공학박사 학위논문

A study on the automated echocardiogram interpretation using deep neural networks for clinical decision support system

임상의사 결정 지원 시스템을 위한 심층 신경망 기반의 심초음파 자동해석에 관한 연구

2022 년 8 월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

손 장 재

Ph. D. Dissertation

임상의사 결정 지원 시스템을 위한 심층 신경망 기반의 심초음파 자동해석에 관한 연구

지도 교수 김 희 찬

이 논문을 공학박사 학위논문으로 제출함

2022 년 06 월

서울대학교 대학원

협동과정 바이오엔지니어링

손 장 재

손장재의 공학박사 학위논문을 인준함

2022 년 08 월

위 원 장 윤 형 진 (인)

부위원장 김 희 찬 (인)

위 원 이 정 찬 (인)

위 원 박 준 빈 (인)

위 원 구 윤 서 (인)

Ph. D. Dissertation

A study on the automated echocardiogram interpretation using deep neural networks for clinical decision support system

BY

JANG JAY SOHN

AUGUST 2022

INTERDISCIPLINARY PROGRAM IN BIOENGINEERING

THE GRADUATE SCHOOL

SEOUL NATIONAL UNIVERSITY

A study on the automated echocardiogram interpretation using deep neural networks for clinical decision support system

Academic adviser Hee Chan Kim, Ph.D.
Submitting a Ph.D. Dissertation of Engineering
June 2022

Interdisciplinary Program in Bioengineering
The Graduate School, Seoul National University

Jang Jay Sohn

Confirming the Ph.D. Dissertation written by
Jang Jay Sohn
August 2022

Chair _____
Hyung Jin Yoon, M.D. / Ph. D.
Vice Chair _____
Hee Chan Kim, Ph.D.
Examiner _____
Jung Chan Lee, Ph. D.
Examiner _____
Jun-Bean Park, M.D. / Ph. D.
Examiner _____
Yunseo Ku, Ph. D.

ABSTRACT

A study on the automated echocardiogram interpretation using deep neural networks for clinical decision support system

Jangjay Sohn

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Echocardiography is an indispensable tool for cardiologists in the diagnosis of heart diseases. By echocardiography, various structural abnormalities in the heart can be quantitatively or qualitatively diagnosed. Due to its non-invasiveness, the usage of echocardiography in the diagnosis of heart disease has continuously increased.

Despite the increasing role in the cardiology practice, echocardiography requires experience in capturing and knowledge in interpreting images. . Moreover, in contrast to CT or MRI images, important information can be missed once not obtained at the time of examination. Therefore, obtaining and interpreting images should be done simultaneously, or, at least, all obtained images should be audited by the experienced cardiologist before releasing the patient from the examination booth. Because of the peculiar characteristics of echocardiography compared to CT or MRI, there have been incessant demands for the clinical decision support system(CDSS) for echocardiography.

With the advance of Artificial Intelligence (AI), there have been several studies regarding decision support systems for echocardiography. The flow of these studies is divided into two approaches: One is the quantitative approach to segment the images and detects an abnormality in size and function. The other is the qualitative approach to detect abnormality in morphology. Unfortunately, most of these two studies have been conducted separately. However, since cardiologists perform quantitative and qualitative analysis simultaneously in analyzing echocardiography, an optimal CDSS needs to be a combination of these two approaches. From this point of view, this study aims to develop and validate an AI-based CDSS for echocardiograms through a large-scale retrospective cohort. Echocardiographic data of 2,600 patients who visited Seoul National University Hospital (1300 cardiac patients and 1300 non-cardiac patients with normal echocardiogram) between 2016 and 2021. Two networks were developed for the quantitative and qualitative analysis, and their usefulnesses were verified with the patient data.

First, a U-net based deep learning network was developed for segmentation in the quantitative analysis. Annotated images by the experienced cardiologist with the left ventricle, interventricular septum, left ventricular posterior wall, right ventricle, aorta, and left atrium, were used for training. The diameters and areas of the six structures were obtained and vectorized from the segmentation images, and the frame information at the end-systolic and end-diastolic phases was extracted from the vector.

The second network for the qualitative diagnosis was developed using a convolutional neural network (CNN) based on Resnet 152. The input data of this network was extracted from 10 frames of each patient based on end-diastolic and end-systolic phase information extracted from the quantitative network. The network not only distinguished the input data between normal and abnormal but also visualized the location of the abnormality on the image through the Gradient-weighted Class Activation Mapping (Grad-CAM) at the last layer.

The performance of the quantitative network in the chamber size and function measurements was assessed in 1300 patients. Sensitivity and

specificity were both over 90% except for pathologies related to the left ventricular posterior wall, interventricular septum, and aorta. The end-systolic and end-diastolic phase detection was also accurate, with an average difference of 0.52 frames for the end-systolic and 0.9 frames for the end-diastolic phases.

In the case of the network for qualitative analysis, 10 input data were selected based on the phase information determined from the first network, and the results of 10 randomly selected images were compared. As a result, the accuracy was 90.3% and 81.2%, respectively, and the phase information selected from the first network contributed to the improvement of the performance of the network. Also, the results of Grad-CAM confirmed that the network trained with 10 images of data extracted based on the phase information from the first network displays the location of the lesion more accurately than the network trained with 10 randomly selected data.

In conclusion, this study proposed an AI-based CDSS for echocardiography in the quantitative and qualitative analysis.

Keywords: Echocardiogram, AI, Clinical Decision Support System, U-net, deep learning, Segmentation

Student Number: 2018-32128

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

AI	Artificial Intelligence
CDSS	Clinical Decision Support System
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
ES	End Systolic
ED	End Diastolic
ECG	Electrocardiogram
EF	Ejection Fraction
LV	Left Ventricle
RV	Right Ventricle
CAM	Class Activation Map
Grad-CAM	Gradient-weighted Class Activation Mapping
CNN	Convolutional Neural Network
DSC	Dice Similarity Coefficient
IoU	Intersection over Unit
LA	Left Atrium
CAM	Class Activation Map
AUROC	Area Under the Receiver Operating Characteristic curve
PLAX	Parasternal Long-AXis

CHAPTER 1

Introduction

1. Introduction
1.1. Echocardiogram
1.1.1. Diagnosis using Echocardiogram

In echocardiography, the diagnosis of various heart diseases can be accomplished by analyzing the reflected ultrasonic wave from the heart. Due to its non-invasiveness, the usage of echocardiography in the diagnosis of heart disease has continuously increased. Moreover, its utilization is also increasing as the number of patients with heart disease increases [1]. In echocardiography, various cross-sectional image planes depend on the various directions of the ultrasonic beam as seen in Figure 1.1. Therefore, types of pathologic findings also vary according to the various imaging planes. In addition, as the heart is a moving structure, an echocardiographic image should be recorded in a moving file, in contrast to the still images of the Computed Tomography(CT) or Magnetic Resonacne Imageing(MRI)

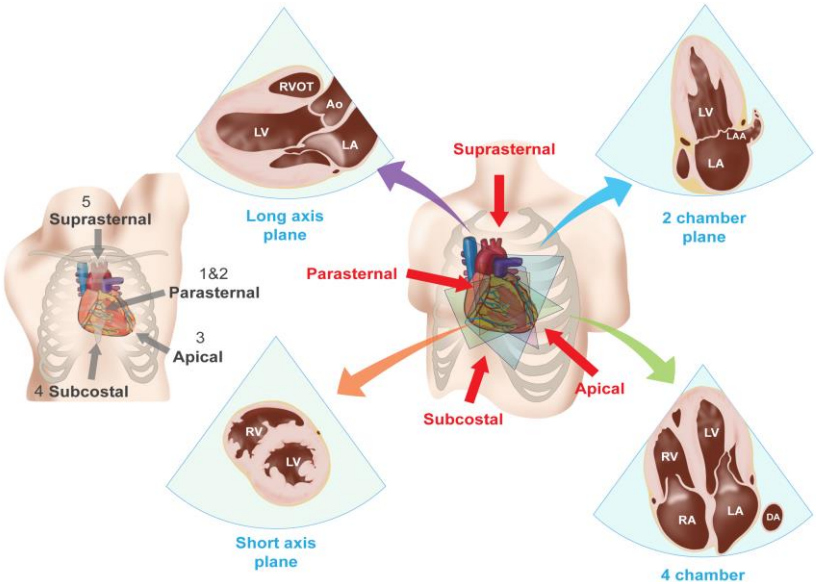


Figure 1.1. Cross-sectional anatomy routinely used in echocardiography

1.1.2. Limitation in Echocardiogram

The technique of image acquisition is very important for a good echocardiographic image. Therefore, it takes quite a time to be an experienced echocardiographer.

In echocardiography, as shown in Figure 1.2, echocardiographers search for the optimal probe position and intended image plane, because the image can be obtained only in a certain position and plane. The echocardiographer should sometimes modify the predefined imaging plane to get the image of the area of interest. However, there is no limitation in getting cross-sectional images in CT or MRI. However, the only obstacle is the presence of artifact-producing objects (prosthesis, calcification, etc.) in the image plane. Images are usually obtained by the image plane. Images are usually obtained by the pre-defined protocol by the paramedical personnel without loss of the anatomic information.

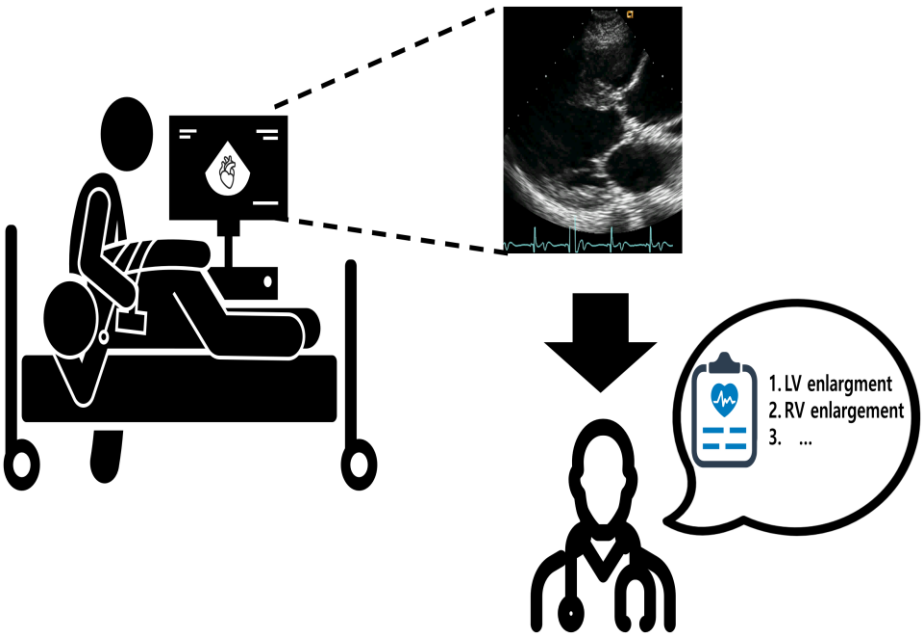


Figure 1. 2. Working of echocardiogram.

Though, echocardiographic images can be obtained only in certain ‘places’ and ‘planes.’ Moreover, information once missed at the time of examination cannot be detected by the analysis of the images after the examination. When the echocardiographic images are obtained by the paramedical personnel, obtained images should be audited by the echocardiographer before the patient is released from the examination booth.

The echocardiographic evaluation consists of quantitative and qualitative analysis. In the quantitative analysis, many studies point to inter-and intra- observer variability [2-4]. Even in the qualitative analysis, because the heart is a moving organ, diverse opinions may be present in a certain specific motion.

To solve these problems, efforts have been made to standardize and apply guidelines for echocardiographic measurements and interpretation. However, even with these guidelines, the interpretation of the image is decisively determined by the expertise of the person obtaining an echocardiogram [5]. As already mentioned above, once an important finding is missed in the initial image acquisition, even the experienced echocardiographer can pick up the abnormality only by reviewing the recorded images. Therefore, not missing a single finding at the time of initial image acquisition is of utmost importance.

The process of interpreting the echocardiogram consists of four processes [6]. The first step is to find out the imaging view. Second, the image quality of the view has to be checked, so that if the quality of the view is suboptimal, the image should be re-obtained. Third, the required parameters are measured from the images. Fourth, the abnormal structure or motion should be looked for.

Among these four processes, for the first and second processes, high-performance products have been released [7]. In the case of the

third and fourth processes, a large number of studies are being conducted for quantification of Ejection Fraction (EF), for diagnostic accuracy, or to find out a prognostic indicator in a specific disease.

1.1.3. Artificial Intelligence in Echocardiogram

Artificial intelligence (AI) has been in progress since 1950s, but the use of AI for medical images has increased only recently. AI technologies such as machine learning can be trained to identify images, quantify them, and discover disease patterns hidden in images, as they can consider relationships between pixels in images, or the clinical metadata [8]. By combining clinical interpretation with information derived from machine learning algorithms, accuracy can be improved through the reduction of inter-observer and intra-observer variability, as well as by providing subtle information [9-11] invisible to the operator. These technologies have the potential to improve the performance of clinical decision support systems (CDSS) and reduce unnecessary treatments and procedures. In particular, AI is likely to be a valuable tool for clinicians without expertise, who cannot diagnose or treat patients with higher accuracy and confidence. CDSS may reduce the likelihood of error occurrence, thereby improving patient management. Although the use of machine learning has developed significantly over the past decade, the entire application still needs to be improved, and further research is needed to improve its implementation in clinical applications.

1.2. Clinical Background

To develop a CDSS that can effectively help make clinical decisions, it is necessary to know how cardiologists use echocardiograms in the diagnosis. In the development of deep learning networks to be applied to CDSS, following the flow of the clinicians' method of interpretation will enable the network to produce faster and more accurate results. The echocardiographic evaluation consists of two analyses: a quantitative and a qualitative diagnosis and it is necessary to know how these two diagnoses are made.

1.2.1. Diagnostic Flow

Cardiologists have anatomical knowledge and interpret the heart structures shown in the echocardiogram quantitatively or qualitatively, judging by the appearance of the heart that differs between systole and diastole as shown in figure 1.3. Along with the echocardiogram, a single-lead electrocardiogram (ECG) is included in the echocardiographic image. While echocardiography provides the structural and functional information of your heart, ECG informs the doctor when the systolic and diastolic periods are timed. This information about the timing in the cardiac cycle is important, especially in the hemodynamic information of the heart. In echocardiography, occasionally, the phase of the respiratory cycle is important in the interpretation. Many vendors of the echocardiographic machine provide information about the respiratory cycle inferred from the ECG voltage change with respiration.

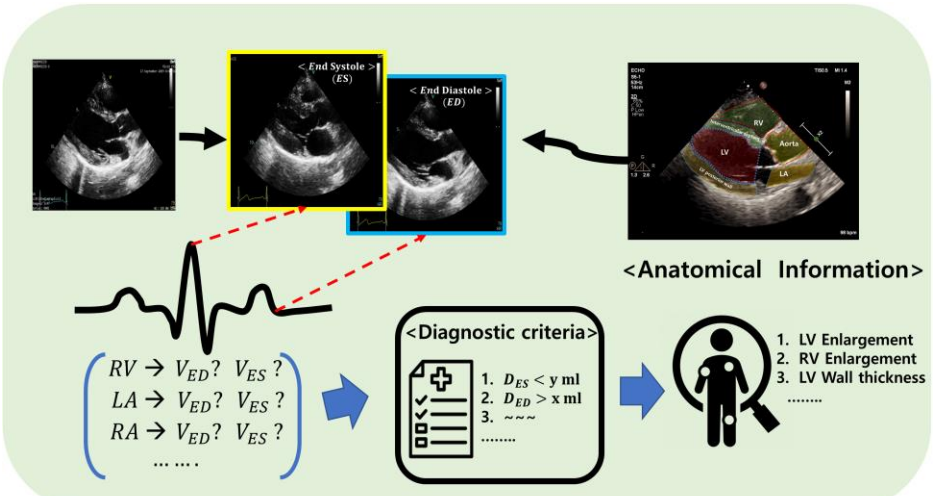


Figure 1.3. Clinical process of reaching echocardiographic diagnosis

For quantitative analysis, cardiologists measure the size of heart structures in systolic and diastolic images in an echocardiogram. Figure 1.4 shows the criteria for quantitative diagnosis, which can be obtained in the parasternal long-axis (PLAX) view used in this study. In the PLAX view, anatomical information of the left ventricle, interventricular septum, left ventricular posterior wall, right ventricle, aorta, and left atrium can be obtained. In normal subjects, the diameters of the chambers decrease during the systole and increase during diastole, whereas wall thickness increases during systole and decreases during diastole. In the disease process, the absolute number or the ratio of increment or decrement of these values becomes abnormal. By providing these absolute values and changes during the cardiac cycle automatically, CDSS can give information about the subtle abnormality that might have been overlooked.

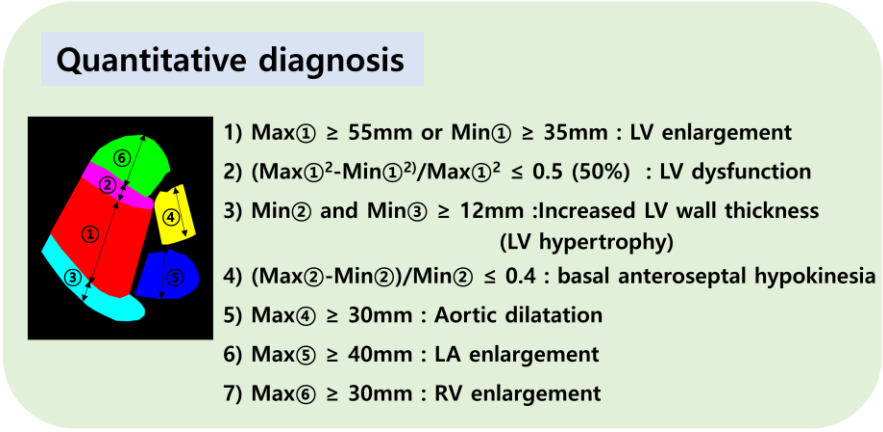


Figure 1.4. Criteria for quantitative diagnosis

In addition to the above quantitative diagnosis, the clinician makes the diagnosis qualitatively. For example, if the patient has a problem with the pericardium, the abnormal septal motion might be accompanied by constrictive pericarditis or pericardial effusion might be seen as shown in Figure 1.5. Among these two, abnormal septal motion, called septal bouncing, is the clue to constrictive pericarditis. For the diagnosis of constrictive pericarditis, the anatomic abnormality is almost always not evident. Diagnosis usually requires an in-depth Doppler examination additionally performed based on the 2D echocardiographic suspicion. Therefore, if the finding of septal bouncing is missed at the time of the scanning procedure, an in-depth Doppler evaluation might not have been done and might have missed the diagnosis of constrictive pericarditis. By CDSS, a reasonable basis for pathology similar to the decisions made by the experienced echocardiographer is to be presented.

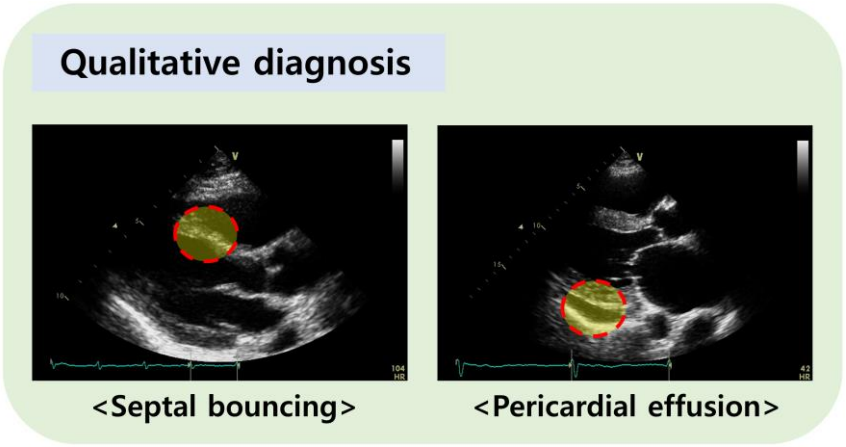


Figure 1.5. Examples of qualitative diagnosis

1.2.2. Previous studies and clinical implications of this study

Imaging diagnosis consists of image acquisition and image diagnosis. In CT and MR, there is no human factor in image acquisition. However, in echocardiography, expertise is needed to get the optimal image for the diagnosis. Therefore, not all cardiologists are capable of performing the examination. Because of the increasing number of examinations, in many Western countries and even in many institutions in our country, there is a tendency to transfer image acquisition tasks to the sonographer. To ensure the quality of image and for the training purpose of the beginners, an AI system for image acquisition such as ′Caption guidance′ has already been released in the market.

Regarding image interpretation, imaging diagnosis is composed of quantitative and qualitative analysis. Several studies have been carried out in image interpretation as shown in Table 1.

For the image interpretation, the first task is the segmentation of the echocardiogram image [12, 13], which compartmentalizes the structures of the heart and compared the difference between the cardiologists′ compartmentalized images. If these segmentation models are well trained, clinicians can automatically get figures related to changes in width or diameter immediately without having to measure them directly in the echocardiogram image.

The second task is to extract cardiovascular parameters [14, 15] or end-diastolic / end-systolic phase [16] from the echocardiogram image and to distinguish between normal and abnormal. It is a task that allows various disease-related indicators to be obtained from echocardiogram

images. In this task, the end-systolic and end-diastolic phases should be detected accurately, so that the clinician can see these specific timings more carefully.

The final task is to determine whether there is a qualitative abnormality. As an end-to-end method, a deep learning network can directly classify the normal and abnormal findings present in the image [17, 18]. Most of these studies suggest what is called a class activation map, indicating the part of the image where the network is significantly referred to.

In the quantitative analysis, the majority of the current studies focused on the calculation of EF. EF is the most important parameter usually calculated in echocardiography. However, other parameters should routinely be obtained in echocardiography. Previous studies used apical 4 and 2 chamber views to obtain EF by summation of disks method (Biplane Simpson′s method). In these views, other standard parameters cannot be obtained. In this current study, the PLAX view was used, in which standard M-mode measurements have been made in daily clinical practice.

In the qualitative analysis, one group of previous studies focused on the accurate diagnosis of a certain ′specific′ disease. These studies intended to get help in the diagnosis when one encountered subtle abnormal findings. The other group of studies was pursued to find out prognostic indicators in images among a large number of patients.

For the CDSS that is useful in daily clinical practice, in addition to providing parameters usually measured in the standard M-mode measurement, CDSS can detect any abnormality and, desirably, suggest

the diagnosis. In this study, the PLAX view was used in which all parameters measured in the standard M-mode measurement can be obtained. We also trained the system not only with the normal image but also with the same number of images with pathologic findings. As CDSS to be developed in this study is capable of both quantitative and qualitative analysis, and diversifies pathologic findings in the qualitative analysis, this CDSS may play a robust role in clinical practice.

Task	Author	Performance	**Characteristic of study**
Segmentation	Sarah Leclerc et al. [12]	LV volume mean correlation 0.95	**Quantitative Analysis**
	Neda Azarmehr et al. [13]	Dice coefficient 0.92±0.05	
Parameter extraction and Abnormal detection	Amirata Ghorbani et al. [14]	Predict presence of pacemaker leads (AUC = 0.89), Predict enlarged left atrium (AUC =0.86),	**Quantitative Analysis**
	Jeffrey Zhang et al. [15]	Detect hypertrophic cardiomyopathy, cardiac amyloidosis, and pulmonary arterial hypertension (AUC=0.93, 0.87 and 0.85, respectively)	
Detecting cardiac Sarcoidosis	Susumu Katsushika et al. [17]	Detect cardiac sarcoidosis (AUC = 0.855, 95% CI: 0.735–0.975)	**Qualitative Analysis**
Detecting valvular diseases	Majid Vafaeezadeh et al. [18]	Detect Cardiac Sarcoidosis (80% rate of model accuracy)	
ES/ED phase detection	Zhibin Liao et al. [16]	Detect ES/ED phase (Average, 0.20 and 1.43 frame difference the ED and ES frames,)	**Quantitative Analysis**

Table 1.1. Studies of quantitative and qualitative analysis of echocardiogram. Related tasks include segmentation, parameter extraction, measurement of dimensions, and phase detection. Although each study shows excellent performance, AI system providing both quantitative and qualitative analysis is lacking.

Author	**Characteristic of study**
Sarah Leclerc et al.	Apical 4 chamber and apical 2 chamber
Neda Azarmehr et al.	Apical 4 chamber views
David Ouyang et al.	Apical 4 chamber views
Amirata Ghorbani et al	Apical 4 chamber
Jeffrey Zhang et al.	Apical 2, 3, 4 chamber, PLAX, PSAX
Susumu Katsushika et al.	Apical 4 chamber
Majid Vafaeezadeh et al.	Apical 4 chamber
Huang et al.	Apical 2, 4 chamber, PLAX, PSAX

Table 1.2. Most of the current echocardiogram studies are conducted mainly on the apical 4 chamber and the apical 2 chamber (PLAX: Parasternal Long Axis view, PSAX: Parasternal Short Axis view)

1.3. Technical Background

1.3.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs), a type of artificial neural network in the field of machine learning, are designed with inspiration from the working of nerves. Unlike a structure in which all neurons in each layer are fully connected to neurons in the previous or next layer in the existing neural network, the filter is placed so that neuron values are calculated in duplicate according to the movement of convolution filters. In addition to the layer performing the convolutional operation, a pooling layer, a dropout layer, and a rectified linear unit layer (ReLU) are used. From LeNet [19] proposed by LeCun to the present, CNN has been widely used in various fields such as video recognition [20], recommendation systems, and natural language processing systems.

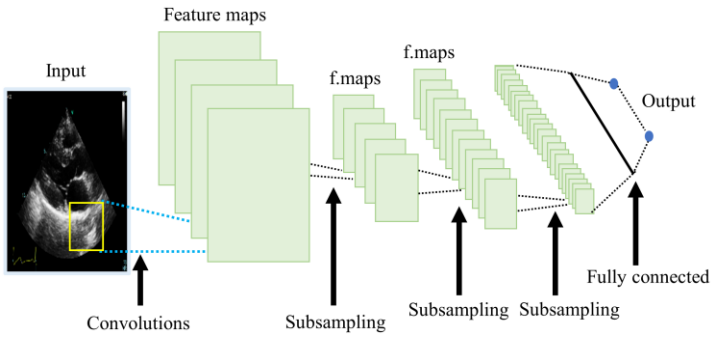


Figure 1.6. A typical shape of Convolutional Neural Network

CNN calculates the loss value in the soft-max loss layer for the probability distribution of each class output from the final layer. CNN proceeds with the learning according to the change in the calculated loss value. Equation (1) is a method of obtaining the soft-max probability distribution for each class, and Equation (2) is a method of calculating the loss value of the probability distribution.

$$p_k = \frac{e^{x_k}}{\sum_{k=1}^{N} e^{x_i}} \qquad (1)$$

$$Loss = \frac{1}{N}\sum_{i=1}^{N} \log(p_i, l_i) \qquad (2)$$

CNN's training proceeds to optimize weight variables such as a filter according to changes in the calculated loss values from the soft-max loss layer. Optimization proceeds using the Stochastic Gradient Descent (SGD) method. SGD updates the weight variables to be used in the next learning step through the change in the loss value according to the weight map changed in the previous learning step. In this process, the learning rate and momentum constants are used. In CNN's soft-max loss layer, computation results are transferred through backpropagation algorithms to higher layers, full-connect layer, activation function, pooling layer, and convolution layer, and each layer has a new weight variable through computation through backpropagation algorithms. The learning rate is a predetermined constant and is multiplied by a constant between 0 and 1 for each predetermined learning step. Finally, it aims to minimize the loss value by updating the parameters inside CNN. In the echocardiogram, CNN is being used in many ways [21, 22].

1.3.1.1. U-net

In this dissertation, the U-net model based on CNN Semantic Segmentation is modified and used to suit the echocardiogram. The U-net model is a structure developed for biomedical image segmentation [23, 24]. Training about a specific object should be conducted in advance to separate the structure to be found from the image from other structures or backgrounds. The function of the U-net structure is to connect the output of the contraction path for each level and the input of the expanding path. These connections allow high spatial resolution capabilities to be confined to a fully convolutional network, resulting in more accurate output based on information. The contracting path is configured by repeatedly applying 3×3 convolution for each convolution layer, and ReLU was used as the active function. For each pooling layer, a 2×2 max pooling operation with strand 2 is similarly used for downsampling. Each time the layer goes down, the number of channels doubles and is downsampled. The expanding path used 3×3 convolution for each convolution layer, doubling the number of channels as the layer climbed, and using ReLU as the activation function. The blue arrow connecting the contracting and expanding paths displays copy and crop, and the input is made to affect the output. In the final layer, 6 classes were classified according to the number of heart structures. In recent studies, U-net showed good results in segmentation [12, 25, 26].

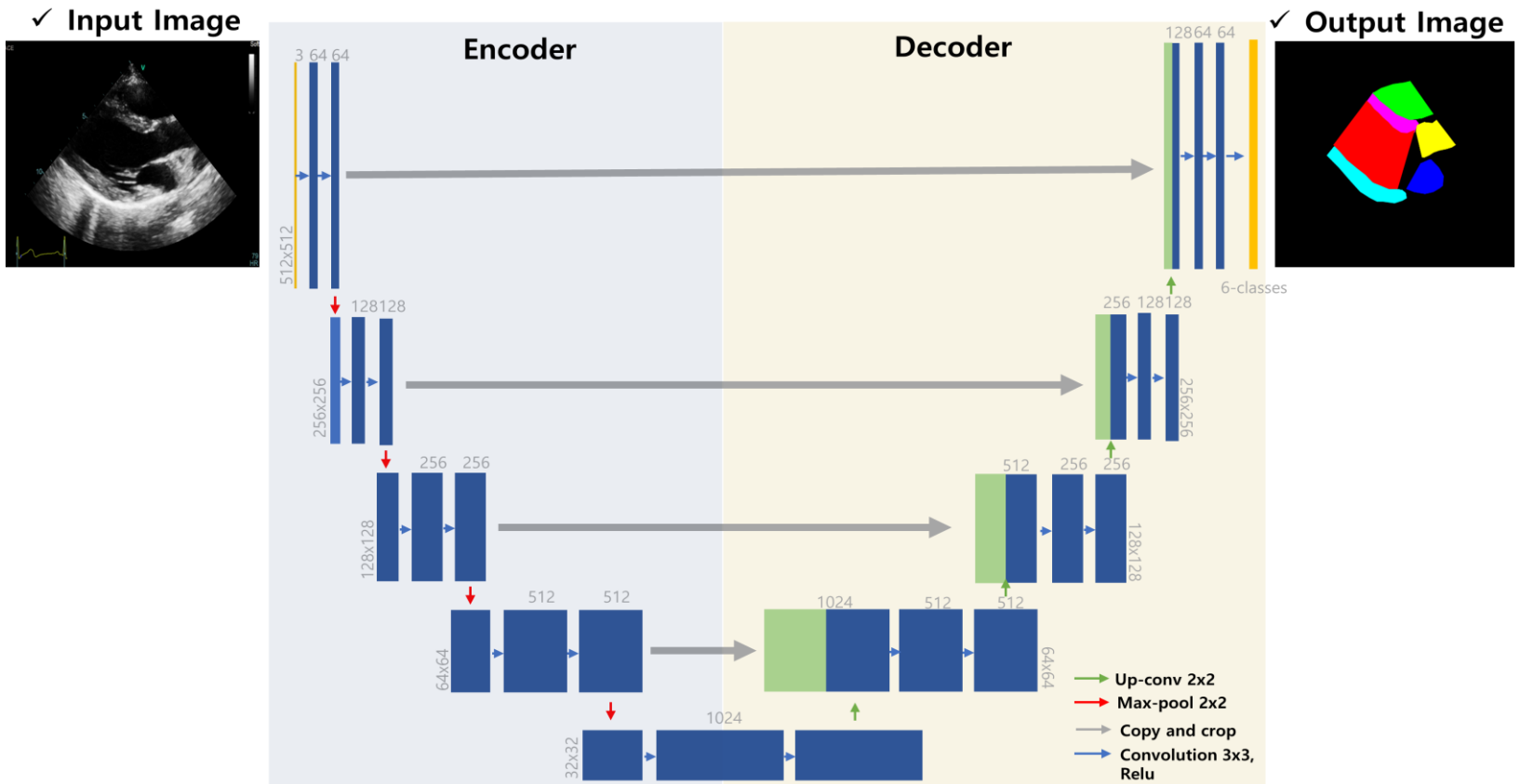


Figure 1. 7. U-net structure for segmentation of 6 chamber

1.3.1.2. Residual Network

ResNet is the architecture that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2015. While the previous year's winning team had 22 layers of GoogleNet architecture, the ResNet architecture consisted of 152 layers, which were approximately 7 times deeper. This deep stacking of the layers of the network allowed the top-5 error to be lowered, and it is still widely used to solve the image classification problem. The existing network structure showed poor results above a certain depth as the neural network was deeply stacked. In other words, the depth of the neural network did not guarantee the best results. To solve these problems, the ResNet research team proposed Residual Block, which allows results to improve as the neural network is deeply stacked. Figure 1.8 shows the Residual Block [27]. The

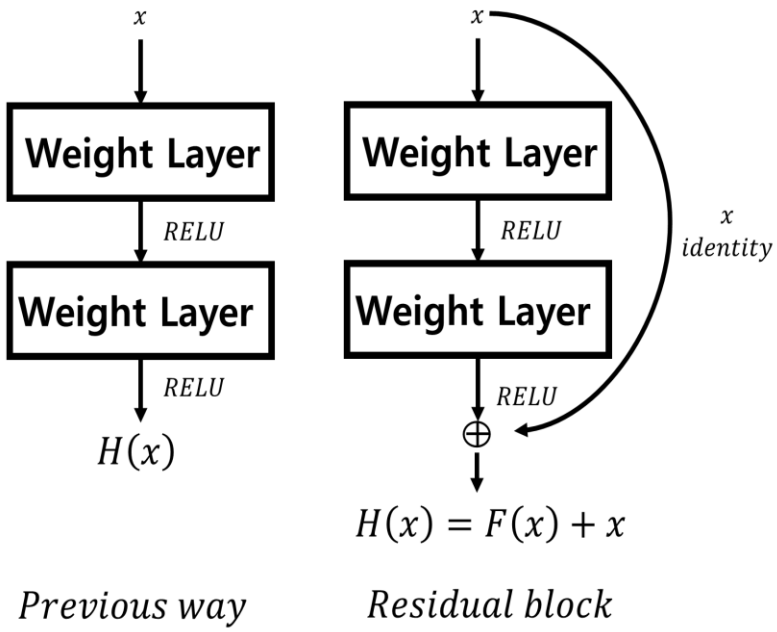


Figure 1.8. Traditional neural network learning and residual block learning

addition of a shortcut that adds input to the output value distinguishes it from the existing neural network layer. .

If the purpose of existing neural networks was to obtain a function $H(x)$ that maps the input value x to the target y, ResNet aims to minimize $F(x)+x$. At this time, since x is a value that cannot be changed, the purpose should be to make $F(x)$ close to zero. $F(x)$ can be expressed as $H(x)-x$, which reduces $H(x)-x$ to a minimum. $H(x)-x$ is called the residual and this network is called ResNet because it minimizes the residual.

Figure 1.12 shows the configuration of the ResNet-152 architecture. The ResNet architecture built in this way solved the existing problem that the results deteriorated as the neural network deepened, and the network was deeply stacked to improve performance. In addition, when the neural network was built up with 18, 34, 50, 101, and 152 layers, the deeper the network, the better the performance. In this paper, the ResNet-152 architecture is utilized as a way to distinguish normal and abnormal echocardiogram. Many previous studies have demonstrated that Residual Network is a suitable structure for finding diseases in the echocardiogram [28, 29].

1.3.1.3. Gradient-weighted Class Activation Mapping (Grad-CAM)

Learning through a deep learning network is a series of nonlinear operations such as deep networks and active functions, therefore, it is difficult to know through which process the correct answer was derived. Hence, several techniques are being studied to estimate and visualize the process of predicting the correct answer by the learned classification model. Among the representative learning visualization techniques, a class activation map (CAM) was studied to view the class of interest predicted by the classification mode [30]. The CAM uses the last convolutional layer before the feature map passes through the fully connected (FC) layer, which loses spatial information and becomes flattened, to identify the region of interest of CNN. We extract k feature maps from the last convolution layer and call them f_k. All feature maps are passed through Global Average Pooling (GAP) to obtain F_k. When we obtain the CAM for class c, we should obtain S_c by the softmax value of the multi-classification function, through the product of weights w_c and F_k toward c, and we obtain M_c. The expression is as follows.

$$F_k = \sum_{x,y} f_k(x,y) \qquad (3)$$

$$S_c = \sum_k w_k^c F_k = \sum_{x,y} \sum_k w_k^c f_k(x,y) \qquad (4)$$

$$M_c = \sum_k w_k^c f_k(x,y) \qquad (5)$$

The region of interest of CNN could be confirmed through the

product of the value corresponding to (x, y) of the last feature map and the weight w_c of class c; $M_c(x, y)$ can be seen as the influence or interest of the classification model on classifying the image into class c. However, the CAM has the disadvantage of being able to see the region of interest only in the last convolution layer, and GAP must be followed after the last convolution in the classification network. Therefore, techniques that can be used in general models have been studied to improve these shortcomings.

The gradient-weighted class activation mapping (Grad-CAM) can check the CAM in all convolutional layers through the slope of the weights generated during the backpropagation of the classification network [31]. The Grad-CAM process is as follows. After backpropagating with k feature maps A_k through the convolution layer and y_c, the predicted value of class c, and the weighted slope value of A_k, the position is multiplied and added to all the values a_k with GAP according to the order of k channels. Then, it passes through the activate function Relu to obtain a Grad-CAM L_c for class c.

$$a_k = \frac{1}{Z} \sum_i \sum_j \frac{\sigma y_c}{\sigma A_{i,j}^k} \qquad (6)$$

$$L_{Grad-CAM}^c = Relu(\sum_k \alpha_k^c A_k) \qquad (7)$$

Unlike the CAM, the Grad-CAM can obtain A_k and a_k at the desired convolution layer, therefore, the CAM at the middle point of the network can be obtained. In the process of calculating a_k, the GAP is calculated separately from the classification network, therefore, GAP is not forced on the network. In this study, we applied it to the last layer of Resnet, a classification network that adopted Grad-CAM. In a structure

consisting of 3×3 convolution Deepwise and 1×1 convolution Pointwise, the feature map after Depthwise was set to A_k to obtain Grad-CAM. Figure 7 shows the Grad-CAM structure.

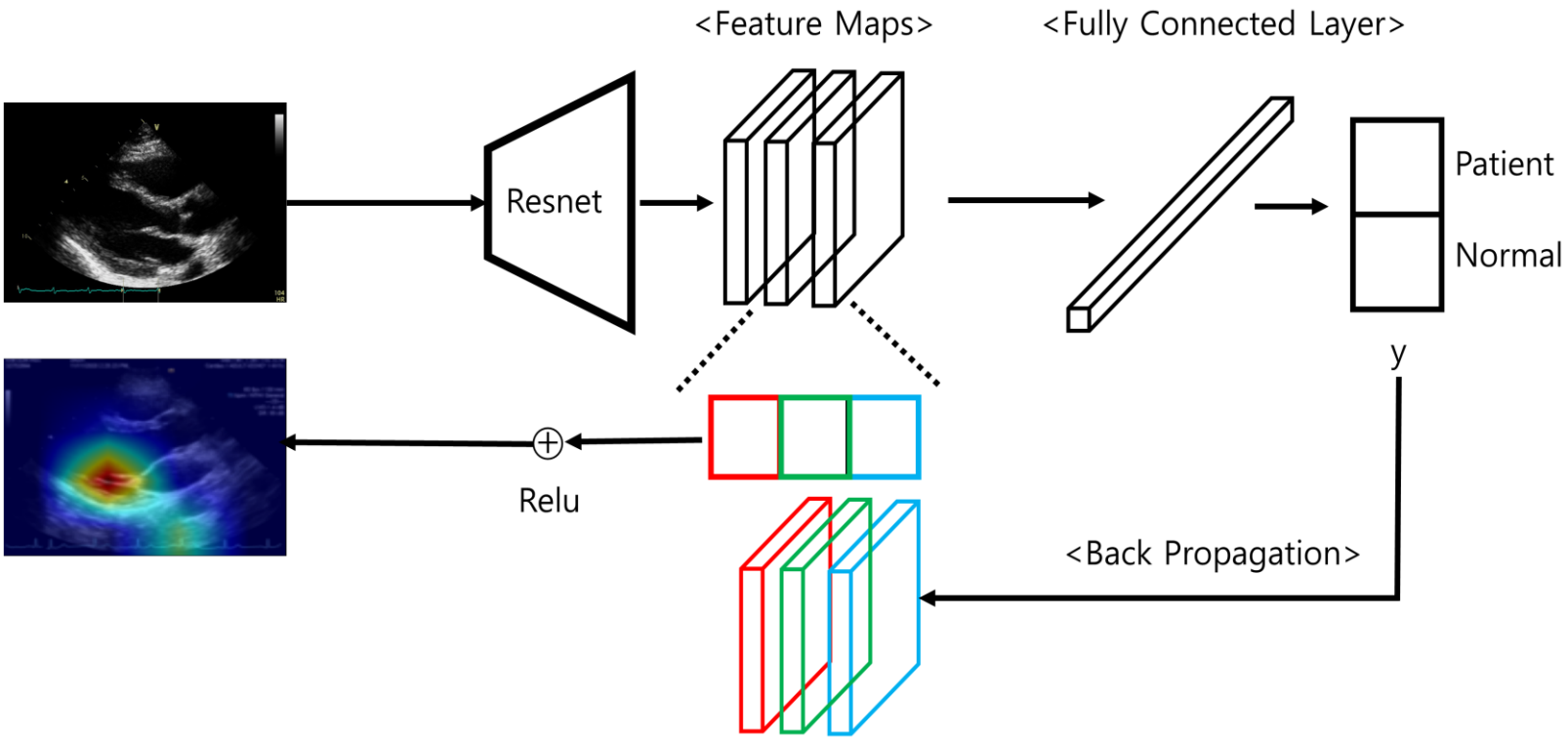


Figure 1.9. Architecture of Grad-CAM used in this study. When any network released the result as an output, Grad-CAM provides the basis for judgement by looking at the feature map of the last convolution layer.

1.4. Unmet Clinical Needs

So far, most of the echocardiogram-related CDSS have been developed separately for qualitative and quantitative diagnosis. However, to develop a reliable and convincing CDSS, it is necessary to develop a network similar to the flow of clinicians' diagnoses, and such a system can be developed using existing deep learning algorithms. It is important to develop a system that can make both quantitative and qualitative diagnoses by analyzing echocardiogram images, and it is necessary to verify this as various heart disease groups can appear in the image.

1.5. Objective

The objective of this dissertation is to develop and validate an AI-based CDSS for echocardiography. An optimal CDSS, which is useful to clinicians must have the following characteristics:

1) Accurate compartmentalization of anatomical structures in images and quantitative diagnosis based on them.
2) Accurate detection of the end-systole and end-diastole of the heart and the presence of pathologic findings as shown in the video.
3) A system for classifying normal and abnormal findings in patients, and the diagnosis based on this should be presented in the video.

CHAPTER 2

Materials & Methods

2. Materials & Methods

2.1. Data Description

The study was approved by the institutional review board of Seoul National University Hospital (IRB No. 2201-110-1293). As a retrospective study, this study analyzed 2,600 echocardiographic data stored in the Seoul National University Hospital Picture Archiving and Communication Systems (PACS) from 2016 to 2021, consisting of data of 1300 non-cardiac patients with normal echocardiograms and 1,300 cardiac patients. The data compositions for the training of quantitative and qualitative networks were different, and for quantitative networks, training used the method of 5-fold cross-validation with 300 annotated images and was verified with 1300 images with pathologic findings as shown in Figure 2.1. In addition, information on end-systolic and end-diastolic frames extracted from the results of quantitative networks was verified with 300 annotated data. On the other hand, for a qualitative network, training was done with data of 1000 out of 1300 normal echocardiograms and 1300 echocardiograms with pathologic findings and verified with data of 600 echocardiograms with pathologic findings as shown in Figure 2.1. As shown in Table 3.1, several state of art echocardiograhic machines from different vendors were used.

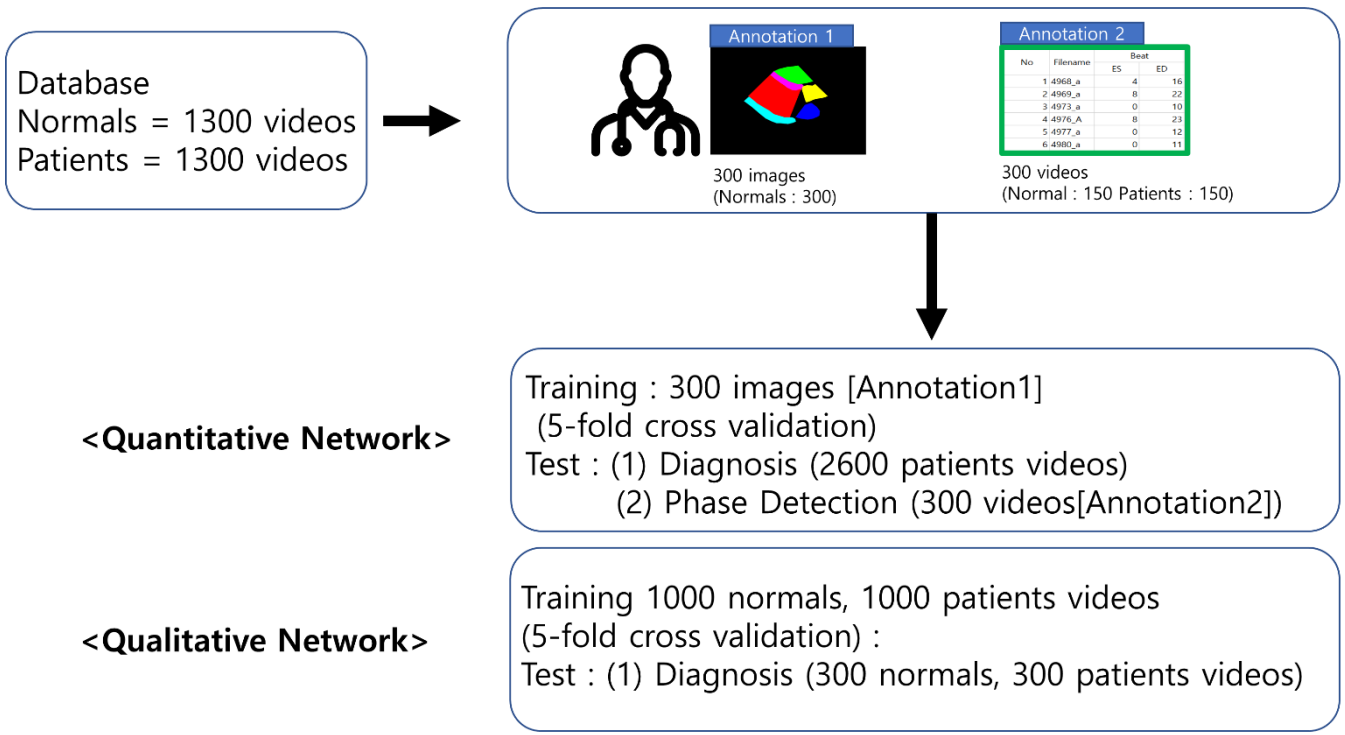


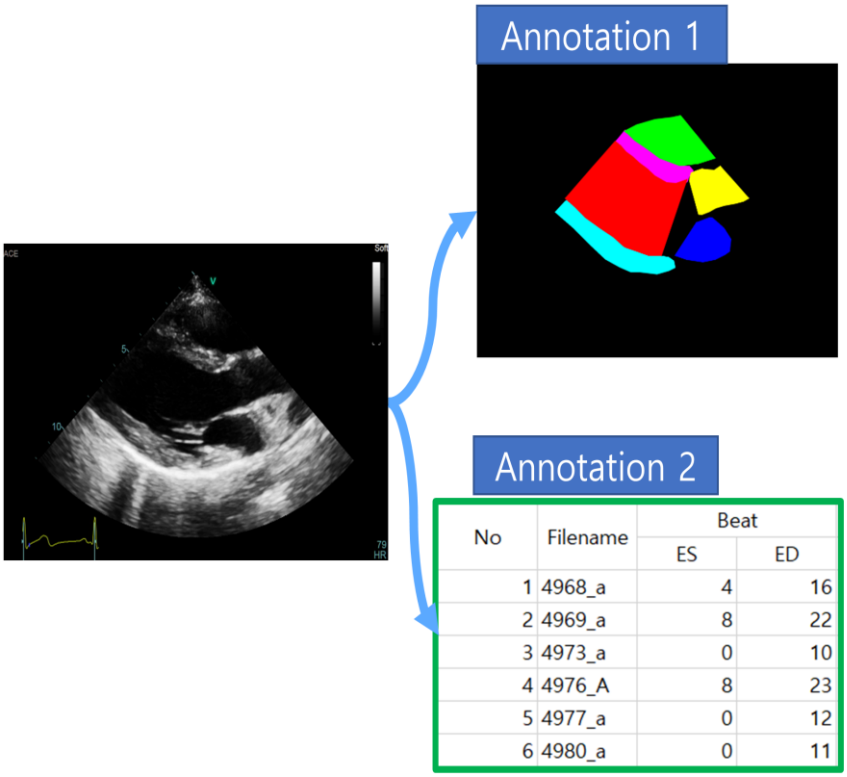
Figure 2.1. Block diagram of the data selection process for quantitative network. To train quantitative networks, 300 images segmented by cardiologists were used for segmentation. 1300 images with pathologic finding and 300 annotated images were used as data to verify them. 1000 patients and normal finding were used to train the qualitative network, and 300 images with pathologic and normal findings were used for verification. All networks were trained using the 5-fold cross validation method.

Table 2.1. Echocardiographic machines used in the study. Various models of machines from 3 different vendors were used.

Manufacturer name	**Model name**	**Number of devices**
GE Healthcare	Vivid S70	12
	Vivid q	1
	Vivid E9	769
	Vivid E95	954
SIEMENS	Vivid 7	13
	ACUSON SC2000	408
	ACUSON SEQUOIA	1
Philips Medical Systems	EPIQ 7C	424
	CX50	18
Total		2600

2.2. Annotated Data

Images segmented by a cardiologist are required to train the segmentation network. Overall, 300 segmented images were provided. In addition, an excel table containing end-systolic and end-diastolic frames was prepared for 300 echocardiograms, as shown in Figure 2.3.



No	Filename	Beat	
		ES	ED
1	4968_a	4	16
2	4969_a	8	22
3	4973_a	0	10
4	4976_A	8	23
5	4977_a	0	12
6	4980_a	0	11

Figure 2.2. Block diagram of the data selection process for qualitative network. Echocardiographic images compartmentalized by a cardiologist (annotation 1) for training in quantitative analysis networks. Excel files (annotation 2) of the end-systolic and end-diastolic frames

2.3. Overall Architecture

Two deep learning networks were used for both quantitative and qualitative analysis. As shown in Figure 2.4, a segmentation task-related network was used for the quantitation network, through which functional abnormalities of the heart could be monitored in real-time, and the information related to systolic and diastolic dimensions could be extracted. In addition, for qualitative analysis, a network was constructed to determine the presence of pathologic findings in the image from the data based on the extracted systolic and diastolic frames. In the second network, the input data was used to indicate the abnormal finding visually.

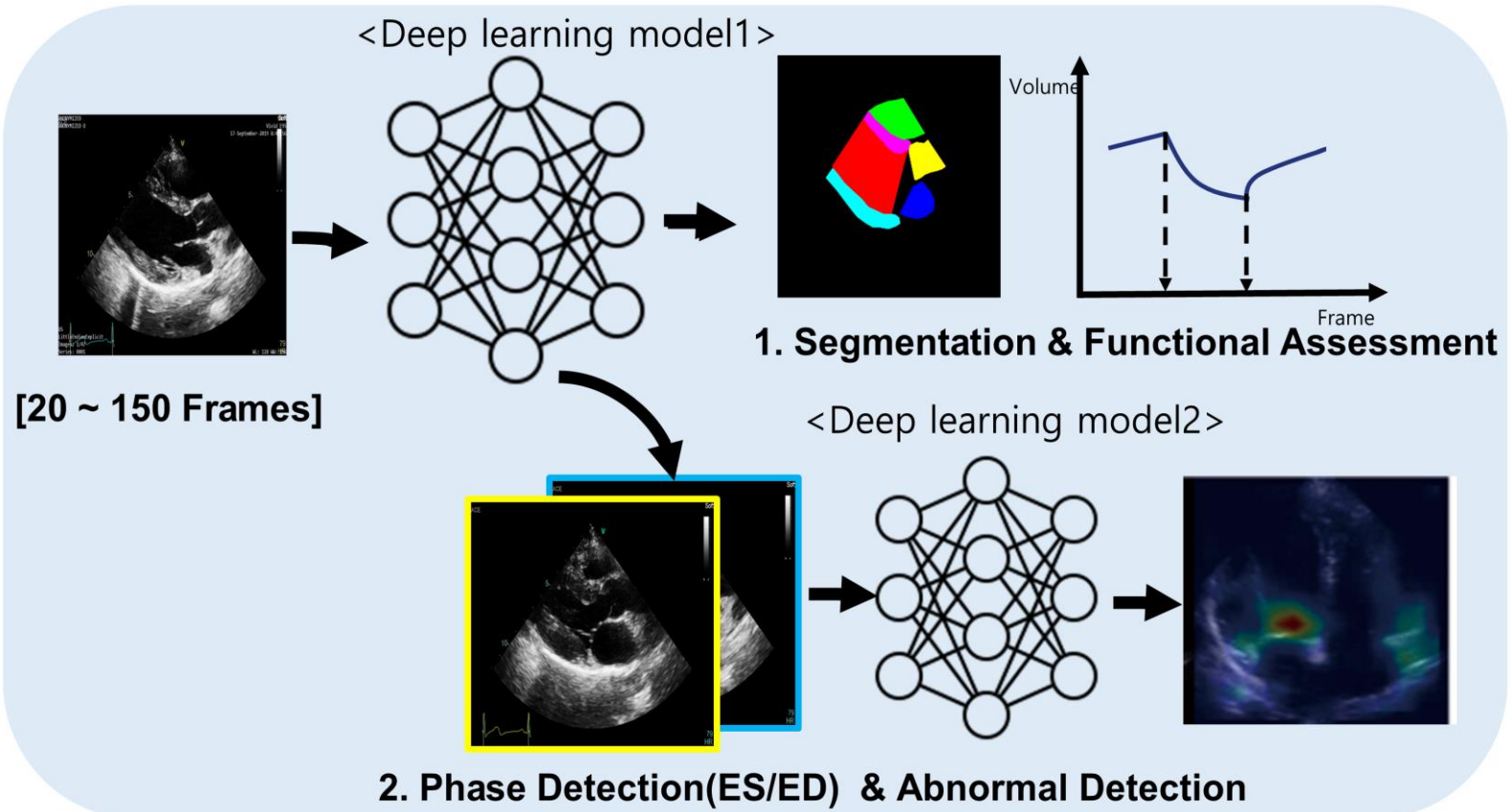


Figure 2.3. An overall network structure that includes networks for quantitative and qualitative analysis

2.3.1. Quantitative Network

Quantitative networks performing segmentation tasks were based on U-net, which already showed high accuracy in the previous studies [32, 33]. Three hundred paired echocardiographic images segmented by a cardiologist were used. Data were trained through 5-fold-cross validation, batch size was 10 and epoch was 300 times. Optimization was performed through Adam optimizer, and the learning rate was set at 0.0001.

After segmentation was completed, each structure's contour was obtained together with the dimension of the short axis. After vectorized dimension of the short axis for each frame, the first minimum and maximum points of the whole vector were extracted and defined as end-systolic, and end-diastolic frames, respectively, and stored as variables for the second qualitative network.

With this quantitative analysis, diagnoses were made using diagnostic criteria shown in Figure 1.4.

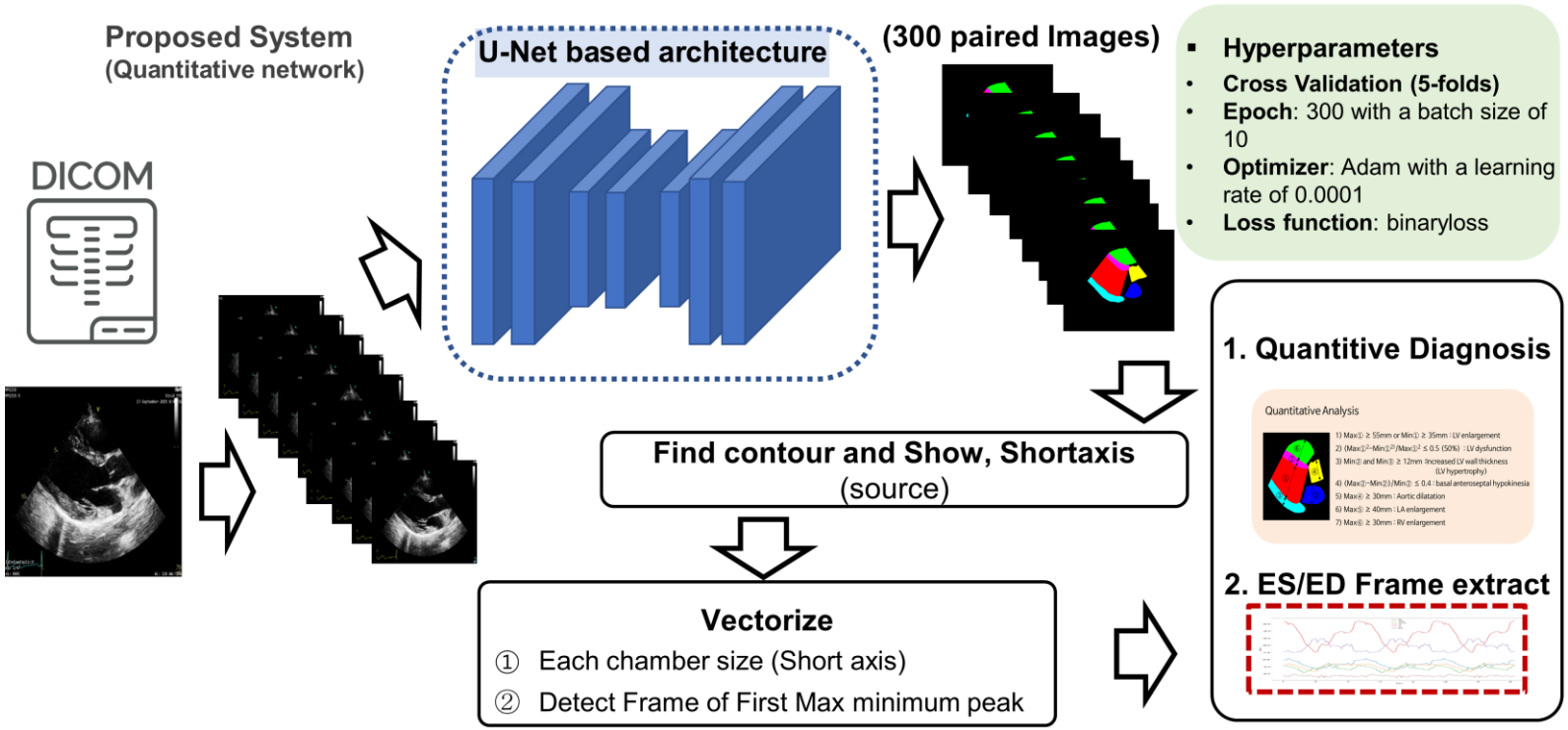


Figure 2.4. Quantitative network for the segmentation and measurement of dimensions.

2.3.2. Qualitative Network

In the qualitative network, based on end-diastolic and end-systolic frame information from the previous network, we used Resnet, a network that selects a total of 10 input images at end-diastole and end-systole as seen in Figure 2.5. Data were trained through 5-fold-cross validation, with a batch size of 10- and 300-times epoch. Optimization was carried out through Adam optimizer, and the learning rate was carried out at 0.0001.

The Grad-CAM methodology is applied to the last layer of Resnet. To check the validity of the input images selected from the quantitative analysis, the training was performed in the same environment with 10 randomly selected images from the input. If the selection process of the end-systolic and end-diastolic frames referenced above is valid, it is assumed that the network will show better diagnostic performance.

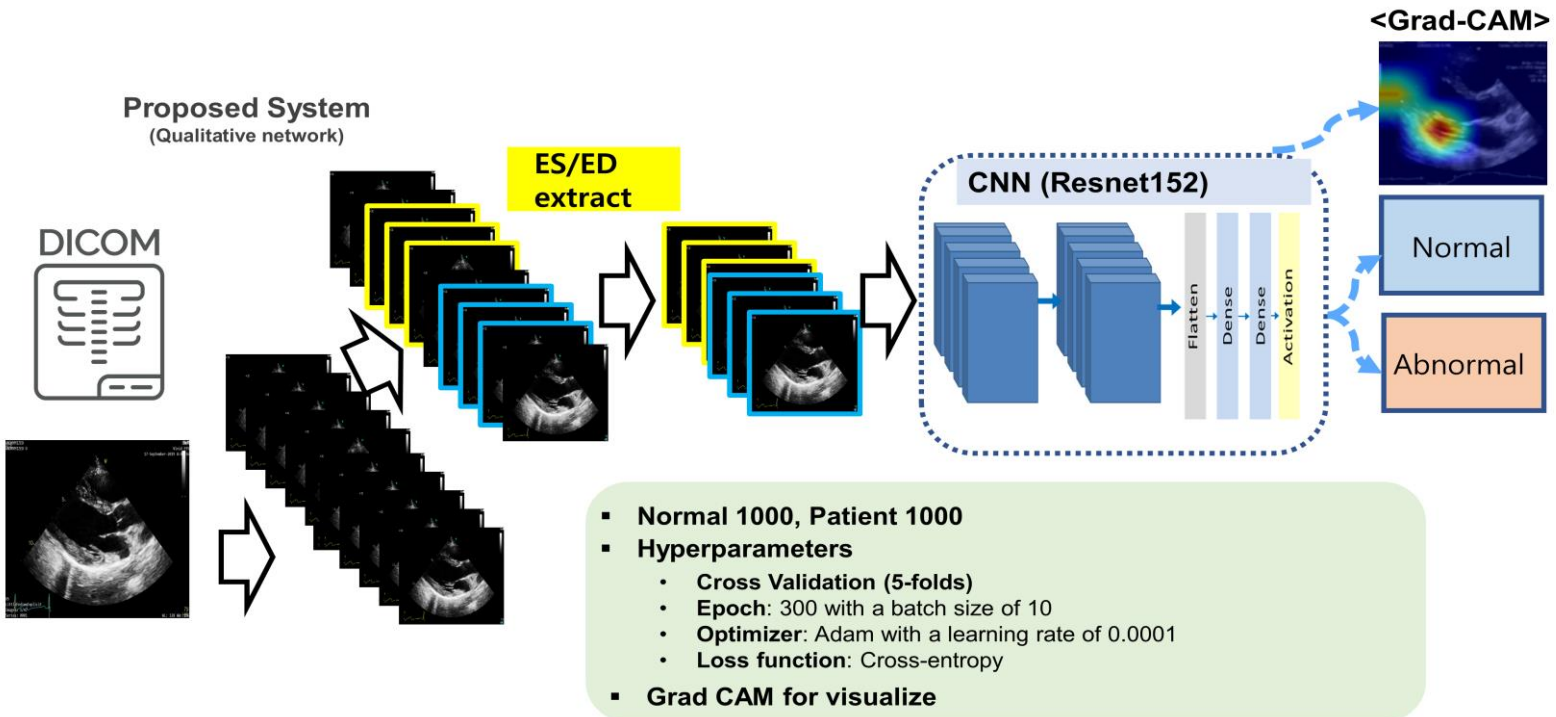


Figure 2.5. Qualitative network for the pathology and visualization of abnormal findings.

2.4. Dice Similarity Score

To compare the segmentation results in this paper, the Dice similarity coefficient (DSC) was calculated and compared with the Ground Truth image segmented by the clinician. As shown in Figure 2.6, the sum of the areas of the two images becomes the denominator, and the double of the intersection of the areas of the two images becomes the numerator. If the images are similar, it has a value closer to 1[34].

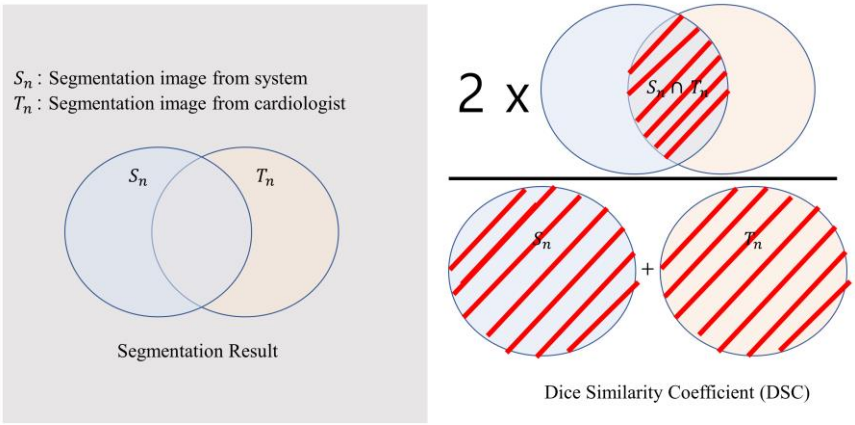


Figure 2.6. Dice similarity coefficient. It is twice the intersection of the two images over the sum of the areas of the two images. It takes a value of 1 when completely overlapping and a value of 0 when completely falling.

2.5. Intersection over Union

Besides the previous DSC coefficients, to compare the segmentation results, the degree of overlap between the predicted value and the ground truth is used as Intersection over Union (IoU). Referring to Figure 2.7, this coefficient is defined as the intersection area over the union area of the two images. A coefficient closer to 1 represents better performance. In most cases of segmentation or bounding box, this IoU value is used as an indicator of accuracy. The meaning is similar to the DSC coefficient, but since each paper uses a different index, the corresponding values were also compared.

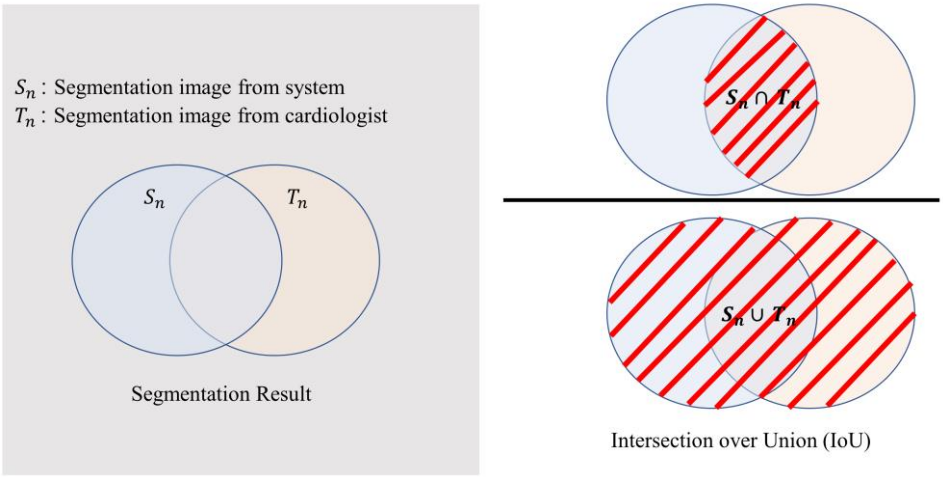


Figure 2.7. Besides the DSC coefficients, in order to compare the segmentation results, the degree of overlap between the predicted value and the ground truth is used as Intersection over Union (IoU). IOU of the two image areas has a value between 0 and 1. Better performance is represented by a number close to 1.

CHAPTER 3

Results & Discussion

3. Results & Discussion

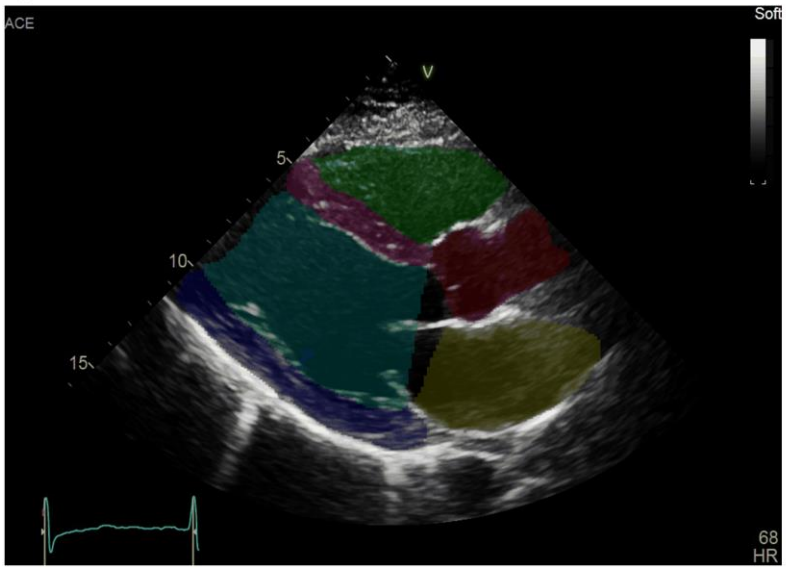
3.1. Quantitative Network Result

The segmentation by the quantitative network was performed on normal and abnormal echocardiograms. As shown in Figure 3.1, we can confirm that dimensional changes of the structures are properly reflected according to the systolic and diastolic phases of the heart. The dimensional change of the structures was tracked well by the network developed in this study, based on U-net. In addition, in Figure 3.2, in the echocardiogram showing decreased LV systolic function, the area change of the segmented LV is reduced reflecting the decreased left ventricular systolic function. This decreased LV systolic function is quantitatively reflected in the vector. In both normal echocardiogram and echocardiogram with pathologic findings, it was assumed that end-systolic and end-diastolic timing can be accurately decided by the minimum and maximum point on the graph, respectively.

However, tracking performance in the interventricular septum, left ventricular posterior wall, and aorta was somewhat inferior to other segments. It is probably because these structures have smaller dimensions than other structures and our tracking system is not fine enough for these structures.

In this study, comparable or slightly better results were obtained in segmentation compared to the previous study as shown in Table 3.2.

a)



b)

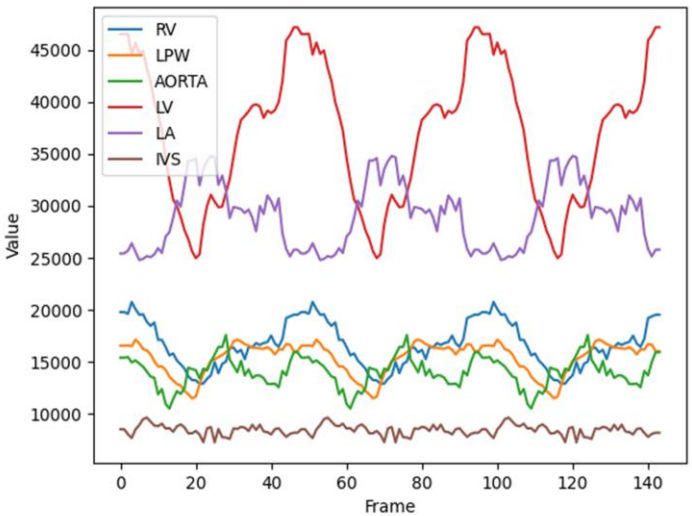
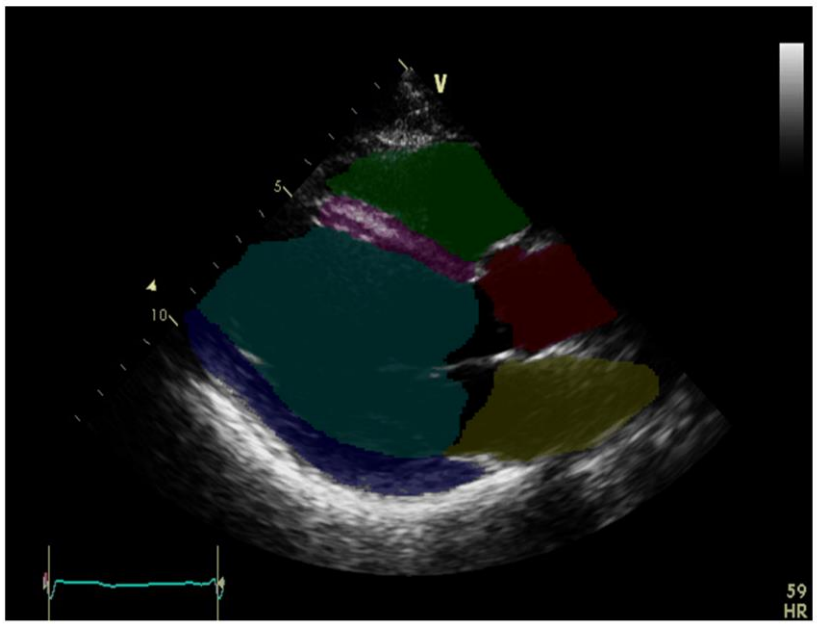


Figure 3.1. a) Segmentation in normal echocardiogram, b) Area of each chamber represented by the number of pixels shows change in area during the cardiac cycle.

a)



b)

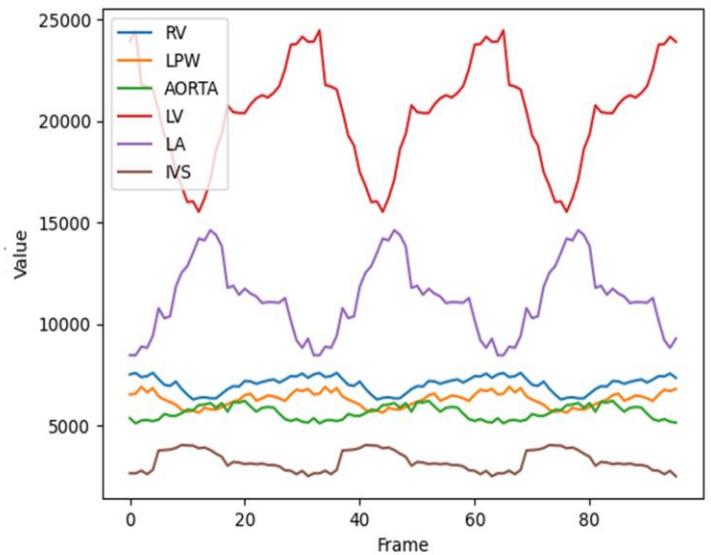


Figure 3.2 a) Segmentation in echocardiogram with decreased LV systolic function, b) Decrements or increments in areas during the cardiac cycle are decreased compared to the normal values.

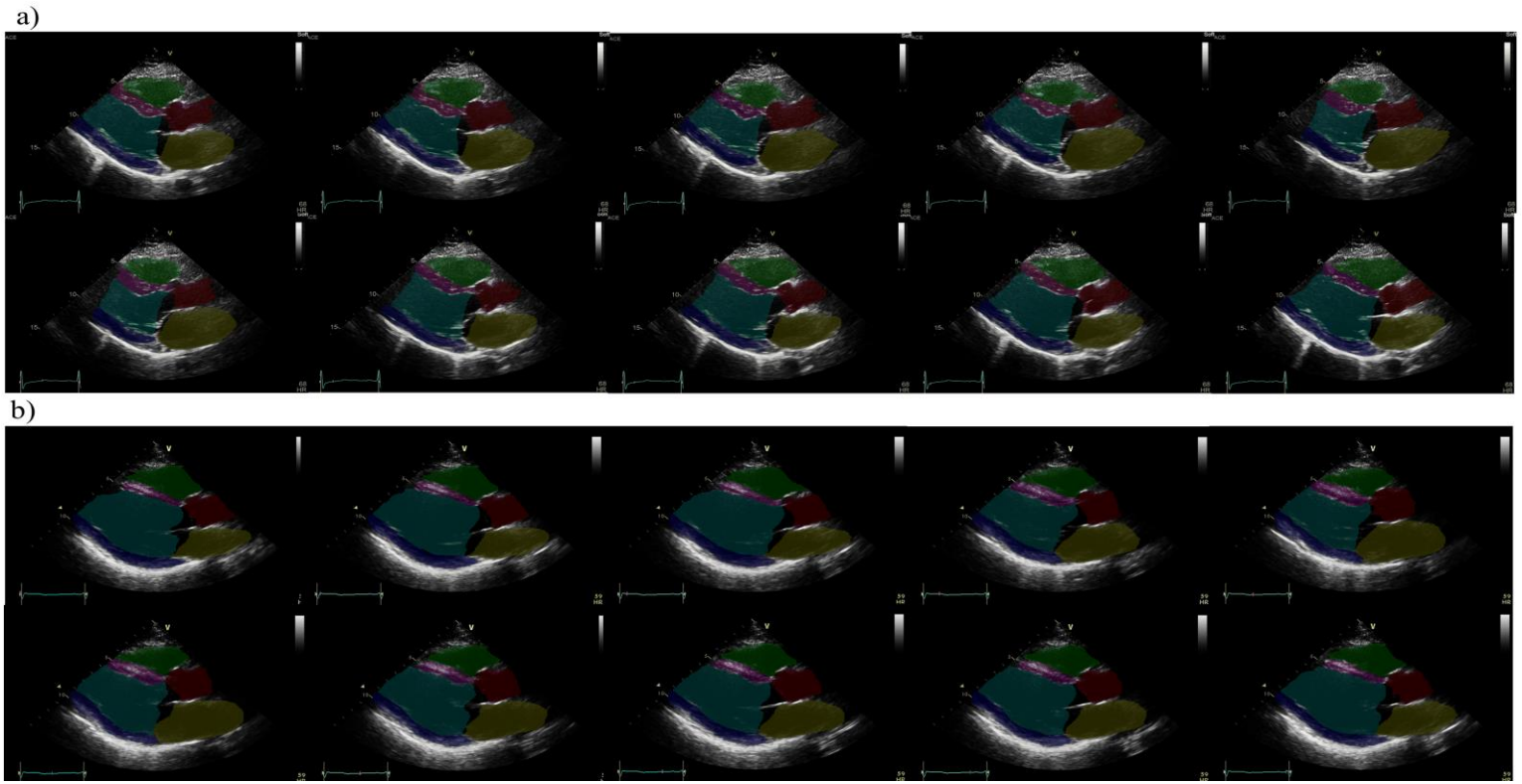


Figure 3.3. Still frames of the whole cardiac cycle a) in normal and b) in patient with dilated LV with decreased systiolic function. Note the optimal segmentation in whole frames.

`	Parameter	Segmentation region	Reference studies	Our study
Jeffrey Zhang et al. [15]	IoU	Left ventricle	0.88	0.92
		Left atrium	0.86	0.89
		Right ventricle	0.85	0.88
		Aortic	0.86	0.84
		Interventricular septum	0.77	0.76
		LV posterior wall	0.75	0.73
	AUC	Averaged	0.89	0.94
Huang et al. [35]	DSC, IoU	Averaged	0.78 (DSC) 0.61(IoU)	0.91(DSC) 0.84(IoU)
	Sensitivity Specificity	Regional Wall Motion Abnormalities	(Sensitivity: 81.8% Specificity: 81.6%)	(Sensitivity: 90.33% Specificity: 87%)

Table 3.1. Segmemtation performance of previous studies

3.1.1. Diagnostic results

Quantitative analysis was performed in 1300 normal echocardiograms and 1300 echocardiograms with pathologic findings. In seven pathologic findings that can be assessed in the PLAX, it was confirmed that the majority of them showed excellent diagnostic performance with relatively high accuracy and specificity.

However, the diagnostic accuracy was low in pathologic findings related to the interventricular septum, left ventricular posterior wall, and aorta, probably related to the inferior tracking performances in these structures.

Diagnostics accuracy is low in detecting basal anteroseptal hypokinesia with a sensitivity of 42.3% and specificity of 62.8%. However, it is well known among the cardiologist that the diagnosis of hypokinesia, to which the performance of our system was compared, is visually estimated, and therefore, it is subjective. Cardiologists have been pursuing an objective way in the diagnosis of hypokinesia. In future, our system can provide this objective way of diagnosing hypokinesia.

	Quantitive analysis	
	Sensitivity	**Specificity**
LV enlargement (n=580) [Max (a) ≥ 55 mm or Min (a) ≥ 35 mm]	91.55% [0.88 0.93]	93.75% [0.91 0.95]
LV dysfunction (n=330) [(Max $(a)^2$ – Min $(a)^2$)/Max$(a)^2$ ≤ 0.5 (50%)]	94.55% [0.91 0.96]	96.08% [0.94 0.97]
Increased LV wall Thickness (n=280) [Min (b) and Min (c) ≥ 12 mm]	72.14% [0.66 0.77]	80.39% [0.77 0.82]
Basal anteroseptal hypokinesia (n=26) [(Max (b) – Min(b)) / Min(b)]	42.31% [0.23 0.63]	62.79% [0.60 0.65]
Aorta dilation (n=257) [Max (d) ≥ 30 mm]	78.99% [0.73 0.84]	75.84% [0.73 0.78]
LA Enlargement (n=554) [Max (e) ≥ 40 mm]	97.65% [0.96 0.98]	95.27% [0.93 0.97]
RV Enlargement (n=158) [Max (f) ≥ 30 mm]	93.67% [0.88 0.96]	87.85% [0.86 0.90]

Table 3.2. Sensitivities and specificities of quantitative diagnosis. (a) Diameter of left ventricle, (b) Diameter of interventricular septum, (c) Diameter of left ventricular posterior wall, (d) Diameter of right ventricle, (e) Diameter of aorta, and (f) Diameter of left atrium)

3.1.2. Phase Detection Result

In addition to performing quantitative analysis, frames at or near the maximum and the minimum dimensions of the left ventricle could be picked up and stored. These data were used to extract the input image of the network for qualitative analysis. Images showing left ventricular maximum and minimum dimensions were regarded as the end-diastolic and end-systolic frames, respectively. The accuracy of detecting end-diastolic and end-systolic frames was evaluated by comparing the difference in the number of frames between the frame detected by the network and that decided by a cardiologist.

There was an average difference of 0.52 and 0.9 frames in end-diastolic and end-systolic frame detection, respectively. In the second network, as the five images before and after the end-diastolic and end-systolic frames were used, this difference in the frame seems to be negligible as shown in Table 3.2 and Figure 3.4.

In the quantitative analysis, it was found that sufficiently accurate phase detection is possible even if a network for phase detection was not separately employed as in the preceding paper [16, 36].

Table 3.3. Phase difference between the network output and the cardiologist's annotation. Differences in frame both in end-diastole and end-systole were less than 1.

Frame timing	Mean ± std (frame)
End- diastole	0.52 ± 1.02
End -systole	0.9 ± 1.32

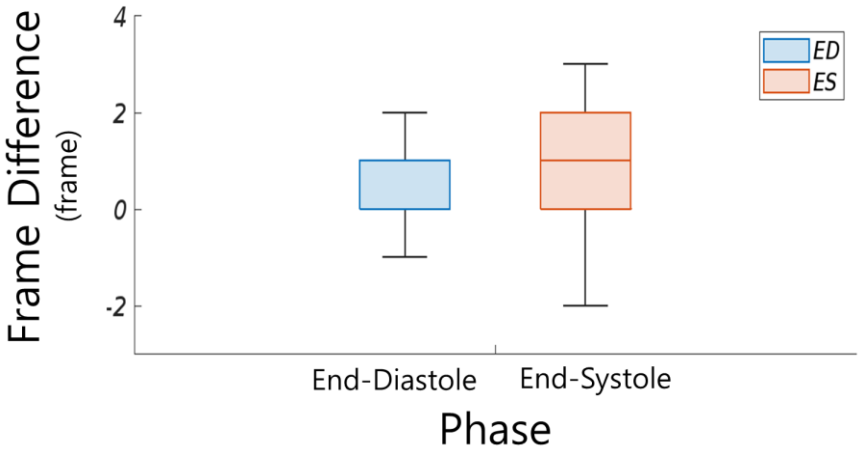


Figure 3.4. Boxplot of end-systolic(ES) and end-diastolic(ED) phase differences between output of quantitative network and the frame annotated by cardiologists

3.2. Qualitative Network Results

To evaluate the accuracy of the qualitative network, accuracy and sensitivity were assessed in 600 patients, and the trained network was compared with 10 randomly selected images.

The area under the receiver operating characteristic (AUROC) was obtained to determine the overall classification performance. The criteria of AUROC values are presented in Table 3.4. The models trained with the random images showed AUROC values of 0.89, indicating good discrimination performance. However, the AUROC value of 0.94 was obtained by the model trained with the input data extracted based on the end-systolic and end-diastolic timing, indicating excellent discrimination performance as shown in Figure 3.4.

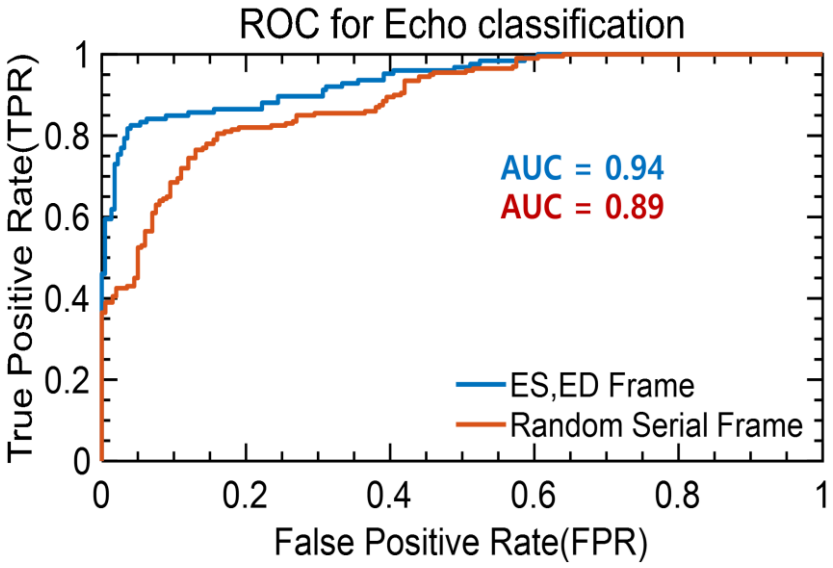


Figure 3.5. The comparison of ROC curves between network trained with random serial input images and ED/ES based input images.

In the qualitative network, the performance of the network improved when the training images were selected based on the systolic and diastolic timings compared to training using random images. Referring to this result, input data based on the systolic and diastolic timing should be selected for the CDSS to be developed in the future as shown in Table 3.4.

Table 3.4. Performance comparison between a network trained with randomly selected 10 images and 10 images based on end-systolic and end-diastolic phase

Input images	Accuracy n (%)	Sensitivity n (%)
Random 10 images	487/600 (81.16%)	235/300 (78.33%)
Network selected images	542/600 (90.33%)	261/300 (87%)

Table 3.5. The rule of thumb for the AUROC. [AUROC: area under the receiver operating characteristic curve]

AUROC=0.5	No discrimination
0.5<AUC≤0.6	Failed discrimination
0.6<AUC≤0.7	Bad discrimination
0.7<AUC≤0.8	Fair discrimination
0.8<AUC≤0.9	Good discrimination
0.9<AUC	Excellent discrimination

Sensitivities and specificities of individual qualitative diagnoses evaluated in this study are shown in Table 3.5. Also in the qualitative diagnosis, pathologic findings related to septum and aorta showed relatively low sensitivity, and this finding is probably related to the suboptimal segmentation of these structures.

Most of the incorrect classification results showed that the segmentation of the LV posterior wall and the interventricular septum was incorrect as shown in Figure 3.6. In most cases, recognizing boundaries is difficult unless you are a skilled cardiologist, and this means that if it is difficult for humans to learn, it is also difficult for a deep learning network.

In learning a quantitative network, learning only with the segmentation image of a normal person may be the cause of this limitation. If images of patients segmented by experienced cardiologists were acquired, made into a database for each disease case, and trained on the network, more accurate segmentation results could be obtained.

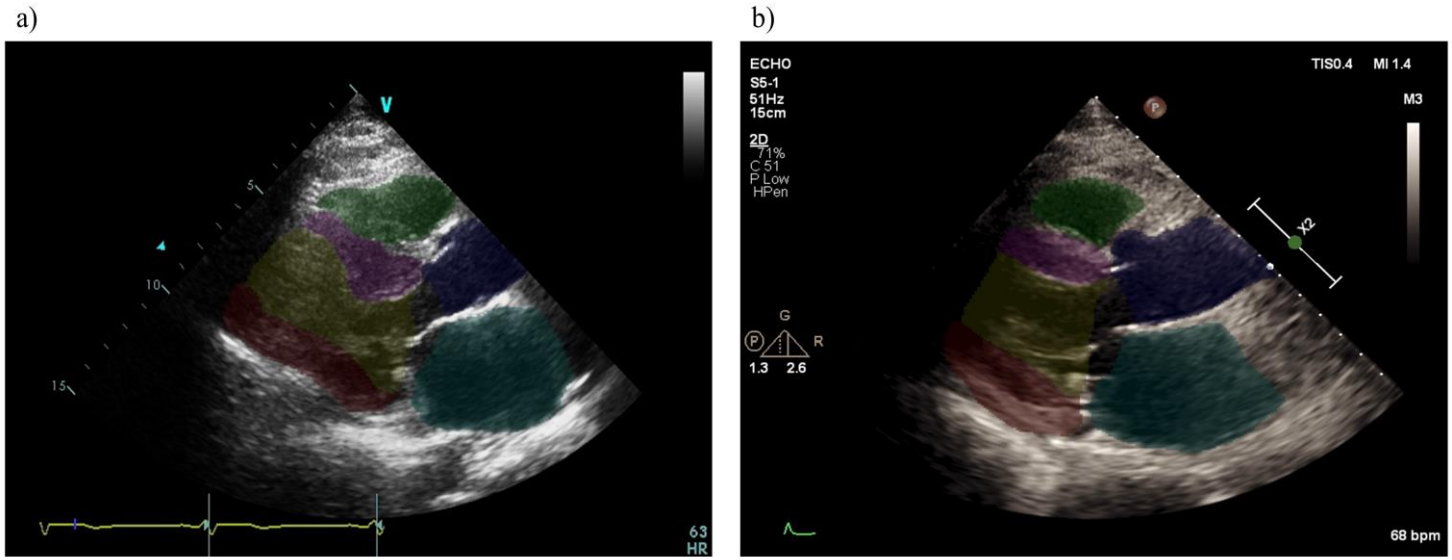


Figure 3.6. a) The label is "Asymmetric septal hypertrophy," however wrongly diagnosis as basal septal hypertrophy, b) The label is "Normal" however wrongly classified as "Increased LV wall thickness."

Table 3.6. Sensitivities and specificities of qualitative diagnosis.

	Qualitative analysis			
	Sensitivity	**Quantitive analysis**	**Specificity**	
	Sensitivity		**Specificity**	
D-shaped left ventricle (n=17)	88.24% [0.63 0.98]		95.20% [0.93 0.97]	
D-shaped left ventricle Basal septal hypertrophy (n=4)	91.55% [0.85 0.93]		93.75% [0.92 0.95]	
Basal septal hypertrophy Asymmetric septal hypertrophy (n=4)	94.55% [0.91 0.96]		96.08% [0.94 0.97]	
Asymmetric septal hypertrophy Mitral valvular disease (n=63)	72.14% [0.66 0.77]		80.39% [0.77 0.82]	
Basal anteroseptal hypokinesia (n=26) Aortic valvular disease (n=78)	82.34% [0.80 0.85]		87.39% [0.86 0.88]	
[(Max (b) – Min(b)) / Min(b)]	[0.29 0.83]		[0.60 0.65]	
Aorta dilation (n=257) Prosthetic mitral valvular disease (n=28)	96.43% 78.99% [0.81 0.99]		90.03% 75.84% [0.87 0.92]	
[Max (d) ≥ 30 mm]	[0.73 0.84]		[0.73 0.78]	
Prosthetic aortic valvular disease (n=15) prosthetic mitral valve dysfunction	93.33% 97.65% [0.96 0.98]		91.45% 95.27% [0.93 0.97]	
Mitral valve prolapse (n=27) RV Enlargement (n=158)	81.48% [0.62 0.94] 95.07%		96.68% [0.95 0.97] 87.85%	
[Max (f) ≥ 30 mm]	[0.88 0.96]		[0.86 0.90]	
Septal bouncing (n=36)	86.11% [0.70 0.95]		93.09% [0.91 0.95]	
Percardial effusion (n=56)	89.29% [0.78 0.96]		94.67% [0.92 0.96]	

3.2.1. Grad-CAM Result

For the quality assessment of the CDSS in the heart disease prediction, Grad-CAM was used to unfold the activation of the network. The brighter the heat map is, the higher the evidence that the network referred to that area. The reason for applying the Grad-CAM is that it is not only important for Resnet to perform well, but it is also important for the user to know the abnormality indicated by the network.

Since the object on the video is moving compared to the object in the still image, still images of the end-systolic and end-diastolic timing are based on the first quantitative network used. In addition to the good performance of the network, Resnet should make the same judgment as the clinician. Even if the network correctly classified the images as having pathologic findings, the cardiologist would not trust the CDSS if the network indicates wrong places on the images as pathologic findings.

As seen in Figure 3.5, selecting images based on the ES/ED timing not only improved the performance of Resnet networks but also showed higher accuracy in indicating the pathology.

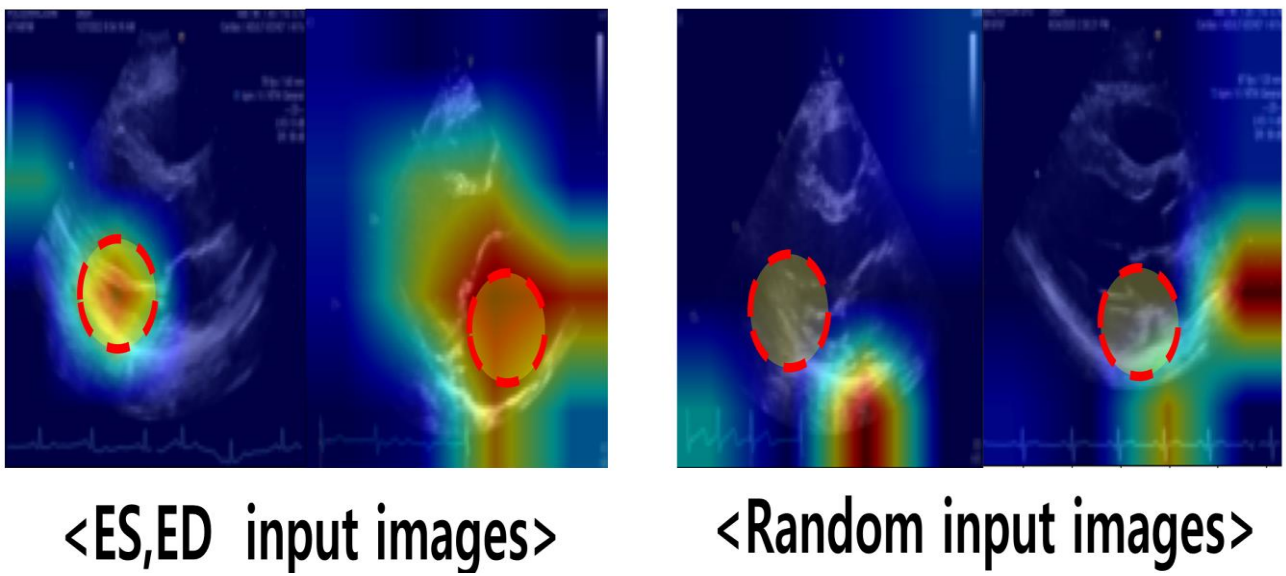


Figure 3.7. The comparison of Grad-CAM between a network trained with 10 images based on end-systolic and end-diastolic timing and the network trained with randomly selected 10 images.

3.3. Limitation

3.3.1. Need for an external dataset for generalizable network

The data collected for training is obtained from Seoul National University Hospital, where the sonographers and cardiologists are highly skilled and high-quality images are obtained. Similar to problems raised in previous papers, for more generalizable networks, data not only from these large hospitals but also from small local hospitals should be used for training.

In addition, an external validation set should be used to test the network, but verification using the data set obtained only from Seoul National University Hospital could also be a limitation.

3.3.2. Future work of the system

In this study, CDSS, a system capable of both quantitative and qualitative analysis was presented and verified to be valid.

This CDSS can only be applied in PLAX. The complete CDSS for other standard views should be incorporated. However, the PLAX image is the most informative compared to other standard views. Moreover, the number of structures in the PLAX view is higher than that in other standard views. Therefore, segmentation would be tougher in PLAX compared to other views, and performance in this system would not degrade when other standard view images are trained.

A heatmap by the Grad-CAM was presented in the still frame. For the real-time appreciation of the pathologic finding by the clinicians, this result should be overlayed on the moving images.

As the heart moves at a faster rate, fast-moving objects were detected incorrectly in quantitative analysis. This system is suboptimal for the quantitative analysis of small subjects. Finer segmentation should be accomplished for the structure with small dimensions to be analyzed accurately.

This CDSS was not tested in the detection of multiple pathologies in a single image. The performance of detecting multiple pathologies should be tested in the future.

CHAPTER 4

Conclusion

4. Conclusion

To the best of our knowledge, this study is the first to develop and validate an AI-based CDSS system for echocardiography in quantitative and qualitative analysis. In quantitative analysis, both sensitivities and specificities were over 90% except pathologies related to interventricular septum, left ventricular, left ventricular posterior wall, and aorta. Also end-systolic and end-diastolic frames can be detected correctly in the quantitative analysis. The qualitative analysis system showed 90% sensitivity and 83% specificity in detecting pathologic findings using end-systolic and end-diastolic input images.

The results suggest that both the interventricular septum and the left ventricular posterior wall change rapidly in shape in the entire image, sometimes the network has difficulty following. However, all of them are improved in segmentation results and diagnostic rules than previous studies, and it can be said as the first attempt to detect most of the diseases related to the parasternal long-axis view at once.

In future, to improve the data generation part of this study, the segmentation of the structure's small dimensions such as the interventricular septum, left ventricular posterior wall, and aorta should be performed finer, having more discrete boundaries, than performed in this study. For this, it is necessary to establish a large database well segmented by a group of experts. CDSS for other standard views must be incorporated to become a complete CDSS for echocardiography.

Abstract in Korean

국문 초록

심초음파 검사는 심장병 진단에 사용되는 중요한 도구이며, 수축기 및 이완기 단계의 심장 이미지를 제공한다. 심초음파 검사를 통해 심방과 심실의 다양한 구조적 이상과, 판막 이상등의 질환을 정량적으로 또는 정성적으로 진단할 수 있다. 심초음파 검사는 비침습적인 특성으로 인하여에 심장 전문의들이 많이 사용하고 있으며, 심장 질환자가 점점 많아지는 추세에 따라 더 많이 사용될 것으로 기대되고 있다.

심초음파 검사는 이러한 안전성과 유용성에도 불구하고, CT 나 MRI 와는 달리 1)정확한 영상을 얻는데 오랜 훈련기간이 필요하고 2) 영상을 얻을 수 있는 부위와 얻을 수 잇는 단면영상이 제한적이어서 검사 시 놓친 소견은 추후 영상을 감수할 경우에도 발견할 수 없는 특징을 가지고 있다. 이에 다라 측정과 해석의 정량화와 함께 검사상 이상소견을 놓치지 않을 수 있는 보완조치에 대한 요구가 많았고, 이러한 요구에 부응하여 심장전문의를 위한 임상 의사결정 지원 시스템에 대한 많은 연구가 진행되고 있다..

인공지능의 발달로 인해 어느정도 이러한 요구에 부응할 수 있게 되었다. 이 연구의 흐름은 두가지로 나뉘게 되는데, 첫째는 심장의 구조물들을 분할하여 크기를 측정하고 특이치를 감지하는 정량적인 연구방법과, 병변이 어느 부위에 있는지 이미지 내에서 확인하는 정성적 접근법으로 나뉜다. 기존에는 이 두 연구가 대부분 따로 진행되어 왔으나, 임상의사의 진단 흐름을 고려해 볼 때 이 두가지 모두가 포함되는 임상 의사 결정 지원 시스템의 개발이 필요한 현실이다.

이러한 관점에서 본 학위 논문의 목표는 대규모 코호트 후향적 연구를 통해 AI 기반의 심장 초음파 임상 의사결정 지원 시스템을

개발하고 검증하는 것이다. 데이터는 2016 년에서 2021 년도 사이에 서울대 병원에서 시행된 2600 예의 심초음파검사 영상(정상소견 1300 명, 병적소견 1300 명)를 이용하였다. 정량적분석과 정성적 분석을 모두 고려하기 위해 두개의 네트워크가 개발되었으며, 그 유효성은 환자 데이터로 검증되었다.

먼저 정량적 분석을 위한 이미지 분할을 위해 U-net 기반 딥러닝 네트워크가 개발되었으며, 개발에 필요한 데이터를 위해 심장전문의가 좌심실, 좌심방, 대동맥, 우심실, 좌심실 후벽 및 심실간 중격의 정보를 이미지에 표시를 하였다. 훈련된 네트워크로부터 나온 이미지로부터 6 개의 구조물의 직경과 면적을 구하여 벡터화 하였으며, 수축기말 및 이완기말 단계의 프레임 정보를 벡터로부터 추출하였다.

둘째로 정성적 진단을 위한 네트워크 개발을 위해 Resnet152 기반의 CNN 을 사용하였다. 이 네트워크의 입력데이터는 정량적 네트워크에서 추출된 수축기말 및 이완기말 정보를 기반으로 10 프레임이 추출되었다. 입력데이터가 정상인지 아닌지 구분하도록 했을 뿐 아니라, 마지막 레이어에서 그라디언트 가중 클래스 활성화 매핑(Grad-CAM)방법론을 이용하여 네트워크가 이미지상의 어느 부위를 보고 이상소견으로 분류했는지 시각화 하였다.

그 결과 먼저 정량적 네트워크 성능을 측정하기 위해 환자 1300 명의 데이터를 통해 각 구조물의 직경과 관련된 심장질환이 얼마나 잘 검출됐는지 확인하였다. 심실중격, 좌심실 후벽, 대동맥과 관련된 병적소견을 제외하고 다른구조물의 민감도와 특이성은 모두 90% 이상이다. 수축기 말기 및 확장기 말기 위상 검출도 정확했는데, 심장전문의에 의해 선택된 프레임에 비하여 수축기 말기의 경우 평균 0.52 프레임, 확장기 말기의 경우 0.9 프레임의 차이를 보였다.

정성분석을 위한 네트워크의 경우, 첫 번째 네트워크로부터 선택된 위상정보를 바탕으로 10 개의 입력데이터를 결정하였고, 무작위로 선택된 10 개의 결과를 비교하였다. 그 결과 정확도가 각각 90.33%, 81.16%로 나타났으며, 1 차 정량적 네트워크 에서 추출된 수축기말, 이완기말 프레임 정보는 환자를 판별하는 네트워크의 성능 향상에 기여했음을 알 수 있다. 또한 Grad-CAM 결과는 첫 번째 네트워크의 프레임 정보를 기반으로 데이터에서 추출된 10 장의 이미지가 입력데이터로 쓰였을 때가 무작위로 추출된 10 장의 이미지로 훈련된 네트워크 보다 병변의 위치를 더 정확하게 표시하는 것을 확인하였다.

결론적으로 본 연구는 정량적, 정성적 분석을 위한 AI 기반 심장 초음파 임상의사 결정 지원 시스템을 개발하였으며, 이 시스템이 실현 가능한 것으로 검증되었다.

주요어: 심방세동심초음파, 인공지능, 임상 의사 결정 지원 시스템, U-net, 딥러닝, 생성적 적대 신경망, 영상 분할

.

주요어: 심방세동심초음파, 인공지능, 임상 의사 결정 지원 시스템, U-net, 딥러닝, 생성적 적대 신경망, 영상 분할

학 번: 2018-32128

BIBLIOGRAPHY

[1] N. Boon *et al.*, "National variations in the provision of cardiac services in the United Kingdom: second report of the British Cardiac Society Working Group, 2005," *Heart,* vol. 92, no. 7, pp. 873-8, Jul 2006, doi: 10.1136/hrt.2006.096255.

[2] V. Mor-Avi *et al.*, "Real-time 3D echocardiographic quantification of left atrial volume: multicenter study for validation with CMR," *JACC Cardiovasc Imaging,* vol. 5, no. 8, pp. 769-77, Aug 2012, doi: 10.1016/j.jcmg.2012.05.011.

[3] L. D. Jacobs *et al.*, "Rapid online quantification of left ventricular volume from real-time three-dimensional echocardiographic data," *Eur Heart J,* vol. 27, no. 4, pp. 460-8, Feb 2006, doi: 10.1093/eurheartj/ehi666.

[4] R. Hoffmann *et al.*, "Analysis of left ventricular volumes and function: a multicenter comparison of cardiac magnetic resonance imaging, cine ventriculography, and unenhanced and contrast-enhanced two-dimensional and three-dimensional echocardiography," *J Am Soc Echocardiogr,* vol. 27, no. 3, pp. 292-301, Mar 2014, doi: 10.1016/j.echo.2013.12.005.

[5] F. C. Cobey, V. Patel, A. Gosling, and E. Ursprung, "The emperor has no clothes: recognizing the limits of current echocardiographic technology in perioperative quantification of mitral regurgitation," *Journal of Cardiothoracic and Vascular Anesthesia,* vol. 31, no. 5, pp. 1692-1694, 2017.

[6] K. Kusunose, "Steps to use artificial intelligence in echocardiography," (in English), *J Echocardiogr,* vol. 19, no. 1, pp. 21-27, Mar 2021, doi: 10.1007/s12574-020-00496-4.

[7] A. Narang *et al.*, "Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use," (in English), *Jama Cardiol,* vol. 6, no. 6, pp. 624-632, Jun 2021, doi: 10.1001/jamacardio.2021.0185.

[8] J. G. Lee *et al.*, "Deep Learning in Medical Imaging: General Overview," *Korean J Radiol,* vol. 18, no. 4, pp. 570-584, Jul-Aug 2017, doi: 10.3348/kjr.2017.18.4.570.

[9] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography," *J Am Coll Cardiol,* vol. 68, no. 21, pp. 2287-2295, Nov 29 2016, doi: 10.1016/j.jacc.2016.08.062.

[10] J. Jeganathan *et al.*, "Artificial intelligence in mitral valve analysis," *Ann Card Anaesth,* vol. 20, no. 2, pp. 129-134, Apr-Jun 2017, doi: 10.4103/aca.ACA_243_16.

[11] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," (in English), *Npj Digit Med,* vol. 1, Mar 21, 2018, doi: ARTN 610.1038/s41746-017-0013-1.

[12] S. Leclerc *et al.*, "Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography," (in English), *Ieee T Med Imaging,* vol. 38, no. 9, pp. 2198-2210, Sep 2019, doi: 10.1109/Tmi.2019.2900516.

[13] N. Azarmehr, X. Ye, F. Janan, J. P. Howard, D. P. Francis, and M. Zolgharni, "Automated Segmentation of Left Ventricle in 2D echocardiography using deep learning," *arXiv preprint arXiv:2003.07628,* 2020.

[14] A. Ghorbani *et al.*, "Deep learning interpretation of echocardiograms," *Npj Digit Med,* vol. 3, p. 10, 2020, doi: 10.1038/s41746-019-0216-8.

[15] J. Zhang *et al.*, "Fully Automated Echocardiogram Interpretation in Clinical Practice," *Circulation,* vol. 138, no. 16, pp. 1623-1635, Oct 16 2018, doi: 10.1161/CIRCULATIONAHA.118.034338.

[16] F. Taheri Dezaki *et al.*, "Cardiac Phase Detection in Echocardiograms With Densely Gated Recurrent Neural Networks and Global Extrema Loss," *IEEE Trans Med Imaging,* vol. 38, no. 8, pp. 1821-1832, Aug 2019, doi: 10.1109/TMI.2018.2888807.

[17] S. Katsushika *et al.*, "Deep Learning Algorithm to Detect Cardiac Sarcoidosis From Echocardiographic Movies," *Circ J,* vol. 86, no. 1, pp. 87-95, Dec 24 2021, doi: 10.1253/circj.CJ-21-0265.

[18] M. Vafaeezadeh, H. Behnam, A. Hosseinsabet, and P. Gifani, "Automatic morphological classification of mitral valve diseases in echocardiographic images based on explainable deep learning methods," *Int J Comput Assist Radiol Surg,* vol. 17, no. 2, pp. 413-425, Feb 2022, doi: 10.1007/s11548-021-02542-7.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," (in English), *P Ieee,* vol. 86, no. 11, pp. 2278-2324, Nov 1998, doi: Doi 10.1109/5.726791.

[20] E. B. Mazomenos, K. Bansal, B. Martin, A. Smith, S. Wright, and D. Stoyanov, "Automated performance assessment in transoesophageal echocardiography with convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018: Springer, pp. 256-264.

[21] F. M. Asch *et al.*, "Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert," (in English), *European Heart Journal,* vol. 40, pp. 2583-2583, Oct 2019. [Online]. Available: <Go to ISI>://WOS:000507313002332.

[22] J. Tromp *et al.*, "Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study," *Lancet Digit Health,* vol. 4, no. 1, pp. e46-e54, Jan 2022, doi: 10.1016/S2589-7500(21)00235-1.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," (in English), *Lect Notes Comput Sc,* vol. 9351, pp. 234-241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[24] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016: Springer, pp. 424-432.

[25] S. Kim *et al.*, "Fully automated quantification of cardiac chamber and function assessment in 2-D echocardiography: clinical feasibility of deep learning-based algorithms," *The International Journal of Cardiovascular Imaging,* vol. 38, no. 5, pp. 1047-1059, 2022.

[26] E. Smistad, I. M. Salte, A. Østvik, S. Leclerc, O. Bernard, and L. Lovstakken, "Segmentation of apical long axis, four-and two-chamber views using deep neural networks," in *2019 IEEE International Ultrasonics Symposium (IUS)*, 2019: IEEE, pp. 8-11.

[27] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," (in English), *Proc Cvpr Ieee,* pp. 770-778, 2016, doi: 10.1109/Cvpr.2016.90.

[28] Y. H. Guo, G. Q. Du, W. Q. Shen, C. L. Du, P. N. He, and S. Siuly, "Automatic myocardial infarction detection in contrast echocardiography based on polar residual network," (in English), *Comput Meth Prog Bio,* vol. 198, Jan 2021, doi: ARTN 10579110.1016/j.cmpb.2020.105791.

[29] F. T. Dezaki *et al.*, "Deep Residual Recurrent Neural Networks for Characterisation of Cardiac Cycle Phase from Echocardiograms," (in English), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support,* vol. 10553, pp. 100-108, 2017, doi: 10.1007/978-3-319-67558-9_12.

[30] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921-2929.

[31] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.

[32] K. Giannakaki *et al.*, "Automatic aortic valve area detection in echocardiography images using convolutional neural networks and U-net architecture for bicuspid aortic valve recognition," in *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2021: IEEE, pp. 1-6.

[33] R. Nova, S. Nurmaini, R. U. Partan, and S. T. Putra, "Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study," *Informatics in Medicine Unlocked,* vol. 24, p. 100601, 2021.

[34] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology,* vol. 26, no. 3, pp. 297-302, 1945.

[35] M. S. Huang, C. S. Wang, J. H. Chiang, P. Y. Liu, and W. C. Tsai, "Automated Recognition of Regional Wall Motion Abnormalities

Through Deep Neural Network Interpretation of Transthoracic Echocardiography," (in English), *Circulation,* vol. 142, no. 16, pp. 1510-1520, Oct 20 2020, doi: 10.1161/Circulationaha.120.047530.

[36] E. S. Lane *et al.*, "Multibeat echocardiographic phase detection using deep neural networks," (in English), *Comput Biol Med,* vol. 133, Jun 2021, doi: ARTN 10437310.1016/j.compbiomed.2021.104373.