

**A Deep Learning Framework for Wall Motion
Abnormality Detection in Echocardiograms**

by

Parisa Asgharzadeh

B.A.Sc, Isfahan University of Technology, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL
STUDIES

(Electrical and Computer Engineering)

The University of British Columbia

(Vancouver)

March 2020

© Parisa Asgharzadeh, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

A Deep Learning Framework for Wall Motion Abnormality Detection in Echocardiograms

submitted by **Parisa Asgharzadeh** in partial fulfillment of the requirements for the degree of **Master of Applied Science in Electrical and Computer Engineering**.

Examining Committee:

Purang Abolmaesumi, Electrical and Computer Engineering
Supervisor

Robert Rohling, Electrical and Computer Engineering
Supervisory Committee Member

Jane Wang, Electrical and Computer Engineering
Additional Examiner

Abstract

Coronary Artery Disease (CAD) is the leading cause of morbidity and mortality in developed nations. In patients with acute or chronic obstructive CAD, Echocardiography (ECHO) is the standard-of-care for visualizing abnormal ventricular wall thickening or motion which would be reported as Regional Wall Motion Abnormality (RWMA). The accurate identification of regional wall motion abnormalities is essential for cardiovascular assessment and myocardial ischemia, coronary artery disease and myocardial infarction diagnosis. Given the variability and challenges of scoring regional wall motion abnormalities, we propose the development of a platform that can quickly and accurately identify regional and global wall motion abnormalities on echo images.

This thesis describes a deep learning-based framework that can aid physicians to utilize ultrasound for wall motion abnormality detection. The framework jointly combines image data and patient diagnostic information to determine both global and clinically-standard 16 regional wall motion labels. We validate the approach on a large cohort of echo studies obtained from 953 patients. We then report the performance of the proposed framework in the detection of wall motion abnormality. An average accuracy of 69.2% for the 16 regions and an average accuracy of 69.5% for global wall motion abnormality were achieved.

To the best of our knowledge, our proposed framework is the first to analyze left ventricle wall motion for both global and regional abnormality detection in echocardiography data.

Lay Summary

Identification of patients with regional wall motion abnormalities is beneficial for early detection of any coronary artery disease not evident by symptoms. Conventional methods for assessment of RWMA, which are based on visual interpretation of endocardial excursion and myocardial thickening, are observer variants and depend on the experience level of the echocardiographer. Considering the variability and challenge of coding RWMA, an effective model for the reduction of the misreading of RWMA is required. Thus, we propose the development of a machine learning platform that can quickly and accurately identify regional wall motion abnormalities on echo images. Such a tool would have several applications to improve the accuracy and consistency of RWMA reporting with bedside echo at the point of care.

Preface

This thesis is predominantly based on a pending journal submission. The presented work involves collaboration among multiple students, professors, sonographers and cardiologists at the University of British Columbia, the Department of Electrical and Computer Engineering and Vancouver General Hospital. The study is conducted under the approval of the University of British Columbia (UBC) Research Ethics Board, certificate number H16-02624, provided by the Vancouver Coastal Health Research Ethics Board.

The author has developed and implemented the proposed framework for wall motion abnormality detection, as well as evaluating the solution on the created local database of corresponding data. Professor Purang Abolmaesumi and Drs. Teresa Tsang and Christina Luong helped with technical guidance and insight into the problem being addressed.

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
Glossary	xiii
Acknowledgments	xv
1 Introduction and Background	1
1.1 Clinical Background	1
1.1.1 Heart	1
1.1.2 Echocardiography	5
1.1.3 Wall Motion Abnormality (WMA)	7
1.2 Machine Learning in WMA Detection	15
1.2.1 Deep Neural Networks	15
1.2.2 Convolutional Neural Networks	17
1.2.3 Recurrent Neural Networks	18
1.2.4 Hyper-Parameter Optimization	21

1.2.5	Applications of ML in WMA detection	21
1.3	Thesis Objective	23
1.4	Contributions	23
1.5	Thesis Outline	24
2	Materials	26
2.1	Ethics	26
2.2	Echocardiography Data	27
2.2.1	Echocardiography Retrospective Data	27
2.2.2	Echo Data Download and Processing	29
2.3	Analytical Measurements	29
2.3.1	Filemaker Measurements	30
2.3.2	Analytical Data Download and Processing	31
2.4	Summary	34
3	Methods	37
3.1	Relevant Cardiac Echo View Selection for WMA	37
3.2	Automatic WMA Detection Deep Convolutional Network	39
3.2.1	Spatial Feature Extractors	39
3.2.2	Temporal Feature Aggregators	41
3.2.3	Spatio-temporal Framework	42
3.2.4	Regularization and Data Augmentation	43
3.3	Summary	45
4	Experiments and Results	46
4.1	Experimental Setup	46
4.1.1	Dataset	46
4.1.2	Wall Motion Abnormality labels	47
4.1.3	Network Architecture	47
4.1.4	Hyper-parameters	48
4.1.5	Regularization and Data Augmentation	48
4.1.6	Training	49
4.1.7	Implementation	50
4.2	Results	51

4.3 Discussion and Summary	54
5 Conclusion	58
5.1 Contributions	59
5.2 Future Work	60
Bibliography	61
A Supporting Materials	69

List of Tables

Table 1.1	The complete list of all 17 segments of left ventricle wall. . . .	10
Table 2.1	The Wall Motion Abnormality data frequency in each region. .	33
Table 3.1	The dataset composition in terms of number of studies from each type of the 14 standard echocardiography views (A#C: apical #-chamber view, PLAX: parasternal long-axis view, RVIF: right ventricular inflow view, S#C: subcostal #-chamber view, IVC: subcostal inferior vena cava view, PSAX-A: parasternal short-axis view at aortic valve, PSAX-M: PSAX view at mitral annulus valve level, PSAX-PM: PSAX view at mitral valve papillary muscle level, PSAX-APEX: PSAX view at apex level, and SUPRA: Suprasternal view).	38
Table 4.1	The co-variance matrix of the 16-segment labels and WMSI. .	53
Table 4.2	The Global WMA accuracy per class comparison of experimented methods.	55
Table 4.3	The A2C relevant RWMA accuracy per class comparison of experimented methods.	56
Table 4.4	The A4C relevant RWMA accuracy per class comparison of experimented methods.	56
Table 4.5	The PLAX relevant RWMA accuracy per class comparison of experimented methods.	57

List of Figures

Figure 1.1	Heart position in the thoracic cavity, located in the mediastinum between the lungs. (This image is available under a Creative Commons Attribution License 2.0 at https://commons.wikimedia.org .)	2
Figure 1.2	Human blood circulation in Heart. (This image is available under a Creative Commons Attribution Licence 2.0 at https://commons.wikimedia.org .)	4
Figure 1.3	Demonstration of different echo modalities.	6
Figure 1.4	Four main cardiac views. (Left) The transthoracic echocardiogram displaying different structures in the heart. (Right) An anatomical diagram of the corresponding view.	8
Figure 1.5	LV regional wall motion analysis. Subfigure (a) shows the effect of WMA severity on motion dysfunction of the wall. It has been advised to assess the wall motion of individual LV segments visually and score them [29].	9

Figure 1.6	ACC/AHA recommended “bull’s-eye” plot of the 17 segment model. The outer ring represents the basal segments, the middle ring represents the segments at midpapillary muscle level, and the inner ring represents the distal level. The anterior insertion of the right ventricular wall into the left ventricle defines the border between the anteroseptal and anterior segments. Starting from this point, the myocardium is subdivided into six equal segments of 60 degrees. The apical myocardium is divided instead into four equal segments of 90 degrees. The apical cap is added in the center of the bull’s-eye.	11
Figure 1.7	(a) Demonstration of colour kinesis where lack of colour changing depicts lack of wall thickening[37]. (b) Three-dimensional echocardiographic images of the heart (apex view) [38]. (c) The main directions of deformation and strain imposed on the LV myocardium. (d) Example of measurements of the main 2D Strain variables. Each of the coloured lines at the left panels denotes one of the six regions measured from the apical 4-chamber window [12].	14
Figure 1.8	Demonstration of a deep convolutional neural network with five hidden layers including convolutional, pooling and fully-connected layers.	17
Figure 1.9	Demonstration of an LSTM block. m_{t-1} stands for the input from a memory cell in time point t ; x_t is an input in time point t ; h_t is an output in time point t that goes to both the output layer and the hidden layer in the next time point.	20
Figure 1.10	A systematic diagram of the workflow for automated RWMA prediction. After data are acquired in the clinic, it is stored in the Vancouver Coastal Health (VCH)’s servers along with the cardiologist’s measurements. The machine learning model will then use these data to make predictions on RWMA. . . .	22
Figure 2.1	A diagram designed to show the relationship between the Echo, Xcelera and Filemaker databases in a routine cardiology study.	28

Figure 2.2	Snapshot of the wall motion abnormality section of FileMaker. By clicking on each region, the cardiologist will type a score from 1 to 5 to each region to label the abnormality of the region.	30
Figure 2.3	A block diagram of the relationship between the Echo database tables.	31
Figure 2.4	The Wall Motion Abnormality data distribution in different regions of the heart: (a) is data distribution in LCX region, (b) is data distribution in LAD region, and (c) is data distribution in RCA region.	35
Figure 2.5	Global WMA data distribution.	36
Figure 3.1	The cardiac view classifier network architecture. Related embeddings are extracted from the individual frames by the Spatial embedding extractor (DenseNet blocks). The embeddings are then fed into the Long Short-Term Memory (LSTM) blocks to extract the temporal information across 10 sequential echo cine frames.	38
Figure 3.2	The proposed WMA network architecture. Spatio-temporal embeddings are extracted from the individual cines. The embeddings are then fed into the FC blocks to connect the reasoning between the information across echo cine frames.	40
Figure 3.3	The network architectures for WMA classification considered in this work. (a) C2D are 2D Covolutions; (b) C2D+LSTM are time distributed 2D Covolutions followed by LSTM units; (c) C2D+GRU are time distributed 2D Covolutions followed by GRU units; (d) D2D are 2D DenseBlocks followed by LSTM units; and (e) D2D are 2D DenseBlocks followed by GRU units. (f) C3D are 3D convolutions. For interpretability, the connections are omitted.	44
Figure 4.1	The value of loss for training and validation over each epoch.	52
Figure A.1	A snapshot of different tabs available in the Filemaker software for cardiologists to records the analytical measurements.	73

Glossary

2D	two-dimensional
3D	three-dimensional
A2C	Apical Two-chamber
A4C	Apical Four-chamber
CAD	Coronary Artery Disease
CCT	Cardiac Computed Tomography
CMR	Cardiac Magnetic Resonance (Imaging)
CNN	Convolutional Neural Networks
CSV	Comma-Separated Values
DCNN	Deep Convolutional Neural Networks
DICOM	Digital Imaging and Communications in Medicine
ECG	Electrocardiogram
ECHO	Echocardiography
FC	Fully-Connected Layers
GRU	Gated Recurrent Unit
LAD	Left Anterior Descending Artery

LCX	Left Circumflex Artery
LSTM	Long Short Term Memory
LV	Left Ventricle
MRN	Medical Record Number
ONEIROS	Open-ended Neuro-Electronic Intelligent Robot Operating System
PACS	Picture Archiving and Communication System
PLAX	Parasternal Long-axis
POCUS	Point-Of-Care Ultrasound
PSAX	Parasternal Short-axis
RCA	Right Coronary Artery
RCL	Robotics and Control Lab
RF	Radio Frequency
ROI	Region of Interest
RWMA	Regional Wall Motion Abnormality
UBC	University of British Columbia
VCH	Vancouver Coastal Health
WMA	Wall Motion Abnormality
WMSI	Wall Motion Score Index

Acknowledgments

I would like to give my enduring gratitude to my supervisor Professor Purang Abolmaesumi for his dedicated role. I have been fortunate to have him as not only a person with a deep insight into our field but a person with persistent patience and endless support that helped me overcome many crisis situations and finish this stage of my life. I would additionally like to thank all of the instructors that I have been lucky to learn from over the last two years. Thank you, Christina Luong and Hany Girgis, for your precious hours teaching me different aspects of cardiology and helping me with my questions.

It has been my pleasure to be part of the Robotics and Control Lab. Thank you all for making it a fantastic place to work. I am really lucky to have worked with you.

To my friend who took time out of her busy life to help edit this thesis, Delaram Behnami, thank you again. Thank you for all your endless guidance and support through the unknown aspects of grad school. A whole-heartedly thanks to my family for their countless love and support.

Finally, I would like to acknowledge the Natural Science and Engineering Research Council (NSERC), and the Canadian Institutes of Health Research (CIHR), who funded this project.

Chapter 1

Introduction and Background

1.1 Clinical Background

1.1.1 Heart

Anatomy

The heart is a muscular organ that pumps blood around the body, located within the thoracic cavity, medially between the lungs in the space known as the mediastinum. Figure 1.1 shows the position of the heart in the thoracic cavity. The heart is located on its own space named the pericardial cavity and is separated from other mediastinal parts in the mediastinum by the pericardium or pericardial sac which is a tough membrane.

The heart looks like a pinecone. A person's heart is typically around the size of his/her fist: normally $12\text{ cm} \times 8\text{ cm} \times 6\text{ cm}$. Considering the differences between the members of each sex, an average female heart is approximately 250-300 grams and an average male heart is approximately 300-350 grams. Exercise is an important factor in the increase in the size of the heart. The behaviour of the cardiac muscle at exercise time is similar to the skeletal muscle. The exercise increases the protein myofilaments, resulting in an increase in the size of particular cells without increasing the number of them, i.e. hypertrophy. Thus, athletes' hearts can pump blood more effectively (same amount of blood pump at a lower heart rate) than

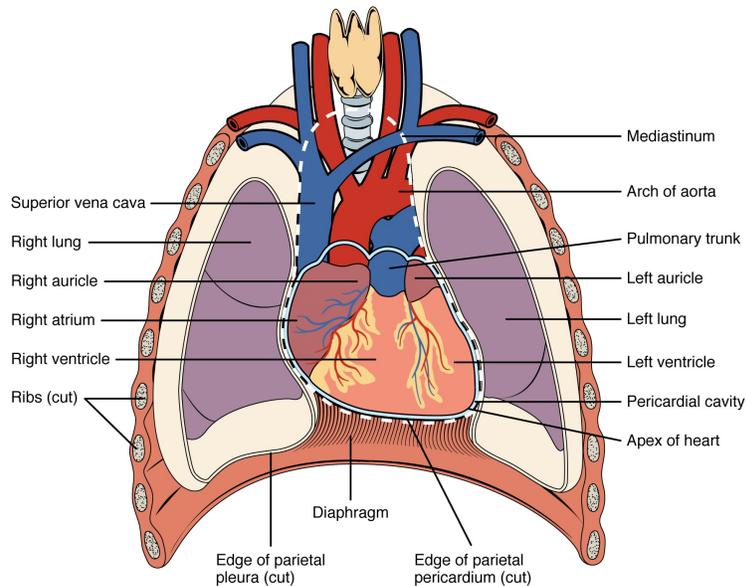


Figure 1.1: Heart position in the thoracic cavity, located in the mediastinum between the lungs. (This image is available under a Creative Commons Attribution License 2.0 at <https://commons.wikimedia.org>.)

non-athletes [43].

There are four chambers in the human heart: one atrium and one ventricle on each side of the heart. The upper chambers, also known as the atrium, serve as receiving chambers and contract to push the blood to the ventricles. On the other hand, the ventricles function as the principal pumping section, pushing the blood to the lungs or the other organs in the body.

The wall of the heart consists of three unequal thickness layers of tissue: epicardium, myocardium, and endocardium. The so-called layers are mainly covered with a thin protective layer named pericardium. The epicardium is mainly made of connective tissue. The myocardium particularly consists of the muscles of the heart, and the endocardium lines the inside of the heart and protects the valves and chambers [43].

Besides, the heart contains four valves being used to keep the blood flow in one direction only. The atrioventricular valve or tricuspid valve is located between the

right atrium and the right ventricle. Moving forward from the right ventricle at the base of the pulmonary trunk, we reach the pulmonary valve. The mitral valve, also known as the bicuspid valve, is located at the opening between the left atrium and left ventricle. The last valve is the aortic valve settled at the base of the aorta that prevents back-flow from the aorta.

Function

The heart, blood, and blood vessels combined are called the circulatory system. An average human body has around 5 litres of blood, which is constantly pumped throughout the body around 100,000 times a day.

There are two blood circulations in the human body called pulmonary and systemic circuits. The pulmonary circuit is the transportation of blood between the heart and lungs, where it receives fresh oxygen and delivers carbon dioxide for exhalation. On the other hand, the systemic circuit is the transportation of oxygenated blood to all other tissues of the body and returning of deoxygenated blood and carbon dioxide to the heart to be sent back to the pulmonary circulation.

A cardiac rhythm or heartbeat is known as the process of blood pumping through the four chambers. The heartbeat can be split into two phases: systole and diastole. In diastole, the atria and ventricles relax and fill with blood. In systole on the other hand, the atria contract and pump blood into the ventricles; after that, as the atria start to relax, the ventricles contract (ventricular systole) and pump blood out of the heart.

In summary, as shown in Figure 1.2, blood flows from the right atrium to the right ventricle, then it is pumped into the pulmonary circuit. The blood in the pulmonary artery branches has low oxygen but a relatively high amount of carbon dioxide. In the pulmonary capillaries, gas exchange occurs (oxygen is entered to blood, carbon dioxide is out). Subsequently, the blood which is high in oxygen and low in carbon dioxide is returned to the left atrium. Then, blood enters the left ventricle, which is pumped to the systemic circuit. Following the exchange in the systemic capillaries, the deoxygenated blood returns to the right atrium and the cycle is repeated.

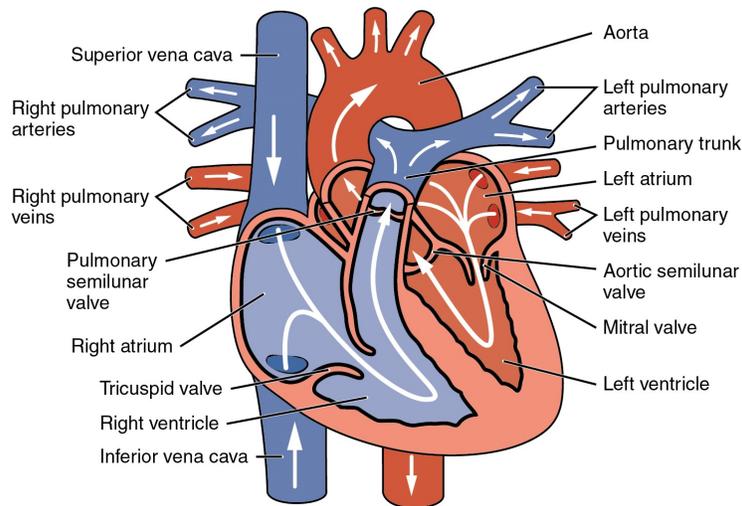


Figure 1.2: Human blood circulation in Heart. (This image is available under a Creative Commons Attribution Licence 2.0 at <https://commons.wikimedia.org>.)

Imaging Modalities

The visual assessment of ventricular function, cardiac chamber dimensions, and ventricular mass is essential for clinical diagnosis, risk assessment, therapeutic decisions, and prognosis in patients with any kind of cardiac disease. Many imaging techniques are applicable for the assessment of the left ventricular function each with their own limitations.

Cardiac Magnetic Resonance (Imaging) (CMR) has been considered as the gold standard for Left Ventricle (LV) assessment. Despite good performance, CMR is expensive, time-consuming, and is not available in most of the medical centers. Moreover, it can not be used for all patients, due to the presence of metal devices or clinical conditions such as claustrophobia and the inability to lay flat in some patients [17].

Cardiac Computed Tomography (CCT) is a non-invasive imaging technique that can be used to obtain information about left ventricular function and morphology, in addition to its main application which is the assessment of coronary artery dis-

ease. It serves as a decent alternative option when other imaging modalities such as echocardiography cannot deliver acceptable images; or where CMR cannot be used due to patient's contraindications [50].

Having compelling developments in ultrasound technology, the routine incorporation of harmonic imaging has been used clinically for the assessment of LV segmental function [21].

A detailed assessment of the global and regional myocardial function is developed using other echocardiographic imaging modalities including automated endocardial border detection using integrated backscatter, tissue Doppler and two-dimensional (2D) speckle tracking imaging of myocardial displacement, velocity, strain and strain rate, and real-time three-dimensional (3D) echo that will be discussed in the next section.

1.1.2 Echocardiography

Echocardiography or cardiac echo, mainly known as echo, is an ultrasound image of the heart. The conventional ultrasound image is created by an ultrasound transducer transmitting and then receiving Radio Frequency (RF) signals. The RF signals are then converted to a digital RF signal, filtered to produce an envelope-detected signal. The resulting signal will produce the final B-mode image after some post-processing methods. Moreover, for doing some measurements, colour flow Doppler is interpreted.

Echocardiography has been routinely used in most of the diagnosis, management, and follow-up of patients related to heart diseases. It is one of the most common diagnostic tests used in cardiology. It provides rich information about the size and shape of the heart (internal chamber size quantification), pumping capacity, and the location and extent of any tissue damage. Moreover, the videos of it help cardiologists to have a good estimation of the heart function, such as the calculation of the cardiac output, ejection fraction, and diastolic function (i.e. how well the heart relaxes). It is also used as a tool for the assessment of how severe is the wall motion abnormality in patients with suspected cardiac diseases.

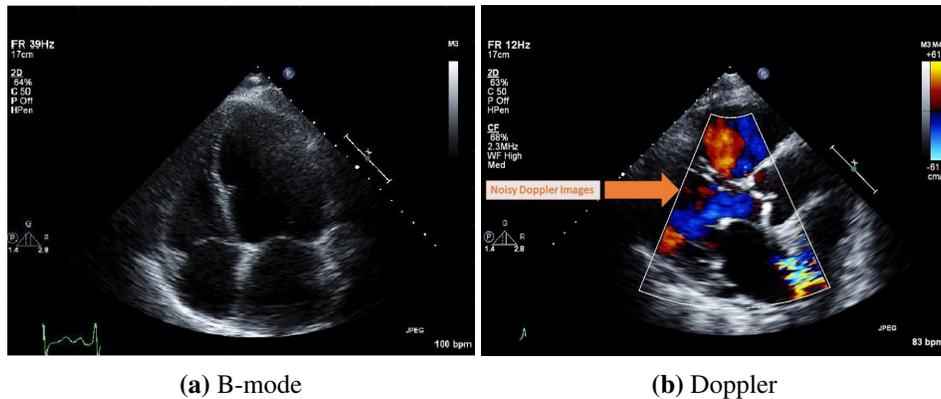


Figure 1.3: Demonstration of different echo modalities.

B-mode

The B-mode echo producing the visual interface of examined anatomy in both 2D and 3D echo files is the most common method of heart imaging. The brightness of each pixel in the images is dependent on the amplitude of the returned echo signal. The difference in brightness of each tissue allows visualization and quantification of anatomical structures, as well as visualization of diagnostic and therapeutic procedures. This is a real-time method for image acquisition, allowing for up to 50-70 images per second in 2D echo.

Colour Doppler

Colour Doppler ultrasound allows cardiologists to clearly observe the blood flow through the heart and the blood vessels. It also allows them to measure obstructions in arteries and the degree of narrowing or leakage of heart valves (regurgitation). It is mainly done by encoding colour Doppler information and overlaying it on 2D echo images. Each colour is a representation of the speed of blood flow within a Region of Interest (ROI).

Cardiac views

Transthoracic Echocardiography (ECHO) is the most common imaging modality used for cardiac assessment. The ultrasound data is acquired from standard cross-

sections of the heart for measurement and examination of multiple variables in cardiac structures and functions. Each of these cross-sections of the heart will result in different views in echo that distinctly highlight specific regions of the heart details. The cardiac echo data is acquired by the manual movement of the imaging probe over chest acoustic windows. A good interpretation of the heart is achieved through correct acquisition using the best fixation on the cross-section. This requires years of experience and expertise.

2D echo incorporates a recording of 2D cardiac images, often referred to as cine series. Each cine illustrates a 2D cross-sectional video of heart and may contain several cardiac cycles. There are 14 cardiac standard views each with their own set of signature features. The main standard cardiac views are:

- Apical views (Apical Two-chamber (A2C), Apical Four-chamber (A4C), etc.)
- Parasternal Short-axis (PSAX) views
- Parasternal Long-axis (PLAX) view.

Figure 1.4 shows the main cardiac views, in which the transthoracic echo view is alongside its anatomical diagram.

The four views most frequently acquired by clinicians are apical four-chamber, parasternal long axis (PLAX), parasternal short axis at the papillary muscle level (PSAX-PM), and subcostal four-chamber (SUBC4). For the purpose of this thesis, we only need three views of the heart: apical two-chamber and four-chamber, and parasternal long axis (PLAX).

1.1.3 Wall Motion Abnormality (WMA)

Clinical Definition

The blood is supplied to different regions of the heart through three main epicardial coronary arteries. The left main coronary artery is split to the Left Anterior Descending Artery (LAD) and the Left Circumflex Artery (LCX). In the blood supply, there may be some variations but the overall pattern is the same: The LAD goes down the interventricular groove and provides blood to the anterior wall (with the help of its diagonal branches), anterior septum and apex of the heart. On the other

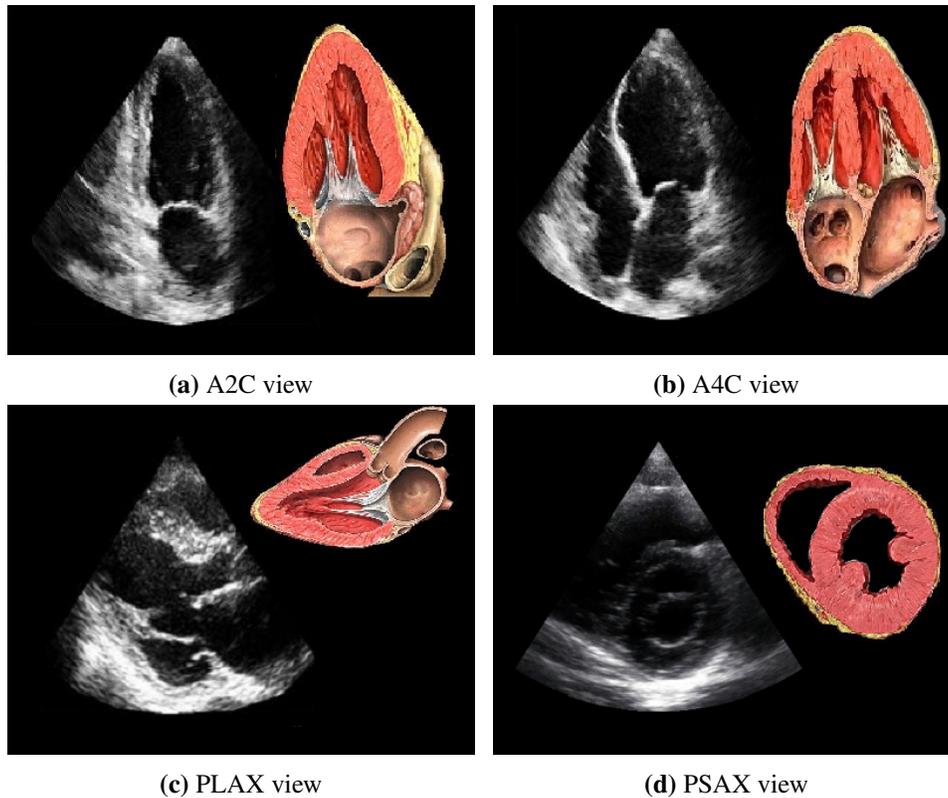


Figure 1.4: Four main cardiac views. (Left) The transthoracic echocardiogram displaying different structures in the heart. (Right) An anatomical diagram of the corresponding view.

hand, the lateral wall is supplied by the left circumflex artery with its marginal branches. The Right Coronary Artery (RCA) arising from the right sinus of Valsalva and going infero-medially down to the atrioventricular groove. The posterior descending artery (PDA), which is a branch of RCA, supplies the inferior wall and the inferior septum.

In echocardiography, regional myocardial function assessment is mainly done by observing the wall thickening and endocardial motion of the myocardial segment.

However, it should be remembered that deformation can also be passive and therefore, may not always accurately reflect myocardial contraction.

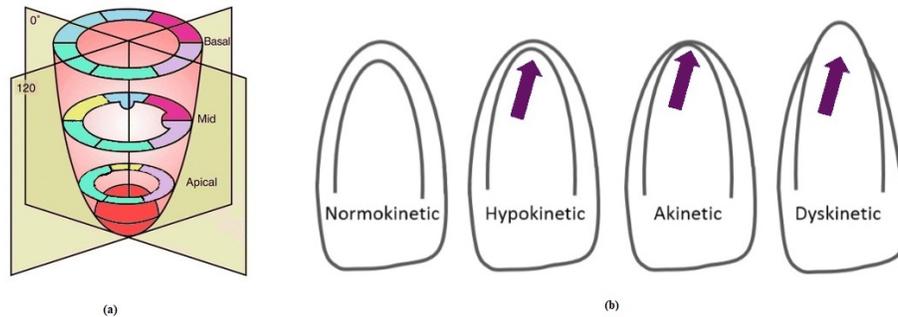


Figure 1.5: LV regional wall motion analysis. Subfigure (a) shows the effect of WMA severity on motion dysfunction of the wall. It has been advised to assess the wall motion of individual LV segments visually and score them [29].

Many models have been used to divide the left ventricular myocardium into segments so that regional wall motion can be accurately explained and quantified. The LV myocardium is divided into three sections as follows:

- base (six segments, each encompassing 60 degrees of the left ventricular short-axis which are described as: basal anterior, basal anterolateral, basal inferolateral, basal inferior, basal inferoseptal)

- mid-section (divided into six segments in a similar manner to the base)

- apex (divided into four 90 degree segments of apical anterior, apical lateral, apical inferior, apical septum).

Each of these sections is divided into some segments that correspond to regions in the LV wall. The most recent recommendation from the American College of Cardiology (ACC) and American Heart Association (AHA) is a 17-segment model [42]. Since the apical cap in the 17 segment model is acontractile and therefore more appropriate for perfusion imaging, the 16 segment model of myocardial segmentation is being used extensively. Table 1.1 shows a complete list of all segments highlighted in Figure 1.6.

Each segment is then assigned a score between 1 to 5, using the following criteria:

- 1 = normal or normokinetic (normal wall thickening and endocardial excur-

Table 1.1: The complete list of all 17 segments of left ventricle wall.

Basal segments	Mid segments	Apical segments
1. Basal anterior	7. Mid anterior	13. Apical anterior
2. Basal antero-septal	8. Mid antero-septal	14. Apical septal
3. Basal infero-septal	9. Mid infero-septal	15. Apical inferior
4. Basal inferior	10. Mid inferior	16. Apical lateral
5. Basal infero-lateral	11. Mid infero-lateral	17. Apex
6. Basal antero-lateral	12. Mid antero-lateral	

sion)

2 = hypokinetic (myocardial thickening $\leq 30-40\%$ in systole)

3 = severely hypokinetic or akinetic (myocardial thickening $\leq 10\%$ in systole)

4 = dyskinetic (segment moves outward in systole)

5 = aneurysmal (segment pouches out in both systole and diastole)

There is a correspondence between the location of regional wall motion abnormalities and the coronary artery territories. Mainly, the anterior septum, anterior wall and the anterior apex are affected by LAD. Any disease in the LCX artery affects the lateral and posterior walls of the left ventricle. Besides, CAD detection is improved by the detection of regional wall motion abnormality in this region mostly due to the pathology of the LCX artery. The RCA at its posterior descending branch supplies the inferior septum and inferior wall of the left ventricle. Any wall motion abnormality in the mentioned areas has a strong correspondence with RCA. Since the location of regional wall motion abnormalities correlates reasonably well with the location disease in coronary arteries, it can be utilized as a valid guide for further management.

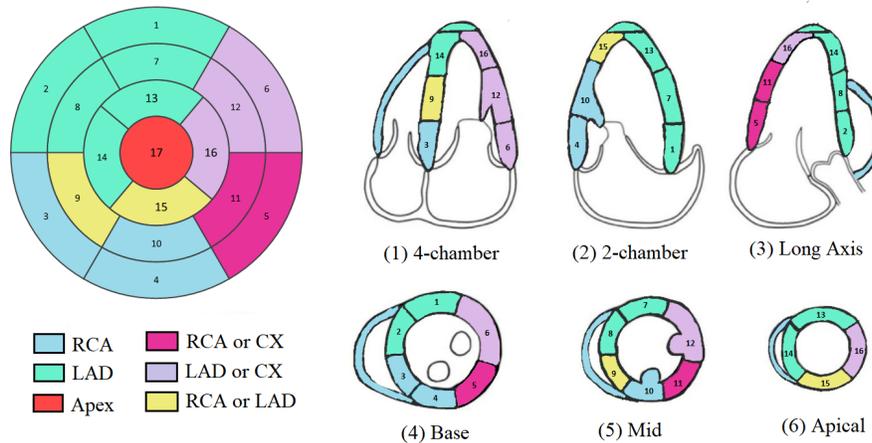


Figure 1.6: ACC/AHA recommended “bull’s-eye” plot of the 17 segment model. The outer ring represents the basal segments, the middle ring represents the segments at midpapillary muscle level, and the inner ring represents the distal level. The anterior insertion of the right ventricular wall into the left ventricle defines the border between the anteroseptal and anterior segments. Starting from this point, the myocardium is subdivided into six equal segments of 60 degrees. The apical myocardium is divided instead into four equal segments of 90 degrees. The apical cap is added in the center of the bull’s-eye.

Clinical Importance

Regional wall motion abnormality refers to the motion of a region of the heart muscle being abnormal. Myocardial infarction or severe ischemia is the most common cause of left ventricular (LV) wall motion abnormalities [21].

Regional wall motion abnormalities mainly occur early in the ischemic cascade followed by Electrocardiogram (ECG) changes. When the myocardium oxygen demand is reduced and subsequently returned to baseline, there would be a resolution of myocardial ischemia and wall motion returns to normal [42]. Wall motion analysis has been a paramount factor in clinical decision-making situations such as in patients with chest pain in the emergency department and patients with congestive heart failure [33]. The accurate evaluation of left ventricular (LV) regional function

is essential for general cardiac assessment, specifically for evaluation of CAD and acute myocardial infraction [27, 29, 33, 58].

Identification of regional wall motion abnormalities (RWMA) can help in the diagnosis of acute myocardial infraction and multi-vessel CAD and coronary syndromes or chronic CAD, direct to the ischemic territory and finally influence patient treatment [8].

Automated LV Wall Motion Assessment Using Echocardiography

Echocardiography, with its high spatial and temporal resolution, is the best choice as a non-invasive method for assessing changes in wall motion. In patients with an acute situation, having ECG inclusive would be useful in the early detection of myocardial ischemia. Equally, in patients with acute chest pain, ECG included, normal regional wall motion may help to exclude underlying myocardial ischemia [21].

Despite the impressive advances in echo technology, detection and quantification of regional left ventricular wall motion abnormalities on echocardiography images is highly subject to the observer skills, require substantial experience [8, 33], and is prone to significant inter-observer variability [8].

To alleviate the wall motion assessment subjectivity, some other echocardiographic imaging modalities are developed to automate the process and reduce the inter- and intra-observer variability. These methods conduct a more comprehensive assessment of global and regional left ventricle function. In the following, we will discuss some of these methods.

- **Border Tracking**

The automated border tracking method is a procedure in which the difference between the ultrasound backscatter emitted from the endocardium and blood in the LV cavity is being used. After image acquisition, the backscatter information along the scan line is analyzed and the pixels are classified. The pixels are colour coded and superimposed onto a 2D image. This leads to real-time tracking of the endocardial border. The figure shows an echo image with and without automated border tracking. The main drawback of automated endocardial border tracking is that it is dependent on good image

quality. The poor quality images lead to poor tracking of the endocardial border which results in turn poor colour-coded tracking images [37, 46].

- Tissue Doppler imaging

In Tissue Doppler imaging (TDI), high amplitude, low-frequency Doppler signals coming from the myocardium and mitral annulus are measured. In Figure 1.7 a tissue Doppler sample volume of the myocardium or the annulus area and the systolic and diastolic velocities at that point are then displayed. Essentially, any area of the myocardium can be studied in this manner. Thus, the quantitative assessment of regional systolic function is achieved by measuring the S wave peak velocity. Any translational movement and tethering affect the myocardial velocity measurements, which in turn leads to the difficulty of discrimination between segments that are akinetic or actively contracting. Besides, the distribution of the velocities is not uniform through the myocardium; as moving from base to apex, it would be harder to set a reference value [18].

- Strain Rate Imaging

Imaging myocardial deformation is a good alternative to conquer the limitations of velocity measurements. Strain and strain rate measurements are a representation of the magnitude and rate of change in length of myocardial fibre which is the energy required in both systole and diastole [12, 14]. Two-dimensional speckle tracking echo is a method for assessment of myocardial motion by tracking natural acoustic markers, known as speckles, generated from interactions between ultrasound and myocardium. By tracking the motion of speckles, deformation is measured. Limitations of speckle tracking in echo include the need for good image quality and the assumption that a given speckle can be tracked from one frame to the next, which may not happen in excessive cardiac motions.

- 3D echocardiography

The development of 3D echo has allowed assessing LV volumes and movement without any dependency on LV geometry and any assumption about LV shape. Thus, the truncation in apical visualization in 2D echo is resolved

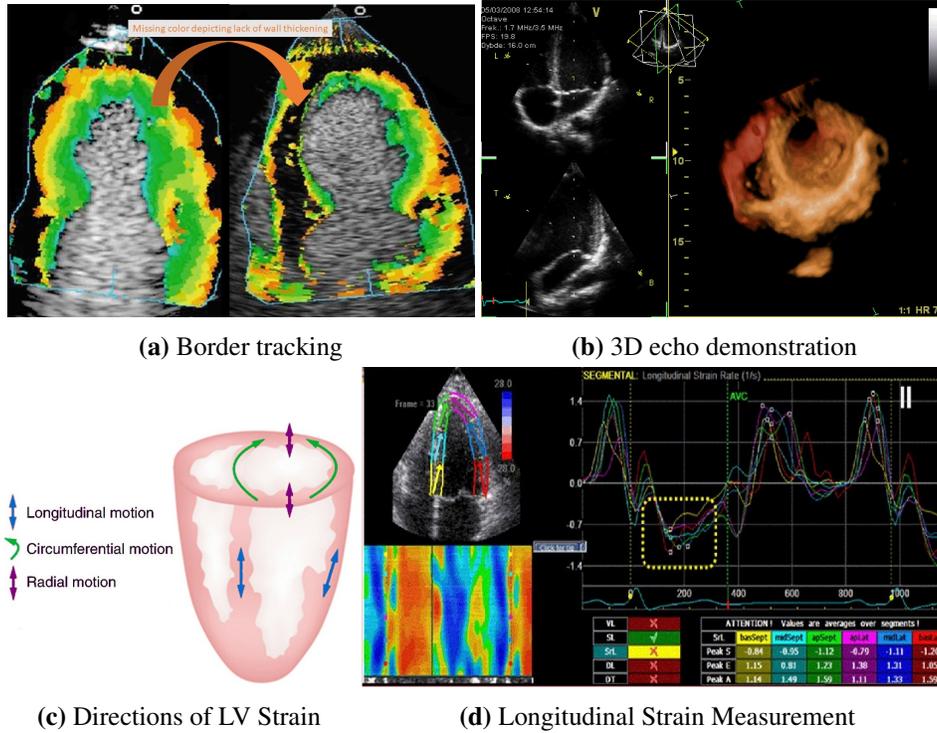


Figure 1.7: (a) Demonstration of colour kinesis where lack of colour changing depicts lack of wall thickening[37]. (b) Three-dimensional echocardiographic images of the heart (apex view) [38]. (c) The main directions of deformation and strain imposed on the LV myocardium. (d) Example of measurements of the main 2D Strain variables. Each of the coloured lines at the left panels denotes one of the six regions measured from the apical 4-chamber window [12].

in 3D echo. 3D echo captures the entire volume of the left ventricle during image acquisition. The acquired images are demonstrated as a rendered volume or surface, wire-framed or 2D tomographic slices [38]. Figure 1.7-b shows an example of a 3D echo. While 3D echo has got more popular than before, it is still limited to the need for good image quality and operator skills and experience. Temporal and spatial resolution development and data manipulation are also needed to enhance the application of this method.

Although echo imaging has had remarkable enhancements, the echo data are

still nontrivial. This results in an adverse effect on echo-based diagnosis due to high inter and intra-observer variability. These limitations are because of intrinsic ultrasound constraints such as noise, frequency vs. depth trade-off, probe tethering and dependency of image quality to the right probe positioning in the planes. However, these limitations are addressed partially by automatic machine learning methods that benefit from high spatio-temporal resolution of echo images [2, 57, 59, 60].

1.2 Machine Learning in WMA Detection

Recently, applications of machine learning in medical imaging have become very prominent. Machine learning can be used in different ways to achieve diverse objectives. Thus, many research groups in the field have focused on semi-automatic and automatic techniques for cardiac assessment (e.g., wall motion abnormality analysis). Deep learning as one powerful branch in machine learning is an effective method for detection and classification for several diseases.

1.2.1 Deep Neural Networks

Neural networks are data processing structures (i.e. functions). They map input x in \mathbb{R}_n to the output \hat{y}_m in \mathbb{R}_m . In classification, \hat{y}_m is the likelihood of each M classes. If we assume each neuron as a function of x , W as parameters (weight) matrix and b as bias vector, then the function of each neuron can be written as:

$$\hat{y}_m = f(Wx + b). \quad (1.1)$$

A typical neural network consists of multiple layers as follows:

- Input Layer
- Hidden Layer(s)
- Output Layer

Deep neural networks have been successfully applied to medical imaging tasks such as image classification, object detection, and image segmentation thanks to the development of Convolutional Neural Networks (CNN). These neural networks

utilize parameterized, sparsely connected kernels which preserve the spatial characteristics of images. Convolutional layers sequentially downsample the spatial resolution of images while expanding the depth of their feature maps. This series of convolutional transformations can create much lower-dimensional and more useful representations of images than what could possibly be hand-crafted. The success of CNNs has spiked interest and optimism in applying deep learning to computer vision tasks. There are many branches of study that hope to improve current benchmarks by applying deep convolutional networks to computer vision tasks. Improving the generalization ability of these models is one of the most difficult challenges. Generalizability refers to the performance difference of a model when evaluated on previously seen data (training data) versus data it has never seen before (testing data). Models with poor generalizability have overfitted the training data. One way to discover overfitting is to plot the training and validation accuracy at each epoch during training.

Training neural networks is an optimization problem. The parameters of f , W and b are optimized with respect to a set of data X , with the labels Y , using a defined loss function. The most common method of optimization takes the derivatives and minimizes the loss function with respect to the weights via iterative back-propagation of gradients in parameters. This is a non-convex optimization, which means during training we may go through some local optima points. Besides, a trained neural network is prone to be over-fitted on the set of training data, meaning you may do well on the train set but perform worse on the unseen test set.

There are some solutions to defeat the over-fitting problem. One is using regularization, which penalizes the parameters W from becoming too specific on training data. Limiting the size of the network (number of hidden units) and also limiting the size of the weights (Weight decay) and early stopping before over-fitting are other possible solutions for this issue.

Deep Neural Networks consist of a number of hidden layers, extracting low to high-level features from images. Convolutional neural networks and recurrent neural networks are two examples of deep neural networks.

1.2.2 Convolutional Neural Networks

Convolutional Neural Networks were first inspired by visual cortex research done by Hubel et al [23]. To date, CNNs are the most powerful tool for image classification and regression problems [34]. Using CNNs the number of network parameters is reduced due to parameter sharing through convolutional kernels.

In CNNs, the input image is a matrix of pixel values that are a representation of the brightness at the given pixel in the image. While traditional neural networks treat the whole image as a one-dimensional array, CNNs include the location of pixels and their neighbours into consideration.

In a convolutional layer, a weight matrix (kernel) is used for the extraction of low-level features. The kernel with its corresponding weights slides over the image matrix to obtain the convolved output. The kernel is like a filter for the extraction of particular information from the input image. By minimizing the loss function the weight of the kernel is learnt.

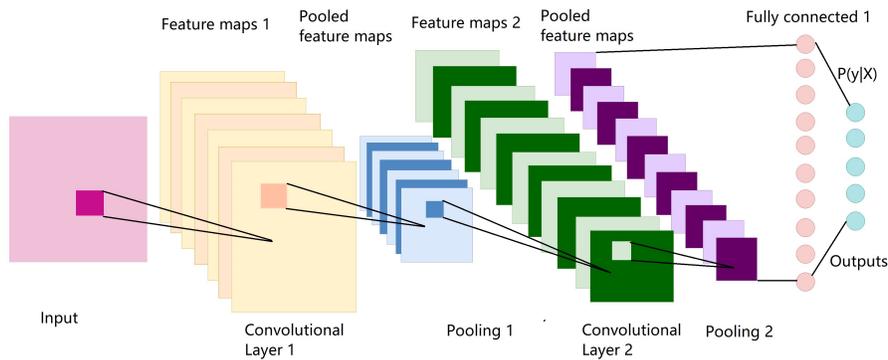


Figure 1.8: Demonstration of a deep convolutional neural network with five hidden layers including convolutional, pooling and fully-connected layers.

Figure 1.8 is a typical configuration of a CNN architecture. The configuration mainly consists of an input image, followed by a sequence of convolutional layers joint with a non-linear function and pooling function. The result is then fed to the last layers containing sequences of fully-connected blocks with the output size being equal to the number of classes in the dataset.

1.2.3 Recurrent Neural Networks

Recurrent neural networks are a branch of neural networks that extract the features in temporal dimension and have been used widely in time-sequence modelling [5]. RNNs process sequential data points through a recurrent hidden state that its activation at each step depends on that of a previous step [19]. An RNN updates its hidden state of h_t by the following equation, where $X = (x_1, \dots, x_N)$ is a sequence data:

$$h_t = \begin{cases} \Theta(h_{t-1}, x_t), & \text{if: } t \neq 0 \\ 0, & \text{if: } t = 0 \end{cases} \quad (1.2)$$

Knowing that x_t is the data value and h_t is the recurrent hidden state at time step t , and Θ denotes the activation function of a hidden layer which is nonlinear such as sigmoid or hyperbolic tangent. Having $y = (y_1, \dots, y_T)$ as the output is optional for RNNs. The conventional RNN model, known as vanilla, updates the recurrent hidden state in equation 1.2 is implemented as:

$$h_t = \Theta(\mathbf{W}x_t + \mathbf{Z}h_{t-1}). \quad (1.3)$$

In the equation, \mathbf{W} and \mathbf{Z} are the coefficient matrices of the input at the current step and the recurrent hidden units activation at the previous step, respectively. By expanding equation 1.2 the hidden vector sequence h_t is calculated as follows:

$$h_t = \Theta(\mathbf{W}_{ih}x_t + \mathbf{W}_{hh}h_{t-1} + b_h), \quad (1.4)$$

where \mathbf{W}_{ih} denotes the input-hidden weight vector and \mathbf{W}_{hh} is the weight matrix of the hidden layer, and b_h is the bias vector in the hidden layer.

While traditional RNN implementation has vanishing gradient problem, meaning that gradients decrease significantly for a deeper temporal model, new types of recurrent hidden units such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) have improved upon this and addressed the problem. While traditional RNN applies a transformation to a weighted sum of inputs in equations 1.3 and 1.4, an LSTM-based recurrent layer creates a memory cell m at each time step whose activation is computed as:

$$h_t = p_t \Theta(m_t), \quad (1.5)$$

where p_t denotes the output gate which determines the portion of the memory cell content in time step t (m_t) to be exposed at the next time step [5, 16]. The expanded, recursive version of updating p_t is as follows:

$$p_t = \sigma(\mathbf{W}_{oi}x_t + \mathbf{W}_{oh}h_{t-1} + \mathbf{W}_{oc}m_{t-1} + b_o), \quad (1.6)$$

In the above equation $\sigma(\cdot)$ is the logistic sigmoid function, W_{oi} is the input-output weight matrix, W_{oh} is the hidden layer-output weight matrix, and W_{oc} is the memory-output weight matrix. Each memory cell, m_t , is updated by the sum of new content, current value of m_t , and discarding part of the present memory:

$$m_t = \mathbf{i}_t \cdot \bar{m}_t + \mathbf{f}_t \cdot m_{t-1}, \quad (1.7)$$

where \cdot is an element-wise multiplication and \bar{m}_t is calculated as:

$$\bar{m}_t = \Theta(\mathbf{W}_{mi}x_t + \mathbf{W}_{ch}h_{t-1} + b_c), \quad (1.8)$$

In equation 1.8, since W term represents weight matrices, W_{mi} is the input-memory weight matrix. Input gate i denotes the degree that new information is to be added and forget gate f determines the degree current information is to be dismissed, as follows:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{ix}x_t + \mathbf{W}_{ih}h_{t-1} + \mathbf{W}_{ic}m_{t-1} + b_i); \quad (8)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{fx}x_t + \mathbf{W}_{fh}h_{t-1} + \mathbf{W}_{fm}m_{t-1} + b_f).$$

All weight matrices, W , and biases, b , are shared between cells across time. A graphical model of an LSTM cell is shown in figure 1.9. GRU is a slightly different structure of LSTM with a fewer number of parameters to avoid over-fitting in models with a low number of training samples [10]. The forget and input gates are combined into a single update gate, known as u , and merge the cell memory and hidden state to a reset gate, r . Also, the activation of the hidden layer in GRU

is an interpolation between the updated activation, h_t , and the previous activation, h_{t-1} :

$$h_t = (1 - u_t)h_{t-1} + \bar{h}_t u_t, \quad (1.9)$$

where u_t denotes the amount of update for the unit content. The update gate formula is:

$$u_t = \sigma(\mathbf{W}_{ui}x_t + \mathbf{W}_{uh}h_{t-1}), \quad (1.10)$$

Given that W_{ui} is the input-update weight matrix and W_{uh} is the update-hidden weight matrix, the updated activation, h_t , will be computed like the traditional RNN in Equation 1.3 as follows:

$$\bar{h}_t = \Theta(\mathbf{W}_{oi}x_t + \mathbf{W}_{oh}(r_t \cdot h_{t-1})). \quad (1.11)$$

Finally, the reset gate, r_t , is computed as:

$$r_t = \sigma(\mathbf{W}_{ri}x_t + \mathbf{W}_{rh}h_{t-1}). \quad (1.12)$$

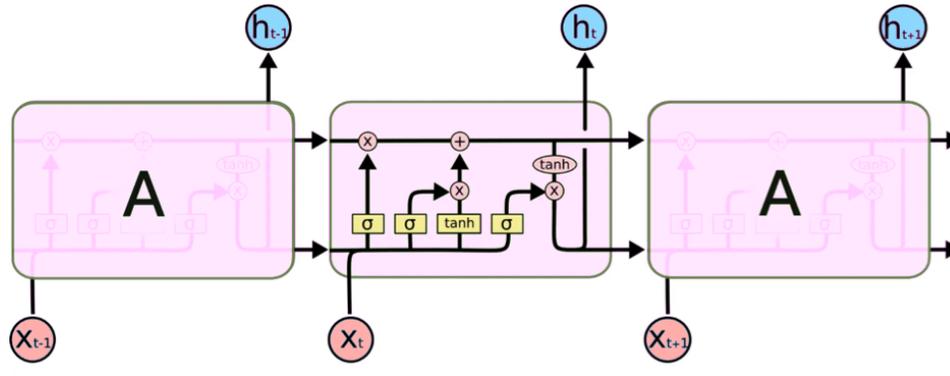


Figure 1.9: Demonstration of an LSTM block. m_{t-1} stands for the input from a memory cell in time point t ; x_t is an input in time point t ; h_t is an output in time point t that goes to both the output layer and the hidden layer in the next time point.

1.2.4 Hyper-Parameter Optimization

Most of the learning algorithms are trained based on a set of hyper-parameters that affect the performance of the model. Generally, hyper-parameters are selected to minimize the generalization error. This objective is essentially done by running different trials of diverse sets of hyper-parameters, comparing the output performances and desiring the best setting. Many approaches have been suggested for hyper-parameter optimization. The most straight forward one is the grid search. Grid search is an exhaustive search in a specified subset of the hyper-parameter space of a learning algorithm [6]. On the other hand, random search replaces the exhaustive enumeration of all combinations by a random selection of them. While Bayesian optimization [53] builds a probabilistic model that maps from hyper-parameter values to the evaluated measurements on a validation set, evolutionary hyper-parameter optimization follows a procedure inspired by the biological concept of evolution [7]. Also, gradient-based optimization [35] and population-based training (PBT) [32] are other methods for hyper-parameter optimization.

1.2.5 Applications of ML in WMA detection

Machine learning with its powerful capabilities has been constantly used for wall motion abnormality detection. The ML methods for wall motion abnormality classification include radial basis functions [11], random forest [13], unsupervised multiple kernel learning [51], dictionary learning [44], support vector machines (SVM)-based wall motion classification (in CMR) images [36]. Recently, deep learning has shown a remarkable role in the classification of several diseases in many medical fields [15, 20, 26]. Conventional machine learning methods mainly require pre-determined features and measurements to identify relevant hidden information in the images [25]. However, deep learning extracts useful features automatically [4, 30]. Moreover, the capability of the deep convolutional layers in extracting low-level features from the original input image is useful in pathology detection in echocardiographic images. Recent research has shown that the Deep Convolutional Neural Networks (DCNN) can be useful for RWMA detection in the clinical setting [28]. The study provides predictions only on coronary infarction territories, using conventional two-dimensional echocardiographic images.

In a work on 3D stress echo [41], Omar et al. proposed a CNN to distinguish between normal and abnormal wall motion. Some works have developed statistical spatio-temporal cardiac atlases, mainly relying on cardiac motion and shape priors [3, 47, 48, 55, 56, 61]. Peressutti et al. extract clinically relevant features by using a motion atlas with non-motion information [45]. Oktay et al. proposed learning cardiac image representations using anatomically constrained neural networks [40]. Also, in CMR, wall motion assessment is done by first segmenting and then feature tracking and strain estimation [49]. Current available DCNNs trained on 2D echo images are only predicting the overall wall motion score as normal and abnormal and none of the above methods predict the severity of abnormality of each region separately. We hypothesize that a deep convolutional neural network trained with echocardiographic images may provide improved detection of RWMA in addition to prediction of the abnormality for all 16 segments in the LV. Figure 1.10 shows a systematic architecture of the proposed method.

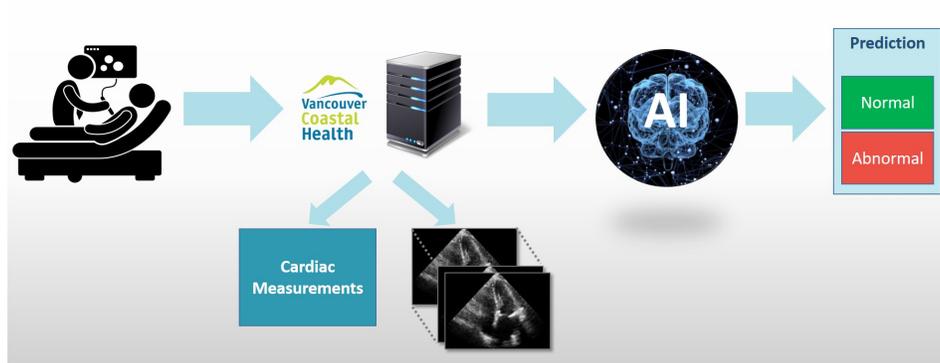


Figure 1.10: A systematic diagram of the workflow for automated RWMA prediction. After data are acquired in the clinic, it is stored in the Vancouver Coastal Health (VCH)’s servers along with the cardiologist’s measurements. The machine learning model will then use these data to make predictions on RWMA.

1.3 Thesis Objective

It is clinically crucial to have an automatic system for evaluation of regional wall motion abnormality that is more objective, consistent, broadly accessible to less trained echocardiographers, and is as precise as an expert. Also, this system could be used when an experienced echocardiographer is not available or could be used as a second opinion by experts. Hence, this system is a practical tool for automated regional wall motion evaluation that scores left ventricular segments as precise as visual determinations by expert echocardiographers. It can be utilized as an extra, broadly used qualitative method for the detection of wall motion abnormalities and as a screening tool for the novice echocardiographers. Moreover, this tool can be used for teaching, unified regional scoring and routine objective evaluation of regional wall motion.

1.4 Contributions

Our research goal has been to develop a framework for regional wall motion abnormality analysis on echo imaging information. To reach this goal the following contributions were made:

Initially a thorough study of the heart and relevant cardiac diseases were performed and the diagnostic imaging techniques were reviewed. In order to improve the diagnosis of wall motion abnormalities and help the cardiologists on this matter, we decided to develop an artificially intelligent model to assist in the identification of the wall motion abnormalities through echo images. This is of a great value given the variability and challenges of scoring such abnormalities.

Subsequently, I developed a machine learning framework for training a neural network, consisting of three views of echocardiography data as input, with abnormality classification for 16 wall segments as output. The network directly analyzes echo data without any need for prior segmentation of the cardiac LV wall.

I trained the network with the data described above from 489 patients and optimized the parameters and hyper-parameters of it. The resultant network precisely identifies regional wall motion abnormalities on echo as compared to the expert human labels with advanced echo training. In independent test dataset, I demonstrated that the neural network can produce accuracy as high as 69.2% for detection

of abnormal wall motion.

The developed model, to the best of our knowledge, is the first to analyze the left ventricle wall motion for both global and regional abnormality detection in echocardiography data. This can be considered a great contribution in enhancing the diagnostic process of the wall motion abnormalities, by expediting the process for physicians.

1.5 Thesis Outline

This thesis covers the background of left ventricle wall motion abnormality, and the relevant technologies for the proposed problem, the details of a system developed, and the evaluation of the proposed system. The outline of the thesis is as follows:

- Chapter 1: Introduction and background

In this chapter, the basic anatomy of the heart is reviewed to gain an understanding of its global function and the importance of the LV. Moreover, a detailed walk-through of the current wall motion abnormality detection system is provided. Furthermore, relevant machine learning techniques used in ultrasound are demonstrated.

- Chapter 2: Materials In this chapter, the dataset obtained from Philip's Xcelera™ and Filemaker™ systems is explained. This dataset is used to train models for the proposed framework. The echo data is mainly acquired from Xcelera™, extracted from routine studies since 2005 at Vancouver coastal health clinics by different ultrasound machines. Consequently, the patient's clinical measurements are acquired from the Filemaker™ database; this dataset contains over 200,000 records and is used to label diseases available in the corresponding echo files.

- Chapter 3: Methods

In this chapter, a deep learning model is proposed to extract features and then classify the 16 segments of LV. The model is used to extract Spatio-temporal features from the echo cine loops. Hence, the proposed model can accurately identify regional wall motion abnormalities on echo images as compared to

the reference standard, experienced human interpreters with advanced echo training.

- Chapter 4: Experiments and Results

To evaluate the presented framework on 489 unique patients, acquired from the routine cardiology care, three streams for tri-plane assessment are fed to the network. The inputs to the network are cine loops captured in the A2C, A4C and PLAX views. The cine loops consist of one full cardiac cycle and are synchronized based on the cardiac phase. The expert annotated RWMA labels will be provided in a supervised learning framework and the network will hence be trained to map the LV regions to a motion score. Then the evaluation results of the proposed framework are reported. Using the proposed model, we investigate the advantages of using a joint-feature model over a single view information model. Finally, we investigate the correlation of occurrence of the abnormality in the 16 segments of LV.

- Chapter 5: Conclusion

This chapter summarizes the objectives of the research and the contributions made and describes potential applications and directions for this research to be continued.

Chapter 2

Materials

The wall motion abnormality detection framework uses the echocardiography dataset obtained from Vancouver Coastal Health (VCH). The datasets and the software applications used for interfacing them are briefly described below.

2.1 Ethics

Ethics approval for evaluation of 3000 studies acquired from routine cardiology care was obtained from the Clinical Medical Research Ethics Board of Vancouver Coastal Health (VCH) and in consultation with the VCH Information Privacy Office. A meticulous process of data anonymization, de-identification and data encryption was fulfilled based on guidelines recommended by the VCH Privacy Office. The patient information such as age, sex, height, weight and related health issues and the echo images were assigned an alpha-numeric code to eliminate any risk of re-identification of participants. The data is encrypted using on-the-fly encryption software named TrueCrypt™ (TrueCrypt.Org, Czech Republic). While the development of TrueCrypt has been discontinued, an independent audit of the software (published in March 2015) has confirmed that there are no significant flaws in the software [1].

2.2 Echocardiography Data

The presented framework is part of the Information Fusion for Echocardiography (INFUSE) project. The data consists of 992 echo studies from 953 unique patients. It consists of two parts: an echo image database and the corresponding measurements and pathology reports. The echo image dataset consists of 2005 to 2015 of Vancouver echo cine videos, measurement screenshots, Doppler cine files and images.

To access the ultrasound studies, the cardiology department's Xcelera™ database is interfaced. It allows for downloading information associated with patient follow-ups, emergency, and investigational echo studies. Each study contains an echo cine that has a variable number of frames, where the mean number of frames is 48. Besides, the collected studies are generated from seven different ultrasound machine models: Philips iE33, GE Vivid 7, Vivid i, Vivid E9, Sequoia, and Sonosite.

The Digital Imaging and Communications in Medicine (DICOM) studies coming from these devices are uploaded to VGH's cardiology department's Xcelera™ server. By the time the upload for echo data is completed, it can be accessed through an Xcelera™ workstation terminal.

2.2.1 Echocardiography Retrospective Data

The Xcelera™ ultrasound software allows users to have access to all saved echo studies through a Graphical User Interface (GUI). The Xcelera™ software also works with DICOM images and can be used as both a server and a viewer for them. In addition to that, it incorporates some advanced features for both cardiologists and sonographers. The features include the image measurement module in the software suite that authorizes the cardiologist to draw or measure the cardiac parameters directly on medical images and save them for future examination. The measurement module has no automatic section and every measurement input to the Filemaker™ database is assessed manually.

The Xcelera™ software contains two separate databases, Echo and Xcelera. The downloaded echo information from the ultrasound machine is written to echo MySQL instance. Moreover, the manual segmentation information made on an image undertaken by a cardiologist is stored in MySQL instance named Xcel-

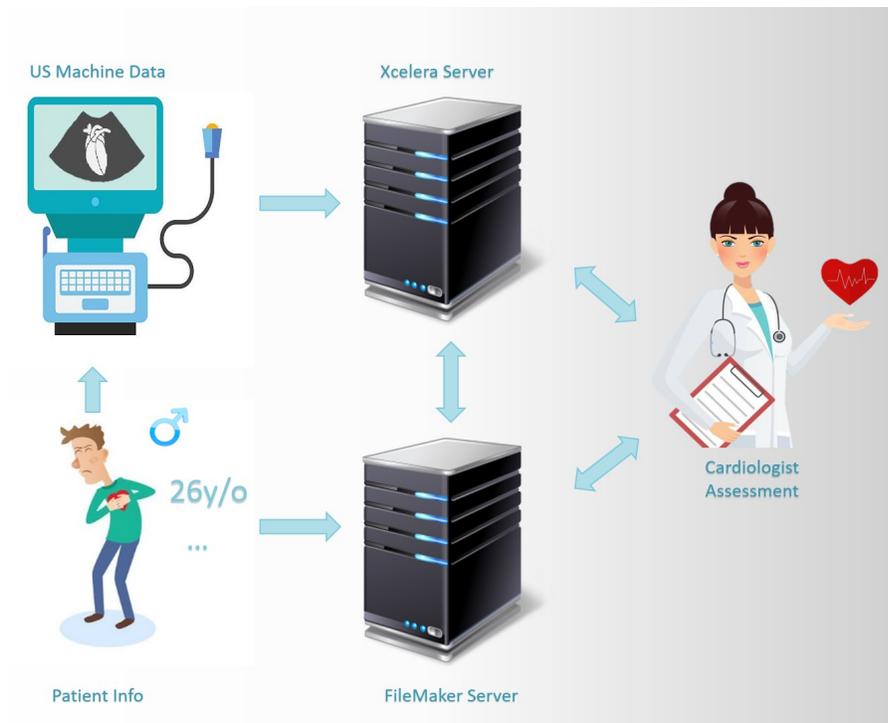


Figure 2.1: A diagram designed to show the relationship between the Echo, Xcelera and Filemaker databases in a routine cardiology study.

era. Xcelera stores patient information keys, which connects both SQL instances. Hence, whenever the measurements are evaluated, they will be stored with a unique identifier.

The key identifier will then be used to link each particular ultrasound, allowing each study to be reopened anytime with the correct measurements attached. The two databases are connected using a key matching method for this large-scale data. For the purpose of this thesis, we only use the Xcelera™ database to obtain the Echo. Figure 2.1 shows an echo saving routine using Xcelera™ (Philips Healthcare, Netherlands)’s database.

2.2.2 Echo Data Download and Processing

Data acquisition is done with the help of the Information Technology team of VGH. A MySQL instance of Xcelera™ is being copied and anonymized and then it will be installed on a computer located at Vancouver Coastal Health's IT department. A subset of available data is then queried. All available echo studies from VGH's Picture Archiving and Communication System (PACS) are copied to Robotics and Control Lab (RCL) servers, encrypted and secured with password as required by our ethics application.

Once all of the data are acquired, a new MySQL database on University of British Columbia (UBC) RCL servers is created. This database contains all the data that have been downloaded from Xcelera™ software in VGH for different projects at RCL.

Each study file located on the RCL servers contains the echo volume pieces of information along with manufacturer name, file name, date of the study, Medical Record Number (MRN), DICOM header information and if any available segmentation coordinates.

Each study is an echo cine series that contains a variable number of frames, where the mean number of frames is 48. Besides, the collected studies were generated from seven different ultrasound machine models: Philips iE33, GE Vivid 7, Vivid i, Vivid E9, Sequoia, and Sonosite. Therefore, the resolution, size of the ultrasound visual area, the probe specification, and imaging settings vary across the machine models. Given these differences, the echo volumes are processed by applying a semi-automatically cropping ultrasound beam shaped black mask, and the frame size is reduced to 128×128 pixels.

2.3 Analytical Measurements

The measurements associated with each study of all patients are stored on VGH's FileMaker™ Pro 6 database. All the findings including diagnostic information for standard measurements, comments, etc along with patient information (e.g., name, age) and exam information (e.g. date, examiner) are entered manually by the sonographer to the FileMaker. Besides, the exams in FileMaker can be linked to the downloaded echo files using a hospital-assigned patient number and the date

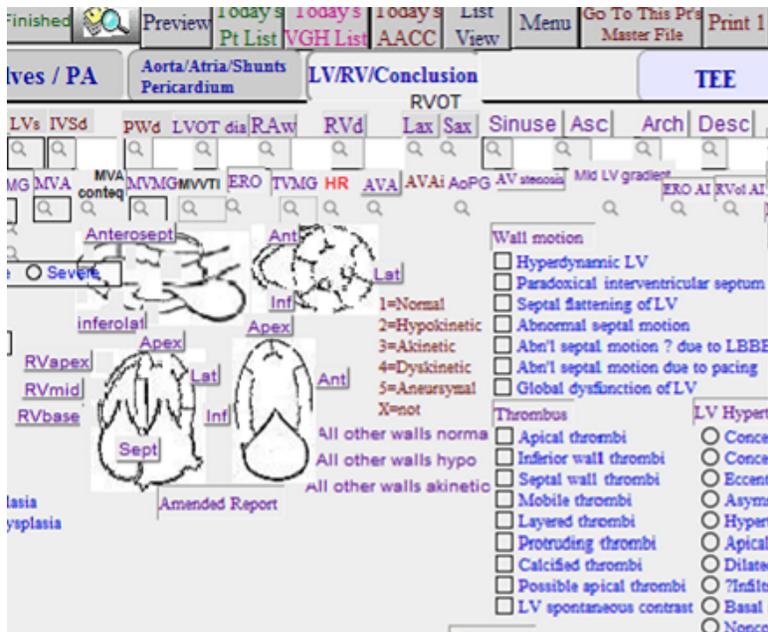


Figure 2.2: Snapshot of the wall motion abnormality section of FileMaker. By clicking on each region, the cardiologist will type a score from 1 to 5 to each region to label the abnormality of the region.

in which the study has been recorded.

2.3.1 Filemaker Measurements

Filemaker™ is a relational database program that integrates a database engine and a graphical user interface (GUI) while keeping security features. It allows all users with a minimal level of technical knowledge to modify the database by dragging new elements into layouts, screens, or forms.

Considering that the Xcelera™ (Philips Healthcare, Netherlands)'s cardiology suite provides searching by MRN only, a custom version of Filemaker™ was created to allow advanced searching by patient physiology in addition to MRN. Thus, this software has a remarkable role within the cardiology department. Moreover, FileMaker™ software is a precious teaching tool for cardiologists because the studies can be retrieved based on certain keywords.

Within the software, there are four main tabs for each echo study. After each

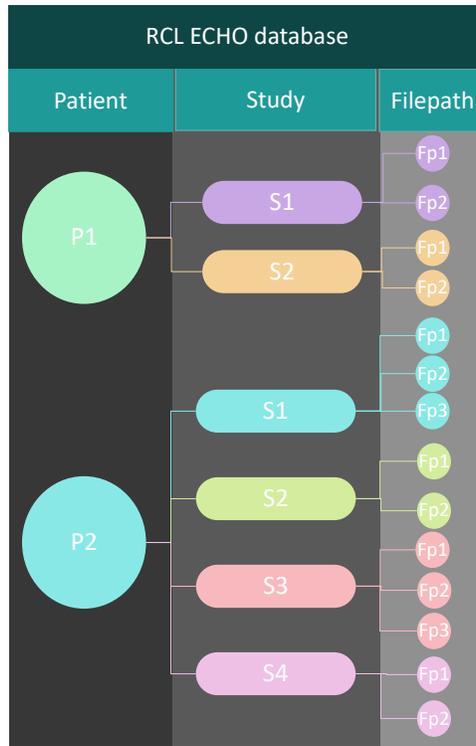


Figure 2.3: A block diagram of the relationship between the Echo database tables.

study, the sonographer fills out the front page, Valves/PA, Aorta/ Atria/Shunts Pericardium pages. As seen in Figure 2.2, there are check boxes and empty values for their report. The cardiologist will then fill out the last tab (LV/RV/Conclusions) after reviewing the first three tabs and the echo images.

2.3.2 Analytical Data Download and Processing

The only process of acquiring FileMaker™ data for this thesis was through exporting the Filemaker™ database into a Comma-Separated Values (CSV) file, where each row represented a unique study and each column represented a unique field.

This CSV file was then imported into an internally created RCL Echo database. Figure 2.3 shows the connection of each table in that database.

As it is depicted, patient table contains all the patients available in the CSV

and is linked to the study table in a one-to-many relationship. The study is then linked to the filepath table in a one-to-many relationship as well. The study table contains all available studies within the FileMaker that we have the ethics for them. Moreover, at the time of each study, there may be many files recorded, these files with their filepath are saved in the filepath table. The decisive components for connecting each file to the correct study is the patientID and the date of the study. If the patient's MRN and date of study in both the echo study file and study table match, the file will be linked to that specific matched study. Besides, once the view classification label of each file is ready it is linked to the filepath table as a one-to-one relationship.

Specific measurements in the FileMaker™ database that were needed for wall motion detection including:

- Global wall motion
- Regional wall motion
 - A2Cbasalant
 - LAXbasalant
 - A4Cbasosept
 - A2Cbasalinf
 - LAXbasalinf
 - A4Cbasolat
 - A2Cmidant
 - LAXmidant
 - A4Cmidsept
 - A2Cmidinf
 - LAXmidinf
 - A4Cmidlat
 - A2Cdistalant
 - A4Cdistalsept
 - A2Cdistalinf
 - A4Cdistallat

were filtered out by creating a query from the database.

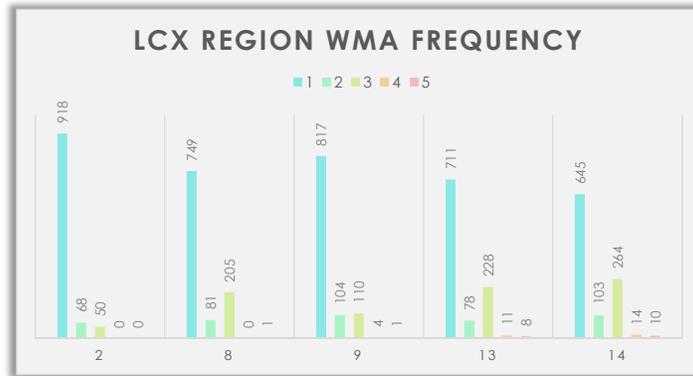
The distribution of regional and global wall motion abnormality from the available studies has been plotted in Figure 2.4, and the frequency of the Wall Motion Score Index (WMSI) is reported in Table 2.1.

Table 2.1: The Wall Motion Abnormality data frequency in each region.

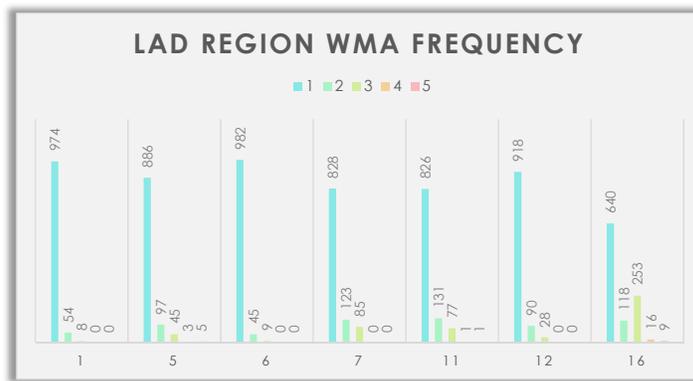
Region	WMSI				
	1	2	3	4	5
Global	541	231	13	0	251
1	974	54	8	0	0
2	918	68	50	0	0
3	909	74	51	1	1
4	817	109	99	2	9
5	886	97	45	3	5
6	982	45	9	0	0
7	828	123	85	0	0
8	749	81	205	0	1
9	817	104	110	4	1
10	783	144	107	0	2
11	826	131	77	1	1
12	918	90	28	0	0
13	711	78	228	11	8
14	645	103	264	14	10
15	715	92	214	6	9
16	640	118	253	16	9

2.4 Summary

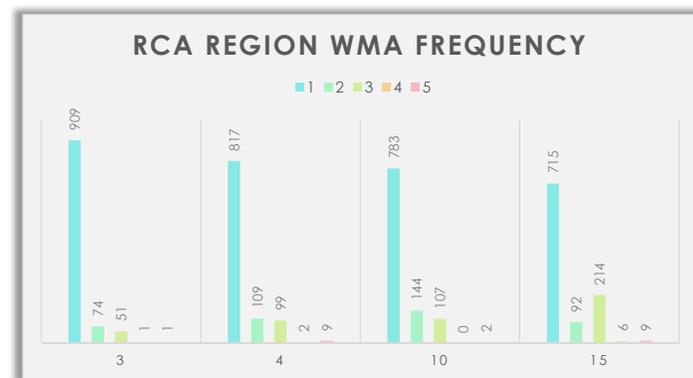
A full description of the dataset obtained from Philip's Xcelera™ and Filemaker™ systems was presented. This dataset is used for the training stage of all models proposed in the next chapter. The echo data are mainly acquired from Xcelera™, extracted from routine studies from 2005 to 2015 at Vancouver Coastal health clinics by different ultrasound machines. Moreover, the patient's clinical measurements are downloaded from the Filemaker™ database; this dataset contains over 200,000 records and is used to label diseases available in the corresponding echo files.



(a) LCX RWMA data distribution



(b) LAD RWMA data distribution



(c) RCA RWMA data distribution

Figure 2.4: The Wall Motion Abnormality data distribution in different regions of the heart: (a) is data distribution in LCX region, (b) is data distribution in LAD region, and (c) is data distribution in RCA region.

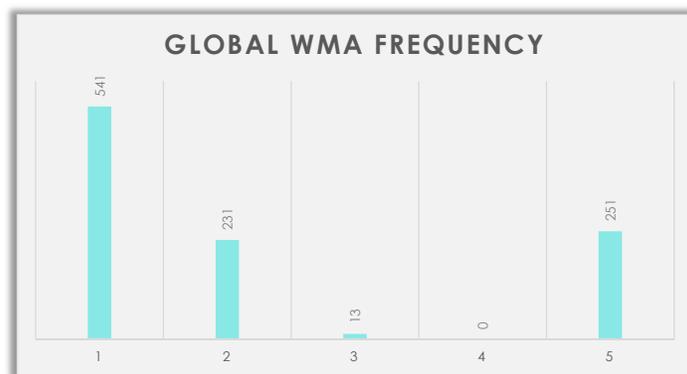


Figure 2.5: Global WMA data distribution.

Chapter 3

Methods

The proposed framework for echo analysis to identify RWMA consists of three phases of view processing, classification, and localization. The objective of the view processing phase is to automatically distinguish the view in which the study has been recorded. In the classification phase, after retrieving key echo views required for the visualization of RWMA (parasternal long-axis, apical 2-chamber, and apical 4-chamber), these clips will be coded for regional and global wall motion abnormalities for training and validation of a convolutional neural network.

3.1 Relevant Cardiac Echo View Selection for WMA

In Point-Of-Care Ultrasound (POCUS), apical four-chamber (AP4), parasternal long axis (PLAX), parasternal short axis at the papillary muscle level (PSAX-PM), and subcostal four-chamber (SUBC4) are the four views most frequently acquired by clinicians. Since RWMA is visible in both parasternal short axis and apical views, we decided to continue with the apical views due to the popularity. For the determination of relevant cine loops, a pre-trained deep learning network is used to predict any of the 14 views. The model was trained on a dataset of 3,151 unique patients who were diagnosed with various heart conditions and diseases during the period of 2005 to 2015. Generally, the dataset contains 16,612 echo cines (with a total of 807,908 frames) from cardiac standard views taken from the four standard imaging windows, namely, Apical, Parasternal, Subcostal, and Suprasternal. The

View	A2C	A3C	A4C	A5C	PLAX	RVIF	S4C	S5C	IVC	PSAX-A	PSAX-M	PSAX-PM	PSAX-APEX	SUPRA
Training set	335	283	359	128	390	131	172	29	218	401	388	187	63	46
Validation set	126	101	105	42	131	49	77	5	67	135	137	60	19	13
Test set	106	95	93	44	137	29	49	15	56	108	147	73	13	11

Table 3.1: The dataset composition in terms of number of studies from each type of the 14 standard echocardiography views (A#C: apical #-chamber view, PLAX: parasternal long-axis view, RVIF: right ventricular inflow view, S#C: subcostal #-chamber view, IVC: subcostal inferior vena cava view, PSAX-A: parasternal short-axis view at aortic valve, PSAX-M: PSAX view at mitral annulus valve level, PSAX-PM: PSAX view at mitral valve papillary muscle level, PSAX-APEX: PSAX view at apex level, and SUPRA: Suprasternal view).

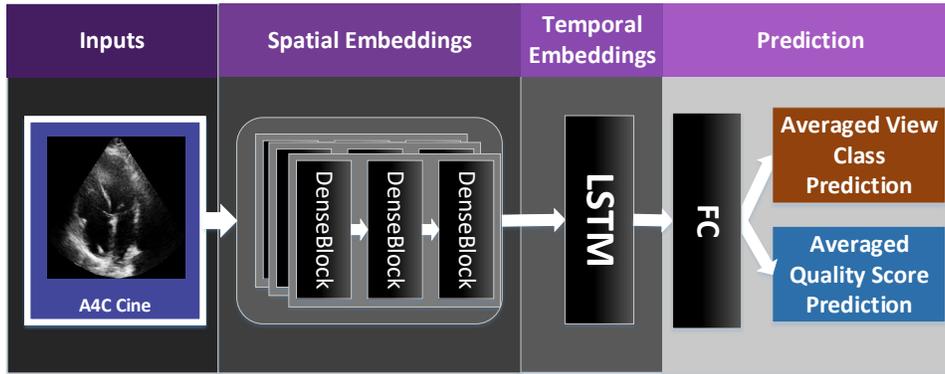


Figure 3.1: The cardiac view classifier network architecture. Related embedding are extracted from the individual frames by the Spatial embedding extractor (DenseNet blocks). The embeddings are then fed into the Long Short-Term Memory (LSTM) blocks to extract the temporal information across 10 sequential echo cine frames.

distribution of the data per class is shown in Table 3.1, and an example of dataset images (one per view class) can be found in the table:

The network architecture is demonstrated in Figure 3.1. The input is a 10-frame randomly extracted from an echo cine array, where each frame is a 120×120 pixel, gray-scale image. As shown in the Figure 3.1 the model consists of four components:

- A seven-layer DenseNet model that extracts per-frame features from the in-

put;

- an LSTM layer with 128 units to capture the temporal dependencies from the generated DenseNet features;
- a regression layer that produces the quality score from the output feature of the LSTM layer for each frame;
- a softmax classifier that predicts the content view from the LSTM features for each frame.

The best accuracy of 89% is achieved by an ensemble of the three very deep models. It is noteworthy that the average performance of the model for A2C, A4C and PLAX view is 89%, 93%, 96%, respectively[60]. For the purpose of this thesis, we only input the samples to the trained model and get the predicted view per study.

3.2 Automatic WMA Detection Deep Convolutional Network

We propose a regression model for WMA detection. The network architecture can be seen in Figure 3.2. The input to the network is sampled from synchronous A2C and A4C and PLAX echo cines. This network consists of two main stages: a convolutional layer and a fully-connected layer. The first stage is composed of convolutional layers (Conv) and pooling layers (Pool); the second stage only contains Fully-Connected Layers (FC).

3.2.1 Spatial Feature Extractors

The convolutional layer mainly consists of 2D or 3D kernels which are convolved with the input images and return the spatial feature-maps (cine representations). The kernels perform a discrete convolution (i.e. a weighted sum of inputs). For instance the 2D kernels of size $2p + 1 \times 2p + 1$ do the following:

$$x_{i,jk}^l = \sum_{m=-p}^p \sum_{n=-p}^p w_{i,mn}^l x_{(j+m)(k+n)}^{l-1}. \quad (3.1)$$

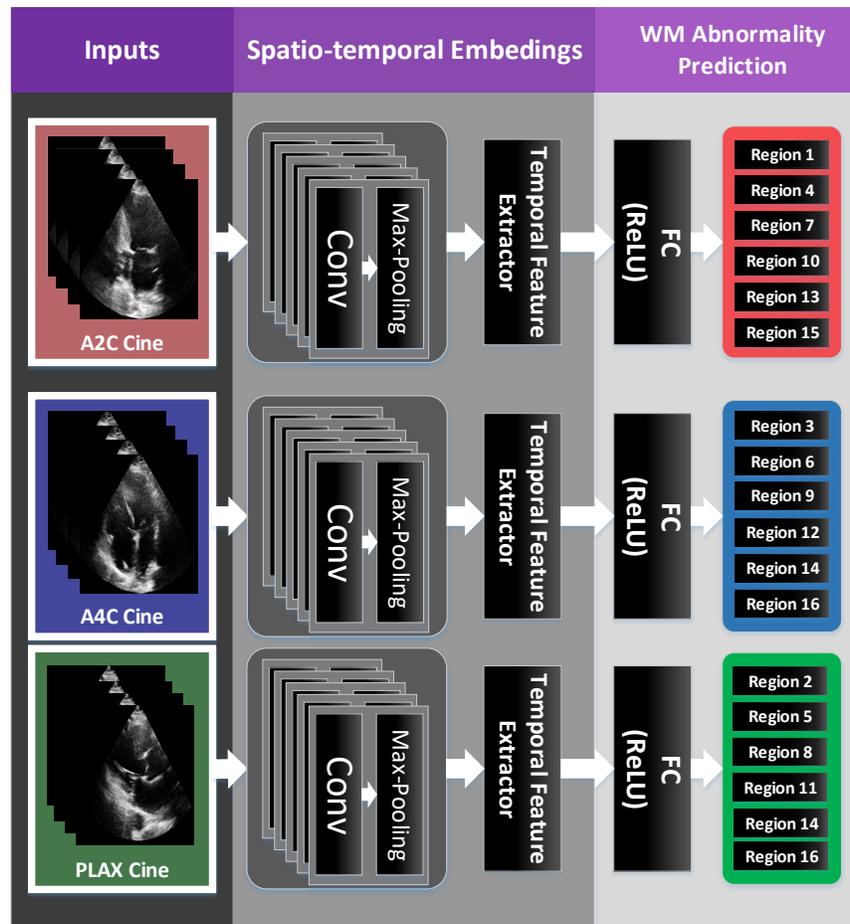


Figure 3.2: The proposed WMA network architecture. Spatio-temporal embeddings are extracted from the individual cines. The embeddings are then fed into the FC blocks to connect the reasoning between the information across echo cine frames.

In the above equation, w_i^l is the weight matrix, and x_i^l is the output feature-map of the i^{th} kernel of the Conv layer l ; x_1^l denotes the input feature-map of the layer. The feature-maps of the previous layer is convolved with the kernel and result in a 2D output. All of the outputs from kernels are stacked and generate a 3D output feature-map.

The total number of parameters in the convolutional layer is calculated by multiplying the number of kernels by the size of each kernel. Compared to FCs a convolutional layer has fewer parameters since only a single kernel is generating the feature-maps. Thus, the convolutional layer has considerably fewer parameters to be optimized.

The output feature-maps are passed to a non-linear activation function. The Rectified Linear Unit (ReLU) is used on all the feature-maps. While having the same performance, the ReLU function is much faster than its rivals such as sigmoid and hyperbolic tangent functions [39]. This function is described below:

$$f(x) = x^+ = \max(0, x) \quad (3.2)$$

A non-linear form of down-sampling is pooling. The spatial variance of feature-maps is reduced using pooling layers. This allows faster convergence in addition to picking the most relevant features [52]. Since pooling layers reduce the size of feature maps, the computation time is reduced as well. Besides, adding pooling layers to the model will increase the generalization as they make the model robust to small translations. Between several methods of pooling, max-pooling has shown a remarkable performance in comparison to its other alternatives [9]. Max-pooling layers have no corresponding weights; thus, the number of parameters in this layer is zero.

3.2.2 Temporal Feature Aggregators

Fully connected layers perform the high-level reasoning in the neural networks, mainly by representing the connection between the spatial feature-maps. In an FC layer, each neuron is connected to every neuron (or activations) in the previous layer. Mathematically, the FC layer is denoted by a matrix multiplication followed

by summation of offset value.

$$f_i^l(x^{l-1}) = \sum_{j=1}^n w_{ij}^l x_j^{l-1} + b_i^l, \quad (3.3)$$

where w_{ij}^l denotes the j -th weight in neuron i of layer l , and b_i^l is the bias value.

The output of an FC layer is passed to the activation function as well as the output of convolutional layers. However, the output of the last FC layer is not filtered by an activation function since it would be the final prediction of the network.

3.2.3 Spatio-temporal Framework

There are several spatio-temporal convolutional variants within this framework. Given that the input is gray-scale cine files of the size $3 \times f \times H \times W$, where 3 is the number of view channels, f is the number of frames, H and W are the frame height and width equal to 128. In the following, different convolutional blocks that are used for feature extraction, will be discussed in detail.

2D convolutions over the cine

One approach would be using 2D convolutions over the entire cine. 2D CNNs disregard the temporal nature of cine files and behave like there is no difference in the relative occurrence of each frame. Thus, this model will reshape the 4D input to a 3D input with a size of $3f \times H \times W$.

The output of the convolutional block is also a 3D tensor. Its size is $N_i \times H_i \times W_i$, where N_i is the number of convolutional filters used in the i^{th} block, and H_i, W_i are the spatial dimensions, which may be smaller than the original input frame due to pooling or striding. The filters are 3D and have the size of $N_{i-1} \times d \times d$, where d is the spatial width and height of the filter. It is worth mentioning that although the filter is 3D, it is convolved spatially in only two dimensions. Consequently, the output of each filter is a 2D output, meaning the temporal information vanishes in the first layers of convolution. This may result in a lack of temporal reasoning in the subsequent layers.

2D convolutions per frame

Another 2D CNN approach is processing all f frames via a series of 2D convolutional blocks. The same filters are applied to all f frames. There is no temporal modelling in the convolutions, however, the global spatiotemporal pooling layer combines the individual information from each frame. This architecture is illustrated in Figure 3.3 as C2D.

3D convolutions over the Cine

3D CNNs extract features in spatial and temporal dimensions by performing 3D convolutions, so they capture the motion information within multiple adjacent frames. The output will be a 4D tensor with the size of $N_i \times L \times H_i \times W_i$, where N_i is the number of filters in i -th layer. The filter size is $N_i \times t \times d \times d$, where d and t denote the spatial and temporal extent of the filter respectively. Moreover, the filters are convolved in both time and space. This architecture is depicted in the Figure 3.3.

2D convolutions+temporal feature extractor over the Cine

The more promising approach is using the 2D convolutions followed by LSTM or GRU units. In other words, the architecture consists of both spatial and temporal feature extraction elements in separate stages. The convolution part can be a plain stack of convolutions or sequence of DenseBlocks [22]. The feature space output of the convolutional blocks is then flattened and fed to the temporal feature extractor part of the network (i.e. LSTM and GRU units). Thus, the output of feature extractor is $X_m^{View_i}$ of length $M \times 1, m = 1 : M$ where $View_i$ is one of the three views used for the WMA detection. The output of all view channels (i.e: $X_m^{A2C}, X_m^{A4C}, X_m^{PLAX}$) fed to the fc layers. The different structures out of this are illustrated in the Figure 3.3 subsections b to e.

3.2.4 Regularization and Data Augmentation

Improving the generalizability of machine learning models has been one of the most challenging tasks. Generalizability denotes the difference in the evaluated performance of the model on training data versus test data. Models that are overfit-

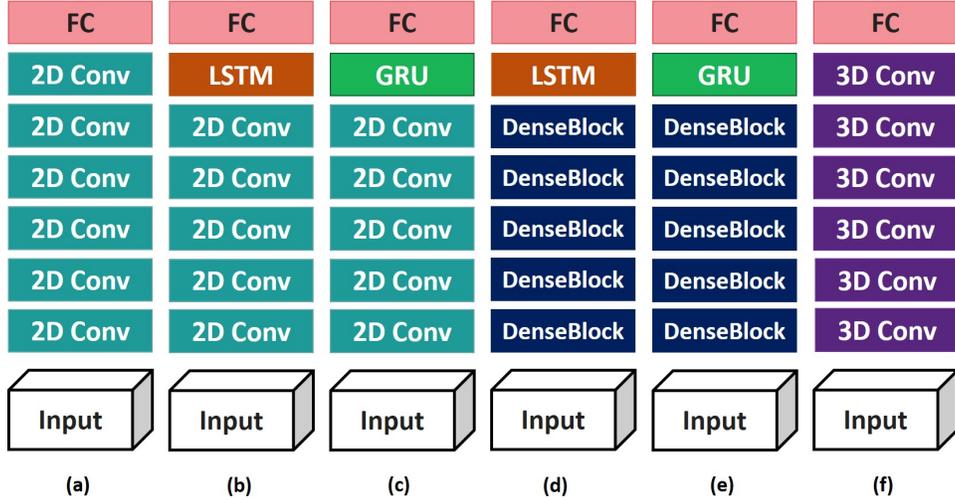


Figure 3.3: The network architectures for WMA classification considered in this work. (a) C2D are 2D Covolutions; (b) C2D+LSTM are time distributed 2D Covolutions followed by LSTM units; (c) C2D+GRU are time distributed 2D Covolutions followed by GRU units; (d) D2D are 2D DenseBlocks followed by LSTM units; and (e) D2D are 2D DenseBlocks followed by GRU units. (f) C3D are 3D convolutions. For interpretability, the connections are omitted.

ted on train data have low generalizability. Monitoring loss and accuracy plot for train and validation set at the end of each epoch is an approach for the detection of over-fitting. Hence, we monitor the plots during training.

There are many strategies to stabilize learning and prevent the model from over-fitting while training. By adding a penalty term to the loss function, it is possible to prevent the coefficients or weights from getting too large. This method is called Regularization. In all experiments, we used a ℓ_2 regularizer term in the loss function in the form of $\lambda \|w\|_2^2$, where $\lambda \in 0.00001 - 0.001$. Moreover, λ will be a hyper-parameter that we will investigate the optimal value for it.

Another method to prevent the over-fitting problem is dropout. Dropout layers limit the co-adaptation of the feature extracting blocks and force the neurons to follow the overall behaviour. In each step, the dropout layer removes some random units from the neurons in the previous layer of the network based on the probability

parameter of the dropout layer as another hyper-parameter. Thus, by removing the units, the network architecture is changed in each training step. This implies that dropout integrates diverse architectures in the model [54]. In other words, a dropout acts like adding random noise to hidden layers of the model. In our design, a dropout layer was deployed after each FC layer. The probability of the dropout can be another hyper-parameter.

3.3 Summary

The model exemplifies the ability of a supervised training algorithm to be applied to the US, to address the clinical need for WMA detection. Different possible structures for extracting features and then classifying the 16 segments of LV were discussed. Convolutional layers are used to extract feature-maps from the echo cine loops and FC layers are used to model the relationship between the feature-maps. Having both spatial and temporal feature extraction blocks, the proposed model can outperform baseline methods in identifying regional wall motion abnormalities on echo images.

Chapter 4

Experiments and Results

In this chapter, a study of wall motion abnormality detection performance using the different spatio-temporal convolutions, described in the previous chapter, is presented. It should be pointed out that the echo dataset used in this study is large enough to enable the training of deep models from scratch over enough number of iterations.

4.1 Experimental Setup

4.1.1 Dataset

The proposed model has a large number of parameters to be trained. Thus, it requires a large annotated dataset. In this research a dataset of 1037 patients collected from the Picture Archiving and Communication System at Vancouver General Hospital is used. The data includes combinations of A2C, A4C and PLAX echo views of each patient. The echo studies are mainly acquired by echo-technicians during routine cardiac exams. However, the ground-truth label of WMA is annotated by expert cardiologists.

In every echo routine, the heart is imaged from different standard imaging views, mainly parasternal long and short axes, apical 2-, 3-, and 4-chamber, sub-costal, and suprasternal, in which the transducer is placed on patient's chest to acquire ultrasound cine files. In this research, the apical two-chamber (A2C), four-

chamber (A4C), and parasternal long axes (PLAX) views are used.

The whole data were divided into a training set and a test set (80:20 split, respectively) so that a total of 953 patients with 992 studies were split to 667 cases (704 studies) as the training set, 95 cases (96 studies) as the validation set, and 191 cases (192) as the test set. The dataset is shuffled randomly and then split into five non-overlapping groups based on the patients. The experiment is done five times, where in each run, one of the five sets is set aside as test data and unseen while training with the other four subsets. Moreover, in each run, the validation portion in the training set is used for searching the optimal hyperparameters.

4.1.2 Wall Motion Abnormality labels

All the labels corresponding to the studies of echo dataset are available in File-Maker. These labels are extracted and an integer score of 1 (normal) to 5 (aneurysmal) was assigned to show the abnormality level of each segment. As discussed in Chapter 1, scores 3, 4 and 5 all correspond to totally abnormal functioning region (akinetic, dyskinetic and aneurysmal). Thus, we consider them all together as abnormal. Distribution of data among the five abnormality-levels is demonstrated in Figure 2.4 for each of the segments.

4.1.3 Network Architecture

To compensate for the differences among patients' frame rate and heart rate, images from only one heart cycle were used. All DICOM (Digital Imaging and Communications in Medicine) images were resized into 128 pixel mat-files. The distribution of WMA labels among the training and test sets were examined using Pearson's χ^2 goodness-of-fit test to verify that train and test set are reasonable representations of the original data (i.e. $p - value > 0.05$).

Three parallel streams of the network with similar architecture in the feature extraction part are trained for each of the three correspondent cardiac views. This disconnects the views from one another, enabling the full use of available information in the cines. While this architecture causes a big increase in the number of parameters in the network, we will show that this structure is successful for WMA detection based on our experiments so far. This is mainly due to having denser

and richer feature vectors from each view, which allows more effective temporal learning.

The network architectures are illustrated in Figure 3.3. Having 5-6 convolutional layers (each followed by ReLU activation functions). All convolutional kernels were convolved with a stride of one on padded inputs to preserve dimensions. Moreover, 2D and 3D pooling layer filters were used afterwards. The features are then fed to FC layers to predict the WMA for each of the regions. Moreover, within each structure there would be another similar channel for optical flow transfer of image is added to add more information about the movement in the videos. The results compare each method with and without optical flow. In results section, if the input data to the method consists of both echo cardiac view and optical flow the method name will have an "OF" at the end of the name of method.

4.1.4 Hyper-parameters

The hyper-parameters are optimized using a grid search. The grid search starts to divide the training set into training and validation sets (80% and 20%, respectively). The loss and accuracy on both training and validation sets are used to find the best combination of the hyper-parameters. Loss is defined as the mean squared error between the predicted label and the true label, while accuracy is the mean absolute error between the ground truth label and the prediction.

We perform a grid search over the initial learning rate, $lr \in \{0.001 - 0.0001\}$ and batch size, using two different optimization algorithms, SGD and Adam. We also experiment with different levels of dropout rate, $dr \in \{0, 0.5\}$ and L2 regularization term (λ) $lreg \in \{0.00001, 0.001\}$. These parameters result in 48 different hyper-parameter settings for the proposed model. All models are trained with the same number of iterations and training is stopped after 1000 epochs.

4.1.5 Regularization and Data Augmentation

There are many strategies to stabilize learning and prevent the model from overfitting while training. By adding a penalty term (regularizer) to the loss function, it is possible to prevent the coefficients or weights from getting too large. In this research, we used a ℓ_2 regularizer term in the loss function in the form of $\lambda \|w\|_2^2$,

where $\lambda = 0.001$. Moreover, λ will be a hyper-parameter that we will investigate the optimal value for it.

Another method to prevent the over-fitting problem is dropout. Dropout layers limit the co-adaptation of the feature extracting blocks and force the neurons to follow the overall behaviour. In each step, the dropout layer removes some random units the neurons from the previous layer of the network based on the probability parameter of the dropout layer as another hyper-parameter. Thus, by removing the units, the network architecture is changed in each training step. This implies that dropout integrates diverse architectures in the model [54]. In other words, a dropout acts like adding random noise to hidden layers of the model. In our design, a dropout layer was deployed after each FC layer. The probability of the dropout is also another hyper-parameter.

Data augmentation is the next approach to prevent over-fitting and add transitional invariance to the model. Therefore, the train samples were augmented on-the-fly while training. In every mini-batch, each sample was translated horizontally and rotated. The number of pixels for translation of the image was generated randomly from a zero-mean Gaussian distribution with a standard deviation of 15% of the image width. Likewise, rotational invariance is added to the images, on-the-fly, with a random degree generated from a zero-mean Gaussian distribution with $\sigma = 25$ degrees and capped to 2σ .

Since the WMA labels are for the original non-augmented files, the maximum for transitional and rotational augmentation is limited. Thus, the upper limit for both transitional and rotational augmentation is estimated by an expert cardiologist on the research team. This is to make sure that data augmentation does not affect the clinical value of the cine files.

4.1.6 Training

Once the hyper-parameters are selected and the architecture of the model is finalized, the proposed network is trained on the entire training set containing both train and validation data subsets. The training was repeated four times not only to emphasize the robustness of the results but to justify the random initialization and different training paradigms. The final performance of the models was evaluated

based on the predictions of the model for the test set. Neither in hyper-parameter selection nor final training test data was not deployed or analyzed in the design of networks. Adam optimizer is used for training the networks.

Adam is an optimization algorithm that can be used extensively instead of the classical stochastic gradient descent procedure to update network weights iteratively. Adam is a gradient-based optimization algorithm using adaptive estimates of lower-order moments for stochastic objective functions. While stochastic gradient descent maintains a single and constant learning rate for all parameter updates during training, Adam maintains a learning rate for each network parameter and adapts separately as learning unfolds.

The suggested parameters for training deep learning models in the paper are:

$$\text{learning-rate} = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \text{ and } \varepsilon = 1e-08. \quad (4.1)$$

A relatively high momentum of 0.9 will be sufficient to reward the persistent reduction in the loss. The optimal initial value for the learning rate in our framework was $1e-5$. After 3000 epochs of training, diagnostic accuracy was calculated using the test set. During training, a small batch-size of 8 triple-cines was favoured.

Early stopping is also deployed to prevent over-fitting. The cost function (loss) is used as a performance measure for it. Training is stopped if the loss does not decrease or increases after 100 epochs. Moreover, an absolute difference in the loss of less than $\gamma = 0.0001$ is considered as no improvement. All the network parameters were randomly initialized using a zero-mean Gaussian distribution. This paradigm as described ensures a substantial convergence of the network. In all trials, training is stopped once the network is converged. Convergence was defined as the state in which no progress was observed in the loss decay. Moreover, since the classes distributions is imbalanced, all the training batch data is sub-sampled using a stratified sampling.

4.1.7 Implementation

Keras was first developed as part of the project Open-ended Neuro-Electronic Intelligent Robot Operating System (ONEIROS) and its main author is François Chol-

let, a Google engineer. [keras.io. Retrieved 2018-02-23.] It is an open-source neural network library in python that is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. Keras was used for training and testing of the proposed network on the Tensorflow backend with Python version 2.7 programming language (Python Software Foundation, Beaverton, Oregon). The experiments were undertaken using Nvidia GeForce GTX 980 Ti GPU with 2816 CUDA cores and a GPU clock of 1 GHz, featuring from the CUDA runtime platform version 8.

4.2 Results

The study population consisted of 953 patients with at least one malfunctioning LV region. Since, the frequency of abnormal cases is roughly 20% of the whole data, reporting only the overall accuracy measurement would not be sufficient for evaluating the network performance. Thus, all the numbers reported in the table are the accuracy per each class. Given that we only have two classes of normal and abnormal, the reported accuracy per class would be the specificity and sensitivity respectively.

Table 4.1 shows the co-variance matrix of the 16-segment labels and WMSI. As it is shown, there is a high correlation between WMSI and regions 7, 8, 9, 13, 14, 15 and 16. Therefore, the wall motion score index had a higher co-variance with the apical regions than in the other regions. Figure 4.1 illustrates the value of the loss function in the training and validation sets for training the proposed model, where the horizontal axis is the number of epochs and the vertical axis is the value of loss function. As shown in this figure, the model converges in the training process near the 1000th epoch. It has been seen that SGD and Adam have similar performance on our data while SGD learns faster than Adam optimizer.

Table 4.3 shows that the Densenet network with LSTM cells and optical flow channel(D2D+LSTM+OF) leads to a lower loss value and a higher accuracy in most of the regions visible in A2C view. Besides, same performance have been observed for the other two views with their corresponding regions in the Tables 4.4 and 4.5.

Quantitative results for detection of global WMA obtained in this study are

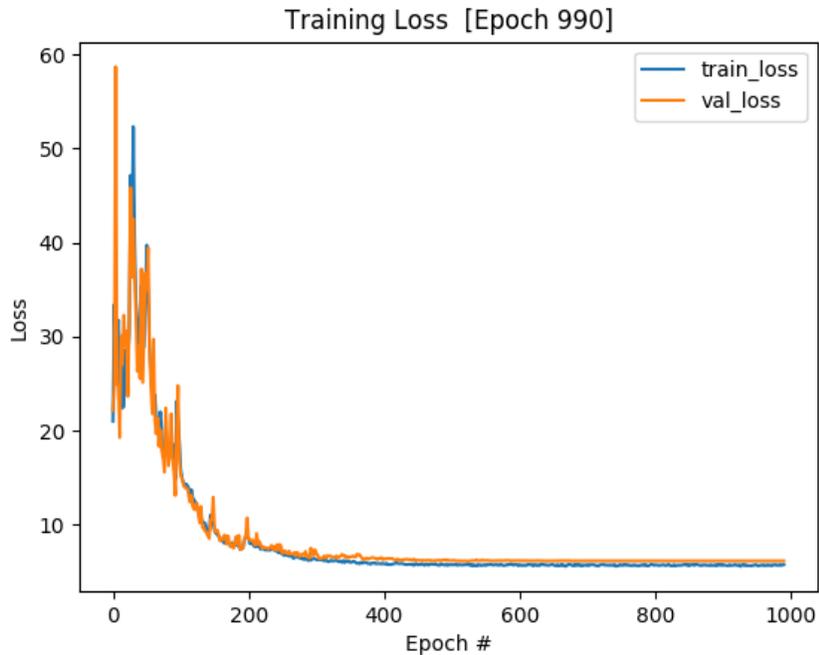


Figure 4.1: The value of loss for training and validation over each epoch.

demonstrated in Table 4.2. The highest performance is achieved using the DenseNets and LSTMs for detection of regional wall motion abnormalities.

As it is seen in the results tables, C3D performs lower than the D2D+LSTM or GRU networks, this difference suggests that the spatiotemporal decomposition of D2D+LSTM makes the optimization easier compared to C3D. The other noteworthy observation, would be the result of adding optical flow transform of the video as a separate channel. However, this doubles the number of parameters, it increases the accuracy per each class in all models, significantly. This proves that the network is analyzing the motion through the cines.

By comparing the results in C2D and D2D structures, it is observed that Dense blocks have been more successful in pertaining the relevant spatial features of the images. A key pattern recognized in our observations is the link between model performance and the quality of images. The performance increases as quality of the cines increase.

	Region1	Region2	Region3	Region4	Region5	Region6	Region7	Region8	Region9	Region10	Region11	Region12	Region13	Region14	Region15	Region16	WMSI	GlobalWM
Region1	0.08																	
Region2	0.05	0.23																
Region3	0.04	0.08	0.25															
Region4	0.03	0.03	0.12	0.43														
Region5	0.03	0.02	0.08	0.18	0.26													
Region6	0.03	0.02	0.03	0.05	0.06	0.08												
Region7	0.08	0.13	0.07	0.02	0.00	0.03	0.37											
Region8	0.06	0.21	0.09	-0.02	-0.03	0.01	0.28	0.65										
Region9	0.05	0.16	0.16	0.06	0.02	0.02	0.18	0.30	0.44									
Region10	0.04	0.05	0.12	0.34	0.16	0.06	0.05	0.04	0.11	0.44								
Region11	0.04	0.02	0.07	0.20	0.22	0.07	0.02	-0.02	0.04	0.22	0.35							
Region12	0.05	0.05	0.05	0.07	0.06	0.07	0.11	0.07	0.08	0.09	0.10	0.18						
Region13	0.07	0.17	0.09	0.02	0.00	0.04	0.33	0.48	0.29	0.08	0.03	0.11	0.72					
Region14	0.05	0.15	0.09	-0.02	-0.02	0.02	0.27	0.49	0.33	0.05	0.00	0.08	0.55	0.78				
Region15	0.06	0.12	0.11	0.12	0.07	0.05	0.26	0.34	0.25	0.18	0.13	0.13	0.55	0.44	0.69			
Region16	0.06	0.14	0.08	-0.01	-0.01	0.03	0.30	0.48	0.29	0.06	0.02	0.12	0.55	0.70	0.46	0.76		
WMSI	0.05	0.10	0.09	0.10	0.07	0.04	0.16	0.21	0.17	0.13	0.10	0.09	0.25	0.25	0.25	0.25	0.14	
GlobalWM	-0.02	-0.06	-0.07	-0.15	-0.09	-0.03	-0.16	-0.26	-0.17	-0.18	-0.14	-0.07	-0.31	-0.38	-0.31	-0.36	-0.17	0.71

Table 4.1: The co-variance matrix of the 16-segment labels and WMSI.

4.3 Discussion and Summary

Several methods have been developed for measuring cardiac wall motion and strain and strain rate. However, a quantitative automatic method is needed to fulfill the inter-observer or intra-observer variability and reproducibility limitations of the those methods. Although strain and 3D echo have superior sensitivity and reproducibility than the routine echo, they require more expertise and several technical difficulties and standardization issues remain. Having portable and handheld echocardiography in emergency departments and intensive care units to evaluate patient hemodynamics results in easier LV wall motion assessment. Moreover, it is more generalizable to a wider range of physicians with various levels of training, expertise and echocardiographic equipment. There are some groups that have proposed automated algorithms for assessment of the the LV function [31]. Though, the majority of them remain semiautomatic, since an observer is needed to input the important landmarks. Thus, a fully automated assessment algorithm is needed to do the WMA assessment without any user interaction. The reported results in the previous section suggest that deep learning with its powerful features in classification is used for development of a fully automated system for wall motion abnormality detection.

The most significant limitation is the difficulty to obtain good images that interpret motion abnormalities clearly. The wall motion scores were assessed by an experienced echocardiographer. While the estimation of the wall motion abnormalities can be subjective, we categorized the patients to normal and abnormal to reduce the observer variability to less than 5%. However, short-axis parasternal views offer the real 3D (360) analysis of cardiac dynamics as opposed to the limited degrees of evaluation obtained with the thin sagittal cuts from the apical views, the acquiring process is much harder for PSAX views. Thus, the measurements from apical views are studied. Given the linear relationship between the model performance and image quality in the cines, misclassified cines generally have unclear LV boundaries, which cause a great deal of variance in the appearance of the heart and its motion. Besides, despite automatic view classification done for these experiments, confusion between the apical views (A2C, three-chamber, A4C and five-chamber) appears to remain a challenge and a potential source of error. Thus,

Method	Global WMA	
	Normal	Abnormal
C2D	53.4%	54.2%
C2D-OF	55.1%	56.2%
D2D[28]	54.5%	54.9%
D2D-OF	56.7%	56.4%
C2D+LSTM	56.9%	55.6%
C2D+LSTM-OF	59.3%	57.1%
C2D+GRU	57.3%	56.0%
C2D+GRU-OF	58.7%	59.1%
D2D+LSTM	64.5%	67.8%
D2D+LSTM-OF	68.8%	70.3%
D2D+GRU	63.3%	66.1%
D2D+GRU-OF	69.4%	69.6%
C3D	61.5 %	62.9%
C3D-OF	65.2%	66.1%

Table 4.2: The Global WMA accuracy per class comparison of experimented methods.

a bottom-up approach for improving WMA accuracy can be through improving the quality of the input data.

The benefit of usage of deep learning to other machine learning methods, is that deep learning creates the matching features needed for classification automatically via its intermediate layers and different structures in deep learning extract different features. Thus, the difference between the deep learning algorithms performance is being addressed in this way and shows why a deep learning model is superior to others.

Since LV localization has been a key step in some ejection fraction estimation approaches proposed for CMR, another approach worth exploring is whether LV

Method	Region Numbers											
	1		4		7		10		13		15	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
C2D	53.0%	53.4%	53.5%	53.9%	54.0%	54.5%	54.4%	54.3%	54.1%	54.6%	53.1%	53.5%
C2D+OF	54.8%	55.1%	55.3%	54.4%	55.1%	55.9%	56.0%	56.5%	55.5%	55.8%	54.5%	56.0%
D2D	54.2%	54.2%	53.0%	54.8%	55.1%	55.9%	54.0%	54.9%	55.2%	55.0%	54.9%	54.8%
D2D+OF	55.7%	55.4%	56.0%	55.1%	56.2%	55.0%	56.7%	56.8%	56.4%	56.6%	55.9%	57.3%
C2D+LSTM	56.4%	54.6%	56.5%	55.5%	57.2%	55.8%	57.3%	56.2%	56.9%	56.0%	56.3%	55.1%
C2D+LSTM+OF	58.5%	56.3%	58.9%	57.1%	59.4%	57.1%	59.9%	57.9%	59.4%	57.7%	59.0%	56.9%
C2D+GRU	56.7%	55.1%	56.9%	55.4%	57.7%	56.1%	57.8%	56.6%	57.4%	56.3%	56.8%	55.4%
C2D+GRU+OF	58.1%	58.2%	58.5%	58.0%	58.9%	59.0%	59.1%	59.5%	59.0%	59.3%	58.0%	58.7%
D2D+LSTM	63.8%	66.8%	64.1%	66.6%	64.6%	67.5%	64.8%	68.9%	65.1%	67.0%	63.5%	65.5%
D2D+LSTM+OF	67.5%	69.4%	68.0%	69.9%	68.5%	70.4%	69.3%	71.4%	69.3%	69.6%	67.8%	68.2%
D2D+GRU	61.7%	65.0%	62.5%	65.1%	63.6%	65.4%	63.3%	67.1%	64.0%	65.8%	62.0%	63.5%
D2D+GRU+OF	67.7%	68.4%	68.3%	68.5%	69.9%	68.8%	69.8%	70.6%	70.3%	68.9%	68.2%	67.1%
C3D	61.5%	62.5%	61.7%	62.9%	62.2%	62.3%	62.7%	63.9%	62.4%	63.7%	61.4%	62.5%
C3D+OF	65.5%	65.4%	66.4%	65.9%	66.7%	65.8%	65.9%	67.2%	67.4%	67.1%	66.6%	65.9%

Table 4.3: The A2C relevant RWMA accuracy per class comparison of experimented methods.

Method	Region Numbers											
	3		6		9		12		14		16	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
C2D	53.4%	53.2%	53.6%	53.4%	55.1%	55.5%	54.0%	54.6%	54.1%	54.6%	53.1%	53.5%
C2D+OF	54.6%	54.5%	55.9%	54.0%	55.8%	56.4%	55.6%	56.9%	55.9%	55.7%	54.0%	56.8%
D2D	54.5%	54.1%	54.2%	54.9%	56.0%	56.4%	55.7%	55.3%	55.0%	55.5%	54.9%	54.4%
D2D+OF	55.4%	55.8%	56.5%	55.8%	56.2%	56.9%	56.5%	57.1%	56.4%	56.5%	55.3%	57.3%
C2D+LSTM	56.3%	55.6%	56.1%	56.1%	56.9%	56.2%	57.2%	55.8%	57.5%	56.6%	55.9%	56.5%
C2D+LSTM+OF	58.4%	56.5%	58.8%	57.3%	59.6%	57.5%	60.5%	58.2%	58.3%	57.0%	59.5%	56.7%
C2D+GRU	56.1%	55.5%	57.3%	55.7%	57.2%	56.5%	57.2%	56.3%	58.0%	56.9%	56.0%	55.2%
C2D+GRU+OF	58.5%	58.3%	58.2%	58.9%	58.3%	58.7%	59.5%	59.9%	60.5%	61.3%	59.5%	58.0%
D2D+LSTM	64.2%	66.5%	64.9%	66.0%	65.7%	67.2%	64.1%	67.7%	65.4%	68.3%	64.2%	64.9%
D2D+LSTM+OF	66.8%	69.9%	69.4%	68.6%	69.7%	71.5%	70.8%	70.9%	69.0%	69.8%	68.1%	68.9%
D2D+GRU	61.2%	65.9%	63.1%	66.4%	63.2%	66.1%	63.9%	66.9%	63.5%	66.7%	63.9%	64.7%
D2D+GRU+OF	67.1%	68.9%	68.0%	67.9%	68.8%	67.3%	69.9%	71.3%	70.4%	69.2%	68.4%	66.9%
C3D	60.9%	61.8%	62.9%	63.4%	62.8%	61.7%	61.9%	63.2%	62.9%	63.2%	61.9%	63.0%
C3D+OF	66.4%	65.9%	67.2%	66.6%	66.2%	64.3%	65.1%	67.9%	66.9%	67.8%	67.3%	65.3%

Table 4.4: The A4C relevant RWMA accuracy per class comparison of experimented methods.

Method	Region Numbers											
	2		5		8		11		14		16	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
C2D	52.8%	53.2%	53.3%	53.7%	54.2%	54.1%	54.4%	54.3%	54.0%	55.1%	53.6%	53.2%
C2D+OF	54.5%	55.0%	54.7%	54.2%	54.9%	56.2%	55.5%	56.1%	55.9%	55.8%	54.5%	56.6%
D2D	53.9%	53.8%	54.5%	54.6%	54.9%	55.0%	55.8%	58.1%	55.2%	55.7%	54.3%	54.8%
D2D+OF	55.3%	55.8%	55.4%	55.9%	56.0%	56.8%	56.3%	57.2%	56.5%	56.6%	55.9%	58.0%
C2D+LSTM	56.9%	54.8%	56.9%	54.9%	57.5%	56.5%	57.1%	55.9%	56.5%	56.7%	56.0%	55.5%
C2D+LSTM+OF	58.0%	57.5%	58.2%	57.9%	59.7%	57.0%	56.3%	57.4%	59.8%	57.5%	58.8%	56.1%
C2D+GRU	56.5%	55.8%	56.4%	55.9%	57.5%	56.9%	59.9%	56.8%	57.9%	56.1%	55.9%	56.3%
C2D+GRU+OF	58.6%	58.9%	58.1%	59.2%	57.5%	59.4%	58.9%	59.3%	60.2%	59.1%	57.5%	57.5%
D2D+LSTM	63.5%	62.1%	63.9%	65.3%	64.5%	67.9%	64.3%	68.2%	65.8%	67.5%	62.9%	66.0%
D2D+LSTM+OF	67.0%	69.9%	68.5%	69.3%	68.8%	71.9%	69.5%	71.0%	69.5%	68.5%	68.9%	67.1%
D2D+GRU	62.9%	63.4%	63.5%	64.9%	63.8%	65.2%	64.7%	65.9%	63.9%	66.6%	63.5%	62.4%
D2D+GRU+OF	67.5%	68.0%	68.9%	67.5%	69.8%	68.6%	67.7%	70.2%	71.1%	68.8%	68.7%	68.8%
C3D	61.2%	61.7%	61.8%	61.6%	62.8%	61.9%	62.2%	64.2%	66.9%	62.5%	61.1%	62.9%
C3D+OF	65.9%	64.9%	66.1%	65.5%	66.8%	64.9%	65.6%	67.7%	67.5%	66.6%	65.9%	66.0%

Table 4.5: The PLAX relevant RWMA accuracy per class comparison of experimented methods.

localization helps with WMA detection in echo. Excluding the motion of the atria and right ventricle decreases variance from the neighbouring chambers. Going through the failed cases, a number of studies might have been misclassified due to wrong localization of the LV. There are some approaches that localize LV automatically with the current segmentation networks [24]. These methods can be used to localize, track and accordingly crop LV throughout the cine.

Our results suggest that the estimation of WMS is an accurate method. Any echocardiography machine can then easily translate this routine information into a robust estimate of WMS without the use of any strain measurement.

Chapter 5

Conclusion

In this thesis, we proposed an automatic system for the evaluation of regional wall motion abnormality. Given the variability and challenge of coding RWMA, the development of a platform that can quickly and accurately identify regional wall motion abnormalities on echo images will be significant asset. Such a tool has several applications to improve the accuracy and consistency of RWMA reporting with bedside echo at the point-of-care, stress echo, and for the workflow of the lab.

The required data were collected from the VGH in two streams consisting 489 unique patients. First part of the data was echo images from these patients and the second piece contained the corresponding measurements and pathology reports. The relevant measurements included diagnostic information for standard measurements, comments, etc along with patient information (e.g., name, age) and exam information. These information were linked to the echo images using the patients unique IDs and dates. Eventually the data were transformed into a local database to make it in a proper format as an input for the model.

Furthermore, we recognize that the categorization of RWMA is imprecise and that there are different degrees of hypokinesis that may correlate with differing severity of flow-limiting lesions. These subtle differences in echo appearance may be difficult to perceive by human interpreters but are detected with a robust machine learning model.

The overarching objective of our research program is the development of integrated tools to assist and automate in the acquisition and interpretation of echo. For

this dissertation, a novel deep learning model is implemented to: identify RWMAAs thus, reducing the time to accurate interpretation; and predict the degree of obstructive coronary disease through multi-view data analysis. The created database allows us to leverage the extensive echo dataset of patients' images, coded RWMAAs, and complete clinical reports.

Automated identification of regional wall motion abnormalities was implemented through the following phases: Using our own previously developed machine learning platform for view classification, we identified and retrieved key echo views required for the visualization of RWMAAs (parasternal long-axis, apical 2-chamber, and apical 4-chamber). These clips were coded for regional wall motion abnormalities for training and validation of a machine learning model. As described in chapter 4, a model that can predict the regional and global wall motion abnormality was trained and hyper-parameters of it were tuned. In the evaluation phase, the model predicted the abnormality of all regions for the test cases.

5.1 Contributions

In this work, a framework for regional wall motion abnormality analysis on echo imaging information has been investigated. The contributions made to reach this goal are summarized below:

A study on the heart and relevant cardiac diseases were performed. The diagnostic imaging techniques were reviewed. In order to improve the diagnosis of wall motion abnormalities, we decided to develop an artificially intelligent model to assist in the identification of the wall motion abnormalities through echo images. This is of a great value given the variability and challenges of scoring such abnormalities.

Then, I developed a deep learning framework consisting of 3 views of echocardiography data as input, with abnormality prediction for 16 wall segments as output. The network directly analyzes echo data without any need for prior segmentation of the cardiac LV wall.

The models were trained using the data consisting 489 patients. The hyper parameters of the model were then optimized. The resultant model precisely identifies regional wall motion abnormalities as well as global WMA on echo as com-

pared to the expert human labels with advanced echo training. In independent test dataset, I demonstrated that the neural network can produce accuracy as high as 69.2% for detection of abnormal wall motion.

The developed model, to the best of our knowledge, is the first to analyze the left ventricle wall motion for both global and regional abnormality detection in echocardiography data. This can be considered a great contribution in enhancing the diagnostic process of the wall motion abnormalities, by expediting the process for physicians.

5.2 Future Work

There is room for future work to improve the accuracy of the current framework. Moreover, the current framework is only predicting the normal versus abnormal regional and global function, it can be extended to predict all severity levels. Besides, we hypothesize that a machine learning approach is able to produce a model that can accurately predict the severity of obstructive coronary disease based on echo images as compared to the reference standard, degree of epicardial stenosis on invasive coronary angiogram or coronary computed tomography angiograms. Also, since only the apical views are used for predictions, it might be worth trying to compare the results with the parasternal views as well.

To conclude, while there are many approaches available for WMA detection using speckle tracking, we believe we are the first to propose a model that predicts all the 16 segments abnormality using apical views of echocardiography. The flexibility and true power of the framework are seen by combining global wall motion abnormality predictions, which results in better performance using only 2DE image information. The benefits advanced by this unified framework are: 1) it can predict both regional and global wall motion abnormality of LV, 2) due to the low computational complexity of the framework, the framework can be applied to portable echo machines that are frequently used in the emergency room and in rural clinics.

Bibliography

- [1] Truecrypt. https://opencryptoaudit.org/reports/TrueCrypt_Phase_II_NCC_OCAP_final.pdf, 2015. Accessed: 2010-08-30. → page 26
- [2] A. H. Abdi, C. Luong, T. Tsang, G. Allan, S. Nouranian, J. Jue, D. Hawley, S. Fleming, K. Gin, J. Swift, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. *IEEE Transactions on Medical Imaging*, 36(6): 1221–1230, 2017. → page 15
- [3] M. Afshin, I. B. Ayed, K. Punithakumar, M. Law, A. Islam, A. Goela, T. Peters, and S. Li. Regional assessment of cardiac left ventricular myocardial function via mri statistical features. *IEEE Transactions on Medical Imaging*, 33(2):481–494, 2013. → page 22
- [4] S. Amari et al. *The Handbook of Brain Theory and Neural Networks*. MIT Press, 2003. → page 21
- [5] S. Azizi, S. Bayat, P. Yan, A. Tahmasebi, J. T. Kwak, S. Xu, B. Turkbey, P. Choyke, P. Pinto, B. Wood, et al. Deep recurrent neural networks for prostate cancer detection: analysis of temporal enhanced ultrasound. *IEEE Transactions on Medical Imaging*, 37(12):2695–2703, 2018. → pages 18, 19
- [6] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012. → page 21
- [7] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, pages 2546–2554, 2011. → page 21
- [8] D. S. Blondheim, R. Beerli, M. S. Feinberg, M. Vaturi, S. Shimoni, W. Fehske, A. Sagie, D. Rosenmann, P. Lysyansky, L. Deutsch, et al.

Reliability of visual assessment of global and segmental left ventricular function: a multicenter study by the israeli echocardiography research group. *Journal of the American Society of Echocardiography*, 23(3): 258–264, 2010. → page 12

- [9] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2559–2566. IEEE, 2010. → page 41
- [10] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *Neural Information Processing Systems 2014 Workshop on Deep Learning*, 2014. → page 19
- [11] C. B. Compas, E. Y. Wong, X. Huang, S. Sampath, B. A. Lin, P. Pal, X. Papademetris, K. Thiele, D. P. Dione, M. Stacy, et al. Radial basis functions for combining shape and speckle tracking in 4d echocardiography. *IEEE Transactions on Medical Imaging*, 33(6):1275–1289, 2014. → page 21
- [12] T. Edvardsen, B. L. Gerber, J. Garot, D. A. Bluemke, J. A. Lima, and O. A. Smiseth. Quantitative assessment of intrinsic regional myocardial deformation by doppler strain rate echocardiography in humans: validation against three-dimensional tagged magnetic resonance imaging. *Circulation*, 106(1):50–56, 2002. → pages xi, 13, 14
- [13] J. Ehrhardt, M. Wilms, H. Handels, and D. Säring. Automatic detection of cardiac remodeling using global and local clinical measures and random forest classification. In *Statistical Atlases and Computational Models of the Heart*, pages 199–207. Springer, 2015. → page 21
- [14] A. Eitan, I. Kehat, D. Mutlak, G. Lichtenberg, D. Amar, and Y. Agmon. Longitudinal two-dimensional strain for the diagnosis of left ventricular segmental dysfunction in patients with acute myocardial infarction. *The International Journal of Cardiovascular Imaging*, 34(2):237–249, 2018. → page 13
- [15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017. → page 21
- [16] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: continual prediction with lstm. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, volume 2, pages 850–855 vol.2, 1999. doi:10.1049/cp:19991218. → page 19

- [17] J. Goebel, F. Nensa, H. P. Schemuth, S. Maderwald, H. H. Quick, T. W. Schlosser, and K. Naßenstein. Detection of regional wall motion abnormalities in compressed sensing cardiac cine imaging. *World Journal of Cardiovascular Diseases*, 8(06):277, 2018. → page 4
- [18] J. Gorcsan III, D. P. Strum, W. A. Mandarino, V. K. Gulati, and M. R. Pinsky. Quantitative assessment of alterations in regional left ventricular contractility with color-coded tissue doppler echocardiography: comparison with sonomicrometry and pressure-volume relations. *Circulation*, 95(10):2423–2433, 1997. → page 13
- [19] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649. IEEE, 2013. → page 18
- [20] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *The Journal of the American Medical Association*, 316(22):2402–2410, 2016. → page 21
- [21] R. S. Horowitz, J. Morganroth, C. Parrotto, C. C. Chen, J. Soffer, and F. J. Pauletto. Immediate diagnosis of acute myocardial infarction by two-dimensional echocardiography. *Circulation*, 65(2):323–329, 1982. → pages 5, 11, 12
- [22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. → page 43
- [23] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962. → page 17
- [24] M. H. Jafari, H. Girgis, N. Van Woudenberg, Z. Liao, R. Rohling, K. Gin, P. Abolmaesumi, and T. Tsang. Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training. *International Journal of Computer Assisted Radiology and Surgery*, 14(6):1027–1037, 2019. → page 57

- [25] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. → page 21
- [26] S. Kida, T. Nakamoto, M. Nakano, K. Nawa, A. Haga, J. Kotoku, H. Yamashita, and K. Nakagawa. Cone beam computed tomography image quality improvement using a deep convolutional neural network. *Cureus Journal of Medical Science*, 10(4), 2018. → page 21
- [27] M. Kinno, P. Nagpal, S. Horgan, and A. H. Waller. Comparison of echocardiography, cardiac magnetic resonance, and computed tomographic imaging for the evaluation of left ventricular myocardial function: Part 1 (global assessment). *Current Cardiology Reports*, 19(1):9, Feb 2017. ISSN 1534-3170. doi:10.1007/s11886-017-0815-4. URL <https://doi.org/10.1007/s11886-017-0815-4>. → page 12
- [28] K. Kusunose, T. Abe, A. Haga, D. Fukuda, H. Yamada, M. Harada, and M. Sata. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC: Cardiovascular Imaging*, 2019. → pages 21, 55
- [29] R. M. Lang, L. P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F. A. Flachskampf, E. Foster, S. A. Goldstein, T. Kuznetsova, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. *European Heart Journal-Cardiovascular Imaging*, 16(3):233–271, 2015. → pages x, 9, 12
- [30] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. → page 21
- [31] K. E. Leung and J. G. Bosch. Automated border detection in three-dimensional echocardiography: principles and promises. *European Journal of Echocardiography*, 11(2):97–108, 2010. → page 54
- [32] A. Li, O. Spyra, S. Perel, V. Dalibard, M. Jaderberg, C. Gu, D. Budden, T. Harley, and P. Gupta. A generalized framework for population based training. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1791–1799, 2019. → page 21
- [33] N. Liel-Cohen, Y. Tsadok, R. Beeri, P. Lysyansky, Y. Agmon, M. S. Feinberg, W. Fehske, D. Gilon, I. Hay, R. Kuperstein, et al. A new tool for automatic assessment of segmental wall motion based on longitudinal 2d

strain: a multicenter study by the israeli echocardiography research group. *Circulation: Cardiovascular Imaging*, 3(1):47–53, 2010. → pages 11, 12

- [34] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. → page 17
- [35] D. Maclaurin, D. Duvenaud, and R. Adams. Gradient-based hyperparameter optimization through reversible learning. In *International Conference on Machine Learning*, pages 2113–2122, 2015. → page 21
- [36] J. Mantilla, M. Garreau, J.-J. Bellanger, and J. L. Paredes. Svm-based classification of lv wall motion in cardiac mri with the assessment of ste. In *10th International Symposium on Medical Information Processing and Analysis*, volume 9287, page 92870N. International Society for Optics and Photonics, 2015. → page 21
- [37] V. Mor-Avi, P. Vignon, R. Koch, L. Weinert, M. J. Garcia, K. T. Spencer, and R. M. Lang. Segmental analysis of color kinesis images: new method for quantification of the magnitude and timing of endocardial motion during left ventricular systole and diastole. *Circulation*, 95(8):2082–2097, 1997. → pages xi, 13, 14
- [38] V. Mor-Avi, L. Sugeng, and R. M. Lang. Real-time 3-dimensional echocardiography: an integral component of the routine echocardiographic examination in adult patients. *Circulation*, 119(2):314–329, 2009. → pages xi, 14
- [39] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010. → page 41
- [40] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O’Regan, et al. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, 2017. → page 22
- [41] H. A. Omar, J. S. Domingos, A. Patra, R. Upton, P. Leeson, and J. A. Noble. Quantification of cardiac bull’s-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography. In *2018*

IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pages 1195–1198. IEEE, 2018. → page 22

- [42] A. H. A. W. G. on Myocardial Segmentation, R. for Cardiac Imaging:, M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation*, 105(4):539–542, 2002. → pages 9, 11
- [43] OpenStax. *Anatomy Physiology*. OpenStax CNX., Feb 26, 2016. → page 2
- [44] N. Ouzir, O. Lairez, A. Basarab, and J.-Y. Tourneret. Tissue motion estimation using dictionary learning: Application to cardiac amyloidosis. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2017. → page 21
- [45] D. Peressutti, M. Sinclair, W. Bai, T. Jackson, J. Ruijsink, D. Nordsletten, L. Asner, M. Hadjicharalambous, C. A. Rinaldi, D. Rueckert, et al. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Medical Image Analysis*, 35:669–684, 2017. → page 22
- [46] D. Platts and M. Monaghan. Colour encoded endocardial tracking: the current state of play. *European Journal of Echocardiography*, 4(1):6–16, 2003. → page 13
- [47] E. Puyol-Antón, M. Sinclair, B. Gerber, M. S. Amzulescu, H. Langet, M. De Craene, P. Aljabar, P. Piro, and A. P. King. A multimodal spatiotemporal cardiac motion atlas from MR and ultrasound data. *Medical Image Analysis*, 40:96–110, 2017. → page 22
- [48] E. Puyol-Antón, M. Sinclair, B. Gerber, M. S. Amzulescu, H. Langet, M. De Craene, P. Aljabar, J. A. Schnabel, P. Piro, and A. P. King. Multiview machine learning using an atlas of cardiac cycle motion. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 3–11. Springer, 2017. → page 22
- [49] E. Puyol-Antón, B. Ruijsink, W. Bai, H. Langet, M. De Craene, J. A. Schnabel, P. Piro, A. P. King, and M. Sinclair. Fully automated myocardial strain estimation from cine mri using convolutional neural networks. In

2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pages 1139–1143. IEEE, 2018. → page 22

- [50] A. Rizvi, R. C. Deaño, D. P. Bachman, G. Xiong, J. K. Min, and Q. A. Truong. Analysis of ventricular function by ct. *Journal of Cardiovascular Computed Tomography*, 9(1):1–12, 2015. → page 5
- [51] S. Sanchez-Martinez, N. Duchateau, T. Erdei, A. G. Fraser, B. H. Bijmens, and G. Piella. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. *Medical Image Analysis*, 35:70–82, 2017. → page 21
- [52] D. Scherer, A. Müller, and S. Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *International Conference on Artificial Neural Networks*, pages 92–101. Springer, 2010. → page 41
- [53] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012. → page 21
- [54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014. → pages 45, 49
- [55] A. Suinesiaputra, A. F. Frangi, T. A. Kaandorp, H. J. Lamb, J. J. Bax, J. H. Reiber, and B. P. Lelieveldt. Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac mr images. *IEEE Transactions on Medical Imaging*, 28(4):595–607, 2009. → page 22
- [56] A. Suinesiaputra, P. Ablin, X. Alba, M. Alessandrini, J. Allen, W. Bai, S. Cimen, P. Claes, B. R. Cowan, J. D’hooge, et al. Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. *IEEE Journal of Biomedical and Health Informatics*, 22(2):503–515, 2017. → page 22
- [57] A. J. Tajik. Machine learning for echocardiographic imaging: embarking on another incredible journey, 2016. → page 15
- [58] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, H. D. White, H. Mickley, F. Crea, F. Van de Werf, et al. Fourth

universal definition of myocardial infarction (2018). *European Heart Journal*, 40(3):237–269, 2019. → page 12

- [59] N. Van Woudenberg, Z. Liao, A. H. Abdi, H. Girgis, C. Luong, H. Vaseli, D. Behnami, H. Zhang, K. Gin, R. Rohling, et al. Quantitative echocardiography: real-time quality estimation and view classification implemented on a mobile android device. In *Simulation, Image processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, pages 74–81. Springer, 2018. → page 15
- [60] H. Vaseli, Z. Liao, A. H. Abdi, H. Girgis, D. Behnami, C. Luong, F. T. Dezaki, N. Dhungel, R. Rohling, K. Gin, et al. Designing lightweight deep learning models for echocardiography view classification. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951, page 109510F. International Society for Optics and Photonics, 2019. → pages 15, 39
- [61] D. M. Vigneault, W. Xie, D. A. Bluemke, and J. A. Noble. Feature tracking cardiac magnetic resonance via deep learning and spline optimization. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 183–194. Springer, 2017. → page 22

Appendix A

Supporting Materials

FileMaker Pro Advanced - [Echo Studies (VCHICLINP01)]

File Edit View Insert Format Requests Scripts Tools Window Help

1 Total
Find Requests New Request Delete Request Perform Find Cancel Find

Layout: Front Page View As: Matching records: Include Qmit Insert: Operators

teaching case
 AAACC IE33 GE VGH IE33 VGH TEE VGH ORTEE VGH STRESS VGH GE UBC
 INPAT ON
 Tape#/Time hour min
 Date of study
 Start/End/Room
 Gender M F BSA Ward
 Techs/Physician cardiology residents/fellows cc:

Echo ID #
 Last Name
 First Name
 Date of Birth
 PHN / Med #
 Hospital #
 Doctor

murmur cardiomegaly LV function congenital heart disease
 MVP trauma coronary disease aortic aneurysm
 stroke/embolus aortic valve disease cardiomyopathy transplant
 endocarditis mitral valve disease thrombus Pregnant
 pericardial disease tricuspid valve disease hypertension Syncope
 arrhythmia's/abn ECG pulmonary valve disease pulmonary hypertension none given
 heart failure/dyspnea prosthetic function RV function Other...

MMMode/2D
 2D
 Doppler
 TEE
 Intraoperative TEE
 Portable
 Dobutamine Echo
 Stress Echo
 Saline Contrast
 Contrast Agent
 Amyl Nitrate
 Emergency Study
 Research Study
 Other...

Sinus Rhythm
 Sinus Tachycardia
 Sinus Bradycardia
 Atrial Fibrillation
 Atrial Flutter
 Uncontrolled ventricular response
 Indeterminate
 Paced Rhythm
 Ventricular Premature Beats
 Preexcitation
 1st Degree Heart Block
 2nd Degree Heart Block
 3rd Degree Heart Block
 LBBB
 Prolonged QT
 Other...

Technically difficult
 Limited Views
 Image quality suboptimal
 Patient in distress
 Liver cyst(s)/mass(es) noted
 Unable to get PA pressure
 Pacemaker wires or catheter in RA/RV.
 Taped under wrong name. Sorry!
 ? wall motion abnormality
 Patient intubated

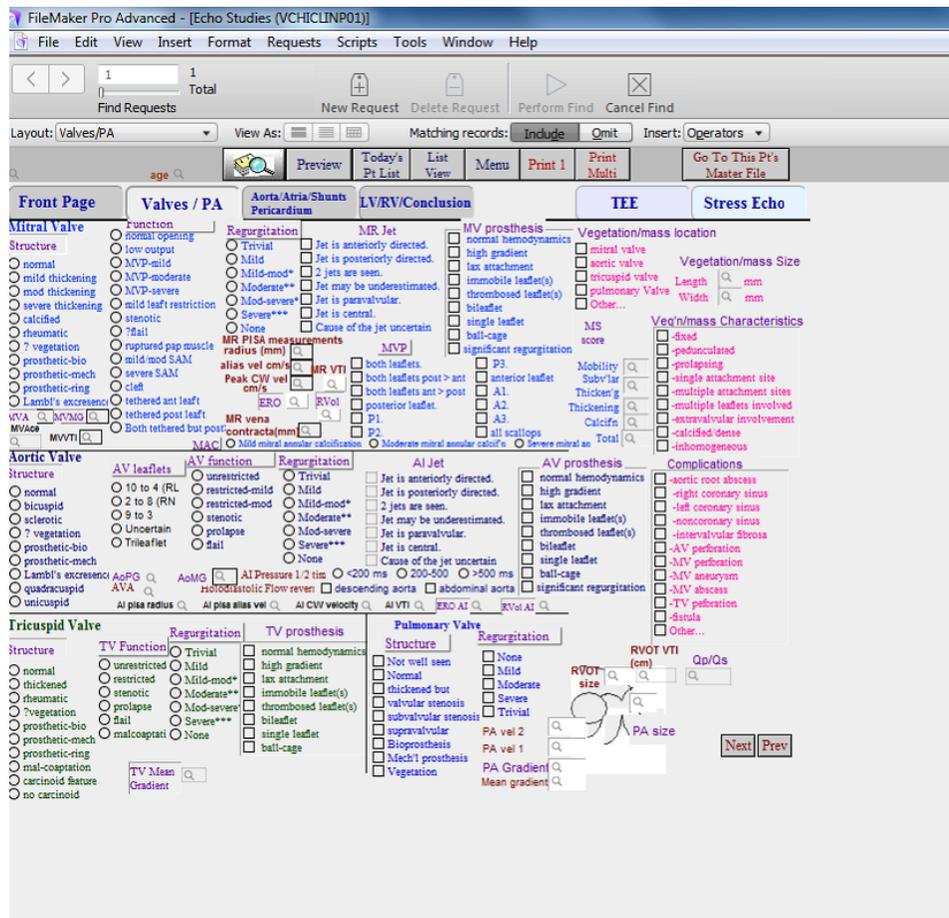
Contrast Study with Definity
 Limited Pulmonary Hypertension Study
 Intracardiac Echo

TTE
 TEE
 Combined TTE & TEE
 TEE in ICU
 TEE in CSICU
 TEE in CCU
 TEE in cath lab for RF ablation
 TEE/TTE in cath lab for TAVI
 TEE in cath lab for LAA closure
 TEE pending
 Portable study
 Portable stat study in cath lab
 Epicardial/epiaortic
 TEE suggested
 TEE not done
 TEE not needed
 Emergency Study
 Tissue Synchronization study

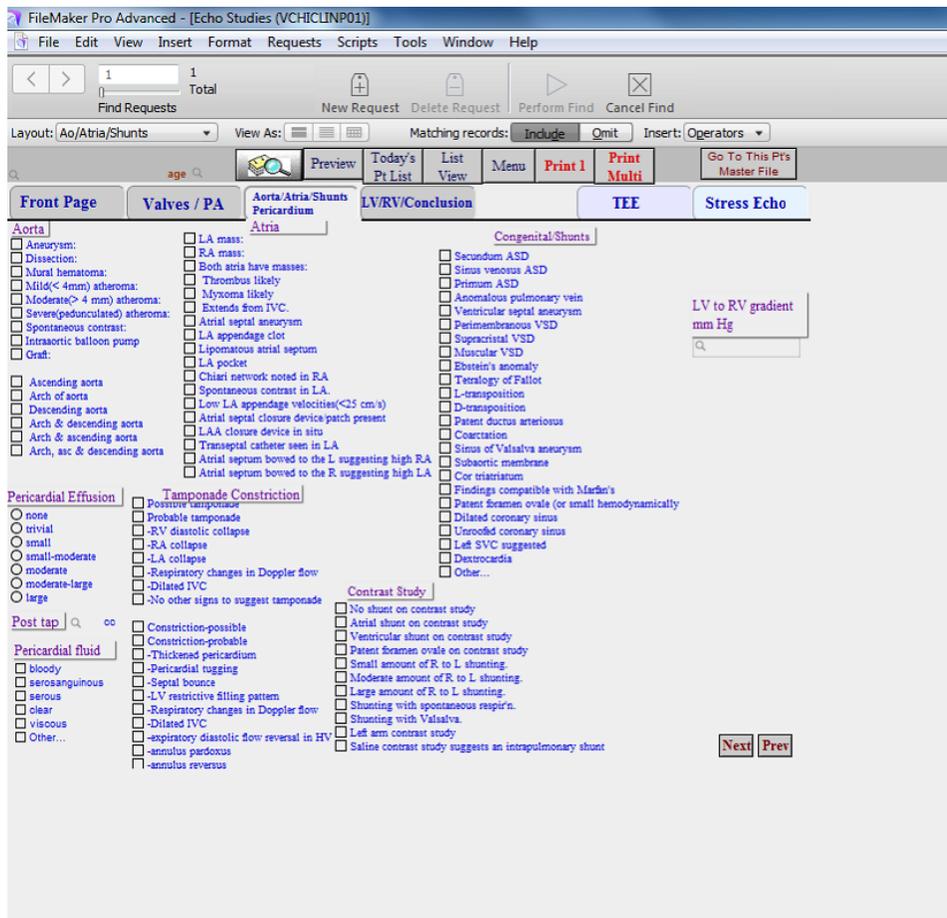
same day request Yes No
 Copy Conclusions
 Risk factors
 None
 Never smoked
 Smoker
 Quit more than 5 years ago
 High BP
 Diabetes
 High Cholesterol
 Other...

Echo Studies
 Next Cond

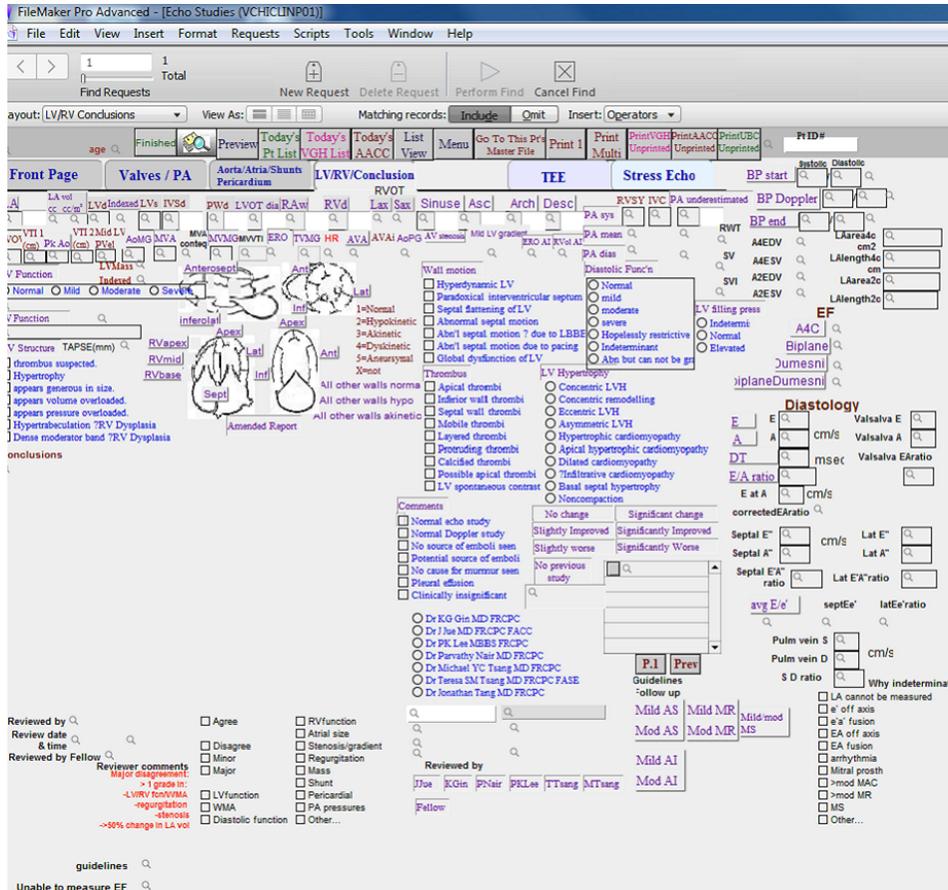
(a) The Front Page



(b) The Valves Page



(c) The Aorta, Atria, Shunts, Pericardium Page



(d) The LV/RV Assessment Page

Figure A.1: A snapshot of different tabs available in the Filemaker software for cardiologists to records the analytical measurements.