A Hybrid Adaptive Multiscale ECG/PCG Fusion Network with Bidirectional Cross-Modal Attention for Automated Cardiac Disease Detection on Wearable Devices

Yijing Wang[1], Tiancheng Cao[1,2], Tong Zou[1], Haining Zhao[1], and Hen-Wei Huang[1,3]

Abstract— **Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality worldwide, underscoring the need for accurate and early diagnostic tools. Electrocardiogram (ECG) and phonocardiogram (PCG) signals are routinely acquired in clinical practice and provide complementary electrophysiological and mechanical information about cardia function. However, most existing automated diagnostic systems rely on a single modality, or employ fixed fusion strategies, limiting their robustness across diverse cardiac conditions. This paper proposes a Hybrid Adaptive Multiscale Cardiac Fusion Network (HAM-CFN) for multi-class cardiac disease classification by jointly modeling ECG and PCG signals. The network integrates multiscale temporal and time–frequency feature extraction, bidirectional cross-modal attention for physiological interaction, and an adaptive dynamic fusion module that assigns sample-specific modality weights. Experiments on the PhysioNet/CinC Challenge 2016 dataset demonstrate that the proposed method outperforms conventional single-modality and traditional fusion approaches with a lightweight architecture, achieving an accuracy of 97.60%, a precision of 97.66%, a specificity of 99.39%, a recall of 97.69%, and an F1-score of 97.67%. These results suggest that adaptive multimodal fusion of ECG and PCG signals can enhance automated cardiac diagnosis in resource constrained and real-world clinical settings.**

Keywords—**ECG, PCG, multimodal, deep learning, cardiac disease.**

I. Introduction

Cardiovascular diseases (CVDs) are the leading cause of global mortality, accounting for 17.9 million deaths in 2019, or 32% of all deaths worldwide [1]. Clinical management of CVDs leverages a suite of tools, ranging from non-invasive electrocardiograms (ECG) and phonocardiograms (PCG) to more advanced imaging, to form a comprehensive assessment [2, 3]. Recent advances in wearable technology enable continuous acquisition of ECG and PCG signals outside clinical settings, providing a noninvasive, cost-effective, and accessible alternative to resource-intensive hospital-based diagnostics [4, 5]. Such continuous monitoring is crucial for early detection of cardiac abnormalities and postoperative surveillance, overcoming the limitations of intermittent clinical examinations.

Currently, most wearable devices focus on single-modality signals [6-13].These wearable health technologies provide methods for continuous and accessible cardiac monitoring outside hospitals, but the lack of multiple indicators is insufficient for comprehensive cardiac assessment. ECG analyses electrical activity of heart, which is effective for identifying electrical abnormalities, while PCG captures mechanical or acoustic activity that can identify systolic and diastolic murmurs for heart valve abnormalities identification [14-16]. However, the diagnostic accuracy of wearable devices remains inferior to clinical standards, partly due to the limited information from single modalities [17]. Therefore, it is meaningful to fuse ECG and PCG signals in wearable systems, combining their complementary characteristics to approach the diagnostic reliability of hospital examinations and enable robust, automated detection of diverse CVDs.

The emergence of recent multimodal studies has made deep learning–based approaches for fusing ECG and PCG signals a major direction of exploration. Liu et al. [18] proposed a 1-D vision transformer with co-attention and dynamic weighted fusion to integrate ECG and PCG for coronary artery disease detection, but it was limited to binary classification and short 0.8-s segments, reducing clinical interpretability. Lin et al. [19] employed a multiscale CNN-BiLSTM for wearable multimodal fusion, yet failed to distinguish clinically important arrhythmia subtypes. Zhang et al. [20] introduced a co-learning–assisted dense fusion framework to handle feature interaction and missing modalities, but its high computational cost hinders deployment on resource-constrained wearable devices. Overall, existing ECG–PCG fusion methods are often computationally heavy, focused on limited disease types, and rely on shallow fusion strategies (e.g., feature concatenation or decision-level fusion), which restricts their clinical applicability and scalability in wearable systems [21].

To address these challenges, this work proposes a Hybrid Adaptive Multiscale Cardiac Fusion Network (HAM-CFN) for five-class cardiac disease detection using ECG and PCG signals. Specifically, HAM-CFN is designed to jointly model multi-scale cardiac representations, cross-modal physiological interactions, and adaptive modality fusion, enabling accurate and robust diagnosis in wearable settings. The proposed architecture consists of three key components: a Multi-Scale Cardiac Encoder (MSCE) that simultaneously extracts 1D temporal features from raw ECG and PCG signals and 2D time–frequency features from ECG scalograms and PCG spectrograms; a Bidirectional Cross-Modal Interaction (BCMI)

This work is supported by the Nanyang Assistant Professorship and the A*STAR Programmatic Seed Fund (M24N9b0130).

Corresponding author: Tiancheng Cao (tiancheng.cao@ntu.edu.sg).

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

[2] Department of Emergency Medicine, Brigham and Women's Hospital, Boston, USA.

[3] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

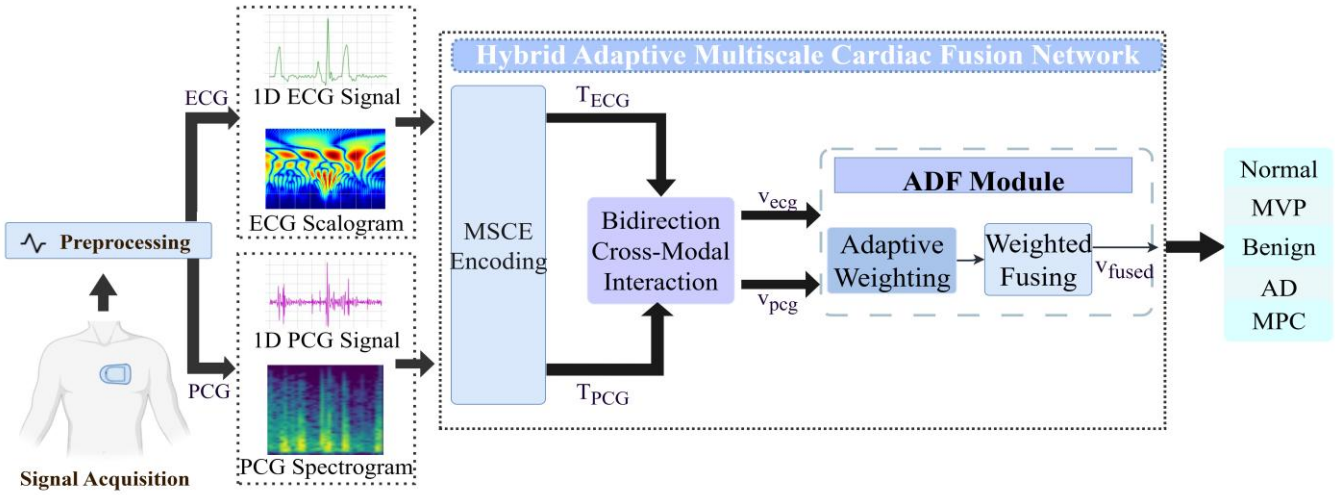

Fig. 1. Overall framework of proposed HAM-CFN.

module that employs multi-head attention to enable deep and symmetric feature exchange between ECG and PCG; and an Adaptive Dynamic Fusion (ADF) module that generates sample-specific fusion weights to balance the contributions of each modality. By integrating complementary electrical and mechanical cardiac information in a unified and computationally efficient manner, the proposed approach achieves reliable multi-class cardiovascular abnormality detection while remaining suitable for resource-constrained devices. Overall, the proposed method provides an efficient and automated solution for robust multi-class cardiac disease detection, achieving more reliable and clinically meaningful diagnosis through complementary ECG–PCG fusion.

II. Methodology

The overall structure of the proposed HAM-CFN is demonstrated in Fig. 1. Initially, the synchronized ECG and PCG signals are preprocessed, including normalization, noise reduction, augmentation, and appropriate segmentation. The processed signals are then transformed into both 1D and 2D representations, which are encoded by the Multi-Scale Cardiac Encoder (MSCE), detailed in Fig. 2a. Subsequently, the Bidirectional Cross-Modal Interaction (BCMI) module facilitates deep fusion and mutual enhancement of ECG and PCG features (Fig. 2b). Finally, the Adaptive Dynamic Fusion (ADF) module generates learnable fusion weights, and the fused features are passed through a compact classifier to achieve cardiac disease identification across five categories (Fig. 2c).

A. Multi-Scale Cardiac Encoder (MSCE)

As in Fig. 2a, first, the convolution-based tokenization was performed on 1D temporal signals of ECG and PCG. We employed a 1D convolution to transform raw signals into 64 token sequences. The convolution uses a kernel size of 32, stride of 31, padding of 15, and produces 64 output channels. Following convolution, the sequence length was adjusted to 64 via center-cropping or zero-padding to generate the 1D token representation $F_{1D} \in \mathbb{R}^{64\times64}$.

Second, for 2D ECG wavelet scalograms and PCG spectrograms, a lightweight convolutional layer was performed for feature extraction. The layer received time-

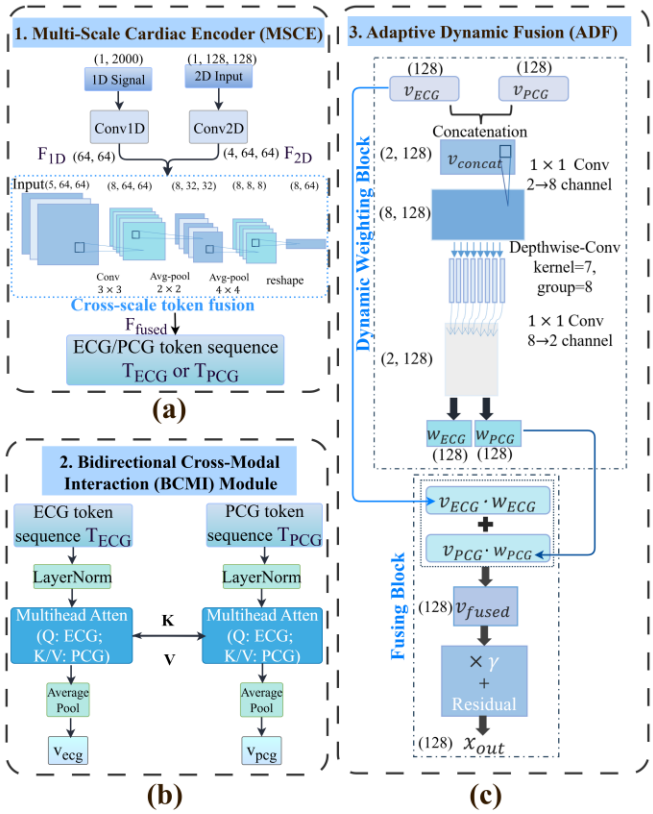

Fig. 2. Workflow of separate modules in the HAM-CFN network.

frequency images $x_{2D} \in \mathbb{R}^{1\times128\times128}$ and generated 4-channel feature maps $F_{2D} \in \mathbb{R}^{4\times64\times64}$.

Finally, a cross-scale token fusion module was used for nonlinear interaction of multi-scale information. The 1D and 2D tokens were initially concatenated along the channel dimension to form a combined token set $F_{fused} \in \mathbb{R}^{5\times64\times64}$:

$$F_{fused} = Concat(F_{1D}, F_{2D}) \quad (1)$$

Afterwards, a 3×3 convolution mapped the 5-channel features to 8 channels, achieving nonlinear combination of cross-scale features. After convolution, 2×2 average pooling was applied to reduce spatial dimensions from 64×64 to 32×32, extracting local invariant features. Adaptive average

pooling reduced feature maps to 8×8 spatial size, which were then reshaped into final multiscale token sequence $T_{ECG}, T_{PCG} \in \mathbb{R}^{8\times64}$ for separate ECG and PCG channels. This design achieved hierarchical abstraction balancing both local and global features.

These multiscale tokens formulated the ECG token sequence and PCG token sequence respectively.

B. Bidirectional Cross-Modal Interaction (BCMI) Module

Unlike common early or late fusion across models, which rely on simple concatenation or weighted voting without interactions between their inherent information, the BCMI module enabled deep interaction and mutual enhancement of ECG and PCG features.

After receiving the token sequences of ECG and PCG from the previous MSCE, which were layer-normalized, this module applied a bidirectional multi-head attention mechanism. The main idea was that each modality acted as both query and context. ECG tokens attended to PCG tokens, and vice versa. ECG's attention to PCG ($A_{ECG\to PCG}$), and PCG's attention to ECG ($A_{PCG\to ECG}$), are as depicted by equations:

$$A_{ECG\to PCG} = MultiHeadAtt(T_{ECG}, T_{PCG}, T_{PCG}) \quad (2)$$

$$A_{PCG\to ECG} = MultiHeadAtt(T_{PCG}, T_{ECG}, T_{ECG}) \quad (3)$$

And the multi-head attention was computed as:

$$MultiHeadAtt(Q,K,V) = Concat(head_1, \dots, head_h)W^O \quad (4)$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (5)$$

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

Here, h is the number of heads, d_k is the head dimension, W^O is the learnable weight matrix that combines the concatenated outputs from all attention heads, and W_i^Q, W_i^K, W_i^V are learned projection matrices that project the input embeddings into different representation subspaces for each attention head. As input vectors to the attention mechanism, Q represents "query" that seeks information, K represents "key" that are matched against queries, and V contains the actual "values" to be aggregated based on attention weights. This symmetrical attention allows the model to learn how information in one modality can enhance and refine the features in the other modality.

The attention output of each attention block was then added back to original tokens via residual connections, followed by passing through a feed-forward network. Separate ECG and PCG attention tokens were generated, as described in Fig. 2b. The outputs of size (8, 64) were passed through an average pooling followed by flattening, denoting as $v_{ECG}, v_{PCG} \in \mathbb{R}^{128}$.

C. Adaptive Dynamic Fusion (ADF) Module

To fuse the two feature vectors v_{ECG}, v_{PCG} after the BCMI module, instead of traditional fixed-weight fusion methods, ADF generates sample-specific fusion weights, which combines a dynamic, sample-adaptive weighting mechanism with a ConvNeXt-inspired depthwise convolutional structure, as seen in Fig. 1.

ECG and PCG features were first concatenated and reshaped to form a feature matrix v_{concat}:

$$v_{concat} = [v_{ECG}, v_{PCG}] \in \mathbb{R}^{2\times128} \quad (7)$$

where ECG and PCG features served as 2 channels.

Afterwards, a 3-layer convolutional block was designed to generate dynamic weights, as shown in Fig. 2c. The block contained a 1×1 pointwise convolution, a 7×7 depthwise separable convolution, and followed by a 1×1 pointwise convolution compression, resulting in $W_{raw} \in \mathbb{R}^{2\times128}$. Although the kernel size seemed large, depthwise separable convolution significantly reduced computational complexity.

To ensure the weights have probabilistic interpretability, we applied layer normalization and $softmax$ to the generated raw weights:

$$W = softmax(LayerNorm(W_{raw}), dim = 1) \in \mathbb{R}^{2\times128} \quad (8)$$

Here, $softmax$ is applied along the channel dimension (dimension 1), ensuring that for each feature position, the sum of ECG and PCG weights is 1. The two channels of W accordingly represent dynamic weights for each modality w_{ECG} and w_{PCG}. The generated weights were used to perform the sample-dependent weighting of the original features:

$$v_{fused} = w_{ECG} \cdot v_{ECG} + w_{PCG} \cdot v_{PCG} \in \mathbb{R}^{128} \quad (9)$$

After dynamic weighted fusion, the fused feature was stabilized through layer normalization. A learnable scaling factor $\gamma \in \mathbb{R}^{128}$(initialized to 1×10^{-6}) was used to adjust the output and added to the original fused features via residual connection, outputting the final transformed features $x_{out} \in \mathbb{R}^{128}$. The final fused feature x_{out} was passed to a small multi-layer perceptron (MLP) classifier to produce 5-class detection.

III. Experiments

A. Dataset

The training-a dataset of PhysioNet/CinC Challenge 2016 database (PCHSD2016) was used in this study, which is the MIT Heart Sounds Database (MITHSDB) [22]. It contains 409 synchronized PCG and ECG recordings, sampled at 2000 Hz. Each recording is clinically annotated into one of five categories: Normal, Mitral Valve Prolapse (MVP), Benign murmurs (Benign), Aortic Disease (AD), or Miscellaneous Pathological conditions (MPC). Following expert quality assessment, 17 noisy recordings were excluded, resulting in 388 clean samples used for model development and evaluation, as listed in Table I.

B. Implementation Details

Signal preprocessing was performed on the dataset. Raw synchronized ECG and PCG signals were preprocessed to counter the unstable signal quality. ECG denoising involved fourth-order Butterworth bandpass (0.5–100 Hz), powerline interference suppression, baseline wander removal, entropy-based Wiener filter [14], and z-score normalization. PCG signals were bandpass filtered (25–400 Hz) and processed with

TABLE I . DATASET DESCRIPTION AND SAMPLE DISTRIBUTION

Subject Type	Number of Used Recordings	Number of Samples	Samples after Augmentation
Normal	116	8896	8896
MVP	122	9659	9659
Benign	114	8559	8559
AD	13	907	8000
MPC	23	1844	8000

spike removal [23], followed by normalization. All signals were resampled to 1000 Hz and segmented into 2-second fragments centered at ECG R-peaks to ensure signal alignment between two modes and physiological relevance. To address severe class imbalance, targeted data augmentation was applied to the minority classes (AD and MPC) via time shifting, warping, scaling, and noise addition. The final dataset contains 43,114 samples with approximately balanced distributions as shown in Table I. In addition to 1D time-domain signals, time-frequency representations were generated to capture non-stationary cardiac characteristics. ECG scalograms were generated via continuous wavelet transform (CWT), focusing on clinically relevant low-to-mid frequencies. While PCG spectrograms were computed with STFT to capture broader sound patterns. All 2D representations were log-scaled, normalized, and resized to 128×128 pixels for model input. The final balanced dataset was split into 80% training and 20% testing set. To ensure a realistic evaluation, a patient-level split is adopted, such that all segments from the same subject appear exclusively in either the training or test set.

Experiments were implemented based on Python 3.9 with PyTorch. The experimental system consists of an Intel(R) Core(TM) Ultra 9 275HX CPU (32 GB RAM) and an NVIDIA GeForce RTX 5070 Ti Laptop GPU. The proposed network employed multi-head attention with four heads and an embedding dimension of 64, achieving a balance between representational capacity and computational efficiency. Detailed hyperparameter settings are summarized in Table II. The AdamW optimizer with an initial learning rate of 1×10^{-3} was adopted, together with a warmup and cosine annealing schedule. Training stabilization strategies, including dynamic label smoothing, gradient clipping, and early stopping, were applied. The overall number of parameters of the model is 110,007, reflecting a compact architecture and efficient computation.

IV. RESULTS AND DISCUSSIONS

Quantitative results in Table III demonstrate that the proposed HAM-CFN model achieves consistently strong performance across all evaluation metrics, outperforming existing methods evaluated on the same PCHSD2016 training-a dataset. Specifically, our model achieved an accuracy of 97.60%, precision of 97.66%, recall of 97.69%, specificity of 99.39%, and F1 score of 97.67%. Notably, while most prior studies focused on binary classification, the proposed approach addresses a more challenging five-class classification task, where inter-class similarity and decision boundaries are more complex. Achieving comparable or even superior performance under this setting indicates the strong discriminative capability and robustness of the proposed framework. The superior performance in multi-class scenarios can be attributed to the complementary design and mutually

TABLE II. MAIN NETOWKR AND TRAINING HYPERPARAMETERS

Hyperparameter	Value
Number of Attention Heads	4
Embedding Dimension	64
Dropout	0.5
Optimizer	AdamW
Initial Learning Rate	1e-3
Learning Rate Scheduler	Warmup + CosineAnneal
Batch Size	16
Epochs	200
Early Stopping Patience	30 epochs
Label Smoothing	0.1
Gradient Accumulation	4 steps
Max Grad Norm	1.0

reinforcing work of the three key modules, which play complementary and mutually reinforcing roles. First, MSCE provides robust multiscale representations by jointly encoding temporal waveforms and time–frequency patterns, ensuring feature completeness within each modality. This multiscale representation is particularly beneficial for distinguishing disease with subtle yet clinically meaningful differences. Building upon these representations, BCMI enables bidirectional cross-modal attention, allowing ECG and PCG to refine each other through physiologically meaningful interactions rather than treating modalities independently or shallow fusion. It is critical for multi-class classification where certain disease categories may be more dominant in one modality. Finally, ADF performs sample-adaptive fusion, dynamically balancing the relative contributions of ECG and PCG at the decision level to further accommodate inter-patient and inter-disease variability. This adaptive mechanism mitigates the bias toward dominant modalities and contributes to balanced performance across all classes, especially under long-tail class distributions. Above these, the three components form a hierarchical fusion pipeline that integrates multiscale encoding, cross-modal interaction, and adaptive decision fusion, which collectively contribute to the robustness and accuracy of the proposed framework.

Moreover, HAM-CFN maintains a lightweight architecture with 110 k parameters. More computational comparisons are listed in Table III, where the number of parameters denoted with * are estimated from their models. Compared with Transformer-based multimodal methods that typically exceed one million parameters, the proposed model achieves higher efficiency without sacrificing accuracy. Although the parameters in [24] is fewer than our model, significantly better metrics have achieved by our work. The lightweight characteristic makes our model more suitable for deployment on portable cardiac surveillance devices and clinical usage.

The confusion matrix in Fig. 3 further illustrates the robustness of the proposed method. Overall, high recognition accuracy is achieved across most categories, indicating that the multimodal fusion strategy effectively captures complementary diagnostic cues from ECG and PCG signals. Certain confusion is observed between clinically similar classes, such as Benign and MVP, which is consistent with their overlapping acoustic characteristics. Nevertheless, the integration of multimodal features helps mitigate this ambiguity and improves classification robustness, particularly for minority classes augmented during training.

TABLE III. PERFORMANCE COMPARISONS OF MODELS TRAINED ON CINC 2016 TRAINING-A DATASET

Methods	Model		N_{class}	Results					$N_{parameters}$
	Input	Method		Acc (%) ↑	Pre (%) ↑	Spe (%) ↑	Recall (%) ↑	F1 (%) ↑	
Liu et al. [18]	1D ECG and PCG	1D vision Transformer	2	94.87	95.40	98.51	92.33	93.60	4.38 M
Zhang et al. [20]	1D ECG and PCG	CNN	2	94.41	-	93.97	94.85	93.60	23.4 M *
AI-Issa et al. [24]	1D time and frequency domain FFT signal.	CNN-LSTM	2	93.76	-	92.42	**99.63**	-	**28 k**
Zhu et al. [25]	1D ECG and PCG	SVM-RNN	2	91.6	-	91.1	92.0	91.5	174 k
Kalatehjari et al. [26]	1D ECG and PCG	CNN-BiLSTM	2	**97.77**	96.78	-	96.73	**97.68**	4.72 M *
Li et al. [27]	1D ECG and PCG	CNN-LSTM	2	87.3	-	84.5	90.3	87.4	0.3 M *
Li et al. [2]	1D ECG and PCG, 2D time-frequency scalograms (CWT).	BiLSTM-GoogLeNet	2	96.13	96.04	90.8	98.48	97.24	2.1 M *
Wang et al. [28]]	1D ECG and PCG	RNN	2	96.1	96.6	-	95.9	96.1	1.6 M *
Marocchi et al. [29]	1D ECG and PCG, 2D spectrograms (STFT), Mel-spectrograms, scalograms (CWT).	InceptionV3	2	92.25	90.77	70.00	98.33	94.40	23.8 M *
Ours	1D ECG and PCG, ECG scalogram, PCG spectrogram.	HAM-CFN	**5**	97.60	**97.66**	**99.39**	97.69	97.67	110 k

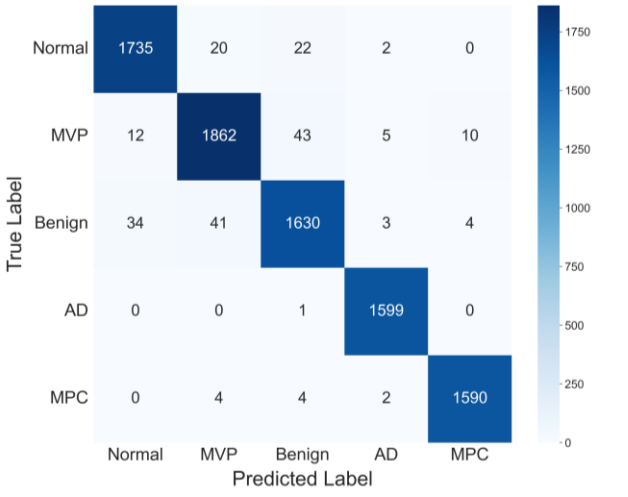

Fig. 3. Confusion matrix of network result.

The results could demonstrate that the fusion of 1D (raw signal) and 2D (spectrogram/scalogram) features contributes to higher robustness and generalization. For their complementary nature, 1D signals preserve precise temporal relationships, such as RR and PR intervals, which are crucial for arrhythmia detection, while 2D time-frequency diagrams reveal time-varying patterns of frequency components, essential for murmur classification and pathological staging. Additionally, certain CVD could manifest as changes in the time-domain waveform segment, when specific frequency band energy changes also occur in the time-frequency domain. Such correlations could be effectively captured by the proposed MSCE. The MSCE also addresses the cross-temporal scale process that cardiac electrical activity changes on a millisecond scale according to action potentials, whereas mechanical activity occurs on a larger scale of time for valve opening and closing. These multiscale characteristics are deeply combined through the convolutional operations in this module. Consequently, the proposed multimodal fusion framework provides an effective and efficient solution for automated multi-class cardiovascular disease classification.

V. CONCLUSION

This paper proposes a novel multimodal fusion framework, HAM-CFN, for the classification of five cardiac disease categories from synchronized ECG and PCG signals. Three innovative modules constitute the network structure, including MSCE, BCMI, and ADF. The designed multimodal fusion strategies address key limitations of existing multimodal cardiac analysis methods, including inadequate fusion of electrical and mechanical cardiac information, lack of multiscale analysis, and high computational complexity. Using training-a dataset from the PhysioNet/CinC Challenge 2016 database, HAM-CFN achieves an accuracy of 97.60%, specificity of 99.39%, and F1 score of 97.67%, outperforming existing multimodal studies. Despite these advances, the generalization of the model requires validation on larger and more diverse datasets, and further research should expand the classification scope to include additional cardiac conditions. Moreover, enhancing the interpretability of model decisions remains essential for clinical applications, necessitating the development of visualization tools for model attention.

REFERENCES

[1] World Health Organization. "Cardiovascular Diseases." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed 9 October, 2025).

[2] J. Li, L. Ke, Q. Du, X. Ding, and X. Chen, "Research on the Classification of ECG and PCG Signals Based on BiLSTM-GoogLeNet-DS," *Applied sciences,* vol. 12, no. 22, p. 11762, 2022, doi: 10.3390/app122211762.

[3] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques," *ICT express,* vol. 8, no. 1, pp. 109-116, 2022, doi: 10.1016/j.icte.2021.08.021.

[4] T. Cao *et al.*, "Cybersecure End-to-End FPGA-Accelerated ECG Monitoring for Precision Diagnosis With Personalized CWT and Adversarial Defense," *IEEE journal of biomedical and health informatics,* vol. 29, no. 12, pp. 8863-8870, 2025, doi: 10.1109/JBHI.2025.3617826.

[5] T. Cao, W. S. Ng, W. L. Goh, and Y. Gao, "DWT-PoolFormer: Discrete Wavelet Transform-based Quantized Parallel PoolFormer Network Implemented in FPGA for Wearable ECG Monitoring," 2024: IEEE, pp. 1-5, doi: 10.1109/BioCAS61083.2024.10798386.

[6] D. Lai, Y. Bu, Y. Su, X. Zhang, and C. S. Ma, "A Flexible Multilayered Dry Electrode and Assembly to Single-Lead ECG Patch to Monitor Atrial Fibrillation in a Real-Life Scenario," *IEEE Sensors Journal,* vol. 20, no. 20, pp. 12295-12306, 2020, doi: 10.1109/JSEN.2020.2999101.

[7] S. Y. Lee, Y. W. Hung, P. H. Su, I. P. Lee, and J. Y. Chen, "Biosignal Monitoring Clothing System for the Acquisition of ECG and Respiratory Signals," *IEEE Access,* vol. 10, pp. 66083-66097, 2022, doi: 10.1109/ACCESS.2022.3183968.

[8] Y. Bu, M. F. U. Hassan, and D. Lai, "The Embedding of Flexible Conductive Silver-Coated Electrodes into ECG Monitoring Garment for Minimizing Motion Artefacts," *IEEE Sensors Journal,* vol. 21, no. 13, pp. 14454-14465, 2021, doi: 10.1109/JSEN.2020.3001295.

[9] B. M. Li *et al.*, "Influence of Armband Form Factors on Wearable ECG Monitoring Performance," *IEEE Sensors Journal,* vol. 21, no. 9, pp. 11046-11060, 2021, doi: 10.1109/JSEN.2021.3059997.

[10] F. Zeng, Y. Lin, P. Siriaraya, D. Choi, and N. Kuwahara, "Emotion Detection Using EEG and ECG Signals from Wearable Textile Devices for Elderly People," *Journal of Textile Engineering,* vol. 66, no. 6, pp. 109-117, 2020, doi: 10.4188/jte.66.109.

[11] S. K. Bashar *et al.*, "Feasibility of atrial fibrillation detection from a novel wearable armband device," *Cardiovascular Digital Health Journal,* vol. 2, no. 3, pp. 179-191, 2021/06/01/ 2021, doi: https://doi.org/10.1016/j.cvdhj.2021.05.004.

[12] N. Giordano, S. Rosati, G. Balestra, and M. Knaflitz, "A Wearable Multi-Sensor Array Enables the Recording of Heart Sounds in Homecare," (in eng), *Sensors (Basel, Switzerland),* vol. 23, no. 13, Jul 7 2023, doi: 10.3390/s23136241.

[13] M. Fynn, K. Mandana, J. Rashid, S. Nordholm, Y. Rong, and G. Saha, "Practicality meets precision: Wearable vest with integrated multi-channel PCG sensors for effective coronary artery disease pre-screening," (in eng), *Comput Biol Med,* vol. 189, p. 109904, May 2025, doi: 10.1016/j.compbiomed.2025.109904.

[14] P. Jyothi and G. Pradeepini, "Heart disease detection system based on ECG and PCG signals with the aid of GKVDLNN classifier," *Multimedia tools and applications,* vol. 83, no. 10, pp. 30587-30612, 2024, doi: 10.1007/s11042-023-16562-9.

[15] R. Hettiarachchi *et al.*, "A Novel Transfer Learning-Based Approach for Screening Pre-Existing Heart Diseases Using Synchronized ECG Signals and Heart Sounds," 2021: IEEE, pp. 1-5, doi: 10.1109/ISCAS51556.2021.9401093.

[16] J. Nedoma *et al.*, "Comparison of BCG, PCG and ECG signals in application of heart rate monitoring of the human body," 2017: IEEE, pp. 420-424, doi: 10.1109/TSP.2017.8076019.

[17] T. Cao, Z. Zhang, W. L. Goh, C. Liu, Y. Zhu, and Y. Gao, "ECG Classification using Binary CNN on RRAM Crossbar with Nonidealities-Aware Training, Readout Compensation and CWT Preprocessing," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 19-21 Oct. 2023 2023, pp. 1-5, doi: 10.1109/BioCAS58349.2023.10389002.

[18] X. Liu *et al.*, "Integrating ECG and PCG Signals through a Dual-Modal ViT for Coronary Artery Disease Detection," *IEEE journal of biomedical and health informatics,* vol. PP, pp. 1-13, 2025, doi: 10.1109/JBHI.2025.3589257.

[19] J. Lin *et al.*, "Portable ECG and PCG wireless acquisition system and multiscale CNN feature fusion Bi-LSTM network for coronary artery disease diagnosis," *Computers in biology and medicine,* vol. 191, pp. 110202-110202, 2025, doi: 10.1016/j.compbiomed.2025.110202.

[20] H. Zhang *et al.*, "Co-learning-assisted progressive dense fusion network for cardiovascular disease detection using ECG and PCG signals," *Expert systems with applications,* vol. 238, p. 122144, 2024, doi: 10.1016/j.eswa.2023.122144.

[21] T. Cao, C. Liu, Y. Gao, and W. L. Goh, "Parasitic-Aware Modelling for Neural Networks Implemented with Memristor Crossbar Array," 2021: IEEE, pp. 122-126, doi: 10.1109/MCSoC51149.2021.00025.

[22] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological measurement,* vol. 37, no. 12, pp. 2181-2213, 2016, doi: 10.1088/0967-3334/37/12/2181.

[23] S. E. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov Model," 2008: IEEE, pp. 345-348, doi: 10.1109/CIC.2008.4749049.

[24] Y. Al-Issa and A. M. Alqudah, "A lightweight hybrid deep learning system for cardiac valvular disease classification," *Scientific reports,* vol. 12, no. 1, pp. 14297-20, 2022, doi: 10.1038/s41598-022-18293-7.

[25] J. Zhu, H. Liu, X. Liu, C. Chen, and M. Shu, "Cardiovascular disease detection based on deep learning and multi-modal data fusion," Biomedical signal processing and control, vol. 99, p. 106882, 2025, doi: 10.1016/j.bspc.2024.106882.

[26] E. Kalatehjari, M. M. Hosseini, A. Harimi, and V. Abolghasemi, "Advanced ensemble learning-based CNN-BiLSTM network for cardiovascular disease classification using ECG and PCG signal," Biomedical signal processing and control, vol. 108, p. 107846, 2025, doi: 10.1016/j.bspc.2025.107846.

[27] P. Li, Y. Hu, and Z.-P. Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," Biomedical signal processing and control, vol. 66, p. 102474, 2021, doi: 10.1016/j.bspc.2021.102474.

[28] X. Wang et al., "Resan: A Residual Dual-Attention Network for Abnormal Cardiac Activity Detection," Computational intelligence, vol. 40, no. 6, p. n/a, 2024, doi: 10.1111/coin.70005.

[29] M. Marocchi, L. Abbott, Y. Rong, S. Nordholm, and G. Dwivedi, "Abnormal Heart Sound Classification and Model Interpretability: A Transfer Learning Approach with Deep Learning," Journal of vascular diseases, vol. 2, no. 4, pp. 438-459, 2023, doi: 10.3390/jvd2040034.