HYBRID CNN-TRANSFORMER NETWORKS WITH CROSS-REPRESENTATION ATTENTION FOR ECG-BASED CARDIOVASCULAR DISEASE CLASSIFICATION

Anh-Dung Ho[1,*]

DOI: https://doi.org/10.57001/huih5804.2026.080

ABSTRACT

Electrocardiogram (ECG) analysis plays an important role in the early diagnosis of cardiovascular diseases. However, the complexity of ECG waveforms presents significant challenges for automated classification. This paper proposes a novel Hybrid CNN-ViT framework for cardiovascular disease classification, where one-dimensional ECG signals are transformed into three complementary two-dimensional representations: Gramian Angular Field (GAF), Recurrence Plot (RP), and Markov Transition Field (MTF). The proposed framework combines CNN-based local feature learning and Transformer-based global contextual modeling through an Adaptive Attention Fusion mechanism to effectively capture complementary ECG characteristics. In addition, a Cross-Channel Attention (CCA) mechanism is introduced to model inter-representation dependencies and enhance discriminative feature interactions among GAF, RP, and MTF representations. Experiments conducted on the PTB-XL dataset demonstrate that the proposed Hybrid CNN-ViT framework achieves a sensitivity of 89.12% and a specificity of 80.38%, outperforming conventional CNN-based models, Transformer-based models, and a recent state-of-the-art (SOTA) method. The results confirm that combining multiple 2D ECG representations with Hybrid CNN-ViT learning and Cross-Channel Attention significantly improves cardiovascular disease classification performance while maintaining a balanced trade-off between sensitivity and specificity, which is critical for biomedical diagnostic applications.

Keywords: *Convolutional Neural Network, Cross-Channel Attention, Electrocardiogram, Gramian Angular Field, Recurrence Plot, Markov Transition Field, Vision Transformer.*

[1]Faculty of Informatic Technology, East Asia University of Technology, Vietnam
[*]Email: dungha@eaut.edu.vn

Received: 18/01/2026
Revised: 21/3/2026
Accepted: 30/3/2026

1. INTRODUCTION

Cardiovascular diseases remain one of the leading causes of mortality worldwide [1]. Early and accurate diagnosis of cardiac disorders through ECG signals plays a crucial role in timely detection and effective treatment of cardiovascular conditions. However, manual ECG analysis largely depends on clinicians' expertise and is susceptible to noise as well as inter-individual physiological variations [2]. Consequently, the application of artificial intelligence (AI) and deep learning techniques for ECG signal classification has emerged as a prominent research trend in recent years [3, 4].

CNNs have been widely employed for ECG feature extraction due to their strong capability in learning local patterns [5, 6]. Nevertheless, the convolutional mechanism inherently focuses on local receptive fields, which limits CNNs in modeling global dependencies and long-range correlations within ECG signals. In contrast, Transformer-based models, particularly ViTs [7], leverage self-attention mechanisms to capture global relationships among feature elements and have demonstrated remarkable performance in image processing and medical diagnosis tasks [8, 9]. Although Vision Transformers are effective in modeling global contextual dependencies through self-attention mechanisms, they may overlook fine-grained local morphological structures that are important for ECG analysis. Therefore, combining CNNs and Transformers within a hybrid framework provides a promising strategy for jointly exploiting local and global ECG characteristics.

In addition, transforming one-dimensional ECG signals into two-dimensional representations has been shown to be an effective strategy for exploiting the strengths of 2D deep learning models [10, 11].

Techniques such as GAF, RP, and MTF have proven effective in characterizing global relationships, recurrent structures, and state transition probabilities of ECG signals [12, 13]. These representations enable models to learn not only morphological features but also the dynamic characteristics of cardiac rhythms.

However, when jointly processing multiple ECG representations (GAF, RP, MTF) or multiple ECG channels, existing models often struggle to effectively capture inter-channel correlations. Suboptimal feature fusion may lead to information dilution or suppression of critical patterns. To address this limitation, this paper proposes a model incorporating a CCA mechanism, which enables the network to automatically learn correlations among feature channels and amplify relevant information during inference.

Unlike conventional self-attention mechanisms that operate independently within each channel, the proposed CCA facilitates cross-channel interactions among feature representations derived from different 2D ECG images. When integrated with 2D CNNs (ResNet50) and a Vision Transformer (ViT-B/16), the proposed Hybrid CNN-ViT framework simultaneously exploits CNN-based local morphological learning and Transformer-based global contextual modeling through an Adaptive Attention Fusion mechanism.

The proposed model consists of three main stages: (1) transforming 1D ECG signals into three 2D representations, namely GAF, RP, and MTF; (2) parallel feature extraction using CNN and ViT-B/16 branches; and (3) feature fusion via a CCA block, followed by classification using a softmax layer. Experimental results on the PTB-XL dataset [14] demonstrate that the the proposed Hybrid CNN-ViT framework achieves a sensitivity of 89.12% and a specificity of 80.38%, outperforming ResNet50+CCA, and the method proposed by authors in [16], which reports corresponding values of 72.3% and 73.9%. These results confirm that combining 2D ECG representations with cross-channel attention constitutes an effective approach, offering strong generalization capability and high reliability for ECG-based cardiovascular disease recognition. These results demonstrate that combining CNN-based local feature learning and Transformer-based global dependency modeling provides complementary ECG representations for cardiovascular disease classification.

2. MATERIALS AND METHODS

2.1. Optimization framework

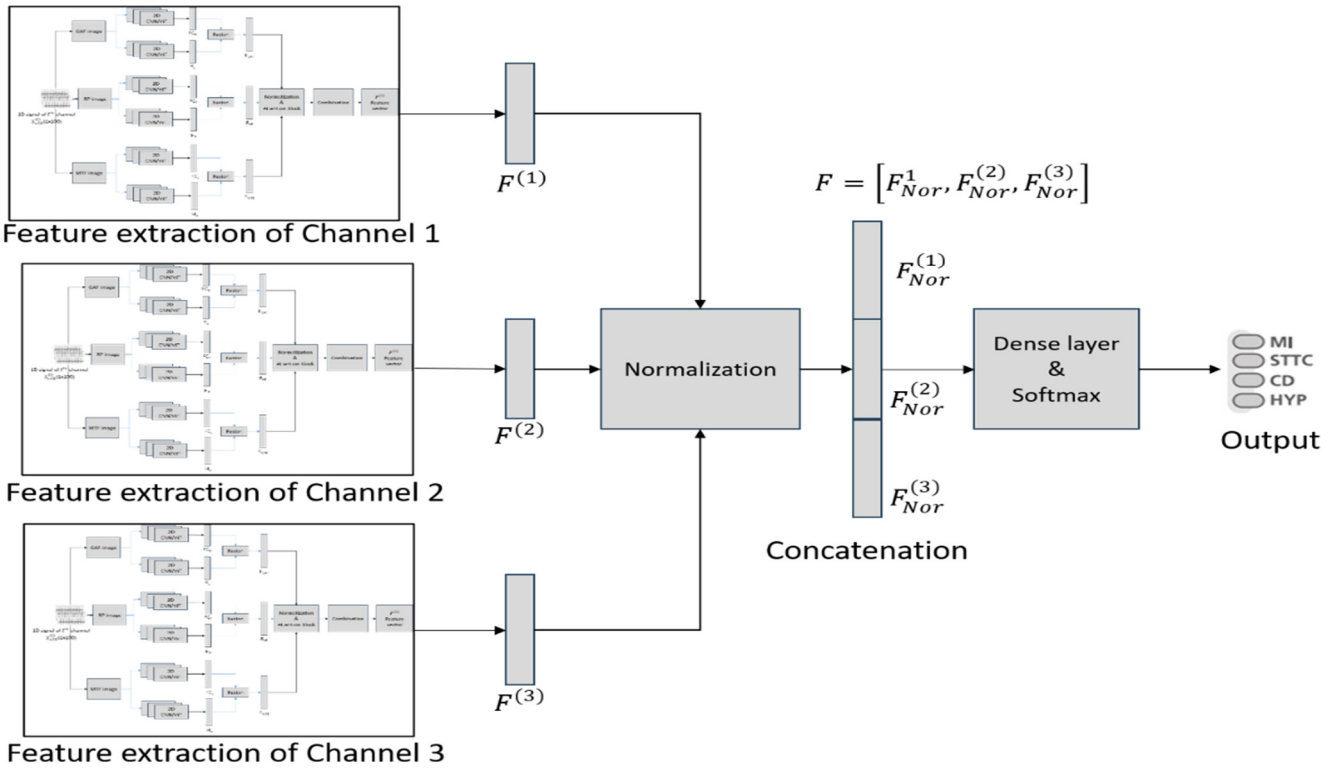

Figure 1. Our proposed ECG-based framework for cardiovascular disease classification model

The proposed Hybrid CNN-ViT framework (Figure 1) is designed to detect and classify abnormal cardiac conditions based on ECG signals. The one-dimensional ECG signal is first transformed into three image-based representations: (1) the GAF, which captures global relationships among signal points; (2) the RP, which characterizes signal recurrence patterns and dynamic structures; and (3) the MTF, which encodes the state transition probabilities between signal amplitude levels.

For each representation branch, the generated 2D image is simultaneously processed by both a CNN encoder and a Vision Transformer (ViT) encoder to extract complementary local and global ECG features. The CNN branch focuses on learning fine-grained local morphological patterns, while the ViT branch captures long-range contextual dependencies through self-attention mechanisms. Subsequently, an Adaptive Attention Fusion block dynamically combines the CNN and ViT features to generate a hybrid feature representation for each ECG modality.

The resulting hybrid features from the GAF, RP, and MTF branches are then normalized and passed through the proposed Cross-Channel Attention (CCA) mechanism to model inter-representation correlations and enhance discriminative feature interactions across multiple ECG representations. Finally, the fused feature vector is forwarded to a fully connected dense layer followed by a Softmax classifier to predict four cardiovascular conditions: myocardial infarction (MI), ST/T change (STTC), conduction disturbance (CD), and hypertrophy (HYP).

2.2. Transformation of ECG Signals into 2D Image Representations

Converting ECG signals from a one-dimensional time series is described as follows. Let the ECG signal at an arbitrary channel be denoted as $x = \{x_1, x_2, \dots, x_N\}$. Three types of 2D image representations are then generated as follows:

a) **Gramian Angular Field (GAF):** This representation captures the global relationships among signal points. First, the signal is mapped from the normalized value domain [-1, 1] to the corresponding phase angles, as defined in Eq. (1):

$$\varphi_i = \arccos(x_i),\ r_i = \frac{i}{N} \tag{1}$$

Subsequently, the GAF matrix is computed according to Eq. (2) as follows:

$$I_{GAP} = \cos(\varphi_i + \varphi_j) = x_i x_j - \sqrt{(1 - x_i^2)(1 - x_j^2)} \tag{2}$$

This G_{ij} matrix represents the pairwise correlations between all signal elements in terms of their phase angles.

b) **Recurrence Plot (RP)**: This representation describes the recurrence behavior and dynamical structure of the signal. Specifically, the RP depicts the reappearance of signal states in phase space, as defined in Eq. (3):

$$I_{RP} = \Theta(\varepsilon - \|x_i - x_j\|) \tag{3}$$

Where Θ denotes the Heaviside step function and ε is the distance threshold..

c) **Markov Transition Field (MTF):** This representation encodes the transition probabilities between signal amplitude levels. The MTF models transitions among discrete states and is formulated as shown in Eq. (4):

$$I_{MTF} = P_{q_i,q_j},\ with\ q_i, q_j \in \{1, 2, \dots, Q\} \tag{4}$$

2.3. ECG Feature Extraction

After transforming the 1D ECG time series into 2D image representations, features are extracted independently from each channel, as illustrated in Figure 2.

After generating the three 2D images, each input image $I_1^{(i)}(3 \times 256 \times 256) \in \{I_{GAF}, I_{RP}, I_{MTF}\}$ is fed into a feature extractor implemented using either a CNN and a ViT-B/16 model, as Eq. (5) and Eq. (6) follow:

$$FC_I = F_{CNN}\left(I_1^{(i)}(3 \times 256 \times 256)\right) \tag{5}$$

$$H_I = F_{ViT-B/16}\left(I_1^{(i)}(3 \times 256 \times 256)\right) \tag{6}$$

In the proposed framework, local features extracted by the CNN branch are denoted as F_{local}, while global contextual features learned by the Vision Transformer branch are represented as F_{global}. Instead of directly concatenating these feature vectors, this study employs an Attention-based Fusion mechanism to adaptively learn the contribution of each feature type. Specifically, the two feature vectors are first combined and projected through a learnable attention layer as Eq. (7) follows:

$$a = \text{Softmax}(W[F_{local} = FC; F_{global} = H]) \tag{7}$$

where W denotes a learnable weight matrix and $a = [a_1, a_2]$ represents the attention coefficients corresponding to the local and global features, respectively. The fused GAF, representation is then computed as Eq. (8) follows:

$$F_{GAF} = a_1 F_{local} + a_2 F_{global} \tag{8}$$

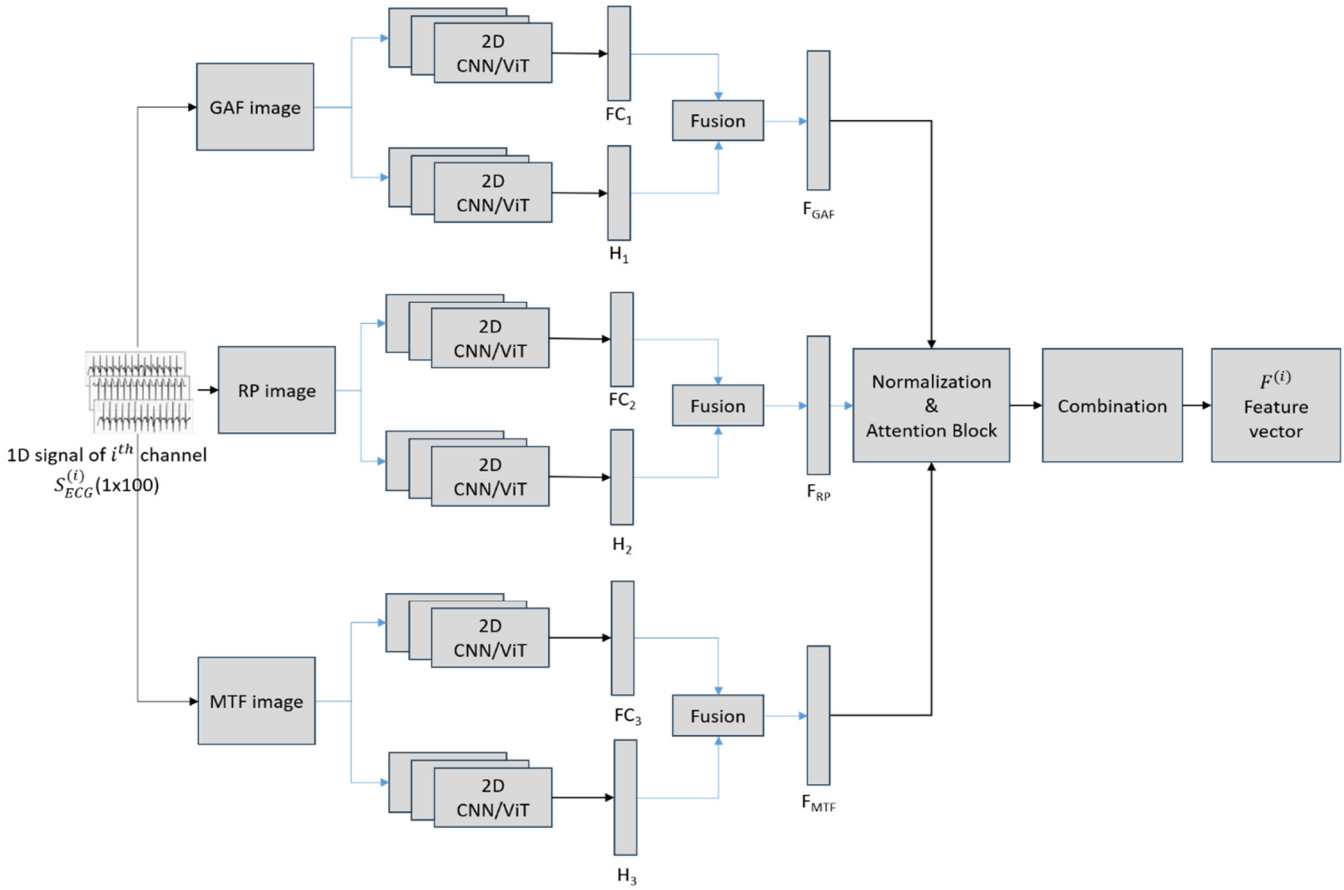

Figure 2. The feature extraction block for one 1D ECG channel

This mechanism enables the model to automatically determine the relative importance of CNN-based local morphological features and ViT-based global contextual dependencies according to the characteristics of each ECG sample. Consequently, the proposed adaptive fusion strategy effectively exploits both fine-grained local patterns, such as QRS complexes and ST-segment variations, and long-range global relationships within ECG signals. This Adaptive Attention Fusion mechanism enhances feature representation capability and improves the overall performance of ECG-based cardiovascular disease classification. Thus from $F_k^{(i)} \in R^L = \{F_{GAF}, F_{RP}, F_{MTF}\}$ is hybrid feature vector that combined of CNN feature and ViT feature.

Feature vectors $(F_{k\_Nor}^{(i)})$ are normalized from $F_k^{(i)} \in R^L;\ k = (1, \dots, 3)$ as follows:

$$F_{k\_Nor}^{(i)} \in R^L = \frac{F_k^{(i)}}{\sqrt{\sum_{k=1}^{3} \left\| F_k^{(i)} \right\|^2}} \quad (9)$$

Three feature vectors are then passed through an attention block to generate three attention coefficients as follows:

$$\alpha_k = Softmax\left(W[F_{1_{Nor}}^{(i)}, F_{2_{Nor}}^{(i)}, F_{3_{Nor}}^{(i)}]\right)$$
$$k = (1, \dots, 3) \quad (10)$$

The feature vector of the i^{th} channel is then fused as follows:

$$F^{(i)} \in R^L = \frac{1}{3} \sum_{k=1}^{3} \alpha_k F_{k\_Nor}^{(i)} \quad (11)$$

2.4. Cross-Channel Attention Across Feature Channels (CCA)

For each channel, after feature extraction, the resulting feature vector is denoted as $F^{(i)}$(i = (1,…,3)). In the dataset used in this study, three 1D ECG channels are available. Therefore, before fusing the three feature vectors as illustrated in Figure 1, a set of Query (Q), Key (K), and Value (V) representations is constructed. For each branch i, the feature vector is projected into three different subspaces through learned weight matrices as follows:

$$Q^{(i)} = F^{(i)} W_Q^{(i)}, \quad K^{(i)} = F^{(i)} W_K^{(i)},$$
$$V^{(i)} = F^{(i)} W_V^{(i)} \quad (12)$$

With three channels, each channel is used as the Query once, while the remaining two channels serve as

the Key and Value. The general formulation of the cross-attention mechanism is given in Eq. (10) as follows:

$$CA\left(F^{(i)} | \{F^{(j)}\}_{i \neq j}\right) = softmax\left(\frac{Q^{(i)}[K^{(j1)}, K^{(j2)}]^T}{\sqrt{d_k}}\right)[V^{(j1)}, V^{(j2)}] \quad (13)$$

The resulting cross-attended feature vector is computed according to Eq. (11) as follows:

$$F_{Cross}^{(i)} = \sum_{i \neq j} \beta_{ij} CA(F^{(i)}, F^{(j)}) \quad (14)$$

The fused representation is obtained using feature concatenation as follows:

$$F = Concat\left[F_{Cross}^{(1)}, F_{Cross}^{(2)}, F_{Cross}^{(3)}\right] \quad (15)$$

The classification process is performed according to Eqs. (13), (14), and (15) as follows:

$$z = WF + b \quad (16)$$

$$\hat{y} = softmax(z) \quad (17)$$

$$L = -\Sigma y_i \log(\hat{y}_i) \quad (18)$$

3. RESULTS AND DISCUSSION

The experiments were conducted on a workstation equipped with an NVIDIA GPU with 11 GB of VRAM, ensuring efficient parallel training and inference for deep learning models. All models were implemented in Python and executed using the PyTorch framework, which facilitates computational optimization and provides flexibility for constructing custom neural network modules. Training and evaluation were performed on the PTB-XL dataset [14], a large-scale, multi-label ECG database that is widely used in automated cardiovascular diagnosis research. Three evaluation settings were considered in this study:

- 2D ECG representations with self-attention, designed to assess the model's ability to extract deep spatial features and to analyze its capacity for learning local dependencies.

- 2D ECG representations with Cross-Channel Attention, where a 2D CNN model (ResNet-50) or a Transformer-based model with ViT-B/16 is combined with channel-wise attention and cross-channel attention mechanisms. This setting is used to evaluate the model's capability to learn both local dependencies and global dependencies in ECG signals.

- A Hybrid CNN-ViT framework with Adaptive Attention Fusion and Cross-Channel Attention, where CNN branches are utilized to capture fine-grained local morphological patterns while Vision Transformer branches are employed to model long-range global dependencies. The extracted local and global representations are adaptively fused to enhance complementary feature learning and improve cardiovascular disease classification performance.

3.1. 2D ECG Representations with a Self-Attention Module

We evaluate a 2D CNN model (ResNet-50) and a Transformer-based model (ViT-B/16), Hybrid CNN-ViT, each integrated with an attention module. The evaluation results are presented in Figure 3.

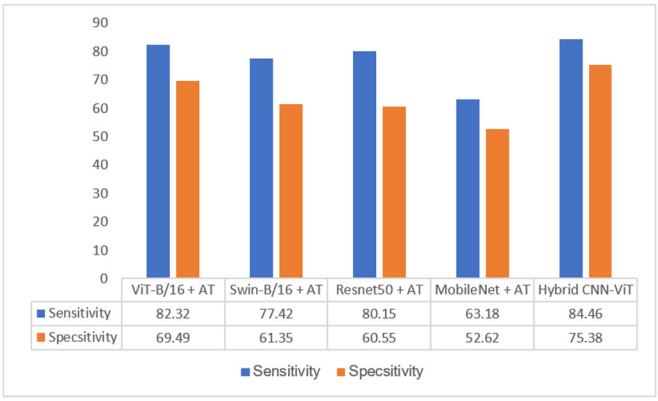

Figure 3. Results of 2D CNN/ViT Models with Self-Attention

The results reported in Table 1 reveal notable differences among Transformer-based models, CNN-based models, and the proposed Hybrid CNN-ViT framework for cardiovascular disease classification using 2D ECG representations. As illustrated in Figure 3, several important observations can be drawn.

- Transformer-based models generally outperform conventional CNN-based architectures in both sensitivity and specificity. In particular, ViT-B/16+AT achieves a sensitivity of 82.32% and a specificity of 69.49%, while Swin-B/16+AT attains 77.42% sensitivity and 61.35% specificity. These results demonstrate the strong capability of Transformer architectures to capture global contextual dependencies and long-range relationships across different regions of ECG image representations through the self-attention mechanism.

- CNN-based models exhibit comparatively lower performance. ResNet50+AT achieves a sensitivity of 80.15% and a specificity of 60.55%, whereas MobileNet+AT records substantially lower values of 63.18% and 52.62%, respectively. Although ResNet50 benefits from deep convolutional structures and residual connections that effectively preserve local morphological information, CNN-based architectures remain inherently

constrained in modeling long-range global dependencies due to their localized convolution operations. In contrast, MobileNet prioritizes computational efficiency and lightweight deployment, but its limited representational capacity reduces its ability to capture subtle and complex ECG patterns.

- The proposed Hybrid CNN-ViT framework achieves the best overall performance, obtaining a sensitivity of 84.46% and a specificity of 75.38%, outperforming both pure CNN-based and pure Transformer-based models. These improvements indicate that combining CNN-based local feature learning with Transformer-based global contextual modeling provides complementary information for ECG classification. Specifically, the CNN branch effectively captures fine-grained local morphological characteristics, such as QRS complexes and ST-segment variations, while the ViT branch models long-range dependencies and global contextual relationships across ECG representations.

Furthermore, the Adaptive Attention Fusion mechanism enables the network to dynamically determine the relative importance of local and global representations according to the characteristics of each ECG sample. This adaptive fusion strategy enhances feature representation capability and improves the robustness of cardiovascular disease classification.

The results in Figure 3 suggest that neither purely convolutional architectures nor purely Transformer-based architectures alone are sufficient to fully capture the complex local-global characteristics of ECG signals. Instead, the proposed Hybrid CNN-ViT framework effectively exploits the complementary strengths of both architectures, resulting in more stable and discriminative feature representations. These findings demonstrate that transforming 1D ECG signals into 2D representations and jointly leveraging CNN-based local learning and Transformer-based global modeling constitutes an effective strategy for cardiovascular disease recognition and classification.

Based on these observations, subsequent evaluations will focus on the proposed Hybrid CNN-ViT framework combined with the Cross-Channel Attention (CCA) mechanism, as presented in Section 3.2.

3.2. 2D ECG Representations with cross-channel attention

In this section, we evaluate the 2D CNN model (ResNet-50) and the Transformer-based model (ViT-B/16) and Hybrid CNN-ViT, which achieved higher performance as reported in the previous section. These models are further combined with channel-wise attention and CCA across channels. The evaluation results are presented in Figure 4.

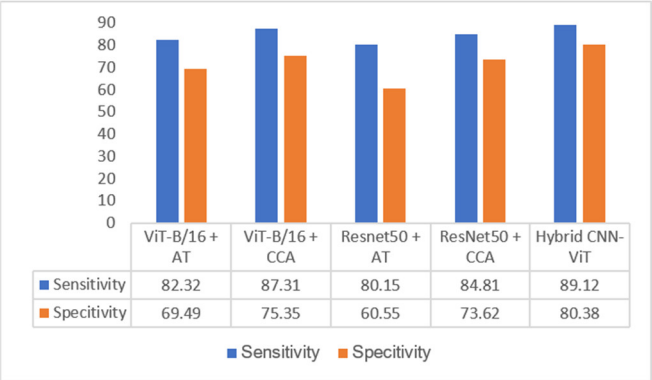

Figure 4. Results of 2D CNN/ViT Models with Self-Attention and CCA

The results shown in Figure 4 demonstrate that both the proposed Cross-Channel Attention (CCA) mechanism and the Hybrid CNN-ViT framework contribute significantly to improving cardiovascular disease classification performance from 2D ECG representations.

- For the Transformer-based model, incorporating the CCA block into ViT-B/16 improves sensitivity from 82.32% to 87.31% and specificity from 69.49% to 75.35%. These improvements indicate that the proposed CCA mechanism effectively enhances inter-representation feature interactions and enables the Transformer architecture to better capture informative global contextual dependencies across ECG representations.

- For the CNN-based model, integrating CCA with ResNet50 increases sensitivity from 80.15% to 84.81% and specificity from 60.55% to 73.62%. The improvement is particularly significant in specificity, demonstrating that CCA effectively suppresses irrelevant feature responses and reduces false positive predictions, thereby improving diagnostic reliability.

- The proposed Hybrid CNN-ViT framework achieves the best overall performance, obtaining 89.12% sensitivity and 80.38% specificity, outperforming both the standalone CNN-based and Transformer-based architectures, including their CCA-enhanced variants. These results demonstrate that combining CNN-based local morphological learning with Transformer-based global contextual modeling provides highly complementary feature representations for ECG analysis.

Furthermore, the Adaptive Attention Fusion mechanism allows the network to dynamically balance local and global features according to the characteristics

of each ECG sample. Consequently, the proposed hybrid architecture can simultaneously capture fine-grained local waveform structures, such as QRS complexes and ST-segment variations, as well as long-range dependencies and global contextual relationships across ECG representations.

The integration of the proposed Cross-Channel Attention (CCA) mechanism and the Hybrid CNN-ViT framework not only improves overall classification performance but also provides a better balance between sensitivity and specificity, which is particularly important for biomedical diagnostic applications where incorrect predictions may lead to serious clinical consequences.

As shown in Table 1, incorporating the CCA mechanism significantly enhances the performance of both CNN-based and Transformer-based architectures. Specifically, ViT-B/16 combined with CCA improves sensitivity from 82.32% to 87.31% and specificity from 69.49% to 75.35%, while ResNet50 combined with CCA increases sensitivity from 80.15% to 84.81% and specificity from 60.55% to 73.62%. These improvements demonstrate that the proposed CCA mechanism effectively strengthens inter-representation feature interactions and enables the models to learn more discriminative ECG characteristics.

Furthermore, the proposed Hybrid CNN-ViT framework achieves the best overall performance, obtaining 89.12% sensitivity and 80.38% specificity. The superior performance of the hybrid architecture indicates that combining CNN-based local morphological learning with Transformer-based global contextual modeling provides highly complementary information for ECG classification. In particular, the Adaptive Attention Fusion mechanism enables the network to dynamically balance local and global feature representations according to the characteristics of each ECG sample, resulting in more robust and stable classification performance across different ECG patterns.

When compared with the existing method proposed by Quancheng Geng et al. [16], which reports a sensitivity of 72.3% and a specificity of 73.9%, the proposed Hybrid CNN-ViT framework achieves substantial improvements of approximately 16.8% in sensitivity and 6.5% in specificity. These results confirm the effectiveness, robustness, and strong generalization capability of the proposed framework compared with existing state-of-the-art methods.

Table 1. Comparison results between our method and SOTA method

Model	Sensitivity	Specsitivity
ViT-B/16 + AT	82.32	69.49
ViT-B/16 + CCA	87.31	75.35
Resnet50 + AT	80.15	60.55
ResNet50 + CCA	84.81	73.62
Hybrid CNN-ViT	**89.12**	**80.38**
Quancheng Geng et al. [16]	72.3	73.9

5. CONCLUSION

This paper presents a novel Hybrid CNN-ViT framework for cardiovascular disease classification using ECG signals transformed into multiple 2D representations, including GAF, RP, and MTF. The proposed framework combines CNN-based local feature learning and Transformer-based global contextual modeling through an Adaptive Attention Fusion mechanism, enabling effective extraction of complementary ECG characteristics. In addition, the proposed Cross-Channel Attention (CCA) mechanism enhances inter-representation feature interactions and improves discriminative feature learning from ECG signals. Experimental results on the PTB-XL dataset demonstrate that the proposed Hybrid CNN-ViT framework achieves superior performance compared with conventional CNN-based models, Transformer-based models, and existing state-of-the-art methods while maintaining a favorable balance between sensitivity and specificity. In future work, we will focus on extending the proposed framework to multi-lead ECG datasets and incorporating multi-task and self-supervised learning strategies to further improve generalization capability and practical deployment in real-world healthcare applications.

REFERENCES

[1]. World Health Organization, *Cardiovascular Diseases Report*. Switzerland, 2023.

[2]. Zheng Z., Chen Z., Hu F., Zhu J., Tang Q., Liang Y., "An Automatic Diagnosis of Arrhythmias Using a Combination of CNN and LSTM Technology," *Electronics*, 9, 121, 2020. https://doi.org/10.3390/electronics9010121

[3]. O. Yildirim, "A novel wavelet sequence-based deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, 96, 189-202, 2018. doi: 10.1016/j.compbiomed.2018.03.016.

[4]. S. Kiranyaz, T. Ince, M. Gabbouj, "Real-time patient-specific ECG classification by 1D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, 63, no. 3, 664-675, 2016. doi: 10.1109/TBME.2015.2468589.

[5]. G. Sannino, G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Generation Computer Systems*, 86, 446-455, 2018. Doi: https://doi.org/10.1016/j.future.2018.03.057.

[6]. Jun T.J., Nguyen H.M., Kang D., Kim D., Kim D., Kim Y., "ECG arrhythmia classification using a 2-D convolutional neural network," *ArXiv:abs/1804.06812*, 2018.

[7]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021.

[8]. Kerol Djoumessi, Ofosu Mensah Samuel, Berens Philipp, "A Hybrid Fully Convolutional CNN-Transformer Model for Inherently Interpretable Medical Image Classification," *arXiv:2504.08481*, 2025. Doi: 10.48550/arXiv.2504.08481.

[9]. Uğraş Berat, Gerek Omer, Saygı İbrahim Talha, "CardioPatternFormer: Pattern-Guided Attention for Interpretable ECG Classification with Transformer Architecture," *arXiv:2505.20481,* 2025. Doi: 10.48550/arXiv.2505.20481.

[10]. Hemaxi Narotamo, Mariana Dias, Ricardo Santos, André V. Carreiro, Hugo Gamboa, Margarida Silveira, "Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches," *Biomedical Signal Processing and Control,* 93, 106141, 2024. Doi: https://doi.org/10.1016/j.bspc.2024.106141.

[11]. Li J, Pang SP, Xu F, Ji P, Zhou S, Shu M., "Two-dimensional ECG-based cardiac arrhythmia classification using DSE-ResNet," *Sci Rep.,* 12(1):14485, 2022. doi: 10.1038/s41598-022-18664-0. PMID: 36008568; PMCID: PMC9411603.

[12]. J. P. Eckmann, S. O. Kamphorst, D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, 4, 9, 973-977, 1987. doi: 10.1209/0295-5075/4/9/004.

[13]. A. S. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, L. A. N. Amaral, "Duality between time series and networks," *PLoS ONE*, 6, 8, e23378, 2011. doi: 10.1371/journal.pone.0023378.

[14]. Wagner P., Strodthoff N., Bousseljot R., Samek W., Schaeffter T., "PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3)," *PhysioNet*. RRID:SCR_007345, 2022.

[15]. Y. Elmir, Y. Himeur and A. Amira, "ECG classification using Deep CNN and Gramian Angular Field," in *2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService)*, Athens, Greece, 137-141, 2023. doi: 10.1109/BigDataService58306.2023.00026.

[16]. Geng Q., Liu H., Gao T., Liu R., Chen C., Zhu Q., Shu M., "An ECG Classification Method Based on Multi-Task Learning and CoT Attention Mechanism," *Healthcare*, 11, 1000, 2023. https://doi.org/10.3390/healthcare11071000