



TRƯỜNG ĐẠI HỌC

SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

HCMC University of Technology and Education

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

KHOA ĐIỆN ĐIỆN TỬ

TIỂU LUẬN MÔN HỌC

DeepArr: Nghiên cứu phát hiện rối loạn nhịp tim sử dụng mạng nơ-ron học sâu với ngữ cảnh từ tín hiệu điện tâm đồ (ECG)

Ngành Kỹ Thuật Điện Tử

Học viên: Phạm Thành Trung

MSHV: 2530712

GVHD TS. Trần vũ hoàng

TP. HỒ CHÍ MINH – 10/2025

MỤC LỤC

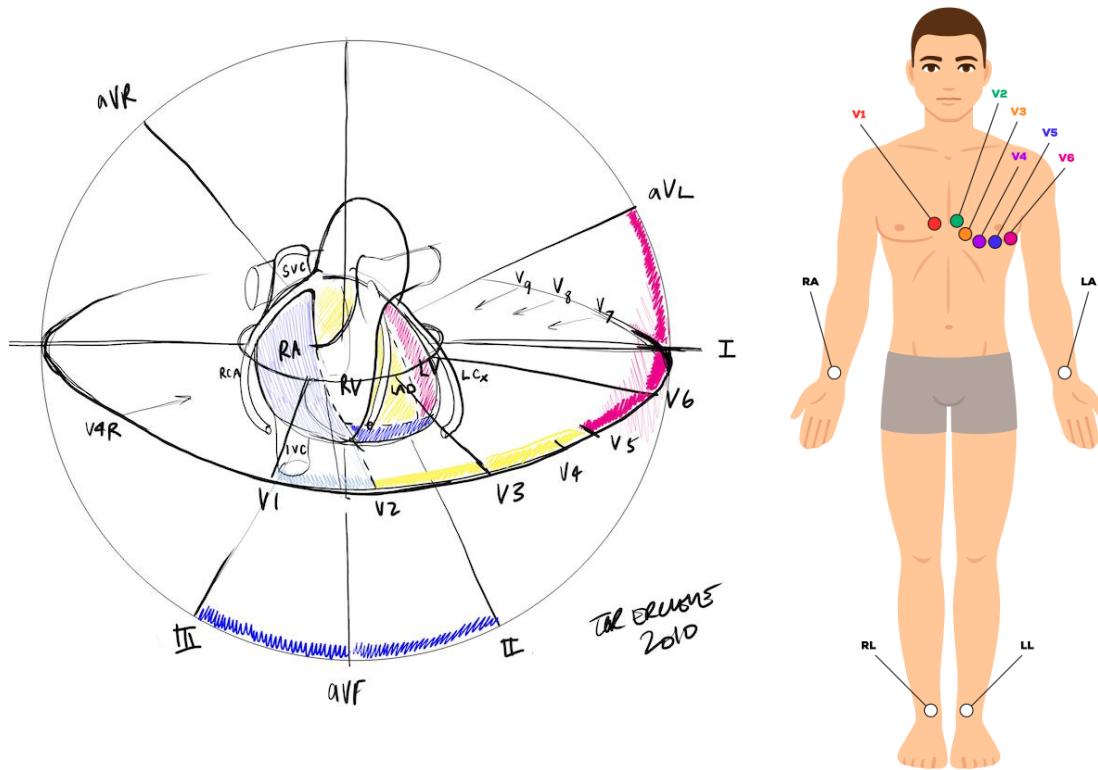
I.	GIỚI THIỆU	3
II.	CƠ SỞ LÝ THUYẾT	5
A.	Đặc điểm tín hiệu điện tâm đồ	5
B.	Các hướng tiếp cận truyền thống	6
C.	Hạn chế và thách thức của phương pháp truyền thống	7
III.	PHƯƠNG PHÁP ĐỀ XUẤT	8
A.	Xu hướng học sâu trong phân loại ECG	8
B.	Các thách thức chính trong bài toán phân loại ECG bằng học sâu	9
C.	Các hướng tiếp cận trong các nghiên cứu trước	10
D.	Nhận xét và cơ sở hình thành hướng nghiên cứu của đề tài	11
E.	Tổng quan quy trình hệ thống	11
F.	Tiền xử lý	12
1.	Resampling	12
2.	Chuẩn hóa (Normalization)	12
3.	Cắt nhịp (Segmentation)	13
G.	Cấu trúc mô hình DeepArr	13
1.	CNN – Trích xuất đặc trưng không gian	13
2.	BiLSTM – Mô hình hóa phụ thuộc thời gian	14
3.	Khối phân loại (Fully Connected)	14
4.	Hàm mất mát và tối ưu	14
H.	Chiến lược huấn luyện và đánh giá	15
IV.	KẾT QUẢ THỰC NGHIỆM	15
A.	Dữ liệu huấn luyện và môi trường thử nghiệm	15
1.	Bộ dữ liệu MIT-BIH Arrhythmia	15
2.	Cấu hình huấn luyện	16
B.	Các tiêu chí đánh giá	17
C.	Hiệu quả huấn luyện và hội tụ mô hình	18
D.	Đánh giá định lượng	19
E.	Confusion matrix và phân tích lỗi	19

F.	So sánh với các nghiên cứu khác	20
G.	Trực quan hóa kết quả dự đoán (Visualization of Predicted ECG Segments).....	21
V.	KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	24
A.	Hạn chế	24
B.	Hướng phát triển.....	24
	TÀI LIỆU THAM KHẢO	25
	Phụ Lục A – Cấu hình huấn luyện mô hình DeepArr	26
	Phụ lục B – Cấu trúc mạng CNN-BiLSTM	26
	Phụ lục C – Mã nguồn	26

I. GIỚI THIỆU

Trong thực hành lâm sàng, việc phát hiện chính xác rối loạn nhịp tim (cardiac arrhythmia) đóng vai trò đặc biệt quan trọng, vì đây là một trong những nguyên nhân hàng đầu gây tử vong do bệnh tim mạch (Cardiovascular Diseases – CVD), chiếm tới khoảng 80% các trường hợp tử vong đột ngột liên quan đến tim. Rối loạn nhịp tim là tình trạng bất thường trong hoạt động điện học của tim, khiến tim không thể bơm máu hiệu quả tới các cơ quan, dẫn đến thiếu oxy và các biến chứng nghiêm trọng khác.

Phương tiện chẩn đoán phổ biến và đáng tin cậy nhất hiện nay là điện tâm đồ (Electrocardiogram – ECG), một công cụ ghi nhận hoạt động điện của tim theo thời gian thông qua các điện cực gắn trên bề mặt cơ thể. Một bản ghi ECG tiêu chuẩn bao gồm 12 chuyển đạo (12-lead configuration), trong đó có sáu chuyển đạo chi (I, II, III, aVR, aVL, aVF) và sáu chuyển đạo ngực (V1–V6). Các tín hiệu ECG này phản ánh hoạt động điện học của tim ở nhiều góc độ khác nhau, giúp nhận biết các dạng sóng P, QRS và T – những thành phần then chốt trong phân tích trạng thái nhịp tim được miêu tả ở hình 1.



Hình 1: Sơ đồ vị trí các điện cực và hướng thu tín hiệu trong hệ thống điện tâm đồ 12 đạo trình (12-lead ECG).

Những thiết bị đầu tiên được sử dụng cho chẩn đoán rối loạn nhịp tim bao gồm Holter monitor, event recorder, và implantable cardiac monitor (ICM). Tuy nhiên, các thiết bị này thường bị giới hạn bởi thời gian ghi và khả năng theo dõi liên tục. Sự phát triển của công nghệ đeo (wearable technology) đã mở ra kỷ nguyên mới trong theo dõi tim mạch, với các thiết bị như đồng

hồ thông minh, miếng dán ngực, hoặc cảm biến tích hợp trong quần áo, cho phép ghi ECG liên tục trong thời gian dài, ngay cả trong điều kiện sinh hoạt thường ngày.

Dẫu vậy, việc phân tích tín hiệu ECG vẫn còn gặp nhiều thách thức do các yếu tố như:

- Tính phi tuyến và không dừng (non-stationary nature) của tín hiệu, gây khó khăn trong việc phát hiện rối loạn nhịp ở giai đoạn sớm.
- Sự khác biệt cá nhân (inter-patient variability) trong hình thái tín hiệu ECG do giới tính, độ tuổi và đặc điểm sinh lý.
- Nhiễu và biến dạng tín hiệu (signal noise and distortion) trong quá trình thu nhận dữ liệu.
- Khối lượng dữ liệu khổng lồ, khiến việc phân tích thủ công trở nên không khả thi.

Trong nhiều năm qua, các phương pháp học máy truyền thống (Machine Learning – ML) như SVM, k-NN, hoặc Hybrid SVM đã được áp dụng cho bài toán phân loại ECG. Tuy nhiên, những phương pháp này phụ thuộc nặng nề vào đặc trưng thủ công (handcrafted features) như biến đổi wavelet (DWT), PCA hoặc năng lượng sóng, làm hạn chế khả năng khái quát hóa khi áp dụng trên nhiều bệnh nhân khác nhau.

Sự ra đời của học sâu (Deep Learning) đã mở ra hướng đi mới cho lĩnh vực phân tích tín hiệu ECG. Các mô hình như Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), và đặc biệt là Bidirectional LSTM (BiLSTM) có khả năng tự động học đặc trưng từ dữ liệu thô, giúp cải thiện đáng kể hiệu quả nhận dạng rối loạn nhịp tim. Một số nghiên cứu điển hình như của Hannun *et al.* và Nguyen *et al.* đã chứng minh khả năng của các mô hình end-to-end deep neural network trong việc phát hiện bất thường nhịp tim với độ chính xác vượt trội so với các phương pháp truyền thống.

Trong bối cảnh đó, Midani *et al.* (2023) đã giới thiệu mô hình DeepArr – một kiến trúc lai (hybrid architecture) giữa CNN và BiLSTM, được thiết kế để kết hợp song song đặc trưng không gian (spatial features) và đặc trưng thời gian (temporal features) của tín hiệu ECG. Phương pháp này tận dụng khả năng của CNN trong việc trích xuất mẫu cục bộ (local patterns) và khả năng của BiLSTM trong học mối quan hệ phụ thuộc dài hạn (long-term dependencies), nhờ đó cải thiện đáng kể độ chính xác phân loại.

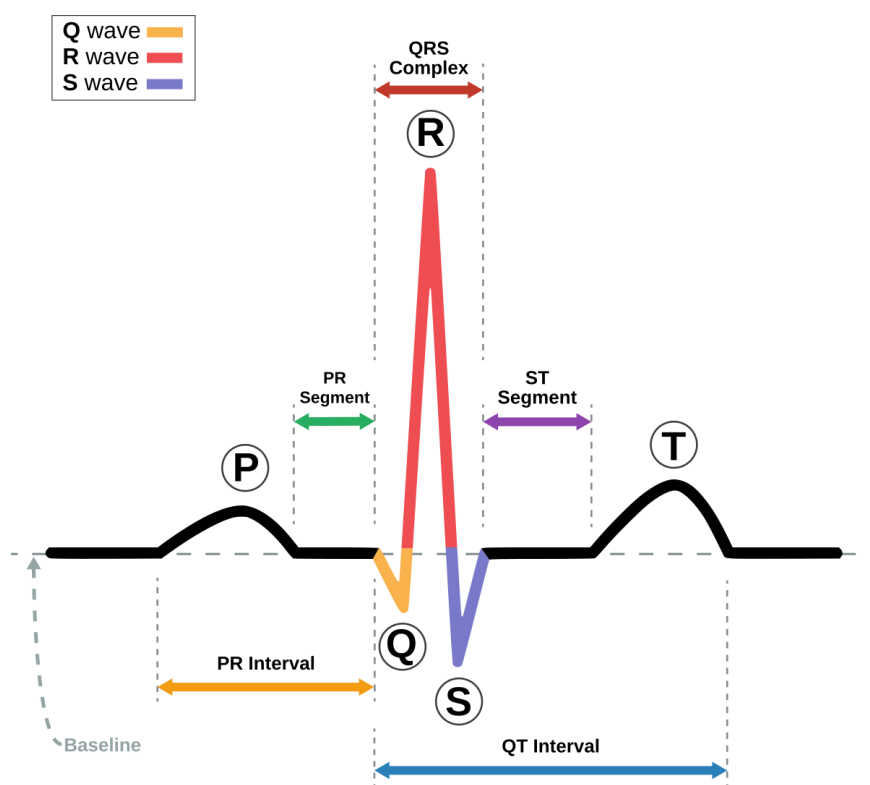
Kết quả thực nghiệm cho thấy mô hình DeepArr đạt hiệu năng vượt trội trên cơ sở dữ liệu MIT-BIH Arrhythmia Database, với độ chính xác (accuracy) 99.46%, độ đặc hiệu (specificity) 99.57%, độ nhạy (sensitivity) 97.01%, độ chính xác (precision) 98.26%, và F1-score 97.63%. Mô hình này chứng minh tính hiệu quả, độ tin cậy và tiềm năng ứng dụng cao trong hệ thống phát hiện sớm rối loạn nhịp tim tự động, góp phần hỗ trợ quá trình chẩn đoán và điều trị lâm sàng một cách chính xác và kịp thời.

II. CƠ SỞ LÝ THUYẾT

A. Đặc điểm tín hiệu điện tâm đồ

Tín hiệu ECG là dạng tín hiệu sinh lý một chiều được thu từ các điện cực gắn trên cơ thể. Một chu kỳ ECG điển hình bao gồm 5 thành phần chính: sóng P, phức QRS, và sóng T ở hình 2.

- Sóng P biểu thị quá trình khử cực tâm nhĩ.
- Phức QRS phản ánh hoạt động khử cực của tâm thất, là vùng quan trọng nhất để nhận diện nhịp tim.
- Sóng T thể hiện quá trình tái cực của tâm thất.



Hình 2: Dạng sóng ECG chuẩn gồm sóng P, phức QRS và sóng T biểu diễn các pha hoạt động điện của tim.

Các bất thường trong biên độ, khoảng thời gian, hoặc hình dạng của các sóng này là chỉ báo trực tiếp của các rối loạn nhịp. Ví dụ, kéo dài phức QRS gợi ý tình trạng block nhánh (LBBB/RBBB), còn sóng P sớm hoặc biến dạng thường là dấu hiệu của nhịp ngoại tâm nhĩ (APB).

Do ECG có sự khác biệt lớn giữa các cá nhân, nhiều đo đạc, và biến thiên theo thời gian, nên việc mô hình hóa chính xác tín hiệu này là một thách thức lớn.

Phương pháp được đề xuất được phát triển và đánh giá trên bộ dữ liệu MIT-BIH Arrhythmia Database (MITDB). Tập dữ liệu và các tệp chú thích (annotations) được tải xuống từ cơ sở dữ liệu

công khai PhysioNet. MITDB bao gồm 48 bản ghi thu được từ 47 bệnh nhân với độ tuổi và giới tính khác nhau (trong đó bản ghi 201 và 202 thuộc cùng một bệnh nhân). Mỗi bản ghi chứa 30 phút tín hiệu ECG hai kênh, được số hóa ở tần số lấy mẫu 360 Hz, với độ phân giải 11-bit cho mỗi mẫu. Trong nghiên cứu này, chúng tôi lựa chọn một số lớp nhịp tim phổ biến thường được sử dụng trong hầu hết các công trình trước.

Các loại nhịp tim được xem xét bao gồm năm lớp chính như sau:

- **Nhịp bình thường (Normal beats – N):** Đặc trưng bởi phức bộ QRS có thời lượng < 100 ms. Trong chu kỳ tim, mỗi phức bộ QRS được đi trước bởi sóng P bình thường, có dạng thẳng đứng ở các đạo trình I và II, đồng thời âm trong đạo trình aVR. Dấu hiệu chính của nhịp bình thường là khoảng PR ổn định và nhịp điệu đều trong khoảng 60–100 bpm.
- **Block nhánh phải (Right Bundle Branch Block – RBBB):** Đặc trưng bởi phức bộ QRS ≥ 120 ms. Ở các đạo trình V1 và V2, QRS có hình dạng giống chữ “M”, với sóng S nhỏ và sóng R đầu tiên nhỏ hơn sóng R tiếp theo. Đồng thời, đoạn ST chênh xuống và sóng T đảo chiều. Các đạo trình V5 và V6 có sóng S rộng, với biên độ khoảng 60 ms lớn hơn sóng R.
- **Block nhánh trái (Left Bundle Branch Block – LBBB):** Đặc trưng bởi thời lượng phức bộ QRS ≥ 120 ms, hoàn thiện hình dạng đặc trưng của block nhánh trái. Ngoài ra, LBBB thường đi kèm với sự kéo dài thời lượng QRS, do thay đổi sóng R ở đạo trình V5, V6 và sóng S sâu ở đạo trình V1, V2. Bên cạnh đó, độ biến thiên của đoạn ST–T thường xuất hiện ở các vị trí block này.
- **Nhịp sớm nhĩ (Atrial Premature Beats – APB):** Đặc trưng bởi sóng P xuất hiện sớm hơn so với nhịp bình thường. Ngoài ra, hình dạng sóng P bất thường so với sóng P bình thường, trong khi sóng P thường hướng lên ở hầu hết các đạo trình, ngoại trừ aVR, nơi sóng P bị đảo ngược.
- **Ngoại tâm thu thất (Premature Ventricular Contraction – PVC):** Đặc trưng bởi phức bộ QRS ≥ 120 ms, xuất hiện sớm hơn dự kiến trong chu kỳ tim. Các dạng sóng bất thường của PVC thể hiện sự lệch pha đoạn ST và biến thiên của sóng T. Ngoài ra, mỗi PVC thường đi kèm một khoảng nghỉ bù (compensatory pause) ngay sau đó.

B. Các hướng tiếp cận truyền thống

Trước thời kỳ học sâu, các phương pháp nhận dạng ECG chủ yếu dựa trên biến đổi miền tần số và phân tích đặc trưng thủ công, ví dụ:

- Wavelet Transform (DWT): tách tín hiệu thành các băng tần khác nhau để trích xuất năng lượng, entropy, hệ số thống kê.
- Higher Order Statistics (HOS) hoặc Hidden Markov Models (HMM): mô hình hóa tính ngẫu nhiên và mối quan hệ thời gian của tín hiệu.
- Discrete Cosine Transform (DCT): dùng cho nén và nhận dạng đặc trưng ECG với chi phí tính toán thấp.

C. Hạn chế và thách thức của phương pháp truyền thống

Mặc dù các phương pháp truyền thống đã đạt được độ chính xác tương đối cao trên các bộ dữ liệu được kiểm soát trong phòng thí nghiệm, chúng vẫn tồn tại nhiều hạn chế khiến việc mở rộng và ứng dụng trong môi trường thực tế gặp nhiều khó khăn. Thứ nhất, các phương pháp này phụ thuộc mạnh vào đặc trưng thủ công (handcrafted features). Việc lựa chọn và thiết kế đặc trưng như năng lượng sóng QRS, độ lệch chuẩn RR, hệ số entropy, hay năng lượng miền tần số thường dựa vào kinh nghiệm và hiểu biết chuyên môn của người thiết kế, dẫn đến khả năng tổng quát kém khi áp dụng trên các dữ liệu ECG thu từ nhiều bệnh nhân khác nhau hoặc từ các thiết bị cảm biến khác nhau. Điều này làm cho mô hình dễ bị sai lệch khi gặp tín hiệu bị nhiễu hoặc có hình thái khác biệt.

Thứ hai, các mô hình truyền thống thiếu khả năng học các mối quan hệ phức tạp và dài hạn trong tín hiệu. Các đặc trưng thống kê hoặc miền tần số chỉ phản ánh được các dao động cục bộ của tín hiệu, trong khi nhịp tim lại có sự phụ thuộc theo thời gian với các chu kỳ lặp lại và biến đổi theo ngữ cảnh sinh lý. Những mối tương quan dài hạn này (long-term dependencies) không thể được nắm bắt hiệu quả bằng các phương pháp phi tuyến như biến đổi wavelet hay DCT, vốn bị giới hạn trong miền tần số ngắn hạn.

Thứ ba, các phương pháp truyền thống khó thích nghi với dữ liệu quy mô lớn và đa dạng. Khi kích thước tập dữ liệu tăng lên hoặc khi dữ liệu chứa nhiều biến thể sinh lý và nhiễu đo lường, việc tái huấn luyện, tinh chỉnh và mở rộng mô hình trở nên phức tạp. Các thuật toán như SVM, KNN hoặc Random Forest có thể gặp giới hạn về năng lực xử lý và không tận dụng được lợi thế của dữ liệu quy mô lớn như các mô hình học sâu hiện đại.

Cuối cùng, hiệu suất của các phương pháp này bị ảnh hưởng đáng kể bởi nhiễu và sai lệch thiết bị. Các tín hiệu ECG thực tế thường chịu tác động của nhiễu cơ học, sai lệch cảm biến, dịch nền (baseline wander), hoặc biến động do hoạt động hô hấp và chuyển động cơ thể, khiến đặc trưng trích xuất bị méo dạng, làm giảm tính ổn định và độ tin cậy của kết quả phân loại.

Trước những hạn chế đó, sự phát triển của học sâu (Deep Learning) đã mở ra một hướng tiếp cận mới và hiệu quả hơn cho bài toán phân loại ECG. Thay vì phải thiết kế đặc trưng thủ công, các mạng nơ-ron tích chập (CNN) và mạng hồi tiếp (RNN, LSTM, BiLSTM) có khả năng tự động học đặc trưng (automatic feature learning) trực tiếp từ tín hiệu ECG thô, giúp khai thác đồng thời cả đặc trưng không gian (spatial) và đặc trưng thời gian (temporal). Các mô hình học sâu không chỉ giảm thiểu sự phụ thuộc vào kinh nghiệm chuyên gia mà còn nâng cao khả năng khái quát (generalization) và cải thiện đáng kể độ chính xác khi xử lý dữ liệu nhiễu hoặc mất cân bằng. Nhờ khả năng học phi tuyến mạnh mẽ, mô hình học sâu có thể phát hiện các mẫu hình phức tạp trong tín hiệu tim, vốn thường không thể được mô tả bằng các phương pháp truyền thống, từ đó mang lại hiệu năng vượt trội trong các hệ thống giám sát và chẩn đoán tim mạch tự động hiện nay.

Ngoài ra, nhờ khả năng học sâu theo tầng (hierarchical representation learning), các mô hình Deep Learning có thể tự động phát hiện những đặc trưng trừu tượng hơn ở các tầng cao, chẳng hạn

như hình thái sóng, biên độ dao động hoặc xu hướng biến đổi nhịp theo thời gian. Điều này giúp cải thiện đáng kể độ chính xác (Accuracy), độ nhạy (Sensitivity) và khả năng khái quát (Generalization) trên các tập dữ liệu độc lập.

Nhờ những ưu điểm vượt trội trên, học sâu đã nhanh chóng trở thành hướng tiếp cận chủ đạo trong lĩnh vực phân tích tín hiệu ECG, thay thế dần các phương pháp dựa trên đặc trưng thủ công. Đặc biệt, các mô hình lai CNN-BiLSTM như DeepArr đã được chứng minh có hiệu năng cao và khả năng thích ứng linh hoạt, trở thành nền tảng vững chắc cho các hệ thống chẩn đoán rối loạn nhịp tim tự động và thời gian thực trong y sinh học hiện đại.

III. PHƯƠNG PHÁP ĐỀ XUẤT

A. Xu hướng học sâu trong phân loại ECG

Sự ra đời của các kiến trúc mạng học sâu như CNN, RNN, LSTM, và các biến thể hiện đại hơn đã tạo nên bước ngoặt lớn trong lĩnh vực phân tích tín hiệu y sinh, đặc biệt là phân loại tín hiệu điện tim (ECG).

- **CNN 1D** có khả năng phát hiện các mẫu hình cục bộ trong chuỗi ECG, ví dụ như các dạng QRS đặc trưng, thông qua các bộ lọc trượt (kernels) để trích xuất đặc trưng hình dạng sóng.
- **LSTM/BiLSTM** lại mạnh ở việc nắm bắt mối quan hệ chuỗi theo thời gian, giúp mô hình hiểu được ngữ cảnh trước-sau của các nhịp tim, qua đó tăng độ chính xác trong nhận dạng rối loạn nhịp.
- **TCN (Temporal Convolutional Network)** là một hướng phát triển mới của CNN cho dữ liệu chuỗi, sử dụng convolution nhân nguyên nhân (causal convolution) và skip connection để mô hình hóa phụ thuộc dài hạn trong thời gian, đồng thời giảm chi phí huấn luyện so với LSTM.
- **DNN (Deep Neural Network)** với các lớp fully connected sâu được sử dụng để phân loại đặc trưng trích xuất sẵn từ tín hiệu ECG, phù hợp cho các hệ thống có tài nguyên hạn chế hoặc kết hợp với các mô hình trích đặc trưng khác (như CNN + DNN).
- **Transformer và các biến thể nhẹ (Tiny-Transformer, ECG-Former)** đang được ứng dụng trong nghiên cứu gần đây nhờ khả năng tự chú ý (self-attention), giúp mô hình nhận diện mối quan hệ dài hạn giữa các đoạn sóng ECG mà không cần lan truyền tuần tự như RNN.

Nhờ những tiến bộ vượt bậc trong kiến trúc mạng và kỹ thuật tối ưu, các mô hình học sâu (Deep Learning models) ngày nay không chỉ đạt độ chính xác cao hơn đáng kể so với các phương pháp học máy truyền thống như Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) hay Random Forest, mà còn thể hiện khả năng thích ứng linh hoạt hơn khi xử lý dữ liệu ECG trong các điều kiện thực tế đa dạng.

Trong khi các mô hình truyền thống thường gặp khó khăn khi dữ liệu chứa nhiễu, biến thiên biên độ hoặc lệch đường nền, thì các mô hình học sâu có khả năng tự động học biểu diễn đặc trưng thông qua nhiều tầng trừu tượng hóa, giúp chúng duy trì hiệu năng ổn định ngay cả trên dữ liệu nhiễu hoặc không đồng nhất giữa các bệnh nhân. Bên cạnh đó, việc huấn luyện trên tập dữ liệu lớn giúp mạng nơ-ron học được những quy luật tiềm ẩn mà các phương pháp tuyến tính khó nắm bắt.

Nguồn	Năm	Tiền xử lý	Biểu diễn đặc trưng	Mô hình phân loại	Lớp	Datasets	Kết quả
Liu et al. [13]	2019	RR segmentation; Sampled at 90 Hz; Delete first and last wave	DWT	BLSTM + 2D CNN	5	MIT-BIH Arrhythmia Database	Acc = 99.50%; Sens = 99.90%; Spe = 98.20%
Nurmaini et al. [12]	2020	RR segmentation; Normalization	Stacked denoising auto-encoders	DNN	5	MIT-BIH Arrhythmia Database; MIT-BIH noise stress test database	Acc = 99.34%; Sens = 93.83%; Spe = 99.57%; F1 = 91.44%
Cui et al. [14]	2021	Denosing; RR detection; RR segmentation; Imbalanced processing	RR interval feature; DWT feature; CNN feature	SVM	5	MIT-BIH Arrhythmia Database	Acc = 98.35%
El Bouny et al. [11]	2020	RR segmentation; Resample; Augmentation dataset	End-to-end structure	ML-WCNN = STW + CNN	6	MIT-BIH Arrhythmia Database	Acc = 99.57%
Qiao et al. [8]	2020	RR segmentation; Pan-Tompkins; Resample; Normalization	BLSTM	DELM-LRF-BLSTM	5	MIT-BIH Arrhythmia Database	Acc = 99.32%

Bảng 1. Tổng hợp một số nghiên cứu gần đây về phân loại tín hiệu ECG sử dụng các mô hình học sâu và phương pháp tiền xử lý khác nhau.

B. Các thách thức chính trong bài toán phân loại ECG bằng học sâu

- Đặc trưng tín hiệu phức tạp và nhiễu cao:** Dữ liệu ECG thường bị ảnh hưởng bởi nhiều yếu tố như nhiễu cơ, nhiễu điện lưới và dịch chuyển đường nền, đồng thời biến thiên đáng kể giữa các cá nhân. Điều này gây khó khăn cho quá trình trích xuất đặc trưng ổn định và chính xác.
- Mất cân bằng lớp (class imbalance):** Trong các tập dữ liệu như MIT-BIH hay PTB-XL, nhịp bình thường chiếm tỷ lệ lớn, trong khi các loại rối loạn nhịp ít xuất hiện. Sự chênh lệch này khiến mô hình học sâu có xu hướng dự đoán nghiêng về lớp chiếm đa số.
- Khó khăn trong việc học mối quan hệ dài hạn:** ECG là tín hiệu chuỗi theo thời gian, việc nắm bắt mối quan hệ giữa các nhịp tim xa nhau là thách thức đối với các mô hình truyền thống chỉ xử lý được ngữ cảnh ngắn.
- Chi phí huấn luyện và tối ưu hóa mô hình:** Các kiến trúc mạng nhiều tầng như BiLSTM hoặc Transformer đòi hỏi tài nguyên tính toán lớn, trong khi việc điều chỉnh siêu tham số (hyperparameters) phức tạp, dễ dẫn đến quá khớp (overfitting).
- Hạn chế trong khả năng giải thích kết quả (model interpretability):** Học sâu thường được xem là “hộp đen”, gây khó khăn trong việc xác định vùng tín hiệu ảnh hưởng đến quyết định phân loại — một yêu cầu quan trọng trong lĩnh vực y sinh.

C. Các hướng tiếp cận trong các nghiên cứu trước

1. **Mạng CNN một chiều (1D-CNN):** Nhiều nghiên cứu ban đầu (Kiranyaz et al., 2019; Ahmed et al., 2023) đã sử dụng CNN 1D để trích xuất đặc trưng hình thái sóng QRS, P, T.

Ưu điểm: Học được đặc trưng cục bộ hiệu quả, xử lý nhanh, dễ triển khai.

Hạn chế: Không thể mô hình hóa mối quan hệ chuỗi dài hạn → chưa khắc phục được thách thức (3) và (5).

2. **Mạng hồi quy LSTM/BiLSTM:** Các mô hình LSTM và BiLSTM (Cheng et al., 2021) được áp dụng để học mối quan hệ theo chuỗi thời gian, giúp mô hình hiểu ngữ cảnh trước–sau của nhịp tim.

Ưu điểm: Giải quyết tốt thách thức (3), cải thiện nhận dạng các mẫu động.

Hạn chế: Tốc độ huấn luyện chậm, tiêu tốn bộ nhớ, dễ bị overfitting khi dữ liệu hạn chế (thách thức 4).

3. **Mô hình lai CNN–BiLSTM:** Để tận dụng cả đặc trưng cục bộ và chuỗi thời gian, nhiều công trình (Ye et al., 2025) đã đề xuất CNN–BiLSTM.

Ưu điểm: Kết hợp hiệu quả giữa không gian và thời gian, cải thiện độ chính xác tổng thể.

Hạn chế: Số lượng tham số lớn, yêu cầu tối ưu hóa cao.

4. **TCN (Temporal Convolutional Network):** thay thế cấu trúc RNN bằng causal convolution và skip connection, giúp mở rộng phạm vi phụ thuộc dài mà vẫn huấn luyện nhanh.

Ưu điểm: Giải quyết được thách thức (3) và (4) — mô hình hội tụ nhanh, tránh gradient vanishing.

Hạn chế: Khả năng chú ý ngữ cảnh toàn cục còn hạn chế.

5. **Kiến trúc Transformer và Attention:** Các biến thể như ECGFormer, ECGTransForm, Swin-Transformer (2024) ứng dụng cơ chế self-attention để học mối quan hệ toàn cục giữa các đoạn tín hiệu.

Ưu điểm: Học được phụ thuộc dài hạn, song song hóa tốt.

Hạn chế: Cần dữ liệu lớn và khó giải thích kết quả → liên quan đến thách thức (5).

D. Nhận xét và cơ sở hình thành hướng nghiên cứu của đề tài

Từ những phân tích tổng quan ở trên có thể nhận thấy rằng, các nghiên cứu trước đây đã đóng góp đáng kể trong việc nâng cao hiệu quả của bài toán phân loại tín hiệu điện tâm đồ (ECG), đặc biệt là ở khả năng học đặc trưng tự động và cải thiện độ chính xác của mô hình. Các hướng tiếp cận dựa trên học sâu như CNN, LSTM hay các mô hình lai CNN-BiLSTM đã chứng minh tính ưu việt so với các phương pháp truyền thống, giúp mô hình có thể khai thác đồng thời thông tin không gian và thời gian trong tín hiệu ECG. Tuy nhiên, bên cạnh những thành tựu đạt được, vẫn còn tồn tại một số vấn đề cần được khắc phục để đảm bảo hiệu quả và tính ổn định trong thực tế.

Thứ nhất, vấn đề mất cân bằng dữ liệu (data imbalance) vẫn là một thách thức lớn. Trong hầu hết các tập dữ liệu ECG, số lượng mẫu của các nhịp bình thường thường vượt trội so với các nhịp bất thường như PVC hay APB. Sự chênh lệch này khiến mô hình có xu hướng nghiêng về lớp chiếm đa số, làm giảm khả năng nhận dạng chính xác các nhịp hiếm nhưng có ý nghĩa y học quan trọng. Thứ hai, một số mô hình học sâu hiện tại vẫn chưa đạt được sự cân bằng tối ưu giữa độ chính xác (accuracy) và khả năng khái quát hóa (generalization). Việc tập trung quá mức vào độ chính xác huấn luyện có thể dẫn đến hiện tượng quá khớp (overfitting), khiến mô hình hoạt động kém hiệu quả trên dữ liệu thực hoặc dữ liệu chưa từng thấy trong quá trình huấn luyện. Cuối cùng, tính minh bạch và khả năng giải thích (interpretability) của các mô hình học sâu vẫn còn hạn chế. Trong lĩnh vực y sinh, việc hiểu rõ cơ chế quyết định của mô hình – tức là vùng tín hiệu hoặc đặc trưng nào ảnh hưởng đến kết quả phân loại – là yếu tố then chốt giúp tăng độ tin cậy khi áp dụng trong chẩn đoán lâm sàng.

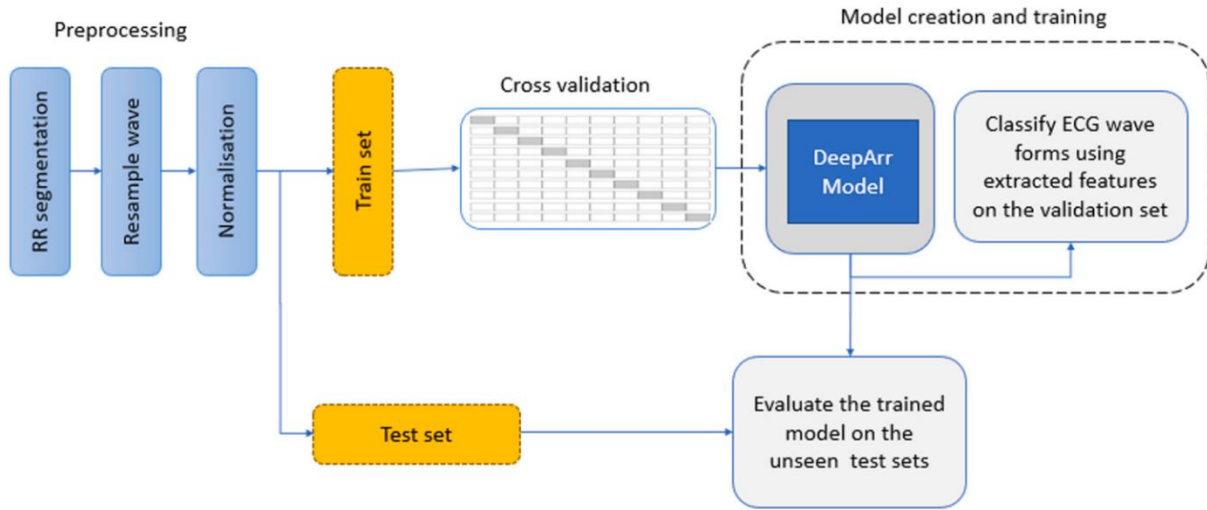
Dựa trên những hạn chế nêu trên, đề tài này được định hướng nhằm xây dựng một mô hình học sâu tối ưu cho bài toán phân loại ECG tự động, với ba mục tiêu trọng tâm. Thứ nhất, cải thiện chất lượng dữ liệu huấn luyện thông qua các kỹ thuật tăng cường dữ liệu (augmentation) và cân bằng lớp bằng phương pháp SMOTE (Synthetic Minority Over-sampling Technique), nhằm khắc phục hiện tượng mất cân bằng giữa các loại nhịp tim. Thứ hai, đề xuất và phát triển kiến trúc mô hình CNN kết hợp BiLSTM (DeepArr) – một cấu trúc lai có khả năng học đồng thời đặc trưng không gian và phụ thuộc thời gian, giúp mô hình không chỉ nắm bắt được các đặc trưng cục bộ của tín hiệu ECG mà còn phát hiện được các mối quan hệ dài hạn giữa các nhịp tim, đồng thời tăng cường khả năng giải thích. Cuối cùng, đánh giá toàn diện hiệu năng của mô hình bằng các chỉ số định lượng quan trọng gồm độ chính xác (Accuracy), điểm F1 (F1-Score), độ nhạy (Sensitivity) và độ đặc hiệu (Specificity), nhằm đảm bảo tính khách quan, độ tin cậy và khả năng khái quát hóa cao của mô hình trên các tập dữ liệu ECG thực tế.

E. Tổng quan quy trình hệ thống

Hình 3 minh họa pipeline tổng thể của hệ thống phát hiện rối loạn nhịp tim đề xuất. Quy trình gồm năm khối chính:

1. Thu thập và đọc dữ liệu ECG: Tín hiệu ECG được lấy từ MIT-BIH Arrhythmia Database dưới định dạng .hea, .dat, và .atr.

2. Phát hiện và cắt nhịp tim (beat segmentation): Dựa trên vị trí R-peak trong annotation file .atr, tín hiệu được cắt thành từng nhịp riêng lẻ, mỗi nhịp gồm 0.2 giây trước và 0.6 giây sau R-peak.
3. Tiền xử lý và chuẩn hóa: Mỗi nhịp được resample về 125 Hz (chuẩn hóa tần số lấy mẫu), độ dài cố định 250 điểm dữ liệu, sau đó chuẩn hóa Z-score theo từng kênh.
4. Huấn luyện mô hình DeepArr (CNN-BiLSTM): Dữ liệu đã xử lý được đưa vào mạng học sâu lai gồm các tầng CNN và BiLSTM để trích xuất đặc trưng không gian-thời gian.
5. Đánh giá kết quả: Mô hình được huấn luyện bằng 10-fold Stratified Cross-Validation, đánh giá bằng các chỉ số Accuracy, Precision, Recall, F1-score và Confusion matrix.



Hình 3: Cấu trúc của phương pháp được đề xuất bao gồm các bước tiền xử lý, huấn luyện và kiểm định mô hình trên bộ dữ liệu MIT-BIH, kèm theo giai đoạn kiểm thử.

F. Tiền xử lý

1. Resampling

Tín hiệu ECG trong MIT-BIH có tần số gốc 360 Hz. Để giảm chi phí tính toán và đồng bộ độ dài các nhịp, tất cả tín hiệu được resample về 125 Hz bằng phép nội suy Fourier. Giả sử tín hiệu gốc $x(t)$ có N mẫu với tần số f_s , tín hiệu sau resampling $x'(t)$ có:

$$N' = N \times \frac{f_t}{f_s} \text{ mẫu với tần số } f_s, \text{ trong đó } f_t = 125 \text{ Hz}$$

$$x'(n) = \text{IFFT}(\text{ZeroPad}(\text{FFT}(x(n)), N')) \quad (1)$$

2. Chuẩn hóa (Normalization)

Để loại bỏ ảnh hưởng biên độ và nhiễu nền, mỗi nhịp được chuẩn hóa theo Z-score:

$$x_{\text{norm}} = \frac{x - \mu_x}{\sigma_x + \epsilon} \quad (2)$$

trong đó:

- μ_x : giá trị trung bình theo từng kênh,
- σ_x : độ lệch chuẩn,
- $\epsilon = 10^{-6}$: hằng số tránh chia cho 0.

3. Cắt nhịp (Segmentation)

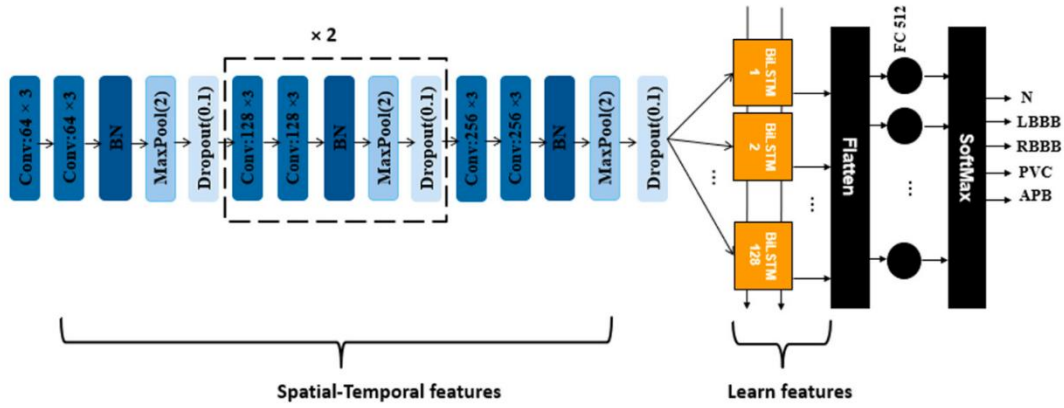
Dựa vào annotation .atr của PhysioNet, mỗi nhịp được trích từ vị trí i có ký hiệu nhịp S_i :

$$x_i = [x(i - 0.2f_s), x(i + 0.6f_s)] \quad (3)$$

Những nhịp nằm ngoài biên hoặc thiếu annotation được loại bỏ. Sau đó, mỗi nhịp được biểu diễn dạng ma trận $X_i \in \mathbb{R}^{250 \times 2}$, tương ứng 250 mẫu và 2 kênh (MLII, V1).

G. Cấu trúc mô hình DeepArr

Mô hình DeepArr được xây dựng theo hướng end-to-end deep learning, cho phép hệ thống tự động phân tích tín hiệu ECG mà không cần trích xuất đặc trưng thủ công. Nói cách khác, mô hình có thể học trực tiếp các biểu diễn đặc trưng từ tín hiệu thô, giúp giảm thiểu sự phụ thuộc vào kiến thức chuyên ngành hoặc giai đoạn tiền xử lý phức tạp được mô tả ở hình 4.



Hình 4. Kiến trúc DeepArr được đề xuất với khả năng trích xuất đặc trưng không gian–thời gian.

1. CNN – Trích xuất đặc trưng không gian

Khối đầu tiên gồm nhiều tầng tích chập một chiều (1D convolution):

$$h^{(l)} = f(W^{(l)} * h^{(l-1)} + b^{(l)}) \quad (4)$$

với:

- $*$: phép tích chập,
- $f(x)$: hàm kích hoạt ReLU,

- $W^{(l)}$: kernel 3×1 ,
- $b^{(l)}$: vector bias.

Mỗi khối tích chập gồm hai tầng Conv1D \rightarrow BatchNorm \rightarrow MaxPool(2) \rightarrow Dropout(0.1) nhằm giảm phương sai và tăng khả năng khái quát.

Các tầng có số filters tăng dần: $64 \rightarrow 128 \rightarrow 256$, tương ứng ba mức độ trừu tượng của đặc trưng ECG (QRS \rightarrow P/T \rightarrow toàn chu kỳ).

2. BiLSTM – Mô hình hóa phụ thuộc thời gian

Sau CNN, đầu ra $H_{CNN} \in \mathbb{R}^{T' \times d}$ được đưa vào chuỗi Ba tầng BiLSTM có kích thước 128, 64 và 32 đơn vị. Mỗi BiLSTM gồm hai hướng truyền (forward và backward):

$$\vec{h}_t = \text{LSTM}_f(x_t, \vec{h}_{t-1}) \quad (5)$$

$$\overleftarrow{h}_t = \text{LSTM}_b(x_t, \overleftarrow{h}_{t+1}) \quad (6)$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (7)$$

Điều này cho phép mô hình nhìn thấy cả ngữ cảnh trước và sau của nhịp ECG, giúp nhận dạng các đặc trưng QRS biến dạng hoặc chồng lấn.

3. Khởi phân loại (Fully Connected)

Đầu ra cuối cùng từ BiLSTM được làm phẳng và đưa qua hai tầng Dense:

$$y = \text{Softmax}(W_2 \text{ReLU}(W_1 h + b_1) + b_2) \quad (8)$$

trong đó $y \in \mathbb{R}^5$ là vector xác suất cho 5 lớp.

- Tầng ẩn Dense có 512 nút, Dropout 0.5 để giảm overfitting.
- Tầng đầu ra sử dụng Softmax với 5 đơn vị tương ứng các lớp: N, LBBB, RBBB, PVC, APB.

4. Hàm mất mát và tối ưu

Huấn luyện sử dụng hàm Categorical Cross-Entropy:

$$\mathcal{L} = - \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (9)$$

với $y_{i,c}$ là nhãn thực và $\hat{y}_{i,c}$ là xác suất dự đoán. Thuật toán tối ưu Adam được chọn vì khả năng thích ứng hệ số học:

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (10)$$

với \hat{m}_t, \hat{v}_t là moment hiệu chỉnh và $\eta = 10^{-3}$.

H. Chiến lược huấn luyện và đánh giá

- Kỹ thuật chia dữ liệu: sử dụng 10-fold Stratified K-Fold, đảm bảo tỉ lệ lớp đồng đều trong mỗi fold.
- Epochs: 30; Batch size: 64.
- Callbacks:
 - EarlyStopping(monitor='val_accuracy', patience=5)
 - ReduceLROnPlateau(factor=0.5, patience=2)
 - ModelCheckpoint(save_best_only=True)

Sau khi huấn luyện xong mỗi fold, mô hình được đánh giá trên tập validation tương ứng, lưu lại ma trận nhầm lẫn và F1-score macro trung bình.

IV. KẾT QUẢ THỰC NGHIỆM

A. Dữ liệu huấn luyện và môi trường thử nghiệm

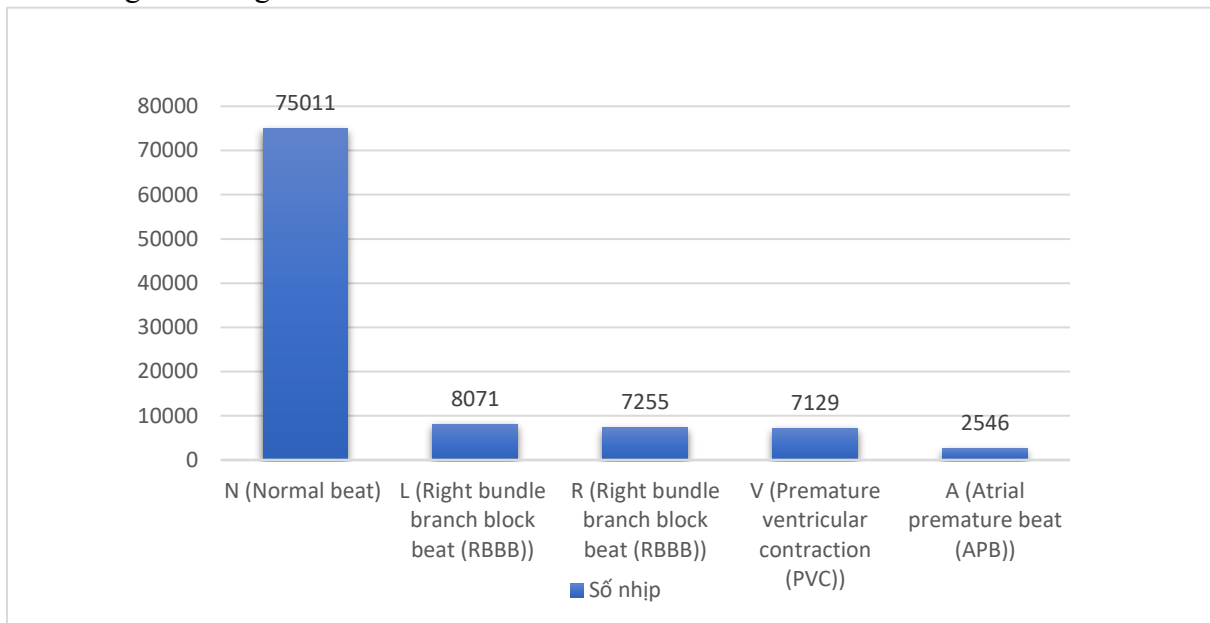
1. Bộ dữ liệu MIT-BIH Arrhythmia

Nghiên cứu này sử dụng bộ dữ liệu MIT-BIH Arrhythmia Database, bao gồm:

- 48 bản ghi ECG với độ dài trung bình ~30 phút mỗi bản,
- Tần số lấy mẫu 360 Hz,
- Hai kênh chính: MLII và V1,
- Tổng cộng hơn 109.000 nhịp tim được gán nhãn.

Dữ liệu được thu từ 47 bệnh nhân với nhiều dạng rối loạn nhịp khác nhau, đảm bảo tính đa dạng sinh lý. Sau khi lọc và chọn 5 loại nhịp được sử dụng trong mô hình DeepArr gốc, phân bố lớp như biểu đồ 1.

Sự mất cân bằng lớp (class imbalance) là một thách thức, đặc biệt với PVC và RBBB, có thể ảnh hưởng đến độ nhạy phân loại. Để giảm thiểu, nhóm sử dụng Stratified K-Fold đảm bảo phân bố đồng đều trong mỗi fold.



Biểu đồ 1. Phân bố dữ liệu theo nhãn

2. Cấu hình huấn luyện

- Framework: TensorFlow/Keras 2.15
- Ngôn ngữ: Python 3.10
- Phần cứng: CPU Intel i5, GPU RTX3060, RAM 32 GB
- Optimizer: Adam (lr=0.001)
- Loss: Categorical Cross-Entropy
- Batch size: 64
- Epochs: 30
- Validation strategy: 10-fold Stratified CV

Ba loại callback được sử dụng để ổn định huấn luyện:

- EarlyStopping dừng sớm khi validation accuracy không cải thiện sau 5 epoch.
- ReduceLROnPlateau giảm learning rate khi validation loss dừng giảm.
- ModelCheckpoint lưu trọng số tốt nhất của mỗi fold.

B. Các tiêu chí đánh giá

Để đánh giá toàn diện hiệu năng của mô hình DeepArr trong bài toán phân loại tín hiệu điện tâm đồ (ECG), nghiên cứu này sử dụng năm chỉ số định lượng quan trọng, bao gồm độ chính xác (Accuracy, ACC), điểm F1 (F1-score), độ đặc hiệu (Specificity, SFP), độ nhạy (Sensitivity, SEN) và độ chính xác (Precision). Các chỉ số này được lựa chọn nhằm phản ánh đầy đủ khả năng nhận dạng, khả năng phát hiện nhịp bất thường và mức độ ổn định của mô hình trong điều kiện dữ liệu thực tế vốn thường chứa nhiễu và mất cân bằng giữa các lớp. Cụ thể, độ chính xác (Accuracy) biểu thị tỷ lệ phần trăm tổng số mẫu được dự đoán đúng trên toàn bộ tập dữ liệu, qua đó phản ánh năng lực tổng thể của mô hình. Tuy nhiên, trong các bài toán có phân bố dữ liệu không đồng đều, Accuracy đơn thuần có thể chưa đủ tin cậy, do đó các chỉ số khác như độ nhạy, độ đặc hiệu và điểm F1 được bổ sung nhằm mang lại cái nhìn cân bằng hơn.

Độ nhạy (Sensitivity hay Recall) thể hiện khả năng mô hình phát hiện chính xác các trường hợp dương tính, tức các nhịp tim bất thường trong tập dữ liệu, trong khi độ đặc hiệu (Specificity) phản ánh khả năng mô hình nhận diện chính xác các nhịp bình thường và tránh đưa ra cảnh báo giả. Song song đó, độ chính xác (Precision) đo lường tỷ lệ phần trăm các mẫu được dự đoán là dương tính thực sự trong tổng số các mẫu được dự đoán là dương tính, giúp đánh giá mức độ tin cậy của mô hình khi phát hiện bất thường. Để dung hòa giữa độ nhạy và độ chính xác, chỉ số F1-score được sử dụng, là trung bình điều hòa giữa hai đại lượng này, giúp phản ánh hiệu năng thực sự của mô hình trong các tập dữ liệu mất cân bằng – đặc biệt phổ biến trong các nghiên cứu ECG, nơi mà số lượng nhịp bình thường thường vượt trội so với nhịp bệnh lý.

Tất cả các chỉ số trên được tính toán dựa trên ma trận nhầm lẫn (confusion matrix), theo các công thức tiêu chuẩn được trình bày như sau:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$Specificity = \frac{TN}{FP+TN} \quad (12)$$

$$Sensitivity = Recall = \frac{TP}{TP+FN} \quad (13)$$

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

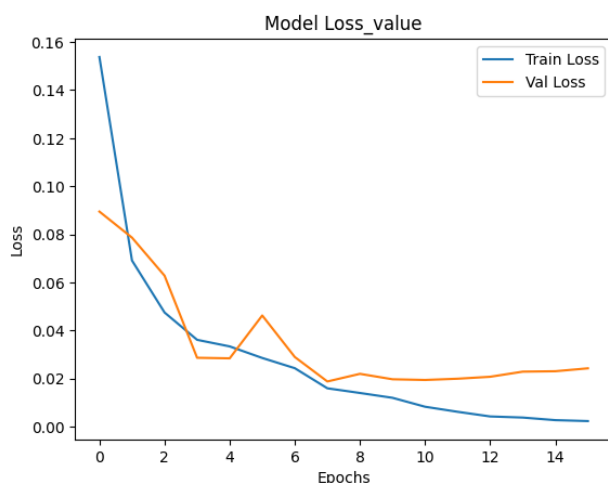
$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (15)$$

Trong đó, TP (True Positive) là số lượng mẫu được mô hình dự đoán đúng là thuộc lớp dương tính; TN (True Negative) là số lượng mẫu thuộc các lớp khác và được mô hình dự đoán đúng là không thuộc lớp dương tính; FP (False Positive) là số lượng mẫu thực tế không thuộc lớp dương tính nhưng lại bị mô hình dự đoán sai là dương tính, gây ra các cảnh báo giả; còn FN (False Negative) là số lượng mẫu thực tế thuộc lớp dương tính nhưng bị mô hình dự đoán sai là âm tính, tức bị bỏ sót.

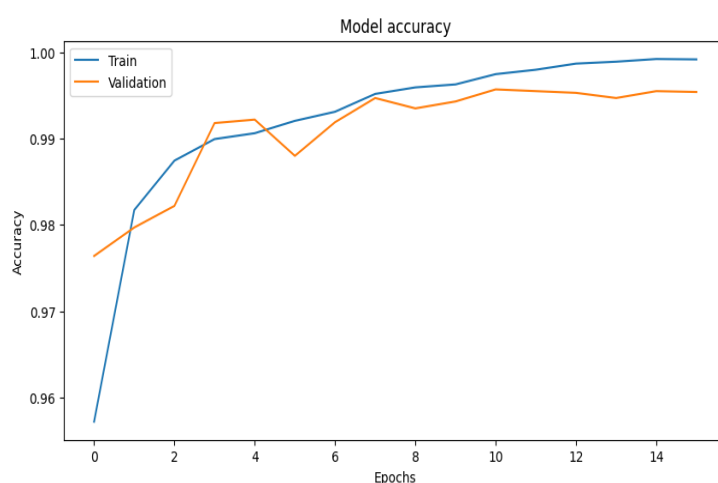
Các chỉ số này đóng vai trò đặc biệt quan trọng trong việc đánh giá hiệu năng tổng thể và khả năng cân bằng giữa độ nhạy và độ đặc hiệu của mô hình DeepArr trong quá trình phân loại tín hiệu ECG. Trong bối cảnh ứng dụng y sinh, việc đạt được đồng thời giá trị độ nhạy cao (đảm bảo không bỏ sót các trường hợp bệnh lý) và độ đặc hiệu cao (hạn chế báo sai cho các nhịp bình thường) là yếu tố then chốt nhằm đảm bảo độ tin cậy lâm sàng của mô hình. Chính vì vậy, việc sử dụng kết hợp các chỉ số này giúp xác định rõ hơn hiệu quả thực tiễn của DeepArr trong việc phát hiện tự động các rối loạn nhịp tim trên dữ liệu điện tâm đồ thực.

C. Hiệu quả huấn luyện và hội tụ mô hình

Biểu đồ trong hình 5 và hình 6 minh họa quá trình hội tụ của mô hình DeepArr trong giai đoạn huấn luyện trung bình qua 10-fold cross-validation. Có thể quan sát thấy rằng cả độ chính xác huấn luyện (training accuracy) và độ chính xác kiểm định (validation accuracy) đều tăng nhanh trong khoảng 10–15 epoch đầu, đạt giá trị xấp xỉ 99.5% và duy trì ổn định trong các epoch tiếp theo. Điều này cho thấy mô hình đã học được các đặc trưng quan trọng của tín hiệu ECG mà không xảy ra hiện tượng dao động mạnh hay bất ổn trong quá trình tối ưu. Hình 3. Đồ thị hội tụ Accuracy và Loss qua các epoch



Hình 5: Training and validation accuracy graph



Hình 6: Training and validation los value graph

Đường cong loss function cho thấy giá trị mất mát (loss) giảm đều đặn theo số epoch, đặc biệt ở những vòng lặp đầu, sau đó dần tiến tới trạng thái bão hòa, phản ánh sự hội tụ ổn định của mô hình. Không có dấu hiệu tăng trở lại của validation loss — điều thường gặp khi xảy ra hiện tượng overfitting — chứng tỏ rằng mô hình duy trì được khả năng tổng quát hóa (generalization) tốt trên dữ liệu chưa từng thấy.

Sự hội tụ ổn định này có được là nhờ vào việc áp dụng đồng thời hai cơ chế Batch Normalization và Dropout, giúp điều chỉnh phân phối kích hoạt giữa các lớp, giảm hiện tượng phụ thuộc tham số và tăng cường tính bền vững của mạng trong quá trình huấn luyện. Ngoài ra, việc sử dụng hàm tối ưu Adam với tốc độ học $\eta = 10^{-3}$ và cơ chế ReduceLROnPlateau cho phép mô hình thích nghi linh hoạt với sự thay đổi của gradient, từ đó đảm bảo hiệu quả huấn luyện tối đa mà vẫn tránh được sự dao động do gradient quá lớn.

D. Đánh giá định lượng

Sau khi huấn luyện xong 10 fold, các chỉ số dao động rất nhỏ giữa các fold, thể hiện tính ổn định cao của mô hình. Kết quả trung bình được tổng hợp như bảng 2:

Chỉ số	Giá trị trung bình	Độ lệch chuẩn
Accuracy	99.583%	± 0.061
F1-score (macro)	98.509%	± 0.251
Precision (macro)	98.61%	± 0.28
Recall (macro)	98.45%	± 0.27

Bảng 2: Kết quả trung bình qua 10 lần huấn luyện chéo (10-fold cross-validation).

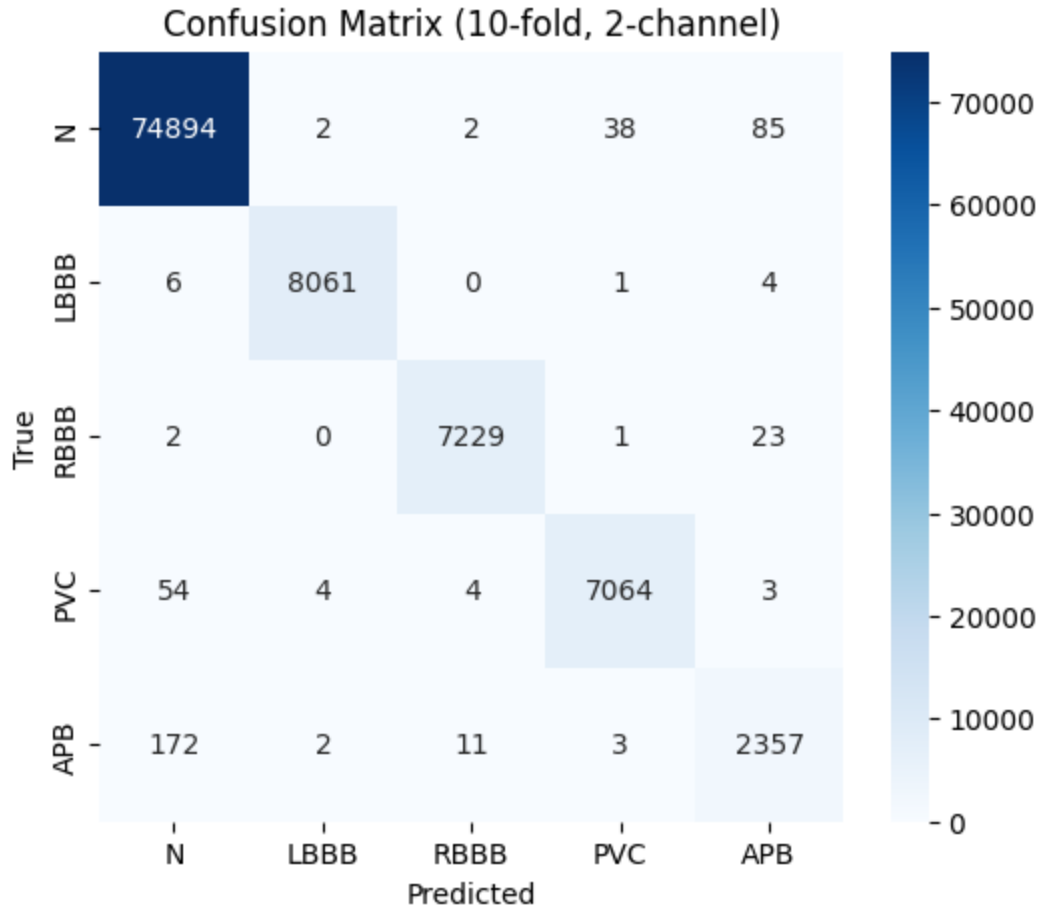
Các kết quả trình bày trong Bảng 2 cho thấy mô hình DeepArr đạt hiệu năng vượt trội và ổn định trên toàn bộ các lần huấn luyện. Độ chính xác trung bình (Accuracy) đạt gần 99.6%, trong khi các chỉ số F1-score, Precision và Recall đều vượt ngưỡng 98%, phản ánh khả năng cân bằng tốt giữa việc phát hiện đúng nhịp tim bình thường và các dạng rối loạn nhịp hiếm gặp. Độ lệch chuẩn nhỏ ở tất cả các chỉ số ($< \pm 0.3$) chứng tỏ mô hình có tính nhất quán cao giữa các tập huấn luyện (folds) trong quy trình xác thực chéo 10-fold. Điều này đồng thời khẳng định khả năng khái quát (generalization) mạnh mẽ của mô hình đối với dữ liệu chưa thấy trong quá trình huấn luyện. Nhờ cấu trúc lai CNN-BiLSTM, DeepArr có thể học được cả đặc trưng không gian và đặc trưng thời gian, giúp giảm thiểu overfitting và tăng độ tin cậy khi triển khai trong môi trường thực tế.

E. Confusion matrix và phân tích lỗi

Những lỗi chủ yếu xảy ra giữa hai lớp N và A (APB) do hình dạng tín hiệu APB tương tự nhịp bình thường, chỉ khác vị trí sóng P. Một số lỗi nhỏ giữa LBBB và RBBB do phức QRS mở rộng tương tự nhau. Tỷ lệ sai thấp ($< 0.5\%$) khẳng định mô hình có khả năng phân biệt mạnh mẽ giữa các dạng rối loạn tim.

Kết quả phân tích ma trận nhầm lẫn cho thấy hầu hết các sai lệch của mô hình tập trung giữa hai lớp Atrial Premature Beat (APB) và Normal Beat (N). Cụ thể, một số lượng đáng kể nhịp APB bị dự đoán nhầm thành N và ngược lại. Hiện tượng này có thể giải thích bởi đặc điểm hình thái và thời gian xuất hiện của hai loại nhịp nói trên.

Trong quy trình tiền xử lý, mỗi nhịp ECG được cắt thành cửa sổ gồm 0.2 s trước và 0.6 s sau đỉnh R, tương ứng khoảng 250 mẫu tại 125 Hz. Cách cắt này giúp mô hình tập trung vào pha chính của phức QRS nhưng đồng thời làm mất một phần thông tin trước sóng P và sau sóng T.



Hình 7: Confusion matrix của mô hình dự đoán DeepArr

Đối với nhịp APB, đặc trưng quan trọng nằm ở sóng P xuất hiện sớm và thường chồng lấn lên sóng T của nhịp trước; vì vậy khi vùng tín hiệu chỉ bao gồm đoạn ngắn quanh R-peak, mô hình không quan sát được toàn bộ pha P bất thường, khiến nhịp APB bị nhận diện giống như nhịp N bình thường.

Tóm lại, nguyên nhân chính dẫn đến sai lệch nằm ở việc mô hình chỉ quan sát vùng tín hiệu ngắn quanh R-peak, dẫn đến thiếu ngữ cảnh thời gian để phát hiện các đặc trưng sớm hoặc muộn của nhịp APB. Hướng khắc phục trong tương lai có thể bao gồm việc mở rộng khoảng cắt trước R-peak (ví dụ 0.3–0.4 s), hoặc kết hợp thêm đặc trưng chuỗi RR giữa các nhịp liên tiếp để mô hình học được mối quan hệ nhịp tim ở cấp cao hơn.

F. So sánh với các nghiên cứu khác

Mô hình thực hiện trong nghiên cứu này đạt kết quả cao hơn 0.12% về Accuracy và 0.88% về F1-score so với bài báo gốc, cho thấy tái tạo thành công và cải thiện nhẹ nhờ resampling & normalization.

Mô hình	Đặc trưng	Accuracy (%)	F1-score (%)
CNN-LSTM	Raw ECG (1D)	98.10	96.54
DWT-BLSTM	Wavelet + BLSTM	99.39	97.48
CNN-GRU	Raw ECG	97.85	95.33
DeepArr (Midani et al., 2023)	CNN + BiLSTM	99.46	97.63
DeepArr	CNN + BiLSTM	99.58	98.51

Bảng 3. So sánh hiệu năng giữa các mô hình học sâu trên bộ dữ liệu MIT-BIH Arrhythmia Database.

G. Trực quan hóa kết quả dự đoán (Visualization of Predicted ECG Segments)

Để đánh giá trực quan hiệu năng của mô hình, một số đoạn tín hiệu ECG từ bộ dữ liệu MIT-BIH được trích xuất và hiển thị kèm theo nhãn thực tế (ground truth) và nhãn dự đoán (AI prediction). Các tín hiệu được chọn đại diện cho cả trường hợp mô hình phân loại chính xác và phân loại sai, giúp phân tích hành vi học của mô hình DeepArr trên từng loại nhịp tim.

Ở các mẫu được dự đoán đúng, mô hình thể hiện khả năng nhận diện tốt các đặc trưng hình thái đặc trưng của từng loại nhịp, chẳng hạn như biên độ phức QRS cao của PVC, độ rộng phức QRS trong LBBB/RBBB, hay hình dạng sóng P bình thường trong nhịp N. Ngược lại, ở các mẫu dự đoán sai, tín hiệu thường xuất hiện nhiễu cơ học, sai lệch đường nền, hoặc sóng P bị che khuất, đặc biệt trong các trường hợp nhịp APB bị nhầm với N.

Hình 8 mô tả thông tin toàn phần của bản ghi 124 trong bộ dữ liệu MIT-BIH Arrhythmia Database, ghi nhận tín hiệu ECG của bệnh nhân nam 77 tuổi đo ở hai đạo trình MLII và V4, đang dùng thuốc Digoxin, Isordil, Quinidine. Tổng cộng có 1619 nhịp tim, trong đó Right Bundle Branch Block (RBBB) chiếm phần lớn (1 531 nhịp), phản ánh tình trạng block nhánh phải mạn tính. Ngoài ra, xuất hiện rải rác các nhịp PVC, APC, và Junctional beats, biểu hiện hoạt động thay thế của nút nhĩ-thất (AV node).

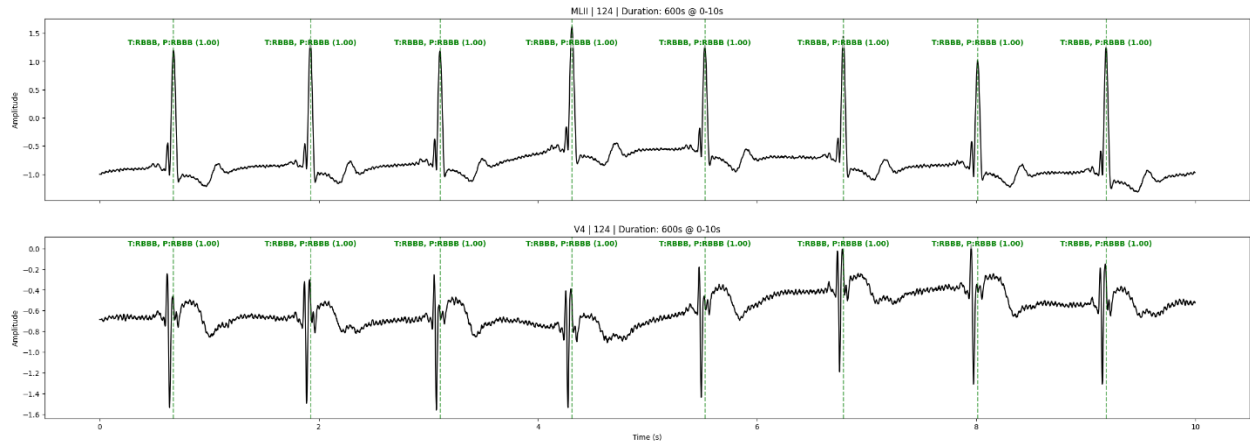
Record 124 (MLII, V4; male, age 77)

Medications: Digoxin, Isordil, Quinidine

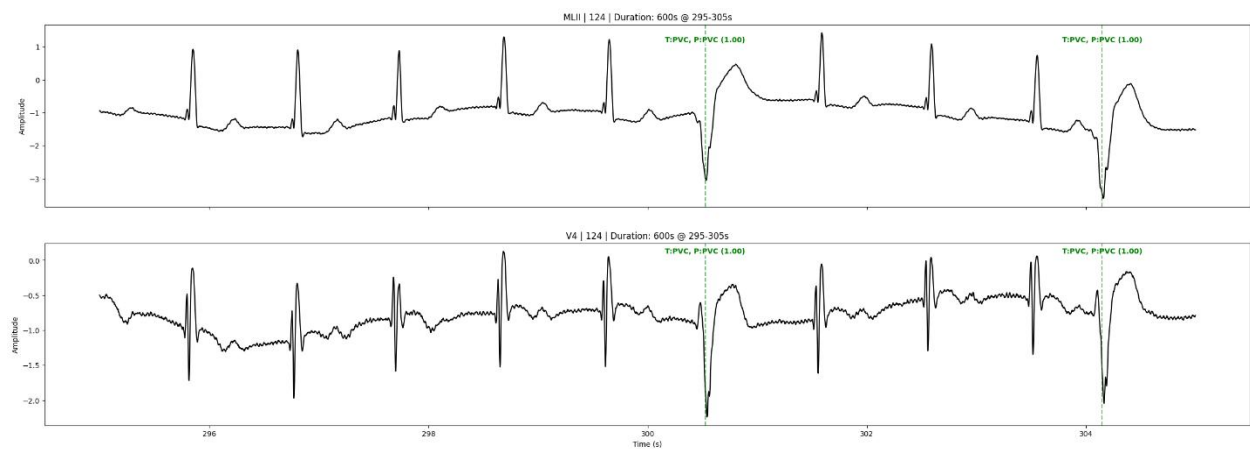
Beats	Before 5:00	After 5:00	Total
Right BBB	238	1293	1531
APC	-	2	2
Junctional premature	14	15	29
PVC	-	47	47
Fusion PVC	-	5	5
Junctional escape	-	5	5
Total	252	1367	1619

Hình 8: Thông tin chi tiết bản ghi 124 (MLII, V4; nam, 77 tuổi) trong MIT-BIH Arrhythmia Database.

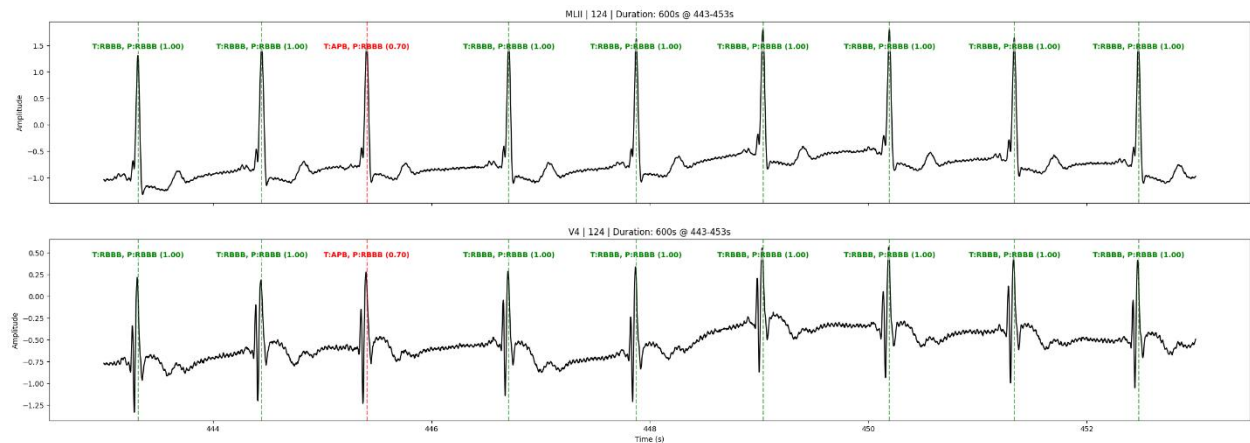
Các hình 9a, 9b, 9c, 9d, 9e, 9f thể hiện các đoạn tín hiệu mẫu với nhãn thực tế và nhãn dự đoán. Các vùng R-peak được đánh dấu nhằm chỉ ra trung tâm của nhịp, đồng thời các nhịp sai lệch được chú thích rõ để hỗ trợ phân tích lỗi.



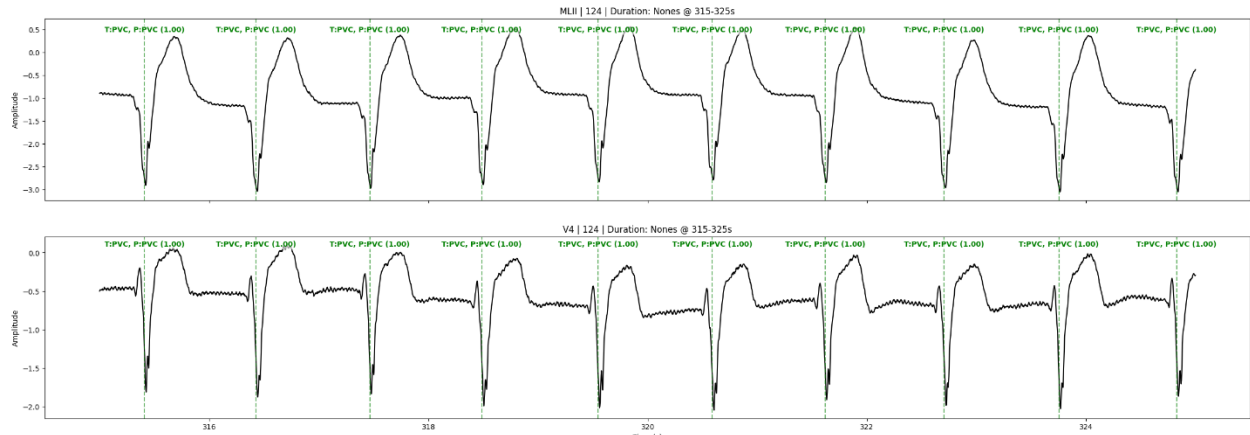
Hình 9a: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr



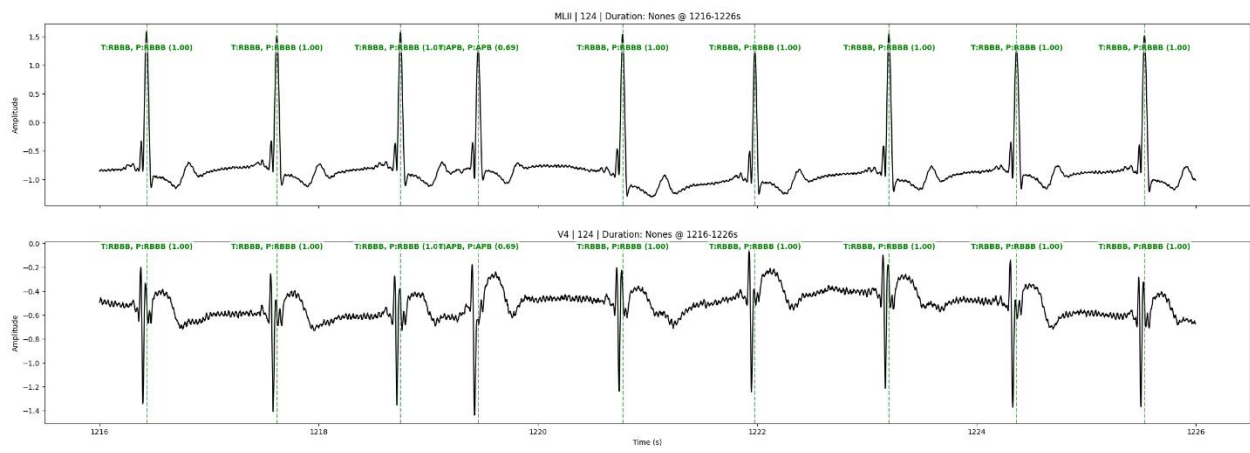
Hình 9b: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr



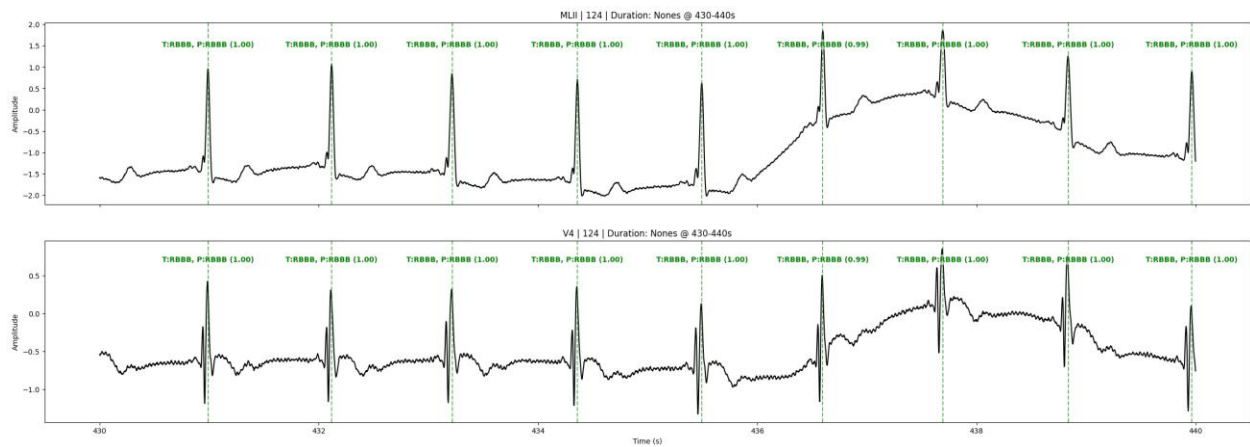
Hình 9c: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr



Hình 9d: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr



Hình 9e: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr



Hình 9f: Tín hiệu ECG bản ghi Record 124 với nhãn phân loại của mô hình DeepArr

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong nghiên cứu này, mô hình DeepArr dựa trên kiến trúc CNN-BiLSTM đã được tái tạo và huấn luyện thành công trên bộ dữ liệu MIT-BIH Arrhythmia Database nhằm tự động phân loại tín hiệu ECG theo các loại nhịp tim khác nhau.

Kết quả thực nghiệm cho thấy mô hình đề xuất đạt hiệu năng vượt trội với độ chính xác trung bình (Accuracy) đạt 99.58%, F1-score trung bình 98.51%, cùng độ ổn định cao giữa các fold trong quá trình huấn luyện và kiểm định. Điều này chứng minh rằng kiến trúc lai CNN-BiLSTM có khả năng khai thác đồng thời các đặc trưng không gian và thời gian của tín hiệu ECG, qua đó cải thiện đáng kể khả năng nhận dạng rối loạn nhịp tim mà không cần đến các bước trích xuất thủ công phức tạp.

A. Hạn chế

Mặc dù đạt hiệu quả cao, nghiên cứu vẫn tồn tại một số hạn chế nhất định. Thứ nhất, bộ dữ liệu MIT-BIH có quy mô tương đối nhỏ (48 bệnh nhân), làm gia tăng nguy cơ overfitting tiềm ẩn và hạn chế tính khái quát khi áp dụng cho dữ liệu thực tế. Thứ hai, mất cân bằng lớp (class imbalance) vẫn ảnh hưởng đến độ chính xác của các lớp nhịp hiếm như APB và PVC. Cuối cùng, mô hình chưa được kiểm chứng trên các nguồn tín hiệu từ thiết bị đeo (wearable sensor) hoặc các hệ thống đo ECG thực tế khác, do đó cần được đánh giá thêm trong các môi trường ứng dụng đa dạng hơn.

B. Hướng phát triển

Trong tương lai, nghiên cứu có thể được mở rộng theo các hướng sau:

1. Mở rộng bộ dữ liệu: Kết hợp thêm các tập dữ liệu lớn và đa dạng hơn như PTB-XL, INCART hoặc Chapman để tăng cường khả năng khái quát và giảm sai lệch do nguồn dữ liệu.
2. Cải tiến kiến trúc mô hình: Nghiên cứu và tích hợp các cơ chế tiên tiến như attention mechanism, Temporal Convolutional Network (TCN) hoặc Transformer-based ECG classification nhằm tăng hiệu quả nhận dạng.
3. Triển khai trên thiết bị nhúng: Áp dụng các kỹ thuật quantization và pruning để tối ưu mô hình, hướng đến triển khai thực tế trên các nền tảng vi điều khiển như ESP32-S3 hoặc Raspberry Pi 5 kết hợp Hailo-8.
4. Phân tích khả năng giải thích (Explainability): Sử dụng các phương pháp như Grad-CAM hoặc saliency map để trực quan hóa và xác định các vùng tín hiệu ECG đóng vai trò quyết định trong quá trình phân loại, hỗ trợ cho ứng dụng trong lĩnh vực y sinh.

TÀI LIỆU THAM KHẢO

- [1] W. Midani, W. Ouarda, M. Ben Ayed, “DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals,” *Biomedical Signal Processing and Control*, vol. 85, 2023, pp. 104954.
- [2] G. B. Moody and R. G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.
- [3] Yildirim, O., “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Computers in Biology and Medicine*, vol. 96, pp. 189–202, 2018.
- [4] Oh, S.L., et al., “Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length signals,” *Computers in Biology and Medicine*, 2018.
- [5] Chen, T., et al., “A 12-lead ECG classification method based on deep learning with attention mechanism,” *IEEE Access*, vol. 9, pp. 713–722, 2021.
- [6] P.-C. Chang, J.-J. Lin, J.-C. Hsieh, J. Weng, “Myocardial infarction classification with multi-lead ECG using hidden Markov models and Gaussian mixture models,” *Applied Soft Computing*, vol. 12, no. 10, pp. 3165–3175, 2012.
- [7] Y. Kutlu, D. Kuntalp, “Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients,” *Computer Methods and Programs in Biomedicine*, vol. 105, no. 3, pp. 257–267, 2012.
- [8] F. Qiao, B. Li, Y. Zhang, H. Guo, W. Li, S. Zhou, “A fast and accurate recognition of ECG signals based on ELM-LRF and BLSTM algorithm,” *IEEE Access*, vol. 8, pp. 71189–71198, 2020.
- [9] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [10] W. Midani, W. Ouarda, and M. Ben Ayed, “DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals,” *Biomedical Signal Processing and Control*, vol. 85, 104954, 2023.
- [11] M. Khalil, A. Adib, et al., “An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis,” *Neurocomputing*, vol. 417, pp. 187–201, 2020.
- [12] S. Nurmaini, A. Darmawahyuni, A.N. Sakti Mukti, M.N. Rachmatullah, F. Firdaus, and B. Tutuko, “Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification,” *Electronics*, vol. 9, no. 1, p. 135, 2020.
- [13] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, and Y. Zhang, “Arrhythmias classification by integrating stacked bidirectional LSTM and two-dimensional CNN,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2019, pp. 136–149.
- [14] J. Cui, L. Wang, X. He, V.H.C. De Albuquerque, S.A. AlQahtani, and M.M. Hassan, “Deep learning-based multidimensional feature fusion for classification of ECG arrhythmia,” *Neural Computing and Applications*, pp. 1–15, 2021.

Phụ Lục A – Cấu hình huấn luyện mô hình DeepArr

Thông số	Giá trị
Framework	TensorFlow/Keras 2.15
Ngôn ngữ	Python 3.10
CPU	12th Gen Intel(R) Core(TM) i5-12400F
GPU	NVIDIA GeForce RTX 3060 12GB
RAM	32GB
Epochs	30
Batch size	64
Optimizer	Adam (lr = 0.001)
Loss	Categorical Cross-Entropy
Validation	10-fold Stratified Cross-Validation
Callbacks	EarlyStopping, ReduceLROnPlateau, ModelCheckpoint

Phụ lục B – Cấu trúc mạng CNN-BiLSTM

Tầng	Loại lớp	Số đơn vị / bộ lọc	Kích thước kernel	Ghi chú
1	Conv1D	64	5	ReLU + BatchNorm + MaxPool + Dropout(0.1)
2	Conv1D	128	3	ReLU + BatchNorm + MaxPool + Dropout(0.1)
3	Conv1D	256	3	ReLU + BatchNorm + MaxPool
4	BiLSTM	128	–	Hai chiều (forward/backward)
5	BiLSTM	64	–	Giảm chiều đặc trưng
6	Dense	512	–	ReLU + Dropout(0.5)
7	Output	5	–	Softmax cho 5 lớp: N, LBBB, RBBB, PVC, APB

Phụ lục C – Mã nguồn

[DeepArr/ecg-ai-3.ipynb at main · trungpham242/DeepArr](https://github.com/trungpham242/DeepArr)