



Tạp chí điện tử
Khoa học và Công nghệ Trường Đại học Công nghệ Đông Á
Website Tạp chí: <https://vjai.org.vn>

Heart Disease Classification Based on 2D ECG Representations Using Cross-Attention Mechanisms in CNN and Vision Transformer (ViT) Models

Ho Anh Dung

Faculty of Information Technology, East Asia University of Technology

ARTICLE INFO

ABSTRACT

Keywords:

convolutional neural network, cross-channel attention, electrocardiogram, Gramian Angular Field (GAF), recurrence plot, Markov Transition Field (MTF), vision transformer.

Electrocardiogram (ECG) signal analysis plays a crucial role in the early diagnosis of cardiovascular diseases. However, the complexity and variability of ECG waveforms make automatic detection and classification of cardiac abnormalities a challenging task. This study proposes a deep learning framework based on two architectures, Convolutional Neural Networks (CNN) and Vision Transformers (ViT), in which one-dimensional ECG signals are transformed into three two-dimensional image representations: Gramian Angular Field (GAF), Recurrence Plot (RP), and Markov Transition Field (MTF). These images are processed in parallel through CNN/ViT branches to simultaneously capture both local and global features of the signal. In particular, the proposed model incorporates a Cross Channel Attention (CCA) mechanism to learn the correlations among feature branches and enhance important inter-channel information. Experiments deployed on the PTB-XL dataset show that the ViT-B/16 model combined with CCA achieves a sensitivity of 87.31% and a specificity of 75.35%, outperforming ResNet50 with CCA (84.81%, 73.62%) and the method in [16] (72.3%, 73.9%). The results demonstrate that integrating 2D ECG representations with the CCA mechanism significantly improves model accuracy and generalization capability while maintaining a balanced trade-off between sensitivity and specificity, which is essential for biomedical diagnostic applications.

* Corresponding author

Email: dungha@eaut.edu.vn

DOI: <https://doi.org/10.65153/44f8j261>

Received: 04/11/2025; Received in revised form: 16/03/2026; Accepted: 30/03/2026

Available online: 03/04/2026

Published by: East Asia University of Technology



Phân loại bệnh tim dựa trên biểu diễn ECG 2D sử dụng cơ chế chú ý chéo trong các mô hình CNN và ViT

Hồ Anh Dũng
Khoa Công nghệ thông tin, Trường Đại học Công nghệ Đông Á

THÔNG TIN BÀI BÁO

TÓM TẮT

Từ khóa:
mạng nơ-ron tích chập, cơ chế chú ý liên kênh, điện tâm đồ, trường góc Gramian, biểu đồ hồi quy, trường chuyển tiếp Markov, transformer thị giác.

Phân tích tín hiệu điện tâm đồ (ECG) đóng vai trò then chốt trong chẩn đoán sớm các bệnh tim mạch. Tuy nhiên, sự phức tạp và đa dạng của dạng sóng ECG khiến việc phát hiện và phân loại tự động các rối loạn tim trở nên thách thức. Nghiên cứu này đề xuất mô hình học sâu sử dụng hai kiến trúc Convolutional Neural Network (CNN) và Vision Transformer (ViT), trong đó tín hiệu ECG 1D được chuyển đổi thành ba biểu diễn ảnh 2D gồm Gramian Angular Field (GAF), Recurrence Plot (RP) và Markov Transition Field (MTF). Các ảnh này được xử lý song song qua các nhánh CNN/ViT nhằm khai thác đồng thời đặc trưng cục bộ và toàn cục của tín hiệu. Đặc biệt, mô hình tích hợp cơ chế Cross Channel Attention (CCA) giúp học mối tương quan giữa các nhánh đặc trưng và khuếch đại thông tin quan trọng giữa các kênh. Thực nghiệm trên cơ sở dữ liệu PTB-XL cho thấy mô hình ViT-B/16 và CCA đạt độ nhạy 87.31% và độ đặc hiệu 75.35%, vượt trội so với ResNet50 và CCA (84.81%, 73.62%) và phương pháp trong [16] đạt (72.3%, 73.9%). Kết quả chứng minh rằng việc kết hợp biểu diễn ECG 2D với cơ chế CCA giúp cải thiện đáng kể độ chính xác và khả năng khái quát hóa của mô hình, đồng thời duy trì sự cân bằng giữa độ nhạy và độ đặc hiệu là yếu tố quan trọng trong các ứng dụng chẩn đoán y sinh.

1. MỞ ĐẦU

Các bệnh lý tim mạch hiện nay vẫn là một trong những nguyên nhân hàng đầu gây tử vong trên toàn cầu [1]. Việc chẩn đoán sớm và chính xác các rối loạn tim thông qua tín hiệu điện tâm đồ (Electrocardiogram – ECG) đóng vai trò quan trọng trong phát hiện sớm và điều trị hiệu quả các bệnh tim mạch. Tuy nhiên, phân tích ECG thủ

công thường phụ thuộc vào chuyên môn của bác sĩ và dễ chịu ảnh hưởng bởi nhiễu hoặc biến thiên sinh lý giữa các cá thể [2]. Do đó, việc ứng dụng trí tuệ nhân tạo (AI) và mô hình học sâu (Deep Learning) trong phân loại tín hiệu ECG đã trở thành xu hướng nổi bật trong những năm gần đây [3][4].

* Tác giả liên hệ

Email: dungha@eaut.edu.vn

DOI: <https://doi.org/10.65153/44f8j261>

Ngày nhận: 04/11/2025; Ngày nhận bản sửa: 16/03/2026; Ngày chấp nhận: 30/03/2026

Ngày online: 03/04/2026

Đơn vị xuất bản: Trường Đại học Công nghệ Đông Á

Trong các phương pháp này, mạng nơ-ron tích chập (Convolutional Neural Network – CNN) đã được áp dụng rộng rãi để trích chọn đặc trưng ECG nhờ khả năng học các đặc trưng cục bộ hiệu quả [5][6]. Tuy nhiên, cơ chế tích chập chỉ tập trung vào vùng lân cận nên CNN bị giới hạn trong việc mô hình hóa mối quan hệ toàn cục (global dependencies) và tương quan dài hạn trong tín hiệu. Ngược lại, mô hình Transformer, đặc biệt là Vision Transformer (ViT) [7], sử dụng cơ chế self-attention giúp nắm bắt mối tương quan toàn cục giữa các phần tử đặc trưng, và đã đạt nhiều thành công trong xử lý ảnh và y học chẩn đoán [8][9].

Bên cạnh đó, một hướng tiếp cận hiệu quả khác là chuyển đổi tín hiệu ECG 1D sang ảnh 2D, cho phép khai thác sức mạnh của các mô hình học sâu 2D [10], [11]. Các kỹ thuật như Gramian Angular Field (GAF), Recurrence Plot (RP) và Markov Transition Field (MTF) đã được chứng minh có khả năng mô tả mối quan hệ toàn cục, cấu trúc lặp lại và xác suất chuyển trạng thái của tín hiệu [12], [13]. Điều này giúp mô hình có thể học được đặc trưng không chỉ về hình thái mà còn về động học của nhịp tim.

Tuy nhiên, khi xử lý đồng thời nhiều loại ảnh (GAF, RP, MTF) hoặc nhiều kênh ECG, các mô hình hiện có thường gặp khó khăn trong việc học sự phụ thuộc giữa các kênh đặc trưng (inter-channel correlation). Việc kết hợp không tối ưu giữa các kênh có thể khiến thông tin quan trọng bị phân tán hoặc triệt tiêu. Để khắc phục vấn đề này, bài báo đề xuất một mô hình kết hợp cơ chế Cross Channel Attention (CCA), cho phép mô hình tự động học mối tương quan giữa các kênh đặc trưng và khuếch đại thông tin liên quan trong quá trình suy luận.

Khác với cơ chế self-attention truyền thống chỉ hoạt động trong từng kênh, CCA giúp mô hình học được sự tương tác chéo giữa các biểu diễn đặc trưng từ các loại ảnh 2D ECG khác nhau. Khi kết hợp với CNN 2D và Vision Transformer (ViT-B/16), mô hình đề xuất tận dụng được đồng thời

khả năng trích chọn đặc trưng cục bộ của CNN và khả năng biểu diễn toàn cục của Transformer, giúp cải thiện đáng kể độ chính xác trong phân loại bệnh tim mạch.

Mô hình đề xuất gồm ba giai đoạn chính: (1) Chuyển tín hiệu ECG 1D thành ba ảnh 2D gồm GAF, RP, và MTF; (2) Trích chọn đặc trưng song song bằng CNN và ViT-B/16; (3) Kết hợp đặc trưng qua khối Cross Channel Attention (CCA) để học sự tương quan giữa các nhánh và phân loại bằng tầng Softmax. Kết quả thử nghiệm trên cơ sở dữ liệu PTB-XL [14] cho thấy mô hình ViT-B/16 và CCA đạt độ nhạy 87.31% và độ đặc hiệu 75.35%, vượt trội hơn các mô hình ResNet50 và AT, và phương pháp của Quancheng Geng và cộng sự trong [16] đạt tương ứng là 72.3% và 73.9%. Điều này chứng tỏ việc kết hợp biểu diễn ECG 2D với cơ chế chú ý chéo giữa các kênh (CCA) là hướng tiếp cận hiệu quả, mang lại khả năng khái quát hóa và độ tin cậy cao trong nhận dạng bệnh lý tim mạch dựa trên tín hiệu ECG.

2. GIẢI QUYẾT VẤN ĐỀ

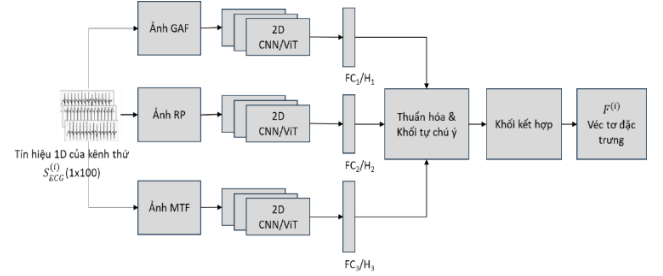
Mô hình được đề xuất (Hình 1) nhằm phát hiện và phân loại các tình trạng tim bất thường dựa trên tín hiệu điện tâm đồ (ECG). Tín hiệu 1D được chuyển đổi thành ba dạng biểu diễn hình ảnh gồm: (1) Gramian Angular Field (GAF) phản ánh mối quan hệ toàn cục giữa các điểm tín hiệu; (2) Recurrence Plot (RP) mô tả sự lặp lại và cấu trúc động học của tín hiệu; và (3) Markov Transition Field (MTF) mã hóa xác suất chuyển trạng thái giữa các mức tín hiệu. Ba loại ảnh này được đưa vào ba nhánh trích chọn đặc trưng song song bằng CNN 2D hoặc Vision Transformer (ViT). Sau đó, các đặc trưng được chuẩn hóa, đưa vào khối chú ý chéo giữa các kênh (Cross-Attention Block) để học sự tương quan giữa các mô hình GAF, RP, và MTF. Cuối cùng, đặc trưng hợp nhất được đưa qua tầng Fully Connected (Dense Layer) và Softmax để dự đoán bốn loại bệnh: MI, STTC, CD, và HYP.

các trạng thái rời rạc và được minh họa như công thức (4) sau đây:

$$I_{MTF} = P_{q_i, q_j}, \text{ với } q_i, q_j \in \{1, 2, \dots, Q\} \quad (4)$$

2.2. Trích chọn đặc trưng dữ liệu ECG

Tín hiệu ECG từ chuỗi tín hiệu 1D sau khi chuyển sang ảnh 2D thì được trích chọn đặc trưng trên từng kênh như minh họa trong Hình 2 sau đây:



Hình 2. Sơ đồ minh họa khối trích chọn đặc trưng dữ liệu trên từng kênh ECG

Sau khi tạo ba ảnh 2D, mỗi ảnh đầu vào $I_1^{(i)} (3 \times 256 \times 256) \in \{I_{GAF}, I_{RP}, I_{MTF}\}$ được đưa qua bộ trích chọn đặc trưng là mạng CNN hoặc ViT-B/16 như trong công thức (5) sau đây:

$$F_k^{(i)} [1 \times L] = FC_i / F_i = F_{CNN/ViT} \left(I_1^{(i)} (3 \times 256 \times 256) \right) \quad (5)$$

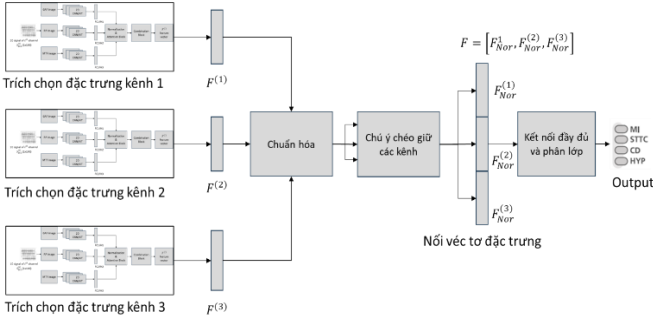
Các feature vector $(F_{k_Nor}^{(i)})$ được chuẩn hóa từ $F_k^{(i)} \in R^L; k = (1, \dots, 3)$ như trong công thức (6) sau đây:

$$F_{k_Nor}^{(i)} \in R^L = \frac{F_k^{(i)}}{\sqrt{\sum_{k=1}^3 \|F_k^{(i)}\|^2}} \quad (6)$$

Ba feature vector được đưa qua khối attention để tạo ba hệ số attention như trong công thức (7) sau đây:

$$\alpha_k = M_{attention} \left(F_{1_Nor}^{(i)}, F_{2_Nor}^{(i)}, F_{3_Nor}^{(i)} \right); k = (1, \dots, 3) \quad (7)$$

Feature vector của một kênh thứ "i" sau đó được kết hợp lại như trong công thức (8) sau đây:



Hình 1. Sơ đồ khối của mô hình phân loại bệnh tim mạch sử dụng tín hiệu ECG.

2.1. Chuyển đổi tín hiệu ECG sang ảnh 2D

Quá trình chuyển tín hiệu ECG từ chuỗi tín hiệu 1D, với tín hiệu ECG tại kênh bất kỳ là $x = \{x_1, x_2, \dots, x_N\}$. Khi đó ba loại ảnh sinh ra như sau:

a. Ảnh GAF (Gramian Angular Field): phản ánh mối quan hệ toàn cục giữa các điểm tín hiệu. Trước tiên, tín hiệu được ánh xạ từ miền giá trị $[-1, 1]$ sang góc pha tương ứng như trong công thức (1) sau đây:

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

Sau đó, ma trận GAF được tính như công thức (2) sau đây:

$$I_{GAF} = \cos(\varphi_i + \varphi_j) = x_i x_j - \text{sqrt}(1 - x_i^2) \text{sqrt}(1 - x_j^2) \quad (2)$$

Ma trận G_{ij} này thể hiện mối quan hệ tương quan giữa mọi cặp phần tử của tín hiệu theo góc pha.

b. Ảnh RP (Recurrence Plot): mô tả sự lặp lại và cấu trúc động học của tín hiệu hay ảnh RP mô tả sự tái xuất hiện của trạng thái tín hiệu trong không gian pha như trong công thức (3) sau đây:

$$I_{RP} = \theta(\varepsilon - \|x_i - x_j\|) \quad (3)$$

Trong đó θ là hàm Heaviside và ε là ngưỡng khoảng cách.

c. Ảnh MTF (Markov Transition Field): mã hóa xác suất chuyển trạng thái giữa các mức tín hiệu hay MTF mô hình hóa sự chuyển tiếp giữa

$$F^{(i)} \in R^L = \frac{1}{3} \sum_{k=1}^3 \alpha_k * F_{k_Nor}^{(i)} \quad (8)$$

2.3. Cơ chế chú ý chéo trên các kênh (Cross Channel Attention - CCA)

Trên mỗi kênh sau khi đã trích chọn đặc trưng là $F^{(i)}$ ($i = (1, \dots, 3)$), trong bộ CSDL chúng tôi sử dụng có ba kênh 1D ECG. Do đó, trước khi kết hợp ba vector đặc trưng lại như trong Hình 1 thì thực hiện tạo bộ ba truy vấn (Query), khóa (Key) và giá trị (Value). Với mỗi nhánh i , ta chiếu đặc trưng sang ba không gian khác nhau thông qua các ma trận trọng số học được như trong công thức (9) sau đây:

$$Q^{(i)} = F^{(i)} W_Q^{(i)}, \quad K^{(i)} = F^{(i)} W_K^{(i)}, \quad V^{(i)} = F^{(i)} W_V^{(i)} \quad (9)$$

Khi có ba kênh, mỗi kênh sẽ làm Query một lần và dùng hai kênh còn lại làm Key và Value. Công thức tổng quát cho cơ chế chú ý chéo được tính bằng công thức (10) sau đây:

$$CA(F^{(i)} | \{F^{(j)}\}_{i \neq j}) = \text{softmax} \left(\frac{Q^{(i)} [K^{(j1)}, K^{(j2)}]^T}{\sqrt{d_k}} \right) [V^{(j1)}, V^{(j2)}] \quad (10)$$

Kết quả tính vector chéo như công thức (11) sau đây:

$$F_{Cross}^{(i)} = \sum_{i \neq j} \beta_{ij} CA(F^{(i)}, F^{(j)}) \quad (11)$$

Kết quả hợp nhất dùng phép nối đặc trưng như trong công thức (12) sau đây:

$$F = \text{Concat} [F_{Cross}^{(1)}, F_{Cross}^{(2)}, F_{Cross}^{(3)}] \quad (12)$$

Phân loại được thực hiện như các công thức (13), (14), (15) sau đây:

$$z = W * F + b \quad (13)$$

$$\hat{y} = \text{softmax}(z) \quad (14)$$

$$L = -\sum y_i \log(\hat{y}_i) \quad (15)$$

3. KẾT QUẢ THỬ NGHIỆM

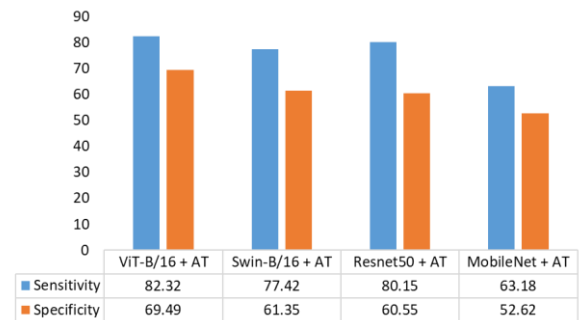
Các thử nghiệm được triển khai trên máy tính trang bị GPU NVIDIA (11 GB VRAM), đảm bảo khả năng huấn luyện và suy luận song song cho các mô hình học sâu. Toàn bộ mô hình được lập trình bằng ngôn ngữ Python và thực thi trên nền tảng PyTorch framework, cho phép tối ưu hiệu năng tính toán cũng như hỗ trợ dễ dàng cho việc xây dựng các khối mạng nơ-ron tùy chỉnh.

Dữ liệu huấn luyện và kiểm thử được thực hiện trên cơ sở dữ liệu PTB-XL, một tập ECG quy mô lớn và được gán nhãn đa lớp, thường được sử dụng trong các nghiên cứu chẩn đoán tim mạch tự động. Ba đánh giá được thực hiện trong nghiên cứu này gồm:

- Biểu diễn 2D ECG kết hợp khối self-attention: nhằm kiểm tra khả năng trích chọn đặc trưng không gian sâu và phân tích năng lực học phụ thuộc cục bộ (local dependencies).
- Biểu diễn 2D ECG kết hợp Cross Channel Attention: Đánh giá mô hình CNN 2D (ResNet-50) hoặc mô hình Transformer (ViT-B/16) kết hợp Attention trên từng kênh và Cross Channel Attention giữa các kênh để phân tích năng lực học phụ thuộc cục bộ (local dependencies) và phụ thuộc toàn cục (global dependencies) của tín hiệu ECG.

3.1. Biểu diễn 2D ECG kết hợp khối self-attention

Thực hiện đánh giá mô hình CNN 2D (ResNet-50) hoặc mô hình Transformer (ViT-B/16) kết hợp khối Attention. Kết quả đánh giá được thể hiện như trong Hình 3 sau đây:



Hình 3. Kết quả sử dụng 2DCNN/ViT kết hợp khối tự chú ý

Kết quả ở Hình 3 cho thấy sự khác biệt rõ rệt giữa nhóm mô hình Transformer (ViT-B/16 và A, Swin-B/16 và AT) và nhóm CNN điển hình (ResNet50 và AT, MobileNet và AT) trong bài toán phân loại bệnh từ tín hiệu ECG 1D được chuyển sang ảnh 2D. Hình 3 cho thấy:

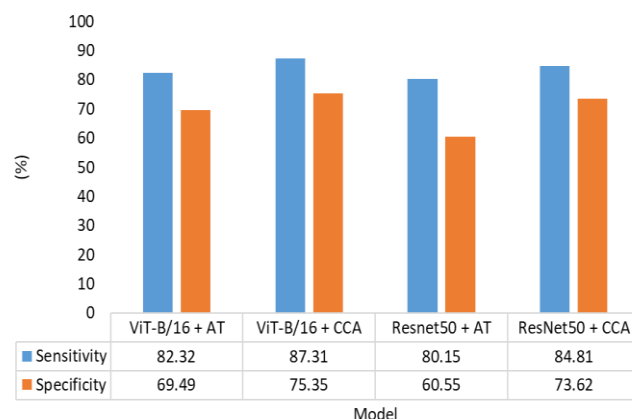
- Hai mô hình Transformer thể hiện hiệu năng vượt trội hơn về cả độ nhạy (Sensitivity) và độ đặc hiệu (Specificity) so với các mô hình CNN. Cụ thể, ViT-B/16 và AT đạt 82.32% độ nhạy và 69.49% độ đặc hiệu, trong khi Swin-B/16 và AT đạt lần lượt 77.42% và 61.35%. Kết quả này phản ánh khả năng nổi bật của các mô hình Transformer trong việc học biểu diễn đặc trưng toàn cục (global representation) và nắm bắt quan hệ dài hạn giữa các vùng trong ảnh ECG 2D, điều mà CNN truyền thống thường gặp hạn chế do cơ chế tích chập.
- Hai mô hình CNN điển hình cho kết quả khiêm tốn hơn: ResNet50 và AT đạt 80.15% độ nhạy và 60.55% độ đặc hiệu, còn MobileNet và AT chỉ đạt 63.18% và 52.62%. ResNet50 cho thấy khả năng trích chọn đặc trưng khá tốt nhờ các khối tích chập sâu và cơ chế residual giúp giữ lại thông tin quan trọng, song vẫn bị giới hạn trong việc mô hình hóa quan hệ không gian và thời gian toàn cục. Trong khi đó, MobileNet dù có ưu điểm về tốc độ và nhẹ mô hình lại không đủ mạnh để xử lý các đặc trưng phức tạp và biến thiên nhỏ trong tín hiệu ECG, dẫn đến độ chính xác thấp hơn đáng kể.

Kết quả trong Hình 3 cho thấy mô hình Transformer, đặc biệt là ViT-B/16 và AT, thể hiện ưu thế rõ rệt nhờ khả năng tự chú ý (self-attention) giúp mô hình nắm bắt được các đặc trưng hình thái ECG một cách toàn diện và ổn định hơn. Ngoài ra, trên từng kênh tổng hợp sau đó lại sử dụng kết hợp với một khối tự chú ý ở vòng ngoài giúp tăng hiệu quả của mô hình ViT.

Điều này chứng minh rằng việc chuyển đổi tín hiệu ECG 1D sang dạng ảnh 2D kết hợp với kiến trúc Transformer và khối tự chú ý trên từng loại ảnh (local feature) là hướng tiếp cận hiệu quả cho bài toán nhận dạng và phân loại bệnh lý tim mạch. Trong đánh giá sau chúng tôi sẽ tiếp tục sử dụng mô hình ViT-B/16 và mô hình ResNet50 để kết hợp với khối CCA và sẽ được trình bày trong mục 3.2 sau đây.

3.2. Biểu diễn 2D ECG kết hợp Cross Channel Attention

Trong phần này chúng tôi thực hiện đánh giá mô hình CNN 2D (ResNet-50) và mô hình Transformer (ViT-B/16) là hai mô hình cho hiệu quả cao hơn như đã trình bày ở mục trên sau đó kết hợp với Attention trên từng kênh và Cross Channel Attention giữa các kênh. Kết quả được thể hiện như trong Hình 4 sau đây:



Hình 4. Kết quả sử dụng 2DCNN/ViT kết hợp self-attention và CCA

Kết quả Hình 4 này cho thấy việc tích hợp cơ chế Cross Channel Attention (CCA) giúp cải thiện rõ rệt hiệu năng của cả hai nhóm mô hình Transformer (ViT-B/16) và CNN (ResNet50) trong bài toán phân loại bệnh dựa trên tín hiệu ECG. Cụ thể:

- Đối với mô hình ViT-B/16, việc kết hợp thêm khối CCA giúp độ nhạy Sensitivity tăng từ 82.32% lên 87.31% và Specificity tăng từ 69.49% lên 75.35%. Mức cải thiện hơn 5% ở cả hai chỉ số cho thấy rằng CCA giúp mô hình khai thác tốt hơn mối tương quan giữa các kênh đặc trưng, nhờ đó nâng cao khả năng nhận biết các biến

thiên nhỏ và dạng sóng bất thường của tín hiệu ECG khi biểu diễn dưới dạng ảnh hai chiều (ECG 2D).

- Tương tự, với mô hình ResNet50, việc bổ sung CCA cũng giúp Sensitivity tăng từ 80.15% lên 84.81% và Specificity từ 60.55% lên 73.62%. Sự cải thiện này đặc biệt đáng chú ý ở chỉ số Specificity, chứng tỏ CCA giúp giảm đáng kể các trường hợp báo sai dương, và tăng độ tin cậy trong chẩn đoán.
- Cả hai mô hình đều đạt hiệu năng cao hơn rõ rệt khi áp dụng Cross Channel Attention, trong đó ViT-B/16 và CCA vẫn là cấu hình mạnh nhất với 87.31% độ nhạy và 75.35% độ đặc hiệu. Kết quả này khẳng định rằng CCA đóng vai trò quan trọng trong việc tăng cường khả năng tương tác và khuếch đại thông tin giữa các kênh đặc trưng, giúp mô hình học được các biểu diễn giàu thông tin hơn cho tín hiệu ECG 2D.

Nhìn chung, việc kết hợp CCA không chỉ cải thiện độ chính xác tổng thể mà còn giúp mô hình đạt được cân bằng tốt hơn giữa hai độ đo Sensitivity và Specificity, điều này rất cần thiết trong các ứng dụng chẩn đoán y sinh nơi sai sót có thể dẫn đến hậu quả nghiêm trọng.

Đáng chú ý, khi so sánh với phương pháp hiện có của nhóm tác giả trong công bố [16] với Sensitivity là 72.3% và Specificity là 73.9% cho thấy mô hình đề xuất ViT-B/16 và CCA đạt hiệu năng vượt trội, với hai tiêu chí này tăng hơn 15% về độ nhạy Sensitivity và 1.4% về độ đặc hiệu Specificity, đồng thời vẫn duy trì được độ ổn định trên nhiều kiểu dữ liệu ECG khác nhau. Kết quả này khẳng định tính hiệu quả và khả năng khái quát hóa cao của phương pháp đề xuất so với các mô hình truyền thống, được thể hiện cụ thể trong Bảng 1 sau đây:

Bảng 1. Kết quả so sánh giải pháp đề xuất với công bố khác

Mô hình	Sensitivity (%)	Specificity (%)
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
Quancheng Geng et al. [16]	72.3	73.9

4. KẾT LUẬN

Bài báo đã trình bày một hướng tiếp cận mới trong nhận dạng bệnh lý tim mạch dựa trên tín hiệu ECG bằng cách kết hợp biểu diễn 2D (GAF, RP, MTF) với các mô hình học sâu CNN và ViT. Cơ chế Cross Channel Attention (CCA) được đề xuất đã chứng minh hiệu quả trong việc tăng cường tương tác giữa các nhánh đặc trưng, giúp mô hình khai thác tốt hơn thông tin không gian–thời gian của tín hiệu ECG. Kết quả thực nghiệm trên cơ sở dữ liệu PTB-XL cho thấy mô hình ViT-B/16 và CCA đạt hiệu quả vượt trội, đồng thời đảm bảo sự cân bằng giữa Sensitivity và Specificity là yếu tố then chốt trong các ứng dụng chẩn đoán y sinh. Hướng phát triển tiếp theo sẽ tập trung vào việc tối ưu hóa mô hình cho các hệ thống phát hiện và nhận dạng, mở rộng sang các tập dữ liệu ECG đa kênh và tích hợp thêm các cơ chế học đa nhiệm để nâng cao tính tổng quát và khả năng triển khai thực tế.

TÀI LIỆU THAM KHẢO

- [1] World Health Organization, *Cardiovascular Diseases Report*, 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* 2020, 9, 121. <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*, vol. 96, pp. 189–202, May 2018, doi: 10.1016/j.compbimed.2018.03.016.
- [4] S. Kiranyaz, T. Ince, and M. Gabbouj, Real-time patient-specific ECG classification by 1D convolutional neural networks, *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, Mar. 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*, Volume 86, 2018, Pages 446-455, ISSN 0167-739X, <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. (2018). ECG arrhythmia classification using a 2-D convolutional neural network. *ArXiv, abs/1804.06812*.
- [7] A. Dosovitskiy et al., An image is worth 16x16 words: Transformers for image recognition at scale, in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2021.
- [8] Kerol, Djoumessi & Ofosu Mensah, Samuel & Berens, Philipp. (2025). A Hybrid Fully Convolutional CNN-Transformer Model for Inherently Interpretable Medical Image Classification. 10.48550/arXiv.2504.08481.
- [9] Uğraş, Berat & Gerek, Omer & Saygı, İbrahim Talha. (2025). CardioPatternFormer: Pattern-Guided Attention for Interpretable ECG Classification with Transformer Architecture. 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*, Volume 93, 2024, 106141, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2024.106141>.
- [11] Li, J., Pang, Sp., Xu, F. *et al.* Two-dimensional ECG-based cardiac arrhythmia classification using DSE-ResNet. *Sci Rep* 12, 14485 (2022). <https://doi.org/10.1038/s41598-022-18664-0>
- [12] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, no. 9, pp. 973–977, Nov. 1987, doi: 10.1209/0295-5075/4/9/004.
- [13] A. S. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, and L. A. N. Amaral, "Duality between time series and networks," *PLoS ONE*, vol. 6, no. 8, e23378, Aug. 2011, doi: 10.1371/journal.pone.0023378.
- [14] Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W., & Schaeffter, T. (2022). PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). *PhysioNet*. RRID:SCR_007345.
- [15] Y. Elmir, Y. Himeur and A. Amira, ECG classification using Deep CNN and Gramian Angular Field, *2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService)*, Athens, Greece, 2023, pp. 137-141, 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*, 2023, 11, 1000. <https://doi.org/10.3390/healthcare11071000>