UNIVERSITA' POLITECNICA DELLE MARCHE

FACOLTA' DI INGEGNERIA

Master's Degree in Biomedical Engineering

Curriculum: E-Health and Clinical Engineering

A PREDICTIVE APPROACH FOR VENTRICULAR ARRHYTHMIAS IN CARDIOPATHIC PATIENTS: STUDY AND ANALYSIS OF ELECTROANATOMIC MAPS AND MACHINE LEARNING

Supervisor:

Prof. **Francesco Piva**

Candidate:

Emanuele Pecorari

Co-Supervisor:

Prof. **Michela Casella**

Academic Year: 2022 / 2023

INDEX

1. INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of death globally. They are usually divided into: Heart disorders that affect the heart, its valves, and the blood vessels that supply blood to the heart muscle and Peripheral vascular diseases that affect the blood vessels of the arms, legs, and trunk [1].

Diseases affecting the blood vessels supplying the brain are referred to as cerebrovascular diseases.

Heart diseases include:

• Diseases of the blood vessels, such as coronary artery disease

• Irregular heartbeats (arrhythmias)

• Congenital heart problems (congenital heart defects)

• Heart muscle disease

• Diseases of the heart valves

Understanding the various disorders affecting the heart requires adequate knowledge of cardiac physiology.

1.1 HEART: ANATOMY AND PHYSIOLOGY

"All the information presented in this chapter are taken from [2]".

The heart is an involuntary muscle about the size of a fist, located at the center of the chest in an area called the mediastinum. It is surrounded by a sac called the pericardium and divided into a right and left part, separated by a septum. Each of the two parts consists of two chambers, an upper atrium and a lower ventricle, see Figure 1. Each atrium is connected to the corresponding ventricle through the atrioventricular opening equipped with valves: the tricuspid valve separates the right cavities, while the mitral valve is located between the left atrium and the left ventricle. The openings that connect the heart chambers to the efferent vessels are protected by valves that prevent blood backflow: the semilunar (pulmonary) valve in the right ventricle for the pulmonary artery and the semilunar (aortic) valve in the left ventricle for the aorta. Blood flow occurs only when there is a pressure difference across the valves, causing them to open. Under normal

conditions, the valves allow blood to flow in one direction. The heart pumps blood to the lungs and to all tissues of the body through a highly organized sequence of contractions of the four chambers. For the heart to function properly, the four chambers must beat in an organized manner [2]."

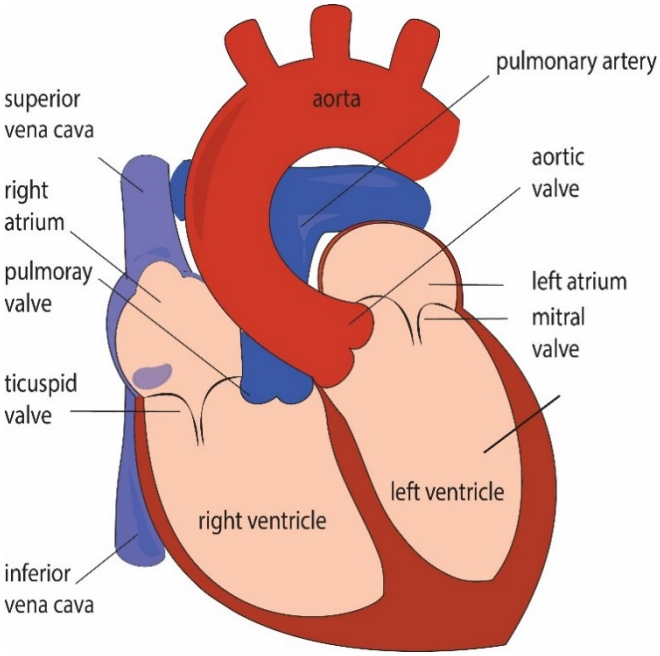

Figure 1. | *Anatomy of the heart*

1.1.1 CARDIAC CONDUCTION SYSTEM

"All the information presented in this chapter are taken from [2]".

As shown in Figure 2, the cardiac conduction system consists of:

• Sinoatrial node (natural pacemaker)

• Internodal pathways (atrial conduction)

• Atrioventricular node

• Intraventricular conduction system (bundle of His, common trunk, and right and left branches)

• Purkinje fibers

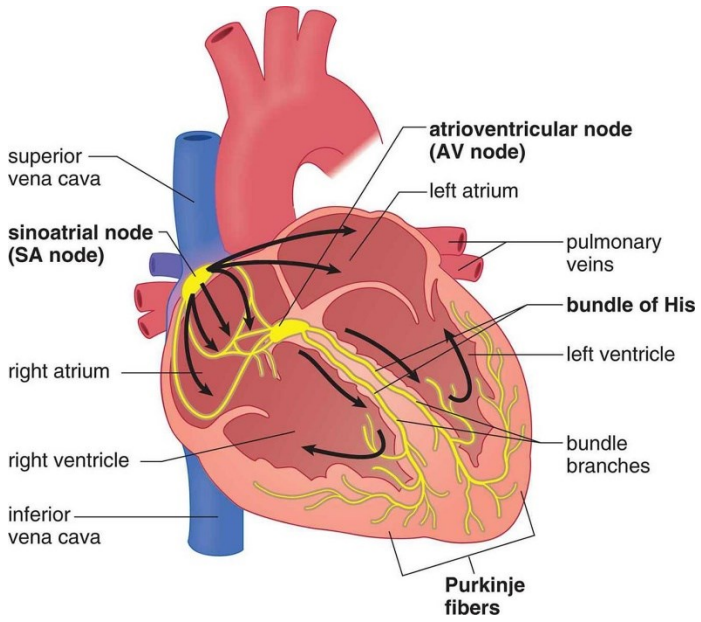

Figure 2 | *Anatomy of cardiac conduction system*

The sinoatrial node is located in the right atrium near the opening of the superior vena cava, just below the endocardium. It has a length of approximately 15 mm and a thickness of 2 mm. It generates impulses at a frequency of 60-100 beats per minute. In addition to having this intrinsic mechanism, it is subject to external control, allowing the heart rate to adapt to the body's various needs. External control of the heart is essentially supported by the autonomic nervous system with its sympathetic (adrenergic) and parasympathetic (vagal) branches. Sympathetic stimulation increases cardiac metabolism, conduction velocity, heart rate, and contractile force, prevailing during physical exercise. During sleep or rest, parasympathetic tone prevails, slowing the heart rate and acting in the opposite direction. From the sinoatrial node, internodal pathways extend, forming the conduction pathways through which the impulse spreads from the sinoatrial node to the atrioventricular node. The atrioventricular node is located posteriorly on the right side of the interatrial septum, near the coronary sinus. It has a length of 22 mm and a thickness of 3 mm. Since the atrial and ventricular myocardium are not electrically connected, electrical activity propagates to the ventricles exclusively through the atrioventricular node. The atrioventricular node serves as a secondary impulse formation center with a spontaneous discharge frequency of 40-60 beats per minute. Its function is to delay the propagation of the impulse from the atria to the ventricles, acting as a filter. The intraventricular conduction system consists of the bundle of His, which runs along the right side of the interventricular septum in a subendocardial position for approximately 12 mm. The common trunk of the bundle of His, along with the atrioventricular

node, forms the atrioventricular junction. From the common trunk, the right and left branches originate, directed towards their respective ventricles. The right branch continues the course of the bundle of His along the septum. The left branch, of larger diameter, perforates the interventricular septum, dividing into an anterosuperior bundle and a posteroinferior bundle. Purkinje fibers constitute the final conduction network branching into the subendocardium of both ventricles, representing the terminal part of the cardiac conduction system. These fibers propagate the electrical impulse to all parts of the ventricular myocardium, allowing for the synchronized contraction of both ventricles.

1.1.2 ELECTRICAL ACTIVITY OF MYOCARDIAL CELLS

The heart is composed of a set of excitable and contractile cells called cardiomyocytes, which come in three different types:

• Working myocardium, which mainly contains contractile material;

• Nodal cells, endowed with automatic excitability, from which the electrical stimulus originates;

• Conduction tissue, with cells organized for rapid and orderly propagation of excitation throughout the working myocardium.

The cyclical functioning of the heart is ensured by the continuous transition of myocardial cells from the resting state to the excited state, due to cellular bioelectrical phenomena. Every mechanical activity, i.e., every contraction of the heart, is preceded by electrical activity, which we record with an ECG (electrocardiogram). Myocardial cells have an electrical potential on their membrane, known as the "resting potential," of -90 mV compared to the surrounding environment. This potential is determined by the different concentrations of ions on both sides of the plasma membrane. These concentration variations are maintained by pumps that expel some ions in exchange for others, consuming energy to counteract the natural tendency towards equilibrium and the electrical neutrality of the cell. In this condition, the myocardial cell is excitable, meaning it can change its resting potential and transmit this rapid change throughout the cell and to neighboring cells. This phenomenon is known as the action potential. When the passage of charges through the membrane reduces the membrane potential, depolarization occurs. This is the starting point for the formation and propagation of the electrical impulse. Once a certain "threshold value" is exceeded, a significant change in potential (from -90 to +50 mV)

occurs very rapidly. The action potential returns rapidly to its initial value, leading to hyperpolarization (increase in membrane potential), making the cell refractory for a certain period during which it cannot be excited again (refractory period). The action potential has different characteristics depending on the type of myocardial tissue in which it occurs. Common myocardial cells (comprising the atria and ventricles, as well as Purkinje fibers) exhibit a "fast response" action potential. As shown in Figure 3, this is determined by the opening of sodium channels entering the cell (phase 0), followed by subsequent repolarization due to the transient passage of chloride ions (phase 1) and a phase (2) in which calcium and sodium ions enter the cell through slow channels (plateau). There is then the final repolarization phase (phase 3), during which the cell returns to its initial conditions due to potassium efflux from the cell and the subsequent restoration of ion concentrations to resting values (phase 4). At the level of the sinoatrial and atrioventricular nodes, the action potential assumes a "slow response": phase 0 begins more slowly, there is no plateau, and phase 3 is more gradual. Additionally, there is a phase 4, represented by a constant and gradual depolarization (due to weak sodium influx and decreased potassium efflux). Once the threshold value is reached, this phase leads to the generation of the action potential. This phenomenon is fundamental for cardiac activity, as it autonomously and independently generates an electrical impulse at well-defined rhythmic intervals. This impulse spreads and triggers a cardiac contraction. For this reason, cells with a slow response are called pacemakers. In summary, an electrical system sends rhythmic impulses that determine cardiac contraction.

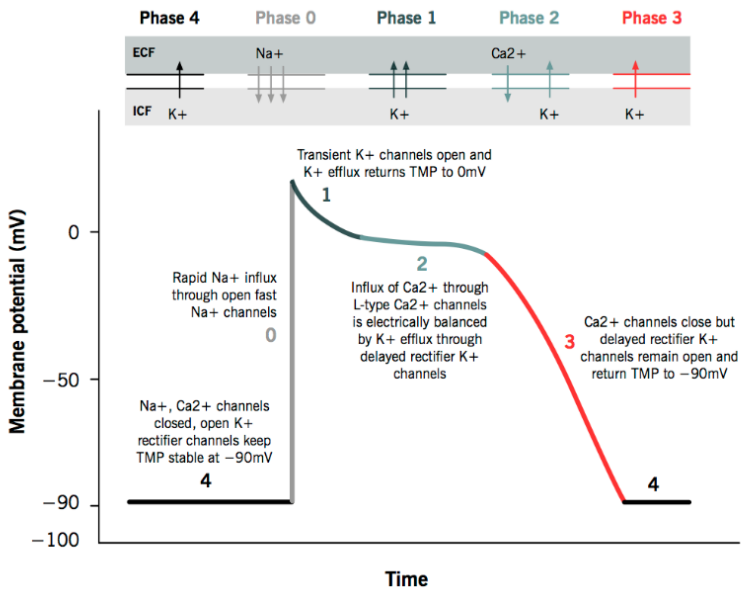

Figure 3 | *Action Potential of Cardiac Muscles*

1.2 ARRHYTHMIAS

The term "arrhythmia" refers to any problem with the rhythm or rate of a person's heartbeat [3]. During an arrhythmia, electrical impulses may be too fast, too slow, or irregular, resulting in an irregular heartbeat. This occurs when the electrical signals that instruct the heart to beat do not work properly. Multiple factors can cause an arrhythmia. For example, anxiety, daily stress, or anything that activates the "fight or flight" stress response can speed up the heartbeat and cause tachycardia. Other common causes of irregular heartbeats associated with cardiac arrhythmias include autonomic imbalances, heart diseases, overstimulation of the vagus nerve, medications that trigger or exacerbate arrhythmias, electrolyte or metabolic disturbances, and genetic disorders. In general, cardiac arrhythmias are grouped based on heart rate [4]. For example, tachycardia is a fast heartbeat, with a rate of more than 100 beats per minute, while bradycardia is a slow heartbeat, with a rate of less than 60 beats per minute. Types of tachycardia include:

- Atrial fibrillation: Chaotic heart signals cause a rapid and disorganized heartbeat. Atrial fibrillation may be temporary and start and stop on its own, but some episodes may not stop without treatment.

- Atrial flutter: Similar to atrial fibrillation, but heartbeats are more organized. Atrial flutter is also associated with the risk of stroke.

- Supraventricular tachycardia: This generic term includes irregular heartbeats originating above the lower chambers of the heart, called ventricles. Supraventricular tachycardia causes episodes of pounding heartbeat that start and stop suddenly.

- Ventricular fibrillation: Fast and chaotic electrical signals cause the lower chambers of the heart to quiver instead of contracting cohesively. This serious problem can lead to death if a regular heart rhythm is not restored within a few minutes. Most people with ventricular fibrillation have an underlying heart disease or have suffered serious trauma.

- Ventricular tachycardia: This fast and irregular heart rate starts with faulty electrical signals in the lower chambers, called ventricles. The rapid heart rate prevents the ventricles from filling properly with blood, so the heart may not be able to pump enough blood to the body. Ventricular tachycardia may not cause serious problems in people with otherwise healthy hearts, but in those with heart diseases, it can be a medical emergency that requires immediate attention.

Bradycardia, a slow heartbeat with a rate of less than 60 beats per minute, can be caused by:

- Sick sinus syndrome: The sinoatrial node determines the heart rate. If the node does not function properly, the heart rate may alternate between too slow and too fast. Sick sinus syndrome can be caused by scars near the sinoatrial node that slow down, interrupt, or block the heart rate signals. The condition is more common among the elderly.

- Conduction block: A blockage in the heart's electrical pathways can cause the slowing or interruption of signals that activate heartbeats. Some blocks may not cause symptoms, while others may cause skipped or slowed heartbeats.

Premature heartbeats are extra beats that occur singly, sometimes in alternating patterns with a regular heartbeat [5]. If the extra beats come from the upper part of the heart, they are called premature atrial contractions (PACs). If they come from the lower part, they are called premature ventricular contractions (PVCs). A premature heartbeat may feel like the heart skipped a beat. Generally, these extra beats are not a cause for concern and rarely indicate a more serious condition. However, a premature beat can trigger a sustained arrhythmia, especially in people with heart disease. Sometimes, having very frequent premature ventricular contractions can lead to weakening of the heart. Premature heartbeats can occur at rest. Stress, intense physical exercise, and the use of stimulants such as caffeine or nicotine can also cause premature heartbeats.

1.3 CARDIOMYOPATHIES

Cardiomyopathy is a disease of the heart muscle that makes it harder for the heart to pump blood to the rest of the body. The various types of the disease have many causes, signs, symptoms, as well as treatments. In most cases, cardiomyopathy causes the heart muscle to become enlarged or stiffened. In rare cases, the tissue of the diseased heart muscle is replaced with scar tissue. As cardiomyopathy worsens, the heart becomes weaker. The heart becomes less able to pump blood throughout the body and unable to maintain a normal electrical rhythm. The result can be heart failure or irregular heartbeats called arrhythmias. A weakened heart can also cause other complications, such as problems with heart valves.

The main types of cardiomyopathy are:

- Dilated cardiomyopathy

- Hypertrophic cardiomyopathy

- Restrictive cardiomyopathy

Some cases of cardiomyopathy have no signs or symptoms and do not require treatment. In other cases, cardiomyopathy develops rapidly with severe symptoms and serious complications occur. In these cases, treatment is necessary [6]. Treatments include lifestyle changes, medications, surgeries, implanted devices to correct arrhythmias, and other non-surgical procedures. These treatments can control symptoms, reduce complications, and prevent the worsening of the disease.

1.3.1 DILATED CARDIOMYOPATHY

Dilated cardiomyopathy (DCM) is the most common type and mainly occurs in adults under the age of 50. It affects the ventricles and atria of the heart, the lower and upper chambers of the heart. Often, the disease begins in the left ventricle, the heart's main pumping chamber. The heart muscle begins to dilate, stretching and thinning. As a result, the inside of the chamber enlarges. The problem often spreads to the right ventricle and then to the atria. When the heart chambers dilate, the heart muscle does not contract normally and cannot pump blood well. When the heart weakens, heart failure can occur. Common symptoms of heart failure include shortness of breath, fatigue, and swelling of the ankles, feet, legs, abdomen, and neck veins. Dilated cardiomyopathy can also lead to problems with heart valves, arrhythmias (irregular heartbeats), and blood clots in the heart.

1.3.2 HYPERTROPHIC CARDIOMYOPATHY

Hypertrophic cardiomyopathy is often caused by abnormal genes in the heart muscle. These genes cause the walls of the heart chamber (left ventricle) to become thicker than normal. The thickened walls can stiffen, reducing the amount of blood drawn in and pumped out to the body with each heartbeat. In obstructive HCM, the thickened part of the heart muscle, usually the wall (septum) between the two lower chambers (ventricles), blocks or reduces blood flow from the left ventricle to the aorta. Most people with HCM have this type. In non-obstructive HCM, the heart muscle is thickened but does not block the flow of blood leaving the heart.

1.3.3 RESTRICTIVE CARDIOMYOPATHY

Restrictive cardiomyopathy tends to affect the elderly. The heart's ventricles stiffen because abnormal tissue, such as scar tissue, replaces normal heart muscle. As a result, the ventricles cannot relax normally and fill with blood, and the atria enlarge. Blood flow in the heart decreases over time. This can lead to problems such as heart failure or arrhythmias. Restrictive cardiomyopathy is the least common form of cardiomyopathy, which can be divided into two main categories:

- Non-obliterative: characterized by the abnormal infiltration of the myocardium by a foreign substance.

- Obliterative: characterized by endocardial and subendocardial fibrosis.

1.4 INCIDCENCE AND MORTALITY

According to Istat data for the year 2020 [7], there were 63,952 reported deaths in Italy due to ischemic heart diseases, of which 34,095 were males and 29,857 were females. According to an article published by the Ministry of Health [8], cardiovascular diseases caused 224,482 deaths (97,952 in men and 126,530 in women), accounting for 38.8% of total deaths. Such a high percentage is partly attributed to the aging population and low birth rates that have characterized the country in recent years. For ischemic heart disease, there were 75,046 deaths (37,827 in men and 37,219 in women), approximately 33% of all deaths due to circulatory system diseases. In men, mortality is negligible until the age of 40, starts to emerge between 40 and 50 years, and then increases exponentially with age. In women, this phenomenon starts around the age of 50-60 and increases rapidly. The disadvantage of men compared to women is more pronounced in reproductive age and tends to decrease with advancing age. The difference in disease frequency between the sexes is also associated with clinical differences, with more frequent sudden deaths and silent heart attacks in women. The term "incidence" refers to the number of new cases of a disease occurring in a population during a specific period, usually a year. Incidence data were obtained from studies conducted as part of the CUORE Project, which enrolled over 21,000 men and women aged 35 to 74 starting in the mid-1980s, with an average follow-up period of 13 years. Rates showed an incidence of coronary events (6.1 per 1,000 per year in men with a 28% 28-day

mortality rate and 1.6 per 1,000 per year in women with a 25% mortality rate). The mortality rate was 27.9% in men and 25.4% in women, increasing significantly with age.

The data are reported in Table 1.

Table 1. | *The project HEART reports a section of the table on incidence and fatality rates.*

Age (years)	Coronary events			
	Man		Women	
	Rates of incidence per year per 1,000	Lethality, %	Rates of incidence per year per 1,000	Lethality, %
35-44	3,2	9,6	0,5	8,3
45-54	4,5	15,3	1,2	11,4
55-64	9,7	33,6	2,8	27,1
65-74	10,1	54,2	4,5	54,5
35-74	**6,1**	**27,9**	**1,6**	**25,4**

1.5 TREATMENTS

The treatments for arrhythmias are described in The New York Times on July 23, 2014 [9], and these treatments include: therapy with electric shock (defibrillation or cardioversion) and implantation of a pacemaker in the patient's chest, antiarrhythmic drugs that can be used to prevent the recurrence of an arrhythmia and to prevent the heart rate from becoming too fast or too slow, cardiac ablation for rhythm problems, implantable cardiac defibrillator to control heart failure, and pacemaker device when the heart beats irregularly. It sends a signal to the heart to beat at a precise and normal rate. The choice among all these different approaches depends on various factors such as the patient's age, pre-existing specific medical conditions, and the different types of arrhythmia.

1.5.1 ANTIARRHYTHMIC DRUGS

The Cardiac Arrhythmia Suppression Trial (CAST) changed the use of antiarrhythmic drugs [10]. CAST was designed to test the hypothesis that suppressing antiarrhythmic drugs for premature ventricular contractions (PVC) and unsustained ventricular tachycardia (VT) might improve mortality in post-myocardial infarction patients with reduced left ventricular function. The preferred drugs Moricizine, Flecainide, and Encainide were known to have potent ventricular arrhythmia suppression properties. However, CAST demonstrated an increase in mortality in patients treated with antiarrhythmic drugs compared to placebo. Perhaps the increased mortality rate was due to the proarrhythmic effects of these drugs, especially in the presence of ischemia and left ventricular dysfunction [11]. Therefore, class 1C drugs are contraindicated in patients with coronary artery disease (CAD) and ischemia. There is concern that an increase in mortality may occur with other antiarrhythmics, especially if administered for relatively benign arrhythmias (e.g., atrial fibrillation, PVC). Quinidine has been shown to increase mortality when given to patients with atrial fibrillation [12]. The Vaughn Williams classification of antiarrhythmic drugs is as shown in Chaudhry et al. [11]. Class 1, sodium channel blockers; Class 1A, depresses phase 0 of the action potential, delays conduction, and prolongs repolarization: phase 3 or 4 (quinidine, procainamide, disopyramide); Class 1B, minimal effect on phase 0 of the action potential in normal tissues, depresses phase 0 in abnormal tissues, shortens repolarization or minimal effect (lidocaine, tocainide, mexiletine, phenytoin); Class 1C, depresses phase 0 of the action potential, slows conduction in normal tissues (Flecainide, Propafenone, Moricizine); Class 2, adrenergic beta-blockers (Acebutolol, Atenolol, Bisoprolol, Carvedilol, Metoprolol, Nadolol, Pindolol, Propranolol); Class 3, prolongs action potential duration by increasing repolarization and refractoriness (Amiodarone, Sotalol, Bretylium, Dofetilide, Azimilide, Ibutilide); Class 4, calcium antagonists (Diltiazem, Verapamil); and others (Digoxin, Adenosine). Following the CAST study, several reports have confirmed the proarrhythmic effects of antiarrhythmic drugs if used capriciously, leading to specific guidelines for the use of antiarrhythmic drugs, especially those that prolong the QT interval and increase proarrhythmia. Usually, class IA and III agents are initiated in the hospital with telemetry monitoring. Class IC agents are relatively safe if used in a normal heart. Similarly, amiodarone, due to its longer half-life (from 43 days to a few months) and low incidence of proarrhythmia, can be initiated at low doses on an outpatient basis in the absence of severe left ventricular dysfunction or bradycardia. Based on the results of the CAST study, the Food and Drug Administration (FDA), the United States, and pharmaceutical industries have taken measures to ensure appropriate prescribing practices and medical credentials when the

new antiarrhythmic drug, Dofetilide (Tikosyn), was released for use in patients with atrial fibrillation.

1.5.2 PACEMAKER AND IMPLANTABLE CARDIAC DEFIBRILLATOR (ICD)

The implantation of a permanent pacemaker (PPM) requires specific levels of evidence and indications based on the guidelines of the American College of Cardiology and the American Heart Association (ACC/AHA) [13]. Class I and Class II indications are appropriate for PPM implantation. Correlation of symptoms with underlying bradyarrhythmias or heart block is necessary. The implantable cardiac defibrillator (ICD) is indicated for sustained ventricular tachycardia (VT) or ventricular fibrillation (VF), for survivors of sudden cardiac death, for the Antiarrhythmics versus Implantable Defibrillators (AVID) study for secondary prevention [14], or for the Multicenter Automatic Defibrillator Implantation Trial I (MADIT I) study for inducible monomorphic VT and secondary prevention [15,16]. These devices can be single or dual-chambered and may have rate-responsive capabilities. The results of studies on biventricular and single-chamber ICDs (Dual-Chamber and VVI [DAVID]) have shown that biventricular ICDs in patients with reduced left ventricular function lead to a higher incidence of congestive heart failure (CHF) and increased mortality [17]. The presumed mechanism is the creation of functional left bundle branch block (LBBB), which can lead to cardiac desynchronization and heart failure. Conversely, restoration of cardiac resynchronization with biventricular pacing may improve CHF symptoms, leading to increased referrals for patients implanted with class II to IV CHF and severe ventricular dysfunction [18]. For primary prevention of sudden death in patients with severe left ventricular dysfunction, intact normal sinus node, AV node, and conduction, and no perceived or anticipated indication for pacing, the preferred ICD device is a single-chamber device (stimulated and inhibited ventricular pacemaker [VVI]). Biventricular ICDs should be reserved for patients with abnormal SA or AV nodal or conduction system or those with frequent supraventricular arrhythmias, such as atrial fibrillation, to avoid shocks secondary to rapid ventricular responses. When biventricular pacing is necessary and the left ventricle is severely compromised, consideration should be given to a biventricular ICD.

1.5.3 ABLATION

Transcatheter radiofrequency ablation (RFA) has been a significant advancement in the treatment of cardiac arrhythmias [19]. RFA has provided the opportunity to address specific cardiac conditions. Introduced over 20 years ago as DC ablation and in the 1980s with radiofrequency, RFA has proven to be a safe and cost-effective treatment for specific cardiac arrhythmias such as atrioventricular nodal reentrant tachycardia (AVNRT), orthodromic reciprocating tachycardia associated with Wolff-Parkinson-White syndrome (WPW), concealed accessory pathways, normal cardiac ventricular tachycardia (especially right ventricular outflow tract tachycardia or fascicular tachycardia), and atrial flutter [20,21]. RFA can also provide adjunctive therapy for ischemic ventricular tachycardia when patients experience frequent ICD shocks or failed antiarrhythmic therapy. Pulmonary vein isolation is a therapeutic option for symptomatic drug-refractory, paroxysmal, or persistent atrial fibrillation and is an acceptable method currently under clinical investigation. However, rate control and chronic anticoagulation are acceptable alternatives for patients with atrial fibrillation, as highlighted by the AFFIRM (Atrial Fibrillation Follow-up Investigation of Rhythm Management) study [22]. Macro reentrant tachycardia, known as orthodromic alternate tachycardia (ORT) or AVRT, alternate tachycardia, occurs when the AV node is used in the anterograde direction and the accessory pathway is used in the retrograde direction. AVRT is a narrow and complex tachycardia but may have small retrograde P waves visible between the QRS and T waves. When the accessory pathway is used in the anterograde direction, antidromic alternate tachycardia (ART) occurs, resulting in wide and complex tachycardia mimicking ventricular tachycardia. Atrial fibrillation is common with WPW syndrome, constant retrograde reentry into the atrium during ventricular depolarization. Due to the potential for rapid conduction over an accessory pathway with atrial fibrillation and WPW, caution is needed with AV nodal blocking agents, digoxin, and calcium channel blockers. Although rare, atrial fibrillation with rapid ventricular response over an accessory pathway can trigger ventricular fibrillation, leading to sudden death. Acute treatment of atrial fibrillation and WPW consists of cardioversion and occasionally intravenous procainamide. The most common location for an accessory pathway is in the left free ventricular wall, but it can also be posteroseptal or right. RFA has been successful in ablating and curing WPW. The success rate is 97% and has been safely achieved in many centers. For symptomatic WPW, especially in young patients, RFA is considered the treatment of choice. Radiofrequency ablation (RFA) has also been extremely useful in curing typical atrial flutter, identified by an atrial rate of 240 beats/min or higher and characteristic negative sawtooth flutter waves on the ECG, typically in the inferior

leads (II, III, and aVF). Mapping studies have shown that typical flutter occurs with counterclockwise rotation of atrial activation descending on the free wall of the right atrium, crossing the isthmus (area between the opening of the coronary sinus and the tricuspid annulus), and ascending the intra-atrial septum. Interruption of conduction on the isthmus by RFA can successfully eliminate typical flutter. However, 25% of patients continued to have atrial tachyarrhythmias, particularly atrial fibrillation. RFA is an acceptable first-line therapy for symptomatic atrial flutter.

1.5.3.1 PROCEDURE OF ABLATION

Catheter ablation procedures are performed in electrophysiology laboratories [23]. In this process, three or four electrode catheters are percutaneously inserted into a femoral vein, internal jugular vein, or subclavian vein and are positioned inside the heart to allow stimulation and recording at crucial sites. The efficacy of trans-catheter ablation is closely linked to the accurate identification of the site of origin of the arrhythmia. Once identified, an electrode catheter is placed directly in contact with the site, and radiofrequency energy is applied through the catheter to destroy it. Radiofrequency energy is delivered with wavelengths ranging from 300 to 750 kHz during trans-catheter ablation procedures. This process induces resistive heating of the tissue in contact with the electrode. Since the degree of tissue heating is inversely proportional to the radius to the fourth power, the lesions created by radiofrequency energy are small. Typical ablation catheters, with a diameter of 2.2 mm (7 French) and a distal electrode length of 4 mm, create lesions approximately 5-6 mm in diameter and 2-3 mm in depth. Larger lesions can be obtained with larger electrodes or irrigated catheters using saline solution. Although electrical damage may contribute, the primary mechanism for tissue destruction through radiofrequency current is thermal damage. Irreversible tissue destruction requires the tissue temperature to reach about 50°C. In most ablation procedures, the power supplied by the radiofrequency generator is manually or automatically adjusted to maintain a temperature between 60 and 75°C at the interface between the electrode and tissue. If the temperature at the electrode-tissue interface exceeds 100°C, plasma clots and dried electrode tissue may form, impeding effective current flow, increasing the risk of thromboembolic complications, and necessitating catheter disposal to allow removal of coagulated material from the electrode.

1.5.3.2 IMPORTANCE OF MAPPING

In the early 1990s, trans-catheter ablation was limited to accessory pathways, atrioventricular nodal reentry tachycardia, and atrial tachycardia/flutter, all based on standard electrophysiological mapping systems. Subsequently, the expansion of ablation included atypical flutters, ventricular tachycardia, and atrial fibrillation (AF). Since conventional fluoroscopic catheter mapping has limited spatial resolution and involves prolonged fluoroscopy, a non-fluoroscopic electroanatomical mapping technique has been developed to overcome these drawbacks. As a result, the past decade has witnessed the development of an increasing number of atrial fibrillation ablation procedures, largely based on anatomical considerations, leading to the explosive development of computerized mapping for increased precision. [24] Three-dimensional (3D) mapping technology enhances our understanding of atrial fibrillation and improves the safety, efficacy, and efficiency of radiofrequency (RF) ablation by enabling the construction of a 3D geometry to guide catheter navigation and lesion placement. [25]. For successful ablation, two factors must come together: detailed mapping and the ability to navigate catheters delivering energy to specific targets. Therefore, 3D mapping systems are an important component of many atrial fibrillation ablation strategies, enabling navigation to relevant anatomical structures for atrial fibrillation ablation. This is why these systems have suddenly become crucial, being used in most EP laboratories. As the catheter moves inside the heart, 3D mapping systems continuously analyse its position and orientation and present this data to the user on the monitor of a graphical workstation, enabling fluoroscopy-free navigation. Compared to conventional mapping, EAM has been shown to reduce fluoroscopy time and radiation exposure while improving procedural success. [26]. CARTO (Biosense, Diamond Bar, CA, USA) and EnSite NavX (St. Jude Medical, Saint Paul, MN, USA) are the most common EAM systems used in clinical practice. These computerized EAM systems accurately reconstruct and identify the target composition of an anticipated ablation field. Arrhythmia management has been revolutionized by the introduction of RFCA procedures. The combination of high procedural success and low complication rates has made RFCA the treatment of choice for most arrhythmias.

2. OBJECTIVES

For years, one of the primary challenges associated with cardiovascular diseases has been the risk of recurrence, i.e., the possibility of symptoms or complications recurring after a period of stability or after a specific intervention or treatment. For conditions such as arrhythmias and heart diseases, extensive and thorough studies have already been conducted. Pharmacological treatments and targeted interventions have been developed, especially in the case of ablation procedures as described previously. The most effective method would be to prevent the risk of recurrence by addressing parameters that are most significant and correlated to the problem. For this reason, many physicians are exploring various parameters that can predict whether a patient may have relapses or not, in order to act accordingly to minimize the chances of recurrence. In a previous analysis, machine learning methods were implemented to provide predictive models for recurrence. In particular, models created based on variables from the patient's medical records and variables from electroanatomical maps obtained from the CARTO 3 system were evaluated. The results highlighted high correlations with recurrence for variables related to the patient's clinical history as well as electroanatomical variables.

In this study, the main purpose is to separately analyse the variables from the clinical records and the electroanatomical variables to try to identify what additional information and correlations come from the CARTO 3 values in order to provide physicians with new assessment tools to consider for reducing the risk of recurrence. All parameters extracted from the electroanatomical mapping exports were divided into 8 different quadrants indicative of specific areas of the ventricle to analyse whether a specific part compared to another could influence recurrence.

The number of patients analysed is too low to establish scientific evidence, but a preliminary analysis has been conducted to develop a promising prototype to continue this research and increase the number of patients. The idea of this work is to use various machine learning techniques, such as logistic regression and support vector machine, to discover which features are most correlated with outcomes and how. All the results obtained were designed to be integrated with the patients' clinical records as additional values that can be used by physicians in making important decisions. The idea is also to compare the results obtained from the various methodologies used to provide a better and more accurate outcome, as well as to determine which of the proposed methodologies is most suitable for this task.

3. MATERIALS AND METHODS

3.1 HARDWARE AND SOFTWARE

For this work, a computer with an i9 processor and 64 GB of RAM was utilized. The program used to conduct various analyses was the MATLAB working environment, specifically version R2023b, both in online and desktop modes. Several toolboxes within the program were employed, including the Statistics and Machine Learning Toolbox and Experiment Manager Toolbox for the analysis part. Additionally, the Parallel Computing Toolbox was utilized for parallel processing to fully leverage all processor cores and perform calculations simultaneously. The main advantage is undoubtedly the reduction in computation times during execution. Furthermore, data from the CARTO 3 system, designated and developed by Biosense Webster, were also utilized in the project for electroanatomical mapping.

3.2 ELECTROANATOMICAL MAPS

Electroanatomical mapping, particularly when referring to the heart, is a process used to identify and track the distribution both over time and in space of electrical signals that occur during a specific cardiac rhythm. Electroanatomical mapping systems have allowed and facilitated challenging interventional ablation procedures for over a decade. Initially, their use was in arrhythmias where the ablation target is difficult to identify, such as ventricular tachycardias in structural heart diseases, atypical atrial flutters, or arrhythmias in patients with complex congenital heart defects. In recent years, electroanatomical mapping systems have also been used to guide catheter-based pulmonary vein isolation, an important component of modern atrial fibrillation (AF) management [27]. Electroanatomical mapping systems integrate three important functionalities:

- Non-fluoroscopic localization of electrophysiological catheters in a three-dimensional (3D) space;

- 3D analysis and visualization of activation sequences calculated from local electrograms or computed, and 3D visualization of electrogram voltage ("scar tissue");

- Integration of these "electroanatomical" information with non-invasive images of the heart (mainly computed tomography or magnetic resonance imaging) [28].

3.2.1 BASIC PRINCIPLES

The basic concepts of mapping are founded on three essential notions: ECG, Position, and Mapping.

ECG: The ECG subsystem supports surface ECG signals and channels for IC-ECG (intracardiac electrogram). Its function is to sample, filter, display, and record ECG signals. Additionally, it transmits ECG signals through the recording system.

Position: The system utilizes two detection technologies:

- CARTO® magnetic detection technology, which provides the position using the magnetic fields generated by the Location Pad and measured by Biosense Webster catheters equipped with an internal magnetic sensor.

- Advanced Catheter Location (ACL) detection, which provides position data for each electrode connected to the system. The position is calculated based on signals received from six patches connected to the patient. Tissue conductivity variation is calibrated using magnetic detection technology, which is not influenced by body conductivity.

Mapping: Mapping technology is used to create maps of the heart chambers for display in the Main Map Viewer. These maps are created by combining precise position data with ECG data. The system allows the use of the following mapping methods:

- Electroanatomic mapping (EA)

- Fast Anatomical Mapping (FAM) technology for rapid anatomical mapping

- Mapping with ultrasound catheter

The maps can be combined. For example, electroanatomic points can be added to a reconstruction created from ultrasound contours.

3.2.2 MAIN COMPONENTS

The Carto system consists of a low-intensity magnetic field generator composed of three coils positioned under the patient's chest (Figure 4), six skin patches, three on the back and three on the patient's chest (Figure 5), a computer for data processing, and a display.

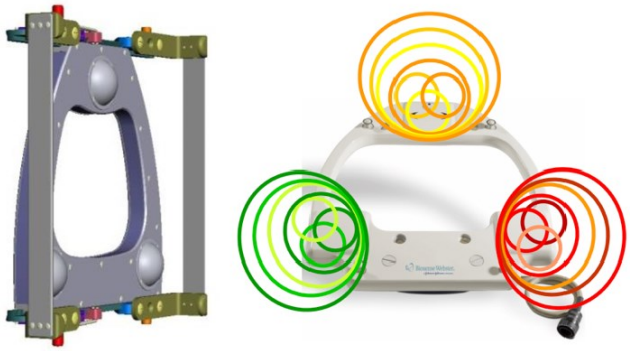

Figure 4 | *Example of location pad of the CARTO 3 System*

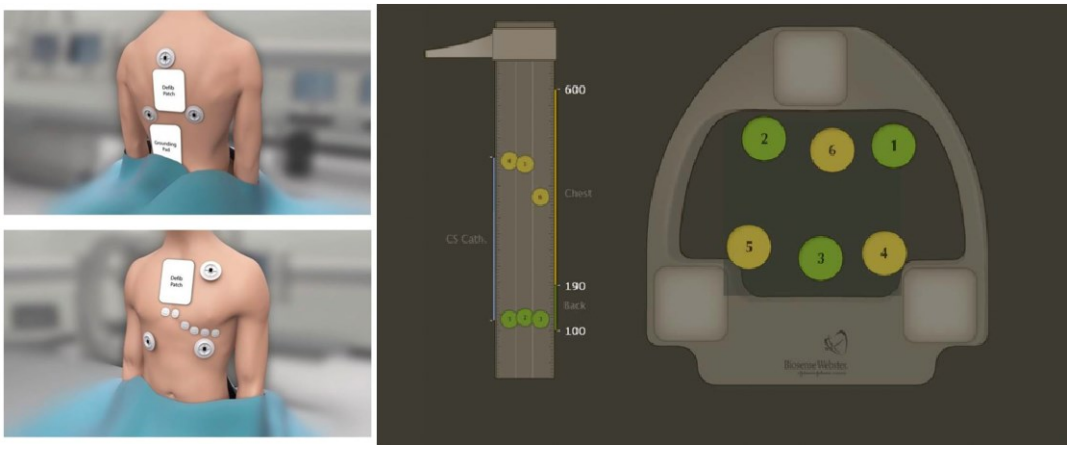

Figure 5 | *The image on the left shows the correct placement of the patches on the patient. The image on the right shows the localization of these patch within the reference system of the Location Pad*

To perform 3D electroanatomical mapping of the cardiac chambers, specialized catheters with location sensors at their tips are required. These sensors consist of spirals positioned orthogonally along the three spatial axes. The Carto system uses magnetic fields to determine the position and

orientation of the catheter and records intracavitary electrocardiograms from the sensors at the catheter tip. By collecting spatial and electrical information from different points, the system reconstructs the geometry of the cardiac chambers in real-time and analyzes arrhythmia mechanisms and ablation substrates. This process is based on the principle that metal spirals generate electric current when exposed to a magnetic field, with the current intensity depending on the intensity of the magnetic field and the orientation of the spirals (see Figure 6).

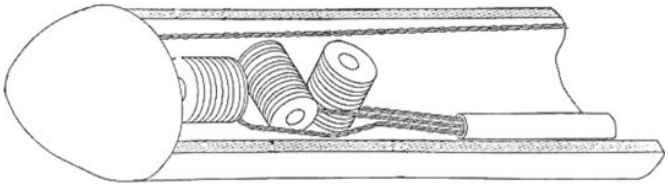

Figure 6 | *Image of the positioning and arrangement of the spirals inside the catheter used to generate a local reference system.*

The Carto system employs a triangulation algorithm similar to that used in GPS. The sensors on the catheter tip measure the current intensity in each spiral (along the x, y, and z axes), allowing the system to determine the distance between the catheter and each magnetic field source, see Figure 7.

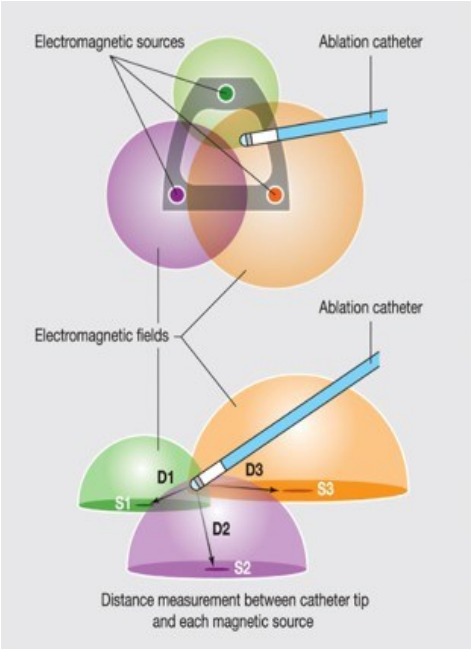

Figure 7 | *Upper and lateral views of the triangulation-based localization system*

These distances are then used to create a spherical cap representing the possible catheter position towards each source. However, the catheter can only be positioned in the area where the spheres intersect, thus determining the three-dimensional position.

The Carto system can also calculate the roll, pitch, and yaw of the catheter, in addition to the x, y, and z coordinates. Intracavitary electrocardiograms are recorded and integrated with positional information for each endocardial site reached, allowing for the creation of the activation map and cardiac geometry. To compensate for artifacts caused by cardiac and respiratory movements, the Carto system makes corrections to the map coordinates, using the surface electrocardiogram as a reference and anatomical tags. The surface electrocardiogram is synchronized with activation data recorded by the catheter during map creation. The anatomical reference, often a skin patch or a catheter fixed inside the heart, is used to correct distortions caused by patient thoracic movements as shown in Figure 8.

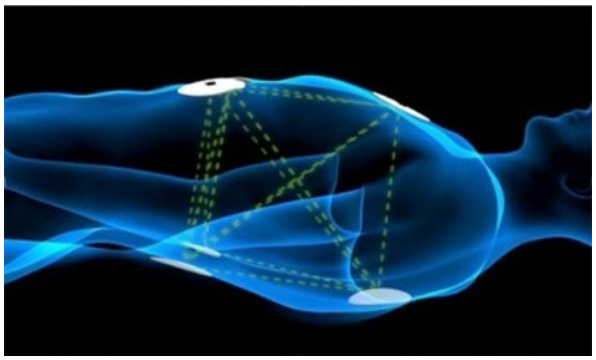

Figure 8 | *Image representing the interaction of patches to correct breathing artifacts*

Furthermore, the system requires the definition of an "area of interest," representing the time interval, relative to a reference point on the surface electrocardiogram, during which local activation occurs, either early or late compared to the reference. The total duration of the area of interest cannot exceed the duration of the cardiac cycle in the case of tachycardia. The Carto system offers the possibility of overlaying the electroanatomical map with CT or MRI images acquired before the procedure, allowing verification of anatomical landmarks, improvement of cardiac geometry, and more precise guidance during ablation. In the latest version, Carto3, two additional modules are available: the CartoUNIVU module, which allows overlaying fluoroscopic images onto the electroanatomical map in real-time, and the CartoSound module, which uses

intracardiac echocardiography as a tool for procedure monitoring and anatomical support in map creation. [29, 30]

3.2.3 EXPORT

All data recorded by the system is collected in an export file containing all the information. This export includes various details for each contact point, meaning every point on the inner heart wall touched by the catheter. Among the many details we have, such as point indices, catheter IDs, the most relevant point-to-point values include:

- Coordinate Position, specifically three coordinates, one for each of the previously described imaginary axes. These are, of course, useful for determining the point's position.

- Angular Coordinate, i.e., the three angles that determine the catheter's orientation relative to the fixed reference system at the time of signal acquisition. These specific pieces of information are not relevant for this project.

- Unipolar Voltage Value uses a single electrode to record electrical activity at a specific point inside the heart. This type of recording measures the amplitude (intensity) of the electrical signal at that point.

- Bipolar Voltage Value, on the other hand, uses two electrodes positioned at a certain distance from each other to record electrical activity between them. This recording provides information about the direction and sequence of propagation of the electrical impulse between the two electrodes.

- Local Activation Time (LAT) value refers to the precise moment when heart cells at a specific point in the heart activate during the cardiac cycle. LAT represents the time elapsed from the beginning of the cardiac cycle or from the onset of the electrocardiographic (ECG) wave to the moment when heart cells in a particular region begin to contract in response to the electrical impulse. In other words, LAT indicates when the electrical impulse reaches that specific point and begins to trigger the contraction of heart cells in that area. This can help determine if a point or an entire region is delayed or early in its activation. With a LAT map, the system color-codes activation time data for each acquired point and overlays this information onto the anatomical geometry. For example, red indicates sites activated early, blue and purple indicate areas activated late, and yellow and green areas indicate intermediate activation times.

- Impedance value is expressed in Ohms and refers to the voltage difference between two points in the electrical circuit (i.e., between the electrodes). This voltage difference is influenced by the resistance of the cardiac tissues. The higher the impedance, the greater the electrical resistance offered by the tissues. This is because when an electrical impulse spreads through the cardiac tissues, it encounters a certain resistance, represented by impedance. This resistance affects the shape and amplitude of the electrical signals recorded by the electrodes. Impedance is primarily used to assess the quality of contact between electrodes or catheters and cardiac tissues. A significant change in impedance could indicate a contact problem or a non-ideal electrode position, which could affect the accuracy of electrical activity measurements.

The export is downloaded as a digital file, available in various formats, including the most commonly used ones, such as text files (.txt) and Excel files (.xlsx).

3.2.4 EXTRACTION ALGORITMH

In this study, one of the main purposes is to extract from the data provided by the system additional parameters that are not directly calculated but can be significant and representative of some pathology or symptom of a possible future relapse. In particular, we have calculated, starting from the data provided by the exports:

- Gradient (ms): The value in milliseconds of the difference between the area with the greatest delay and the area with the greatest advancement relative to their distance.

- Voltage: Point-to-point difference of the bipolar potential with unipolar potential with possible identification of more organized regions, where multiple points with high differences are grouped in the same area.

- Impedance: the average of the Impedance values.

To calculate these new parameters, a specific algorithm has been created. The main problem with these mapping systems is that only some points are measured from the entire image displayed on the monitor, while other intermediate values are estimated and calculated by the system and therefore are not present in the exports. Then there is the problem that each patient's heart is of different sizes and positioned with a slight difference in orientation compared to another. It was therefore decided to normalize these values by projecting them onto a sphere. More precisely, a sphere of fixed dimensions is constructed, the entire surface of the sphere is precisely divided into

14400 faces of equal area, the centroid of the cloud of points resulting from the measurements made is calculated, the centroid is positioned at the center of the sphere, and all points are projected from the centroid to the inner surface of this sphere. Each face of the surface of the sphere can have only two values: null value if the face has not been hit by any projection, otherwise, the face takes on the values corresponding to the projected point. The number of faces in which to divide the sphere was chosen following many tests, and that value was selected for which no more than one projected point corresponded to each face, this in order not to reduce the accuracy of the analysis and not have to make any kind of approximation. The main advantages of this projection are two: one referred to the position problem described above, and the other to the possibility of calculating the required values. The first is because the coordinates of each face are known a priori and the dimensions of both the sphere and the faces are fixed, and therefore are the same for each measurement. This allows us to compare all patient mappings with each other, something that would not have been possible without normalization in terms of space and orientation. The second is that without an algorithm like this, the export would be difficult to understand and therefore unusable. Many values are reported in the export for each point, but being able to interpret the distances and differences between the values simply by reading them is too difficult and not at all enjoyable. Instead, thanks to this algorithm, we can automatically extract not only the individual point-to-point values but also the values, as in the case of the required parameters, which are calculated as iterations between values of different points. For the evaluation of the zones with the greatest delay and with the greatest advancement, the LAT value is used. Values below a certain threshold (last 20% between maximum and minimum values) are considered late, and adjacent delay points are part of the same delay area. This allows calculating the extension and position of the late potentials as a percentage of the non-late ones. Following the same criteria, the positions and extensions of the areas with early potentials are calculated. By applying a similar process, the amplitudes of the point-to-point potentials, both bipolar and unipolar, are calculated. The difference is calculated for each individual point, and those with values above a certain threshold (last 20% between maximum and minimum values) are considered critical. In this way, it is possible to evaluate the presence of regions where multiple points with significant differences, i.e., critical values, are grouped in the same area. Below, in Figure 9, two images resulting from the execution of the designed program are shown. In addition to the numerical results useful for the continuation of this project, it was thought that the possibility of visually seeing what is happening would also be useful for the clinician.

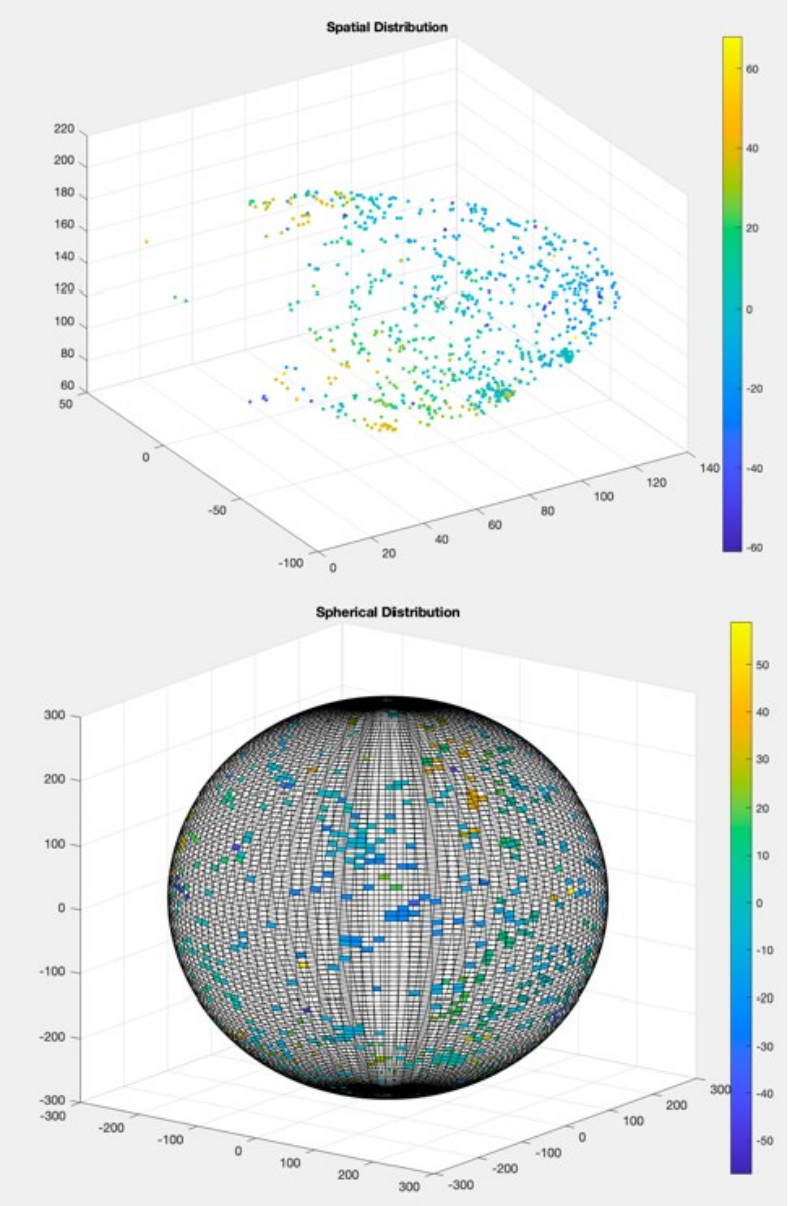

Figure 9 | *Two images resulting from the extraction algorithm. The image at the top shows the spatial distribution of points in the left ventricle measured during an examination with the CARTO 3 system. The image at the bottom represents the spherical distribution of the same points.*

An additional algorithm has been further implemented to divide the sphere into eight different quadrants, as can be seen in Figure 10, each representing a specific position of the patient's ventricle. In this way, for each parameter, we have eight different values associated with the specific zone where they are measured.

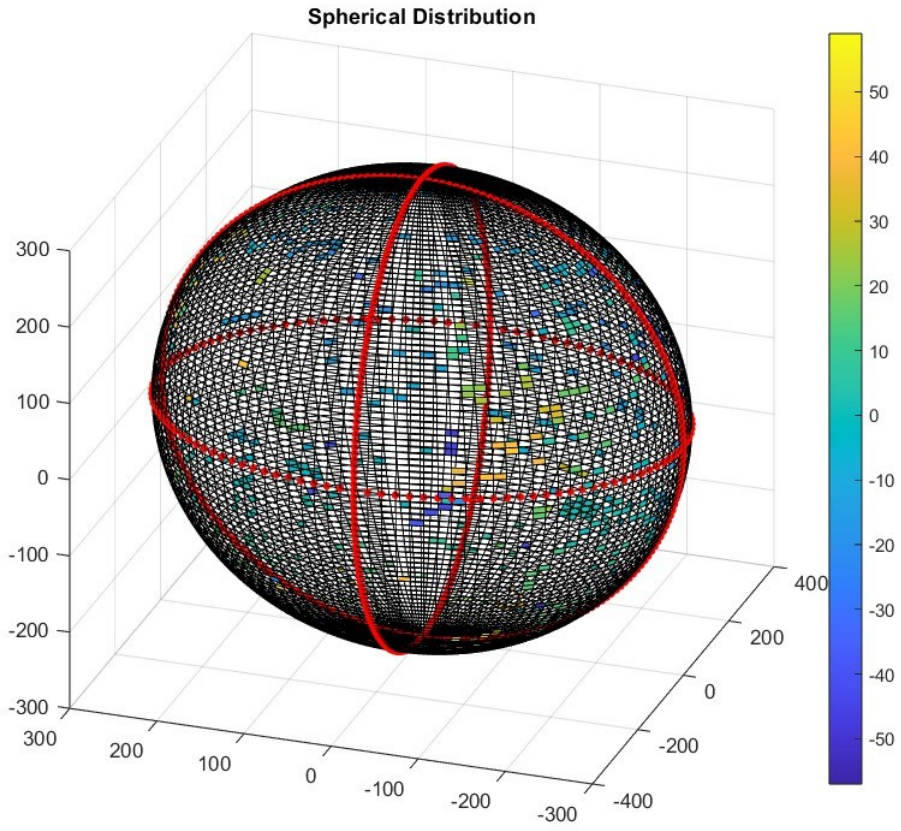

Figure 10 | *Images resulting from the extraction algorithm. The image shows the division into eight quadrants of the sphere where the points of the left ventricle measured during an examination with the CARTO 3 system are projected.*

3.3 STUDY COHORT

Retrospectively, data from consecutive patients referred to the Department of Clinical Cardiology and Arrhythmology Department of the Univeristy Hospital of the Ospedali Riuniti di Ancona were collected from September 2018 to July 2023. Patients for this study were selected using various inclusion and exclusion criteria. The primary inclusion criterion is the presence of any cardiac pathology, ischemic or non-ischemic, in the clinical record. Additionally, among all patients in the department, those who underwent electroanatomical mapping, in addition to the traditional clinical tests we will discuss in detail later, were selected for this study. To be included in the study cohort, the patient must have undergone follow-up, as otherwise predicting the outcome would be impossible. Regarding the exclusion criteria, relevance was fundamental; characteristics in the entire clinical record that, according to the literature, are not indicative for predicting cardiac diseases were disregarded and not included in the database. Characteristics

with an excessive number of missing values, i.e., parameters not common to all patients, were excluded. Features with non-heterogeneous binary logical values were also eliminated because it is pointless to use a parameter that always contains the same value as input for our model; they would have been parameters without any statistical significance.

3.4 CLINICAL TESTS

All patients in the database, in addition to the medical history conducted by the physician, underwent several clinical tests from which specific values considered risk factors for this field of study were extracted. The clinical examinations in question are echocardiogram, electrocardiogram, magnetic resonance imaging, electroanatomical mapping, and blood tests. Naturally, a follow-up was then conducted for each patient to ascertain those who had a relapse and those who did not**.**

3.5 DATA PREPARATION

The input values for our models, regardless of the machine learning technique used, are always the same. In particular, we have divided the inputs from the clinical records and the inputs from the data extracted from the mapping system. Regarding the first set of data, they consist of numerical or binary values, depending on the type of variable, and they refer to one or more specific parameters emerged from the medical history, follow-up, or clinical tests. The numerical values are expressed in percentage or in a specific muscle unit, while binary values 0 and 1 often represent the logical value "yes/no," indicating whether that parameter is present or not in the patient. Binary values also express membership in two different classes, such as males and females in the "gender" parameter. When there are more than two classes, the variable represents a value between one and the number of classes to indicate membership. According to the literature, there are various parameters that can be indicative of some type of heart disease, which is why only a few values considered valid for the study have been chosen from the patient's entire clinical record. Specifically, the values in question refer to: risk factors and pathological conditions, symptoms, characteristics of heart diseases, echocardiogram, electrocardiogram, magnetic resonance imaging, electroanatomical mapping, reason for intervention, ablation intervention, and blood tests. As for the second set of data, they are all numerical values

representing the value of Gradient, Voltage, or Impedance in that specific area. The "recurrence" parameter in the database is selected as the expected value, aiming to predict it accurately. The value we want as the output of our models is logical, where the value 1 corresponds to the patient who had a relapse and the value 0 to the patient who did not. Specifically, initially, 223 features were provided for each patient. Following careful analysis and after applying inclusion and exclusion criteria, 166 features were discarded, and 57 were then considered for analysis. One feature, namely "Recurrence," is the variable to be predicted. The other 60 features used for the study are summarized in Table 2. The "Variables" column shows the names or acronyms of the 57 features used for this study. For better understanding, all acronyms or abbreviations are written out in full and listed below in Table 3. The variable type is indicated in the "Type" column. The letters V, B, and C correspond respectively to: Value, to indicate if the variable takes real values, Binary, where values 0 and 1 generally imply the presence or absence of that parameter, and Class, in case the variable takes values for a certain range. The last three columns report the mean values for V-type variables, the count of parameter presence in the case of B-type variables, and the most present class number in the case of C-type variables. In the "Total" row, these values are reported for all 220 patients, in the "No" row, the values refer to patients who do not relapse, and in the "Yes" row, those refer to patients who relapse.

Table 2. | *The variables used for the study are reported, with an indication of the variable type: N for numerical value, B for binary, C for categorical. Average values are also provided, namely the count of the presence or absence of the variable for all subjects and for those who had a relapse and those who did not.*

Variables	**Type**	**Total (n=220)**	**Recurrence**	
			No (n=175)	**Yes (n=45)**
Age (years)	V	55	54	62
Sex	B	M=180/F=40	M=137/F=38	M=43/F=2
BMI	V	0,0026	0,0026	0,0028
Hypertension	B	87	62	25
Diabetes mellitus	B	15	10	5
Smoking	B	69	49	20
Family history of MCI	B	14	10	4
OSAS	B	10	7	3
BPCO	B	8	5	3
Vascular disease	B	34	22	12
Prior TIA/STROKE	B	14	9	5
Previous angioplasty	B	38	28	10
Bypass surgery	B	13	7	6

FA	B	35	21	14
HF	B	121	80	41
NYHA Class	C	0	0	2
HFpEF	B	62	44	18
HFmEF	B	16	16	0
HFrEF	B	59	34	25
COVID19	B	25	22	3
Anemia	B	7	5	2
Palpitations	B	67	55	12
Dyspnea	B	41	32	9
Lightheadedness	B	23	16	7
Syncope	B	13	9	4
Chest pain	B	28	22	6
Fatigue	B	16	9	7
Dilated cardiomyopathy	B	41	30	11
Ischemic cardiomyopathy	B	46	30	16
Myocarditis	B	20	15	5
Valvular cardiomyopathy	B	24	19	5
VT Idiopathic	B	61	59	2
FE (%)	V	49,12	50,98	41,91
LAV (ml/m2)	V	33,71	32,66	39,04
RVD (mm)	V	37,48	37,40	37,92
TAPSE (mm)	V	22,48	22,69	21,62
Mitral valve insufficiency	B	174	137	37
Tricuspid valve insufficiency	B	160	130	30
Aortic valve insufficiency	B	54	43	11
PAPs (mmHg)	V	28,55	27,87	31,12
LV aneurysm	B	22	15	7
Rhythm	B	39	24	15
T-wave inversion	B	36	33	3
LVEDV (ml/m2)	V	92,58	92,40	93,80
LGE	B	107	97	10
BEV	B	118	112	6
Arrhythmic storm	B	49	23	26
TV Paroxysmal	B	74	57	17
Ablation (Yes/No)	B	155	113	42
Presence of late potentials	B	94	61	33
Substrate ablation	B	139	102	37
Bipolar endocardiacal low-voltage area	B	76	55	21
Bipolar endocardiacal scar area	B	59	39	20
Inducibility (Yes/No)	B	13	9	4
HB (g/dl)	V	13,90	13,97	13,63
RDW	V	13,37	13,24	13,88
Blood glucose	V	98,87	93,82	118,69
Creatinine	V	1,01	0,97	1,16

Table 3. | *The full meaning of the acronyms or abbreviations used to indicate the considered variables is provided. In cases where the word 'or' is used between two acronyms, it is because some variables are presented in the Italian language*

BMI	Body Mass Index
MCI	Unexpected Cardiac Death
OSAS	Obstructive Sleep Apnea Syndrome
BPCO or COPD	Chronic Obstructive Pulmonary Disease
Prior TIA/STROKE	Preceding Transient Ischemic Attack or Stroke
FA or AF	Atrial Fibrillation
HF	Heart Failure
Classe NYHA	New York Heart Association functional classification
HFpEF	Heart Failure with Preserved Ejection Fraction
HFmEF	Heart Failure with Mid-Range Ejection Fraction
HFrEF	Heart Failure with Mid-Range Ejection Fraction
FE o EF (%)	Ejection Fraction
LAV (ml/m2)	Left Atrial Volume
RVD (mm)	Right Ventricular Diameter
TAPSE (mm)	Tricuspid Annular Plane Systolic Excursion
PAPs (mmHg)	Pulmonary Artery Pressure
LV	refered to the Left Ventricol
LVEDV (ml/m2)	Left Ventricular End-Diastolic Volume
LGE	Late Gadolinium Enhancement
BEV	Ventricular Ectopic Beats
TV or VT	Ventricular Tachycardia
HB (g/dl)	concentration of Hemoglobin in the Blood
RDW	Red cell Distribution Width

3.5.1 FEATURE SELECTION

Before beginning the various machine learning analyses, a feature selection algorithm was implemented to assess whether the different variables from the patient's clinical records and those extracted from the CARTO exports are significant for the analysis to be conducted. For both sets of variables, an analysis using the p-value was performed. P-value, also known as probability value or asymptotic significance, is a probability value associated with a given statistical model. It represents the probability that, if the null hypothesis is true, a set of statistical observations, commonly referred to as the statistical summary, would be greater than or equal in magnitude to the observed results.

The results are reported in Table 4 and 5.

Table 4 | *Feature selection of the variables of CARTO*

Feature	P Value
GR1	0,012
GR2	0,040
GR3	0,028
GR4	0,036
GR5	0,025
GR6	0,035
GR7	0,047
GR8	0,036
VLT1	0,046
VLT2	0,038
VLT3	0,043
VLT4	0,049
VLT5	0,042
VLT6	0,046
VLT7	0,035
VLT8	0,044
IMP1	0,014
IMP2	0,032
IMP3	0,020
IMP4	0,042
IMP5	0,033
IMP6	0,047
IMP7	0,040
IMP8	0,034

Table 5 | *Feature selection of the variables of medical records*

Feature	Pvalue
Sex	0,022
Diabetes mellitus	0,037
BPCO	0,018
Vascular disease	0,014
Bypass surgery	0,025
FA	0,040
HF	0,038
NYHA Class	0,034
HFmEF	0,043
Anemia	0,039
Dilated cardiomyopathy	0,045
Valvular cardiomyopathy	0,070
VT Idiopathic	0,039
Mitral valve insufficiency	0,044
Tricuspid valve insufficiency	0,028
PAPs(mmHg)	0,041
Rhythm	0,012
T-wave inversion	0,032
LVEDV (ml/m2)	0,019
LGE	0,040
Ablation (Yes/No)	0,032
Presence of late potentials	0,034
Bipolar endocardiacal scar area	0,045
Hypertension	0,029

Regarding the variables from CARTO, the p-values have shown values below 0.05 for all, indicating that all are considered significant. For the variables from medical records, the top twenty-four with p-values below 0.05 were selected to match the same number of features.

3.5.2 MULTICOLLINEARITY

Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a regression model. Multicollinearity can lead to skewed or misleading results when we try to determine how well each independent variable can be used most effectively to predict a dependent variable. In general, multicollinearity can lead to wider confidence intervals that produce less reliable probabilities in terms of the effect of independent variables in a model.

We can represent the multicollinearity results by heatmaps. If features have low correlation, we maintain both features, if they have high correlation instead, we maintain the feature that have the highest correlation with the gold standard. In table 6 are reported the result of this analysis that give us the certainty that all the variables are independent one from the other so we can maintain all of them.

Table 6 | *Multicollinearity of Carto's variables*

Multicollinearity																							
1	0,33	0,30	0,32	0,26	0,25	0,46	0,12	0,03	0,12	0,08	0,06	0,11	0,10	0,10	0,08	0,07	0,11	0,07	0,03	0,08	0,01	0,13	0,23
0,33	1,00	0,32	0,26	0,34	0,26	0,18	0,28	0,07	0,08	0,05	0,06	0,03	0,03	0,07	0,01	0,14	0,14	0,16	0,09	0,12	0,17	0,16	0,20
0,3	0,32	1,00	0,20	0,21	0,12	0,18	0,03	0,06	0,08	0,11	0,08	0,06	0,05	0,04	0,05	0,10	0,02	0,08	0,01	0,09	0,05	0,10	0,13
0,32	0,26	0,20	1,00	0,20	0,26	0,19	0,07	0,01	0,05	0,04	0,01	0,05	0,03	0,02	0,04	0,06	0,04	0,20	0,17	0,02	0,17	0,12	0,09
0,26	0,34	0,21	0,20	1,00	0,19	0,29	0,29	0,30	0,03	0,09	0,03	0,09	0,04	0,05	0,07	0,07	0,11	0,12	0,13	0,02	0,06	0,08	0,13
0,25	0,26	0,12	0,26	0,19	1,00	0,25	0,23	0,08	0,15	0,15	0,07	0,06	0,04	0,03	0,06	0,04	0,20	0,14	0,17	0,10	0,03	0,15	0,14
0,46	0,18	0,18	0,19	0,29	0,25	1,00	0,22	0,03	0,07	0,09	0,11	0,09	0,09	0,09	0,04	0,05	0,10	0,14	0,09	0,10	0,04	0,21	0,15
0,12	0,28	0,03	0,07	0,29	0,23	0,22	1,00	0,12	0,11	0,02	0,06	0,07	0,09	0,05	0,02	0,03	0,17	0,07	0,09	0,09	0,07	0,07	0,12
0,03	0,07	0,06	0,01	0,30	0,08	0,03	0,12	1,00	0,26	0,23	0,05	0,35	0,33	0,30	0,21	0,01	0,06	0,02	0,04	0,07	0,04	0,03	0,06
0,12	0,08	0,08	0,05	0,03	0,15	0,07	0,11	0,26	1,00	0,12	0,23	0,19	0,48	0,44	0,34	0,11	0,04	0,09	0,01	0,05	0,05	0,07	0,06
0,08	0,05	0,11	0,04	0,09	0,15	0,09	0,02	0,23	0,12	1,00	0,05	0,33	0,22	0,15	0,08	0,06	0,01	0,14	0,04	0,04	0,08	0,04	0,04
0,06	0,06	0,08	0,01	0,03	0,07	0,11	0,06	0,05	0,23	0,05	1,00	0,04	0,20	0,01	0,01	0,02	0,07	0,05	0,04	0,04	0,04	0,02	0,06
0,11	0,03	0,06	0,05	0,09	0,06	0,09	0,07	0,35	0,19	0,33	0,04	1,00	0,54	0,41	0,23	0,03	0,06	0,04	0,04	0,04	0,04	0,02	0,05
0,1	0,03	0,05	0,03	0,04	0,04	0,09	0,09	0,33	0,48	0,22	0,20	0,54	1,00	0,86	0,62	0,07	0,09	0,06	0,05	0,03	0,02	0,03	0,06
0,1	0,07	0,04	0,02	0,05	0,03	0,09	0,05	0,30	0,44	0,15	0,01	0,41	0,86	1,00	0,63	0,02	0,02	0,01	0,03	0,03	0,00	0,03	0,01
0,08	0,01	0,05	0,04	0,07	0,06	0,04	0,02	0,21	0,34	0,08	0,01	0,23	0,62	0,63	1,00	0,04	0,08	0,03	0,03	0,05	0,06	0,10	0,13
0,07	0,14	0,10	0,06	0,07	0,04	0,05	0,03	0,01	0,11	0,06	0,02	0,03	0,07	0,02	0,04	1,00	0,32	0,24	0,17	0,19	0,17	0,36	0,32
0,11	0,14	0,02	0,04	0,11	0,20	0,10	0,17	0,06	0,04	0,01	0,07	0,06	0,09	0,02	0,08	0,32	1,00	0,20	0,18	0,10	0,20	0,34	0,54
0,07	0,16	0,08	0,20	0,12	0,14	0,14	0,07	0,02	0,09	0,14	0,05	0,04	0,06	0,01	0,03	0,24	0,20	1,00	0,07	0,29	0,01	0,18	0,10
0,03	0,09	0,01	0,17	0,13	0,17	0,09	0,09	0,04	0,01	0,04	0,04	0,04	0,05	0,03	0,03	0,17	0,18	0,07	1,00	0,06	0,05	0,08	0,03
0,08	0,12	0,09	0,02	0,02	0,10	0,10	0,09	0,07	0,05	0,04	0,04	0,04	0,03	0,03	0,05	0,19	0,10	0,29	0,06	1,00	0,16	0,17	0,07
0,01	0,17	0,05	0,17	0,06	0,03	0,04	0,07	0,04	0,05	0,08	0,04	0,04	0,02	0,00	0,06	0,17	0,20	0,01	0,05	0,16	1,00	0,25	0,41
0,13	0,16	0,10	0,12	0,08	0,15	0,21	0,07	0,03	0,07	0,04	0,02	0,02	0,03	0,03	0,10	0,36	0,34	0,18	0,08	0,17	0,25	1,00	0,35
0,23	0,20	0,13	0,09	0,13	0,14	0,15	0,12	0,06	0,06	0,04	0,06	0,05	0,06	0,01	0,13	0,32	0,54	0,10	0,03	0,07	0,41	0,35	1,00

3.6 TRAINING, VALIDATION AND TESTING

The database used consists of 220 patients, with an 80% split for training and a 20% split for testing. With limited samples available, we chose to use K-fold cross-validation to evaluate the model. For validation purposes, it is standard practice to consciously partition the data into training and test sets. Data in the test set are not used for model training, allowing us to simulate real-world scenarios where the model encounters previously unseen variables. Cross-validation is a widely used technique that extends this approach. It involves repeatedly dividing the data into distinct training and test (validation) sets, known as "fold," with different combinations. This iterative process can be performed as many times as necessary to generate an average assessment of the model's performance. In doing so, it reduces the risk of overfitting the model to the data, thereby improving the model's ability to make predictions that generalize well to a broader population [31]. The number of folds is 3. The parameters used to evaluate the performance of the various models are AUC, Sensitivity, Specificity, Accuracy, and Precision. The area under the Receiver Operating Characteristic (ROC) curve (AUC) is a metric used in binary classification

tasks. The ROC curve plots the false positive rate against the true positive rate, and the AUC quantifies the area under the curve. An AUC of 0.5 corresponds to random classification, while an AUC of 1.0 indicates a model making perfect predictions. Sensitivity, also known as Recall or True Positive Rate, represents the ability of a classification model to correctly identify all positive cases in the test data. High sensitivity indicates that the model is effective at detecting the presence of the class of interest, minimizing false negatives. Sensitivity = TP / (TP + FN). Specificity measures the ability of a model to correctly identify negative cases. It expresses the percentage of true negatives relative to the total negative cases and indicates how accurately the model can distinguish examples that do not belong to the class of interest. Specificity = TN / (TN + FP). Precision is a measure of the fraction of positive instances correctly identified by the model relative to the total number of instances identified as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives. Accuracy focuses on the accuracy of positive predictions. Precision = TP / (TP + FP). Accuracy, on the other hand, is a general measure of the overall correctness of the model. It represents the fraction of all correct estimates relative to the total number of estimates. Accuracy evaluates the overall correctness of the model, considering both positive and negative predictions. Accuracy = (TP + TN) / (TP + TN + FP + FN). [32]

3.7 MACHINE LEARNING METHODS

Artificial intelligence (AI) in the medical field has been gaining momentum in recent years, introducing new methods and technologies capable of revolutionizing medicine. Over the past 30 years, there have been technological advancements that could make this a reality, including exponential increases in computing power, big data processing technologies, access to large clinical data sets using electronic health records, and machine learning (ML) [33]. In the field of medicine, ML has the potential to improve the accuracy of diagnostic algorithms and personalize patient treatment. The fundamental concept of ML is to employ algorithms that acquire input data, apply computer analysis to predict output values within an acceptable range of accuracy, discern patterns and trends within the data, and ultimately learn from experience. Although ML is not a new concept and has been around since the advent of modern computing, the idea of a thinking machine has been proposed to harness the computational capacity of computers to discover patterns and draw conclusions that may be difficult to reach through conventional statistical methods. These traditional methods often rely on human operators to formulate and provide a

basis of rules or hypotheses regarding correlations for further computer analysis [34]. ML relies on statistical foundations or incorporates them to support its operation [35].

3.7.1 REGRESSION

Linear regression is arguably the simplest ML algorithm. The central concept in regression analysis is to establish a connection between one or more numeric features and a single numeric target. Linear regression is an analytical method employed to address regression problems by employing a straight line to characterize a dataset. In the case of univariate linear regression, which focuses on predicting a target value using just a single feature, it can be represented in a slope-intercept form:

$$Y = \beta_0 + \beta_1 X$$

In this representation, β_1 serves as the slope weight, describing how much the line rises on the y-axis for each increment in x. The intercept, β_0, indicates the point where the line intersects the y-axis. [36, 37] Linear regression models a dataset using this slope-intercept form, with the machine's task being to ascertain values of a and b that enable the determined line to best correlate the provided x values with the y values. To be more precise:

$$\beta_1 = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2} \quad \text{and} \quad \beta_0 = \bar{Y} - \beta_1 \bar{X}$$

Where X, Y are the detected values and $\bar{X}$,$\bar{Y}$ are the theoretical values. Multiple linear regression is similar; however, there are multiple weights in the algorithm, each describing to what degree each feature influences the target. Basically, there is rarely a single function that fits a dataset perfectly. To measure the error associated with a fit, the residuals are measured. Conceptually, residuals are the vertical distances between predicted values, $\bar{Y}$ and actual values, Y. For multiple linear regression the model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Logistic regression is a classification algorithm where the goal is to find a relationship between features and the probability of a particular outcome. Instead of employing the straight line generated by linear regression to estimate class probability, logistic regression uses a sigmoidal curve to estimate class probability. This curve is determined by the sigmoid function:

$$y = \frac{1}{1 + e^{-x}}$$

which produces an S-shaped curve that transforms discrete or continuous numeric features (x) into a single numerical value (y) between 0 and 1. The key advantage of this approach is that probabilities are bounded within the range of 0 and 1 (i.e., probabilities cannot be negative or exceed 1). Logistic regression can be either binomial, where there are only two possible outcomes, or multinomial, where there can be three or more possible outcomes. [33, 34] In statistics, the logistic model (or logit model) is a statistical model that shapes the probability of an event taking place by expressing the log-odds for the event as a linear combination of one or more independent variables. In regression analysis, logistic regression (or logit regression) entails estimating the parameters of a logistic model, which are the coefficients in the linear combination. Formally, in binary logistic regression, there is a single binary dependent variable, coded using an indicator variable, where the two values are labeled "0" and "1," while the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The logistic function is therefore represented as follows:

$$logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Notice that the right hand side of the equation above looks like the multiple linear regression equation. However, the technique to estimate the regression coefficients in a logistic regression model is different from that used to estimate the regression coefficients in a multiple linear regression model. In logistic regression the coefficients derived from the model (e.g., β_1) indicate the change in the expected log odds relative to a one unit change in X_1, holding all other predictors constant. Defined p as the probability, the multiple logistic regression model can be written as follows:

$$\text{p=} \frac{e^{\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn}}{1 + e^{\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn}}$$

p is the expected probability that the outcome is present; X_1 through X_n are distinct independent variables; and β_1 through β_n are the regression coefficients. In this thesis work, the first machine learning method was chosen to use a model based on multinomial logistic regression. The choice of a logistic regression was made because the variable to be predicted is binary, so the outputs we expect can only be 0 or 1. Multinomial because the parameters used as predictors are multiple. In

MATLAB, in the function used for the regression, it was specified that the model was linear and that the distribution was binomial. Several models equal to all the combinations of five variables were proposed, for each combination the test performance was calculated to see which was the best.

3.7.2 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machines (SVMs) are one of the cornerstones of machine learning and are particularly powerful for binary classification. In this chapter, we will delve into the theory behind SVMs, including the details of how to find the optimal hyperplane and the optimization problem. Furthermore, we will explore the use of nonlinear kernels, including the polynomial and sigmoid kernels. Support Vector Machines (SVMs) are a machine learning model used for both classification and regression tasks. The primary goal of an SVM is to find an optimal hyperplane in a multidimensional space that can effectively separate different data classes. This hyperplane is chosen to maximize the margin between classes, which is the distance between the hyperplane and the nearest data points from each class, referred to as "support vectors". The equation of a hyperplane in an N-dimensional space is given by:

$$x^T\beta + \beta_0 = 0$$

Where:

- β is a weight vector that determines the orientation of the hyperplane.
- x is an input vector.
- β_0 is the bias term that regulates the position of the hyperplane relative to the origin.

In a Figure 11, we can see how this technique works. The decision boundary is the central line, while the two lateral lines bound the shaded maximal margin of width $2M = 2/||\beta$ [33, 34] The points labeled ξ_{1j^*} are on the wrong side of their margin by an amount $\xi_{j^*} = M\xi_j$; points on the correct side have $\xi_{1^*} = 0$. The margin is maximized subject to a total budget $\xi_i \leq$ constant. Hence, ξ_{j^*} is the total distance of points on the wrong side of their margin.

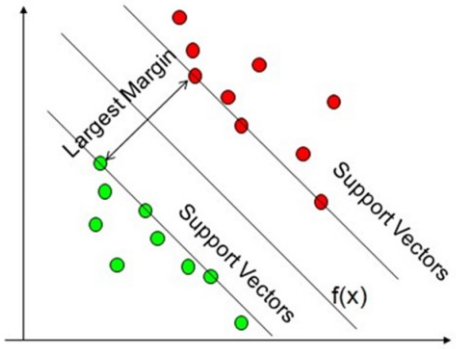

Figure 11. | *Explanatory image of the theory behind the Support Vector Machine method in a linear case.*

Consider a p-dimensional real-valued space (e.g., $\mathbb{R}p$). An optimal separating hyperplane is essentially an p-1 dimensional affine space residing within the larger p-dimensional space. For p=2, this affine space is simply a one-dimensional line, while for p=3, it is a two-dimensional plane. For higher dimensions, this affine space is known as a hyperplane. This is certainly challenging (if not impossible) to visualize, but it can be conceptually grasped. Note that "affine" refers to a hyperplane that doesn't necessarily pass through the origin (or the zero element) of the larger space. If we consider elements in the p-dimensional space, that is, $x = (X_1, \dots, X_2) \in \mathbb{R}p$, such an affine hyperplane p-1 dimensional is defined by the following equation:

$$\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = 0 \qquad \text{or} \qquad \beta_0 + \sum_{j=1}^{p} \beta_j X_j = 0$$

We can construct a maximum-margin hyperplane (MMH), which is the separation hyperplane that is farthest from any training observations. First, you compute the perpendicular distance from each training observation xi to a given separation hyperplane. The closest perpendicular distance from a training observation to the hyperplane is known as the margin. MMH is the separation hyperplane where the margin is the largest. This ensures that it is the farthest minimum distance from any training observation. The classification procedure is then simply a matter of determining which side a test observation falls on. Such a classifier is known as a maximum-margin classifier (MMC). We hope that a wide margin on the training observations also leads to a wide margin on test observations and therefore provides a good classification rate. However, note that we must be cautious to avoid overfitting when the number of feature dimensions is high. In this case, overfitting means that the MMH fits the training data very well but can perform quite poorly when exposed to test data. One of the key features of MMC (and subsequently of SVM) is that the position of the MMH depends solely on the support

vectors, which are the training observations that lie directly on the margin boundary, but not on the hyperplane. Another example in Figure 12.

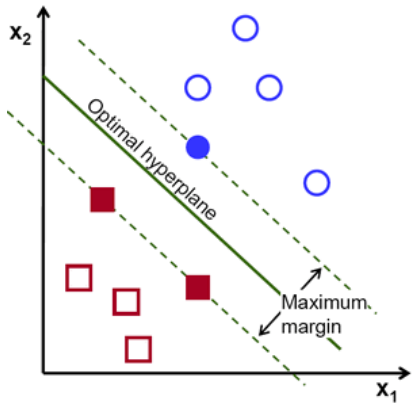

Figure 12. | *Image related to explaining the search for the best hyperplane.*

This means that the position of the MMH does NOT depend on other training observations. The MMH is the solution to the following optimization procedure:

$$\max_{\beta, \beta_0} M$$

$$subject\ to \quad y_i(x_i^T\beta + \beta_0) \geq M, \quad i = 1, \ldots, N$$

Where:

- β is a weight vector that determines the orientation of the hyperplane.
- β_0 is the bias term that regulates the position of the hyperplane relative to the origin.
- (x_i, y_i) are the training points, with x_i representing the input vector and y_i the class label (+1 or -1).

The analysis with Support Vector Machine was conducted using specific MATLAB functions. Specifically, the classification problem was solved, and the output variable to be predicted is always of binary logical type, with only values 0 and 1. The "ClassName" is defined and used to name the output classes, which helps the algorithm understand that it is a classification. By defining this parameter as "[False True]", it helps the algorithm understand that the variable to be predicted is binary, and therefore the expected outputs will be "False" for patients who do not relapse and "True" for patients who relapse. Another parameter defined was the

"BoxConstraint", which is a positive number determining the "stiffness" of the margin. Larger "BoxConstraint" values correspond to tighter margins, making the model more sensitive to misclassification errors on training data. Conversely, smaller "BoxConstraint" values lead to wider margins, making the model more tolerant to misclassification of training data but potentially at the expense of lower generalization ability. The "Solver" parameter is another parameter that was defined, specifying the algorithm used to solve the optimization problem associated with training the SVM. The solver determines how the optimization problem underlying the search for support vectors and associated weights is solved. The two main values are: "SMO" (Sequential Minimal Optimization), that is the default algorithm. It is based on sequential minimal optimization and is particularly effective for moderate-sized problems. The sequential minimal approach divides the optimization problem into smaller subproblems, iteratively optimizing the weights associated with pairs of training examples. "ISDA" (Iterative Single Data Algorithm): this solver focuses on one training example at a time. It is useful when working with very large datasets where storing the complete kernel matrix could be prohibitive. The "KernelFunction" parameter specifies the type of function to be used, while the "KernelScale" parameter optimizes the predictors for the specific "KernelFunction". To perform a linear analysis, the "ClassName" was specified, the "KernelFunction" was defined as "Linear" to indicate linear analysis, "BoxConstraint" was set to 100 to strongly penalize classification errors, aiming to obtain a separation hyperplane that separates the classes more rigorously. The "Solver" was left at its default value.

4. RESULTS

Below are reported all the results obtained from the various analysis performed. In particular, four different tables are reported for the linear regression analysis. The first two tables refer to the variables extracted from CARTO. In the first table, the different indices used to evaluate the performance of the created model are listed, while in the second table, the various coefficients associated with each variable and the intercept are reported. Indeed, being a linear method, for each model, there are values indicating how the variables related to that model influence the predictability of the output, i.e., how they are correlated with it and whether they are positively or negatively correlated. The other two tables refer to the variables from the medical records. Specifically, they report the performance of the models created with all the variables and with the variables after feature selection. The same results are reported for the second technique used. Linear support vector machine to also assess which method produces better results.

4.1 LINEAR LOGISTIC REGRESSION

In the following table 7, the best performance values of the model are reported.

Table 7. | *The performance parameters of the best model resulting from the analyses conducted with logistic regression are reported. The performances are expressed in terms of AUC, Accuracy, Precision, Sensitivity, and Specificity.*

AUC Train	0,805
AUC Test	0,830
Accuracy	0,770
Precision	0,588
Sensibility	0,344
Specificity	0,623

Table 8, instead, shows the parameters of the equation resulting from the logistic regression analysis. In the first row, we find the intercept value, while in the other rows, the variables and their associated coefficients are listed. Table 9 and 10 shows the performance of the

analysis on the medical records.

Table 8. | *Coefficients of the polynomial resulting from logistic regression analysis.*

Intercept	-0,637
GR1	0,114
GR2	0,016
GR3	0,011
GR4	0,02
GR5	0,135
GR6	-0,066
GR7	0,222
GR8	0,009
VLT1	-1,587
VLT2	1,056
VLT3	0,882
VLT4	-10,58
VLT5	-0,523
VLT6	-1,902
VLT7	-1,018
VLT8	4,195
IMP1	-0,007
IMP2	-0,008
IMP3	-0,547
IMP4	-0,009
IMP5	-0,393
IMP6	0,014
IMP7	0,001
IMP8	-0,009

Table 9. | *Performance of the model with 57 feature*

AUC Train	0,705
AUC Test	0,720
Accuracy	0,650
Precision	0,528
Sensibility	0,324
Specificity	0,542

Table 10. | *Performance of the model with 24 feature*

AUC Train	0,745
AUC Test	0,760
Accuracy	0,670
Precision	0,541
Sensibility	0,366
Specificity	0,592

4.2 LINEAR SUPPORT VECTOR MACHINE

The best performance values of the model are reported in the following Table 11.

Table 11. | *The performance parameters of the best model resulting from the analyses conducted with linear support vector machine are reported. The performances are expressed in terms of AUC, Accuracy, Precision, Sensitivity, and Specificity.*

AUC Train	0,803
AUC Test	0,828
Accuracy	0,768
Precision	0,528
Sensibility	0,374
Specificity	0,602

For the purpose of interpreting the models, i.e., the equations related to each model, Table 12 reports the intercepts and coefficients of each equation corresponding to the model.

Table 12. | *Coefficients of the polynomial resulting from linear support vector machine analysis.*

Incercept	-1,052
GR1	0,810
GR2	0,071
GR3	-0,010
GR4	0,040
GR5	0,807
GR6	-0,313
GR7	-0,780
GR8	0,247
VLT1	-1,006
VLT2	-0,057
VLT3	0,049
VLT4	-0,225

VLT5	0,339
VLT6	-0,145
VLT7	-0,431
VLT8	1,137
IMP1	0,017
IMP2	-0,127
IMP3	-0,508
IMP4	-0,567
IMP5	-0,528
IMP6	0,594
IMP7	-0,175
IMP8	-0,647

Table 13 and 14 report the performance of the model with the feature of the medical record.

Table 13. | *Performance of the model with 57 feature*

AUC Train	0,703
AUC Test	0,738
Accuracy	0,642
Precision	0,548
Sensibility	0,344
Specificity	0,572

Table 14. | *Performance of the model with 24 feature*

AUC Train	0,732
AUC Test	0,75
Accuracy	0,68
Precision	0,533
Sensibility	0,372
Specificity	0,584

5. DISCUSSION

To evaluate the performance of the different methods used and to assess the differences between the two sets of variables employed, we used various metrics including the AREA UNDER THE CURVE (AUC), Accuracy, Precision, Sensitivity, and Specificity. Each metric holds a specific meaning that allows us to understand if the algorithm is functioning correctly and yielding good results. From Table 7, we can observe the performance obtained using logistic regression in the database composed of variables extracted from electroanatomic mapping. In the first two columns, we have the AUC values for the test and training sets respectively. Significant values are achieved with an AUC greater than 0.75, indicating that the model's performance can be considered good and thus yielding important correlations. Regarding Accuracy, it measures the percentage of correct predictions out of the total predictions made by the model, with higher values indicating that the model is correctly capturing all predictions. In the case of logistic regression, we observe a value of 75%, indicating that the model can correctly predict a good portion of the database. Precision, on the other hand, measures the percentage of predictions identified as positive and are actually positive. In our case, we have a value close to 60%, indicating good correctness in predicting positive instances. Sensitivity values, however, are lower, around 35%, indicating that the model is missing many true positive instances, resulting in a high number of false negatives. Finally, the Specificity value rises to around 60%, indicating that the model mostly succeeds in minimizing false positives, i.e., in not erroneously classifying negative instances as positive. This scenario of low sensitivity can occur when the model, as in our case, does not have enough data to learn and thus is unable to effectively discriminate between classes. In conclusion, we can say that with linear regressions, we obtain simple and easily interpretable models with excellent performance.

In Table 11, instead, we can check the same performance obtained with the same database but using another machine learning method, the linear support vector machine, to compare the performance and results. Again, as we can see from Table 12, we obtain very similar values compared to the linear logistic regression method. In fact, almost all the metrics assume the same values. Therefore, we can say that both trained models already offer good performance which can be increased by increasing the number of patients.

In the tables 9 and 10, 13 and 14, we can observe the performance obtained from linear logistic regression models and support vector machine models using only clinical records. The difference between the two tables in each technique used is that in the first table, analyses were performed using all 57 variables present in the database, while in the second table, the same analyses were performed after careful feature selection as previously described. The values in the first table are low mainly due to having many features but few patients. Therefore, the model is not fully able to find correlations between the variables and the output of our interest. For this reason, we decided to reduce the number of features to 24, considering only those that according to the p-value could be more correlated to the output. In this case, we can notice that we obtain higher AUC values, indicating that the model is more likely to provide interesting correlations. However, from the obtained values of Precision, Sensitivity, and Specificity, we can say that the model manages to discretely classify true negatives but has many shortcomings in classifying true positives. This again indicates the presence of few cases compared to the number of variables.

As for Tables 8 and 12, they show the coefficients associated with variables extracted from electroanatomic mapping. In particular, we have the Gradient (GR), which represents the difference value between the zone with the greatest delay and the zone with the greatest advance relative to their distance. The Voltage (VLT) represents the point-to-point difference of the bipolar potential with the unipolar potential. The Impedance (IMP) represents the voltage difference between two points of the electrical circuit, a voltage difference influenced by the resistance of the cardiac tissues. All these variables have their values expressed in each quadrant into which we have divided the sphere, representing our ventricular reconstruction. In this way, we can evaluate not only which of the three variables has more influence, but we can also assess if a certain position is more influential than another.

From Table 8, we can make a comparison between the variables. The coefficients are almost all comparable to each other because when multiplied by the assigned variable, they have approximately the same order of magnitude. This is important because sometimes coefficients may be larger than others, but when multiplied by the values of their variable, it turns out they have a lower impact. However, we must make some specific considerations for each variable. Considering the Gradient values along the eight faces we have selected, we can notice that along faces 1, 5, and 7, we have more significant coefficients and thus more correlated to the output of our interest. Regarding the Voltage values, we notice greater weight in quadrant 1, 4, 7, and 8. Finally, concerning the Impedance values distributed in the different quadrants, we can certainly notice that in quadrants 1, 3, and 5, we obtain more significant coefficients. After this analysis,

we can say that the quadrants that appear most frequently are the first, which is present for all three variables, and the seventh quadrant, which is significant in two out of the three variables.

In Table 12, we can make a similar consideration but based on the second machine learning method used. Again, concerning the Gradient values, we can notice that we obtain greater coefficients in the 1, 5, 6, and 7 quadrants. The most significant Voltage values, on the other hand, are found in quadrants 1, 5, 7, and 8. While the most correlated Impedance coefficients are present in quadrants 1, 3, 5, and 7. Therefore, we can say that the 1, 5, and 7 quadrants are the most recurrent in all three variables.

6. CONCLUSIONS

The present thesis aimed to investigate significant variables that could support clinical decision-making to reduce or predict the risk of recurrence of cardiovascular diseases after an initial treatment or intervention. In particular, an algorithm was created to extract new parameters that are not directly calculated by the electroanatomic mapping system, and their correlation with recurrence was evaluated. In a previous study, predictive models based on a database including variables related to patients' clinical records and variables extracted from the electroanatomic mapping system were created, revealing high correlations that can be taken into consideration. In addition, in this work, we aimed not only to give weight to the new variables extracted from the electroanatomic mapping system but also to find correlations, where present, related to the position of the variable itself. A specific algorithm was created to divide the sphere in which all points mapped by catheters within the cardiac cavity are projected into eight different quadrants. Each of these quadrants is linked to a specific position. Two different machine learning methods were implemented to evaluate the performance and results obtained. From the various analyses, it emerged that the use of these new parameters, such as Gradient, representing the difference between the area with the greatest delay and the area with the greatest advance relative to their distance; Voltage, the point-to-point difference of the bipolar potential with unipolar potential with possible identification of more organized regions, where more points with high differences are grouped in the same area; and Impedance, are extremely correlated to Recurrence. Additionally, it emerged that different specific zones of the cardiac chamber have a greater impact than others. Therefore, targeted treatment in those specific zones, which are more relevant, can be considered.

The main problem is that we do not have a sufficient number of data, so the analyses can be considered preliminary. However, with an increase in the number of patients and therefore the number of observations, more interesting results will be possible. Nevertheless, it remains a promising prototype for the future that can lead to more truthful and accurate results. Another consideration is that these new parameters, despite high correlation, provide additional support but are not a substitute for clinical records, which must always be considered by physicians. We are confident that further parameters can be found to add to the recurrence prediction analyses, and many of these will be closely correlated with them, as already demonstrated in this work.

7. BIBLIOGRAPHY

1. Mitchell, L. Brent. “Panoramica sui sintomi delle malattie cardiovascolari.” Manuali MSD Edizione Professionisti, Manuali MSD, 20 Sept. 2023, https://www.msdmanuals.com/it-it/casa/disturbi-cardiaci-e-dei-vasi-sanguigni/sintomi-delle-malattie-cardiovascolari/panoramica-sui-sintomi-delle-malattie-cardiovascolari.
2. Anatomia dell’uomo, edi-ermes
3. Mitchell, L. Brent. “Panoramica Sulle Aritmie - Disturbi Dell’apparato Cardiovascolare.” Manuali MSD Edizione Professionisti, Manuali MSD, 20 Sept. 2023.
4. Arrhythmia. National Heart, Lung, and Blood Institute. https://www.nhlbi.nih.gov/health-topics/arrhythmia. June 13, 2023.
5. What is arrhythmia? American Heart Association. https://www.heart.org/en/health-topics/arrhythmia. June 13, 2023.
6. Cardiomyopathy. American Heart Association.
7. Istat, “Mortalità per Territorio Di Evento”. Mortalità per Territorio Di Evento, 2020, dati.istat.it/Index.aspx?DataSetCode=DCIS_CMORTE1_EV
8. Malattie Cardiovascolari - Salute.Gov.It,
9. The New York Times. (2014, July 23). Arrhythmias treatment.
10. Cardiac Arrhythmia Suppression Trial (CAST) Investigators. (1989). Preliminary report: Effect of encainide and flecainide on mortality in a randomized trail of arrhythmia suppression after myocardial infarction. New England Journal of Medicine, 321, 406–412.
11. Chaudhry, G., Muqtada, M. D., & Haffajee, C. I. (2000). Antiarrhythmic agents and proarrhythmia. Critical Care Medicine, 28, N158–N164.
12. Coplen, S. E., Antman, E. M., Berlin, J. A., et al. (1990). Efficacy and safety of quinidine therapy for maintenance of sinus rhythm after cardioversion. A meta-analysis of randomized control trials. Circulation, 82, 1106–1116.
13. Gregoratos, G., Abrams, J., Epstein, A. E., & American College of Cardiology/American Heart Association Task Force on Practice Guidelines/North American Society for Pacing and Electrophysiology Committee to Update the 1998 Pacemaker Guidelines: ACC/AHA/NASPE 2002 guideline update for implantation of cardiac pacemakers and antiarrhythmia devices. (1998). A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/NASPE Committee to Update the 1998 Pacemaker Guidelines). Circulation, 106, 2145–2161.

14. Zipes, D. P., Wyse, G., Friedman, P. L., et al. (1997). A comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. New England Journal of Medicine, 337, 1576–1583.
15. Mushlin, A. I., Hall, W. J., Zwanziger, J., Multicenter Automatic Defibrillator Implantation Trial, et al. (1998). The cost-effectiveness of automatic implantable cardiac defibrillators: Results of MADIT. Circulation, 97, 2129–2135.
16. Moss, A. J., Zareba, W., Hall, J., Multicenter Automatic Defibrillator Implantation Trial II Investigators, et al. (2002). Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. New England Journal of Medicine, 346, 877–888.
17. Wilkoff, B. L., Cook, J. R., Epstein, A. E., Dual Chamber and VVI Implantable Defibrillator Trial Investigators, et al. (2002). Dual-chamber pacing or ventricular backup pacing in patients with an implantable defibrillator: The Dual Chamber and VVI Implantable Defibrillator (DAVID) Trial. JAMA, 288, 3115–3123.
18. Abraham, W. T., Fisher, W. G., Smith, A. L., et al. (2002). Cardiac resynchronization in chronic heart failure. New England Journal of Medicine, 346, 1845–1853.
19. Zipes, D. P., DiMarco, J. P., Jackman, W. M., et al. (1995). Guidelines for clinical intracardiac electrophysiological and catheter ablation procedures. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Clinical Intracardiac Electrophysiologic and Catheter Ablation Procedures), developed in collaboration with the North American Society of Pacing and Electrophysiology. Journal of the American College of Cardiology, 26, 555–573.
20. Tracy, C. M., Masood, A., DiMarco, J. P., et al. (2000). ACC/AHA clinical competence statement on invasive electrophysiology studies, catheter ablation, and cardioversion. Journal of the American College of Cardiology, 36, 1725–1736.
21. Jackman, W., Wang, X., Friday, K. J., et al. (1991). Catheter ablation of accessory atrioventricular pathways (Wolff–Parkinson–White syndrome) by radiofrequency current. New England Journal of Medicine, 324, 1604–1611.
22. Wyse, D. G., Waldo, A. L., DiMarco, J. P., Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) Investigators, et al. (2002). A comparison of rate control and rhythm control in patients with atrial fibrillation. New England Journal of Medicine, 347, 1825–1833.
23. Calkins H, Sousa J, El-Atassi R, et al. Diagnosis and cure of the Wolff– Parkinson–White syndrome or paroxysmal supraventricular tachycardias during a single electrophysiologic test. N Engl J Med 1991;324:1612-8.
24. Pappone C, Santinelli V. Remote navigation and ablation of atrial fibrillation. J Cardiovasc Electrophysiol (2007) 18(Suppl. 1):S18–S20.

25. Pappone C, Vincenzo Santinelli, Non-fluoroscopic mapping as a guide for atrial ablation: current status and expectations for the future, European Heart Journal Supplements, Volume 9, Issue suppl_I, December 2007, Pages I36–I47, https://doi.org/10.1093/eurheartj/sum059
26. Samuel J. Asirvatham e Matthew J. Swale. «Imaging and Cardiac Ablation». In: JACC: Cardiovascular Imaging 4.7 (lug. 2011), pp. 727–729. issn: 1936878X. doi: 10.1016/ j.jcmg.2010.12.010. url: https://linkinghub.elsevier.com/retrieve/pii/ S1936878X11003536.
27. Ziad Issa JMM, Douglas P. Zipes. Clinical Arrhythmology and Electrophysiology: A Companion to Braunwald's Heart Disease. 1 ed. Philadelphia, PA: Saunders, an imprint of Elsevier Inc.; 2009
28. Bhakta, D., & Miller, J. M. (2008). Principi di mappatura elettroanatomica. *Giornale indiano di stimolazione ed elettrofisiologia, 8*(1), 32–50.
29. Bunch TJ, Weiss JP, Crandall BG, et al. Image integration using intracardiac ultrasound and 3D reconstruction for scar mapping and ablation of ventricular tachycardia. Journal of cardiovascular electrophysiology 2010;21:678-84.
30. Webster B. Carto 3 System Fact Sheet. https://wwwbiosensewebstercom/documents/carto3-factsheetpdf?Cache=1%2F19%2F2015+3%3A56%3A28+PM 2014.
31. Arlot, S. and Celisse, A., 2010. A survey of cross-validation procedures for model selection
32. Altman, D.G. and Bland, J.M., 1994. Diagnostic tests. 1: Sensitivity and specificity. BMJ: British Medical Journal, 308(6943), p.1552
33. Sadegh-Zadeh, Kazem. "In dubio pro aegro." Artificial Intelligence in Medicine 2, no. 1 (1990): 1-3.
34. Turing, Alan M. Computing machinery and intelligence. Springer Netherlands, 2009.
35. Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), pp.199-231.
36. James G, Witten D, Hastie T, Tibshirani R, An Introduction to Statistical Learning: With Applications in R. New York: Springer; 2013. [Google Scholar]
37. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York, NY: Springer; 2009. [Google Scholar]