

# Deep learning supported echocardiogram analysis: A comprehensive review

Sanjeevi G.<sup>a</sup>, Uma Gopalakrishnan<sup>a,\*</sup>, Rahul Krishnan Parthinarupothi<sup>a</sup>, Thushara Madathil<sup>b</sup>

<sup>a</sup> Center for Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri, India

<sup>b</sup> Department of Cardiac Anesthesiology, Amrita Institute of Medical Sciences and Research Center, Kochi, India

## ARTICLE INFO

### Keywords:

Echocardiogram  
Deep learning  
Decision support system

## ABSTRACT

An echocardiogram is a sophisticated ultrasound imaging technique employed to diagnose heart conditions. The transthoracic echocardiogram, one of the most prevalent types, is instrumental in evaluating significant cardiac diseases. However, interpreting its results heavily relies on the clinician's expertise. In this context, artificial intelligence has emerged as a vital tool for helping clinicians. This study critically analyzes key state-of-the-art research that uses deep learning techniques to automate transthoracic echocardiogram analysis and support clinical judgments. We have systematically organized and categorized articles that proffer solutions for view classification, enhancement of image quality and dataset, segmentation and identification of cardiac structures, detection of cardiac function abnormalities, and quantification of cardiac functions. We compared the performance of various deep learning approaches within each category, identifying the most promising methods. Additionally, we highlight limitations in current research and explore promising avenues for future exploration. These include addressing generalizability issues, incorporating novel AI approaches, and tackling the analysis of rare cardiac diseases.

## 1. Introduction

The World Health Organization (WHO) estimates that cardiovascular disorders are increasingly becoming the leading cause of death in industrialized nations. Unfortunately, 40% of heart attack deaths are reported among people under the age of 55 in the developing world [1, 2]. In response to this global health crisis, researchers worldwide are exploring innovative solutions, with many teams harnessing the power of artificial intelligence (AI) to predict and detect cardiovascular diseases on time [3–6].

Cardiac imaging is pivotal in evaluating heart function and guiding clinicians in determining the most effective treatment strategies. The most prevalent cardiac imaging techniques include computed tomography (CT), magnetic resonance imaging (MRI), and echocardiography [7,8]. Echocardiography, an ultrasound-based heart imaging, offers several advantages over CT and MRI. It is cost-effective, suitable for bedside use, and allows for real-time interpretation by radiologists. By manipulating the angle of the transducer probe, echocardiography can visualize the heart muscle and its movements from various perspectives. This enables the assessment of cardiac wall and muscle dimensions, volumes, and movements, which can help identify abnormalities. Portable echocardiography devices enhance patient convenience by enabling bedside heart function assessments [9,10]. Another significant advantage of echocardiography scanners is their ability to provide real-time images of the heart. Various types of echocardiograms,

such as transthoracic echocardiograms (non-invasive), transesophageal echocardiograms (invasive), stress echocardiograms (exercise-based), and Doppler echocardiograms (blood flow assessment), are recommended based on clinical requirements.

The Transthoracic Echocardiogram (TTE) is the most prevalent type of echocardiogram. It is non-invasive and conducted entirely outside the patient's body. TTE is instrumental in assessing and diagnosing various cardiac disorders, including coronary artery disease, myocardial infarction (MI), cardiomyopathy, amyloidosis, and heart chamber hypertension. Other echocardiograms, such as transesophageal and Doppler echocardiograms, detect blood clotting and analyze blood flow. The stress echocardiogram evaluates cardiac function pre and post-exercise. This comprehensive review primarily focuses on TTE analysis, the most widely employed modality, and technique for diagnosing a broad spectrum of heart diseases.

The clinician's expertise plays a crucial role in echocardiogram analysis. Clinicians are required to manually select end-systole and end-diastole frames and then measure the cardiac parameters of the appropriate region. However, this manual method has several drawbacks. Firstly, it exhibits high inter- and intra-reader variability. Secondly, the manual distinction is a time-consuming process that increases the workload, potentially leading to distraction and incorrect or delayed diagnosis. Lastly, cardiological expertise, a critical resource, is often inaccessible in resource-limited settings [9].

\* Corresponding author.

E-mail address: [umag@am.amrita.edu](mailto:umag@am.amrita.edu) (U. Gopalakrishnan).

<https://doi.org/10.1016/j.artmed.2024.102866>

Received 17 June 2023; Received in revised form 20 March 2024; Accepted 30 March 2024

Available online 4 April 2024

0933-3657/© 2024 Elsevier B.V. All rights reserved.

Deep Learning (DL) [11] has emerged as an essential tool for clinicians in this context. DL models have shown promising outcomes in recognizing, measuring, and analyzing echocardiograms. These models could streamline clinical decision-making by providing interactive feedback to guide clinicians with less experience [1,2,7]. Most earlier reviews have broadened their scope when surveying automated cardiac disease diagnosis. Some reviews focused on diagnosing diseases using various cardiac imaging techniques, including cardiac CT, MRI, and several echocardiographic methods. Other reviews have explored the extensive domain of artificial intelligence, encompassing traditional machine learning and deep learning algorithms. Furthermore, there has been significant progress in AI techniques since the most recent review of echocardiogram-based diagnosis was published. As a result, there is a compelling need to provide researchers with the necessary knowledge and understanding in this field to facilitate more rapid translational research. This paper specifically focuses on a review of various DL applications in the automated analysis of transthoracic echocardiograms, which will be a valuable resource for researchers in this fast-changing area of research.

Prior to this review, De Siqueira et al. [12,13] covered the application of AI, including DL and machine-learning algorithms in the automated analysis of the different types of echocardiograms such as TTE, transoesophageal, doppler, stress, and fetus echocardiogram. Litjens et al. [14] reviewed DL applications in cardiac imaging modalities, including echocardiogram, cardiac CT and MRI, optical coherence tomography (OCT), and single-photon emission computerized tomography (SPECT). Bizopoulos et al. [10] covered DL applications in cardiology, including decision-making using patient clinical data, cardiac signals such as phonocardiogram (PCG), electrocardiogram (ECG), and cardiac imaging modalities such as cardiac CT and MRI, fundus photography, echocardiogram, and OCT. Zamzmi et al. [9] reviewed manual and computerized methods to analyze the different echocardiogram modes such as B-mode, M-mode, and Doppler mode. Karatzia et al. [15] also covered the broad application of AI in cardiology. We have conducted an exhaustive review of primary studies on DL algorithms employed in the automated processing of TTE to evaluate the advancements and challenges in DL-driven clinical decision-making.

We scoured major scientific databases to select studies on DL techniques applied to TTE. Each study was assessed and summarized, and primary studies with similar objectives were grouped. This process led us to discover state-of-the-art DL approaches. We identified the limitations of current methods and the challenges that future research may encounter when using DL in this context. This comprehensive review includes over 97 papers that cover a broad spectrum of DL applications for TTE analysis.

Our comprehensive analysis reveals scope for further research in the following areas:

1. Open research challenges for rare cardiac diseases:
  - Most studies have used DL to diagnose common cardiac diseases such as motion abnormality detection, hypertrophy, and hypertension. There is a need for further exploration of DL techniques to diagnose less common cardiac conditions such as peripheral, pericardial, and congenital heart diseases.
  - Compiling extensive datasets for rare cardiac diseases poses a challenge. Researchers could consider different learning algorithms, such as unsupervised, deep reinforcement, and few-shot learning [16], to enhance model performance where data collection is costly and challenging.
  - The advent of novel, portable echocardiogram devices at the edge level has spurred active research into developing lightweight decision support systems.
2. Model interpretability: Clinicians need explanations of the clinical decision-making process of the DL model to understand its workings better. A reliable and transparent decision-making process, facilitated by an explainability module, is essential.

3. Generalizable model across demographics: DL algorithms should be trained and validated on multicenter datasets encompassing various demographics and different types of ultrasound equipment.
4. Image quality enhancement: Most studies improved image quality by removing speckle noise and shadowing artifacts [17–19]. Researchers may explore the possibilities of DL solutions to solve artifacts occurring in echocardiogram images in different situations, such as improving the temporal resolution of TTE videos for high heart rate patients, improving the spatial resolution of patients with thick chest walls and lung diseases, and removing shadowing artifacts on patients with calcium deposits and metallic valves.

The contributions of this paper are

- A detailed overview of the metrics employed to assess the efficacy of diverse DL models in the automated analysis of TTE is provided in (Section 3.1).
- An exhaustive analysis of cutting-edge automated DL models is presented, categorized as follows:
  1. TTE view classification (Section 3.2)
  2. Enhancement of image quality (Section 3.3)
  3. Segmentation or identification of cardiac structures (Section 3.4)
  4. Quantification of cardiac functions (Section 3.5)
  5. Diagnosis of cardiac diseases (Section 3.6)
- An overview of various publicly accessible TTE datasets (Section 4).
- An outline of potential future research directions in automated TTE analysis using DL (Section 5).

The structure of the remaining sections of this article is as follows: The subsequent section provides the methodology employed for this review, encompassing the search strategy and the article selection process. In Section 3, we categorize and delve into various applications of DL in TTE analysis. Section 4 furnishes details about the diverse publicly available TTE datasets. Section 5 provides insights into the limitations and research challenges and also proposes potential directions for future research in the realm of DL applied to TTE analysis. Finally, Section 6 encapsulates the findings and concludes the article.

## 2. Materials and methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [20] to conduct our study and present our search and selection methods to reproduce this review. PRISMA guidelines were created to assist researchers in enhancing the reporting of meta-analyses and systematic reviews. Fig. 1 illustrates the summary of the article selection process.

### 2.1. Search strategy

The keywords and the databases are determined to search for scientific papers in Scopus. The terms “Echocardiogram” OR “Echocardiography” were initially used to search for relevant and important articles. “Deep Learning” was added to search filters to refine the search terms further. Additionally, specific application areas were added to narrow down the relevant articles. The entire search string was as follows: (i). (“Echocardiogram” OR “Echocardiography”) AND (“Deep Learning”) AND (“View Classification”), (ii) (“Echocardiogram” OR “Echocardiography”) AND (“Deep Learning”) AND (“Cardiac Segmentation”), (iii). (“Echocardiogram” OR “Echocardiography”) AND (“Deep Learning”) AND (“Cardiac Disease”), and (iv). (“Echocardiogram” OR “Echocardiography”) AND (“Deep Learning”) AND (“Quantification”). The literature papers were then taken from the following databases: Institute

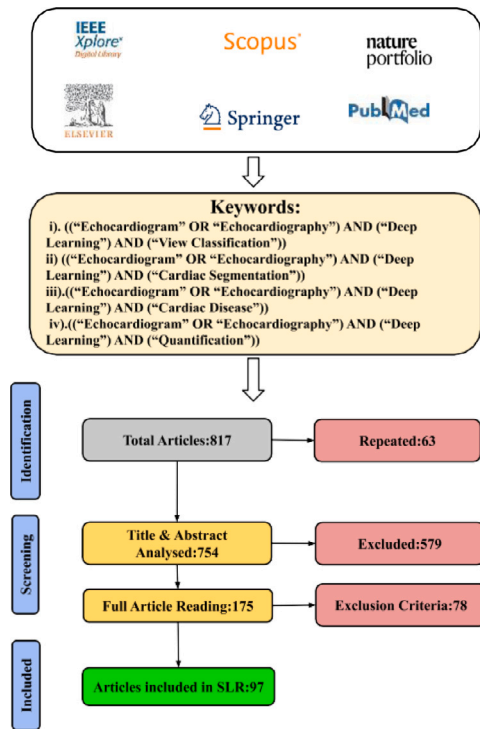


Fig. 1. Flow diagram of article inclusion based on PRISMA. (This search was conducted in January 2024.)

of Electrical and Electronics Engineers (IEEE) Xplore, Nature, PubMed, Springer, and ScienceDirect. This comprehensive review aims to understand and present the evolution of DL networks in automated TTE analysis and decision-making. The search was limited to English-language papers published between 2017 and December 2023.

## 2.2. Selection strategy

The following inclusion (I) and exclusion (E) criteria were used to choose the articles:

- Articles that analyzed TTE using deep learning algorithms (I1);
- Complete, English-written articles (I2);
- Articles presenting original research (I3);
- Articles without the specific TTE view (E1);
- Articles that were either incomplete or did not specify the deep learning technique employed (E2);
- Articles that failed to report the experiment’s findings (E3).

A total of 817 papers were found after searches in the scientific databases. The 63 duplicates were removed from the analysis. The 579 papers were disregarded after reading the abstracts and titles of 754 articles. Hence, 175 articles were chosen for extensive reading. After reading all the articles, 78 were excluded because they met the exclusion standards (E1, E2, and E3). Finally, the comprehensive review contained 97 articles.

## 3. Application of deep learning in TTE analysis

In a hospital setting, clinical decision-making using TTE involves several steps, including acquiring images from an appropriate view based on the patient’s condition and clinician’s suggestion, analyzing specific cardiac structures to determine cardiac function or calculating parameters like End-Systolic or End-Diastolic volumes, and ejection fraction (EF) using computations. Subsequently, disease diagnosis and

treatment planning are carried out. This workflow is complex, time-intensive, and requires significant manual intervention. The slight differences between various TTE view images make it challenging to identify a specific view. Ensuring the images are free from noise is vital for clear interpretation. Accurately identifying cardiac structures and interpreting cardiac function or disease requires years of experience. This manual process can be burdensome, potentially leading to delayed diagnoses.

Automating repetitive tasks in clinical decision-making can significantly reduce clinicians’ workload and enhance patient outcomes. DL has been employed in the healthcare sector in recent years to aid medical decision-making through various analyses. Several DL frameworks or pipelines have been proposed to automate different aspects of TTE analysis (Fig. 2). Table 1 summarizes DL algorithms and applications in TTE analysis. This section first elaborates on the essential evaluation metrics used for assessing the effectiveness of DL applications and subsequently discusses the different categories of automated cardiac analysis tasks. These evaluation metrics are a comparative measure of the performance of diverse methodologies employed in TTE analysis. The organization of this section is as follows:

- Evaluation metrics;
- TTE view classifications or view identification;
- TTE image quality improvement and dataset augmentation to address data scarcity;
- Cardiac structure segmentation or identification from TTE;
- Cardiac function quantification from TTE; Cardiac disease diagnosis, which is further divided into:

- Motion abnormality detection
- Hypertrophy or heart wall thickness detection
- Cardiac valve disease diagnosis
- Other cardiac disease detection.

### 3.1. Evaluation metrics

The various metrics used to assess the effectiveness of tasks of TTE image or video with DL are summarized in this section. These measurements are categorized into classification assessment metrics, segmentation assessment metrics, and quantification assessment metrics. Table 2 summarizes evaluation metrics.

#### 3.1.1. Classification metrics

The primary metric used to evaluate classification tasks is accuracy, representing the proportion of correct predictions to total predictions [9]. Higher accuracy indicates better model performance. Additionally, studies often assess performance using the area under receiver operating curve (AUC) and Cohen’s kappa coefficient (CKC) metrics, which provide similar insights as accuracy. AUC evaluates a model’s ability to distinguish between classes, with higher values indicating better discrimination. Fig. 3 illustrates the AUC curve. CKC measures reliability in categorical classification, considering agreement probability between raters. It is computed between model predictions and ground truth in view classification tasks [21].

#### 3.1.2. Image quality metrics

The various metrics used to evaluate noise reduction and image augmentation techniques are mean square error (MSE), signal-to-noise ratio (SNR), Peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and contrast-to-noise ratio (CNR) [17]. CNR measures contrast improvement, SNR measures signal strength relative to uncertainty, and PSNR indicates image quality by comparing signal power to noise power. Higher SNR, CNR, and PSNR values suggest adequate filtering or compression. SSIM quantifies similarity between images, with 1 indicating high similarity and 0 indicating complete dissimilarity. Cross-entropy loss, ranging from 0 to 1, is also used to assess the model’s noise reduction capability, with 0 representing optimal performance [23].

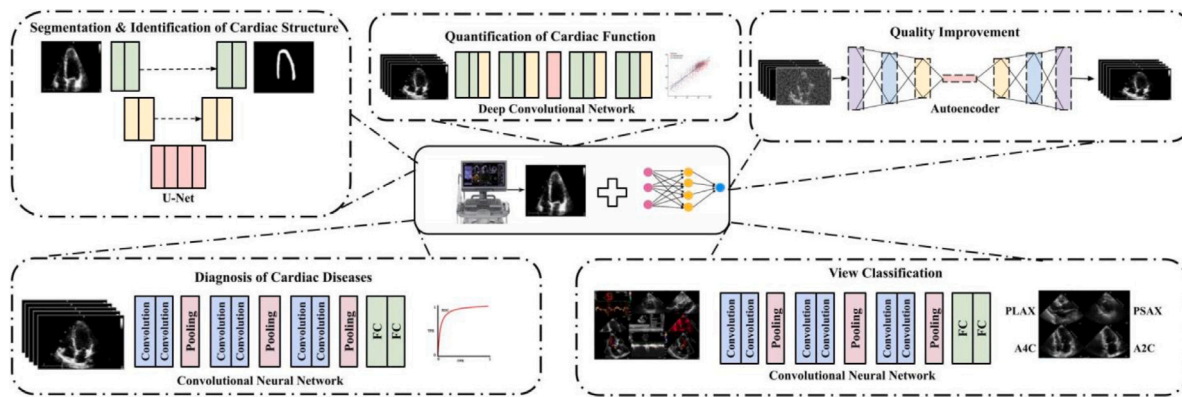


Fig. 2. An overview of the automated TTE analysis process, such as segmentation or identification of cardiac structure, quantification of cardiac function, quality improvement, diagnosis of cardiac disease, and view classification.

Table 1  
Different deep learning algorithms used in TTE analysis.

Deep learning algorithm	Description	Application
Convolutional neural network (CNN)	Each node is linked to nodes in the layer below to execute a convolution operation. This substantially lowers the number of parameters in the network and enables these networks to recognize features irrespectively of where they are located in an image. Some fully connected layers are typically added for prediction at the network's end.	View classification, disease detection, and quality prediction.
Fully convolutional neural network (FCN)	Provide entire images, such as segmentation masks, as output despite not having fully connected layers. U-Net is the most often used fully convolutional network in TTE images.	Cardiac structure segmentation, image quality improvement.
Recurrent neural network (RNN)	Best choice for sequence data because they feed their output as input, allowing them to incorporate information from earlier points in the series. Convolutional layers can be combined with RNNs with a flexible structure.	Prediction of ejection fraction (EF). Segmentation of cardiac structure from image sequence
Generative adversarial network (GAN)	A network comprises discriminators and generators, often CNNs. The discriminator determines if the image is generated or real one. The generator learns how to produce realistic images by simultaneously optimizing generated and real images.	Most applications are for image augmentation and generation.

### 3.1.3. Cardiac structure segmentation or identification metrics

The most common metrics for evaluating the segmentation results are the dice similarity index (DSI), Hausdorff distance (HD), and intersection over union (IOU). DSI assesses model performance by measuring pixel-level agreement between model output and ground truth, with a value of 1 indicating a perfect match. HD quantifies the distance between points in model and annotation sets, with lower values indicating better segmentation. IOU measures overlap between segmented and ground truth masks, with 1 indicating perfect overlap and 0 indicating no overlap [9,24].

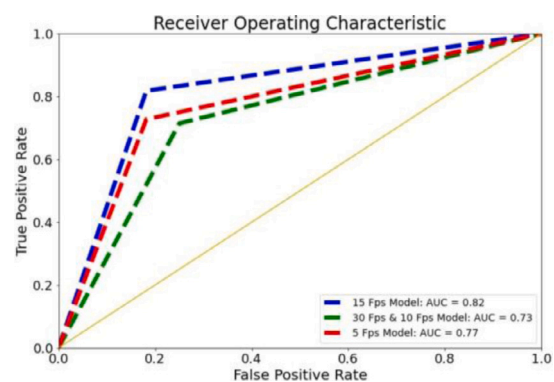


Fig. 3. Illustration of AUC curve.  
Source: Reproduced from [22].

### 3.1.4. Quantification metrics

The standard statistical measures used to assess the quantification of cardiac function are correlation, Bland-Altman agreement (B& A) coefficient of determination ( $R^2$ ), mean bias error (MBE), and mean absolute deviation (MAD). Correlation indicates the relationship between variables, ranging from  $-1$  to  $+1$ . The complete relationship between two variables is represented by either  $+1$  or  $-1$ . Using the mean difference and bounds of agreement, B& A calculates the degree of agreement between two sets of measurements or data.  $R^2$  quantifies how well a model predicts results, with 1 indicating perfect prediction. MBE measures the average bias between ground truth and model prediction, while MAD calculates the average deviation from the mean. Lower MBE and MAD values signify better DL model performance [9,25].

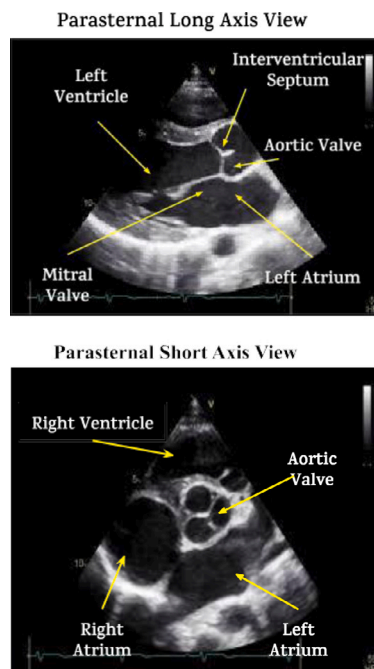
### 3.2. TTE view classification

The initial and crucial step in the analysis of TTE for further diagnosis is the identification of the TTE view. During a comprehensive TTE examination, the heart is imaged using various windows. These windows, which could be parasternal, apical, subcostal, or suprasternal, are determined by the position of the transducer. Furthermore, the transducer's position also dictates the passage of the tomographic plane through the heart, whether long, short, four, or five chambers. This section provides an overview of a common TTE view and the DL algorithms used for their classification. The parasternal window enables excellent imaging of the right ventricle, tricuspid valve, pulmonary valve, and right ventricular outflow tract, among other anterior cardiac structures. Additionally, it allows quick evaluation of the pericardium

**Table 2**  
Categories of evaluation metrics.

Category	Metrics name	Range
Classification metrics	Acc.	0–1 (↑)
	AUC	0–1 (↑)
	CKC	0–1 (↑)
Image quality metrics	MSE	(↓)
	SNR	(↑)
	PSNR	(↑)
	SSIM	0–1 (↑)
	CNR	(↑)
Segmentation or identification metric	DSI	0–1(↑)
	HD	0–1(↓)
	IOU	0–1(↑)
Quantification metrics	Corr.	–1 to 1*
	R <sup>2</sup>	0–1(↑)
	MBE	(↓)
	MAE	(↓)

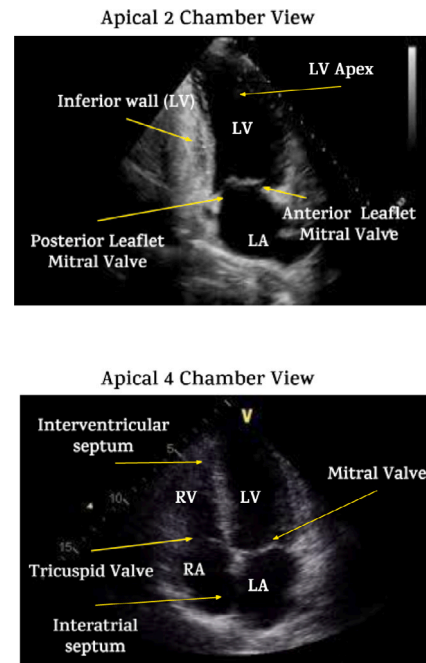
↑ - Higher is better; ↓ - Lower is better; \* The values near the extreme are good.



**Fig. 4.** Illustration of cardiac structure in parasternal long axis view and parasternal short axis view.

and the left ventricle (LV) in the short axis. A quick assessment of patients who are in unstable conditions frequently uses this window. The apical window simultaneously visualizes all four heart chambers, such as two atriums and ventricles, the mitral valve, and the tricuspid valve, enabling clinicians to analyze the essential cardiac structures [26].

The most commonly used echocardiogram views in clinical practice are the apical 4-chamber view (A2C), parasternal long axis view (PLAX), apical 2-chamber view (A4C), and parasternal short axis view (PSAX) [26]. Fig. 4 shows the PLAX, PSAX, and Fig. 5 shows A2C and A4C, respectively. Additionally, clinicians or radiologists actively focus on substructures within a TTE image by adjusting the transducer's zoom level and rotating the transducer, which results in numerous variations of these images. Identifying the TTE view is the essential first step in cardiac disease assessment. Various studies have tried to classify the views from echocardiogram images. This step is challenging since several points of view slightly diverge.



**Fig. 5.** Illustration of cardiac structure in apical-2-chamber view and apical-4-chamber view.

Early work in this area has tried to use basic artificial neural networks and support vector machine techniques. Balaji GN et al. [27] employed a backpropagation neural network (BPNN) and support vector machine algorithms for classifying various views. More recent studies in TTE view classification have explored the use of Convolutional Neural Networks (CNNs). A CNN is designed to process and extract high-level and low-level features from images directly [28].

Gao et al. [29] developed a 2D CNN model that utilizes temporal and spatial data from TTE videos. The model integrates two networks through the linear combination of class vectors generated from each network. It is capable of classifying eight types of TTE video views. In parallel, Madani et al. [30] proposed a 2D CNN classification model tested on fifteen TTE views. These views included PLAX, right ventricular inflow, short axis at mid or mitral level, basal short axis, A4C, A5C, A2C, A3C/long apical axis, subcostal inferior vena cava, subcostal four-chamber, subcostal/abdom, continuous-wave Doppler, pulsed-wave Doppler, and m-mode. The model achieved an accuracy of 91.7%. When an echocardiography specialist evaluated the same inputs, the accuracy ranged from 70.2% to 84.0%. Additionally, occlusion testing and saliency mapping of echocardiogram images were performed to enhance the interpretability of the DL model's decision-making process.

In contrast with previous studies, Kusunose et al. [31] used a CNN model with a unique approach to data handling for classifying five TTE image views. The model was trained and tested using two types of data. The first type consisted of ten equally spaced images from the cardiac cycle of the TTE, with predictions averaged over these ten images. The second type involved averaging ten images from a single patient for model training. The CNN model achieved an impressive view classification accuracy when ten images were input. Meanwhile, Howard et al. [32] evaluated the performance of various CNN models, including 2D CNN, time-distributed CNN, time-streamed CNN, and 3D CNN, in classifying echocardiogram views. The study incorporated 14 TTE views (A2C, A3C, A4C LV, A4C RV, A5C, PLAX, PLAX valves, PLAX TV, PS AV, PSAX LV, IAS, Apical, subcostal suprasternal). The two-stream networks utilized both echocardiogram and optical images and demonstrated superior accuracy, achieving a CKC of 0.957 and

an error rate of merely 3.9%. Notably, the time-distributed CNN and time-streamed CNN outperformed both the conventional 2D and 3D CNNs.

In addition to the vanilla CNN, some specialized CNNs were also explored for TTE view classification for improved performance on a broader set of TTE views. Gu et al. [33] introduced a compact architecture known as the Efficient-Evidential Network (Efficient-EvidNet), designed to classify echocardiogram views while providing a sampling-free uncertainty prediction. Evidential uncertainty aids in identifying outliers and filtering out incorrect data, enhancing the model's overall performance. Efficient-EvidNet can classify 13 TTE views, including A2C, PLAX, PSA, SUPR, A3C, RVIF, PSM, A4C, PSPM, A5C, SIVC, and PSAP. Remarkably, reporting only on data with low evidential uncertainty yields a test accuracy of 97.6%. Similarly, Gao et al. [34] introduced a DL framework comprising three processes for cardiac view recognition. Initially, a spatial transform network is employed to comprehend the variability in the heart's structure throughout a cardiac cycle, thereby reducing intra-class variability. Subsequently, a channel attention mechanism is developed to recalibrate channel-wise feature responses. Finally, graph-based image embedding converts structured signals based on similarities among cardiac views. This proposed method successfully classified nine TTE views (A2C, A3C, A4C, A5C, SB, SL, PSLA, SUB4C, SUPAO).

Further advancement was done by Huang et al. [35]. They tried a deep convolutional model for classifying 29 TTE views. This model, a modification of the capsule net [36], incorporates an autoencoder component to interpret the feature maps of the final layer for human comprehension. In a series of studies, Zhang et al. [37], Lin et al. [38], and Huang et al. [39] proposed an automated interpretation of TTE, with view classification serving as the initial step in their pipeline. The CNN models they developed were able to classify 23 viewpoints (Parasternal (Long axis — remote, Long axis, Long axis — zoom of left atrium, Long axis — centered over left atrium, RV inflow, Short axis at apex, Short axis at papillary muscle, Short axis at mitral valve, Short axis at aortic valve, Short axis at aortic valve — zoom), Apical (2-chamber — no occlusions, 2-chamber — occluded left atrium, 2-chamber — occluded left ventricle, 3-chamber — no occlusions, 3-chamber — occluded left atrium, 3-chamber — occluded left ventricle, 4-chamber — no occlusions, 4-chamber — occluded left atrium, 4-chamber — occluded left ventricle, 5-chamber), Subcostal, Superasternal) three views (A2C, A4C, Apical long axis), and four views (PLAX, PSAX, A4C, A2C) with an average accuracy of 86%, 96%, and 99%, respectively. Lau et al. [40] introduced a method for estimating LV parameters using DL, including view classification. The 3D model they developed achieved an AUC score of 0.97 in view classification.

The process of identifying views is pivotal in the automated analysis of TTE. The articles in this section introduce various methods for identifying viewpoints and addressing the complexity of the task. A DL framework that can automatically identify a wide range of views could prove invaluable for radiologists or medical trainees seeking to navigate complex viewpoints. Table 3 summarizes the performance of CNNs in view classification. Most studies [29–35] have shown promising results, demonstrating that CNNs can reliably perform view classification.

Key research findings:

1. State-of-the-art accuracies have been reported by Huang et al. [35] and Zhang et al. [37], who have also incorporated a wider range of TTE views, such as 29 and 23 views, respectively.
2. TTE view classification poses multiple challenges due to several factors:
  - Dataset scarcity: The number of publicly available TTE datasets with a broader spectrum of TTE views is limited, making it challenging to train a CNN that generalizes well to new data.

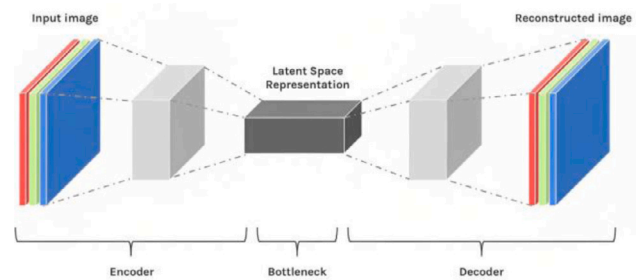


Fig. 6. Illustration of autoencoder.  
Source: Reproduced from [42].

- Viewpoint variability: The appearance of the heart and surrounding structures can significantly vary from one view to another due to factors such as patient positioning, transducer placement, and the patient's body habitus.
- Class imbalance: The distribution of views in a typical echocardiogram dataset is highly imbalanced. Some views, like the parasternal long-axis view, are much more common than others, such as the subcostal views. This imbalance needs to be corrected through various techniques to enable better generalizability of the models.

### 3.3. TTE image quality improvement

The TTE allows measuring heart parameters to detect abnormalities and understand how the heart chambers and valves pump blood. However, factors such as body fat, the position of the ultrasound probe, chest hair, and intervening tissues or bones can introduce noise into TTE images. This noise, including shadowing artifacts, Gaussian noise, and speckle noise, can sometimes lead to missed diagnoses. Compared to CT and MRI, echocardiograms have significant limitations, including a small field of view and low CNR and SNR [18]. Various TTE image quality improvement techniques have been proposed to overcome these limitations, with most noise reduction studies suggesting using autoencoder-based models. Fig. 6 illustrates the autoencoder model representation. This section also discusses applying DL-based data augmentation methods to solve data scarcity.

Variability in the clarity and visibility of TTE images can arise due to differences in probe handling by various clinicians. To address this, Abdi et al. [43] developed a CNN model. The model calculates a quality score for these images, which ranges from 0 (unacceptable) to 5 (good). Building on this, Labs et al. [44] developed QA-NET, a multi-output regression model for assessing the quality of echocardiogram images. This model accurately predicts quality scores for clarity, visibility, depth gain, and foreshortenedness.

Further advancements were made by Diller et al. [23], who evaluated the effectiveness of DL algorithms in reducing acoustic shadowing artifacts and denoising TTE images in patients with Congenital Heart Disease (CHD). They developed deep neural network-based autoencoders to denoise and eliminate acoustic shadowing artifacts. The autoencoders significantly improved image quality across diagnostic subgroups. Similarly, Jalali et al. [45] developed a technique to decompose A4C images to reduce the impact of speckle artifacts. The method separates the image into a texture component, consisting of highly oscillating and repeated patterns, and a cartoon image, containing smooth areas and sharp edges. Convolutional sparse coding was used to solve the decomposition optimization function, significantly reducing speckle noise in TTE images. Numerical results demonstrated the effectiveness of this approach.

Sanjeevi et al. [17] proposed a real-time convolutional autoencoder pipeline for noise reduction in TTE videos. The pipeline was evaluated on TTE video data with added Gaussian noise. The developed model

**Table 3**  
Summary of TTE view classification using deep learning.

Study	No. of view	TTE view	Model	No. of samples	Hyper parameters	Acc.
Balaji et al. [27]	4	PLAX PLSX A2C A4C	BPNN & SVM	200 Images	NA	87.5%
Gao et al. [29]	8	PLAX A2C A3C A4C A5C PSA of mitral PSA of aorta PSA of papillary	Fusion CNN	432 Videos	Opt.: Gradient Descent LR: 0.01 BS: 100 Initial bias: 0.1	92.1%
Madani et al. [30]	15	***	VGG-16	267 Videos	Opt.:RMSProp Epochs: 45 BS: 46	91.7%
Kusunose et al. [31]	5	PLAX PLSX A2C A4C A5C	2D CNN	340 Videos	Opt.: Adam Epochs: 50 LF: CE	98.1%
Howard et al. [32]	14	***	Xception Modified C3D Time Distributed, Time Stream CNN	374 Videos	Opt.: Adam BS: 20 LF: CE	CKC = 0.957
Gu et al. [33]	13	***	Efficient-EvidNet	3151 Images	Opt.: Adam LR = $2.5 \times 10^{-4}$ Epochs: 30 Pretrain: ImageNet	91.6%
Gao et al. [34]	9	***	Incep-tionV3+GNN	684 Videos	Opt.: Adam Epochs: 500 BS: 8 Alpha: 0.4 LR: 0.0001 Pretrain: ImageNet	97.7%
Huang et al. [35]	29	***	CNN+ Autoencoder	26425 Images	NA	98.2%
Huang et al. [39]	4	PLAX PLSX A2C A4C	3D CNN	11317 Images	Opt.: Adam LF: CE LR: 0.0001 Plateau WD rate: 0.9	99.1%
Lin et al. [38]	3	A2C A4C Apical-long axis	Xception	6953 Images	NA	96%
Zhang et al. [37]	23	***	VGG-16	277 Videos	Opt.: Adam LR: 0.00001 BS: 64 WD: $1 \times 10^{-8}$ Epochs: 20	86%
Yu et al. [41]	2	A2C A4C	ResNet18	724 Videos	Opt.: Adam LF: CE Reg.: L2	AUC = 1.00

Acc. - Accuracy; Comp. — Comparison; BPNN — Backpropagation Neural Network; SVM — Support Vector Machine; NA — Not Available; CNN — Convolutional Neural Network; Opt. — Optimizer; BS — Batch Size; LF — Loss Function; LR — Learning Rate; WD — Weight Decay; GNN — Graph Neural Network; AUC — Area Under Curve; CKC — Cohen's Kappa Coefficient; Reg. — Regularization; \*\*\* - TTE views are presented in specific study paragraphs.

can eliminate noise, reconstruct images using 70% of their original features, and outperform the conventional Gaussian filter in reducing Gaussian noise. Similarly, Beevi et al. [46] developed a Convolutional-based Improved Despeckling Autoencoder (CIDAE) to remove speckle noise in TTE images of patients with regional wall motion abnormalities. The developed model achieved the best MSE, PSNR, and SSIM values among all speckle noise variation levels.

The studies employed different approaches to improve the TTE image quality. Liao et al. [18] developed a Quality Transfer StarGAN (QT-StarGAN) network to transfer the quality of echocardiogram images. The network was trained using echo images and corresponding quality labels. The authors initially used StarGAN [47] for the quality transfer task but found that it led to reconstructing artifacts from the transferred image quality. They made significant improvements to

the generator and discriminator networks of StarGAN, enabling the generation of images that closely resemble clinically collected data. The TTE view detection task was used as a quantitative evaluation, and the developed QT-StarGAN significantly improved classification accuracy. Building on this, Ting et al. [19] proposed a DL-based fusion technique to merge two spatially distinct TTE images, specifically the apical and parasternal views. Their method uses a mutual information estimating network to optimize the mutual information between the source and fused images and an autoencoder framework to generate the fused image. The fused image produced by this technique reduced noise and stitching artifacts.

To address the need for manual labeling of images and data collection for DL tasks, Tiago et al. [48] introduced an image creation pipeline based on a Generative Adversarial Network (GAN) to generate 3D echocardiogram images with corresponding ground truth labels. The pipeline uses the heart's complex anatomical segmentations as sources for ground truth labels. The feasibility of the generated images was evaluated by segmenting the left atrium (LA), LV, and myocardium using segmentation methods. A quantitative analysis of the 3D segmentation provided by the models trained with the generated images indicated the potential use of this GAN approach to generate 3D data and train DL models for various clinical tasks.

Meanwhile, Monkam et al. [49] proposed a data augmentation method, DF-Aug (Disentanglement and Fusion Augmentation), to address the shortage of annotated data for TTE segmentation. The technique involves image disentanglement and fusion, separating an image into its texture and cartoon components using Texture + Cartoon decomposition. The texture component of a randomly selected unlabeled image is then integrated into each image in the annotated dataset, creating a new annotated image with unique properties for the LV segmentation task. This process expands the size of the available training dataset. The proposed method improved the performance of the U-Net task of LV segmentation, increasing the Dice and IOU scores by 4.48%, 3.55%, and 2.36%, 1.76%, respectively, for the CAMUS dataset [50] and the in-house dataset.

Narang et al. [51] developed DL-based software to guide nurses in obtaining TTE images for quantifying cardiac function. Eight nurses with no experience with echocardiography were trained using this DL software to perform echocardiogram scans. These scans were then compared with those performed by sonographers. Using the DL software, each of the eight nurses scanned 30 patients, yielding results considered to be of diagnostic quality in 237 of 240 cases (98.8%) for pericardial effusion, left ventricular function and size, and in 222 of 240 cases (92.5%) for right ventricular size. This DL-based software enables beginners to obtain TTE for clinical evaluation, demonstrating the potential of DL in enhancing the accessibility and quality of TTE imaging.

Image quality is a crucial factor in diagnosing diseases using TTE. Table 4 summarizes the improvements in TTE image quality based on DL models. Research has explored the application of DL models in various areas, such as image quality assessment [43,44] quality enhancement [19,48], noise reduction [17,23,45,46], and data augmentation to address issues arising from dataset limitations. Analyzing noisy images or videos can be challenging as noise can distort the structures or patterns of image features, potentially leading to inaccurate clinical interpretations or conclusions.

Key research findings:

- DL models have been utilized for image quality analysis, which ensures that only high-quality images are passed on for subsequent analysis, thereby streamlining and expediting clinical decision-making.
- Autoencoders have demonstrated promising results for noise reduction and image quality improvement in TTE images or videos. However, autoencoders have only reduced synthesized speckle noise, Gaussian noise, and shadowing artifacts.
- The introduction of the GAN model in TTE data augmentation has aided in model training for LV segmentation by increasing the number of samples in the dataset.
- There are open research challenges for applying DL models in real-time image quality improvement:
  - Temporal resolution improvement: The temporal resolution of TTE video is compromised in patients with high heart rates. The interval between heartbeats shortens when the heart rate is high, leaving less time for the echocardiogram device to capture an image before the next heartbeat. This could result in visual blurring or missing events in the TTE video, making it difficult to identify cardiac issues. In such scenarios, DL solutions for frame interpolation [55] can be explored, where new frames are interpolated between the old frames to compensate for blurring and missing events.
  - Spatial resolution improvement: The image quality is often poor for patients with lung diseases, specifically Chronic Obstructive Pulmonary Disease (COPD). Individuals with COPD often have hyper-inflated lungs, which can result in a thicker chest wall. This might make it more difficult for ultrasonic waves to reach the heart, leading to poor image resolution. Open research challenges exist to explore the DL models to improve the disease-specific image quality. DL models may incorporate the super-resolution technique [56] to enhance image quality and resolution. This can be achieved by training a DL model on a batch of high-resolution images to predict the missing data in low-resolution images.

### 3.4. Cardiac structure segmentation or identification

The process of quantifying the shape and deformation of the heart wall is crucial for assessing cardiac function in echocardiogram images or videos. This assessment relies heavily on accurately segmenting the heart wall in TTE image frames. Image segmentation, which divides an image into different regions, aids in evaluating the heart chambers. The evaluation of the LV is critical as it is responsible for pumping blood throughout the body. In TTE, the automatic segmentation of the LV can serve as a valuable tool for various tasks. These tasks include calculating LV parameters (such as volume, area, and EF), detecting regional motion abnormalities, identifying cardiac hypertrophy, and diagnosing cardiac ischemia. Furthermore, identifying heart valves, including the mitral valve, aortic valve, and tricuspid valve, enhances the prognosis of these valves.

#### 3.4.1. Structure segmentation

The field of cardiac structure segmentation from TTE using DL has significantly evolved. This area is rich and diverse, featuring a variety of state-of-the-art models. In the early stages, U-Net models were predominantly used for LV segmentation. These models were later enhanced by incorporating elements such as residual connections, atrous spatial pyramid pooling, and attention mechanisms. To boost performance further, CNN models were integrated with recurrent neural networks (RNN), specifically long short-term memory (LSTM). Recently, transformers have also been employed for cardiac segmentation.

Most of the literature has utilized the DL network called U-Net [28], a variant of a CNN. Some studies have utilized SegNet [57], an encoder-decoder-based CNN, specifically for segmenting LV [24,50,58]. Fig. 7 illustrates the U-Net. The SegNet model mirrors the architecture of autoencoders and outputs the segmented part from TTE images. A skip connection between the encoding and decoding layers in U-Net makes it different from SegNet.

Smistad et al. [59] introduced a fully convolutional network (FCN) model specifically for segmenting the LV from TTE images. They discovered that pretraining the FCN with an automatic Kalman filter

**Table 4**  
Summary of TTE image quality improvement using deep learning.

Study	Problem	TTE view	No. of samples	Model	Hyper parameters	Metrics	Value/Result
Abdi et al. [43]	Quality Assessment	A4C	6916 Images	CNN	Opt.: PSO [52] Epochs: 54 ±6 Reg.: L2 BS: 36 LR: 0.0002 Momt.: 0.95	MAE	0.71
Labs et al. [44]	Quality Assessment	A4C PLAX	33784 Images	QA-Net	Opt.: Multi-label Opt.[53] LR: 0.002 Momt.: 0.95 Decay Rate: 0.1 Epochs: 40 BS: 32	MAE	0.375
Diller et al. [23]	Noise Removal Artifact Removal	A4C	152 Videos	Autoencoder	Opt.: Adadelta Epochs: 200-400  BS: 16 LF: BCE	CE	0.2899 0.2744
Jalali et al. [45]	Noise Reduction (Speckle)	A4C A3C PSAX	6 Videos	Conv. Sparse Encoding	Opt.: ADMM	MAE CNR SSIM	0.0641 4.5572 dB 0.9989
Sanjeevi et al. [17]	Noise Reduction (Gaussian)	A4C	53 Videos	Autoencoder	Opt.: Adam Epochs: 500 BS: 5 LF: BCE	PSNR SSIM MSE	68.5 dB 0.71 0.0085
Ting et al. [18]	Image fusion	A4C PLAX	9 Videos	Autoencoder	Opt.: SGD BS: 1 LR: 0.03 Momt.: 0.9 WD: $5 \times 10^{-6}$	SNR CNR	26.41 dB 20.79 dB
Liao et al. [19]	Data Augmentation	14 Views*	3, 157 Images	QT-StarGAN	Opt.: Wasserstein GAN with gradient penalty [54]	–	The model generate images visually similar to clinical data.
Tiago et al. [48]	Data Augmentation	3D	13 Videos	3D Pix2pix	Opt.: Adam Epochs: 200 LR: 0.0002	–	The generated images shows clear heart structures
Monkam et al. [49]	Data Augmentation	A4C	CAMUS 500 Patients	Cartoon+ Texture Decomp.	Opt.: Chambolle and Pock's algorithm $\lambda = 0.8$	–	Improved segmentation dice and IOU scores
Beevi et al. [46]	Noise Reduction (Speckle)	PLAX PSAX A2C A4C	33096 Images	CIDAE	Epochs: 200 Opt.: Adam LR: 0.001	PSNR SSIM MSE	26.89 db 0.88 133.08

Momt. — Momentum; MAE — Mean Absolute Error; BCE — Binary Cross Entropy; CE — Cross Entropy; CNR — Contrast to Noise Ratio; SNR — Signal to Noise Ratio; SSIM — Structural Similarity Index Measure; PSNR — Peak Signal-to-Noise Ratio; SGD — Stochastic Gradient Descent; ADMM — Alternating direction method of multipliers; Conv. — Convolutional; Decomp. - Decomposition;

\* Indicates the study used 14 different views of TTE images (View Classification).

segmentation method could substantially reduce the need for manual annotation. Building on this, Leclerc et al. [24] contributed to the field by presenting the largest publicly available, fully annotated dataset (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) CAMUS, for TTE assessment. They evaluated various DL models (U-Net, Automated Constrained Neural Network (ACNN), U-Net++, Stacked Hourglasses (SHG)) and non-DL models (Structured Random Forests (SRF), B-Spline Explicit Active Surface Model (BEASM)) on LV segmentation from TTE. The U-Net model outperformed the others. Further, Leclerc et al. [50] developed a unique attention mechanism in the Refining U-Net (RU-Net), a combination of two U-Nets, to enhance the 2D TTE segmentation of the LV endocardium and epicardium.

Jafari et al. [60] proposed a recurrent fully convolutional network (RFCN) with optical flow for segmenting the LV in TTE videos.

The RFCN uses convolutional bi-directional Long Short-Term Memory units and optical flow to analyze temporal information and estimate motion between successive frames. Further advancements made by Lin et al. [61] presented convolutional long-short-term memory attention-gated U-Net (CLA-U-Net), which incorporates a convolutional long-short-term memory (C-LSTM) block to capture the temporal information between video frames and a channel attention mechanism to suppress noise and enhance desirable features.

Zyuzin et al. [58] improved upon the U-Net architecture by integrating residual blocks for segmenting the LV endocardium, epicardium, and LA. Their model demonstrates the superiority of the residual blocks added to U-Net architecture over the conventional U-Net architecture without additional skip connections. On top of this, Amer et al. [62] introduced the Residual dilated U-Net (ResDUNet) based on an embedded U-Net [63]. The model, which uses residual blocks instead of

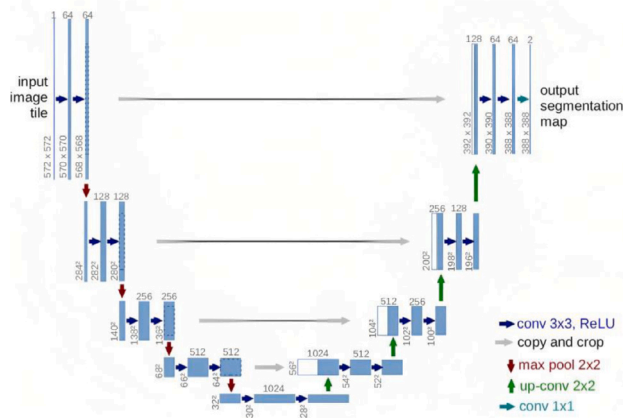


Fig. 7. Illustration of U-Net.  
Source: Reproduced  
from [28].

U-Net blocks, was evaluated on the CAMUS dataset. Similarly, Leclerc et al. [64] introduced a multi-stage network, LU-Net, built around the U-Net architecture, which is divided into two phases: the first extracts the LV region and its mask, and the second uses the retrieved image to segment the area. Ali et al. [65] proposed a hybrid net deep Res-U network to extend the research. The developed ResU, an encoder in the U-Net, is a modified version of ResNet-50. They are combining U-Net and ResNet's advantages, outperforming state-of-the-art techniques.

Azarmehr et al. [66] evaluated the performance of different segmentation models, including U-Net, Seg-Net, and Fully Convolutional DenseNet (FC-DenseNet), in segmenting the LV endocardium. The U-Net model emerged as the best performer with an average DSI of  $0.93 \pm 0.04$  and a Hausdorff distance of  $4.52 \pm 0.90$ . On the other hand, Zhu et al. [67] compared a CNN and active contour (AC) in LV segmentation. The CNN achieved performance metrics values, such as a DSI of 0.85 and HD of 5.77, comparable to AC. However, while achieving similar accuracy with less training data and runtime, the AC method is semi-automated and requires some preprocessing steps. On top of this, Kim et al. [68] conducted a comparative study of popular segmentation algorithms, including Res-U-Net, Simple U-Net, and Dense-U-Net, to evaluate the performance in LV segmentation.

Arafati et al. [69] presented a multi-label segmentation method using adversarial training and deep convolutional networks. They addressed the generalization issue by training a GAN for pixel classification. Their evaluation of the model's performance against state-of-the-art methods confirmed its generalizability. The authors segmented all cardiac chambers on their test dataset, such as LV, right ventricle (RV), LA, and right atrium (RA), on their test dataset. Building on this, Thi et al. [70] addressed two primary challenges of the DL technique for image segmentation: the difficulty in obtaining enough training data and the time required to construct high-quality models. They presented a pipeline that uses active learning to simplify labeling tasks and neural architecture search to create a suitable DL model. Their approach achieved the same Intersection over Union (IOU) accuracy of 87.0% using only a quarter of the initial training dataset.

Ibtehaz et al. [71] introduced the MultiResUNet for biomedical image segmentation. Azizi et al. [72] updated the MultiResUNet model by including the Attention and Atrous Spatial Pyramid Pooling blocks to partition the LV. Similarly, Chen et al. [73] proposed UDeep, an encoder-decoder model considering multi-scale, low-level fine-grained, and high-level semantic information. The encoder extracts semantic features at multiple scales using the atrous spatial pyramid pooling module [74] and the computation-efficient separated Xception backbone [75]. A decoder module consisting of upsampling fusion modules is used simultaneously to fuse features of various levels.

Zeng et al. [76] introduced the multi-attention efficient feature fusion network (MAEF-Net), which uses a multi-attention mechanism, including channel and spatial attention, and incorporates a spatial pyramid feature fusion and deep supervision mechanism. Similarly, the Deep Pyramid Local Attention Neural Network (PLANet) was proposed by Liu et al. [77] to improve performance. The model uses a pyramid local attention module and a label coherence learning process. To extend this, Su et al. [78] developed a DL model based on joint attention. This model comprises two variants: one excels in segmentation performance, while the other, despite a slight decrease in segmentation performance, offers faster embedding with smaller model size.

Monkam et al. [79] proposed a CNN-based multitask learning architecture to segment the LV, LV wall simultaneously, and LV+LV wall. Using U-Net and DeeplabV3, they achieved an average DSI of 0.88, concluding that a single CNN model could define several cardiac structures with superior performance. To extend this, A DL network called VDS-UNET was proposed by Huang et al. [80] to segment the LA, LV, and Mitral Valve (MV). The model builds on the VGG-16 convolution layers and adds depth supervision to the expansion path.

Further advancements made by Li et al. [81] proposed the multi-task DL model EchoEFNet, which simultaneously segments the LV, detects the mitral annulus, and identifies the apical landmark. Utilizing ResNet50 with dilated convolution as the backbone, this network preserves spatial characteristics while extracting high-dimensional features. The branching network concurrently detects landmarks and segments the LV, employing a multi-scale feature fusion decoder for LV parameter calculation. It achieved a percentage of correct key points (PCK) accuracy for landmark detection on the CAMUS and CAMEcho datasets of 0.73 and 0.96, respectively.

Liao et al. [82] introduced two DL models utilizing Transformers [83] as the base architecture for LV segmentation. The first model combines Swin Transformer and K-Net, while the second employs Segformer. The authors achieved reliable segmentation results through post-processing by removing regions with the largest pixel square or perimeter. In a series of studies, Delgri et al. [84] and Hamila et al. [85] proposed a motion abnormality detection framework that includes LV segmentation using SegNet. Zhang et al. [37] and Huang et al. [39] used the U-Net for LV segmentation in their DL framework. Similarly, Lin et al. [38] and Ouyang et al. [25] proposed a DL framework to analyze TTE, including LV segmentation, using LSTM-Unet and Deeplab V3, respectively.

In contrast with previous studies, the researchers explored the cardiac valve segmentation from TTE. Lai et al. [86] developed a Bi-Directional ConvLSTM U-Net (BDCU-Net) trained to segment the aortic valve area in PSAX TTE images. In parallel, Rodríguez et al. [87] tried different DL models such as U-Net, Unet-mini, VGG-Unet, and ResNet50-Unet for segmenting the coronary artery specifically for patients with Kawasaki disease. The ResNet50-Unet model with transfer learning achieved the best results.

### 3.4.2. Structure identification

Nizar et al. [88] proposed a CNN-based model to detect the aortic valve during a live echocardiogram examination. The authors evaluated the effectiveness of two object detection models: the Faster Regional based CNN (R-CNN) and the Single Shot Multibox Detector (SSD). The Faster R-CNN with Inception v2 model achieved the best result. Similarly, Chandra et al. [89] proposed a model using the Mobilenet as the backbone of the Yolo model to track the mitral and tricuspid valves.

The segmentation or identification of heart structures from TTE images or videos is essential for quantifying cardiac parameters and extracting features or biomarkers to detect cardiac diseases [24,50,58, 62,66]. Table 5 summarizes the performance of various DL models used for segmenting and identifying cardiac structures in TTE images and videos. The introduction of the CAMUS and EchoNet-Dynamic datasets has enabled researchers to explore different DL frameworks for LV segmentation.

**Table 5**

Summary of segmenting and identifying cardiac structures from TTE using deep learning (The DSI is the primary metric in the DSI/Acc/IOU column, and the other metrics are explicitly mentioned).

Study	TTE View	No. of samples	Model	Hyper parameters	Cardiac structure	DSI/IOU/Acc.	HD
Smistad et al. [59]	A4C	1500 Videos	U-Net	Opt.: SGD Epochs: 10 LR: 0.01 Momt.: 0.9	LV	0.87	5.90
Jafari et al. [60]	A4C	566 Videos	U-Net+ Bidirectional Conv. LSTM	Opt.: Adam LR: 0.0001 BS: 10 LF: Dice Loss	LV	0.92	–
Leclerc et al. [24]*	A4C A2C	CAMUS 500 Patients	U-Net	Opt.:Adam Epochs: 30 LR: 0.0001 BS: 10 LF: CE+WD	LV	0.95	5.70
Leclerc et al. [50]	A4C A2C	CAMUS 500 Patients	RU-Net	LF: Multi-class dice loss	LV	0.92	5.90
Zyuzin et al. [58]	A4C A2C	CAMUS 500 Patients	U-Net+ ResNet	Opt.: Adam LR: 0.001 BS: 32	LV LA	0.94 0.90	–
Azarmehr et al. [66]	A4C	61 Videos	U-Net Seg-Net FC-Dense-Net	Opt.: Adam Epochs : 250 LR: 0.00001 LF: Negative log likelihood loss	LV	0.93	4.52
Amer et al. [62]*	A4C A2C	CAMUS 500 Patients	ResDUnet	Opt.: Adam Epochs: 1000 LR: 0.0001 LF: BCE	LV	0.95	4.50
Leclerc et al. [64]*	A4C A2C	CAMUS 500 Patients	LU-Net	Opt.: Adam LR: 0.001 or 0.0001 Epochs: 20 LF: Multi class dice loss	LV	0.95	5.30
Arafati et al. [69]*	A4C A2C	100 Videos	VGG + GAN	Opt.: Mini-batch GD LF: PWCE + CE Epochs: 90 Pretrain: PASCAL dataset (VGG)[90]	LV RV LA RA	0.92 0.86 0.89 0.91	5.19 7.16 5.20 4.86
Ali et al. [65]	A4C A2C	CAMUS 500 Patients	Deep Res U	Opt.: Adam LF: CE WD: $1 \times 10^{-4}$ LR: 0.0001	LV	0.97	2.56
Zhu et al. [67]*	A4C A2C PSAX	1500 Videos	CNN AC	NA	LV	0.85	5.88
Thi et al. [70]	A2C	252 Videos	CNN	Various Parameters**	LV	IOU = 87%	–
Lin et al. [61]	A4C	EchoNet- Dynamic 10030 Videos	CLA-U-Net	Opt.: SGD LR: 0.00001 Momt: 0.9 Epochs: 20. LF: BCE	LV	0.93	–
Chen et al. [73]	A4C	EchoNet- Dynamic 10030 Videos	UDeep	Opt.: Adam LR: 0.005 betas: (0.9, 0.999) eps: $1 \times 10^{-8}$ $\beta$ : $5 \times 10^{-2}$ , $\lambda$ : $1.05 \times 10^{-3}$ /epoch  LF: PSP Pretrain:Image Net	LV	0.92	–

(continued on next page)

Table 5 (continued).

Study	TTE View	No. of samples	Model	Hyper parameters	Cardiac structure	DSI/IOU/Acc.	HD
Azizi et al. [72]	A4C	EchoNet-Dynamic 10030 Videos	Modified MultiResUNet	Opt.: Adam Epochs: 50 BS: 16 LR: 0.001 LF: BCE	LV	0.92	–
Kim et al. [68]*	A4C A2C PSAX	500 Videos	U-Net Res-U-Net Dense-U-Net	Opt.: Adam LF: PWCE LR: 0.0001 Epochs: 200 BS: 5	LV	0.83 - 0.95	–
Zeng et al. [76]*	A4C	EchoNet-Dynamic 10030 Videos	MAEF-Net	Opt.: SGD BS: 16 Epochs: 50 LR: $10^{-5}$ Momt.: 0.9 LF: CE+Soft Dice loss	LV	0.93	–
Monkam et al. [79]	A4C	CAMUS 500 Patients	U-Net DeepLab V3	Opt.: Adam LR:0.01 BS: 8 Epochs: 100 $\beta1 = 1 \times 10^{-8}$ $\beta2 = 0.9$ LF: BCE	LV LV wall LV+LV wall	0.88	–
Liu et al. [77]*	A4C	CAMUS 500 Patients EchoNet-Dynamic 10030 Videos	PLA-Net	Opt.: mini-batch SGD Momt.: 0.9 WD: $1 \times 10^{-4}$ BS: 24 Epochs: 500 LR: 0.1 to 0.0001	LV	0.95	4.4
Hunag et al. [80]	A4C A3C A2C	49 Videos	VDS-UNET	Opt.: RMSProp LF: Weighted CE Epochs: 200 LR:0.001 BS: 1000	LV LA MV	0.935 0.915 0.757	–
Lai et al. [86]	PSAX	58 Videos	Bidirectional ConvLSTM-Unet	BS: 1000	Aortic Valve	IOU = 91%	–
Rodríguez et al. [87]	PSAX	1531 Videos	ResNet50+ Unet	Opt.: Adadelta Transfer Learning	Coronary Artery	IOU = 54.54%	–
Nizar et al. [88]	A5C	46 Videos	Faster RCNN SSN	Various Parameters**	Aortic Valve	98.6%	–
Chandra et al. [89]	A4C	40 Videos	MobileNet+ Yolo	BS: 3 LR :0.0001 Epochs: 40 Pretrain: ImageNet	Mitral Valve Tricuspid Valve	98% 90%	–
Ouyang et al. [25]	A4C	EchoNet-Dynamic 10030 Videos	EchoNet	Opt.: SGD LF: Pixel-level BCE Pretrain: Kinetics-400 dataset	LV	0.92	–

(continued on next page)

## Key research findings:

1. Numerous studies have demonstrated the effectiveness of U-Net and its variants in segmenting the LV from TTE images. Models based on U-Net have achieved state-of-the-art performance in this task.
2. Modifications to U-Net, such as adding or combining residual blocks, atrous spatial pyramid pooling, feature fusion, and attention mechanisms, have improved performance over the original U-Net in LV segmentation.
3. The study by Ali et al. [65] developed the Deep Res U, a combination of U-Net and ResNet, which achieved the highest Dice score of 0.97 in LV segmentation on the CAMUS dataset.
4. Combining recurrent neural networks with convolutional neural networks, specifically Long Short-Term Memory, can facilitate the training of models with temporal data associated with TTE videos. Recent studies have also demonstrated the efficiency of transformers in LV segmentation.
5. Despite extensive research in cardiac structure segmentation, there are still many gaps:

Table 5 (continued).

Study	TTE View	No. of samples	Model	Hyper parameters	Cardiac structure	DSI/IOU/Acc.	HD
Huang et al. [39]	PLAX PSAX A4X A2C	2736 Images	U-Net	Opt.: Adam LR: 0.0001 Plateau WD rate: 0.9 LF: 1- Dice similarity	LV	0.75	–
Delgri et al. [84]	A4C	HMC-QU 160 Videos	SegNet	Opt.: Adam BS: 32 Epochs: 25 LR: 0.001	LV	99.4%	–
Hamila et al. [85]	A4C	165 Videos	SegNet	Opt.: RMSProp LF: MSE Epochs: 100 BS: 256	LV	97.1%	–
Lin et al. [38]	A2C A4C ALX	6953 Images	LSTM-UNet	NA	LV	0.89	–
Zhang et al. [37]	PLAX PSAX A2C A3C A4C	791 Images	U-Net	LF: PWCE+ Distance based loss penalty	LV	IOU = 92%	–
Yu et al. [41]	A4C	1610 Videos	U-Net++	Opt.: Adam LF: CE Reg.: L2	LV	0.86	7.44
Qu et al. [91]	A4C	9676 Images	LACNet	NA	LV	0.96	–
Liao et al. [82]	A4C	Echonet-Dynamic 10030 Videos	Swin Transformer+ K-Net Segformer	Epochs: 50 Opt.: AdamW LF: CR LR: 0.00006 Pretrain: ADE20K	LV	92.92% 92.80%	–
Su et al. [78]	A4C	Echonet-Dynamic 10030 Videos	JANet	Epochs: 80 Pretrain: ImageNet LR: 0.0001 to 0.000001 LF: CE+Dice loss	LV	93.69%	–
Li et al. [81]	A2C A4C	CMUEcho (764 Videos) CAMUS (500 Patients)	EchoEF Net	Epochs: 40 BS: 24 Opt.:Adam LR: 0.001 gamma=0.1	LV	0.96 0.93	3.0 5.4

DSI — Dice Similarity Index; HD — Hausdorff Distance; IOU — Intersection Over Union; LV — Left Ventricle; BCE — Binary Cross Entropy; PWCE — Pixel-wise Cross Entropy; GD — Gradient Descent; PSP — Pseudo-Segmentation Penalty; ALX — Apical long axis.

\* Indicates the studies computed cardiac parameters for comparing segmentation results.

\*\* Indicates the studies used different models with different hyperparameters.

- There is no public dataset available for segmenting cardiac chambers other than the left ventricle.
- Segmentation of other cardiac structures, such as the left atrium, right atrium, and right ventricle, would be useful in quantifying cardiac parameters relative to these structures.

### 3.5. Quantification of cardiac function

This section covers the comprehensive review of studies computing and using cardiac parameters for clinical evaluation. To assess cardiac dysfunction, one of the most crucial measurements of cardiac function is left ventricle ejection fraction (LVEF), which is the ratio of change in the LV end-systolic and end-diastolic volumes. Also, measuring other LV clinical parameters such as area, length, mass, and thickness is essential. Most studies [37,38] used the segmentation output to compute the parameters. The area and axis length parameters are manually extracted from the segmented output. The EF is calculated using the

single plane method applied to the A4C view TTE image, as represented by Eq. (1).

$$\begin{aligned}
 EDV &= \frac{8A_d^2}{3\pi L_d} \\
 ESV &= \frac{8A_s^2}{3\pi L_s} \\
 EF &= \frac{EDV - ESV}{EDV}
 \end{aligned} \tag{1}$$

EDV and ESV represent the ventricular volumes at end-diastole and end-systole, respectively. The LV area at end-diastole and end-systole are denoted as  $A_d$  and  $A_s$ , respectively. The  $L_s$  and  $L_d$  refer to the long-axis lengths of the LV at the end of systole and diastole, respectively.

Detecting end-diastolic and end-systolic image frames is crucial for predicting the EF. Several researchers have developed innovative approaches to address this challenge. Dezaki et al. [92] treated

the identification of cardiac phases (end-systolic and end-diastolic) as a regression problem. They developed DL-based models integrating RNNs to capture the temporal correlations between individual frames in a sequence and CNNs for image feature extraction. The authors evaluated four RNN architectures: LSTM, bi-directional LSTM, Gated Recurrent Unit (GRU), Bi-GRU, and two CNN architectures, ResNet and DenseNet. The combination of GRU and DenseNet emerged as the optimal model for phase detection, with an average frame mismatch for the end-diastolic and end-systolic frames of 0.20 and 1.43, respectively.

In a separate study, Lane et al. [93] developed deep neural networks to detect end-systolic and end-diastolic frames in A4C multi-beat TTE videos of any length. These models were trained and evaluated on patient data from multiple centers, including the MultiBeat, PACS-dataset, and EchoNet-Dynamic datasets. The authors employed a long-term recurrent convolutional network. Compared to the ground truth, the developed model achieved an average frame difference of 0.11 and  $-0.09$  for end-systolic and end-diastolic frames, respectively. On top of this, Farhad et al. [94] introduced a DL model named DeepPhase, designed to identify the end-systolic, end-diastolic, and non-ES/ED phases from TTE images. The model was trained and validated using three distinct TTE image datasets: the CAMUS, EchoNet Dynamic, and CardiacPhase datasets. DeepPhase demonstrated robust performance, achieving AUC values of 0.96 and 0.82 on the CAMUS and CardiacPhase datasets. The authors further proposed a cropping method to enhance the model's functionality.

Recent studies have sought to predict the EF directly from TTE videos. For this purpose, Ouyang et al. [25] developed EchoNet, a video-based DL model that uses residual connections and spatiotemporal convolutions across echocardiogram frames to predict EF. The model achieved a mean absolute error (MAE) of 4.1%. Building on this, Ghorbani et al. [95] predicted the EF and the end-diastolic and systolic volumes. They compared the performance of a direct "end-to-end" DL prediction of the EF with a DL model that calculates the EF from the end-systolic volume and end-diastolic volume, following the typical human workflow. They found that the "end-to-end" DL model outperformed the latter.

Reynaud et al. [96] introduced a novel architecture for TTE video interpretation based on a residual auto-encoder network and a modified version of the BERT (Bidirectional Encoder Representations from Transformers) model for token classification. The model, capable of processing videos of any duration, automatically computes the EF and identifies end-systolic and end-diastolic frames. It achieved an MAE of 5.95. In parallel, Farzy et al. [97] proposed hierarchical vision transformers to extract the EF. This approach could estimate the EF without the need for LV segmentation.

Building on this, Muhtaseb et al. [98] introduced the Echo Convolutional Transformer (EchoCoTr) method to measure the EF using TTE videos. This method leverages the power of CNNs and visual transformers to analyze echocardiogram video sequences and predict cardiac EF. The EchoCoTr method demonstrated superior performance, achieving an  $R^2$  of 0.82 on the EchoNet-Dynamic dataset. On the other hand, Mokhtari et al. [99] developed a graph neural network called EchoGNN to estimate the EF from TTE videos. This model infers a latent echo graph from the frames of one or more echo image sequences and then calculates weights for the nodes and edges of the graph, emphasizing the importance of specific frames for EF estimation.

Lin et al. [38] proposed an automated TTE analysis framework that includes view classification, segmentation, regional wall motion abnormality detection, and quantification of cardiac function. They categorized patients into three clinically significant LVEF groups: preserved ( $>50\%$ ), decreased ( $<40\%$ ), and medium ( $40\%–50\%$ ). The accuracy of this prediction was 77%, comparable to echocardiographic reports from clinicians. Similarly, Zhang et al. [37] introduced a DL model encompassing LV segmentation, cardiac function quantification, and disease detection. The output from the segmentation was utilized to calculate the longitudinal strain of the LV and estimate the EF, LV

mass, and chamber volumes. These parameters were computed using a mathematical model based on the American echocardiogram guidelines. The researchers expanded the automation to patients diagnosed with specific diseases, such as cardiac amyloidosis and hypertrophic cardiomyopathy. They found that the automated evaluations of LV mass and left atrial volume differed significantly between patients and controls for these conditions.

Duffy et al. [100] evaluated the efficacy of a DL approach in detecting increased LV wall thickness. For the domestic external dataset, the mean absolute errors (MAEs) of intraventricular septum thickness, LV internal dimension, and LV posterior wall thickness were 1.7 mm, 3.8 mm, and 1.8 mm, respectively. For the external international dataset, the corresponding MAEs were 1.7 mm, 2.9 mm, and 2.3 mm. Additionally, Lau et al. [40] studied the correlation between outcomes and TTE markers produced by DL. The authors created a 3D CNN for the left atrial dimension, left ventricular wall thickness, chamber diameter, and EF measurement in the TTE.

Recent studies emphasize the importance of computing the RV parameters, as RV dysfunction can act as a prognostic marker across various cardiovascular diseases. These parameters are significant predictors of clinical symptoms in patients with pulmonary hypertension. Notably, RV dysfunction is prevalent in approximately half of the individuals with heart failure and a decreased LVEF. As a result, an essential component of cardiovascular care should include screening for RV dysfunction. In their study [101], Bohoran et al. proposed computing the RV end-diastole and end-systole volume using an attention-based model on tabular data. The developed transformer model receives data on age, cardiac phase, gender, and morphological measures (areas) as input from eight different views of TTE images. A feature tokenizer module converts all features to numerical and categorical.

Building on this, Tokodi et al. [102] aimed to predict RVEF from TTE videos. Multiple spatiotemporal CNNs were trained to predict RVEF using the TTE videos. An ensemble model was created by combining the three highest-performing networks, and it was then assessed using a different set of data. In the internal validation set, the ensemble model predicted RVEF with a mean absolute error of 4.57 percentage points, while in the external validation set, it was 5.54 percentage points. In the latter case, the model's accuracy is 78.4% in identifying RV dysfunction.

Table 6 encapsulates the performance of DL models in quantifying cardiac function, specifically the EF. Initial studies [37,38] computed heart parameters using a mathematical model based on the Simpson bi-plane method or American echocardiography recommendations [103]. These studies underscored the importance of precise heart chamber segmentation by quantifying heart function based on the results of LV segmentation. The manual nature of applying these mathematical models for quantifying cardiac function introduces challenges. The process is subject to interobserver variability, and cardiac parameters, specifically, EF, can fluctuate from one heartbeat to the next. The complexity of the procedure is further compounded by the recommendations from the European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE) to monitor up to five consecutive heartbeats [103].

Key research findings:

1. Most studies have predicted the left ventricle ejection fraction (EF) from TTE videos. The study by Muhtaseb et al. [98] achieved the state-of-the-art  $R^2$  score in EF prediction in EchoNet-Dynamic Dataset. Recent studies [101,102] have shifted focus towards right ventricle parameter computation. However, there is a need for active research to predict whole cardiac quantification for improved patient outcomes.
2. The capabilities of CNNs and transformers have been harnessed for cardiac quantification [25,95–98]. The initial layers of CNNs possess small receptive fields, but they gradually expand their field of view through convolution operations. In contrast, Vision Transformers (ViTs) leverage the self-attention mechanism.

**Table 6**  
Summary of cardiac function quantification from TTE using deep learning.

Study	TTE View	No. of samples	Model or Technique	Hyper Parameters	Cardiac Parameters	Metrics	Value
Ouyang et al. [25]	A4C	EchoNet-Dynamic 10030 Videos	EchoNet	Opt.: Adam LF: MSE LR: 0.0001 Momt.: 0.9 BS: 16 Epochs: 45	EF	$R^2$	0.81
Ghorbani et al. [95]	A4C	3312 Videos	EchoNet	Opt.: Adam LF: SE Reg.: WD	EF EDV ESV	$R^2$	0.50 0.74 0.70
Reynaud et al. [96]	A4C	EchoNet-Dynamic 10030 Videos	Transformer	LF: MSE	EF	$R^2$	0.52
Farzy et al. [97]	A4C	EchoNet-Dynamic 10030 Videos	Hierarchical Vision Transformers	Opt.: AdamW LR: $1 \times 10^{-4}$ BS: 8 WD: $1 \times 10^{-4}$ LF: MSE Pretrain: ImageNet 22k dataset	EF	$R^2$	0.59
Muhtaseb et al. [98]	A4C	EchoNet-Dynamic 10030 Videos	EchoCoTr	Opt.: AdamW BS: 25 Epochs: 45 LR: $1 \times 10^{-4}$ Pretrain: Kinetics-400	EF	$R^2$	0.82
Mokhtari et al. [99]	A4C	EchoNet-Dynamic 10030 Videos	EchoGNN	Opt.: Adam LR: $1 \times 10^{-4}$ BS: 80 Epochs: 2500	EF	$R^2$	0.76
Lin et al. [38]	A4C A2C ALX	6953 Videos	Mathematical Model	–	EF	Mean Bias	4.3* 4.0**
Zhang et al. [37]	PLAX PSAX A2C A3C A4C	8666 Images	Mathematical Model	–	Card. Str.; EF; LV Strain	Absolute Deviation	17% 6% 1.4%
Duffy et al. [100]	A4C PLAX	23745 Videos	DeepLab V3	Opt.: Adam LR: 0.001 Epochs: 50 LF: Weighted MSE	Ventricular& Posterior Wall thickness; LV diameter	$R^2$	0.96* 0.90**
Bohoran et al. [101]	PLAX RV Inflow PSAX A4C Subcostal views	100 data points	Transformer	Epochs:500 BS:1 LR: 0.01 Opt.: Adamax	RVES RVED	$R^2$	0.97
Li et al. [81]	A2C A4C	CAMUS 500 Patients	Mathematical model	NA	EDV ESV EF	r	0.97 0.97 0.91
Lau et al. [40]	A2C A4C PLAX	54780 Videos	DROID	NA	EF EDD ESD IVS PWT LAAP	$R^2$	0.90 0.86 0.90 0.62 0.49 0.74
Tokodi et al. [102]	A4C	RVFNet 3583 Videos	Ensemble network	Opt.: Adam BS: 32 LR: 0.0003 gamma=0.1	RVEF	MAE	5.54

MSE — Mean Square Error; SE — Squared Error; EF — Ejection Fraction;  $R^2$  - Coefficient of Determination; EDV — End Diastolic Volume; ESV — End Systolic Volume; card. Str.— Cardiac Structures; LVEDD — Left ventricular end-diastolic dimension; LVESD —left ventricular end-systolic dimension; IVS — Interventricular septal wall thickness; PWT — Posterior wall thickness; LAAP — Left atrial anteroposterior dimension; RVEF — Right Ventricle Ejection Fraction.

\* Indicates model performance in internal test data

\*\* Indicate model performance in external test data.

However, unlike CNNs, ViTs lack inductive bias, which means they generally require a large amount of data for training. This data may not always be readily available, especially in medical imaging. Therefore, adopting a method that combines the strengths of both CNNs and ViTs when dealing with spatio-temporal data is highly beneficial [98].

### 3.6. Diagnosis of cardiac diseases

This section presents the DL framework used to detect cardiac diseases such as motion abnormality, hypertrophy, and cardiac valve diseases from TTE images or videos.

#### 3.6.1. Motion abnormality detection

Regional Wall Motion Abnormality (RWMA) refers to the development of unusual or absent contractility in a specific region of the heart muscle. The presence of RWMA in the ventricular muscle is often the initial indication of ischemia. LV wall motion abnormalities are closely associated with Myocardial Infarction (MI). The detection of RWMA is a diagnostic marker for identifying coronary artery disease and cardiovascular disease, which electrocardiograms or cardiac biomarkers may not capture. Early identification of patients with RWMA can benefit clinicians, enabling them to administer prompt treatment and potentially prevent severe cardiac conditions [22].

Studies [22,38,39,84,85,104–106] have employed deep CNNs to detect wall motion abnormalities. In particular, 3D CNNs and fusion methods were utilized [22,38,105,106]. A classification model based on 3D CNN considers the temporal features in an image sequence or video, unlike a 2D CNN, which only learns the spatial features from an image. Using video or image sequence data, a 3D CNN can learn spatial and temporal features, facilitating the analysis of automated raw echocardiogram videos [22]. However, a 3D CNN has a higher computational complexity than a 2D CNN as it processes temporal and spatial features. Fusion CNNs, similar to 3D CNNs, also learn from temporal data. Fusion networks, such as late fusion CNN, early fusion CNN, and slow fusion CNN, integrate data from the temporal domain. Convolutional filters in the first layer of the DL model can be extended early in the network process to perform feature fusion. Alternatively, two single-frame networks can be separated, and their outputs combined later in the processing to perform feature fusion [107].

Omar et al. [104] developed a framework that preprocesses TTE videos with an asymmetric characteristics method and feeds them into a slow fusion CNN model. Meanwhile, Kusunose et al. [105] examined the effectiveness of deep convolutional networks such as DenseNet, ResNet, Inception-ResNet, Inception, and Xception for improved RWMA detection. Among these, ResNet outperformed the others in detecting RWMA, achieving an area under the curve (AUC) of 0.97.

On top of this, Huang et al. [39] proposed a pipeline with a 3D CNN for view selection, a U-Net model for LV segmentation, and a 3D CNN for RWMA detection. Conversely, Saeed et al. [106] presented a DL methodology for MI classification in TTE videos. They utilized 3D ResNet-18 and (2+1)D ResNet-18 networks pre-trained using EchoNet-Dynamic and CAMUS datasets for EF prediction. The models achieved an MI detection accuracy of 83% on the HMC-QU dataset. Similarly, Sanjeevi et al. [22] developed the EC3D-Net, a 3D CNN for RWMA detection from TTE videos. The model achieved an AUC score of 0.82 in RWMA detection on raw TTE videos without any preprocessing steps. Additionally, they optimized the frames per second (fps) rate of TTE videos to minimize the computational complexity of a 3D CNN.

Degerli et al. [84] designed a three-phase approach for early detection of MI. This approach includes LV wall segmentation, feature extraction from the segmentation output as preprocessing stages, and MI detection. They employed supervised machine learning techniques for motion classification, with the support vector machine yielding superior results. Additionally, they made the HMC-QU dataset one of the first publicly accessible datasets for RWMA detection. To extend

this, Sun et al. [108] introduced a DL model called STGA-MS for diagnosing RWMA. This model could identify the presence of RWMA across various myocardial segments, corresponding to three unique TTE views. The model was built upon the characteristics of cardiac motion and comprised three key modules: Multiscale Downsampling Module, Segment Feature Extraction Module, and Spatial–Temporal Grouping Attention (STGA) Module.

Hamila et al. [85] proposed a method for diagnosing MI based on the RWMA. Their approach involved a SegNet that segments the LV chambers. The 3D CNN used in their study detected MI with an accuracy of 90.9%. Similarly, Lin et al. [38] developed a DL framework for the automated interpretation of echocardiograms. This framework includes view selection, segmentation, quantification of cardiac function, and RWMA detection using a 3D CNN. The dataset used in their study is a combination of standard and bedside echocardiograms. To generalize the DL framework, both types of echocardiograms were used.

Table 7 summarizes that a CNN gives state-of-the-art RWMA detection from TTE data. Complex preprocessing steps are employed before detecting RWMA [38,39,84,85,104–106]. These steps include manually collecting specific end-diastolic and end-systolic frames from TTE videos, segmenting the LV, and utilizing DL models to classify TTE viewpoints. These preprocessing steps are time-consuming and heavily reliant on human input. This makes detecting RWMA a complex and lengthy process, thereby limiting the rapid detection of cardiac diseases and the planning of treatments. Our previous work [22] developed a 3D CNN-based model that detects RWMA from raw echocardiograms without requiring complex preprocessing steps for the A4C view. The developed model's performance metrics were comparable to the state-of-the-art model.

#### 3.6.2. Hypertrophy

Cardiac hypertrophy, characterized by the thickening of the heart muscle, is the most prevalent heart disease. This condition, known as hypertrophic cardiomyopathy (HCM), can make it more challenging for the heart to pump blood [110]. In the case of cardiac amyloidosis, a type of restrictive cardiomyopathy, amyloid protein deposits replace healthy heart muscle. This condition can affect the conduction system, disrupting the transmission of electrical signals through the heart [37]. Hypertensive heart disease, resulting from chronic high blood pressure, is associated with anomalies in the LA, LV, and coronary arteries. Hypertension induces anatomical and functional changes in the myocardium, increasing the heart's workload [37].

Ouyang et al. [25] developed EchoNet to identify cardiomyopathy. Building on this, Ghorbani et al. [95] demonstrated the application of a DL model, aided by CNN, to predict systemic characteristics influencing cardiovascular risk. The EchoNet model detected enlarged LA, pacemaker electrodes, and LV hypertrophy. Meanwhile, Zhang et al. [37] utilized individual CNNs to classify pulmonary arterial hypertension, cardiac amyloidosis, and hypertrophic cardiomyopathy in their framework. Similarly, Nasimova et al. [110] developed a CNN model to classify cardiomyopathy, demonstrating an impressive accuracy of 98.2% in identifying hypertrophic cardiomyopathy and dilated cardiomyopathy.

Yu et al. [41] proposed a diagnostic network to identify cardiac dysfunction. It could automatically identify four conditions: Normal, hypertrophic cardiomyopathy, hypertensive heart disease, and cardiac amyloidosis. To improve on this, Hwang et al. [111] developed a model using CNN-LSTM to detect common causes of left ventricular hypertrophy, such as hypertrophic cardiomyopathy, hypertensive heart disease, and light-chain cardiac amyloidosis. The total diagnostic accuracy of the DL algorithm (92.3%) was significantly higher than that of echocardiography experts (80.0% and 80.6%). Similarly, Duffy et al. [100] evaluated a DL methodology for assessing ventricular hypertrophy. The DL model accurately identified cardiac hypertrophic cardiomyopathy

**Table 7**  
Diagnostic ability of deep learning models in motion abnormality detection.

Study	TTE view	No. of samples	Model	Hyper parameters	AUC/Acc.
Omer et al. [104]	A4C	35 Videos	Slow fusion CNN	WI: Xavier-improved Epochs: 60 BS: 20 LR: 0.001	85.4%
Kusunose et al. [105]	PSAX	400 Videos	ResNet	Opt: Adam LF: CE Epochs: 100	0.97
Huang et al. [39]	PLAX PSAX A2C A4C	6454 Images	3D CNN	Opt.: Adam LR: 0.0001 Plateau WD rate: 0.9 LF: Mean BCE	0.89
Degerli et al. [84]	A4C	HMC-QU 160 Videos	SVM	Kernal: Radial Basis Function Cost value: 10	85%
Hamila et al. [85]	A4C	165 Videos	3D CNN	Opt.: RMSProp Loss: BCE LR: 0.001 Epochs per fold: 100 BS: 8	90%
Lin et al. [38]	A4C A2C ALX	6953 Images	3D CNN	Opt.: SGDM Momt.: 0.9 WD: 0.0001 LR: 0.00001 LF: CE	0.88
Saeed et al. [106]	A4C A2C	HMC-QU 160 Videos	(2+1)D CNN ResNet-18	Opt.: Madgrad [109] LR: 0.00001 Epochs: 50 BS: 16 LF: CE Pretrain: CAMUS, EchoNet-Dynamic	83%
Sanjeevi et al. [22]	A4C	HMC-QU 130 Videos	EC3D-Net	Opt.: Adam LF: BCE Epochs: 100 LR: 0.00001 Decay: 0.000001	0.82
Sun et al. [108]	A2C A3C A4C	137 Videos	STGA-MS	LR: 0.001 Opt.: Adam BS: 6 Epochs: 50	0.67

AUC — Area Under Curve; WI — Weight Initialization; SGDM — Stochastic Gradient Descent Momentum

and amyloidosis. The model successfully determined the causes of hypertrophy and minor alterations in the LV wall geometric metrics.

Liu et al. [112] developed AIEchoDx, an end-to-end DL framework that classifies four prevalent cardiovascular diseases. These include atrial septal defect, hypertrophic cardiomyopathy, dilated cardiomyopathy, and prior myocardial infarction. The framework achieved impressive AUC values for each disease and employs class activation mapping to visualize the decision-making process. Li et al. [113] provided an automated DL model with a fusion architecture based on different TTE view images to extend the research. The developed fusion model classified patients with hypertrophic cardiomyopathy, cardiac amyloidosis, hypertensive heart disease, or other conditions. An Inception-ResnetV2 model was optimized using six TTE images. A meta-learner operating under a fusion architecture was trained using each model output. The fusion model achieved the AUC of HCM: 0.93, CA: 0.90, and HTN/other: 0.92, respectively.

In contrast with previous studies, Madani et al. [114] employed semi-supervised and supervised DL models to classify left ventricular hypertrophy. Remarkably, the semi-supervised GAN model with CNN

achieved an accuracy of 92.3% using only 4% of the labeled data, while the supervised model achieved an accuracy of 91.2%. To extend this, A data-efficient method for automated LVH classification has been presented by Farhad et al. [115]. Using a modified Siamese network, the authors classified LVH and normal images using zero-shot and few-shot algorithms. In contrast to conventional zero-shot learning techniques, the developed method uses a cutoff distance for classification rather than text vectors. The model achieved an 8% increase in precision for zero-shot learning and up to 11% improvement for few-shot learning approaches.

### 3.6.3. Cardiac valve diseases

Ginsberg et al. [116] developed an automated method for assessing Aortic Stenosis (AS). AS is a form of heart valve disease where the opening of the valve connecting the lower left heart chamber to the body's major artery is restricted and incomplete. This restriction reduces or restricts blood flow from the heart to the aorta and the rest of the body. The authors employed multi-task training to detect the severity of AS. The established technique yielded mean F1 scores of 96.5%

for identifying AS and 73% for classifying AS into mild, moderate, and severe categories. Similarly, Akmadi et al. [117] developed a DL framework to evaluate the feasibility of detecting and classifying the severity of AS using TTE. The developed model achieved 91.5% and 95.2% accuracy in AS detection and 78.1% and 83.8% in AS severity classification, respectively, on private and public datasets.

Building on this, Holste et al. [118] developed a pipeline to diagnose the severity of AS. The ensemble 3D model was pre-trained using self-supervised contrastive learning. The model was validated using two geographically distinct cohorts of TTE data from California and New England. Saliency maps highlighting the LA, mitral annulus, and aortic valve as the predicted areas enhanced the model's comprehensibility. In parallel, Cheng et al. [119] developed a 3D CNN to automatically detect aortic valve (AV) regurgitation from TTE videos. AV regurgitation is when the heart's aortic valve does not completely seal, causing the blood in the LV to leak backward. According to feature importance studies, the anterior leaflet tip of the mitral valve during valve opening was crucial for detecting AV regurgitation.

Yuan et al. [120] explored the prediction of the Coronary Artery Calcification (CAC) score from TTE videos. They developed a video-based CNN to predict CAC scores with spatiotemporal convolutions and residual connections across video frames. The developed CNN model was highly selective in identifying patients with no calcium. The model's performance in predicting patients with intermediate calcium scores (> 200 Agatston units) and high calcium scores (> 400 Agatston units) resulted in AUC values of 0.75 and 0.74, respectively. Similar to computed tomography CAC, TTE video-predicted CAC influenced differences in 1-year survival. To extend research, Majid et al. [121] introduced a deep transformer network called CarpNet to classify Carpentier's categorization of mitral valve (MV) disease. The detection of cardiac valve diseases from TTE is crucial. From the studies discussed above, we can infer that DL models can detect valve diseases from TTE. Future studies could explore detecting rare cardiac valve diseases using DL and work towards enhancing the performance of current models.

#### 3.6.4. Other cardiac diseases

Researchers have employed DL techniques to diagnose common cardiac conditions using TTE. This section delves into the automated diagnosis of cardiac diseases using TTE facilitated by DL. The spectrum of cardiac diseases examined includes LVEF abnormality, atrial fibrillation, septal flash, congenital heart disease, septal defect, pericardial diffusion, and heart function preserved EF.

Cheng et al. [119] developed a 3D CNN to automatically detect abnormal left ventricular function from TTE videos. They also investigated which anatomical features or temporal frames provided the most relevant data for disease classification. The 3D CNN achieved an accuracy of 86% for reduced LV function. Feature importance studies identified the mitral valve and LV myocardium as crucial for detecting LV dysfunction. Building on this, Silva et al. [122] developed a 3D CNN model to classify the level of abnormality in the LVEF. The model classified LVEF abnormality into the following categories: unhealthy (45%), intermediate (45%–55%), healthy (55%–75%), and abnormal (>75%).

To extend the research, Begnami et al. [123] proposed a dual-stream model that includes view-specific feature extraction blocks for A2C and A4C views. This model utilizes shared RNN layers to estimate LVEF abnormality directly from TTE videos, eliminating the need for LV segmentation and identification of crucial cardiac frames. LVEF classifications were determined based on the risk of developing heart failure: below 40% was considered high risk for LVEF, and above 40% was considered low risk.

Lu et al. [124] introduced a regression model to automatically detect anomalies in TTE images, distinguishing between normal and abnormal cardiomyopathy cases. In parallel, Dezaki et al. [125] proposed Echo-Rhythm Net, to automate the identification of atrial fibrillation (an irregular heartbeat that occurs when both atriums fire their

electrical signals simultaneously) based on echocardiography images without needing an electrocardiogram. The proposed framework comprises three main components: a self-supervised encoder, a temporal self-similarity matrix layer, and a supervised detector. In a test dataset, Echo-Rhythm Net detected atrial fibrillation with an accuracy rate of 79%, outperforming a trained echocardiographer who scored an atrial fibrillation detection accuracy of 63%.

Wang et al. [126] proposed an end-to-end framework to assess multi-view echocardiograms for detecting congenital heart diseases, a broad category of birth disorders that impair the heart's normal functioning. To minimize network parameters, depthwise separable convolution-based multi-channel networks were used. The authors also introduced an adaptive soft attention approach to examine raw video data directly, reducing the need for key-frame selection. Four different neural aggregation techniques were examined for combining video frame information. The developed video-based model, which does not require key-frame selection or view annotation during testing, was diagnosed with an accuracy of 92.1% in a testing set.

Qu et al. [91] introduced a Linear Attention Cascaded Network (LACNet) designed to analyze echocardiograms and automatically detect Septal Flash (SF), a condition often found in patients with a complete Left Bundle Branch Block (LBBB). LBBB is characterized by delayed or blocked electrical impulses to the heart's left side. SF, a leftward motion of the ventricular septum occurring before ejection, is commonly observed in LBBB patients. The proposed method simultaneously extracts spatial and temporal characteristics using a cascaded CNN encoder and an LSTM decoder. The model employs a Spatial Transformer Network module to eliminate image consistency and reduce data complexity using linear attention layers. The LV area-time curve, derived from segmentation results and serving as an independent disease predictor, was incorporated to diversify the input data.

To extend, Nurmaini et al. [127] developed a stacked residual-dense network to detect septal defects automatically from prenatal and post-natal TTE. The model comprises three modules: defect segmentation, classification, and decision-making. A variant of the ResNet architecture was used as the backbone for cardiac chamber segmentation. The model was capable of detecting Atrial Septal Defect (ASD), Ventricular Septal Defect (VSD), and Atrioventricular Septal Defect (AVSD). In unseen test data, the model achieved a Dice Coefficient score of 75.74% and an accuracy of 92% in septal defect detection.

Cheng et al. [128] proposed a DL framework for detecting Pericardial Effusion (PE) from TTE. The developed pipeline consists of three modules: Moving Window View Selection (MWVS), segmentation, and width computation from a segmented mask. The MWVS model, which utilized the ResNet architecture, was used to classify the TTE views. The Mask-RCNN was employed for cardiac chamber segmentation. A maximal width calculator, a computer vision approach, was then used to determine the maximum width of the PE from the segmented mask. The model achieved an AUC of 0.92 in PE detection. Akerman et al. [129] utilized a 3D CNN model to diagnose Heart Failure with Preserved Ejection Fraction (HFpEF) from TTE. The model was evaluated on TTE data from 6823 patients and achieved an AUC of 0.91.

Table 8 summarizes the performance of DL models in detecting various cardiac diseases or abnormalities from TTE. The results of the studies presented in sections 3.6.2, 3.6.3, and 3.6.4 show promise for the automated and early diagnosis of various cardiac diseases from TTE. Fully automated DL-based diagnostic support systems hold substantial potential to provide cost-effective and high-quality healthcare in resource-limited settings.

Key research findings:

1. Most of the research work has been focused on diagnosing cardiac diseases such as wall motion abnormality, hypertrophy, and some common cardiac diseases, including cardiac valve disease. Different research groups have attempted to diagnose motion abnormality and cardiomyopathy, achieving state-of-the-art performance. Abnormalities related to the aortic and mitral valves were also explored.

**Table 8**  
Diagnostic ability of deep learning models to detect cardiac diseases.

Study	TTE View	No. of samples	Model	Hyper parameters	Disease	Metrics	Value
Ouyang et al. [25]	A4C	EchoNet-Dynamic 10030 Videos	EchoNet	Opt.: Adam LF: MSE LR: 0.0001 Momt.:0.9 BS: 16 Epochs: 45	Cardiomyopathy	AUC	0.96%
Ghorbani et al. [96]	A4C	3312 Videos	EchoNet	Opt.: Adam LF: CE Reg.: WD	Pacemaker electrodes; Enlarged LA; LVH	AUC	0.890 0.860 0.750
Zhang et al. [37]	PLAX A2C A3C A4C	6793 Images	VGG	Opt.: Adam LR: 0.00005 mini BS: 64 Epochs: 20 WD: $1 \times 10^{-8}$	CA PAH CHD	AUC	0.93 0.85 0.87
Madani et al. [114]	A4C	455 Videos	CNN+GAN	Opt.: Adam LR: 0.0003	LVH	Acc.	92.3%
Nasimova et al. [110]	A4C	3632 Videos	CNN	Opt.: Adam LR:0. 0001 BS: 64 Reg.: L2 (0. 001) Epsilon value: 0. 001	HCM DCM	Acc.	98.2%
Yu et al. [41]	A4C PLAX	1610 Videos	ResNet	Opt.: Adam LF: CE Reg.: L2	HHD HCM CA	AUC	0.88 0.90 0.94
Hwang et al. [111]	PLAX PSAX A3C A4C A2C	930 Videos	CNN-LSTM	Opt.: RMSprop LF: BCE WI: 'He' Init. Lr: $1 \times 5e^{-5}$ Epochs: 50/view	HHD HCM CA	AUC	0.962 0.982 0.996
Liu et al. [112]	A4C	1807 Videos	AIEchoDx	Opt.: SGD LR: 0.001 Momt: 0.9 LF: CE Pretrain: ImageNet	ASD DCM HCM Prior MI	AUC	0.99 0.98 0.99 0.98
Duffy et al. [100]	A4C PLAX	23745 Videos	3D-ResNet	Opt.: Adam LF: BCE LR: 0.01. Epochs: 100 BS: 14	CA HCM	AUC	0.79 0.89
Ginsberg et al. [116]	PLAX PLSX	2247 Videos	Deep Residual Network	Opt.: Adam LR:0.0001 Epochs: 15 BS: 8 LF: CE or Gaussian	Aortic Stenosis	F1 Score	96.5%
Yuan et al. [120]	PLAX	2881 Videos	CNN	Opt.: Adam Pretrain: EchoNet-Dynamic LR: 0.001 BS: 10 Epochs: 40 LF: Squared loss between the prediction and log (CAC score + 1)	CAC score	AUC	0.81
Cheng et al. [119]	A4C	3554 Videos	3D CNN	Opt.: Adam LF: CE LR: 0.001 BS: 16 Epochs: 50	Impaired LV function; AV Regurgitation	Acc.	86% 83%

(continued on next page)

Table 8 (continued).

Study	TTE View	No. of samples	Model	Hyper parameters	Disease	Metrics	Value
Silva et al. [122]	A4C	30 Videos	3D CNN	LR: $5 \times 10^{-6}$ BS: 64 Epochs: 20000 steps Reg.: L2	EF Abnormality	Acc.	78%
Begnami et al. [123]	A2C A4C	1186 Videos	CNN+RNN	NA	EF Abnormality	Acc	83.1%
Lu et al. [124]	A4C A2C PLAX	927 Videos	Regression Model	Opt.: Adam Nesterov Momentum Reg.: L2 LF: MSE	Cardiomyopathy	AUC	0.84
Dezaki et al. [125]	PLAX	3947 Videos	Echo-Rhythm Net	Opt.: Adam LF: BCE LR: 0.001 Epochs: 40	AF	Acc.	79%
Wang et al. [126]	PLAX PSAX A4C SXLAX SSLAX	1308 Videos	CNN	Opt.: Adam Epochs: 120 Reg.: L2 LF: CE	Congenital Heart disease (ASD, VSD)	Acc.	92.1%
Qu et al. [91]	A4C	238 Videos	LACNet	NA	Septal Flash	AUC	0.95
Farhad et al. [115]	A2C A4C	LVH Echo 72 images	Cardiac Siamese	Epochs:30 BS:2 Opt: Adam LF: Contrastive loss	LVH	Precision	0.91
Li et al. [113]	PLAX A2C A3C A4C PSAX	187521 Images	Fusion model	BS: 32 LR: 0.000001 WD: 0.3 LF: CE Epochs: 50	HCM CA HTN/Other	AUC	0.93 0.90 0.92
Majid et al. [121]	PLAX	1773 Videos	CarpNet	NA	Mitral valve Carpentier functional	Acc.	0.71
Akmadi et al. [117]	PLAX PSAX A4C A2C	TMED-2 577 Videos	Transformer	Attention Head: 8 Opt.: Adam LR: 0.0001 LR schedule: Cosine Annealing Epochs: 100	AS	Acc.	94.5%
Holste et al. [118]	PLAX	32742 Videos	3D-Resnet18	Opt.: Adam LR: 0.0001 BS: 88 LF: Sigmoid CE	AS	AUC	0.97
Nurmaini et al. [127]	PLAX PLSX A4C A5C Subcostal	1526 Videos	Stacked residual dense network	NA	Septal defect	Acc	92%
Cheng et al. [128]	PLAX PSAX A4C SC	6434 Images	Maximal width calculator	NA	PE	AUC	0.92
Akerman et al. [129]	A4C	7321 Videos	3D CNN	Opt.: Adam LF: CE LR: 0.0003	HFpEF	AUC	0.92

LVH — Left Ventricle Hypertrophy; PAH — Pulmonary Arterial Hypertension; CHD — Hypertrophic Cardiomyopathy; HHD — Hypertensive Heart Disease; HCM — Hypertrophic Cardiomyopathy; DCM — Dilated Cardiomyopathy; ALCA — Light-Chain Cardiac Amyloidosis; CA — Cardiac Amyloidosis; ASD — Atrial Septal Defect; AV — Aortic Valve; CAC — Coronary Artery Calcification; AF — Atrial Fibrillation; VSD — Ventricular Septal Defect; SXLAX — Subxiphoid long-axis view of two atria; SSLAX — Suprasternal long-axis view of the aortic arch; PE — Pericardial effusion; HFpEF — Heart failure with preserved ejection fraction.

2. Open research challenges exist in utilizing DL algorithms to diagnose other rare cardiac diseases such as pericardial diseases (acute pericarditis, cardiac tamponade, constrictive pericarditis), peripheral diseases, and congenital heart diseases (Ebstein anomaly, Tetralogy of Fallot (TOF)) from TTE.

3. Automated diagnosis of rare cardiac diseases from TTE presents the following challenges and opportunities:

- Complexity of collecting large datasets to train the DL models and the time-consuming nature of data annotation.

- Recent studies [114,115] have used semi-supervised and few-shot learning models in hypertrophy diagnosis. Researchers may explore these different learning algorithms, such as unsupervised and few-shot learning methods, to build decision support systems in low-data settings. Specifically, the few-shot learning method aims to generalize the model with a low number of samples in the dataset [56].
- Generative Adversarial Network models were used to augment and generate data for LV segmentation training models. However, there is a need to explore fine-grained image-generative models and methods for generating image data to train models to diagnose rare cardiac diseases.

#### 4. Dataset

This section details the publicly available datasets for TTE analysis, such as for LV segmentation, cardiac function quantification, and cardiac disease detection. Most studies used the CAMUS dataset to segment LV and quantify cardiac function. Some studies used the HMC-QU and EchoNet-Dynamic datasets for MI detection and quantification of cardiac function, respectively. A considerable number of studies also used their private dataset to analyze TTE. The following describes these benchmark datasets for automated TTE analysis.

##### 4.1. HMC-QU

The Qatar University and Hamad Medical Corporation Hospital collaborated to generate the HMC-QU benchmark dataset for MI detection. The dataset consists of 160 A4C TTE videos collected between 2018 and 2019. The patient's TTE is taken within 24 h of hospital admission or before coronary angioplasty. Normal (non-MI) patients underwent a necessary health check. The cardiologists from HMC Hospital created the labels for MI detection. Ultrasound machines made by Phillips and GE Vivid (GE-Healthcare-USA) are used to acquire TTEs. Each video has a spatial resolution ranging from  $422 \times 636$  to  $768 \times 1024$  pixels and a temporal resolution of 30 frames per second. Also, the dataset provides LV segmentation masks for cardiac cycle frames in TTE videos [84,130].

##### 4.2. CAMUS

The CAMUS dataset includes TTE from 500 patients obtained at the University Hospital of St. Etienne (France). The dataset has three TTEs: good, medium, and low. The dataset was used to segment the LV and LA and quantify the EF from the TTE. EchoPAC analysis program exported 2D A4C and A2C sequences for each patient. These TTE views were chosen to allow the calculation of EF values using Simpson's biplane method [24].

##### 4.3. Echonet-dynamic

The dataset contains 10,030 TTE A4C videos from patients who had imaging procedures at Stanford University Hospital from 2016 to 2018 as part of routine clinical care. Text and other content beyond the scanning region were removed from each video by cropping and masking it. The produced images were downsampled into  $112 \times 112$  pixel videos using cubic interpolation. The dataset was utilized to calculate the EF, segment the LV, and evaluate cardiomyopathy [25].

##### 4.4. EchoNet-LVH

The EchoNet-LVH dataset, comprising 12,000 annotated TTE PLAX videos, serves as a foundational resource for investigating heart chamber dimensions and wall thickness. These videos are labeled by human experts who provide detailed annotations, including measurements, tracings, and computations. The dataset incorporates ground truth values for the thickness of the left ventricular posterior wall (LVPW), the left ventricular internal dimension (LVID), the intraventricular septum (IVS), and LVH classification [40,100].

##### 4.5. TMED-2

The TMED-2 dataset incorporates TTE scans conducted at the Tufts Medical Centre from 2011 to 2020. The dataset classifies the severity of Aortic Stenosis (AS) into four categories: no AS, mild AS, mild-to-moderate early AS, and major AS (moderate to severe). The dataset encompasses PLAX, PSAX, A4C, and A2C images, with each patient contributing between 50 and 100 images. The dataset is divided into three subsets: The fully labeled set includes images from 577 patients, each annotated with both image-level view labels and patient-level AS severity. The view-only labeled set comprises images from 703 patients, each labeled only with the view. The unlabeled set contains images from 5287 patients, none labeled with either view or severity [117].

##### 4.6. RVENet

The primary objective of the RVENet dataset is to aid in developing and testing DL for predicting RVEF using TTE videos. The RVENet collection comprises 3583 A4C view TTE videos in DICOM format, derived from 944 examinations of 831 patients. Each patient underwent one or more 3D TTE at the Heart and Vascular Centre of Semmelweis University between November 2013 and March 2021 [102] (see Table 9).

#### 5. Discussion

Despite the superior imaging quality of MRI and CT, echocardiograms remain a prevalent tool for diagnosing cardiac disorders due to their availability, portability, convenience, and cost-effectiveness compared to other modalities. These attributes make echocardiograms particularly useful in low-resource settings. However, the manual interpretation of an echocardiogram is subjective and can vary among clinicians. In these scenarios, automated diagnostic support systems offer significant potential to reduce subjectivity and provide cost-effective, high-quality healthcare, especially in resource-limited environments.

The TTE is the most commonly used type of echocardiogram in clinical decision-making [12]. Comprehensive review results indicate that TTE research is progressing towards automated analysis. Over the past six years, there has been rapid advancement in adapting DL models for automating various tasks such as view classification, noise reduction, EF quantification, heart chamber segmentation, and detecting different cardiac diseases. Fig. 8 illustrates the publication timeline.

The earlier studies [92,93], conducted between 2018 and 2021 involved complex steps in cardiac parameter quantification, such as identifying frames, manually segmenting the LV, and using mathematical equations such as Simpson biplane or American Society of Echocardiography recommendation for computation. More recent studies [25, 95–99] from 2022 to 2023 have automatically predicted parameters from TTE videos using the DL framework. For detecting motion abnormalities from TTE, earlier studies required complex preprocessing steps such as LV segmentation, manual collection of specific cardiac cycle frames from TTE videos, and DL models to classify TTE viewpoints [38, 39,84,85,105,106]. However, recent studies [22] have automatically detected motion abnormalities from TTE videos.

The A4C view of TTE was utilized in 95% of the shortlisted articles. The A4C TTE image or video allows for simultaneous examination of the atrial septum, mitral and tricuspid valves, interventricular septum, ventricles, and atrium. The A4C view is one of the best viewpoints for determining the overall and regional contractility of the LV. Moreover, it provides segmentation results for the heart's four chambers, enhancing the quantification of cardiac function and improving the prognosis for heart failure.

Diagnosing cardiac disorders from TTE requires years of experience. DL frameworks have empowered clinicians to detect these disorders

**Table 9**  
Summary of publicly available TTE dataset.

Dataset	No. of Samples	Location	System & Data	Ground truth	Study
EchoNet-Dynamic	10030 Videos	Stanford University Hospital	Acuson SC2000 Siemens Epiq 5G, Epiq 7, Philips & A4C with 30 FPS	EF, ESV, EDV, Segmentation mask for ES and ED frames	[25,61,72,73,76,77,96–99,106]
CAMUS	500 Patients	University Hospital of St. Etienne (France)	GE M5S probe from GE Vivid E95 ultrasound scanners (GE Vingmed Ultrasound, Horten Norway) (GE Healthcare, US) & A4C and A2C	Segmentation masks for ES and ED frames	[24,24,58,62,64,65,77,106]
HMC-QU	160 Videos	The Hamad Medical Corporation Hospital and Qatar University	Ultrasound Machine by Phillips and GE Vivid (GE-Healthcare-USA)	Myocardial Infarction Annotation, and LV wall mask far cardiac cycle frames	[17,22,84,106]
EchoNet-LVH	12000 Videos	Stanford University Hospital	Acuson SC2000 Siemens Epiq 5G, Epiq 7, Philips	LVH annotation	[40,100]
TMED-2	577 Videos	Tufts Medical Center	Toshiba, Siemens, Philips	Aortic stenosis annotation	[117]
RVENet	3583 Videos	Heart and Vascular Center of Semmelweis University (Budapest, Hungary)	Vivid E95 system, GE Vingmed Ultrasound, Horten, Norway; iE33, EPIQ CVx, 7C, or 7G systems, Philips, Best, The Netherlands)	RV dysfunction	[102]

more easily. Existing methods primarily focus on segmenting and evaluating the LV chamber, which is crucial for blood flow and diagnosing cardiac diseases. Fully automated approaches in the literature are specific to a particular TTE view and are based on certain assumptions, such as chamber shape or view. These methods may fail if these assumptions are not met. The LA, Right Atrium (RA), and RV chambers have received less attention due to their complex shapes and unclear boundaries. In [69], the authors segmented all four cardiac chambers. Analyzing these chambers enhances the diagnosis and provides a complete assessment of cardiac chamber dysfunction. This leads to a rise in applying novel frameworks by AI research groups to analyze the TTE data. Future studies focus on developing fully automated DL support systems that can assess the complex structures of the RA, LA, and RV chambers in addition to LV assessment. Clinical cardiac parameters collected from these chambers are crucial for identifying cardiac diseases such as hypertrophy, hypertension, and chamber enlargement.

The accuracy of outcomes using computer vision has improved due to increased computational power in recent years. Although the application of DL in TTE has seen significant advancements, there are several potential areas for further research. Future studies could utilize fine-grained transfer learning models trained on large public datasets like CAMUS and EchoNet-Dynamic. The studies presented in this review explored the supervised learning method, which primarily depends on the clinician's annotation. The introduction of unsupervised learning methods in TTE analysis could identify different patterns and biomarkers of TTE for decision-making. Further research in applying reinforcement learning methods to TTE should also be explored. The advantage of using reinforcement learning in analyzing unlabeled medical datasets is the ability to develop meaningful visualizations

for transfer learning. Deep Reinforcement Learning (DRL) adds the representative power of deep neural networks to the reinforcement learning framework. The application of DLR in other imaging techniques has shown great potential [131]. We observe that there are open challenges in the application of unsupervised and reinforcement learning in automated TTE analysis.

The clinical decisions made by the DL model need to be explained to clinicians to understand the model's workings. Most of the current DL models create a prediction based on extracted features from the images, but they do not provide a feature explanation for the detected output. Decisions made without transparency and comprehensibility may not be reliable (black-box models). Only a limited number of studies [22,30,43,95,112] have used explainability methods such as Grad-Cam [132], activation mapping [133], and SmoothGrad [134] techniques to observe the DL models' decision-making. Another advanced explainability method is Integrated Gradients [131], which has the advantage of satisfying the sensitivity axiom over other methods. It determines the output's gradient about the feature map. Therefore, future research should focus on incorporating the explainability method into their DL-based support system for clinical decision-making.

Most published papers on the automated analysis of TTE primarily discuss pilot applications without conducting in-depth validation on datasets collected using various ultrasound equipment models from different brands. Future research should validate these models on multicenter datasets, i.e., datasets from diverse demographics, to enhance the generalizability of DL models [93,135]. Additionally, there is a need to focus on predicting patient prognosis from imaging data by incorporating clinical patient information. There are also open research areas in constructing models that optimize time, power, and memory

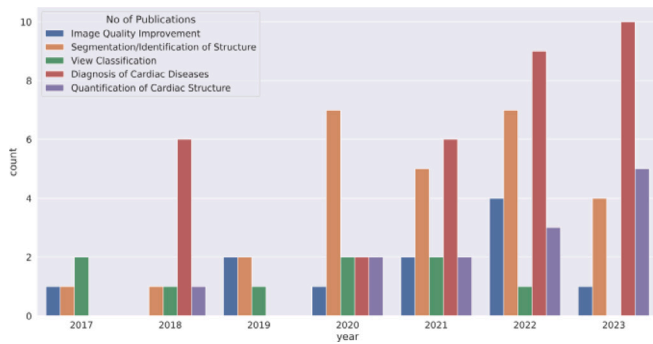


Fig. 8. The publication timeline with the number of articles.

to develop lightweight models suitable for low-resource settings, such as edge-level portable echocardiogram devices. Cardiologists and radiologists play a pivotal role in developing AI-enabled cardiac decision support systems. As they are the end-users, physician-in-the-loop systems should be guided by substantial clinical inputs from them, which can lead to potential improvements in cardiac assessment.

In summary, this review presents both the state-of-art and open research challenges in automated Transthoracic Echocardiogram analysis using Deep Learning. Specifically, for translational clinical tools to gain better acceptance, it is essential that the following open challenges are addressed and opportunities utilized for further enhancement of this domain:

1. **Dataset scarcity:** There is a lack of publicly available datasets for exploring a broad spectrum of view classification and segmentation of cardiac structures beyond the left ventricle.
2. **TTE Resolution Improvement:** DL can be explored to enhance spatial or temporal resolution in specific disease conditions. For patients with high heart rates, the frame interpolation can be utilized by DL to compensate for missing events. DL models can use the super-resolution method to improve spatial resolution for patients with thick chest walls.
3. **End-to-End Cardiac Analysis:** The automated TTE analysis, including cardiac chambers and valves, in a single clinical tool would be highly beneficial. This will combine the benefits of multiple specialized analyses into a single framework.
4. **Unsupervised and Reinforcement Learning:** Exploring these learning methods can reveal new patterns or biomarkers.
5. **Transfer Learning Model:** Studies can leverage fine-grained transfer learning weights from large cardiac datasets such as CAMUS, Echonet-Dynamic, and Echonet-LVH.
6. **Model Validation:** DL models should be validated on multi-center data to enhance their generalizability.
7. **Patient clinical data:** In addition to TTE, the patient's clinical data, such as demography and vital parameter data, could be utilized for a better prognosis of cardiac condition.
8. **Lightweight Model Development:** The creation of lightweight models can optimize power and time in edge-level echocardiogram devices, especially with the introduction of portable systems.
9. **Diagnosis of Rare Cardiac Diseases:** DL-aided diagnosis of rare cardiac diseases such as acute pericarditis, cardiac tamponade, constrictive pericarditis, peripheral diseases, and specific congenital heart diseases are relatively less explored research areas.

## 6. Conclusion

This paper presents a comprehensive systematic review of research advancements in applying deep learning tools for automating TTE analysis. Significant improvements have been observed over the past six years in the following areas:

- TTE view classification
- TTE image quality improvement
- Segmentation or identification of cardiac structure
- Cardiac parameter quantification for clinical assessment
- Diagnosis of cardiac diseases, including wall motion abnormality, hypertrophy, and valve diseases.

We have summarized publicly accessible TTE datasets, identified limitations in existing studies, and suggested potential future research approaches. The significant limitations of current studies include the specificity of automated methods to certain assumptions, a primary focus on LV chamber analysis, a lack of transparency and interpretability in the developed models, and a scarcity of large public datasets and validation on multi-center data.

Future studies should aim to enhance the efficiency of current models for detecting and measuring cardiac functions, classify heart diseases with consideration given to memory, time, and power consumption, and address the limitations in the publicly available large datasets. To ensure the production of sufficiently generalizable results, studies should utilize multi-center data and attempt to incorporate patients' clinical variables into automated decision-support systems.

Significantly, there are numerous opportunities for research in this field. Applying unsupervised and reinforcement learning models to interpret echocardiograms will open up novel possibilities for future studies. Therefore, researchers striving to improve DL performance on automated echocardiogram analysis should focus on these areas to achieve scalable translational tools to assist clinicians in their diagnoses.

## CRedit authorship contribution statement

**Sanjeevi G.:** Writing – review & editing, Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Uma Gopalakrishnan:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization. **Rahul Krishnan Parthinarupothi:** Project administration, Investigation, Formal analysis, Conceptualization, Validation, Writing – review & editing. **Thushara Madathil:** Conceptualization, Methodology, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] Gahungu N, Trueick R, Bhat S, Sengupta PP, Dwivedi G. Current challenges and recent updates in artificial intelligence and echocardiography. *Curr Cardiovasc Imag Rep* 2020;13:1–12.
- [2] Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med* 2020;7(25).
- [3] Durga P, Rangan Ekanath, Pathinarupothi Rahul Krishnan. Real-time identification & alert of ischemic events in high risk cardiac patients. In: 2016 IEEE international conference on computational intelligence and computing research. IEEE; 2016, p. 1–5.
- [4] Reddy KV, Kumar N. Automated prediction of sudden cardiac death using statistically extracted features from electrocardiogram signals. *Int J Electr Comput Eng* (2088-8708) 2022;12(5).
- [5] Reddy L, Thangam S. Predicting relapse of the myocardial infarction in hospitalized patients. In: 2022 3rd international conference for emerging technology. IEEE; 2022, p. 1–7.
- [6] Pravin V, Srinivasan N, Rohith P, Arvind UV, Vijayan D. Automatic identification of heart abnormalities using PCG signals. In: International conference on computer, communication, and signal processing. Cham: Springer; 2022, p. 314–24.
- [7] Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. *Echo Res Pract* 2018;5(4):R115–25.

- [8] Sehly A, Jaltotage B, He A, Maiorana A, Ihdayhid AR, Rajwani A, et al. Artificial intelligence in echocardiography: The time is now. *Rev Cardiovascul Med* 2022;23(8):256.
- [9] Zamzmi G, Hsu LY, Li W, Sachdev V, Antani S. Harnessing machine intelligence in automatic echocardiogram analysis: Current status, limitations, and future directions. *IEEE Rev Biomed Eng* 2020;14:181–203.
- [10] Bizopoulos P, Koutsouris D. Deep learning in cardiology. *IEEE Rev Biomed Eng* 2018;12:168–93.
- [11] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [12] de Siqueira VS, de Castro Rodrigues D, Dourado CN, Borges MM, Furtado RG, Delfino HP, et al. Machine learning applied to support medical decision in transthoracic echocardiogram exams: a systematic review. In: 2020 IEEE 44th annual computers, software, and applications conference. IEEE.; 2020, p. 400–7.
- [13] de Siqueira VS, Borges MM, Furtado RG, Dourado CN, da Costa RM. Artificial intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: A systematic review. *Artif Intell Med* 2021;120:102165.
- [14] Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC: Cardiovascul Imag* 2019;12(8 Part 1):1549–65.
- [15] Karatzia L, Aung N, Aksentijevic D. Artificial intelligence in cardiology: Hope for the future and power for the present. *Front Cardiovascul Med* 2022;9.
- [16] Cai A, Hu W, Zheng J. Few-shot learning for medical image classification. In: Artificial neural networks and machine learning–ICANN 2020: 29th international conference on artificial neural networks, Bratislava, Slovakia, September (2020) 15–18, Proceedings, Part I. Vol. 29. Springer International Publishing; 2020, p. 441–52.
- [17] Sanjeevi G, Pathinarupothi RK, Uma G, Madathil T. Deep learning pipeline for echocardiogram noise reduction. In: 2022 IEEE 7th international conference for convergence in technology. IEEE; 2022, p. 1–6.
- [18] Liao Z, Jafari MH, Girgis H, Gin K, Rohling R, Abolmaesumi P, et al. Echocardiography view classification using quality transfer star generative adversarial networks. In: Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October (2019) 13–17, Proceedings, Part II. Vol. 22. Springer International Publishing.; 2019, p. 687–95.
- [19] Ting J, Punithakumar K, Ray N. Multiview 3-d echocardiography image fusion with mutual information neural estimation. In: 2020 IEEE international conference on bioinformatics and biomedicine. IEEE.; 2020, p. 765–71.
- [20] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021;88:105906.
- [21] Cohen's kappa. 2023, [Online]. Available: [https://en.wikipedia.org/wiki/Cohen%27s\\_kappa](https://en.wikipedia.org/wiki/Cohen%27s_kappa). [Accessed 5 January 2023].
- [22] Sanjeevi G, Gopalakrishnan U, Pathinarupothi RK, Madathil T. Automatic diagnostic tool for detection of regional wall motion abnormality from echocardiogram. *J Med Syst* 2023;47(1):13.
- [23] Diller GP, Lammers AE, Babu-Narayan S, Li W, Radke RM, Baumgartner H, et al. Denoising and artefact removal for transthoracic echocardiographic imaging in congenital heart disease: utility of diagnosis specific deep learning algorithms. *Int J Cardiovascul Imag* 2019;35:2189–96.
- [24] Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging* 2019;38(9):2198–210.
- [25] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;580(7802):252–6.
- [26] Mohamed AA, Arifi AA, Omran A. The basics of echocardiography. *J Saudi Heart Assoc* 2010;22(2):71–6.
- [27] Balaji GN, Subashini TS, Chidambaram N. Automatic classification of cardiac views in echocardiogram using histogram and statistical features. *Procedia Comput Sci* 2015;46:1569–76.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October (2015) 5–9, Proceedings, Part III. Vol. 18. Springer International Publishing; 2015, p. 234–41.
- [29] Gao X, Li W, Loomes M, Wang L. A fused deep learning architecture for viewpoint classification of echocardiography. *Inf Fusion* 2017;36:103–13.
- [30] Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 2018;1(1):6.
- [31] Kusunose K, Haga A, Inoue M, Fukuda D, Yamada H, Sata M. Clinically feasible and accurate view classification of echocardiographic images using deep learning. *Biomolecules* 2020;10(5):665.
- [32] Howard JP, Tan J, Shun-Shin MJ, Mahdi D, Nowbar AN, Arnold AD, et al. Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography. *J Med Artif Intell* 2020;3.
- [33] Gu AN, Luong C, Jafari MH, Van Woudenberg N, Girgis H, Abolmaesumi P, et al. Efficient echocardiogram view classification with sampling-free uncertainty estimation. In: Simplifying medical ultrasound: second international workshop, ASMUS 2021, held in conjunction with MICCAI 2021, Strasbourg, France, September 27 2021, Proceedings. Vol. 2. Springer International Publishing; 2021, p. 139–48.
- [34] Gao Y, Zhu Y, Liu B, Hu Y, Yu G, Guo Y. Automated recognition of ultrasound cardiac views based on deep learning with graph constraint. *Diagnostics* 2021;11(7):1177.
- [35] Huang M, Lin WC, Chen YD, Hsiao TA, Liu PY, Tsai WC. Explainable deep neural network for echocardiography view classification. *European Heart Journal-Cardiovascular Imaging* 2022;23(Supplement1). jeab289-012.
- [36] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: Advances in neural information processing systems, vol. 30, 2017.
- [37] Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation* 2018;138(16):1623–35.
- [38] Lin X, Yang F, Chen Y, Chen X, Wang W, Chen X, et al. Echocardiography-based AI detection of regional wall motion abnormalities and quantification of cardiac function in myocardial infarction. *Front Cardiovascul Med* 2022;9.
- [39] Huang MS, Wang CS, Chiang JH, Liu PY, Tsai WC. Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography. *Circulation* 2020;142(16):1510–20.
- [40] Lau ES, Di Achille P, Kopparapu K, Andrews CT, Singh P, Reeder C, et al. Deep learning-enabled assessment of left heart structure and function predicts cardiovascular outcomes. *J Am Coll Cardiol* 2023;82(20):1936–48.
- [41] Yu X, Yao X, Wu B, Zhou H, Xia S, Su W, et al. Using deep learning method to identify left ventricular hypertrophy on echocardiography. *Int J Cardiovascul Imag* 2021;1–11.
- [42] Understanding latent space in machine learning. 2023, [Online]. Available: <https://towardsdatascience.com/understanding-latentspace-in-machine-learning>. [Accessed 5 January 2023].
- [43] Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. *IEEE Trans Med Imaging* 2017;36(6):1221–30.
- [44] Labs RB, Vrettos A, Loo J, Zolgharni M. Automated assessment of transthoracic echocardiogram image quality using deep neural networks. *Intell Med* 2022.
- [45] Jalali M, Behnam H, Shojaeifard M. Echocardiography image enhancement using texture-cartoon separation. *Comput Biol Med* 2021;134:104535.
- [46] Bevi A, Ratheesha S, Kalady S, Chackola JJ. Denoising transthoracic echocardiographic images in regional wall motion abnormality using deep learning techniques. *Soft Comput* 2023;1–17.
- [47] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 8789–97.
- [48] Tiago C, Gilbert A, Beela AS, Aase SA, Snare SR, Šprem J, et al. A data augmentation pipeline to generate synthetic labeled datasets of 3D echocardiography images using a GAN. *IEEE Access* 2022;10:98803–15.
- [49] Monkam P, Jin S, Tang B, Zhou X, Lu W. A disentanglement and fusion data augmentation approach for echocardiography segmentation. In: 2022 IEEE international ultrasonics symposium. IEEE; 2022, p. 1–4.
- [50] Leclerc S, Smistad E, Grenier T, Lartizien C, Ostvik A, Cervenansky F, et al. RU-Net: A refining segmentation network for 2D echocardiography. In: 2019 IEEE international ultrasonics symposium. IEEE.; 2019, p. 1160–3.
- [51] Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol* 2021;6(6):624–32.
- [52] Díez J, Luaces O, del Coz JJ, Bahamonde A. Optimizing different loss functions in multilabel classifications. *Progr Artif Intell* 2015;3:107–18.
- [53] Everingham M, Eslami SMA, Van Gool L, et al. The pascal visual object classes challenge: A retrospective. *Int J Comput Vis* 2015;111:98–136. <http://dx.doi.org/10.1007/s11263-014-0733-5>.
- [54] Gulrajani I, et al. Improved training of wasserstein gans. In: NIPS. 2017, p. 5767–77.
- [55] van Amersfoort J, Shi W, Acosta A, Massa F, Totz J, Wang Z, et al. Frame interpolation with multi-scale deep loss functions and generative adversarial networks. 2017, arXiv preprint arXiv:1711.06045.
- [56] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 2015;38(2):295–307.
- [57] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(12):2481–95.
- [58] Zyuzin V, Mukhtarov A, Neustroev D, Chumarnaya T. Segmentation of 2D echocardiography images using residual blocks in U-Net architectures. In: 2020 Ural symposium on biomedical engineering, radioelectronics and information technology. Yekaterinburg, Russia; 2020, p. 499–502. <http://dx.doi.org/10.1109/USBREIT48449.2020.9117678>.

- [59] Smistad E, Østvik A. 2D left ventricle segmentation using deep learning. In: 2017 IEEE international ultrasonics symposium. IEEE; 2017, p. 1–4.
- [60] Jafari MH, Girgis H, Liao Z, Behnami D, Abdi A, Vaseli H, et al. A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20 2018, Proceedings. Vol. 4. Springer International Publishing; 2018, p. 29–37.
- [61] Lin Z, Tsui PH, Zeng Y, Bin G, Wu S, Zhou Z. CLA-U-Net: Convolutional long-short-term-memory attention-gated U-Net for automatic segmentation of the left ventricle in 2-D echocardiograms. In: 2022 IEEE international ultrasonics symposium. IEEE; 2022, p. 1–4.
- [62] Amer A, Ye X, Zolgharni M, Janan F. ResDUNet: Residual dilated UNet for left ventricle segmentation from echocardiographic images. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society. Montreal, QC, Canada; 2020, p. 2019–22. <http://dx.doi.org/10.1109/EMBC44109.2020.9175436>.
- [63] Bergstra JS, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: Proc. adv. neural inf. process. syst. 2011, p. 2546–54.
- [64] Leclerc S, et al. LU-Net: A multitask attention network to improve the robustness of segmentation of left ventricular structures in 2-D echocardiography. IEEE Trans Ultrasonics, Ferroelectr, Freq Control 2020;67(12):2519–30. <http://dx.doi.org/10.1109/TUFFC.2020.3003403>.
- [65] Ali Y, Janabi-Sharifi F, Beheshti S. Echocardiographic image segmentation using deep Res-U network. Biomed Signal Process Control 2021;64:102248.
- [66] Azarmehr N, Ye X, Sacchi S, Howard JP, Francis DP, Zolgharni M. Segmentation of left ventricle in 2D echocardiography using deep learning. In: Medical image understanding and analysis: 23rd conference, MIA 2019, Liverpool, UK, July (2019) 24–26, Proceedings. Vol. 23. Springer International Publishing; 2020, p. 497–504.
- [67] Zhu X, Wei Y, Lu Y, Zhao M, Yang K, Wu S, et al. Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging. Comput Methods Programs Biomed 2021;199:105914.
- [68] Kim S, Park HB, Jeon J, Arsanjani R, Heo R, Lee SE, et al. Fully automated quantification of cardiac chamber and function assessment in 2-D echocardiography: clinical feasibility of deep learning-based algorithms. Int J Cardiovasc Imag 2022;38(5):1047–59.
- [69] Arafati A, Morisawa D, Avendi MR, Amini MR, Assadi RA, Jafarkhani H, et al. Generalizable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. J R Soc Interface 2020;17(169):20200267.
- [70] Thi THD, Minh TN, Van PN, Tran QL. Fully automated machine learning pipeline for echocardiogram segmentation. In: 2021 13th international conference on knowledge and systems engineering. IEEE; 2021, p. 1–6.
- [71] Ibtihaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw 2020;121:74–87.
- [72] Azizi F, Sani AF, Priambodo R, Karunianto WC, Ramadhan MML, Rachmadi MF, et al. Modified MultiResUNet for left ventricle segmentation from echocardiographic images. In: 7th international workshop on big data and information security. Institute of Electrical and Electronics Engineers Inc; 2022, p. 33–8.
- [73] Chen E, Cai Z, Lai JH. Weakly supervised semantic segmentation of echocardiography videos via multi-level features selection. In: Pattern recognition and computer vision: 5th chinese conference, PRCV 2022, Shenzhen, China, November (2022) 4–7, Proceedings, Part II. Cham: Springer Nature Switzerland; 2022, p. 388–400.
- [74] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Trans Pattern Anal Mach Intell 2017;40(4):834–48.
- [75] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 1251–8.
- [76] Zeng Y, Tsui PH, Pang K, Bin G, Li J, Lv K, et al. MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography. Ultrasonics 2023;127:106855.
- [77] Liu F, Wang K, Liu D, Yang X, Tian J. Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography. Med Image Anal 2021;67:101873.
- [78] Su C, Zhou Y, Ma J, Chi H, Jing X, Jiao J, et al. JANet: A joint attention network for balancing accuracy and speed in left ventricular ultrasound video segmentation. Comput Biol Med 2024;169:107856.
- [79] Monkam P, Jin S, Lu W. Multi-task learning framework for echocardiography segmentation. In: 2022 IEEE international ultrasonics symposium. IEEE; 2022, p. 1–3.
- [80] Huang H, Ge Z, Wang H, Wu J, Hu C, Li N, et al. Segmentation of echocardiography based on deep learning model. Electronics 2022;11(11):1714.
- [81] Li H, Wang Y, Qu M, Cao P, Feng C, Yang J. EchoEFNet: Multi-task deep learning network for automatic calculation of left ventricular ejection fraction in 2D echocardiography. Comput Biol Med 2023;156:106705.
- [82] Liao M, Lian Y, Yao Y, Chen L, Gao F, Xu L, et al. Left ventricle segmentation in echocardiography with transformer. Diagnostics 2023;13(14):2365.
- [83] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in neural information processing systems, vol. 30, 2017.
- [84] Degerli A, Zabihi M, Kiranyaz S, Hamid T, Mazhar R, Hamila R, et al. Early detection of myocardial infarction in low-quality echocardiography. IEEE Access 2021;9:34442–53.
- [85] Hamila O, Ramanna S, Henry CJ, Kiranyaz S, Hamila R, Mazhar R, et al. Fully automated 2D and 3D convolutional neural networks pipeline for video segmentation and myocardial infarction detection in echocardiography. Multimedia Tools Appl 2022;81(26):37417–39.
- [86] Lai KW, Shoaib MA, Chuah JH, Nizar MHA, Anis S, Ching SLW. Aortic valve segmentation using deep learning. In: 2020 IEEE-EMBS conference on biomedical engineering and sciences. IEEE; 2021, p. 528–32.
- [87] Rodríguez IM, Mantecón T, Fernández-Cooke E, Grasa C, Barrios A, Toral B, et al. Coronary artery segmentation on echocardiograms for Kawasaki disease diagnosis. In: 2022 E-health and bioengineering conference. IEEE; 2022, p. 1–4.
- [88] Nizar MHA, Chan CK, Khalil A, Yusof AKM, Lai KW. Real-time detection of aortic valve in echocardiography using convolutional neural networks. Curr Med Imag 2020;16(5):584–91.
- [89] Chandra V, Sarkar PG, Singh V. Mitral valve leaflet tracking in echocardiography using custom Yolo3. Procedia Comput Sci 2020;171:820–8.
- [90] Loshchilov I, Hutter F. Decoupled weight decay regularization. 2017, [Online]. Available: <https://arxiv.org/abs/1711.05101>.
- [91] Qu M, Wang Y, Li H, Yang J, Ma C. Automatic identification of septal flash phenomenon in patients with complete left bundle branch block. Med Image Anal 2022;82:102619.
- [92] Dezaki FT, Liao Z, Luong C, Girgis H, Dhungel N, Abdi AH, et al. Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss. IEEE Trans Med Imaging 2018;38(8):1821–32.
- [93] Lane ES, Azarmehr N, Jevsikov J, Howard JP, Shun-Shin MJ, Cole GD, et al. Multibeam echocardiographic phase detection using deep neural networks. Comput Biol Med 2021;133:104373.
- [94] Farhad M, Masud MM, Beg A, Ahmed LA, Memon S. Cardiac phase detection in echocardiography using convolutional neural networks. Sci Rep 2023;13(1):8908.
- [95] Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. NPJ Digit Med 2020;3(1):10.
- [96] Reynaud H, Vlontzos A, Hou B, Beqiri A, Leeson P, Kainz B. Ultrasound video transformers for cardiac ejection fraction estimation. In: Medical image computing and computer assisted intervention—MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part VI. Vol. 24. Springer International Publishing; 2021, p. 495–505.
- [97] Fazly L, Haryono A, Nissa NK, Hirzi NM, Rachmadi MF, Jatmiko W. Hierarchical vision transformers for cardiac ejection fraction estimation. In: 7th international workshop on big data and information security. Institute of Electrical and Electronics Engineers Inc; 2022, p. 39–44.
- [98] Muhtaseb R, Yaqub M. EchoCoTr: Estimation of the left ventricular ejection fraction from spatiotemporal echocardiography. In: Medical image computing and computer assisted intervention—MICCAI 2022: 25th international conference, Singapore, September (2022) 18–22, Proceedings, Part IV. Cham: Springer Nature Switzerland; 2022, p. 370–9.
- [99] Mokhtari M, Tsang T, Abolmaesumi P, Liao R. EchoGNN: Explainable ejection fraction estimation with graph neural networks. In: Medical image computing and computer assisted intervention—MICCAI 2022: 25th international conference, Singapore, September (2022) 18–22, Proceedings, Part IV. Cham: Springer Nature Switzerland; 2022, p. 360–9.
- [100] Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, et al. High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. JAMA Cardiol 2022;7(4):386–95.
- [101] Bohoran TA, Kampaktis PN, McLaughlin L, Leb J, Moustakidis S, McCann GP, et al. Right ventricular volume prediction by feature tokenizer transformer-based regression of 2D echocardiography small-scale tabular data. In: International conference on functional imaging and modeling of the heart. Cham: Springer Nature Switzerland; 2023, p. 292–300.
- [102] Tokodi M, Magyar B, Soos A, Takeuchi M, Tolvaj M, Lakatos BK, et al. Deep learning-based prediction of right ventricular ejection fraction using 2D echocardiograms. JACC: Cardiovasc Imag 2023.
- [103] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, 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. Eur Heart J-Cardiovasc Imaging 2015;16(3):233–71.
- [104] Omar HA, Patra A, Domingos JS, Leeson P, Noble AJ. Automated myocardial wall motion classification using handcrafted features vs a deep CNN-based mapping. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society. IEEE; 2018, p. 3140–3.
- [105] Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. Cardiovasc Imag 2020;13:374–81.

- [106] Saeed M, Yaqub M. End-to-end myocardial infarction classification from echocardiographic scans. In: Simplifying medical ultrasound: third international workshop, ASMUS 2022, held in conjunction with MICCAI 2022, Singapore, September 18 2022, Proceedings. Cham: Springer International Publishing; 2022, p. 54–63.
- [107] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 1725–32.
- [108] Sun S, Wang Y, Yu Q, Qu M, Li H, Yang J. STGA-MS: AI diagnosis model of regional wall motion abnormality based on 2D transthoracic echocardiography. *Heliyon* 2024;10(1).
- [109] Defazio A, Jelassi S. Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization. 2021, arXiv preprint arXiv:2101.11075.
- [110] Nasimova N, Muminov B, Nasimov R, Abdurashidova K, Abdullaev M. Comparative analysis of the results of algorithms for dilated cardiomyopathy and hypertrophic cardiomyopathy using deep learning. In: 2021 International conference on information science and communications technologies. IEEE; 2021, p. 1–5.
- [111] Hwang IC, Choi D, Choi YJ, Ju L, Kim M, Hong JE, et al. Differential diagnosis of common etiologies of left ventricular hypertrophy using a hybrid CNN-LSTM model. *Sci Rep* 2022;12(1):20998.
- [112] Liu B, Chang H, Yang D, Yang F, Wang Q, Deng Y, et al. A deep learning framework assisted echocardiography with diagnosis, lesion localization, phenogrouping heterogeneous disease, and anomaly detection. *Sci Rep* 2023;13(1):3.
- [113] Li J, Chao CJ, Jeong JJ, Farina JM, Seri AR, Barry T, et al. Developing an echocardiography-based, automatic deep learning framework for the differentiation of increased left ventricular wall thickness etiologies. *J Imag* 2023;9(2):48.
- [114] Madani A, Ong JR, Tibrewal A, Mofrad MR. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit Med* 2018;1(1):59.
- [115] Farhad M, Masud MM, Beg A, Ahmad A, Ahmed LA, Memon S. A data-efficient zero-shot and few-shot Siamese approach for automated diagnosis of left ventricular hypertrophy. *Comput Biol Med* 2023;163:107129.
- [116] Ginsberg T, Tal RE, Tsang M, Macdonald C, Dezaki FT, van der Kuur J, et al. Deep video networks for automatic assessment of aortic stenosis in echocardiography. In: Simplifying medical ultrasound: second international workshop, ASMUS 2021, held in conjunction with MICCAI 2021, Strasbourg, France, September 27 2021, Proceedings. Vol. 2. Springer International Publishing; 2021, p. 202–10.
- [117] Ahmadi N, Tsang MY, Gu AN, Tsang TSM, Abolmaesumi P. Transformer-based spatio-temporal analysis for classification of aortic stenosis severity from echocardiography cine series. *IEEE Trans Med Imag* 2023.
- [118] Holste G, Oikonomou EK, Mortazavi BJ, Coppi A, Faridi KF, Miller EJ, et al. Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur Heart J* 2023;44(43):4592–604.
- [119] Cheng LH, Bosch PB, Hofman RF, Brakenhoff TB, Bruggemans EF, van der Geest RJ, et al. Revealing unforeseen diagnostic image features with deep learning by detecting cardiovascular diseases from apical 4-Chamber ultrasounds. *J Am Heart Assoc* 2022;11(16):e024168.
- [120] Yuan N, Kwan AC, Duffy G, Theurer J, Chen JH, Nieman K, et al. Prediction of coronary artery calcium using deep learning of echocardiograms. *J Am Soc Echocardiogr* 2022.
- [121] Vafaezadeh M, Behnam H, Hosseinsabet A, Gifani P. CarpNet: Transformer for mitral valve disease classification in echocardiographic videos. *Int J Imag Syst Technol* 2023.
- [122] Silva JF, Silva JM, Guerra A, Matos S, Costa C. Ejection fraction classification in transthoracic echocardiography using a deep learning approach. In: 2018 IEEE 31st international symposium on computer-based medical systems. IEEE; 2018, p. 123–8.
- [123] Behnami D, Luong C, Vaseli H, Abdi A, Girgis H, Hawley D, et al. Automatic detection of patients with a high risk of systolic cardiac failure in echocardiography. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20 2018, Proceedings. Vol. 4. Springer International Publishing; 2018, p. 65–73.
- [124] Lu A, Dehghan E, Veni G, Moradi M, Syeda-Mahmood T. Detecting anomalies from echocardiography using multi-view regression of clinical measurements. In: 2018 IEEE 15th international symposium on biomedical imaging. Washington, DC, USA; 2018, p. 1504–8. <http://dx.doi.org/10.1109/ISBI.2018.8363858>.
- [125] Dezaki FT, Ginsberg T, Luong C, Vaseli H, Rohling R, Gin K, et al. Echo-rhythm net: Semi-supervised learning for automatic detection of atrial fibrillation in echocardiography. In: 2021 IEEE 18th international symposium on biomedical imaging. IEEE; 2021, p. 110–3.
- [126] Wang J, Liu X, Wang F, Zheng L, Gao F, Zhang H, et al. Automated interpretation of congenital heart disease from multi-view echocardiograms. *Med Image Anal* 2021;69:101942.
- [127] Nurmainsi S, Sapitri AI, Tutuko B, Rachmatullah MN, Rini DP, Darmawahyuni A, et al. Automatic echocardiographic anomalies interpretation using a stacked residual-dense network model. *BMC Bioinform* 2023;24(1):365.
- [128] Cheng CY, Wu CC, Chen HC, Hung CH, Chen TY, Lin CHR, et al. Development and validation of a deep learning pipeline to measure pericardial effusion in echocardiography. *Front Cardiovascul Med* 2023;10.
- [129] Akerman AP, Porumb M, Scott CG, Beqiri A, Chartsias A, Ryu AJ, et al. Automated echocardiographic detection of heart failure with preserved ejection fraction using artificial intelligence. *JACC: Adv* 2023;2(6):100452.
- [130] Kiranyaz S, Degerli A, Hamid T, Mazhar R, Ahmed REF, Abouhasera R, et al. Left ventricular wall motion estimation by active polynomials for acute myocardial infarction detection. *IEEE Access* 2020;8:210301–17.
- [131] Zhou SK, Le HN, Luu K, Nguyen HV, Ayache N. Deep reinforcement learning in medical imaging: A literature review. *Med Image Anal* 2021;73:102193.
- [132] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 618–26.
- [133] Jung H, Oh Y. Towards better explanations of class activation mapping. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 1336–44.
- [134] Smilkov D, Thorat N, Kim B, Viégas F, Wattenberg M. Smoothgrad: removing noise by adding noise. 2017, arXiv preprint arXiv:1706.03825.
- [135] Tromp J, Seekings PJ, Hung CL, Iversen MB, Frost MJ, Ouwkerk W, et al. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health* 2022;4(1):e46–54.