 3-5 The performance of the network after each test

	NPV	**PPV**	**Accuracy**	**Specificity**	**Sensitivity**
Original	95.57	90.48	95.2	99.25	60.8
Test 1	97.1	60.6	94.0	96.2	75.2
Test 2	97.6	43.4	86.9	87.5	81.6
Test 3	85.4	94.2	88.2	96.8	75.6
Test 4	89.9	97.1	93.2	97.4	89.0

Synthesizing or creating new images can be more effective than adding authentic data. The GAN generates images following a procedure of "fooling" the system into believing it is

real data. The iterative process allows the generator to learn the features of the data. This means the new images are more "positive" than any authentic positive images.

It can be considered an enhancement and boost to the performance of the classification model.

3.6 Conclusion

Atherosclerosis screening from CCTA images is a critical task in cardiovascular disease diagnosis and follow-up, even though it is a popular research subject. It is usually done after a step of plaque segmentation and classification. Screening straight from CCTA images can be faster and more efficient, especially in emergency cases.

In this chapter, we managed to do atherosclerosis screening straight from CCTA using transfer learning, a powerful tool in deep learning. After testing different pre-trained models, we achieved the best performance using ResNet101. The model scored 95.21% of accuracy, 99.25% of specificity, 60.8% sensitivity, 90.48% positive predictive value, and 95.6% negative predictive value. The negative predictive value qualifies the model to be used clinically to discharge cases of chest pain in emergency rooms.

The poor training can explain the low sensitivity of the model on positive images, which is due to the lack of them as opposed to negative images in the dataset. Intrinsic data imbalance is a problem open for research in medical applications. Many solutions have been suggested and tried throughout the literature.

Using generated data is a novel approach to data augmentation. However, it has proven efficient in different applications. In this section, we used GAN-generated images to balance the dataset. This was to increase the number of positive images in the test dataset under the assumption that the data imbalance caused the sensitivity of our model to be low.

The original model scored 95.21% accuracy and 90.48% PPV but only 60.8% sensitivity. The original test dataset contains 125 positive images and 1066 negative images. To overcome the issue, we used a Generative Adversarial Network to generate new images

resembling the original data. Through various experiments with the generated data, we were able to improve the accuracy, sensitivity, and Positive Predictive Value of our model. The final results, 93.2% accuracy, 97.13% Positive Predictive Value, and 89.0% sensitivity, are comparable to state-of-the-art methods and have significant clinical implications for the fast and accurate detection of this disease.

Chapter 4

Cardiac Structure Identification from Cardiac MRI

4 Cardiac Structure Identification From CMRI

4.1 Introduction

Like all medical applications, cardiac segmentation is a critical task, having an automatic efficient approach is not always enough; clinicians usually ask an explanation to go along with the answers provided by the model.

In order to provide machine learning algorithms with more credibility, recent research is headed towards finding new ways to make the algorithms more dependable. This can be done through a number of ways, making them interpretable, explainable, or trustworthy. Even though the terms can be found used interchangeably, they do not mean the same things. [152]

Interpretability in a model suggests it can be easy to understand how the model operates and the reasoning behind its results [153] .

Explainability offers a level of understanding of how a model operates and reaches decisions without explaining in detail how every step works [154]. Complex models, such as black boxes, can be difficult to comprehend, but techniques like partial dependence plots, SHapley Additive exPlanations (SHAP) dependence plots, and surrogate models provide insights into the relationship between input data and model outputs.

Trustworthy AI/ML, on the other hand, refers to the ability to understand and interpret the reasoning behind a model's predictions. This understanding helps to clarify the nature and behavior of the AI/ML model and improve trust and transparency in its decision-making process.

In our attempt to perform an automatic yet comprehensible segmentation of cardiac structures we simulate the steps taken by an expert from the moment they see the image, to the moment they decide on its segmentation, which is illustrated in the top half of Figure 4-1. Then, we replicated each step using a machine learning technique (lower half of Figure 4-1). Our proposal to mimic the doctor's approach and reasoning makes our AI trustworthy.

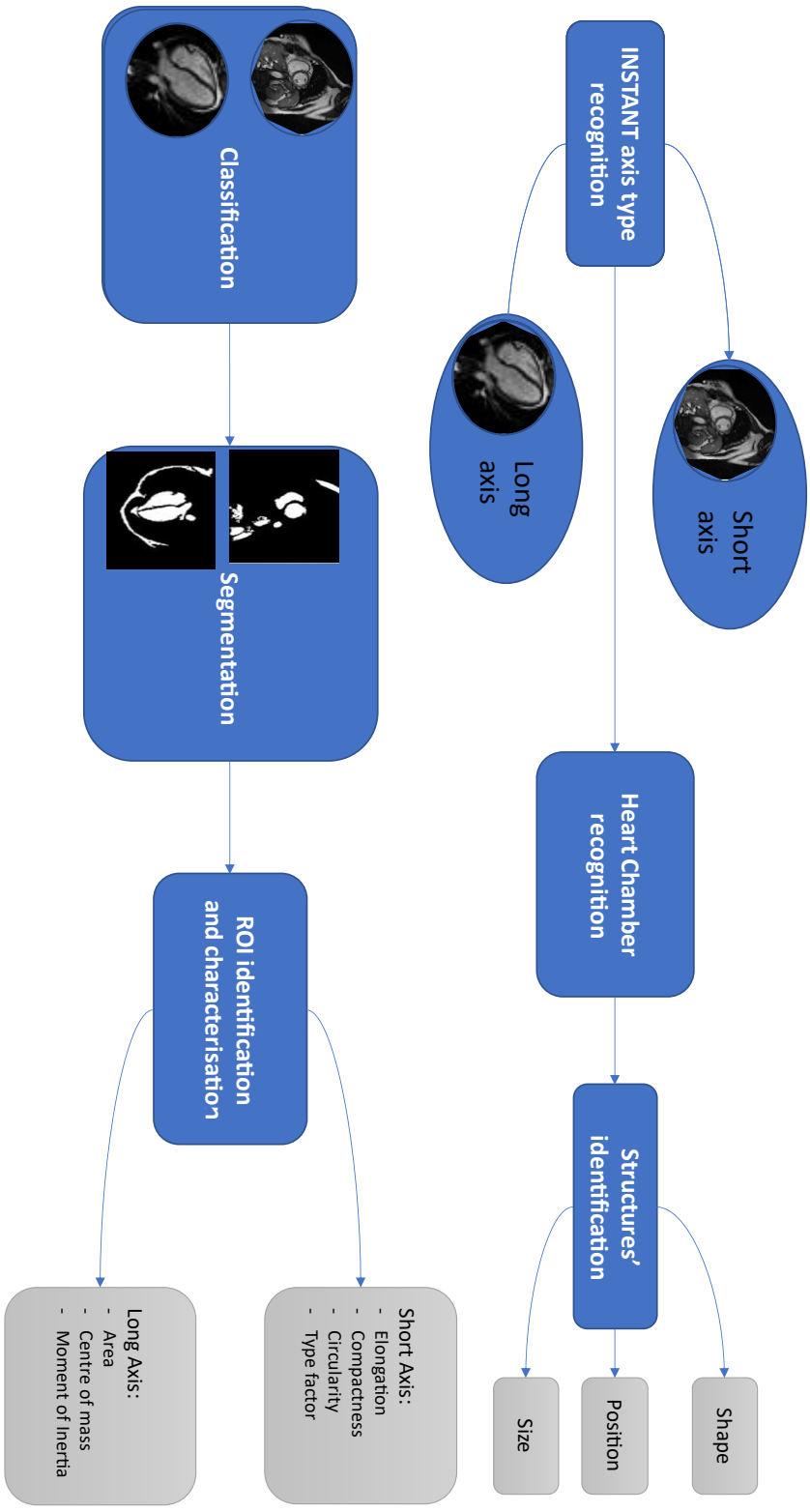

Figure 4-1 Thinking process for ventricle identification vs. computerized workflow

The remainder of this chapter is dedicated to providing a thorough explanation of each step involved in the process of segmenting cardiac structures. Starting with identifying the view of the image, both as done by an expert, and using a classification network. This step is critical as it ensures that the segmentation process is applied to the correct view of the image. The next step is the segmentation of the ventricles using texture. Finally, the distinction between the left and right ventricles is made.

4.2 View Identification

There are two distinguished views in cardiac MRI: long axis and short axis. The two are essentially different in the type of information they offer. Each view had been studied thoroughly in the literature. Telling them apart can make the rest of the segmentation/ identification process easier.

4.2.1 Long Axis vs. Short Axis

4.2.1.1 Long Axis View

The long axis is the line that runs through the center of the mitral valve and the left ventricular apex. This line provides a reference for viewing the left ventricle and its surrounding structures in a standardized manner.

There are two standard long-axis views in cardiac MRI: the horizontal (four chamber) view and the vertical view. The horizontal long-axis view is taken from a plane that passes through the long axis of the left ventricle and provides a good overall view of the left ventricle and the right ventricular free wall. It also provides clear visualization of the septal and lateral walls of the left ventricle and the size of the chambers. Additionally, the mitral and tricuspid valves can be easily seen in this view. (Refer to Figure 4-2 for reference).

The vertical long axis view, on the other hand, is taken perpendicular to the horizontal view and intersects the lower third of the mitral valve and the apex of the left ventricle. This view provides a different perspective on the structure and function of the left ventricle and is useful in evaluating certain aspects of heart function.

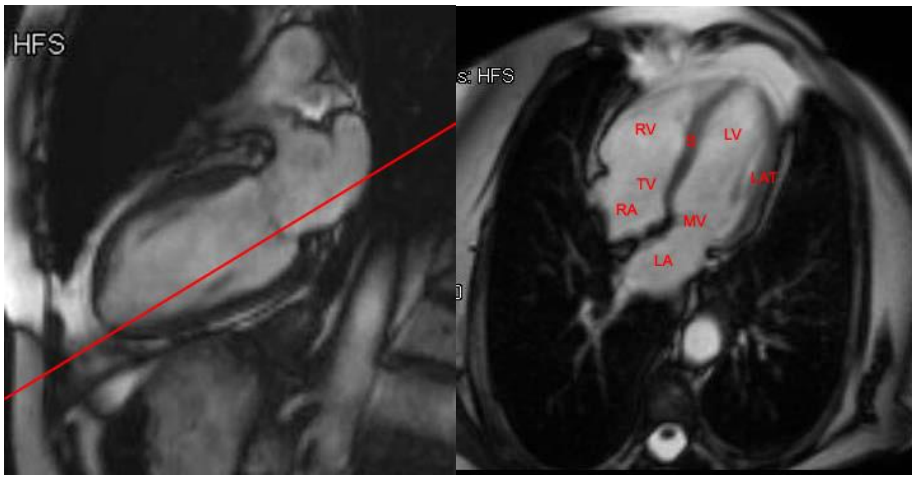

Figure 4-2 A horizontal long axis (four chamber view) of a cardiac MRI, the image on the right represents the plane obtained at the red line from the image on the left. LA = Left Atrium, LAT = Lateral wall, LV = Left Ventricle, MV = Mitral Valve, RA = Right Atrium, RV = Right Ventricle, S = Septum, TV = Tricuspid Valve

4.2.1.2 Short Axis View

The short axis planes are positioned roughly perpendicular to the long axis of the left ventricle, providing cross-sectional views of both the left and right ventricles. These images are valuable in making volumetric measurements of the ventricles, as they show the true size and shape of the heart chambers. The short axis views offer important information for clinicians to make accurate assessments of the heart's function and health. By capturing cross-sectional images of the heart, it is possible to calculate the volume of the ventricles and track changes over time. These measurements are crucial for diagnosing and monitoring conditions such as heart failure or cardiomyopathies. (Refer to Figure 4-3 for reference).

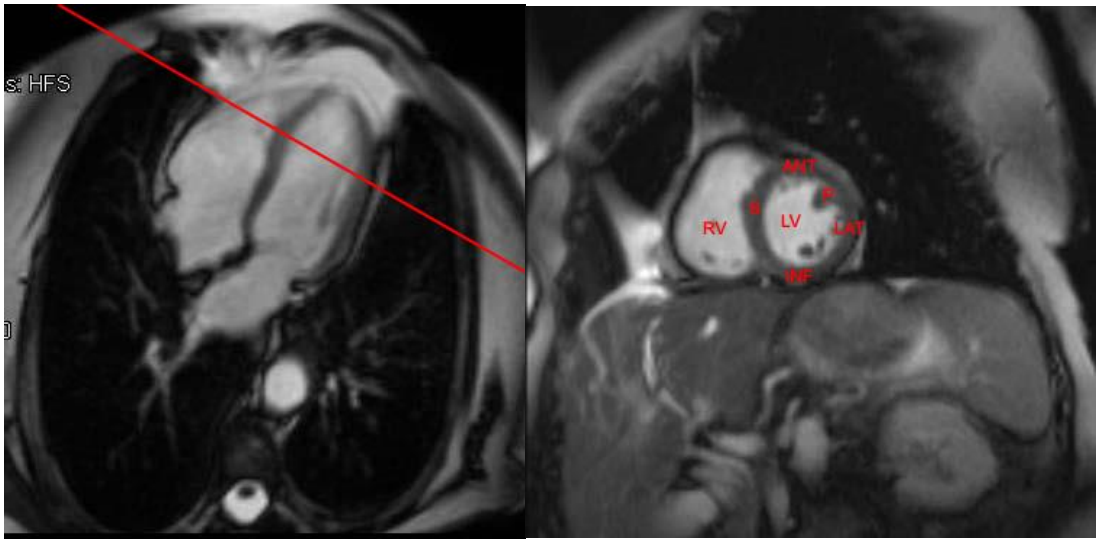

Figure 4-3 A short axis view of a cardiac MRI. The image on the right is obtained through the plane represented by the red line of the left. ANT = Anterior wall, INF = Inferior wall, LAT = Lateral wall, LV = Left Ventricle, P = Papillary muscle, RV = Right Ventricle, S = Septum

4.2.2 Automatic View Classification Deep Learning

For the task of classification, we chose to use deep learning.

Classification is the process of assigning input data to predefined categories or classes based on some criterion. In deep learning, a neural network is trained on a large dataset, learning to extract relevant features and representations of the data. During the training process, the network is presented with a set of inputs and their corresponding outputs, and adjusts its internal weights and biases to minimize the prediction error.

Once trained, the network can be used to make predictions on new, unseen data by passing it through the network and using the learned weights and biases to generate an output. This

allows the network to perform the classification task, assigning inputs to the appropriate class based on its learned representations of the data.

4.2.2.1 Dataset

The data used in this study is obtained from a publicly available dataset hosted on www.ub.edu/mnms , provided by the Multi-Centre, Multi-Vendor, and Multi-Disease Cardiac Segmentation (M&M) Workshop organized as part of the Statistical Atlases and Computational Modelling of the Heart (STACOM). The data was collected from three countries, Spain, Germany, and Canada, using four different types of MRI scanners manufactured by Siemens, General Electric, Philips, and Canon. The sample size consists of 375 subjects, including patients with hypertrophic and dilated cardiomyopathies and healthy subjects.

The training set consists of 150 annotated images generated by two vendors and 25 unannotated images from a third vendor. The testing set has 200 images, including 50 new cases from each of the previous vendors and 50 from a new vendor. The data was annotated by experienced clinicians from the respective institutions, the annotation included contours of the left ventricle (LV), the right ventricle (RV) and left ventricular myocardium (MYO) [155].

To make the dataset suitable for the task at hand, we manually separated the images into long axis and short axis views based on the original dataset's labelling. As expected, the number of short axis images is much higher (6,350 images) compared to long axis images (1,612). Due to the elongated shape of the heart, this is a common aspect of the cardiac MR. We kept the unbalanced ratio through the training set and test set to imitate the results of the Cardiac MRI.

The new dataset partition contains 1,612 long axis images and 6,350 short axis images in the training set. In the testing set, there are 323 long axis images and 1,270 short axis images.

4.2.2.2 Model Training

The deep learning model we used for the classification task is ResNet-101 which was previously described in detail in a previous chapter. ResNet has been widely recognized as a network that is relatively simple to fine-tune, making it a convenient choice for transfer learning.

To avoid overfitting, we utilized data augmentation and a dropout rate of 30%. The model was trained for 5 epochs and achieved an accuracy of 99.0%.

To validate the results and ensure the model was not simply memorizing the images, rather than properly recognizing them (a recurrent problem with small datasets), we conducted tests on unseen images. The model continued to perform well, with a validation accuracy of 98.31%. To further demonstrate the validity of the results, confusion matrices for both the test and validation sets are provided in Figure 4-4.

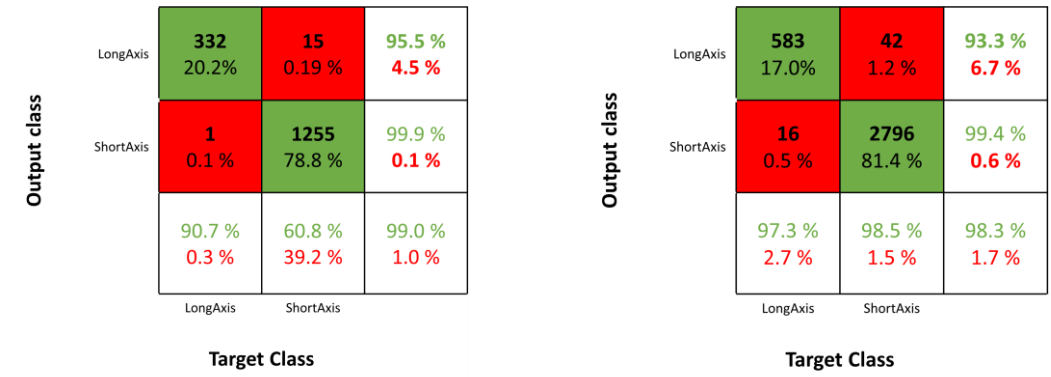

Figure 4-4 The ResNet-101 model's confusion matrices on both the test set (on the right), and validation set (on the left)

This step is rather interpretable, thanks to a vast body of research in the literature that supports the idea that classification tasks are never random. A neural network operates by uncovering latent features within the images, which become increasingly complex with each iteration. These features range from edges and shapes to textures and more, as depicted in Figure 4-5. This process demonstrates how a neural network progresses from

recognizing simple features to more complex ones, enabling it to accurately classify images.

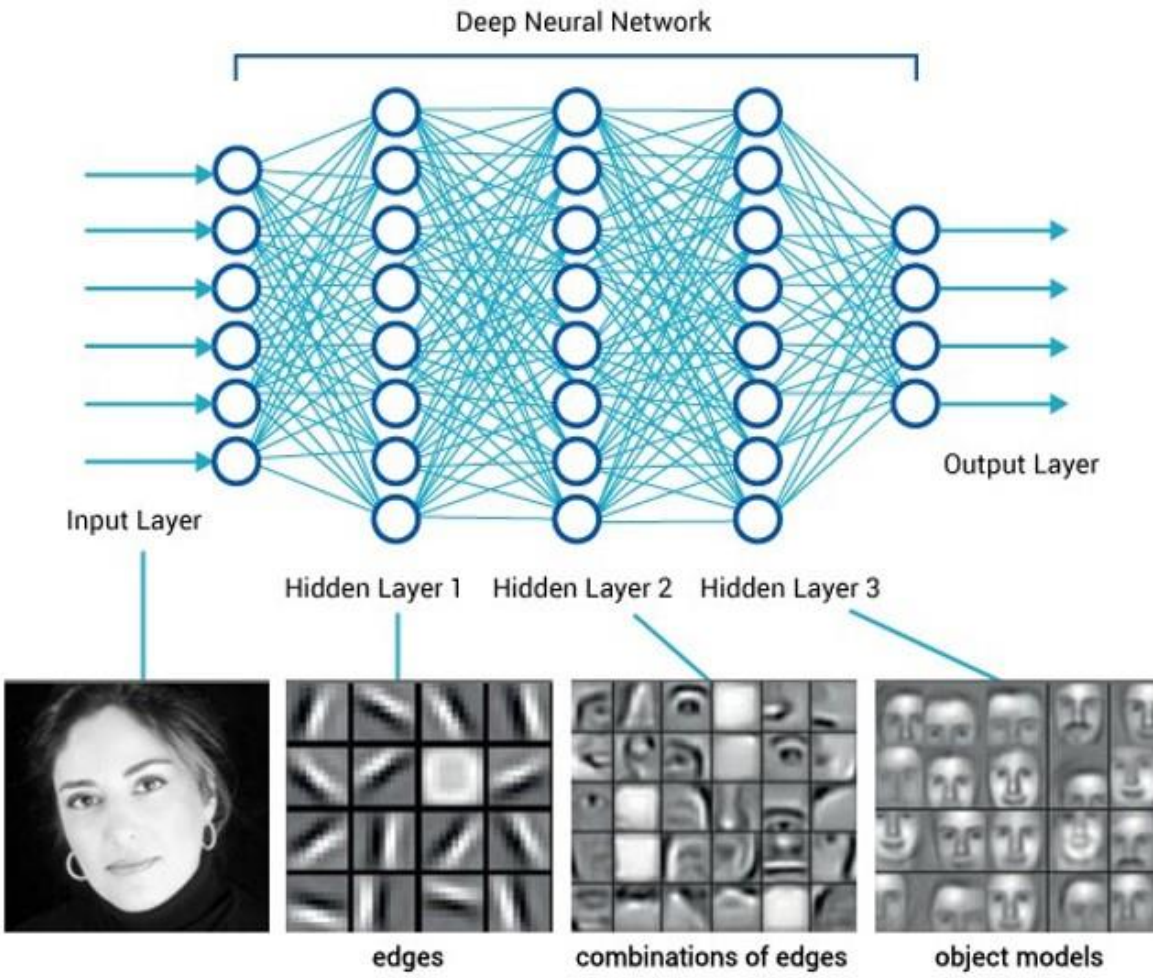

Figure 4-5 The output of different layers during a classification task

4.3 Ventricle Segmentation

The ventricles are easily recognizable to a trained expert due to their distinct texture compared to other cardiac structures. To replicate this, we utilized an algorithm called Particle Swarm Optimization (PSO).

Particle Swarm Optimization is a computational method inspired by the social behavior of birds and fish. Developed in 1995 by Dr. Eberhart and Dr. Kennedy [156], PSO uses a population-based approach to optimize a problem.

The algorithm adheres to three key principles: cohesion, alignment, and separation. Cohesion refers to the tendency of particles to be attracted to the center of the group, alignment refers to particles following the same path as their neighbors, and separation refers to the avoidance of collisions through self-separation with a safety distance.

The PSO algorithm is based on the principle that all particles have a specific objective to reach, which is determined by an "objective function" that should be optimized or provided by the user, depending on the application. This technique is commonly used to solve multilevel thresholding segmentation problems.

The reason we resolved to a relatively old technique is because it is gaining back popularity lately. Even compared to other techniques, PSO-based image segmentation can create segments with great details according to a recent study [157].

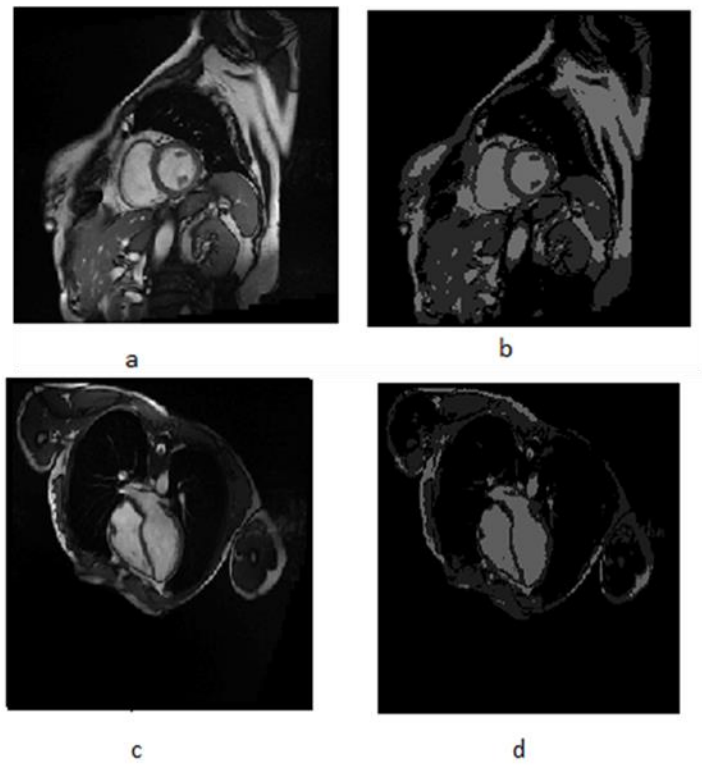

Figure 4-6 (a, c) Original images. (b, d) Images after segmentation with PSO [3]

The algorithm segments the short axis view into three classes, as depicted in Figure 4-6 (b). The first class is made up of low intensity pixels and includes the lung and the background of the image. The second class includes pixels of average intensity, such as the myocardium, and the third class includes pixels of high intensity, such as the left and right ventricles.

When it comes to the long axis view shown in Figure 4-6(d) , the first class, made up of low intensity pixels, includes the two lungs and the background. The second class includes pixels of average intensity, such as the myocardium, and the third class includes pixels of high intensity, such as the left atrium, ventricle, right atrium, and descending aorta.

In order to locate the left ventricle and right ventricle, the process begins by identifying the lung from the first class (Figure 4-7 E). By using the detected lung's position, it becomes possible to subsequently locate the right and left ventricles from the third class (Figure 4-7-F).

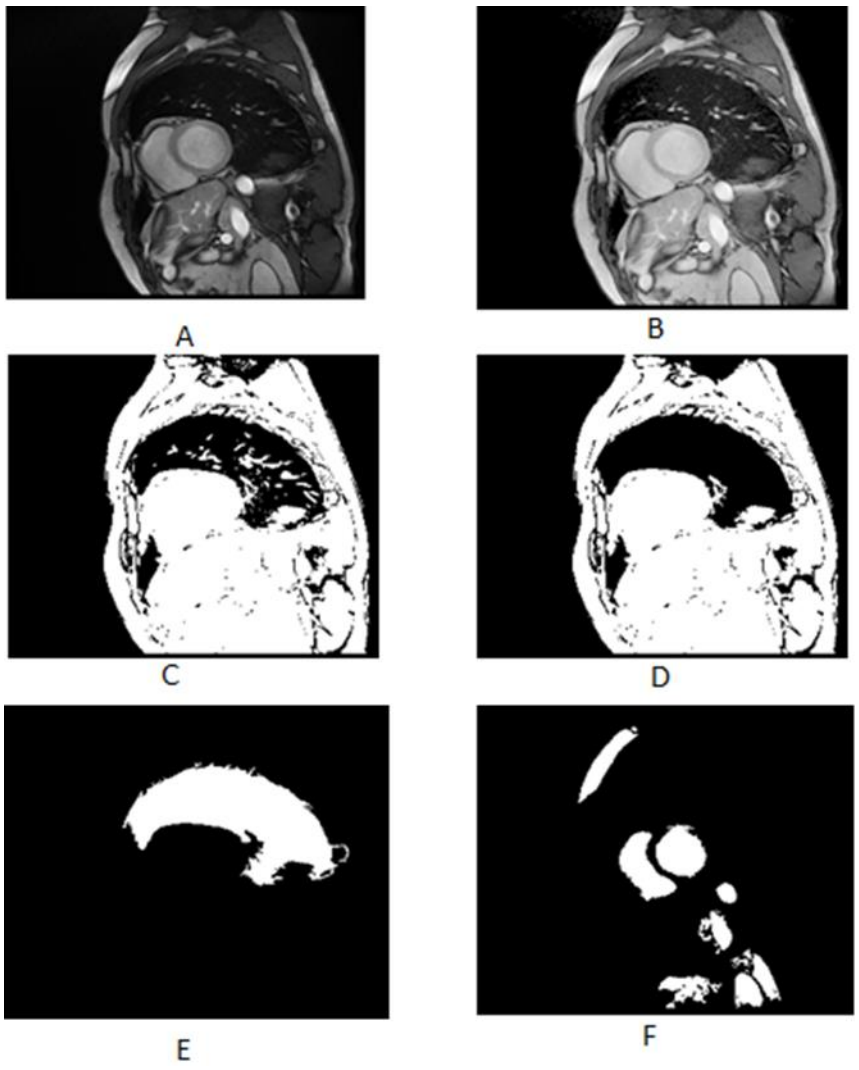

Figure 4-7 Segmentation outcome after different number of iterations. [3]

As illustrated in Figure 4-8, the segmentation of the long axis view results in three areas of focus: the aorta, the left heart (consisting of the left atrium and ventricle), and the right heart (consisting of the right atrium and ventricle).

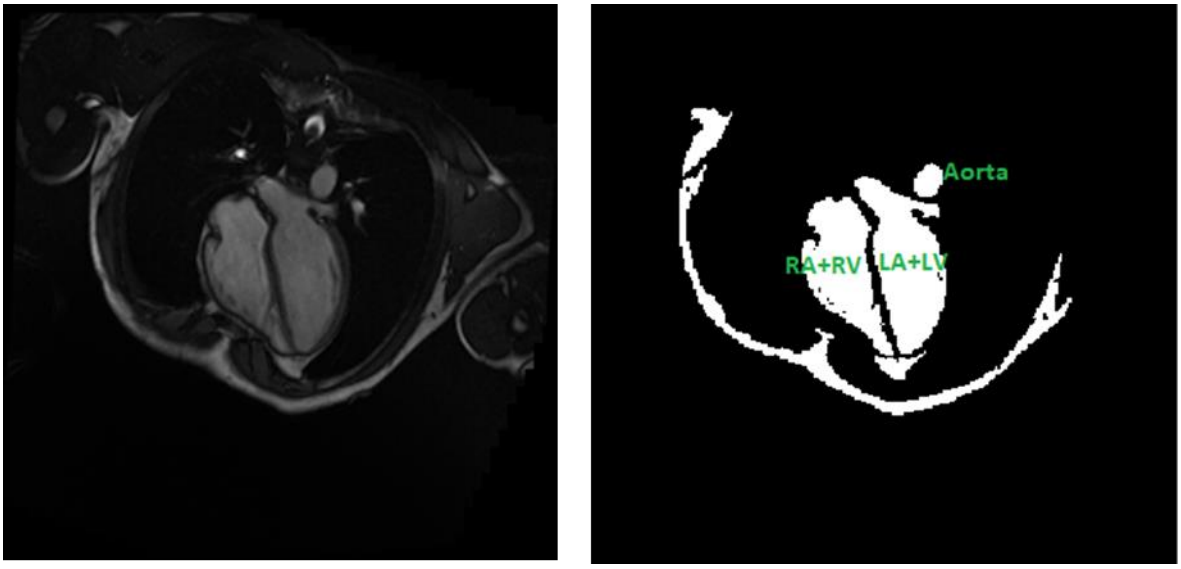

Figure 4-8 The segmentation result on a long axis image

So, to sum up, the Particle Swarm Optimization divides the short and long axis views into three classes based on the intensity of the pixels. The first class includes low intensity pixels such as the lung and background, the second class includes pixels of average intensity such as the myocardium, and the third class includes pixels of high intensity such as the ventricles. By identifying the lung in the first class, the right and left ventricles can be subsequently located in the third class. Additionally, the long axis view has three regions of interest: the aorta, the left heart (consisting of the left atrium and ventricle), and the right heart (consisting of the right atrium and ventricle).

4.4 Cardiac Structure Identification

“Structure identification” refers to the process of recognizing and labeling the various structures within an image of the heart. This is a crucial step in medical imaging analysis as it allows for a more accurate assessment of the heart.

Doctors use various criteria to accurately identify cardiac structures, including but not limited to their shape, size, and position.

To improve the accuracy of structure identification, it is best to use an automated model. This requires a more refined selection of criteria, which can be achieved by using a statistical analysis model to determine the most valuable and relevant features for characterizing and recognizing the structures.

In this study we used Analysis Of Variance, a statistical analysis model, to carefully pick the most informative features from a pool of potential features that we proposed.

4.4.1 Analysis Of Variance (ANOVA)

Analysis of Variances , or ANOVA, is a statistical method that was created by Ronald Fisher in 1918 [158]. Its purpose is to examine the differences in means among two or more groups, by assessing the variation in the data and determining its source (hence its name). Specifically, ANOVA compares the variation between groups with the variation within groups.

4.4.1.1 Assumptions of ANOVA

Like any other parametric test, the ANOVA test is built around a number of assumptions:

- All groups are independent, meaning there is no subject present in a group that can be in a different group.
- All groups/levels have equal sample sizes.
- The dependent variable must be normally distributed for ANOVA to be conducted, meaning that the middle scores are the most frequent, and extreme scores are the least frequent.
- The population variances must be equal, meaning that the deviation of scores is similar across all populations.

4.4.1.2 ANOVA F-value

The ANOVA test statistic is represented by the letter F. It is calculated by dividing the variance caused by the treatment by the variance due to random chance. The resulting F-value can indicate whether there is a significant difference between the

levels of the independent variable. A p-value less than 0.05 and a higher F-value suggest that the treatment variables are significant.

Another way to express the formula for F is:

$$\boldsymbol{F} = \frac{MSE}{MST} \qquad (4\text{-}1)$$

where:

F is the ANOVA coefficient,

MST is the Mean Sum of Squares due to Treatment,

MSE is Mean Sum of Squares due to Error.

4.4.1.3 Running an ANOVA Test

When running an ANOVA test, the first step is to assume that all groups have the same variance (null hypothesis). Next, the Sum of Squares (MST) is calculated. The ANOVA effect size is then determined, which is used to calculate the degree of freedom (d.f). The F-value is then computed using equation above, which is used to determine if the group of variables are jointly significant and can be used to support or reject the null hypothesis. Finally, the decision is made to either accept or reject the null hypothesis.

4.4.2 Results And Discussion

After segmentation, a long-axis view image includes the right heart (made up of the right ventricle and right atrium), the left heart (made up of the left atrium and ventricle), and the aorta. In order to distinguish between these groups, eight different features were selected and tested for significance.

The ANOVA test was used to identify the most significant features that distinguish the different groups within the long-axis view image. Table 4-1 and Figure 4-9 provide a detailed breakdown of the results of this analysis, including the sum square, degree of freedom, mean square, F-value, and the final verdict for each feature. These values are

important for understanding the statistical significance of the results and how they support or reject the null hypothesis.

The F-value represents the ratio of the variance between groups to the variance within groups. A higher F-value indicates that there is a larger difference between the groups, which supports the alternative hypothesis that the groups are not equal. The F-value is then compared to the F critical value, which is determined based on the degree of freedom and the level of significance (usually set at 0.05). A feature is considered significant when the F-value is greater than the F critical value.

Additionally, when the F-value is high and the critical value is not met, it means that the feature is a good criterion to distinguish between the two structures, as there is no overlap between them. This makes it easier to differentiate and segment the different groups in the image.

Table 4-1 Feature extraction for long axis segmentations

	source	**Sum square**	**Degree of freedom**	**Mean square**	**F-value**	**Prb> f**	**Verdict**
Centre of Mass X	g	8221	3	2470.34	1.04	0.3857	Feature not significant
	error	10549	40	2637.3			
	Total	113713	43				
Centre of Mass Y	g	27859.5	3	9286.5	8.6	0.0002	The aorta is significantly different from every other group
	error	43202.4	40	1080.06			
	Total	71061.9	43				
Area	g	5.4749 1007	3	1.82497 1007	6.05	0.0017	Left heart and right heart are different than other groups
	Error	1.20749 1008	40	3.01872 1006			
	Total	1.75498 1008	43				

Elongation Factor	g	15.7963	3	5.26542	5.74	0.0023	Feature not significant
	Error	36.7085	40	0.91771			
	Total	52.5048	43				
Heywood Circularity Factor	g	3.519	3	1.17301	4.29	0.0102	Feature not significant
	Error	10.9277	40	0.27319			
	Total	14.4468	43				
Moment of Inertia XX	g	6.97888 1012	3	2.32629 1012	1.59	0.2061	Feature not significant
	Error	5.84062 1013	40	1.46015 1012			
	Total	6.5351 1013	43				
Moment of Inertia XY	g	6.7484 1012	3	2.24947 1012	3.45	0.0253	Can be used to identify left heart
	Error	2.60511 1013	40	6.51278 1011			
	Total	3.27995 1013	43				
Moment of Inertia YY	g	6.7484 1012	3	2.24947 1012	3.45	0.0253	Can be used to identify left heart
	Error	2.60511 1013	40	6.51278 1011			
	Total	3.27995 1013	43				

Not all of the features studied above were found to have significant variation and were thus not suitable for classification purposes. One example of this is the aorta, which was found to have a significantly lower center of mass Y with an F-value of 0.0002 compared to the other objects on the segmented image. This means that the center of mass Y for the aorta was lower than the other objects, and this difference was statistically significant.

On the other hand, the left heart was found to have a higher moment of inertia XY and moment of inertia YY compared to the other subjects. This means that the left heart has a higher rotational inertia in the XY plane and YY plane, respectively, when compared to

the other subjects. These features can be used to distinguish the left heart from the other objects.

Position can also be used to differentiate between the left heart and the right heart, but it was not used in this study due to the inconsistent nature of the dataset. This means that the position of the left heart and the right heart in the image may vary, making it less reliable as a feature to distinguish between the two.

Table 4-2 and Figure 4-10 contain features that were tested for short-axis characterization. All four features that were assessed for this task were found to be useful in characterizing and distinguishing between the right and left ventricles on short-axis images. Specifically, the right ventricle was found to have high elongation and Heywood circularity, which are measures of how elongated and circular the right ventricle is respectively. On the other hand, the left ventricle can be characterized by high compactness and type factor, which are measures of how compact and a shape of the left ventricle respectively.

Table 4-2 Feature extraction for short axis segmentation

	Source	**Sum square**	**Degree of freedom**	**Mean sq.**	**F-value**	**Prb> f**	**Verdict**
Elongation	g	7.7261	1	7.72607	66.58	7.03777 10-10	Significant feature
	error	4.4094	38	0.11604			
	Total	12.1355	39				
Compactness	g	0.20517	1	0.20517	64.61	1.01530 10-09	Significant feature
	error	0.12067	38	0.00318			
	Total	0.32584	39				
Heywood Circularity	g	0.5774	1	0.5774	42.74	1.04602 10-07	Significant feature
	Error	0.51339	38	0.01351			
	Total	1.0908	39				
Type Factor	g	0.42483	1	0.42483	155.14	5.47648 10-15	Significant feature
	Error	0.10406	38	0.00274			
	Total	0.52889	39				

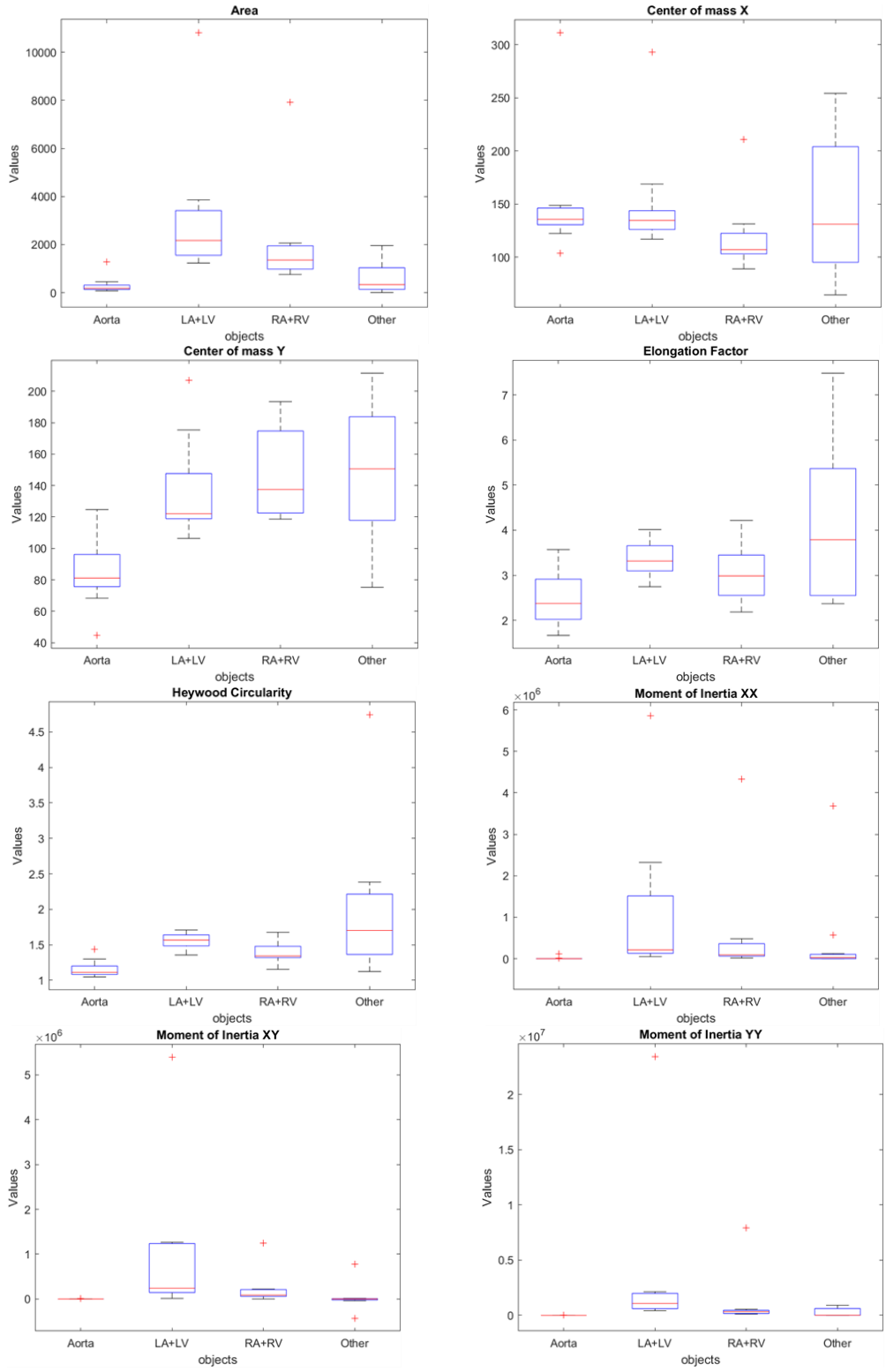

Figure 4-9 Box plot representations of the ANOVA for long axis features

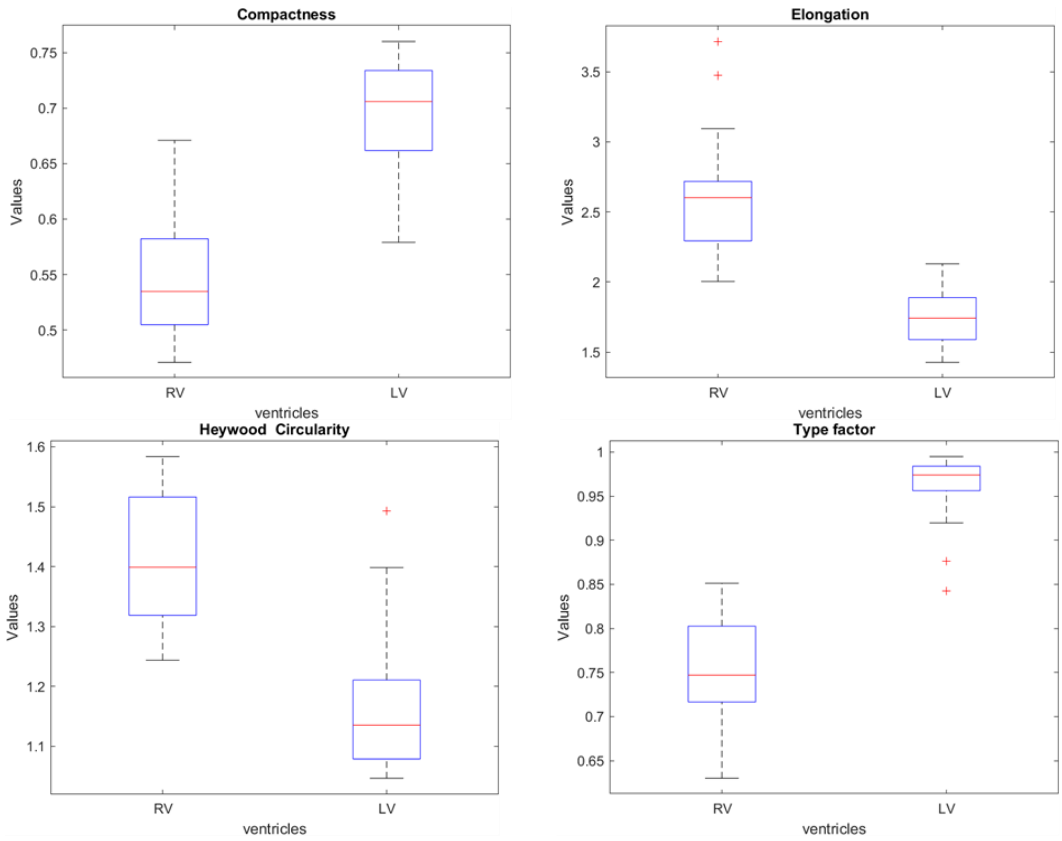

Figure 4-10 Box plot representations of the ANOVA for short axis features

4.5 Conclusion

Trustworthy AI is the new urging problem. After models started achieving human-like and even better-than-human results, experts started wondering to what extent they could rely on fully automated algorithms for critical tasks. Nowadays, providing a visual explanation of the process behind the decision-making is essential towards making the final result more acceptable to the clinicians.

In this work we propose an AI model that mimics doctor reasoning, which would hopefully have the potential of AI models, but is also easy to follow and transparent throughout.

Our main contribution is separating the initial images into long-axis and short-axis views. This step not only follows how a human expert would proceed, but it also allows for a separate analysis of the two image types going forward, which allows for a more precise study.

The next step would be segmenting the ventricles. For that we used Particle Swarm Optimization for general segmentation, a method that proved particularly handy with multi class images. After segmentation, we ended up with various structures on the images, including heart chambers, the lungs, the aorta, and other objects.

For accurate identification of the different objects, we applied a characterization based on shape descriptors and ANOVA for features selection to detect the location of each region of interest: lung, left and right ventricle in the short-axis view, the aorta, the left heart (left atrium and ventricle), and the right heart (right atrium and ventricle) in the long axis view.

The purpose of such characterization is to provide a tool which eventually allows advanced segmentation locating objects of interest after a global segmentation using classification methods: FCM, KFCM, EM and operator's manual and automatic thresholding.

General Conclusion

Starting this PhD program, the main objective was to use Deep Learning for cardiac applications, a broad research subject.

While getting familiarized with both cardiac imaging and Artificial Intelligence, many research problems emerged, but we came to the conclusion that there exist two main challenges in this field: the shortage of labeled data, particularly for the heart, and the lack of trustworthy Artificial Intelligence (AI) models.

To tackle the data shortage problem, we focused on automatic screening for atherosclerosis from Coronary CT Angiography, a serious cardiovascular disease. The traditional method of screening requires segmentation, which is time-consuming. However we explored the use of transfer learning to screen the disease straight from CCTA images, resulting in a faster outcome. We have achieved an encouraging performance with scores 95.21% of accuracy, 99.25% of specificity, 60.8% sensitivity, 90.48% positive predictive value. The poor sensitivity could be due to the lack of positive images in the testing dataset.

This is a problem commonly faced in all medical applications as a result of intrinsic data imbalance. In order to balance the dataset, we generated new positive images using Generative Adversarial Network (GAN). The generated images were added to the training set, while the test set remained composed of only real images, to insure accurate validation.

The model used on the newly balanced dataset improved sensitivity from 60.8% to 89.0%, which suggests that : (1) data imbalance was the reason behind the poor performance of the model, and (2) synthesized data can be used for data augmentation to solve the problem of data unavailability.

In the future, we would like to carry the work by studying the impact of using only generated data for the model training, how it would impact the overall performance of the model. We would like to study how dependent Deep learning models on data distributions are.

The second part of the thesis focused on heart structure segmentation using an AI model that mimics doctor reasoning; creating a mind map that describes every step the doctor use when manually segmenting and identifying the heart structures, then replacing each step with a corresponding machine learning algorithm. The goal was to achieve superior performance while maintaining a transparent algorithm.

The resulting model started with separating the long axis view images from the short axis view images, using a ResNet-101; then we performed multiclass segmentation using Particle Swarm Optimization, which resulted in segmented images with the different heart structures. For the final step, the structure identification was done after a number of features, judged useful based on an ANOVA statistical study.

The model we created describes a clear interpretation of how a human mind solves such a complicated task as cardiac segmentation. Our proposal to mimic the doctor's approach and reasoning makes our AI trustworthy, although not Explainable AI.

This model needs to be finalized, tested, and implemented into a fully functioning software for clinical use, to assess how well-received it would be amidst doctors.

5 Reference List

1. World Health Organization. [Online].; 2021 [cited 2022 May 23. Available from: HYPERLINK "https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)" https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) .

2. Laidi A, Ammar M, Daho MEH, Mahmoudi S. GAN Data Augmentation for Improved Automated Atherosclerosis Screening from Coronary CT Angiography. EAI Endorsed Transactions on Scalable Information Systems. 2022; 10(1). DOI: 10.4108/eai.17-5-2022.173981

3. Laidi A, Ammar M, Daho MEH, Mahmoudi S. Toward an automatic detection of cardiac structures in short and long axis views. Biomedical Signal Processing and Control. 2023; 79(2): p. 104187. DOI: 10.1016/j.bspc.2022.104187

4. Weinhaus AJ, Roberts KP. Anatomy of the human heart. In Handbook of cardiac anatomy, physiology, and devices: Springer; 2005. p. 51-79.

5. BETTS JG, DESAIX P, JOHNSON E, JOHNSON JE, KOROL O, KRUSE D, et al. Anatomy & Physiology: Rice University; 2013.

6. Roger VL. Epidemiology of myocardial infarction. Medical Clinics of North America. 2007; 91(4): p. 537-552. DOI: 10.1016/j.mcna.2007.03.007

7. Falk E. Pathogenesis of Atherosclerosis. Journal of the American College of Cardiology. 2006; 47(8 Suppl): p. C7–12. DOI: 10.1016/j.jacc.2005.09.068

8. Portegies MLP, Koudstaal PJ, Ikram MA. Cerebrovascular disease. In Michael J. Aminoff FBDFS, editor. Handbook of Clinical Neurology.: Elsevier; 2016. p. 239-261. DOI: 10.1016/B978-0-12-802973-2.00014-8

9. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J, editors. Harrison's principles of internal medicine. 19th ed.: Mcgraw-hill New York, NY, USA; 2015.

10. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J, editors. Harrison's Principles of Internal Medicine. 20th ed.: McGraw Hill Education; 2018.

11. Carli MFD, Kwong RY, Solomon SD. Noninvasive Cardiac Imaging: Echocardiography, Nuclear Cardiology, and Magnetic Resonance/Computed Tomography Imaging. In Harrison's Principles of Internal Medicine, 20e. New York, NY: McGraw-Hill Education; 2018.

12. AL-SHAMSI M. Addressing the physicians' shortage in developing countries by accelerating and reforming the medical education: Is it possible? Journal of Advances in Medical Education & Professionalism. 2017; 5(4): p. 210–219.

13. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 1943; 5(4): p. 115-133. DOI: https://doi.org/10.1007/BF02478259 .

14. Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv: arXiv:1609.04747.

15. Curry HB. The method of steepest descent for non-linear minimization problems. Quarterly of Applied Mathematics. 1944; 2(3): p. 258-261.

16. Robbins H, Monro S. A stochastic approximation method. The annals of mathematical statistics. 1951; p. 400-407. DOI: https://doi.org/10.1214/aoms/1177729586

17. Goyal P, Dollàr P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, et al. Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv preprint arXiv:1706.02677. 2017. DOI: https://doi.org/10.48550/arXiv.1706.02677

18. Qian N. On the momentum term in gradient descent learning algorithms. Neural networks. 1999; 12(1): p. 145-151. DOI: https://doi.org/10.1016/S0893-6080(98)00116-6

19. Hinton G, Srivastava N, Swersky K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent.: Coursera; 2012.

20. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014. DOI: https://doi.org/10.48550/arXiv.1412.6980

21. LeCun Y, Bengio Yao. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. 1998; 3361(10): p. 255–258.

22. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in neural information processing systems. 2012; 25. URL: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

23. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR. 2015. DOI : https://doi.org/10.48550/arXiv.1409.1556

24. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. In IEEE Conference on Computer Vision and Pattern Recognition; 2015; Boston, MA. p. 1-9. DOI: 10.1109/CVPR.2015.7298594

25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770-778. DOI: https://doi.org/10.48550/arXiv.1512.03385

26. Shelhamer E, Long J Darrell T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431-3440. DOI: 10.1109/TPAMI.2016.2572683

27. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention; 2015: Springer. p. 234-241. URL: https://doi.org/10.48550/arXiv.1505.04597

28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In Advances in neural information processing systems; 2017. DOI: https://doi.org/10.48550/arXiv.1706.03762

29. Wu B, Xu C, Dai X, Wan A, Zhang P, Yan Z, et al. Visual transformers: Token-based image representation and processing for computer vision. arXiv preprint arXiv:2006.03677. 2020. DOI: https://doi.org/10.48550/arXiv.2006.03677

30. Peng P, Lekadir K, Gooya A, Shao L, Petersen SE, Frangi AF. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine. 2016; 29(2): p. 155-195. DOI: 10.1007/s10334-015-0521-4

31. Petitjean C, Dacher JN. A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis. 2011; 15(2): p. 169-84. DOI: 10.1016/j.media.2010.12.004

32. Petitjean C, Zuluaga MA, Bai W, Dacher JN, Grosgeorge D, Caudron J, et al. Right ventricle segmentation from cardiac MRI: A collation study. Medical image analysis. 2015; 19(1): p. 187-202. DOI**:** 10.1016/j.media.2014.10.004

33. Ringenberg J, Deo M, Devabhaktuni V, Berenfeld O, Boyers P, Gold J. Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI. Computerized Medical Imaging and Graphics. 2014; 38(3): p. 190-201. DOI: 10.1016/j.compmedimag.2013.12.011

34. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Transactions on medical imaging. 2006; 25(8): p. 987-1010. DOI: 10.1109/TMI.2006.877092

35. Mazaheri S, Sulaiman PSB, Wirza R, Khalid F, Kadiman S, Dimon MZ, et al. Echocardiography Image Segmentation: A Survey. In Advanced Computer Science Applications and Technologies; 2013: IEEE. p. 327-332. DOI: https://doi.org/10.1109/ACSAT.2013.71

36. Axel L, Montillo A, Kim D. Tagged magnetic resonance imaging of the heart: a survey. Medical image analysis. 2005; 9(4): p. 376-393. DOI: https://doi.org/10.1016/j.media.2005.01.003

37. Frangi AF, Niessen WJ, Viergever MA. Three-dimensional modeling for functional analysis of cardiac images, a review. IEEE transactions on medical imaging. 2001; 20(1): p. 2-25. DOI: 10.1109/42.906421

38. Tavakoli V, Amini AA. A survey of shaped-based registration and segmentation techniques for cardiac images. Computer Vision and Image Understanding. 2013; 117(9): p. 966-989. DOI: https://doi.org/10.1016/j.cviu.2012.11.017

39. Kang DJ, Kuo CCJ, Slomka PJ, Dey D, Germano G. Heart chambers and whole heart segmentation techniques: review. In ournal of Electronic Imaging. 1st ed.: SPIE; 2012. p. 010901. DOI: https://doi.org/10.1117/1.JEI.21.1.010901

40. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep Learning for Cardiac Image Segmentation: A Review. Frontiers in Cardiovascular Medicine. 2020; 7: p. 25. DOI: 10.3389/fcvm.2020.00025

41. Bizopoulos P, Koutsouris D. Deep Learning in Cardiology. IEEE Reviews in Biomedical Engineering. 2018; 12: p. 168-193. [PDF] Deep Learning in Cardiology | Semantic Scholar

42. Jamart K, Xiong Z, Maso Talou GD, Stiles MK, Zhao J. Mini Review: Deep Learning for Atrial Segmentation From Late Gadolinium-Enhanced MRIs. Frontiers in Cardiovascular Medicine. 2020; 7: p. 86. DOI: 10.3389/fcvm.2020.00086

43. Huang S, Liu J, Lee LC, Venkatesh SK, Teo LLS, Au C, et al. An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images. Journal of digital imaging. 2011; 24(4): p. 598-608. DOI: https://doi.org/10.1007/s10278-010-9315-4

44. Lu YL, Connelly KA, Dick AJ, Wright GA, Radau PE. Automatic functional analysis of left ventricle in cardiac cine MRI. Quantitative imaging in medicine and surgery. 2013; 3(4): p. 200. DOI: 10.3978/j.issn.2223-4292.2013.08.02

45. Jolly MP. Automatic Segmentation of the Left Ventricle in Cardiac MR and CT Images. International Journal of Computer Vision. 2006; 70(2): p. 151-163. DOI: https://doi.org/10.1007/s11263-006-7936-3

46. Zhukov L, Bao Z, Gusikov I, Wood J, Breen DE. Dynamic deformable models for 3D MRI heart segmentation. In Medical Imaging 2002: Image Processing; 2002. p. 1398-1405. DOI: https://doi.org/10.1117/12.467105

47. Montagnat J, Delingette H. 4D deformable models with temporal constraints: application to 4D cardiac image segmentation. Medical image analysis. 2005; 9(1): p. 87-100. DOI: https://doi.org/10.1016/j.media.2004.06.025

48. Nasr-Esfahani M, Mohrekesh M, Akbari M, Soroushmehr SR, Nasr-Esfahani E, Karimi N, et al. Left ventricle segmentation in cardiac mri images using fully convolutional neural networks. In 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018. p. 1275-1278. DOI: 10.1109/EMBC.2018.8512536

49. Radau P, Lu Y, Connelly K, Paul G, Dick A, Wright G. Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge. 2009; 49. DOI: https://doi.org/10.54294/g80ruo

50. Emad O, Yassine IA, Fahmy AS. Automatic localization of the left ventricle in cardiac MRI images using deep learning. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015. p. 683-686. URL: https://ieeexplore.ieee.org/document/7318454

51. Molaei S, Shiri ME, Horan K, Kahrobaei D, Nallamothu B, Najarian K. Deep convolutional neural networks for left ventricle segmentation. In 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017. p. 668-671. URL: https://ieeexplore.ieee.org/document/8036913

52. Ngo TA, Carneiro G. Fully automated segmentation using distance regularised level set and deep-structured learning and inference. Deep Learning and Convolutional Neural Networks for Medical Image Computing. 2017: p. 197-224. DOI:10.1007/978-3-319-42999-1_12

53. Luo G, Dong S, Wang K, Zhang H. Cardiac left ventricular volumes prediction method based on atlas location and deep learning. In International Conference on Bioinformatics and Biomedicine (BIBM); 2016: IEEE. p. 1604-1610. DOI: 10.1109/bibm.2016.7822759

54. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Medical image analysis. 2016; 30: p. 108-119. DOI: https://doi.org/10.1016/j.media.2016.01.005

55. Yang H, Sun J, Li H, Wang L, Xu Z. Deep Fusion Net for Multi-atlas Segmentation: Application to Cardiac MR Images. In International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016: Springer. p. 521-528. DOI: https://doi.org/10.1007/978-3-319-46723-8_60

56. Yang X, Bian C, Yu L, Ni D, Heng PA. Class-Balanced Deep Neural Network for Automatic Ventricular Structure Segmentation. In International workshop on statistical atlases and computational models of the heart; 2017: Springer. p. 152-160. DOI: http://dx.doi.org/10.1007/978-3-319-75541-0_16

57. Rohé MM, Sermesant M, Pennec X. Automatic multi-atlas segmentation of myocardium with svf-net. In International workshop on statistical atlases and computational models of the heart; 2017: Springer. p. 170-177. URL: https://hal.inria.fr/hal-01575297/document

58. Rohé MM, Datar M, Heimann T, Sermesant M, Pennec X. SVF-Net: Learning Deformable Image Registration Using Shape Matching. In International conference on medical image computing and computer-assisted intervention; 2017: Springer. p. 266-274. DOI: https://doi.org/10.1007/978-3-319-66182-7_31

59. Tan LK, Liew YM, Lim E, McLaughlin RA. Cardiac left ventricle segmentation using convolutional neural network regression. In EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2016: IEEE. p. 490-493. DOI:10.1109/IECBES.2016.7843499

60. Fonseca CG, Backhaus M, Bluemke DA, Britten RD, Chung JD, Cowan BR, et al. The Cardiac Atlas Project—an imaging database for computational modeling and

statistical atlases of the heart. Bioinformatics. 2011; 27(16): p. 2288-2295. DOI: 10.1093/bioinformatics/btr360

61. Liao F, Chen X, Hu X, Song S. Estimation of the Volume of the Left Ventricle From MRI Images Using Deep Neural Networks. IEEE transactions on cybernetics. 2017; 49(2): p. 495-504. URL: https://ieeexplore.ieee.org/document/8232462

62. Wolterink JM, Leiner T, Viergever MA, I. Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2018: Springer. p. 101-110. DOI: 10.1016/j.compmedimag.2021.101864

63. He Y, Qin W, Wu Y, Zhang M, Yang Y, Liu X, et al. Automatic left ventricle segmentation from cardiac magnetic resonance images using a capsule network. Journal of X-ray Science and Technology. 2020; 28(3): p. 541-553. DOI: 10.3233/XST-190621

64. Zhao W, Ye J, Yang M, Lei Z, Zhang S, Zhao Z. Investigating capsule networks with dynamic routing for text classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; 2018; Brussels, Belgium: Association for Computational Linguistics. URL : https://aclanthology.org/D18-1350/

65. Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Advances in neural information processing systems. 2017; 30. URL: https://proceedings.neurips.cc/paper/2017/file/2cad8fa47bbef282badbb8de5374b894-Paper.pdf

66. Tran PV. A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. ArXiv. 2016; abs/1604.00494. DOI: https://doi.org/10.48550/arXiv.1604.00494

67. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis. 2019; 51: p. 21-45. DOI: https://doi.org/10.1016/j.media.2018.10.004

68. Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, et al. Semi-supervised learning for network-based cardiac mr image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention; 2017: Springer. p. 253-260. DOI: https://doi.org/10.1007/978-3-319-66185-8_29

69. Avendi MR, Kheradvar A, Jafarkhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. Magnetic resonance in medicine. 2017; 78(6): p. 2439-2448. DOI: 10.1002/mrm.26631

70. Giannakidis A, Kamnitsas K, Spadotto V, Keegan J, Smith G, Glocker B, et al. Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3d convolutional neural network. In 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS); 2016: IEEE. p. 42-46. URL: https://ieeexplore.ieee.org/document/7907443

71. Zheng Q, Delingette H, Duchateau N, Ayache N. 3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation. IEEE transactions on medical imaging. 2018; 37(9): p. 2137-2148. DOI:10.1109/TMI.2018.2820742

72. Poudel RP, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. International Workshop on Reconstruction and Analysis of Moving Body Organs. 2016: p. 83-94. DOI:10.1007/978-3-319-52280-7_8

73. Patravali J, Jain S, Chilamkurthy S. 2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2017: Springer. p. 130-139. DOI:10.1007/978-3-319-75541-0_14

74. Du X, Yin S, Tang R, Zhang Y, Li S. Cardiac-DeepIED: Automatic Pixel-Level Deep Segmentation for Cardiac Bi-Ventricle Using Improved End-to-End Encoder-Decoder Network. IEEE journal of translational engineering in health and medicine. 2019; 7: p. 1-10. DOI: 10.1109/JTEHM.2019.2900628

75. Kong B, Zhan Y, Shin M, Denny T, Zhang S. Recognizing End-Diastole and End-Systole Frames via Deep Temporal Regression Network. In International conference on medical image computing and computer-assisted intervention; 2016: Springer. p. 264-272. DOI: https://doi.org/10.1007/978-3-319-46726-9_31

76. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein aKH. Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features. In Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM ; 2017. DOI: https://doi.org/10.1007/978-3-319-75541-0_13

77. Lieman-Sifry J, Le M, Lau F, Sall S, Golden D. Fastventricle: Cardiac segmentation with ENet. In International Conference on Functional Imaging and

Modeling of the Heart; 2017: Springer. p. 127-138. DOI: https://doi.org/10.1007/978-3-319-59448-4_13

78. Paszke A, Chaurasia A, Kim S, Culurciello E. Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147. 2016. DOI: https://doi.org/10.48550/arXiv.1606.02147

79. Winther HB, Hundt C, Schmidt B, Czerner C, Bauersachs J, Wacker F, et al. v-net: Deep Learning for Generalized Biventricular Mass and Function Parameters Using Multicenter Cardiac MRI Data. JACC: Cardiovascular Imaging. 2018; 11(7): p. 1036-1038. DOI: DOI: 10.1016/j.jcmg.2017.11.013

80. Vesal S, Ravikumar N, Maier A. Dilated Convolutions in Neural Networks for Left Atrial Segmentation in 3D Gadolinium Enhanced-MRI. In International workshop on statistical atlases and computational models of the heart; 2018: Springer. p. 319-328. DOI: https://doi.org/10.1007/978-3-030-12029-0_35

81. Li C, Tong Q, Liao X, Si W, Sun Y, Wang Q, et al. Attention based hierarchical aggregation network for 3D left atrial segmentation. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2018: Springer. p. 255-264. DOI: https://doi.org/10.1007/978-3-030-12029-0_28

82. Xia Q, Yao Y, Hu Z, Hao A. Automatic 3D Atrial Segmentation from GE-MRIs Using Volumetric Fully Convolutional Networks. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2018: Springer. p. 211-220. https://doi.org/10.1007/978-3-030-12029-0_23

83. Yang X, Wang N, Wang Y, Wang X, Nezafat R, Ni D, et al. Combating uncertainty with novel losses for automatic left atrium segmentation. In International workshop on statistical atlases and computational models of the heart; 2018: Springer. p. 246-254. DOI:10.1007/978-3-030-12029-0_27

84. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Lecture Notes in Computer Science; 2015: Springer. p. 346–361. DOI: https://doi.org/10.1007/978-3-319-10578-9_23

85. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2881-2890. DOI:10.1109/CVPR.2017.660

86. Bian C, Yang X, Ma J, Zheng S, Liu YA, Nezafat R, et al. Pyramid network with online hard example mining for accurate left atrium segmentation. In international

workshop on statistical atlases and computational models of the heart; 2018: Springer. p. 237-245. DOI: https://doi.org/10.1007/978-3-030-12029-0_26

87. Vigneault DM, Xie W, Ho CY, Bluemke DA, Noble JA. Ω-Net (Omega-Net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. Medical image analysis. 2018; 48: p. 95-106. DOI: https://doi.org/10.1016/j.media.2018.05.008

88. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of the Society for Cardiovascular Magnetic Resonance. 2018; 20(1): p. 1-12. DOI: 10.1186/s12968-018-0471-x

89. Xiong Z, Fedorov VV, Fu X, Cheng E, Macleod R, Zhao J. Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network. IEEE transactions on medical imaging. 2018; 38(2): p. 515-524. DOI: 10.1109/TMI.2018.2866845

90. Preetha CJ, Haridasan S, Abdi V, Engelhardt S. Segmentation of the left atrium from 3D gadolinium-enhanced MR images with convolutional neural networks. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2018: Springer. p. 265-272. DOI: 10.1007/978-3-030-12029-0_29

91. Chen C, Bai W, Rueckert D. Multi-task learning for left atrial segmentation on GE-MRI. In International workshop on statistical atlases and computational models of the heart; 2019: Springer. p. 292-301. DOI:10.1007/978-3-030-12029-0_32

92. Savioli N, Montana G, Lamata P. V-FCNN: volumetric fully convolution neural network for automatic atrial segmentation. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2019: Springer. p. 273-281. DOI: https://doi.org/10.1007/978-3-030-12029-0_30

93. Jia S, Despinasse A, Wang Z, Delingette HPX, Jaïs P, Cochet H, et al. Automatically segmenting the left atrium from cardiac images using successive 3D U-Nets and a contour loss. In International Workshop on Statistical Atlases and Computational Models of the Heart; 2019: Springer. p. 221-229. DOI: https://doi.org/10.1007/978-3-030-12029-0_24

94. Mortazi A, Karim R, Rhode K, Burt J, Bagci U. CardiacNET: Segmentation of Left Atrium and Proximal Pulmonary Veins from MRI Using Multi-View CNN. In Medical Image Computing and Computer-Assisted Intervention; 2017: Springer. p. 377-385. DOI: https://doi.org/10.1007/978-3-319-66185-8_43

95. Yang G, Chen J, Gao Z, Zhang H, Ni H, Angelini E, et al. Multi view sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images. In 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018. p. 1123-1127. DOI: 10.1109/EMBC.2018.8512550

96. Shen Y, Fang Z, Gao Y, Xiong N, Zhong C, Tang X. Coronary Arteries Segmentation Based on 3D FCN With Attention Gate and Level Set Function. IEEE Access. 2019; 7: p. 42826-42835. DOI:10.1109/ACCESS.2019.2908039

97. Huang W, Huang L, Lin Z, Huang S, Chi Y, Zhou J, et al. Artery Segmentation by Deep Learning Neural Networks on Computed Tomographic Coronary Angiographic Images. In IEEE Engineering in Medicine and Biology Society (EMBC); 2018. p. 608-611. DOI: 10.1109/EMBC.2018.8512328

98. Hong Y, Commandeur F, Cadet S, Goeller M, Doris MK, Chen X, et al. Deep learning-based stenosis quantification from coronary CT Angiography. In Medical Imaging 2019: ImageProcessing, International Society for Optics and Photonics; 2019. DOI: 10.1117/12.2512168

99. Liu J, Jin C, Feng J, Du Y, Lu J, Zhou J. A Vessel-Focused 3D Convolutional Network for Automatic Segmentation and Classification of Coronary Artery Plaques in Cardiac CTA. In Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges; 2019: Springer. p. 131-141. DOI: https://doi.org/10.1007/978-3-030-12029-0_15

100. Pan LS, Li CW, Su SF, Tay SY, Tran QV, Chan WP. Coronary artery segmentation under class imbalance using a U-Net based architecture on computed tomography angiography images. Scientific Reports. 2021; 11(1): p. 1-7. DOI: https://doi.org/10.1038/s41598-021-93889-z

101. Lei Y, Guo B, Fu Y, Wang T, Liu T, Curran W, et al. Automated coronary artery segmentation in Coronary Computed Tomography Angiography (CCTA) using deep learning neural networks. In Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications; 2020: SPIE. p. 279-284. DOI: https://doi.org/10.1117/12.2550368

102. Gao Z, Liu X, Qi S, Wu W, Hau WK, Zhang H. Automatic segmentationof coronary tree in CT angiography images. International Journal of Adaptive Control and Signal Processing. 2019; 33(8): p. 1239-1247. DOI: https://doi.org/10.1002/acs.2762

103. Lesageab D, Angelinib ED, Blochb I, Funka-Leaa G. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Medical Image Analysis. 2009; 13(6): p. 819-845. DOI: https://doi.org/10.1016/j.media.2009.07.011

104. Kirişli H A, Schaap M, Metz CT, Walsum Tv, Giessen AGvd, Weustink AC, Mollet NR, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. Medical Image Analysis. 2013; 17(8): p. 859-876. DOI: 10.1016/j.media.2013.05.007

105. Zreik M, Hamersvelt RWv, Wolterink JM, Leiner T, Viergever MA, Isgum I. A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography. IEEE Transactions on Medical Imaging. 2019; 38(7): p. 1588-1598. DOI: 10.1109/TMI.2018.2883807

106. Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson K. Coronary Calcium Score and Cardiovascular Risk. Journal of The American College of Cardipology. 2018;: p. 434-47. DOI: 10.1016/j.jacc.2018.05.027

107. Kurkure U, Chittajallu DR, Brunner G, Le YH, Kakadiaris IA. A supervised classification-based method for coronary calcium detection in non-contrast CT. The International Journal of Cardiovascular Imaging. 2010; 26: p. 817–828. https://doi.org/10.1007/s10554-010-9607-2

108. Brunner G, Chittajallu DR, Kurkure U, Kakadiaris IA. Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data. The International Journal of Cardiovascular Imaging. 2010; 26: p. pages829–838. DOI: https://doi.org/10.1007/s10554-010-9608-1

109. Sanchez CI, Niemeijer M, Isgum I, Dumitrescu A, Suttorp-Schulten MS, Abramoff MD, et al. Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans. Medical image analysis. 2012; 16: p. 50-62. DOI: https://doi.org/10.1016/j.media.2011.05.004

110. Shahzad R, Walsum Tv, Schaap M, Rossi A, Klein S, Weustink AC, et al. Vessel specific coronary artery calcium scoring: an automatic system. Academic Radiology. 2013; 20(1): p. 1-9. DOI: https://doi.org/10.1016/j.acra.2012.07.018

111. Wolterink JM, Leiner T, Takx RAP, Viergever MA, Isgum I. Automatic Coronary Calcium Scoring in Non-Contrast-Enhanced ECG-Triggered Cardiac CT With

Ambiguity Detection. IEEE Transactions on Medical Imaging. 2015; 34(09): p. 1867-1878. DOI: DOI: 10.1109/TMI.2015.2412651

112. Isgum I, Rutten A, Prokop M, Staring M, Klein S, Pluim JPW, et al. Automatic Coronary Calcium Scoring in Low-Dose Chest Computed Tomography. IEEE Transactions on Medical Imaging. 2010; 31(12): p. 2322-2334. DOI: DOI: 10.1109/TMI.2012.2216889

113. Yang G, Chen Y, Ning X, Sun Q, Shu H, Coatrieux JL. Automatic coronary calcium scoring using noncontrast and contrast CT images. Medical Physics. 2016; 43(5): p. 2174. DOI: 10.1109/TMI.2015.2412651

114. Merkow J, Marsden A, Kriegman D, Tu Z. Dense Volume-to-Volume Vascular Boundary Detection. In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016; 2016: Springer. p. 371-379. DOI; https://doi.org/10.1007/978-3-319-46726-9_43

115. Moeskops P, Wolterink JM, Velden BHMvd, Gilhuijs KGA, Leiner T, Viergever MA, et al. Deep learning for multi-task medical image segmentation in multiple modalities. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016; 2016. p. 478-486. DOI: https://doi.org/10.1007/978-3-319-46723-8_55

116. Wolterink JM, Leiner T, Vos BDd, Hamersvelt RWv, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. Medical Image Analysis. 2016; 34: p. 123-136. DOI: https://doi.org/10.1016/j.media.2016.04.004

117. Lessmann N, Ginneken Bv, Zreik M, Jong PAd, Vos BDd, Viergever MA, et al. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions. IEEE Transactions on Medical Imaging. 2018; 37(2): p. 615-625. DOI: 10.1109/TMI.2017.2769839

118. Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ. Automated Agatston Score Computation in non-ECG Gated CT Scans Using Deep Learning. In SPIE–the International Society for Optical Engineering; 2018. DOI: 10.1117/12.2293681

119. Vos BDd, Wolterink JM, Leiner T, Jong PAd, Lessmann N, Išgum I. Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. IEEE Transactions on Medical Imaging. 2019; 38(9): p. 2127-2138. DOI:10.1109/TMI.2019.2899534

120. Zhao F, Wu B, Chen F, Yi XCH, Hou Y, He X, et al. An automatic multi-class coronary atherosclerosis plaque detection and classification framework. Medical & Biological Engineering & Computing. 2019; 57: p. 245–257. https://doi.org/10.1007/s11517-018-1880-6

121. Jawaid MM, Riaza A, Rajanib R, Reyes-Aldasoroa CC, Slabaugha G. Framework for detection and localization of coronary non-calcified plaques in cardiac CTA using mean radial profiles. Computers in Biology and Medicine. 2017; 89: p. 84-95. DOI: https://doi.org/10.1016/j.compbiomed.2017.07.021

122. Wei J, Zhou C, Chan HP, Chughtai A, Agarwal P, Kuriakose J, et al. Computerized detection of noncalcified plaques in coronary CT angiography: evaluation of topological soft gradient prescreening method and luminal analysis. Medical Physics. 2014; 41(8). DOI: 10.1118/1.4885958

123. Zuluaga MA, Hush D, Leyton EJFD, Hoyos MH, Orkisz M. Learning from Only Positive and Unlabeled Data to Detect Lesions in Vascular CT Images. In International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI; 2011. p. 9-16. DOI: https://doi.org/10.1007/978-3-642-23626-6_2

124. Hampe N, Wolterink JM, Velzen SGMv, Leiner T, Išgum I. Machine Learning for Assessment of Coronary Artery Disease in Cardiac CT: A Survey. Frontiers in Cardiovascular Medicine. 2019; 6: p. 172. DOI: 10.3389/fcvm.2019.00172

125. Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. Medical image analysis. 2018; 44: p. 72-85. DOI: 10.1016/j.media.2017.11.008

126. Candemir S, White RD, Demire M, Gupta V, Bigelow MT, Prevedello LM, et al. Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary CT angiography with a deep 3-dimensional convolutional neural network. Computerized Medical Imaging and Graphics. 2020; 83: p. 101721. DOI: https://doi.org/10.1016/j.compmedimag.2020.101721

127. Acharya UR, Meiburger KM, Koh JEW, Vicnesh J, Ciaccio EJ, Lih OS, et al. Automated plaque classification using computed tomography angiography and Gabor transformations. Artificial intelligence in medicine. 2019; 100: p. 101724. DOI: 10.1016/j.artmed.2019.101724

128. White RD, Erdal BS, Demirer M, Gupta V, MTB, Dikici E, et al. Artificial Intelligence to Assist in Exclusion of Coronary Atherosclerosis During CCTA Evaluation of Chest Pain in the Emergency Department: Preparing an Application

for Real-world Use. Journal of Digital Imaging. 2021; 34(3): p. 554-571. DOI: 10.1007/s10278-021-00441-6

129. Marcovitch H. Black's medical dictionary. 42nd ed.: A & C Black; 2009.

130. Ong TK, Chin SP, Liew CK, Chan WL, Seyfarth MT, Liew HB, et al. Accuracy of 64-row multidetector computed tomography in detecting coronary artery disease in 134 symptomatic patients: influence of calcification. American heart journal. 2006; 151(6): p. 1323.e1-1323.e6. DOI: https://doi.org/10.1016/j.ahj.2005.12.027

131. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In International Conference on Computer Vision (ICCV); 2017. p. 618-626. DOI: https://doi.org/10.1007/s11263-019-01228-7

132. Demirer M, Gupta V, Bigelow M, Erdal B, Prevedello L, White R. Image dataset for a CNN algorithm development to detect coronary atherosclerosis in coronary CT angiography; 2019. DOI: 10.17632/fk6rys63h9.1

133. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. In IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248-255. DOI:10.1109/CVPR.2009.5206848

134. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common Objects in Context. In Computer Vision – ECCV 2014; 2014; Cham: Springer International Publishing. p. 740-755. DOI: https://doi.org/10.1007/978-3-319-10602-1_48

135. Mottaghi R, Chen X, Liu X, Cho NG, Lee SW, Fidler S, et al. The Role of Context for Object Detection and Semantic Segmentation in the Wild. In Computer Vision and Pattern Recognition (CVPR); 2014. DOI:10.13140/2.1.2577.6000

136. Mathworks. [Online]: Mathworks [cited 2021]. Available from: https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html .

137. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV). 2015; 115(3): p. 211-252. DOI: https://doi.org/10.1007/s11263-015-0816-y

138. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 2818-2826. DOI:10.1109/CVPR.2016.308

139. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence; 2017: AAAI Press. p. 4278–4284. DOI: https://doi.org/10.1609/aaai.v31i1.11231

140. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering. 2009; 21(9): p. 1263-1284. DOI: DOI:10.1109/TKDE.2008.239

141. Fernàndez A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. Learning from Imbalanced Data Sets: Springer; 2018. DOI: https://doi.org/10.1007/978-3-319-98074-4

142. Holte RC, Acker L, Porter BW. Concept Learning and the Problem of Small Disjuncts. In International Joint Conference on Artificial Intelligence; 1989. p. 813-818. URL: http://www.ijcai.org/Proceedings/89-1/Papers/130.pdf

143. Liu XY, Wu J, Zhou ZH. Exploratory Under Sampling forClass Imbalance Learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2009; 39(2): p. 539-550. DOI: DOI:10.1109/tsmcb.2008.2007853

144. Engstrom L, Tran B, Tsipras D, Schmidt L, Madry A. A rotation and a translation suffice: Fooling cnns with simple transformations.: arXiv; 2018 [cited 2022 06 14. Available from: https://arxiv.org/abs/1712.02779v1 .

145. Tanaka FHKdS, Aranha C. Data augmentation using GANs. arXiv preprint arXiv:1904.09135. 2019. DOI: https://doi.org/10.48550/arXiv.1904.09135

146. Lan L, You L, Zhang Z, Fan Z, Zhao W, Zeng N, et al. Generative Adversarial Networks and Its Applications in Biomedical Informatics. Frontiers in Public Health. 2020; 8: p. 164. DOI: 10.3389/fpubh.2020.00164

147. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In Advances in Neural Information Processing Systems 27 (NIPS); 2014. p. 2672–2680. DOI:10.1145/3422622

148. Deng L. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning. Technometrics. 2006; 48(1): p. 147-148. DOI:10.1198/tech.2006.s353

149. Skandarani Y, Painchaud N, Jodoin PM, Lalande A. On the effectiveness of GAN generated cardiac MRIs for segmentation. In Medical Imaging with Deep Learning; 2020. DOI: https://doi.org/10.48550/arXiv.2005.09026

150. Wulan N, Wang W, Sun P, Wang K, Xia Y, Zhang H. Generating electrocardiogram signals by deep learning. Neurocomputing. 2020; 404(3): p. 122-136. DOI:10.1016/j.neucom.2020.04.076

151. Skandarani Y, Lalande A, Afilalo J, Jodoin PM. Generative Adversarial Networks in Cardiology. Canadian Journal of Cardiology. 2022; 38(2): p. 196-203. DOI: https://doi.org/10.1016/j.cjca.2021.11.003

152. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. Canadian Journal of Cardiology. 2021; 38(2): p. 204-213. DOI:10.1016/j.cjca.2021.09.004

153. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA); 2018: IEEE. p. 80-89. DOI:10.1109/DSAA.2018.00018

154. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics. 2015; 24(1): p. 44-65. DOI:10.1080/10618600.2014.907095

155. Campello VM, Gkontra P, Izquierdo C, Martın-Isla C, Sojoudi A, Full PM, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m&ms challenge. IEEE Transactions on Medical Imaging. 2021; 40(12): p. 3543-3554. DOI:10.1109/TMI.2021.3090082

156. Ghamisi P, Couceiro MS, Benediktsson JA. Extending the fractional order Darwinian particle swarm optimization to segmentation of hyperspectral images. Image and Signal Processing for Remote Sensing. 2012; 8537: p. 8537. DOI:10.1117/12.978776

157. Song X, Li H. Segmentation Based on Particle Swarm Optimization. In International Conference on Applications and Techniques in Cyber Security and

Intelligence; 2020: Springer. p. 731-736. DOI: https://doi.org/10.1007/978-3-030-53980-1_107

158. Lani J. ANOVA (Analysis of Variance). Statistics Solution. 2010. URL ; https://www.statisticssolutions.com/wp-content/uploads/wp-post-to-pdf-enhanced-cache/1/anova.pdf