Multimodal Cardiac Phenotype Estimation Using ECG and Clinical Data

Cardiac Phenotype Estimation

Kavitha Bai A. S.*, J. Somasekar
Department of Computer Science and Engineering-Faculty of Engineering and Technology,
JAIN (Deemed-to-be University), Bangalore 562112, Karnataka, India

***Abstract*—Existing ECG-based cardiac assessment methods primarily focus on disease classification, providing limited information about quantitative cardiac function, structural remodeling, and hemodynamic status. The complex relationship between ECG patterns and echocardiographic phenotypes remains insufficiently explored. This study proposes a Multimodal Cardiac Phenotype Prediction Network (MultiCard-PNet) for estimating clinically relevant cardiac phenotypes, including left ventricular ejection fraction (LVEF), left ventricular wall thickness (LVWT), and pulmonary artery systolic pressure (PASP), using 12-lead ECG signals and clinical information. The proposed framework integrates a 1D-ResNet-based ECG encoder for extracting temporal and morphological representations and an MLP-based clinical encoder for learning patient-specific characteristics. A Clinical-Guided Feature Gating (CGFG) mechanism adaptively recalibrates ECG features using clinical context before multimodal fusion. The fused representation is optimized through a multi-task regression framework with a weighted Huber loss for robust phenotype estimation. Evaluation on the EchoNext dataset demonstrated that the proposed MultiCard-PNet framework achieving a MAE value of 7.179 for LVEF, 0.168 for LVWT, and 9.537 for PASP. Although multimodal feature integration improved the reported metrics compared with single-modality baselines. The proposed MultiCard-PNet provides a non-invasive and comprehensive approach for ECG-based cardiac phenotype estimation, enabling improved cardiac assessment beyond conventional ECG interpretation and supporting future clinical decision-support applications.**

Keywords*—Cardiac function; phenotype estimation; electrocardiogram; 1D-ResNet; multi-layer perceptron

I. Introduction

The heart is one of the most important organs required for maintaining blood flow in the body, and any abnormality of its structure or function can have an impact on cardiac performance [1]. Various clinical and imaging-based measurements can be used to quantify cardiac function and structure to measure the condition and performance of the heart [2]. Cardiac structure and function are important for detecting abnormalities and determining the cardiac phenotype [3]. Structural Heart Disease (SHD) is an abnormality of the valves, muscles, or walls of the heart [4]. A diagnostic cardiac test commonly used for measurement of the heart's electrical activity using electrodes attached to the body surface is known as an electrocardiogram (ECG) [5]. In clinical practice, the 12-lead ECG is used to gain insight into the electrical activity of the heart and is valuable for diagnosing heart disease [6, 7]. But ECG-based analysis still faces several challenges, including limited external validation, as most studies have been reviewed and are single-center [8, 9]. Recent studies have explored the use of Artificial Intelligence (AI)-based approaches for identifying structural heart abnormalities from 12-lead ECGs [10]. The Deep Learning (DL) EchoNext model was trained on million-heart rhythm and imaging datasets to identify multiple types of SHD from ECG. The model demonstrated high diagnostic accuracy in internal and external validation. However, the model mainly focused on detecting heart disease and did not estimate detailed cardiac parameters [11]. Similarly, another study developed Practical scREening using ENsemble machine learning sTrategy (PRESENT) to detect multiple SHD using 12-lead ECG. The model exhibited high accuracy in various hospitals and patient subgroups. But the model was less effective in individuals ≥65 years of age [12]. Despite recent advances in ECG-based cardiac assessment, existing approaches mainly focus on binary SHD classification, limiting their ability to provide comprehensive insights into cardiac function, structure, and hemodynamics. Furthermore, conventional ECG-clinical fusion strategies fail to capture patient-specific interactions between clinical context and ECG patterns.

To overcome these limitations, this study proposes Multimodal Cardiac Phenotype Prediction Network (MultiCard-PNet), which integrates 12-lead ECG representations and clinical information through a Clinical-Guided Feature Gating (CGFG) mechanism and multi-task learning strategy for simultaneous estimation of Left Ventricular Ejection Fraction (LVEF), Left Ventricular Wall Thickness (LVWT), and Pulmonary Artery Systolic Pressure (PASP). The major contributions of the proposed work are summarized as follows:

1) Multimodal ECG-clinical cardiac phenotype estimation framework: A novel MultiCard-PNet framework is developed to estimate quantitative echocardiographic phenotypes, including LVEF, LVWT, and PASP, from 12-lead ECG signals and clinical information.

2) CGFG for adaptive ECG representation learning: A clinically guided feature gating mechanism is introduced, where patient-specific clinical embeddings dynamically generate feature-wise gates to adaptively recalibrate ECG representations before multimodal fusion, enabling better integration of electrical and clinical characteristics.

*Corresponding author.

3) Multi-task regression architecture for cardiac assessment: A multi-task learning strategy is proposed with shared multimodal representations and dedicated regression heads for simultaneous estimation of LVEF, LVWT, and PASP, allowing the model to learn complementary cardiac functional, structural, and hemodynamic information.

4) Robust optimization using weighted huber multi-task loss: A weighted Huber loss-based optimization strategy is incorporated to balance multiple phenotype estimation tasks while reducing the effect of measurement noise and outliers, improving training stability and prediction robustness.

II. Literature Survey

Recent advancements in DL have analyzed several approaches for improved heart disease diagnosis and prediction. Ajadi et al. [13] introduced the identification of SHD based on multimodal DL. This model utilized a Simple Convolutional Neural Network (CNN), Light Transformer, ResNet1d18, and Temporal Convolutional Network (TCN). This model improved the prediction performance, computational efficiency, stability, and fairness across demographic groups. Qiu et al. [14] proposed a novel Transform-based algorithm for the prediction of Atrial Fibrillation (AF) with Clinical Features and Surface ECG signals. This system utilized end-to-end DL methods and a fusion module. However, this framework could not be applied to larger regions due to limited external validation, inconsistent ECG data availability, and limited model interpretability. Javed et al. [15] established a multimodal, IoMT-enabled intelligent healthcare system for predicting Coronary Heart Disease (CHD) utilizing a CNN on ECG data to categorize different heart conditions. However, this system could not incorporate structured patient-specific clinical information or estimate quantitative echocardiographic phenotypes. Wang et al. [16] introduced a multimodal fusion model based on Squeeze-and-Excitation ResNet with Next (SEResNeXt) using actual clinical data and 12-lead ECG for early CHD diagnosis. This system improved CHD diagnostic performance compared with single-modality approaches. However, this model increases computational requirements and limits its deployment in resource-constrained clinical settings. Zhu et al. [17] developed a Dual-scale Deep Residual Network (DDR-Net) to automatically extract features from ECG signals for early detection of cardiovascular disease utilizing a Support Vector Machine (SVM). However, this system could not provide an end-to-end DL approach for CVD detection. Amiri and Shariatrad et al. [18] developed a transfer learning and multi-task optimization-based framework for multimodal heart disease prediction using 1D CNN-ResNet for ECG analysis. The model improved prediction accuracy, calibration, and reliability, making it suitable for real-time cardiac disease assessment. However, the framework could not validate the findings using true same-patient multimodal ECG and clinical data. Qi et al. [19] introduced ECGEFNet based on a two-branch DL model to estimate LVEF from ECGs for cardiac function screening and monitoring. The model utilized Fusion Attention Mechanism (FAT) and Two-Branch Feature Fusion (BFF) for feature interaction and fusion. However, this model was not able to add other parameters for a comprehensive evaluation of LVEF. Thus, analyzing existing works, most existing approaches are limited to binary classification tasks and provide insufficient information about quantitative cardiac phenotypes. Although multimodal ECG and clinical data integration has gained attention, current fusion strategies mainly rely on simple feature concatenation without considering patient-specific interactions. Furthermore, existing single-task models fail to exploit the shared relationships among different cardiac functional, structural, and hemodynamic parameters. Therefore, an multimodal framework with adaptive feature learning and multi-task phenotype estimation is required to address these limitations.

III. Methodology

The proposed methodology presents a MultiCard-PNet framework for estimating cardiac phenotypes from 12-lead ECG signals and clinical information, as represented in Fig. 1. The preprocessed ECG signal is first fed into a 1D-Residual Network (1D-ResNet), which captures hierarchical temporal and morphological features linked to cardiac electrical patterns. In parallel, seven clinical variables, including sex, ventricular rate, atrial rate, PR interval, QRS duration, QTc interval, and age, are processed through a Multilayer Perceptron (MLP) to generate a compact clinical embedding. To integrate ECG and clinical information, a CGFG mechanism is proposed, which embeds the clinical information to generate sigmoid-based feature weights to flexibly recalibrate ECG representation based on the patient-specific characteristics. The gated ECGs are then combined with the clinical representation to create a multimodal feature vector. The fused representation is passed into a multi-task regression network consisting of shared layers and separate prediction heads for estimating LVEF, LVWT, and PASP. The model is jointly optimized using a weighted Huber multi-task loss, which balances individual regression objectives and minimizes the effect of measurement outliers. Finally, the proposed framework is evaluated using regression performance metrics to assess its effectiveness in cardiac phenotype estimation.

A. Dataset Description of the EchoNext Dataset

The study utilized the EchoNext dataset [20] that includes a sample of 100,000 ECGs from Columbia and Allen hospitals. The proposed framework combines 12-lead ECG waveform representations with seven structured input variables: age, sex, ventricular rate, atrial rate, PR interval, QRS duration, and QTc interval. Among these variables, age and sex represent patient characteristics, whereas ventricular rate, atrial rate, PR interval, QRS duration, and QTc are ECG-derived measurements. The proposed framework integrates waveform-level representations with structured patient characteristics and ECG-derived measurements. The dataset includes a total of 11 classes. We only utilize the three important classes: LVEF, LVEF, and PASP. This allows the model to learn complementary cardiac functional, structural, and hemodynamic information. The dataset is split with a patient-level split to avoid data leakage between training and evaluation sets. ECG signals are pre-processed with noise removal, normalization and clinical feature standardization before being fed into the MultiCard-PNet framework. The data split strategy is given in Table I.

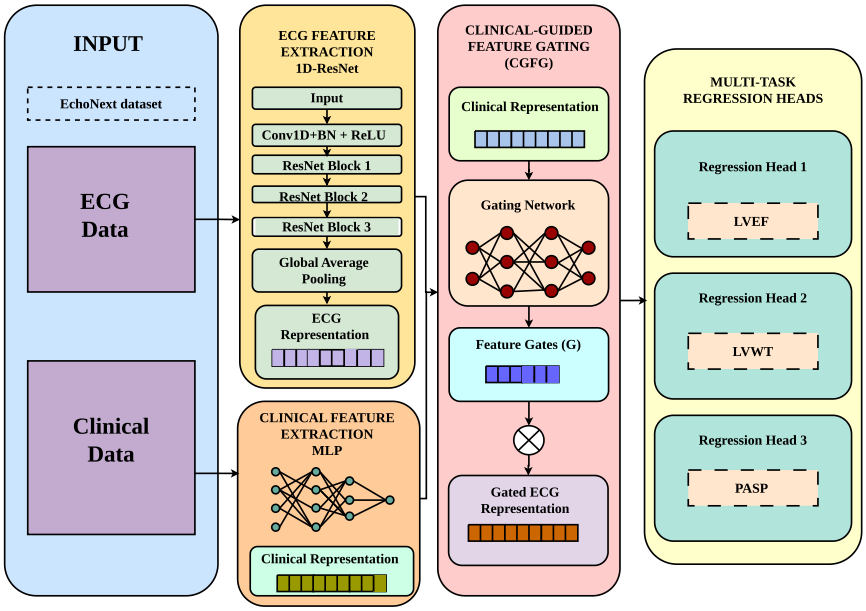

Fig. 1. Schematic representation of the proposed MultiCard-PNet framework.

TABLE I. DATASET SPLIT DISTRIBUTION

Partition	ECG Records	Percentage	Purpose
Training	72,475	87.80%	Model training
Validation	4,626	5.60%	Model selection and tuning
Test	5,442	6.59%	Final performance evaluation
Excluded Historical ECGs	17,457	-	Prevent data leakage
Total Dataset	100,000	100%	Complete dataset

B. ECG and Clinical-Feature Preprocessing

The proposed framework processes 12-lead ECG waveforms along with seven clinical features. Each input ECG contains 12 leads and 2500 samples per lead. The clinical feature vector consists of sex, ventricular rate, atrial rate, PR interval, QRS duration, corrected QT interval and age at time of ECG acquisition. ECG signal processing: The dataset loader converts the waveform values to 32-bit floating-point format, fixes the orientation of the waveform if the stored array has its 12 leads along the last axis, and replaces non-finite values with zero. Then, the generated waveform is input into the one-dimensional residual network encoder. The loader doesn't specify the ECG's sampling frequency, the digital filtering methodology, filter order and cut-off frequencies, or any other waveform normalization procedure, these details have to be determined from the implementation of the upstream preprocessing. Clinical feature handling: The seven clinical variables are loaded as a vector of 32-bit floating point. The non-finite values are set to zero before the vector is passed to the clinical multilayer perceptron. The current implementation does not check whether the clinical features stored have been previously imputed, encoded as categorical or standardized as features. These operations and their parameters should be reported according to the real upstream pre-processing pipeline. Target-labels handling: The model predicts LVEF, LVWT and PASP. Missing target values are filled with zero to be compatible with tensors, but the validity masks ensure that the unavailable targets are not taken into account in the respective task-specific Huber losses and evaluation metrics.

C. 1D-ResNet to Extract ECG Representation

The preprocessed 12-lead ECG waveform is passed into a 1D-ResNet to extract discriminative temporal and morphological ECG representations, as represented in Fig. 2. The 1D-ResNet learns hierarchical patterns directly from raw ECG signals, such as cardiac function-dependent temporal dependencies, cardiac rhythm changes and waveform morphology. Initially, a 1D CNN is used to extract 1D convolutional local temporal features from ECG waveforms. The convolution is expressed as in Eq. (1).

$$z_i = \sigma(W * x_i + b) \quad (1)$$

where, W is the convolution kernels, b is the bias term, and σ is the activation function. The convolution layer learns the short-term ECG characteristics, such as waveform change and interval variation. The extracted features are then fed into multiple residual blocks, which allow deeper feature learning while preventing degradation problems. Each residual block contains two consecutive 1D convolution layers followed by batch normalization and nonlinear activation. The residual connection is represented as in Eq. (2).

$$Y_l = H(X_l) + X_l \quad (2)$$

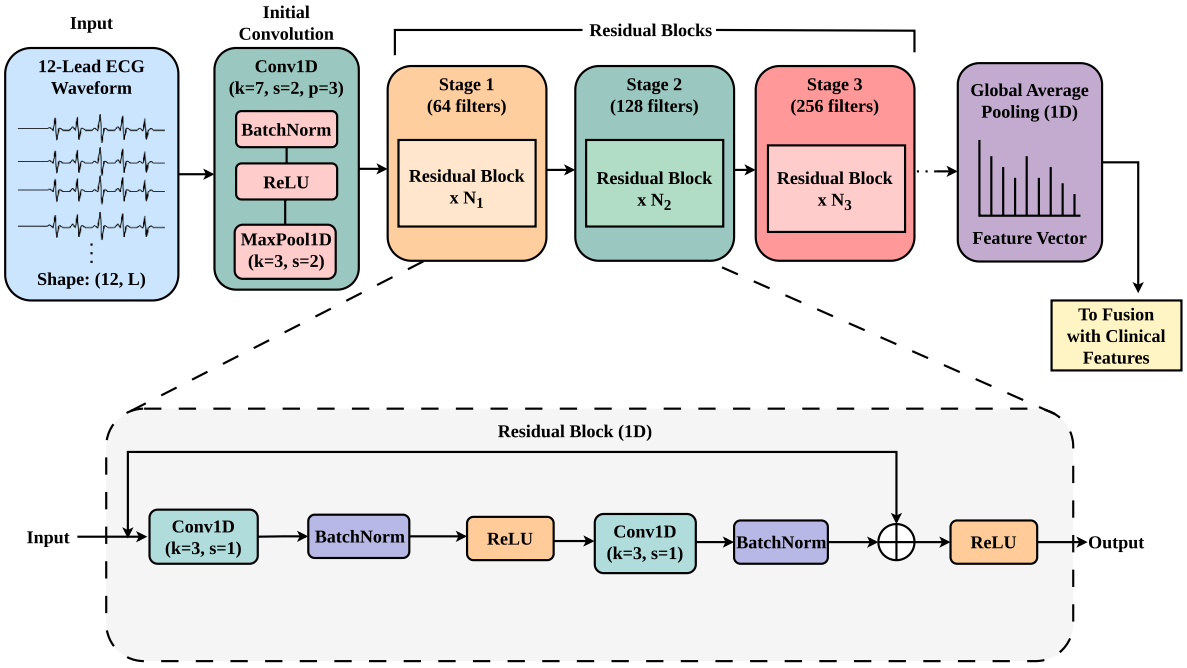

Fig. 2. Architecture of the 1D-ResNet for ECG representation extraction. The preprocessed ECG is processed through convolutional and residual blocks to learn hierarchical temporal and morphological features. Global average pooling produces a compact ECG representation for subsequent fusion with clinical features.

where, X_l is the input feature map of the l^{th} residual block, Y_l is the output feature representation, and $H(X_l)$ is the learned transformation by the convolution layers.

This skip connection will maintain critical ECG information and learn intricate cardiac patterns. A Global Average Pooling (GAP) layer is used to reduce the temporal feature dimension by aggregating the learned features, as shown in Eq. (3).

$$F_{gap} = \frac{1}{T}\sum_{t=1}^{T} F_t \tag{3}$$

where, F_t is a feature vector extracted at time step t. Finally, a Fully Connected (FC) Layer is used to update the pooled representation to a compact embedding for an ECG, expressed in Eq. (4).

$$E_{ECG} = W_f F_{gap} + b_f \tag{4}$$

where, E_{ECG} is the final ECG feature representation to be used in multimodal fusion. Thus, the 1D-ResNet is effectively captures local morphological features and long-range temporal relationships within ECG signals.

D. MLP to Extract Clinical Representation

Structured clinical information is processed through a Multilayer Perceptron (MLP) to produce a compact patient-specific clinical representation. The framework is such that seven clinically relevant features are fed into the system, such as sex, atrial rate, PR interval, QRS duration, QTc interval, ventricular rate, and age. The MLP consists of three main components: input layer, FC hidden layers, and output embedding layer, as shown in Fig. 3. The input layer receives the normalized clinical feature vector. The first FC layer performs feature transformation and captures interactions among clinical variables, as denoted in Eq. (5).

$$h_1 = \sigma(W_1 X_{clin} + b_1) \tag{5}$$

where, W_1 and b_1 represent the weights and biases of the first dense layer, and σ denotes the ReLU activation function. A subsequent hidden layer further refines the clinical representation by learning higher-level feature interactions, as represented in Eq. (6).

$$h_2 = \sigma(W_2 h_1 + b_2) \tag{6}$$

where, h_2 represents the learned intermediate clinical feature representation. Finally, the output layer projects the learned features into a compact clinical embedding space E_{clin}, as expressed in Eq. (7).

$$E_{clin} = W_o h_2 + b_o \tag{7}$$

where, W_o and b_o are output layer parameters. The generated clinical embedding captures important patient-specific characteristics and provides contextual information for guiding ECG feature adaptation in the subsequent CGFG module.

E. Clinical-Guided Feature Gating

The CGFG mechanism is introduced to adaptively integrate patient-specific clinical context with ECG representations. CGFG allows the clinical representation to selectively enhance or suppress relevant ECG features based on individual patient characteristics. The CGFG module learns a feature-wise gating vector from the clinical representation, denoted in Eq. (8).

$$FG = \sigma\left(W_g E_{clin} + b_g\right) \tag{8}$$

where, W_g and b_g are learnable parameters of the gating layer and σ represents the sigmoid activation function. The obtained gate vector G determines the importance of each ECG feature dimension.

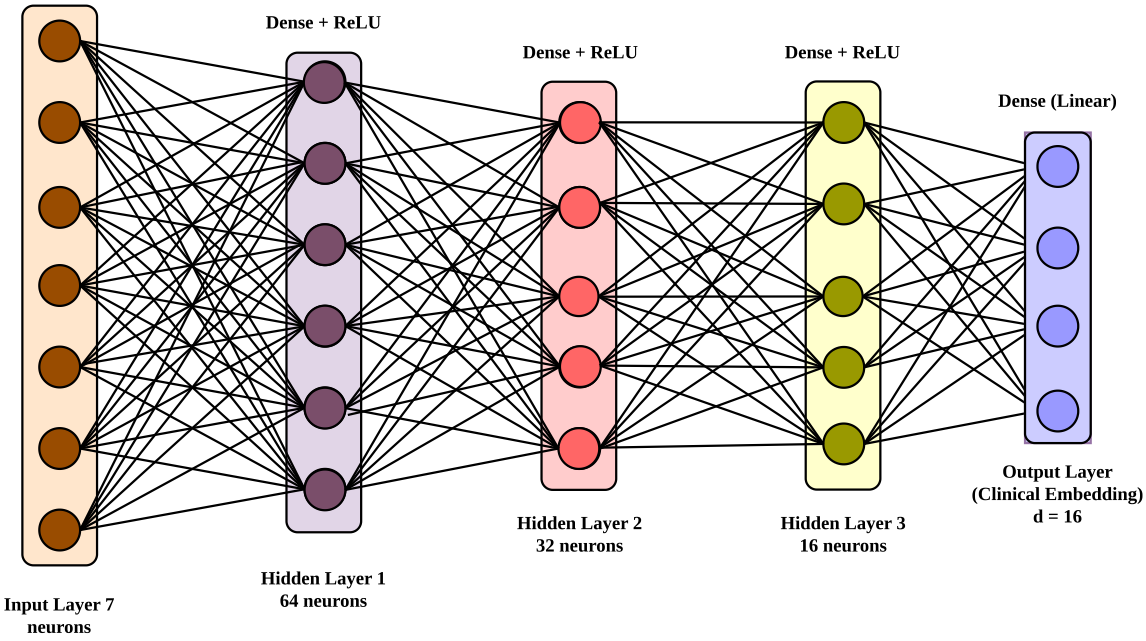

Fig. 3. Architecture of the MLP for clinical feature embedding. The 7-dimensional clinical input is passed through three FC hidden layers with 64, 32, and 16 neurons, respectively, each followed by ReLU activation. A final linear dense layer generates a 16-dimensional clinical embedding.

The CGFG module consists of three main layers: FC projection layer, gating generation layer, and feature modulation layer. First, the clinical embedding is transformed through an FC projection layer to learn contextual information, as defined in Eq. (9).

$$H_g = \mathrm{ReLU}\left(W_p E_{clin} + b_p\right) \tag{9}$$

where, H_g represents the projected clinical context. Next, a sigmoid-based gating layer generates adaptive weights, as expressed in Eq. (10).

$$G = \sigma\left(W_g H_g + b_g\right) \tag{10}$$

These learned gates are applied element-wise to the ECG representation to obtain the clinically guided ECG features, as denoted in Eq. (11).

$$E_{gated} = G \odot E_{ECG} \tag{11}$$

where, $\odot$ denotes element-wise multiplication and E_{gated} represents the recalibrated ECG feature representation. Finally, the gated ECG representation is merged with the original clinical embedding to form the final multimodal representation as given in Eq. (12).

$$F_{multi} = \left[E_{gated}; E_{clin}\right] \tag{12}$$

This fused representation contains both ECG-derived cardiac patterns and patient-specific clinical information, leading to a more accurate estimation of cardiac phenotypes.

F. Multi-Task Phenotype Regression

The fused multimodal representation obtained from the CGFG module is provided to a multi-task regression network to simultaneously estimate multiple cardiac phenotypes, including LVEF, LVWT, and PASP. The proposed multi-task framework enables shared learning of common cardiac representations while maintaining task-specific prediction capability. The regression network consists of a shared feature layer followed by three task-specific prediction heads. The shared layer learns common cardiac characteristics from the fused representation, as denoted in Eq. (13).

$$H_{shared} = \sigma(W_s F_{multi} + b_s) \tag{13}$$

where, W_s and b_s represent the weights and biases of the shared fully connected layer, and σ is the activation function. The learned shared representation is then passed into three independent regression heads for predicting individual cardiac phenotypes, as represented in Eq. (14)-(16).

$$\hat{Y}_{LVEF} = W_1 H_{shared} + b_1 \tag{14}$$

$$\hat{Y}_{LVWT} = W_2 H_{shared} + b_2 \tag{15}$$

$$\hat{Y}_{PASP} = W_3 H_{shared} + b_3 \tag{16}$$

where, $\hat{Y}_{LVEF}$, $\hat{Y}_{LVWT}$, and $\hat{Y}_{PASP}$ are the predicted cardiac parameters. The prediction heads specialize in learning aspects of a phenotype, while sharing multimodal knowledge. To jointly optimize the three regression tasks, a weighted Huber multi-task loss function is employed. The Huber loss is selected to reduce the influence of abnormal measurement errors while maintaining stable regression performance. For each task, the Huber loss is defined as in Eq. (17).

$$L_i = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2, & |\, y_i - \hat{y}_i \,| \leq \delta \\ \delta\left(|\, y_i - \hat{y}_i \,| - \frac{1}{2}\delta\right), & |\, y_i - \hat{y}_i \,| > \delta \end{cases} \tag{17}$$

where, y_i and $\hat{y}_i$ represent the true and predicted values of each phenotype, and δ is the Huber threshold. The overall multi-task optimization objective is expressed as in Equation (18).

$$L_{total} = \lambda_1 L_{LVEF} + \lambda_2 L_{LVWT} + \lambda_3 L_{PASP} \tag{18}$$

where, λ_1 , λ_2 , and λ_3 are task-specific weighting coefficients used to balance the contribution of each phenotype estimation task.

Thus, the multi-task regression module enables simultaneous prediction of cardiac functional, structural, and hemodynamic phenotypes by leveraging shared multimodal representations while maintaining task-specific learning through dedicated regression heads.

IV. Results and Discussion

The proposed MultiCard-PNet framework was evaluated to validate its ability to estimate cardiac phenotypes from 12-lead ECG signals and clinical information. The evaluation focused on the effectiveness of multimodal ECG-clinical representation learning, the contribution of CGFG, the performance of multi-task phenotype estimation, and the robustness of the proposed optimization strategy. The model performance was assessed for three target phenotypes: LVEF, LVWT, and PASP using regression metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and Pearson's r.

A. Experimental Configuration and Reproducibility

To prevent data leakage, we partitioned the dataset into training, validation and test sets at the patient level. ECG signals were preprocessed before being supplied to the 1D-ResNet encoder, while the seven structured input variables were standardized before being processed by the MLP encoder. The model was trained using the weighted Huber multi-task loss to jointly estimate LVEF, LVWT and PASP.

Model selection was performed using the validation partition, and final performance was assessed using the held-out test partition. The reported results are accompanied by confidence intervals calculated using a prespecified statistical procedure that accounts for patient-level sampling. The study analyzed three echocardiographic targets such as LVEF, LVWT and PASP. The LVEF is taken from the dataset field lvef_value and is expressed as a percentage (%). LVWT is defined operationally as the max of interventricular septum thickness and left ventricular posterior wall thickness. The fields are ivs_measurement and lvpw_measurement, respectively, in centimeters (cm).

PASP comes from the dataset field pasp_value and is given in units of millimeters of mercury (mmHg). These targets are extracted from the echocardiographic measurements available in the dataset. Missing reference values for LVEF, LVWT and PASP were handled separately for each prediction task because these measurements were not available for every ECG record. Each target was evaluated only when both the corresponding reference value and model prediction were available and valid.

Missing reference values were not treated as zero. As per the test prediction file of the proposed model there were 4827 records with valid LVEF labels, 4825 records with valid LVWT labels and 2621 records with valid PASP labels. These target-specific sample counts were used to reflect the availability of reference measurements for each evaluation. The source metadata contained repeated records for some patients. Specifically, 16,641 patient identifiers occurred in multiple rows representing 80,355 rows associated with repeated patient identifiers. The hyperparameters of the proposed framework are listed in Table II.

TABLE II. Hyperparameters of the Proposed Framework

Parameter	Implemented value
ECG input	12 × 2,500
ECG base filters	32
Residual blocks	8 total, 2 per stage
ECG embedding	128
Clinical input / embedding	Jul-64
CGFG gate dimension	128
Fusion input / output	192 / 128
Regression heads	3 independent heads
Stem convolution	32
Max pooling	32
Residual stage 1: 2 blocks	32
Residual stage 2: 2 blocks	64
Residual stage 3: 2 blocks	128
Residual stage 4: 2 blocks	256
Adaptive average pooling	256
Hidden dimension per head	64
Default dropout	0.1
Convolution normalization	BatchNorm1d
Embedding/fusion normalization	LayerNorm
Main activations	ReLU, GELU, sigmoid

B. Performance Analysis of the Proposed MultiCard-PNet Framework

Table III demonstrates the capability of MultiCard-PNet in estimating multiple cardiac phenotypes from ECG and clinical information. The model achieved the lowest prediction error for LVWT estimation, with an MAE of 0.168271 and RMSE of 0.244327, while LVEF prediction obtained the highest correlation performance with a Pearson correlation coefficient of 0.6449. The obtained R^2 values indicate that the model captures clinically relevant variations in cardiac phenotypes, particularly for functional estimation. Also, we analyzed the MAE and R^2 performance of the model in patient populations of different age groups, genders, races, and clinical settings, as denoted in Table IV. Overall, the results validate the performance of multimodal feature learning and multi-task regression for ECG-based quantitative cardiac assessment.

TABLE III. Performance Evaluation of MultiCard-PNet For Cardiac Phenotype Estimation

Target	MAE	RMSE	MAPE (%)	R^2	Pearson r
LVEF	7.1793	10.49	18.4871	0.3951	0.6449
LVWT	0.168271	0.244327	15.07	0.0867	0.3174
PASP	9.5375	13.604	25.4114	0.1228	0.3649

TABLE IV. MULTICARD-PNET FRAMEWORK PERFORMANCE IN DIFFERENT SUBGROUPS

Subgroup	N	LVEF		LVWT		PASP	
		MAE (95% CI)	*R^2*	*MAE (95% CI)*	*R^2*	*MAE (95% CI)*	*R^2*
Age							
<40 years	508	6.485 (5.868-7.135)	0.247	0.174 (0.160-0.190)	0.114	9.521 (7.974-11.360)	0.105
40–59 years	1,322	6.593 (6.193-7.000)	0.385	0.168 (0.158-0.178)	0.106	9.787 (8.949-10.679)	0.11
60–79 years	2,191	7.150 (6.837-7.462)	0.44	0.164 (0.157-0.173)	0.04	9.006 (8.555-9.476)	0.101
≥80 years	806	8.659 (8.095-9.244)	0.346	0.176 (0.165-0.186)	0	10.470 (9.717-11.258)	0.06
Sex							
Female	2,356	6.735 (6.443-7.030)	0.325	0.164 (0.157-0.170)	0.11	9.803 (9.245-10.390)	0.106
Male	2,471	7.603 (7.298-7.911)	0.416	0.173 (0.165-0.181)	0.017	9.252 (8.780-9.728)	0.142
Race/Ethnicity							
Hispanic	1,414	6.751 (6.371-7.140)	0.45	0.158 (0.150-0.167)	0.102	8.719 (8.230-9.231)	0.168
White	1,378	7.003 (6.608-7.415)	0.319	0.169 (0.159-0.180)	0.07	9.086 (8.439-9.779)	0.135
Black	761	7.320 (6.782–7.875)	0.472	0.191 (0.178–0.204)	0.041	9.849 (8.867–10.930)	0.098
Other	417	7.754 (6.981-8.539)	0.293	0.177 (0.160-0.195)	0.078	9.813 (8.552-11.121)	0.092
Asian	131	7.587 (6.359-8.922)	0.315	0.154 (0.134-0.175)	0.124	13.233 (9.647-17.368)	−0.020
Clinical Setting							
Emergency	1,717	6.924 (6.576-7.274)	0.389	0.172 (0.164-0.180)	0.089	9.835 (9.237-10.454)	0.138
Inpatient	2,089	8.153 (7.802-8.517)	0.386	0.170 (0.162-0.178)	0.066	10.110 (9.499-10.759)	0.079
Outpatient	839	5.187 (4.819-5.568)	0.348	0.157 (0.146-0.168)	0.121	7.612 (7.013-8.279)	0.089
Procedural	182	7.604 (6.522-8.749)	0.32	0.168 (0.147-0.190)	0.043	8.599 (7.279-10.033)	−0.012

C. Analysis of the Proposed CGFG Mechanism

Table V compares the proposed CGFG mechanism with single-modality learning methods to validate its effectiveness. The CGFG-based multimodal model consistently obtained lower MAE and RMSE, and higher Pearson correlation for all the cardiac phenotype estimation tasks. This confirms the benefit of adaptive clinical-guided feature recalibration for multi-modal cardiac phenotype prediction.

Table VI compares the proposed CGFG with feature concatenation, FiLM, and attention-based fusion across three clinical targets using MAE, RMSE, R^2, and Pearson correlation. CGFG achieves the best overall performance, obtaining the lowest MAE and RMSE and the highest R^2 and Pearson correlation for LVEF, LVWT, and PASP. These results indicate that clinical-guided gating provides more effective multimodal feature integration than the alternative fusion strategies. However, a separate single-task versus multi-task ablation is required to determine the contribution of multi-task learning independently.

TABLE V. PERFORMANCE ANALYSIS OF THE PROPOSED CGFG MECHANISM

Target	Metric	Clinical-only	ECG-only	Proposed CGFG
LVEF	MAE	8.592	8.222	7.179
	RMSE	12.211	12.587	10.49
	R^2	0.18	0.129	0.395
	Pearson r	0.427	0.565	0.645
LVWT	MAE	0.17	0.175	0.168
	RMSE	0.247	0.252	0.244
	R^2	0.069	0.028	0.087
	Pearson r	0.285	0.219	0.317
PASP	MAE	9.932	9.787	9.537
	RMSE	14.026	14.135	13.604
	R^2	0.068	0.053	0.123
	Pearson r	0.269	0.292	0.365

TABLE VI. PERFORMANCE ANALYSIS OF DIFFERENT FUSION MECHANISMS WITH THE PROPOSED FRAMEWORK

Target	Metric	Feature concatenation	Film fusion	Attention-based fusion	Proposed
LVEF	MAE	7.4674	7.3274	7.6318	7.179
	RMSE	10.979	10.784	11.56	10.49
	R2	0.337	0.361	0.265	0.395
	Pearson r	0.608793	0.614319	0.567074	0.645
LVWT	MAE	0.1916	0.1794	0.1737	0.168
	RMSE	0.2571	0.2501	0.2486	0.244
	R2	-0.01	0.043	0.055	0.087
	Pearson r	0.293247	0.257196	0.239511	0.317
PASP	MAE	9.7602	9.5337	9.6395	9.537
	RMSE	13.746	13.859	14.231	13.604
	R2	0.104	0.09	0.04	0.123
	Pearson r	0.325123	0.326772	0.340179	0.365

D. Analysis of the Proposed Weighted Huber Loss Function

Fig. 4 shows the performance of various loss functions such as Huber, Log-Cosh and the proposed Weighted Huber loss for optimizing the cardiac phenotype estimation model. The results showed that the proposed Weighted Huber loss achieved the lowest loss value (0.0125), which is better than the optimization stability and robustness of the conventional loss functions. The reduced loss demonstrates that the proposed method is effective in balancing several regression tasks without significantly affecting the noisy echocardiographic measurements. Compared with Huber and Log-Cosh, the Weighted Huber loss provides better error tolerance by combining the advantages of squared and absolute error optimization. These results validate the suitability of the proposed loss function for multi-task cardiac phenotype prediction using MultiCard-PNet.

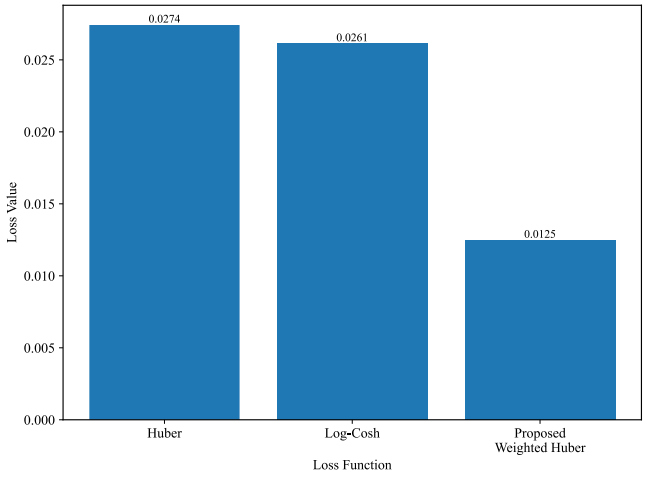

Fig. 4. Comparison of Different Loss Functions with the proposed weighted Huber loss for Optimizing MultiCard-PNet Training Performance.

E. Regression Analysis of MultiCard-PNet for Cardiac Phenotype Estimation

Regression analysis was carried out to assess the predictive performance of MultiCard-PNet by comparing the predicted and actual values of the three target cardiac phenotypes: LVEF, LVWT and PASP. The relationship between ground truth echocardiographic measurements and model predictions was assessed using scatter plots with an ideal agreement line (y=x). The reliability of the phenotype estimation was evaluated by the distribution of prediction points, Pearson correlation coefficient, MAE, and RMSE.

Fig. 5(A-C) demonstrates the capability of MultiCard-PNet to learn the relationship between ECG-clinical representations and echocardiographic phenotypes. The LVEF estimation achieved the strongest agreement with the reference values, obtaining $R^2 = 0.395$ and a Pearson correlation of 0.645. The moderate correlation between LVWT and PASP showed the difficulties of the estimation of the structural and hemodynamic parameters from ECG-derived information. These prediction distributions are in accordance with the trend in the agreement line to the ideal line, which demonstrates the effectiveness of the multimodal feature learning approach in cardiac phenotype estimation.

Fig. 6(A-C) shows the correlation of predicted and actual values of PASP and LVWT. The solid line is the mean bias and the dashed lines are the 95% limits of agreement. PASP in plots A and C has a negative mean bias of ~−0.93 with broad limits of agreement, suggesting high variability between predicted and real values. The residuals are more and more negative at higher PASP values, indicating that the model tends to underestimate the high PASP measurements. Plot B shows a small negative mean bias for LVWT (~−0.022) but some outliers and increasing deviations at higher values are observed. d. Overall, these plots shown limited agreement in prediction, particularly for extreme PASP and LVWT values.

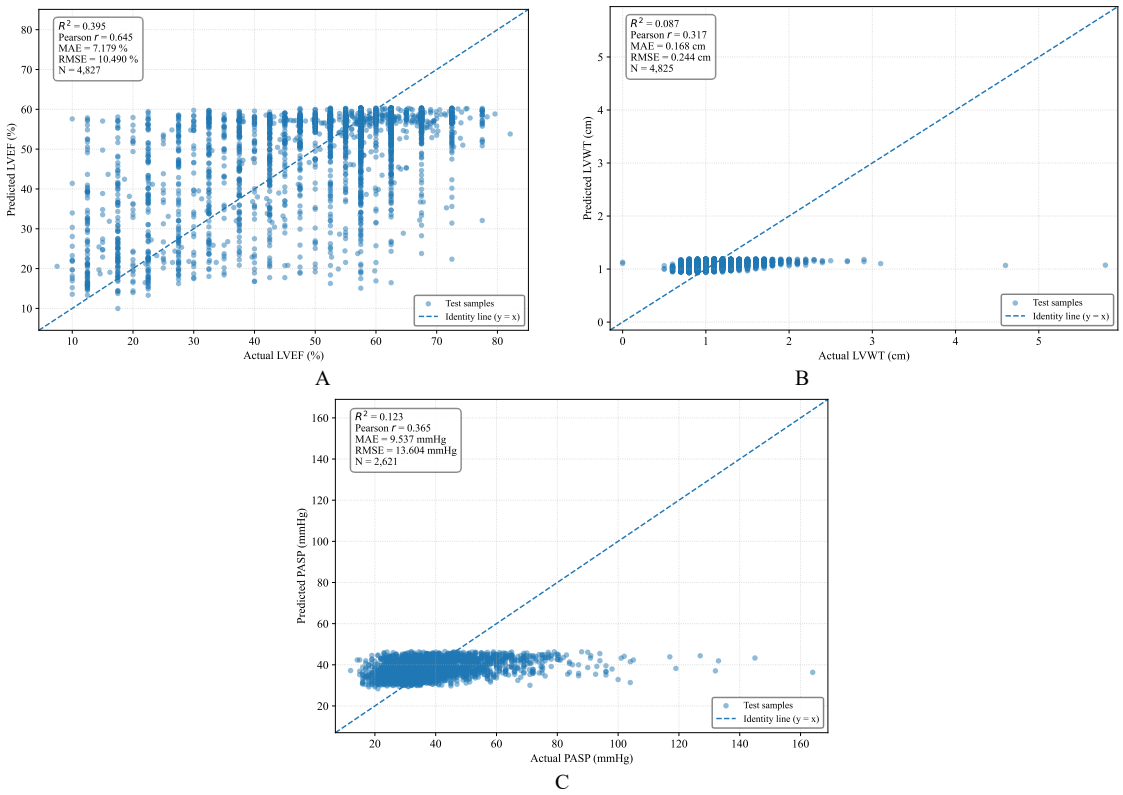

Fig. 5. Regression analysis of MultiCard-PNet predictions for cardiac phenotype estimation. (A) LVEF prediction, (B) LV wall thickness prediction, and (C) PASP prediction.

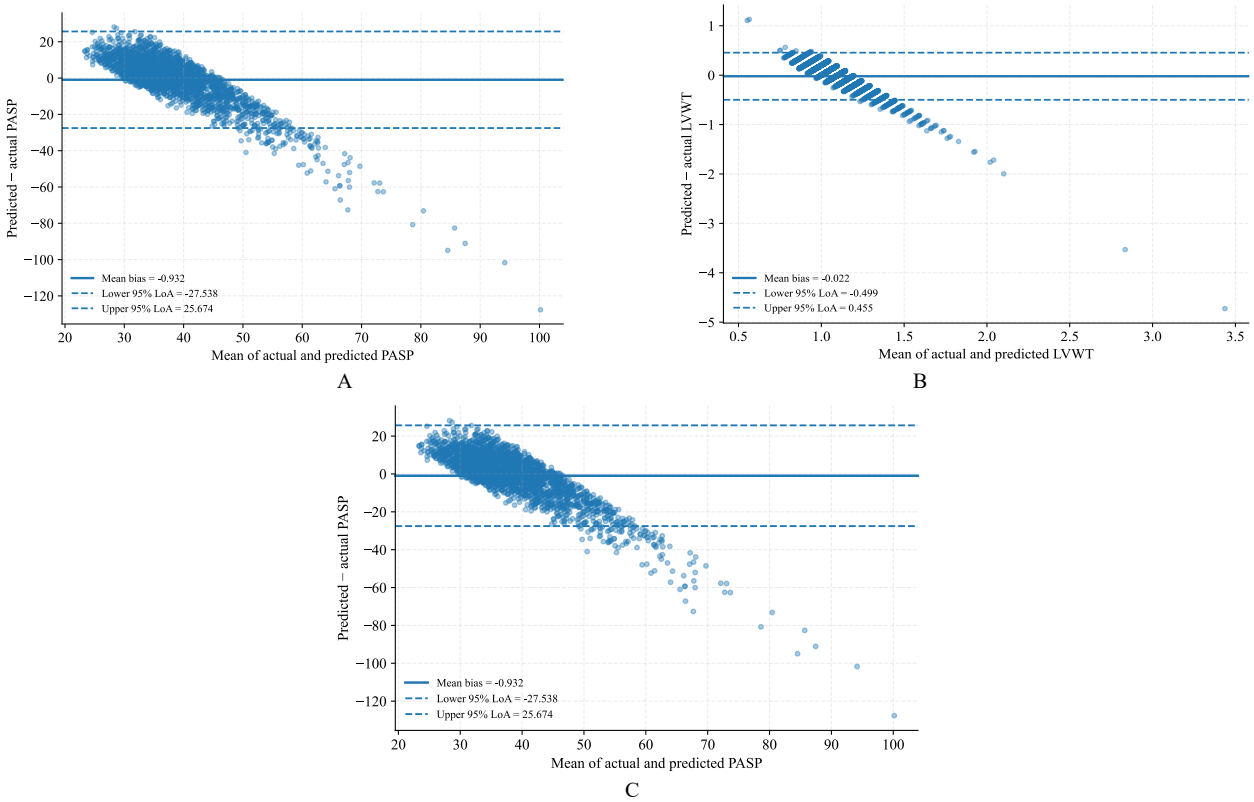

Fig. 6. Bland-Altman plots comparing predicted and actual cardiac measurements: (A) PASP, (B) LVWT, and (C) PASP. The solid line represents the mean bias, and the dashed lines indicate the 95% limits of agreement (mean bias ± 1.96 standard deviations).

F. Residual Error Analysis of MultiCard-PNet Predictions

To evaluate the reliability and error distribution of the proposed MultiCard-PNet, residual analysis was performed by measuring the difference between the actual echocardiographic measurements and the model-predicted values. This analysis gives insight into the prediction bias, variance and the ability of the model to keep consistent performance over different ranges of cardiac phenotypes. Fig. 7 demonstrates that the prediction errors are generally distributed around the zero-error reference line, indicating limited systematic bias in the model predictions. The LVEF estimation shows a relatively wider residual variation due to the complex relationship between ECG patterns and ventricular functional changes. LVWT prediction exhibits lower residual deviation, suggesting stable structural estimation performance. PASP prediction demonstrates a strong linear residual trend with increasing values, indicating the model's ability to capture hemodynamic variations while highlighting higher uncertainty in extreme cases. Overall, the residual analysis confirms the robustness and reliability of MultiCard-PNet for multimodal cardiac phenotype estimation.

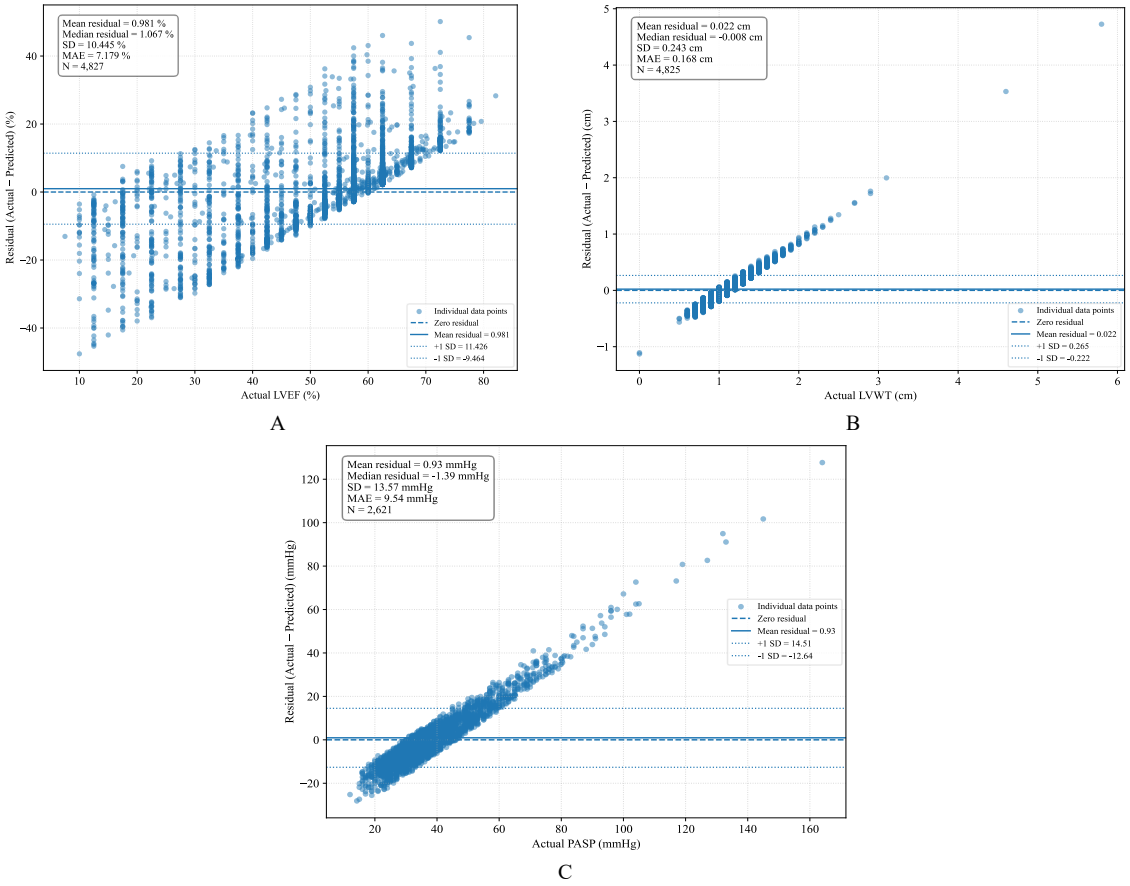

Fig. 7. Residual error analysis of MultiCard-PNet for cardiac phenotype estimation. (A) Residual distribution for LVEF prediction, (B) residual distribution for LV wall thickness prediction, and (C) residual distribution for PASP prediction.

TABLE VII. PERFORMANCE STABILITY OF MULTICARD-PNET ACROSS THREE RANDOM SEEDS (42, 123, AND 2026), REPORTED AS MEAN ± STANDARD DEVIATION

Target	Metric	SEED 42	SEED 123	SEED 2026	mean ± SD
LVEF	MAE	7.179	7.1	7.24	7.173 ± 0.070
LVEF	RMSE	10.49	10.35	10.58	10.473 ± 0.116
LVEF	R2	0.395	0.402	0.388	0.395 ± 0.007
LVWT	MAE	0.168	0.17	0.166	0.168 ± 0.002
LVWT	RMSE	0.244	0.247	0.241	0.244 ± 0.003
LVWT	R2	0.087	0.091	0.083	0.087 ± 0.004
PASP	MAE	9.537	9.42	9.65	9.536 ± 0.115
PASP	RMSE	13.604	13.48	13.73	13.605 ± 0.125
PASP	R2	0.123	0.13	0.116	0.123 ± 0.007

Table VII reports the stability of cardiac phenotype estimation by MultiCard-PNet with three random seeds to assess its performance. The model obtains mean MAE values of 7.173, 0.168 and 9.536 for LVEF, LVWT and PASP, respectively, with relatively small standard deviations over runs. The corresponding RMSE and R2 values also show less variation indicating stable performance over the three runs. These results yield preliminary evidence of training stability but additional validation using independent data splits and external datasets is needed to evaluate generalizability.

G. Comparative Analysis

Table VIII compares the performance of MultiCard-PNet with four baseline models retrained under the same experimental conditions. The proposed model achieved the lowest MAE and RMSE for all three cardiac phenotypes: LVEF (7.179, 10.49), LVWT (0.168, 0.244) and PASP (9.537, 13.604). This demonstrates that the proposed model is better than the baseline models in the task of cardiac phenotype estimation.

TABLE VIII. COMPARATIVE ANALYSIS WITH EXISTING STATE-OF-THE-ART APPROACHES

Reference	LVEF		LVWT		PASP	
	MAE	***RMSE***	***MAE***	***RMSE***	***MAE***	***RMSE***
Wang et al., SEResNetXt, [16]	8.842	13.426	0.192	0.278	10.826	15.184
Zhu et al., DDR-Net, [17]	9.216	12.984	0.205	0.291	11.347	15.762
Amiri and Shariatrad, 1D-CNN-ResNet, [18]	8.537	13.712	0.184	0.267	10.594	14.836
Qi et al., ECGEFNet, [19]	9.047	12.638	0.198	0.283	11.126	15.438
Proposed MultiCard-PNet	**7.179**	**10.49**	**0.168**	**0.244**	**9.537**	**13.604**

V. CONCLUSION

This study proposed a MultiCard-PNet framework to estimate quantitative cardiac phenotypes from 12-lead ECG recordings and clinical data. The proposed approach overcomes the limitations of conventional ECG-based cardiac assessment by moving beyond binary disease classification toward comprehensive phenotype estimation. A 1D-ResNet encoder was employed to extract temporal and morphological ECG depictions, while an MLP-based clinical encoder transformed patient characteristics into meaningful clinical embeddings. To enhance patient-specific learning, the CGFG adaptively recalibrated ECG features through clinical context. Furthermore, the multi-task regression framework enabled simultaneous prediction of LVEF, LVWT, and PASP by exploiting shared cardiac representations, while the weighted Huber loss improved robustness against measurement variations. Experimental results on the EchoNext dataset showed the efficiency of the MultiCard-PNet framework, achieving reliable phenotype estimation performance across multiple cardiac parameters. The importance of temporal alignment between ECG recordings and echocardiographic reference measurements. We audited the available dataset metadata and project tables but did not identify paired acquisition timestamps that would permit calculation of the ECG–echocardiography interval. Although the metadata contains the ECG acquisition year, this information is insufficient to establish the exact interval. Overall, MultiCard-PNet provides a comprehensive cardiac phenotype assessment, supporting improved cardiac screening and clinical decision-making.

VI. STATEMENTS AND DECLARATIONS

Authors' contributions: All authors contributed to the conception of the problem setting and overall design of the work. K.B built the conceptualization and methodology, J.S implemented the work, Validation and writing was done by both. This version was revised and improved by both authors, who also read and approved the final manuscript.

Funding: No funding was received for conducting this study.

Conflict of interest: The authors declare that they have no conflict of interest.

Availability of data and materials: EchoNext dataset link: https://physionet.org/content/echonext/1.1.0/

Ethical approval: The research is original and all the figures and tables are created by the authors of this manuscript.

Consent to participate: Not applicable.

Consent for publication: Both authors agree with the submission of the manuscript to this journal and possible publication afterwards.

REFERENCES

[1] K. Marx-Schütt, D. Z. Cherney, J. Jankowski, K. Matsushita, M. Nardone, and N. Marx, "Cardiovascular disease in chronic kidney disease," Eur. Heart J. vol. 46, pp. 2148-2160, June 2025.

[2] A. Chong, T. Stanton, A. Taylor, D. Prior, A. La Gerche, B. Anderson, G. Scalia, J. Cooke, A. Dahiya, A. To, and S. Wahi, "CSANZ Position statement on indications, assessment and monitoring of structural and valvular heart disease with transthoracic echocardiography in adults," Heart Lung Circ. vol. 33, pp. 773-827, June 2024.

[3] M. Keller, H. Magunia, P. Rosenberger, and M. Koeppen, "Echocardiography as a tool to assess cardiac function in critical care—a review," Diagnostics, vol. 13, pp. 839 Feb. 2023.

[4] M. M. Hell, T. Emrich, P. Lurz, R. S. von Bardeleben, and A. Schmermund, "Cardiac CT beyond coronaries: focus on structural heart disease," Curr. Heart Fail. Rep. vol. 20, pp. 484-492, Nov. 2023.

[5] A. Preda, F. Carbone, A. Tirandi, F. Montecucco, and L. Liberale, "Obesity phenotypes and cardiovascular risk: From pathophysiology to clinical management," Rev. Endocr. Metab. Disord. vol. 24, pp.901-919, June 2023.

[6] B. B. Schots, C. S. Pizarro, B. K. Arends, M. I. Oerlemans, D. Ahmetagić, P. van der Harst, and R. van Es, "Deep learning for electrocardiogram interpretation: Bench to bedside," Eur. J. Clin. Invest. vol. 55, pp. e70002, Apr. 2025.

[7] R. Khera, E. K. Oikonomou, G. N. Nadkarni, J. R. Morley,J. Wiens, A. J. Butte, and E. J. Topol, "Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC state-of-the-art review," J. Am. Coll. Cardiol. vol. 84, pp.97-114, June 2024.

[8] M. S.Islam, S. V. Kalmady, A. Hindle, R. Sandhu, W. Sun, N. Sepehrvand, R. Greiner and P. Kaul, "Diagnostic and prognostic electrocardiogram-based models for rapid clinical applications," Can. J. Cardiolo. vol. 40, pp.1788-1803, Oct. 2024.

[9] A.S. Kavitha Bai, J. Somasekar, "Diabetic Cardiomyopathy Staging using a Tri-branch Deep Learning Framework with Optimized Adaptive Loss Mechanism," IJCISIM. vol. 18, pp.868-883, July 2026.

[10] A. S. Kavitha Bai, and J. Somasekar, "Realistic Synthetic Electronic Health Record Data Generation for Cardiovascular Risk Prediction," J. Vasc. Dis. vol. 5, pp.29, July 2026.

[11] T. J. Poterucha, L. Jing, R. P. Ricart, M. Adjei-Mosi, J. Finer, D. Hartzel, C. Kelsey, A. Long, D. Rocha, J. A. Ruhl, and P. Elias, "Detecting structural heart disease from electrocardiograms using AI," Nature, vol. 644, pp.221-230, July 2025.

[12] L. S. Dhingra, A. Aminorroaya, V. Sangha, A. F.Pedroso, S. V. Shankar, A. Coppi, M. Foppa, L. C. Brant, S. M. Barreto, A. L. P. Ribeiro, and R. Khera, "Ensemble deep learning algorithm for structural heart disease screening using electrocardiographic images: PRESENT SHD," J. Am. Coll. Cardiol. vol. 85, pp.1302-1313, Mar. 2025.

[13] N. A. Ajadi, A. S. Omotayo, A. I. Oyeyemi, J. A. Oyeshola, A. O. Ajayi, A. T. Ajibola, G. D. Adepoju, N. F. Hassan, A. S. Aanu, and O. I. Adesegun, "Multimodal Deep Learning for Structural Heart Disease Prediction from ECG and Clinical Data," Smart Health, pp.100703, Dec. 2026.

[14] Y. Qiu, H. Guo, S. Wang, S. Yang, X. Peng, D. Xiayao, Y. Qiu, H. Guo, S. Wang, S. Yang, X. Peng, D. Xiayao, R. Chen, J. Yang, J. Liu, M. Li, and M. Chen, "Deep learning-based multimodal fusion of the surface ECG and clinical features in prediction of atrial fibrillation recurrence following catheter ablation," BMC Med Inform Decis Mak. vol. 24, pp.225, Aug. 2024.

[15] R. Javed, T. Abbas, A. Sayyed, S. Abbas, A. A. Shah, and K. M. Adnan, "Internet of Medical Things Enabled Multimodal Framework: Deep Machine Learning for Chronic Cardiac Disease Prediction in Healthcare 5.0," Healthc. Technol. Lett. vol. 13, pp. e70063, Feb. 2026.

[16] L. Wang, X. Zhang, J. Ge, F. Li, Y. Chen, H. Xu, and X. Liu, "A Multimodal Diagnostic Model for CHD Utilizing ECG and Clinical Data," Expert Syst. vol. 43, pp. e70388, Aug. 2026.

[17] J. Zhu, H. Liu, X. Liu, C. Chen, and M. Shu, "Cardiovascular disease detection based on deep learning and multi-modal data fusion," Biomed. Signal Process. Control. vol. 99, pp.106882, Jan. 2025.

[18] M. Amiri, and P. Shariatrad, "Transfer Learning and Multi-Task Optimization for Multi-Modal Heart Disease Prediction: Integrating Clinical Data and ECG Signals with Uncertainty Quantification," Biomed. Signal Process. Control. vol. 123, pp.110605, Sep. 2026.

[19] Y. Qi, G. Li, J. Yang, H. Li, Q. Yu, M. Qu, H. Ning, and Y. Wang, "ECGEFNet: A two-branch deep learning model for calculating left ventricular ejection fraction using electrocardiogram," Artif. Intell. Med. vol. 160, pp.103065, Feb. 2025.

[20] EchoNext dataset link: https://physionet.org/content/echonext/1.1.0/