Vietnamese Heart Segmentation And Cardiovascular Diseases Dataset

Nguyen Le Quoc Bao & Le Tuan Hy

March 17, 2024

Abstract

Cardiovascular disease is a leading cause of mortality worldwide. The application of Deep Learning algorithms in cardiac medicine is increasingly gaining attention in this field. Image segmentation and analysis of computed tomography images are crucial initial steps for important applications such as three-dimensional cardiac structure reconstruction, disease pre-diagnosis, and treatment planning. However, existing datasets often have one or more limitations: they may be outdated, lack comprehensive coverage of cardiac components, sourced from abroad, or lack labeling for cardiovascular diseases. To serve the Vietnamese cardiovascular patient group, the team conducted a comprehensive study of publicly available datasets worldwide, researched raw volumetric images from Toshiba Aquiline ONE scanner, and performed manual segmentation and labeling of cardiac diseases under the guidance, supervision, and verification of physicians and experts. The resulting VHSCDD dataset overcomes the aforementioned limitations, facilitating experimentation, comparison, and development of Deep Learning algorithms in this field.

1 Introduction

Cardiovascular disease is a leading cause of global mortality, according to the World Health Organization (WHO). Approximately 17.9 million people died from heart disease in 2016. This number continues to increase annually [1]. According to statistics, cardiovascular disease accounts for 31% of deaths in Vietnam, totaling over 170,000 fatalities [2]. However, 90% of heart diseases can be prevented with timely diagnosis [3, 4]. Software applications supporting analysis and prognosis are rapidly developing due to advancements in imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) [5]. Analysis, evaluation, and reconstruction of cardiac anatomy play a crucial role in early stroke prevention, preoperative planning, and disease monitoring [6, 7, 8]. Manual segmentation by imaging diagnostic experts is the gold standard in medicine [7]. However, this process is time-consuming, exhausting, prone to errors [3, 9], and sometimes infeasible in terms of human resources and costs, especially with large datasets [10]. Currently, traditional reconstruction algorithm (e.g Marching Cubes isosurface) and machine learning are being applied for semi-automatic segmentation. With the emergence of Deep Learning, its potential is believed to surpass traditional methods in image segmentation and disease analysis [1]. However, cardiovascular diseases often involve complex structural deformations, especially congenital heart diseases, leading to a scarcity of complete datasets. This limits the full potential of Deep Learning algorithms [7]. This report consists of three main sections: a comprehensive review of major cardiovascular dataset research, a presentation of the VHSCDD dataset research, and an overview of cardiovascular diseases.

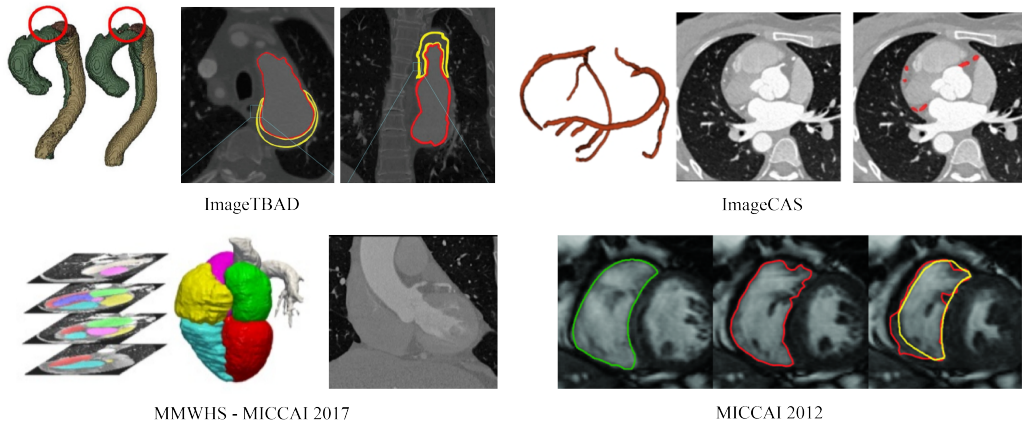

Figure 1: Current cardiovascular dataset overview

2 Researching Related Works

In the past decade, three international competitions have been organized, providing three different datasets [6, 7, 8]. All these 3 datasets focus on either the left ventricle or the right ventricle. Since these three competitions were organized before 2012, no participants utilized Deep Learning. By 2017, a fourth competition was held providing a new dataset named Multi-Modality Whole Heart Segmentation (MMWHS) labeling 7 important but basic regions of the cardiac structures (see 1 and 1). At this time, Deep Learning was newly researched and applied in this field. Additionally, there are two datasets focusing on segmenting coronary arteries and the aorta. The details of the six datasets are studied in the order below.

2.1 The Sunnybrook Cardiac MR Left Ventricle Segmentation challenge - MICCAI 2009

The MICCAI 2009 dataset is provided by the Sunnybrook Health Sciences Centre, Toronto, Canada, publicly available online, containing 45 sets of MRI slice images from various patients with different pathologies, grouped into three categories. Each category consists of 15 cases: 4 cases of localized myocardial infarction, 4 cases of non-ischemic heart failure, 4 cases of left ventricular hypertrophy, and 3 normal cases [6, 11].

2.2 The Left Ventricle Segmentation Dataset and Challenge - MICCAI STACOM 2011

The competition focuses on comparing different methods for left ventricle segmentation. The dataset is created by collecting computed tomography slice images from 200 patients with coronary artery disease and myocardial infarction (100 images for training and 100 images for testing). The slice images are obtained using Steady-State Free Precession (SSFP). Machine parameters vary case by case, creating diversity among scanner types and image parameters suitable for real-world scenarios [7].

2.3 The Right Ventricle Segmentation Dataset - MICCAI 2012

The competition aims to compare methods for segmenting the right ventricle based on a dataset consisting of 48 cardiac slice images, labeled by a cardiologist (16 images for training, 32 images for testing). From June 2008 to August 2008, patients were referred to the Rouen University Hospital center for undergoing slice image scanning and data

Name	Year	Train	Test	Region	P-Labelled
Sunnybrook	2009	45	NA	Left ventricle	Yes
STACOM	2011	100	100	Left ventricle	No
MICCAI RV	2012	16	32	Right ventricle	No
ImageTBDA	2015	100	NA	Aortic duct	Yes
MMWHS	2017	40	80	Whole heart (7 regions)	No
ImageCAS	2018	1000	NA	Coronary arteries	Yes

Table 1: Detailed analysis of current cardiovascular datasets

collection. Exclusion criteria for patient subjects were: under 18 years old, congenital heart disease, cardiac arrhythmia disorders [8].

2.4 Multi-modality Whole Heart Segmentation (MMWHS) challenge - MICCAI 2017

The dataset was collected from two modern 64-slice CT scanners (Philips Medical Systems, Netherlands) using standard CT imaging procedures in Shanghai, China. Cardiac MRI data were acquired from two hospitals in London, United Kingdom. One dataset was collected from St. Thomas' Hospital on a Philips 1.5T scanner (Philips Healthcare, Best, The Netherlands) and from Royal Brompton Hospital on a Siemens Magnetom Avanto 1.5T scanner (Siemens Medical Systems, Erlangen, Germany). The dataset consists of 120 images including 60 CT/CTA images and 60 MRI images. This dataset was manually labeled with only 7 basic regions: left/right ventricles, left/right atria, aorta, pulmonary artery, left ventricular myocardium [12].

2.5 ImageTBDA: Type-B Aortic Dissection

The dataset comprises a total of 100 3D CTA images collected from the People's Hospital of Guangdong Province from January 1, 2013, to April 23, 2015. Images were collected from two types of scanners (Siemens SOMATOM Force and Philips 256-slice Brilliance iCT systems). Segmentation labeling was performed by a team of two expert cardiovascular imaging technologists. Each image was segmented by one expert and verified by another expert. The labeling time for each image was approximately 1-1.5 hours [13].

2.6 ImageCAS: large-scale dataset and benchmark for coronary artery segmentation

The proposed dataset includes 3D CTA images captured by a Siemens 128-slice dual-source scanner from 1000 patients diagnosed with arterial diseases. Data were collected from real clinical cases at the People's Hospital of Guangdong Province from April 2012 to December 2018. The dataset only contains patients over 18 years old with a medical history of stroke, transient ischemic attack, and/or peripheral arterial disease [14].

3 VHSCDD Dataset

3.1 Survey in Vietnam

Through research conducted in Vietnam, the team has observed that segmenting the entire heart would provide more practical applications than focusing on individual parts. Cardiovascular pathologies often involve multiple adjacent regions. For example, the inter-ventricular septum (related to both left and right ventricles), arterial trunks, aortic root (related to the aortic arch and pulmonary artery), Anomalous origin of coronary

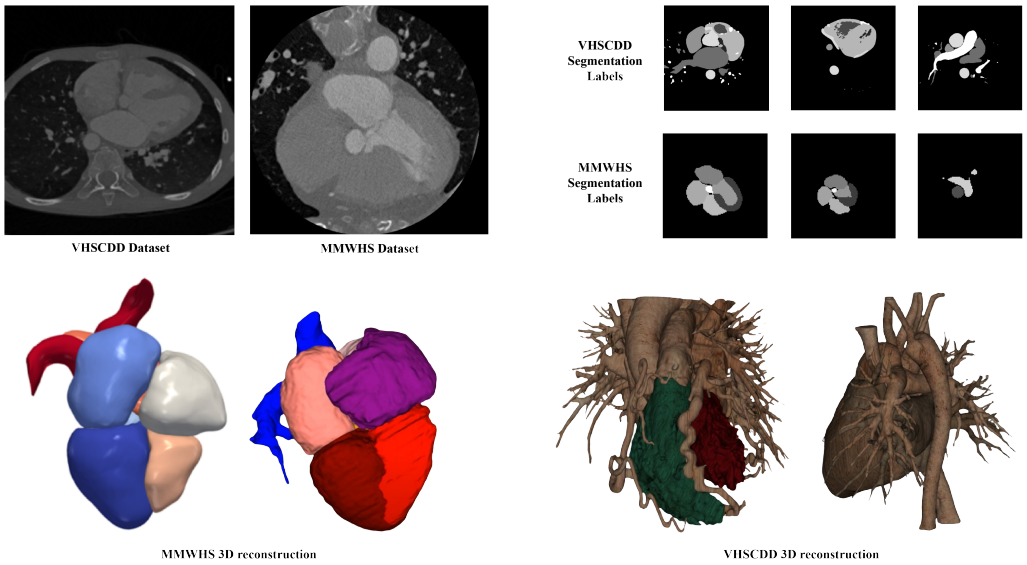

Figure 2: Comparison between MMWHS and our VHSCDD dataset

arteries (related to coronary arteries and the aortic arch). MMWHS is the most comprehensive dataset available to train Deep Learning algorithms. However, this dataset only consists of 7 basic regions: right/left ventricles, right/left atria, aorta, pulmonary artery, myocardial wall; and it does not include data from patients with cardiovascular pathologies. Particularly, the 3D reconstruction model based on this dataset lacks details down to the level of individual pulmonary veins. This deficiency renders the Deep Learning algorithm trained on this dataset lacking practical applicability, especially when applied to Vietnamese individuals with diverse pathologies.

3.2 Methodology

The raw CT/CTA slice image data was collected using the advanced Toshiba Aquilion ONE CT scanner from Cho Ray Hospital, from patients undergoing chest CT scans from May 2023 to December 2023. The acquisition procedure utilized helical CT technique, covering areas from the chest to the pelvic bone. Imaging parameters included a slice thickness of 5.0mm, peak tube voltage (KVP) of 120.0, and an exposure time of 500ms. The data files also incorporate information regarding the scan time and important patient information: patient name, patient ID, and date of birth. CT images provide anatomical details with a resolution in the plane of 0.858mm $\times$ 0.858mm and a pixel bit depth of 16. When reading the stack of slice images, a three-dimensional image volume with dimensions of 512 $\times$ 600 $\times$ 600 is obtained.

After obtaining the raw dataset, we performed the following preprocessing steps to create optimal conditions for testing Deep Learning algorithms: anonymization of data, normalization of Hounsfield density, and noise removal. Subsequently, under the guidance and supervision of two cardiologists - Dr. Nguyen Van Nghia, Dr. Le Thi Phuong Nga, and an imaging diagnostician Dr. Dang Thanh, manual segmentation was conducted using ITK-SNAP software for eleven detailed anatomical structures of the cardiovascular system: left ventricle, right ventricle, left atrium, right atrium, ascending aorta, aortic arch, vena cava, myocardium, coronary artery, and auricle. Manual segmentation required approximately 4-5 hours per set. Results were sent to physicians and experts for weekly review, and any discrepancies were promptly adjusted from September 2023 to December 2023. A total of 50 successfully manually segmented datasets (each set comprising

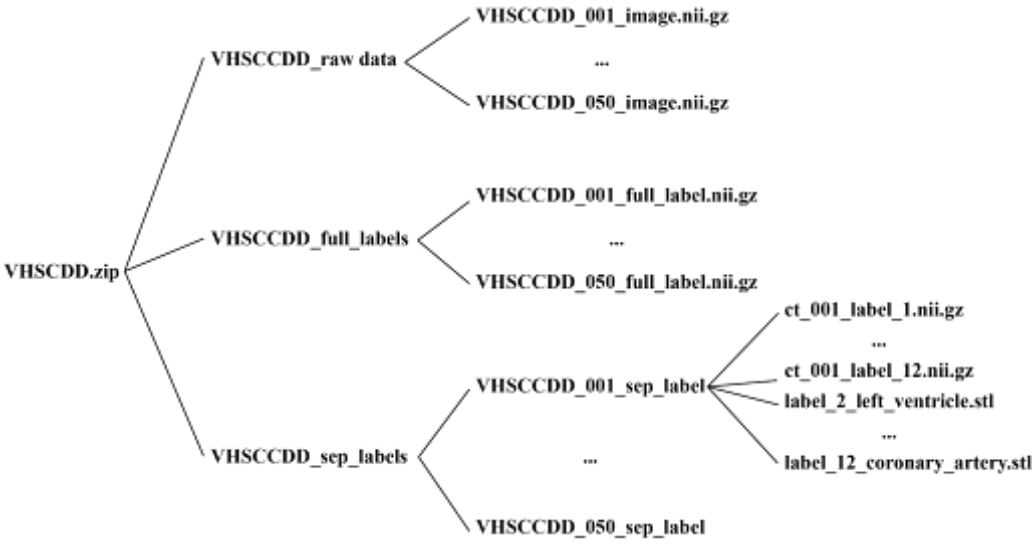

Figure 3: Data Tree Structure of VHSCDD Dataset

512 slice images of size 600×600 pixels) were approved by the physicians/experts. We employed a "3D cardiac structure reconstruction" algorithm to construct 3D models, facilitating easier observation of cardiovascular pathologies. The labeling of cardiovascular pathologies was supervised by Dr. Phuong Nga and verified by Dr. Nguyen Van Nghia. This labeling process spanned one month, from November 18, 2023, to January 9, 2023.

The VHSCDD dataset has been completed meeting criteria regarding detail level, with 11 segmented regions, providing optimal conditions for achieving the highest efficacy in three-dimensional cardiac reconstruction algorithms. Figure 2 compares two datasets, MMWHS and VHSCDD. Through observational analysis, the VHSCDD dataset demonstrates superiority as it can label even the smallest capillaries in the blood circulation and coronary arteries. This capability facilitates easier training for disease classification due to the dataset's numerous and detailed regions, thus exhibiting high practical applicability.

4 Cardiovascular Pathologies

The statistics regarding the prevalence of cardiovascular diseases in the VHSCDD dataset are presented in 2. The number of images with diseases is arranged in descending order. Due to the limited time span of actual data collection, the overall quantity remains relatively restricted for disease classification purposes. We propose a balanced training ratio between normal and diseased states for each condition to ensure objectivity in algorithmic analysis.

4.1 Ventricular Septal Defect

A ventricular septal defect (VSD) is a defect in the septum, the wall between the two ventricles of the heart. The ventricular septum is a complex structure comprising muscular, membranous, infundibular, and trabecular parts. Typically, at birth, this septum is intact, preventing the mixing of blood between the two ventricles. A large VSD causes a significant left-to-right shunt, leading to feeding difficulties, poor weight gain, cyanosis of the skin, lips, and nail beds, slow growth, rapid or labored breathing, fatigue, and swelling of the feet, ankles, or abdomen, along with rapid heartbeats. Sometimes, a ventricular septal defect may remain undetected until adulthood, manifesting symptoms of

Pathology	Positive	Negative
Ventricular Septal Defect (VSD)	18	32
Double Outlet Right Ventricle (DORV)	11	39
Patent Ductus Arteriosus (PDA)	18	32
Common Arterial Trunk (CAT)	18	32
Coronary Artery Anomalies	18	32
Arterial Hypertension	18	32
Double Superior Vena Cava (DSVC)	18	32
Transposition of Great Arteries (TGA)	18	32
Coarctation of the Aorta (CA)	18	32
Total Anomalous Pulmonary Venous Return (TAPVR)	18	32
Pulmonary Artery Sling (PAS)	18	32
Double Aortic Arch (DAA)	18	32

Table 2: Statistics of pathological labels in VHSCDD

heart failure such as dyspnea.

4.2 Double Outlet Right Ventricle

In a double outlet right ventricle (DORV), both the aorta and pulmonary artery connect to the right ventricle. Ventricular septal defects almost always accompany this anomaly, and clinical manifestations are determined by the location of the defect and whether there is pulmonary valve narrowing. In a normal heart, the aorta connects to the left ventricle below, and the pulmonary artery connects to the right ventricle below. However, in children with DORV, both the aorta and pulmonary artery may partially or entirely connect to the right ventricle. These children also have a ventricular septal defect between the two ventricles. This defect, known as a "ventricular septal defect," results in the mixing of oxygen-rich and oxygen-poor blood. Children with this condition may experience inadequate oxygenation of blood circulation. The skin may become gray or blue, indicating cyanosis. Excessive blood flow through the pulmonary artery to the lungs can lead to heart failure and an increased risk of malnutrition. Diagnosis is made through electrocardiography, imaging, and cardiac catheterization. Medical therapy is beneficial, but surgical intervention is always necessary.

4.3 Patent Ductus Arteriosus

Patent ductus arteriosus (PDA) is the persistence of the fetal structure linking the aorta and pulmonary artery after birth. Small patent ductus arteriosus may close on their own. Large PDAs, however, pose circulation problems in infants. If there are no structural heart abnormalities or increased pulmonary vascular resistance, blood flow through the ductus arteriosus will be from left to right (from the aorta to the pulmonary artery). Blood will flow directly from the aorta to the pulmonary artery, leading to increased blood flow to the pulmonary circulation and increased blood return to the left side of the heart. If a large patent ductus arteriosus remains untreated, abnormal blood flow from the major arteries within the heart, increased pressure in the heart chambers, weakening of the heart muscle, and other complications may occur. Symptoms may include slow growth, poor feeding, rapid heartbeat, and rapid breathing. A continuous murmur over the left upper chest is common. Diagnosis is made by echocardiography, CT, and MRI. The use of cyclooxygenase inhibitors (ibuprofen lysine or indomethacin) with or without fluid restriction may be attempted in premature infants with significant shunts, but this therapy is ineffective in full-term infants or older children with PDA. Surgical intervention or catheter-based closure is indicated if the patent ductus arteriosus persists.

4.4 Common Arterial Trunk

A common arterial trunk occurs when, during fetal development, the primitive arterial trunk fails to divide into the aorta and pulmonary artery, resulting in a single, large, riding trunk above the membranous or peri-membranous ventricular septum. Consequently, oxygen-rich and oxygen-poor blood mix to supply the body, lungs, and coronary arteries. This condition often accompanies a ventricular septal defect. Symptoms include cyanosis and heart failure, difficulty feeding, sweating, and rapid breathing. The first heart sound (S1) is normal, and the second heart sound (S2) is loud and single; murmurs may vary. Before deep heart failure develops, peripheral pulses will be bound due to the large flow from the ascending aorta to the pulmonary artery. Diagnosis is made by echocardiography, CT angiography, or cardiac catheterization. Medical treatment for heart failure, followed by early corrective surgery, is recommended.

4.5 Coronary Artery Anomalies

Coronary artery anomalies are one of the leading causes of death worldwide and the second most common cause of stroke in athletes. The coronary arteries have two main branches: the left coronary artery (LCA) originating from the left sinus of Valsalva and the right coronary artery (RCA) originating from the right sinus of Valsalva. Any misplacement of these coronary artery branches is considered a coronary artery anomaly. For example, the right coronary artery originates from the left sinus of Valsalva or from the non-coronary sinus. The dataset can segment detailed coronary artery structures down to capillaries, but it cannot accurately reflect narrow plaque thickness or label fibrous plaques or calcifications.

4.6 Arterial Hypertension

Pulmonary artery aneurysm is a rare abnormality of the pulmonary artery system, where the pulmonary artery is dilated locally due to the weakening of the vessel wall. Pulmonary artery aneurysms may appear at any location in the pulmonary artery, but they are commonly found in the central and left pulmonary arteries. Pulmonary artery aneurysms can cause symptoms such as cough, dyspnea, chest pain, hemoptysis, cyanosis, dizziness, or syncope. Pulmonary artery aneurysms can be detected by MRI, CT, or echocardiography. The treatment of pulmonary artery aneurysms depends on the cause, location, size, and symptoms of the disease. Medical management, endovascular intervention, or surgery may be utilized.

4.7 Double Superior Vena Cava

Double superior vena cava is a rare anatomical variant originating from the left superior vena cava, resulting from the regression of the main embryonic vessel towards the failed fetal vein. The left-sided SVC drains into the right atrium in 90% of cases via an enlarged coronary sinus; alternate locations include the inferior vena cava, hepatic vein, and left atrium. The atypical drainage of the left SVC into the left atrium leads to right-to-left shunting, potentially causing cyanosis and being associated with cases of venous obstruction or infection. The right-sided SVC is observed in 82-90% of cases of double SVC. The absence of the right SVC or the right SVC draining into the left atrium is associated with an increased incidence of congenital heart diseases such as ASD, VSD, and TOF.

4.8 Transposition of the Great Arteries

Transposition of the great arteries (TGA) is the abnormal positioning of the large arteries, where the "big" arteries, the aorta and pulmonary artery, are switched in their origin

from the heart. The aorta is connected to the right ventricle, and the pulmonary artery is connected to the left ventricle, entirely opposite to the anatomy of a normal heart. With these arteries reversed, deoxygenated blood (blue) from the body returns to the right atrium, enters the right ventricle, and then flows into the aorta, returning to the body. Oxygen-rich blood (red) returns to the left atrium from the lungs, enters the left ventricle, and is pumped back to the lungs - contrary to the normal circulation.

4.9 Coarctation of the Aorta

Coarctation of the aorta is a congenital condition where the aorta is narrowed, usually concentrated in the aortic isthmus, called the aortic ligament. A common scenario is the coarctation of the aortic arch, which may present as a small size in newborns with problems with the ductus. With coarctation, the left ventricle must work harder than normal, generating high pressure to push blood through the narrowed aortic arch, delivering blood to the lower part of the body. If severe coarctation occurs, the left ventricle may not be strong enough to overcome the constriction, leading to inadequate blood supply to the lower body. This condition can cause various problems, including hypertension in the upper limbs, left ventricular hypertrophy, and even disrupted blood supply to abdominal organs and lower limbs. Symptoms of coarctation of the aorta vary depending on the degree of narrowing and may include headaches, chest pain, cold feet, fatigue, and weakness. A faint murmur may be heard over the narrowed area. Diagnosis is typically made through echocardiography, CT, or MRI. To treat coarctation of the aorta, methods include balloon angioplasty with stent placement or surgery.

4.10 Total Anomalous Pulmonary Venous Return

Total anomalous pulmonary venous return (TAPVR) is a congenital heart defect. In infants with TAPVR, oxygen-rich blood does not return to the left atrium from the lungs. Instead, oxygen-rich blood returns to the right side of the heart. Here, oxygen-rich blood mixes with oxygen-poor blood. This results in the baby receiving less oxygen than necessary for the body. To survive with this defect, infants with TAPVR often have a hole between the right and left atria (atrial septal defect) allowing mixed blood to flow to the left side of the heart and be pumped to the rest of the body. Some babies may have other heart defects in addition to an atrial septal defect. Because babies with this defect may need surgery or other procedures immediately after birth, TAPVR is considered a serious congenital heart defect.

4.11 Pulmonary Artery Sling

A pulmonary artery ring is a rare condition identified as a form of pulmonary artery malformation. This condition causes tracheal compression due to the encirclement of the trachea by the constricted pulmonary artery ring. This condition is typically detected in children and can cause symptoms such as wheezing, stridor, cough, rapid breathing, and difficulty breathing. Treatment for this condition may involve surgical tracheal reconstruction or placing a pulmonary artery stent into the trachea to relieve narrowing.

4.12 Double Aortic Arch

A double aortic arch is an abnormality of the aortic arch in which two aortic arches form a complete ring that can compress the trachea and/or esophagus. The aortic arch is the first segment of the aorta leaving the heart to supply blood to the body's organs. The vascular ring formed by the double aortic arch is a malformation of the aortic arch. The vascular ring formed by the double aortic arch encircles part or all of the trachea or esophagus. There are cases where both occur. These defects are present from birth (congenital). But symptoms may occur in infancy or later in childhood, adolescence.

References

[1] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: a review," *Frontiers in cardiovascular medicine*, vol. 7, p. 25, 2020.

[2] H. Ta, B. Lin, and L. Palaniappan, "Vietnamese and vietnamese-american health statistics, 2003-2019," *CARE Data Brief*, 2020.

[3] C. A. Miller, P. Jordan, A. Borg, R. Argyle, D. Clark, K. Pearce, and M. Schmitt, "Quantification of left ventricular indices from ssfp cine imaging: impact of real-world variability in analysis methodology and utility of geometric modeling," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 5, pp. 1213–1222, 2013.

[4] H. C. McGill Jr, C. A. McMahan, and S. S. Gidding, "Preventing heart disease in the 21st century: implications of the pathobiological determinants of atherosclerosis in youth (pday) study," *Circulation*, vol. 117, no. 9, pp. 1216–1227, 2008.

[5] S. Park and M. Chung, "Cardiac segmentation on ct images through shape-aware contour attentions," *Computers in Biology and Medicine*, vol. 147, p. 105782, 2022.

[6] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri. midas j," *Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[7] A. Suinesiaputra, B. R. Cowan, J. P. Finn, C. G. Fonseca, A. H. Kadish, D. C. Lee, P. Medrano-Gracia, S. K. Warfield, W. Tao, and A. A. Young, "Left ventricular segmentation challenge from cardiac mri: a collation study," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: Second International Workshop, STACOM 2011, Held in Conjunction with MICCAI 2011, Toronto, ON, Canada, September 22, 2011, Revised Selected Papers 2.* Springer, 2012, pp. 88–97.

[8] C. Petitjean, M. A. Zuluaga, W. Bai, J.-N. Dacher, D. Grosgeorge, J. Caudron, S. Ruan, I. B. Ayed, M. J. Cardoso, H.-C. Chen *et al.*, "Right ventricle segmentation from cardiac mri: a collation study," *Medical image analysis*, vol. 19, no. 1, pp. 187–202, 2015.

[9] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," *arXiv preprint arXiv:1604.00494*, 2016.

[10] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Medical image analysis*, vol. 43, pp. 66–84, 2018.

[11] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Medical image analysis*, vol. 30, pp. 108–119, 2016.

[12] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of mri," *Medical image analysis*, vol. 31, pp. 77–87, 2016.

[13] Z. Yao, W. Xie, J. Zhang, H. Qiu, and H. Yuan, "Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection," *Frontiers in Physiology*, vol. 12, p. 732711, 2021.

[14] A. Zeng, C. Wu, G. Lin, W. Xie, J. Hong, M. Huang, J. Zhuang, S. Bi, D. Pan, N. Ullah *et al.*, "Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images," *Computerized Medical Imaging and Graphics*, vol. 109, p. 102287, 2023.