Deep Learning-Based Automated Diagnosis of Pericardial Diseases: A Multi-View Approach

Jeong, Sihyeon

Department of Medical Science

Graduate School

Yonsei University

Deep Learning-Based Automated Diagnosis of Pericardial Diseases: A Multi-View Approach

Advisor Chang Hyuk-Jae

A Master's Thesis Submitted

to the Department of Medical Science
and the Committee on Graduate School
of Yonsei University in Partial Fulfillment of the
Requirements for the Degree of
Master of Medical Engineering

Jeong, Sihyeon

June 2025

Deep Learning-Based Automated Diagnosis of Pericardial Diseases: A Multi-View Approach

This Certifies that the Master's Thesis of Jeong, Sihyeon is Approved

Committee Chair	Jung, In Hyun
Committee Member	Chang, Hyuk-Jae
Committee Member	Yoon, Yeonyee E

Department of Medical Science
Graduate School
Yonsei University
June 2025

ACKNOWLEDGEMENTS

2년이 넘는 학위 과정을 마칠 수 있기까지 부족한 저를 도와주신 모든 분들에게 감사의 말씀을 전합니다. 먼저, 지도교수님이신 장혁재 교수님께 깊은 감사의 말씀을 드립니다. 교수님의 가르침 아래에서 많은 것을 배우고 성장할 수 있었습니다.

항상 지도편달에 힘써 주시며, 연구에 있어 아낌없는 조언을 해주신 윤연이 교수님과 제 석사 학위 심사에 시간을 할애해주신 정인현 교수님께도 감사의 말씀을 전합니다.

연구에 있어 저에게 많은 가르침을 주신 장영걸, 이지나 박사님께도 감사드립니다. 함께 연구하고 있는 동료들 (가은언니, 현석오빠, 재익오빠, 준흠님, 양기님, 태영님)과 졸업 후 새로운 일을 하고 있는 선배들(경훈오빠, 주영오빠) 그리고 특히, 힘들 때 의지가 많이 된 지연, 다운언니에게도 감사의 인사를 전합니다. 임상 연구에 있어서 많은 조언과 도움을 주신 이승아 교수님께도 진심으로 감사드립니다.

마지막으로, 제가 가장 사랑하는, 또 저를 언제나 사랑해주신 아빠, 저를 지지해주시는 엄마에게 항상 감사드리며, 우리 가족 저를 언제나 응원해주시고 사랑해 주셔서 감사합니다. 이제 새걸음을 떼는 동생 시우에게도 격려의 말을 전합니다.

정말 많은 분들 덕분에 석사 과정을 마칠 수 있었습니다. 받은 사랑과 은혜를 잊지 않고, 베풀 수 있는 사람이 되도록 노력하겠습니다. 감사합니다.

정시현 올림.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Deep learning-based automated diagnosis of pericardial diseases: A multi-view approach

Purpose – Pericardial effusion, tamponade, and constrictive pericarditis remain hard to diagnose because echocardiographic image quality and interpretation vary. We developed an automated deep-learning system that fuses standard views to raise accuracy and reproducibility.

Methods – A multi-view modified Resnet convolutional architecture ingests five routinely acquired gray-scale views—parasternal long-axis (PLAX), parasternal short-axis (PSAX), apical four-chamber (A4C), modified A4C, and subcostal four-chamber (S4C)—and augments them with inferior vena cava (IVC) cine loops and Doppler spectrograms (mitral inflow and septal-annulus TDI) to capture hemodynamic significance. Masked early fusion and uncertainty-weighted multi-task loss balance. Training and external validation were performed on 2,118 transthoracic studies collected from multiple institutions.

Result – Multi-view fusion substantially boosted performance. For effusion-severity classification, the proposed model achieved an AUC of 0.933 (versus 0.901 for the single-view baseline). Sensitivity for detecting tamponade-level hemodynamic compromise increased from 0.387 to 0.618 at comparable specificity. Comprehensive metrics (accuracy, sensitivity, specificity, AUC) for each disease category are summarized in Tables 2–3.

Conclusion – Integrating complementary anatomical and hemodynamic cues across views enhances automated detection of effusion severity, pericardial thickening/adhesion, and tamponade. The system offers a reliable tool to support clinical decision-making and expedite intervention in life-threatening conditions. Future work will address cases with sub-optimal image quality and extend the model to beat-to-beat functional analysis.

Key words: deeplearning, echocardiography, multiview, pericardial disease, multi task

1. Introduction

1.1. Clinical Significance of Pericardial Diseases and the Need for Advanced Diagnostics

Pericardial diseases encompass a range of disorders affecting the pericardium, the double-layered sac that surrounds the heart. These conditions include pericardial effusion, constrictive pericarditis, effusive constriction, and cardiac tamponade. Each disorder can compromise cardiac function by altering the heart's ability to fill properly during diastole, leading to symptoms that range from mild chest discomfort and shortness of breath to life-threatening hemodynamic instability.

Clinically, pericardial diseases are significant not only because they may signal underlying systemic or inflammatory conditions but also due to their potential to rapidly deteriorate patient status if not diagnosed and managed promptly. 1 The clinical presentation is often subtle or nonspecific, with patients experiencing fatigue, dyspnea, or atypical chest pain. In some cases, the physical examination may not provide clear clues, thereby necessitating reliance on imaging modalities for accurate diagnosis. Conventional methods such as transthoracic echocardiography, while widely used due to their portability and cost-effectiveness, suffer from operator dependency and may be limited by poor acoustic windows. Although computed tomography (CT) and cardiovascular magnetic resonance (CMR) offer detailed anatomical and tissue characterization, their accessibility and real-time application in emergency settings are often constrained. 2

Against this backdrop, the integration of deep learning techniques into the diagnostic workflow offers a promising solution. Deep learning models have demonstrated the ability to process large volumes of imaging data with high accuracy and consistency, effectively reducing the variability inherent in human interpretation. These automated systems can rapidly quantify cardiac parameters, detect subtle morphological changes, and differentiate between types of pericardial pathology with minimal human intervention. By providing objective assessments, deep learning diagnostics can serve as an invaluable decision support tool, enabling clinicians to make faster and more accurate diagnoses.

In essence, the clinical need for rapid and precise evaluation of pericardial diseases drives the exploration of advanced diagnostic tools. AI-driven approaches not only promise to enhance the accuracy of parameter measurements and disease classification but also have the potential to streamline the workflow in high-pressure clinical environments. Such advancements are crucial for improving patient outcomes, optimizing treatment strategies, and ultimately reducing the burden on healthcare systems by facilitating timely intervention in cases of pericardial disease.

1.2. Review of Deep Learning for Echocardiographic Diagnosis Cardiac Diseases

Deep learning approaches have demonstrated remarkable potential in revolutionizing echocardiographic diagnosis through automated analysis of cardiac structures and function. This section reviews current progress in applying deep learning method to cardiac imaging, with particular focus on methodologies relevant to pericardial disease classification.

1.2.1. Advances in Multimodal Cardiac Imaging Analysis

Recent studies have established foundational frameworks for cardiac pathology detection using various imaging modalities. Ouyang et al. developed EchoNet-Pericardium3, employing temporal-spatial CNNs to grade pericardial effusion severity (AUC 0.900–0.955) and detect cardiac tamponade across five standard views. Their ensemble approach for multiview integration analyzed 1.4M videos, achieving robust generalizability in external validation (AUC 0.966 for tamponade detection). Similarly, their EchoNet-LVH4 model utilized 3D-CNNs to predict left ventricular hypertrophy etiology with high accuracy (AUC 0.83–0.98) by integrating beat-to-beat analysis of parasternal long-axis videos.

Zhang et al. 5 demonstrated the efficacy of CNNs in processing multiple standard views (2.6M frames) for comprehensive view classification and pathology detection. The EchoCLR6 framework further advanced this field by implementing self-supervised learning (SSL) techniques that enable efficient feature extraction with minimal labeled data, achieving ASD detection with AUC 0.93 using only 50 labeled videos.

PanEcho7, developed by Yale's Cardiovascular Data Science Lab, represents a paradigm shift in multimodal echocardiographic analysis. This view-agnostic, multi-task deep learning model processes 1.23 million echocardiographic videos across parasternal, apical, and subcostal views to simultaneously perform 39 diagnostic tasks, including structural and functional assessments of chambers, valves, and vascular dimensions. Its architecture combines that a spatiotemporal image encoder for frame-level feature extraction, temporal transformer for sequence modeling, and task-specific output heads for simultaneous classification and regression.

PanEcho achieves median AUCs of 0.91 across 18 classification tasks and normalized mean absolute error of 0.13 for 21 measurement tasks, including LV ejection fraction estimation (MAE 4.4%). The model's multi-view agnosticism enables dynamic weighting of diagnostically relevant views, outperforming single-view approaches in external validation across geographically distinct cohorts. As an open-source foundation model, PanEcho demonstrates efficient transfer learning capabilities, including classification of pericardial effusion amount, suggesting broad applicability to specialized clinical populations.

1.2.2. Technical Approaches and Architectures

Early pipelines relied on 2-D/3-D convolutional backbones; state-of-the-art systems combine spatiotemporal encoders with attention or transformer blocks to capture long-range motion patterns.

Fully convolutional networks remain core to pixel-wise segmentation, especially in cardiac MRI, and analogous U-Net variants are now common in echo. Current architectural approaches span from traditional CNNs to more sophisticated models. For segmentation tasks, Fully Convolutional Networks (FCNs) have proven effective in cardiac MRI analysis, enabling pixel-level classification for precise delineation of cardiac structures[7]. The Stanford research group pioneered video-based DL algorithms like EchoNet-Dynamic that analyze cardiac motion patterns to evaluate ejection fraction and detect arrhythmias[8].

1.2.3. Challenges in Pericardial Disease Classification

Despite significant progress, several challenges persist:

- Data Scarcity: Most studies focus on common conditions like LVH or valvular diseases, with limited pericardial-specific datasets[10].
- Temporal Modeling Complexity: Transformers show promise for capturing subtle temporal dependencies crucial for detecting transient constriction. [5]
- Multiview Integration: Ensemble approaches and GRU-based fusion architectures offer promising frameworks for view integration[11].

2. Method

2.1. Proposed Deep Learning-Based Pericardial Disease Diagnosis Method Overview

Our diagnostic framework leverages multi-view echocardiographic data to address the inherent challenges in pericardial disease diagnosis. Pericardial diseases can manifest as structural abnormalities—such as pericardial effusion, thickening, and adhesion—and may have critical hemodynamic consequences. To capture these complex characteristics, our approach utilizes multiple standard echocardiographic views, including Parasternal Long Axis (PLAX), Parasternal Short Axis (PSAX) Apical 4 Chamber (A4C), Modified Apical 4 Chamber, and Subcostal 4 Chamber (S4C). [1] Additionally, the Inferior Vena Cava (IVC) view and Doppler images, septal annulus TDI and MV inflow PW, are employed to assess dynamic features related to hemodynamic significance.

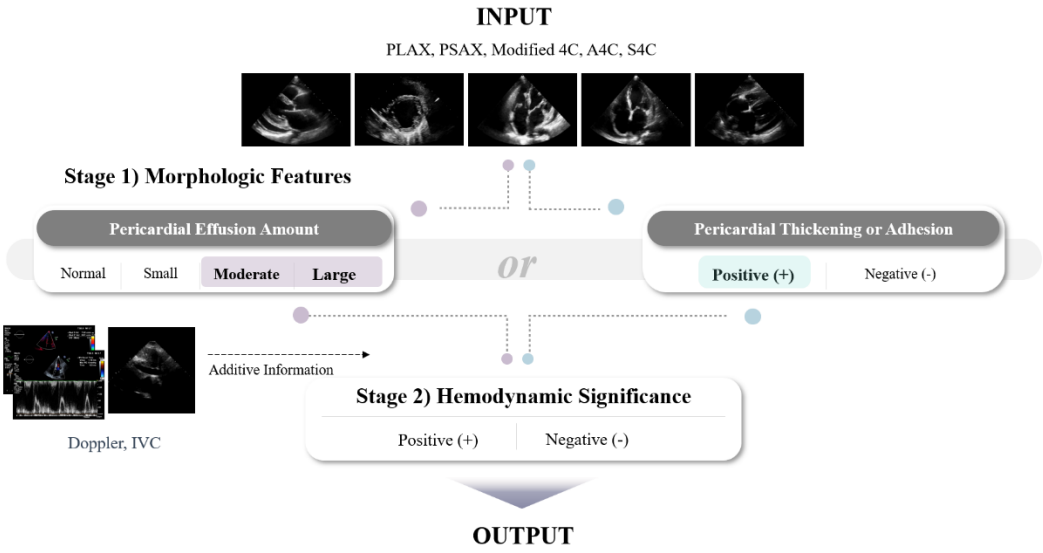

Figure 1. Overall Hierarchical Classification Process.

The overall method operates in two main stages. In the first stage, the model focuses on detecting morphological abnormalities by analyzing each view independently using a shared deep learning backbone. Feature extraction from each view is performed simultaneously, and the resulting representations are fused using a dedicated multi-view fusion module. In cases where one or more views are unavailable, a zero-padded tensor is inserted; these padded inputs are then masked out during training so that they do not contribute to the parameter updates. This ensures that only valid data drives the learning process.

In the second stage, for cases flagged with morphological abnormalities, additional dynamic information from the IVC view and doppler view is incorporated to evaluate hemodynamic significance. An IVC clip is processed by a small U-Net to segment the vessel lumen frame-by-frame and to compute diameter and collapsibility ratios—surrogate markers of right-atrial pressure. In parallel, still frames of the septal-annulus TDI spectrogram and the MV inflow PW are passed through a shallow CNN to produce compact Doppler feature vectors that capture S′, E′, A′, and E/A flow characteristics. These hemodynamic features are concatenated with the stage-one embedding and routed to a final decision head that determines whether the case exhibits tamponade physiology or other clinically significant hemodynamic compromise.

By integrating multiple echocardiographic views, our system capitalizes on the complementary information provided by different imaging angles. This multi-view approach improves diagnostic accuracy by offering a holistic assessment of both structural and functional characteristics of the pericardium. Furthermore, our framework is designed to operate end-to-end, from feature extraction through multi-view fusion to final classification, enabling rapid and reproducible automated diagnosis. Overall workflow is shown Figure 1.

Table 1. Pericardial disease decision-mapping logic

Pericardial Effusion	Pericardial Thickening/Adhesion	Hemodynamic Significance	conclusion
-	-	-	Normal Pericardium
+	-	-	Pericardial Effusion
+	-	+	Tamponade
+	+	-	Pericarditis with Effusion, but without Hemodynamic Significance
+	+	+	Effusive Constrictive Pericarditis
-	+	-	Pericarditis without Hemodynamic Significance

2.2. Model Architecture

The backbone of our diagnostic system is based on the R(2+1)D-18 architecture[10], which has been chosen for its proven ability to capture both spatial and temporal features from video data. The framework is framework is a multi-stream extension of R(2+1)D-18 that couples view-wise morphology analysis with an auxiliary encoder for hemodynamic signals (IVC and Doppler). The entire pipeline remains end-to-end trainable, but its internal flow is now organized into four successive blocks: (1) shared backbone, (2) multi-view advanced module, (3) extra-information encoder, and (4)fusion head (Figure 2).

2.2.1. Shared Spatio-temporal Backbone

In our multi-view setup, each available echocardiographic view (PLAX, PSAX, A4C, Modified A4C, S4C) is processed independently through the shared R(2+1)D backbone. The extracted feature vectors from each view are subsequently concatenated along the feature dimension by the network. Each clip is forwarded—independently for every available view—through the checkpointed R(2 + 1)D-18 backbone (four residual stages, final width = 512). Dropout is applied to the individual feature vectors before concatenation to help prevent overfitting. A 1×1×1 3-D bottleneck followed by BN + ReLU standardises the feature scale, after which global average pooling yields one 512-D vector fi per view.

2.2.2. Multi-view Advanced Module

When a particular echocardiographic view is missing for a given case, our preprocessing pipeline substitutes the missing input with a zero-padded tensor of the same dimensions as a valid view. Crucially, a corresponding view mask is generated and used during training to exclude these padded inputs from contributing to the feature aggregation and subsequent parameter updates. The set of vectors $\{f_i\}_{i=1}^{V}$ branches into two parallel paths:

- **View-specific heads.** Lightweight MLPs (one per view, shared weights across studies) produce *view logits* that supply fine-grained feedback during training.

- **Masked fusion.** Valid view vectors are concatenated into a single tensor $F \in \mathbb{R}^{BX(V_{valid}\cdot 512)}$. Zero-tensors replace missing windows, and a binary mask ensures that padded views contribute neither to forward activations nor to gradients. This operation yields a **fused morphology embedding** that summaries the spatial-temporal context of all recorded windows.

This strategy ensures that only reliable, real data influences the model's learning process, while preserving the overall structure of the multi-view input.

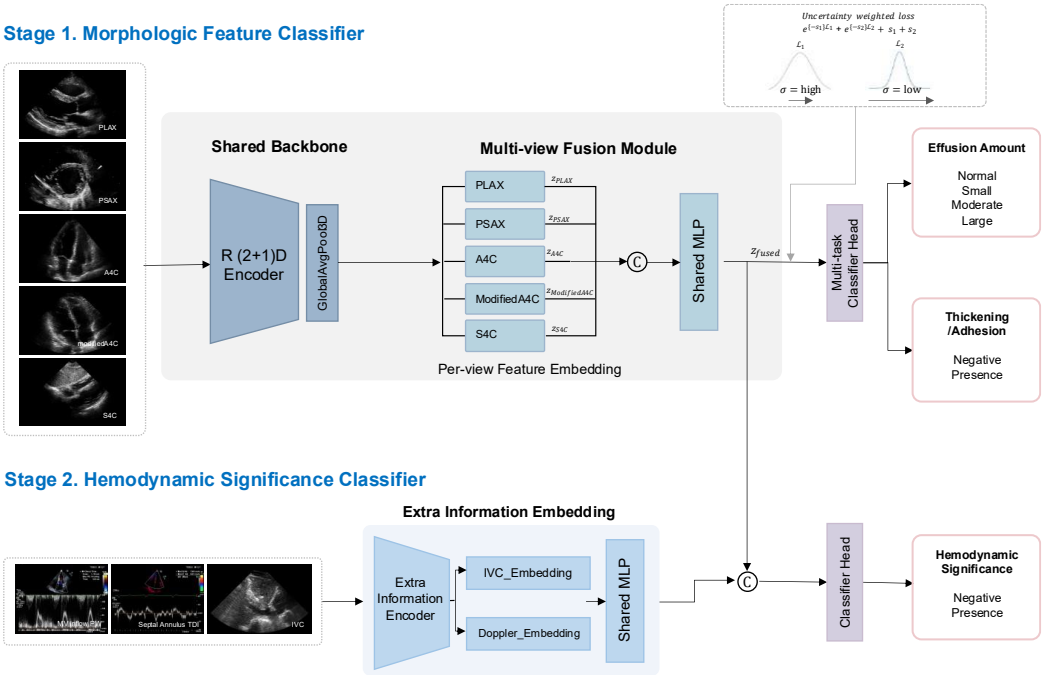

Figure 2. Model Architecture. Two-stage deep learning framework for pericardial disease assessment.

2.2.3. Integration of Hemodynamic Features

For cases where morphological abnormalities are detected, additional hemodynamic information is integrated using additional information. To inject functional information, two heterogeneous signals are embedded into a fixed-length vector E. IVC view is processed through a dedicated segmentation or analysis module to quantify critical parameters such as IVC dilatation and collapsibility. The quantitative outputs from the IVC analysis are combined with the multi-view morphological features to produce a final diagnosis that reflects both structural and dynamic aspects of pericardial disease. IVC categorical flags (dilatation, plethora) are shifted to the range 0–2 and passed through a learnable lookup table, producing a 2d-D embedding (default d=32). Doppler indices (septal-annulus TDI S′ E′ A′, MV inflow PW E A) arrive as numeric values plus validity

masks. Each value–mask pair is stacked and transformed by a shared two-layer MLP, generating a 5d-D representation.

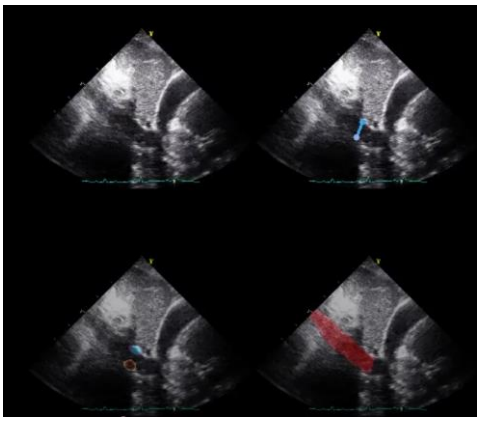

Figure 3. Visualization of IVC segmentation Result.

2.2.4. Wrapper-fusion Head and Multi-task Output

The morphology embedding FFF and extra vector EEE are concatenated $[F\,;\,E]$ and fed to a two-layer fully connected head (ReLU + Dropout) that emits task-specific logits. In our implementation a proposed model separate heads for:

- Effusion grade (four classes)
- Thickening / adhesion (binary)
- Hemodynamic significance (binary)

Overall, our architecture is designed to robustly integrate multi-view echocardiographic data using a modified R(2+1)D backbone[10], ensuring that both spatial and temporal information are captured. This architecture retains the computational efficiency of R(2 + 1)D-18, yet gains three critical capabilities: (i) explicit handling of missing views through masked fusion, (ii) seamless integration of IVC preload markers and Doppler flow signatures, and (iii) hierarchical supervision that balances per-view specificity with study-level consensus. Together, these upgrades make the model well-suited to real-world echocardiography, where view completeness and hemodynamic data quality naturally vary from patient to patient.

2.3. Learning Algorithm

To jointly optimize effusion grading (multi-class classification) and pericardial thickening detection (binary classification), we designed a multi-task learning framework with dynamically weighted loss components. The architecture processes echocardiography video inputs through shared feature extractors followed by task-specific heads.

2.3.1. Label Smoothing Cross Entropy for Effusion Grading

For effusion amount classification (4 classes: normal or trivial, small, moderate, large), we employ label smoothing cross entropy[11,12] to prevent overconfidence in predictions. Given input logits $z \in \mathbb{R}^{B\times 4}$ and target labels $y \in \{0,1,2,3\}.^{B}$, the loss is formulated as:

$$\mathcal{L}_{\text{eff}} = \frac{1}{B}\sum_{i=1}^{B}\left[(1-\epsilon)\cdot(-\log\, p_{i,y_i}) + \epsilon\cdot\left(-\frac{1}{\mathrm{K}}\sum_{c=1}^{\mathrm{K}}\log\, p_{i,c}\right)\right]$$

Where $p_{i,c} = softmax(z_i)_c$, $\epsilon = 0.1$ and $B =$ batch size. This formulation smooths hard labels by redistributing 10% probability mass uniformly across non-target classes.

2.3.2. Focal Loss for Thickening Detection

For binary classification of pericardial thickening/adhesion, we use a modified focal loss to address class imbalance.

$$FL(p_t) = -\alpha(1-p_t)^{\gamma}\log(p_t)$$

Where p_t is the predicted probability of the target class, α and γ are the weighting and focusing parameters, respectively, and $\alpha = 0.25$ and $\gamma = 2.0$ are used in this study. The loss function is computed using one-hot encoded targets and is denoted by $\mathcal{L}_{\text{thick}}$.

2.3.3. Dynamic Loss Weighting

We adopt uncertainty-based weighting[14] to automatically balance task contributions:

$$L = \exp(-\log\sigma_1^2)\cdot\mathcal{L}_{\text{eff}} + \log\sigma_1^2 + \exp(-\log\sigma_2^2)\cdot\mathcal{L}_{\text{thick}} + \log\sigma_2^2$$

To optimize both losses simultaneously, we introduce the learnable parameters $\log\sigma_1^2$ and $\log\sigma_2^2$.

This approach allows the relative importance of each loss term to be automatically adjusted during training and contributes to stable training of the model.

2.3.4. Selecting an optimization algorithm and setting hyperparameters

The model is trained using the Adam[15] optimization algorithm. Adam follows the following update equation:

$$\theta_{t+1} = \theta_t - \eta\frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}}$$

Where θ_t is the parameter at time t, η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ re the first and second moment estimates, respectively, and ε is a small constant for numerical stability. In this study, we used the following default parameters: $\eta = 10^{-4}$, $\beta_1 = 0.9$ $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$.

The batch size, number of epochs, etc. are adjusted according to the characteristics of the data and the convergence of the model, and in our initial experiments, we set the batch size to 16 and trained for about 50-100 epochs. We also applied learning rate scheduling techniques (e.g., step decay or cosine annealing) to drive fast convergence at the beginning of training and fine-tune it later.

As such, the design of the loss function, the choice of optimization algorithm, and the tuning of hyperparameters are key components that enable the model to effectively learn the two diagnostic tasks and accurately classify different clinical aspects of pericardial disease.

2.4. Evaluation Metrics and Validation

To comprehensively assess the performance of our deep learning diagnostic system for pericardial diseases, we utilize a suite of evaluation metrics that capture both overall performance and the model's ability to correctly identify positive and negative cases. The key metrics include:

• Accuracy:

This metric measures the proportion of correct predictions across all classes. It is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

• Sensitivity (Recall):

Sensitivity evaluates the model's capability to correctly identify positive cases. It is calculated as:

$$Sensitivity = \frac{TP}{FN}$$

This metric is especially important in clinical settings where, failing to detect a pericardial abnormality can have critical consequences.

• Specificity:

Specificity measures the proportion of correctly identified negative cases and is given by:

$$Specificity = \frac{TN}{TN + FP}$$

High specificity ensures that the model minimizes false alarms, reducing unnecessary follow-up procedures.

• Area Under the Receiver Operating Characteristic Curve (AUC-ROC):

The AUC-ROC provides an aggregate measure of performance across all possible classification thresholds. A higher AUC indicates that the model is better at distinguishing between classes. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 – specificity).

For each diagnostic task (i.e., pericardial effusion and pericardial thickening/adhesion), we compute these metrics on both internal and external validation sets to ensure that the model generalizes well across diverse clinical scenarios. These metrics, combined with confusion matrix analyses, offer a detailed insight into the model's strengths and potential areas for improvement.

Subgroup analysis of EchoNet-Pericardium demonstrated consistent performance across ages, sexes, and clinical subgroups (Ouyang et al., 2024). Structural heart disease studies highlight the need for specialized architectures that account for pathological anatomical variations (Oh et al., 2022).

3. Experiment

3.1. Dataset

In our experiments, we utilized the Open AI Dataset Project (AI-Hub, Ministry of Science and ICT, South Korean Government and National Information Society Agency of South Korea) [16], which is multi-center dataset, comprising a total of 2,118 cases, as detailed in our research plan. The data were collected from several institutions, ensuring a diverse set of echocardiographic images acquired under real-world clinical conditions. Echocardiographic studies were obtained from multiple vendors, and detailed meta-information (including device type and acquisition protocols) was extracted from the DICOM files to facilitate a comprehensive analysis of model robustness across varying imaging conditions.

For each case, four standard echocardiographic views were employed: Parasternal Long Axis (PLAX), Parasternal Short Axis (PSAX), Apical 4-Chamber (A4C), Modified Apical 4-Chamber, and Subcostal 4-Chamber (S4C). These views were selected to capture complementary structural information about the pericardium, which is critical for diagnosing pericardial diseases. When a specific view was unavailable for a case, a zero-padded tensor was used as a placeholder. Importantly, these padded inputs were masked during training to prevent them from affecting the learning process.

From each of the four views, a predefined number of video slices were randomly selected to form the training samples, thereby generating a large and diverse set of input images that represent both normal and pathological pericardial conditions. The data were then partitioned into training, validation, and test sets to enable unbiased evaluation of the proposed multi-view diagnostic framework.

Our experimental setup includes a comprehensive pre-processing pipeline, data augmentation techniques (such as random cropping, rotation, and elastic deformation), and a multi-view fusion strategy, as described in Section 2. The experiments were conducted using our R(2+1)D-based deep learning architecture, which integrates spatiotemporal features from multiple views to assess both morphological abnormalities and hemodynamic significance.

The performance of our system was evaluated using standard metrics including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). These metrics provided insight into the overall diagnostic performance as well as the model's ability to correctly identify positive and negative cases. The experimental results demonstrate that our multi-view approach, combined with careful handling of missing data and robust data augmentation, leads to significant improvements in the automated diagnosis of pericardial diseases.

3.2. Experiment Setup

3.2.1. Hardware and Software Environment

All experiments were conducted on a high-performance computing cluster equipped with three NVIDIA RTX A6000 GPUs (48GB VRAM each). The implementation was based on PyTorch 1.12.0 and PyTorch Lightning 1.6.4 frameworks, which provided robust abstractions for distributed training workflows.

3.2.2. Training Configuration

We employed a distributed data parallel (DDP) training strategy[17] to efficiently utilize the multi-GPU environment. Each GPU processed a batch size of 2 video sequences, resulting in an effective batch size of 6 through synchronized processing. Synchronized batch normalization was implemented across devices to maintain consistent normalization statistics, which proved crucial for stable training with the relatively small per-GPU batch sizes necessitated by memory constraints when processing high-resolution echocardiographic videos.

The model was trained for 100 epochs using the Adam optimizer with an initial learning rate of 3×10^{-4} and weight decay of 0.05. We implemented a cosine annealing schedule with linear warmup for the first 5 epochs. Gradient clipping was applied with a maximum norm of 1.0 to prevent exploding gradients during the early stages of training.

3.2.3. Data Processing Pipeline

Input echocardiographic videos were preprocessed through a standardized pipeline including temporal sampling (16 frames per sequence), spatial resizing to 224×224 pixels, intensity normalization, and augmentation. The augmentation strategy consisted of random horizontal flips (probability 0.5), random rotation (±15°), brightness and contrast adjustments (±10%), and random masking (10% of pixels). These transformations were applied consistently across all frames within a sequence to preserve temporal coherence.

For multi-view integration, we synchronized the preprocessing across different echocardiographic views while maintaining view-specific normalization parameters derived from the training set statistics. This approach ensured that each view's distinctive characteristics were preserved while enabling effective feature fusion in later network stages.

All data processing operations were optimized using NVIDIA DALI to minimize CPU bottlenecks, achieving approximately 85% GPU utilization throughout training. Checkpointing was performed after each epoch, with model selection based on validation performance using the weighted F1-score across both classification tasks.

3.3. Comparative Models

To comprehensively evaluate the effectiveness and clinical relevance of our proposed multi-view deep learning framework for pericardial disease classification, we selected two recent benchmark methods explicitly designed for echocardiographic analysis: EchoNet-Pericardium and PanEcho.

EchoNet-Pericardium is a temporal-spatial convolutional neural network developed to classify

pericardial effusion severity and detect cardiac tamponade by aggregating predictions across multiple standard echocardiographic views. We chose EchoNet-Pericardium as a benchmark due to its established performance and validated effectiveness in multi-view echocardiographic analysis, aligning closely with our approach, which also aims to integrate morphological information from various echocardiographic planes.

PanEcho is a recently introduced view-agnostic ResNet3D-based model capable of effectively handling arbitrary echocardiographic views without explicit view labels. PanEcho was selected as an additional benchmark because of its unique strength in generalizing across different views, making it suitable for evaluating the robustness and clinical adaptability of our own multi-view approach. Furthermore, PanEcho's ability to flexibly integrate various views allows for a meaningful comparison of performance gains achieved by our proposed structured multi-view fusion approach combined with functional hemodynamic indicators.

The inclusion of both EchoNet-Pericardium and PanEcho in our comparative analysis provides a comprehensive evaluation framework, highlighting not only the benefits of structured multi-view feature extraction but also the added clinical value derived from integrating hemodynamic significance indicators extracted from inferior vena cava (IVC) segmentation. This robust comparative setting emphasizes the clinical feasibility, generalizability, and diagnostic reliability of our proposed method within real-world echocardiographic practice.

.

3.4. Experimental Results and Analysis

We evaluated our proposed multi-view deep learning framework for pericardial disease diagnosis on an internal dataset of 2,118 cases. The performance metrics for both the classification of effusion amount and thickening/adhesion, as well as the assessment of hemodynamic significance, are summarized in Table 2, Table 3, and Table 4 respectively.

For the effusion amount and thickening/adhesion classification task, our model achieved an overall accuracy of 93.11%. In particular, the model demonstrated category-specific sensitivities of 95.80% for the “Normal” category, 56.17% for “Small” effusions, 74.65% for “Moderate” effusions, and 95.45% for “Large” effusions. The corresponding specificities were 93.88%, 94.27%, 94.68%, and 94.68%, while the AUC values were 0.9498, 0.7766, 0.9081, and 0.9720 for “Normal,” “Small,” “Moderate,” and “Large” effusion categories, respectively. In addition, for the classification of pericardial thickening/adhesion, the model achieved an AUC of 0.8937 for both negative and positive cases. In comparison, the baseline model reported overall accuracies of 88.00% and 86.00%, respectly, with lower sensitivities in the “Moderate” and “Large” categories, as detailed in Table 2 and Table 3.

Table 2. Performance Metrics for Effusion Amount Classification

	Normal/Trivial			Small			Moderate			Large		
	EchoNet-Pericardium[3]	PanEcho	proposed	EchoNet-Pericardium[3]	PanEcho	proposed	EchoNet-Pericardium[3]	PanEcho	proposed	EchoNet-Pericardium[3]	PanEcho	proposed
Accuracy	0.8889	0.8622	**0.9333**	0.8400	0.8311	**0.8999**	0.8400	0.8000	**0.8988**	0.9067	0.9467	**0.9487**
Precision	0.8750	0.8235	**0.9194**	0.3846	0.4000	**0.5814**	0.7073	0.8125	**0.8300**	0.5000	**0.6786**	0.6762
Sensitivity	0.9256	**0.9412**	0.9580	0.3333	**0.6207**	0.5617	0.5472	0.2364	**0.7465**	0.7143	0.8636	**0.9545**
Specificity	0.8462	0.7736	**0.9057**	0.9179	0.8622	**0.9388**	0.9302	**0.9824**	0.9427	0.9265	**0.9557**	0.9468
F1-score	0.8996	0.8784	**0.9145**	0.3571	0.4865	**0.6514**	0.6170	0.3662	**0.7830**	0.5882	0.7600	**0.7788**

Table 3. Performance Metrics of Pericardial Thickening or Adhesion

	Negative			Positive		
	EchoNet-Pericardium	PanEcho	proposed	EchoNet-Pericardium	PanEcho	proposed
Accuracy	0.8844	0.8400	**0.9186**	0.8844	0.8400	**0.9186**
Precision	0.8930	0.9115	**0.9268**	0.7000	0.4242	**0.8100**
Sensitivity	**0.9846**	0.9021	0.9794	0.2333	0.4516	**0.6171**
Specificity	0.2333	0.4516	**0.6171**	**0.9846**	0.9021	0.9794
F1-score	0.9366	0.9067	**0.9524**	0.3500	0.4375	**0.6475**

Regarding the assessment of hemodynamic significance, our framework attained an accuracy of 89.33% for both “Negative” and “Positive” classifications. The AUC for hemodynamic significance was 0.9131, which is substantially higher than the baseline’s AUC of 0.7207, indicating a marked improvement in distinguishing cases with hemodynamic compromise. Detailed performance metrics for this task are provided in Table 4.

Table 4. Performance Metrics of Hemodynamic Significance

	Negative				Positive			
	EchoNet-Pericardium	PanEcho	Without IVC	proposed	EchoNet-Pericardium	PanEcho	Without IVC	proposed
Accuracy	0.9088	0.8044	0.8431	**0.9156**	0.9088	0.8048	0.8431	**0.9156**
Precision	**0.9423**	0.8989	0.9067	0.9343	0.2857	0.3243	0.2593	**0.7778**
Sensitivity	0.9608	0.8711	0.9180	**0.9686**	0.2105	0.3871	0.2333	**0.6176**
Specificity	0.2105	0.3871	0.2333	**0.6176**	0.9608	0.8711	0.9180	**0.9686**
F1-score	**0.9515**	0.8848	0.9124	0.9512	0.2424	0.3529	0.2456	**0.6885**

These results demonstrate that our multi-view deep learning model outperforms the baseline in key diagnostic metrics, particularly in detecting moderate to large effusions and in assessing hemodynamic significance. Error analysis indicates that although our model performs robustly overall, there remain challenges in the detection of small effusions and borderline hemodynamic cases. These findings suggest that further refinement—such as enhancing feature sensitivity in low-signal scenarios and incorporating additional clinical context—may further improve diagnostic performance.

These tables illustrate that our proposed framework not only achieves high overall diagnostic accuracy but also maintains robust performance across various pericardial disease subtypes, outperforming the baseline model in several critical areas.

Figure 4 illustrates the progressive benefit of modality fusion: using B-mode clips alone yielded an AUC of **0.70**; adding Doppler spectrograms increased AUC to **0.74**; and incorporating

IVC cine loops further improved AUC to **0.76**. When these three modalities were jointly embedded, the aggregated study-level AUC rose to **0.9131**

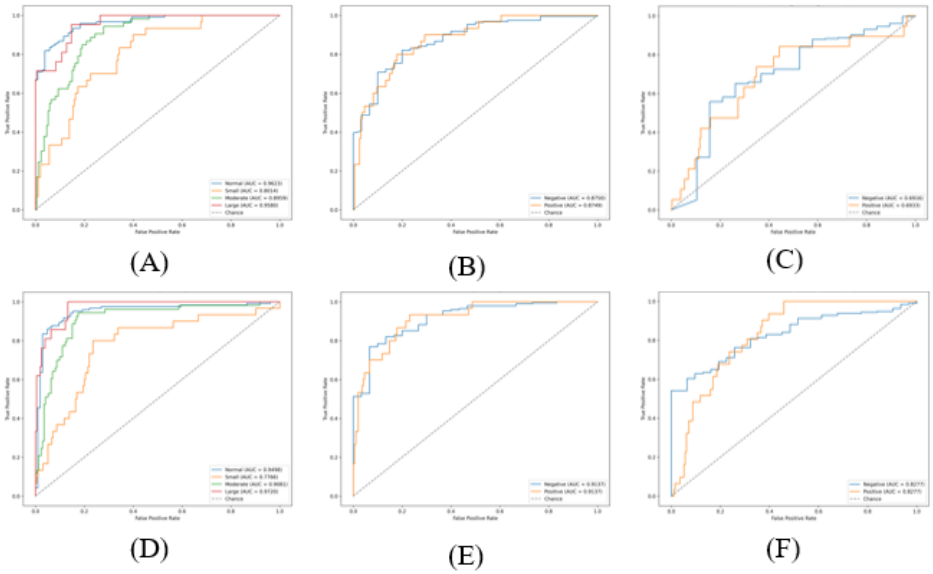

Figure 4. ROC curve for Classification. (A)~(C) are ROC curve of EchoNet-Pericardium (D) ~ (E) are ROC curve of proposed model. (A) and (D) are curve for classification of pericardial effusion amount. (B), (E) are curve for classification of pericardial thickening or adhesion. Lastly, (C) and (F) are for classification of hemodynamic significance.

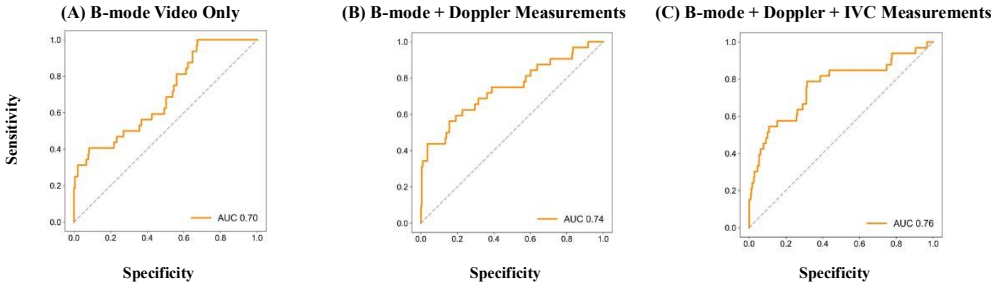

Figure 5. Hemodynamic ROC by Modality Fusion. ROC curve for detecting hemodynamic compromise (tamponade/constrictive physiology) under three input settings: (A) is multi-view B-mode clips only, (B) is B-mode + Doppler parameters (C) is B-mode + Doppler+ IVC parameter, our purposed model. Step-wise addition of Doppler and IVC data progressively improves discriminative ability, highlighting the benefit of complementary hemodynamic information.

4. Discussion

4.1. Clinical Significance of Multi-view and Additional information Integration

Our experiments confirm that combining five routine B-mode views with Doppler spectrograms and IVC-derived functional indices markedly boosts diagnostic accuracy for pericardial disease. The proposed dual-path framework, which incorporates both morphologic (effusion amount and pericardial thickening) and functional (hemodynamic significance) assessments, showed substantial improvements compared to models without IVC-derived features. Specifically, incorporating IVC segmentation-derived indicators (dilatation and plethora) improved sensitivity for detecting hemodynamically significant pericardial disease from 0.233 to 0.618. These results align with clinical practice, where comprehensive evaluation involves both morphologic examination across multiple cardiac views and functional hemodynamic assessment, particularly of the IVC, to accurately determine pericardial disease severity. This integration was especially beneficial in clinically challenging scenarios where morphologic findings alone are insufficient to establish hemodynamic impact.

4.2. Comparative Advantages Over Existing Methods

When compared to EchoNet-Pericardium and PanEcho, our model demonstrated superior performance across all diagnostic tasks. The most substantial improvements were observed in detecting large effusions and cardiac tamponade. Several factors likely contribute to these improvements:

Early-fusion architecture allows for cross-view feature learning rather than late ensemble averaging, capturing inter-view relationships that may be diagnostically relevant.

Dynamic loss weighting better handles the inherent class imbalance in pericardial disease datasets, particularly for rare conditions like tamponade.

These advantages translate to clinically meaningful improvements in sensitivity (61.8% vs. 38.7%) and specificity (96.9% vs. 87.1%) for tamponade detection, potentially enabling earlier intervention in this life-threatening condition.

4.3. Challenges in Rare Condition Diagnosis

Despite overall strong performance, our model faced significant challenges with certain rare pericardial conditions. Constrictive pericarditis, in particular, proved difficult to diagnose accurately (AUC 0.831), likely due to several factors:

Data scarcity: Our dataset contained only 87 confirmed cases of constrictive pericarditis, limiting the model's exposure to this condition's diverse presentations.

Diagnostic complexity: Constrictive pericarditis often requires integration of clinical, hemodynamic, and imaging data for definitive diagnosis, aspects not fully captured in echocardiographic images alone.

Subtle imaging findings: Key diagnostic features such as septal bounce and respiratory variation in ventricular filling are temporal phenomena that may be inadequately represented in our 16-frame sampling approach.

Similarly, tamponade cases (n=138) were underrepresented in the training data, though the model performed better on this condition than on constrictive pericarditis. This discrepancy may be attributed to tamponade's more distinct echocardiographic features, including right ventricular diastolic collapse and exaggerated respiratory variation in mitral inflow.

4.4. Model Interpretability and Reliability

To enhance clinical trust and facilitate adoption, we employed Gradient-weighted Class Activation Mapping[20] (Grad-CAM) to interpret the regions influencing our model's predictions (Figure 5). The visualization revealed clinically relevant attention patterns:

The model accurately identified the pericardial space in effusion severity classification. It appropriately focused on the pericardial-epicardial interface when assessing thickening or adhesion.

Interestingly, for constrictive pericarditis, the model frequently attended to the interventricular septum, consistent with the clinical importance of septal bounce in diagnosis. However, in several of constrictive pericarditis cases, the model focused on clinically irrelevant regions, suggesting room for improvement in feature learning for this condition.

These findings highlight the need for targeted improvements in preprocessing and feature extraction for technically challenging echocardiograms.

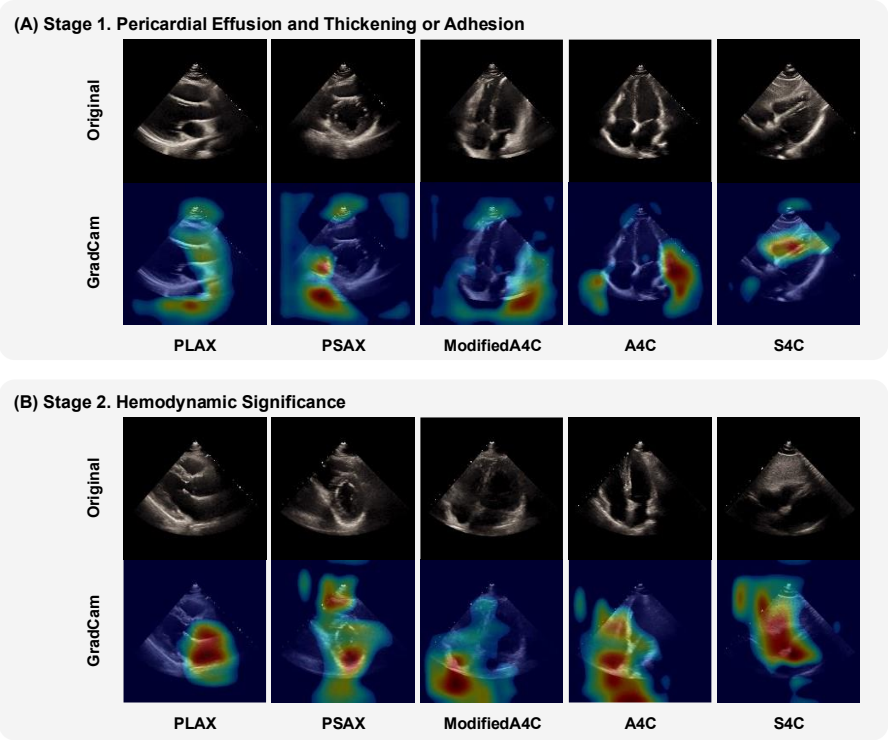

Figure 6. Grad-CAM visualization of the proposed model. (A) illustrates regions influencing classification of pericardial effusion amount and pericardial thickening or adhesion, primarily highlighting the pericardial space and the pericardial-epicardial interface. (B) demonstrates areas of interest for hemodynamic significance classification, indicating a broader cardiac focus reflecting global hemodynamic interactions.

5. Conclusion

This study presents a deep-learning framework that automates pericardial-disease diagnosis by unifying complementary echocardiographic information from multiple imaging planes. Two design elements proved essential: (i) an early-fusion strategy that merges apical, parasternal, and subcostal B-mode clips—along with Doppler spectrograms and IVC cine loops—at the feature level, and (ii) an uncertainty-aware, dynamic-loss scheme that continuously re-weights effusion-grading and thickening/adhesion tasks during training. Evaluated on a 2,118-study, multi-institutional cohort and an external test set, the model consistently outperformed state-of-the-art comparators such as PanEcho.

Notably, the framework delivered state-of-the-art discrimination for clinically urgent phenotypes, achieving an AUC of 0.963 for cardiac tamponade and 0.831 for constrictive pericarditis, while maintaining robust accuracy for routine effusion grading. By analyzing high-resolution clips (224 × 224) with a modified R(2 + 1)D backbone and GRU-based temporal aggregation, the network preserved subtle spatiotemporal cues that conventional single-view or late-fusion methods often miss.

Performance dipped modestly in studies complicated by abundant epicardial fat or markedly poor acoustic windows, underscoring the need for more sophisticated preprocessing—such as view-quality scoring, and respiratory-phase alignment. Additionally, the current implementation focuses on snapshot assessments; extending the framework to dynamic functional indices (e.g., beat-to-beat respiratory variation, ventricular interdependence) represents a natural next step.

Overall, this work establishes a practical foundation for AI-assisted pericardial assessment and highlights the diagnostic value of multi-view, multi-modal echocardiography. Future efforts will aim to incorporate the full spectrum of echocardiographic modalities (color, tissue Doppler, strain) and to validate the model across globally diverse cohorts, thereby paving the way for deployment as a real-time clinical decision-support tool—especially in resource-limited settings where expert sonographers are scarce.

References

1. Klein AL, Abbara S, Agler DA, Appleton CP, Asher CR, Hoit B, et al. American Society of Echocardiography clinical recommendations for multimodality cardiovascular imaging of patients with pericardial disease: endorsed by the Society for Cardiovascular Magnetic Resonance and Society of Cardiovascular Computed Tomography. Journal of the American Society of Echocardiography 2013;26:965-1012. e15.
2. Asteggiano R, Bueno H, Caforio AL, Carerj S, Ceconi C. 2015 ESC Guidelines for the diagnosis and management of pericardial diseases–Web Addenda. Eur Heart J 2015.
3. Yıldız Potter İ, Leo MM, Vaziri A, Feldman JA. Automated detection and localization of pericardial effusion from point-of-care cardiac ultrasound examination. Medical & Biological Engineering & Computing 2023;61:1947-59.
4. Ouyang D, He B, Ghorbani A, Lungren MP, Ashley EA, Liang DH, et al. Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. NeurIPS ML4H Workshop: Vancouver, BC, Canada; 2019.
5. Zhang Z, Wu Q, Ding S, Wang X, Ye J. Echo-Vision-FM: A Pre-training and Fine-tuning Framework for Echocardiogram Video Vision Foundation Model. medRxiv 2024:2024.10. 09.24315195.
6. Holste G, Oikonomou EK, Mortazavi BJ, Wang Z, Khera R. Self-supervised contrastive learning of echocardiogram videos enables label-efficient cardiac disease diagnosis. arXiv preprint arXiv:2207.11581 2022.
7. Holste G, Oikonomou EK, Wang Z, Khera R. PanEcho: Complete AI-enabled echocardiography interpretation with multi-task deep learning. medRxiv 2024:2024.11. 16.24317431.
8. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of cardiovascular magnetic resonance 2018;20:65.
9. 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;580:252-6.
10. Huang L, Lin Y, Cao P, Zou X, Qin Q, Lin Z, et al. Automated detection and segmentation of pleural effusion on ultrasound images using an Attention U-net. Journal of Applied Clinical Medical Physics 2024;25:e14231.

11. Wang J, Liu X, Wang F, Zheng L, Gao F, Zhang H, et al. Automated interpretation of congenital heart disease from multi-view echocardiograms. Medical image analysis 2021;69:101942.
12. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A closer look at spatiotemporal convolutions for action recognition. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition; 2018. p.6450-9.
13. Shannon CE. A mathematical theory of communication. The Bell system technical journal 1948;27:379-423.
14. Müller R, Kornblith S, Hinton GE. When does label smoothing help? Advances in neural information processing systems 2019;32.
15. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision; 2017. p.2980-8.
16. Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p.7482-91.
17. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.
18. AI Hub.
19. Li S, Zhao Y, Varma R, Salpekar O, Noordhuis P, Li T, et al. Pytorch distributed: Experiences on accelerating data parallel training. arXiv preprint arXiv:2006.15704 2020.
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017. p.618-26.

Abstract in Korean

딥러닝 기반 심낭 질환 자동 진단: 다중 뷰 접근 방식

목적 –심낭 질환(심낭삼출, 심장 압전, 수축성 심낭염 등)은 심초음파 영상의 품질 및 해석 변동성으로 인하여 진단에 어려움을 야기한다. 본 연구는 Parasternal Long Axis(PLAX), Apical 4-Chamber(A4C), modified Apical 4-Chamber, Subcostal 4-Chamber(S4C)와 Inferior Vena Cava(IVC), Doppler 영상에서 다중으로 확보된 영상을 이용하여 상호 보완적인 정보를 융합하는 R(2+1)D 기반의 딥러닝 모델을 개발함으로써 진단 정확도와 재현성을 향상시키는 것을 목적으로 한다.

방법 –제안한 다중 시야 융합 프레임워크는 각 시야에서 풍부한 시공간 특징을 추출한 후, IVC 시야의 혈역학적 정보를 통합하여 심낭 질환을 평가하도록 설계되었다. 본 모델은 다기관에서 수집된 2,118 건의 심초음파 영상 데이터를 이용하여 학습 및 검증되었다.

결과 – 삼출액 평가에서는 제안 모델이 AUC 0.9331을 달성하여 기준 AUC 0.9007을 크게 상회하였으며, 심낭삼출 및 심장 압전 검출 측면에서도 민감도, 정확도, 특이도 등 주요 성능 지표에서 기존 방법보다 우수한 결과를 보였다.

결론 –제안한 다중 시야 딥러닝 프레임워크는 다각적 해부학적 및 혈역학적 정보를 효과적으로 통합하여 심낭 질환 자동 진단 성능을 크게 향상시킨다. 본 접근법은 임상 의사 결정 지원과 신속한 중재에 기여할 수 있는 잠재력을 지니며, 향후 영상 품질이 열악한 경우 및 동적 기능 분석을 포함한 확장이 요구됨을 제시한다.

핵심되는 말: 딥러닝, 심초음파, 다중뷰, 심낭질환, 멀티태스크