

**ĐẠI HỌC ĐÀ NẴNG**  
**TRƯỜNG ĐẠI HỌC BÁCH KHOA**  
**KHOA ĐIỆN TỬ - VIỄN THÔNG**

**BÁO CÁO**  
**TIẾN ĐỘ PBL4**

**ĐỀ TÀI:**

**IMLE-NET: AN INTERPRETABLE MULTI-LEVEL  
MULTI-CHANNEL MODEL FOR ECG  
CLASSIFICATION**

Sinh viên thực hiện: **TRẦN DUY NGUYÊN**

Mã số sinh viên: **106220264**

Lớp: **22KTMT2**

**Đà Nẵng, 01/2026**

## I. TÌM HIỂU CHI TIẾT BỘ DỮ LIỆU PTB-XL

Bộ dữ liệu PTB-XL là cơ sở dữ liệu điện tâm đồ lâm sàng quy mô lớn, cung cấp nền tảng vững chắc cho việc huấn luyện mô hình IMLE-Net.

- Quy mô dữ liệu: Bao gồm 21.837 bản ghi ECG từ 18.885 bệnh nhân, với sự cân bằng về giới tính (52% nam, 48% nữ) và độ tuổi rộng từ 0 đến 95 tuổi.
- Cấu trúc tín hiệu: Mỗi bản ghi kéo dài 10 giây, thu thập đầy đủ từ 12 đạo trình tiêu chuẩn (I, II, III, aVL, aVR, aVF, V1-V6).
- Phân loại chẩn đoán: Các bản ghi được dán nhãn theo 5 nhóm (superclasses) chính:
  - NORM: Điện tâm đồ bình thường.
  - CD: Rối loạn dẫn truyền.
  - MI: Nhồi máu cơ tim.
  - HYP: Phì đại.
  - STTC: Thay đổi đoạn ST/T.

- Tính đa dạng: Ngoài các nhóm ở trên, dữ liệu còn hỗ trợ phân tích các nhánh con (subclasses) như Anteroseptal MI (ASMI) và Inferior MI (IMI) để đánh giá khả năng chẩn đoán chi tiết của mô hình.

## II. PHƯƠNG PHÁP XỬ LÝ DỮ LIỆU

### 1. Tiền xử lý tín hiệu và Hạ tần số lấy mẫu (Downsampling)

- Hạ tần số: Tín hiệu gốc từ bộ dữ liệu PTB-XL được ghi lại ở tần số 400 Hz, nhưng để tối ưu hóa hiệu năng tính toán mà vẫn bảo toàn các đặc trưng hình thái (morphology), dữ liệu được hạ xuống còn 100 Hz.
- Định dạng dữ liệu: Sau khi hạ tần số, mỗi bản ghi 10 giây tương đương với 1000 điểm dữ liệu thời gian trên mỗi đạo trình.
- Trích xuất: Các bản ghi được chuyển đổi từ định dạng file sóng y tế sang các mảng dữ liệu số (NumPy arrays) để chuẩn bị cho các bước tính toán tiếp theo.

### 2. Phân đoạn nhịp tim bằng Cửa sổ trượt (Sliding Window Segmentation)

Thay vì dựa vào việc phát hiện đỉnh R (R-peak detection) vốn dễ sai sót khi gấp tín hiệu nhiều hoặc biến thiên nhịp tim (HRV) bất thường, nhóm lựa chọn phương pháp sliding window with no overlap.

- Cơ chế: Để có được phân đoạn nhịp thứ  $k$ , cửa sổ sẽ trải dài từ  $(k - 1) \times W$  đến  $k \times W$  trên toàn bộ tín hiệu.

- Thông số: Với tổng chiều dài  $T = 1000$  và kích thước cửa sổ  $W = 50$ , mỗi đoạn được chia thành  $N = 20$  phân đoạn nhịp tim cố định.

$$N = \frac{T}{W}$$

### 3. Chuẩn hóa và Cấu trúc hóa đầu vào

- Chuẩn hóa toàn cục (Global Standardization): Thay vì chuẩn hóa từng kênh đơn lẻ, mô hình thực hiện chuẩn hóa dựa trên giá trị trung bình ( $\mu$ ) và độ lệch chuẩn ( $\sigma$ ) tính toán trên toàn bộ dữ liệu huấn luyện đã được làm phẳng

$$z = \frac{x - \mu}{\sigma}$$

Việc này giúp duy trì sự khác biệt về biên độ thực tế giữa các đoạn, tạo điều kiện cho cơ chế chú ý nhận diện đúng tầm quan trọng của từng kênh.

- Tensor đầu vào: Dữ liệu được cấu trúc lại dưới dạng tensor 3 chiều kích thước  $12 \times 1000 \times 1$ . Các trọng số của lớp CNN đầu tiên sẽ được chia sẻ (shared weights) trên tất cả các phân đoạn và các kênh để giảm tham số và tăng tính tổng quát hóa.

### 4. Ánh xạ nhãn và Chiến lược phân tách

- Gom nhóm nhãn (Label Aggregation): Các mã bệnh lý SCP được ánh xạ về 5 siêu lớp chính phục vụ bài toán đa nhãn. Các nhãn con như ASMI và IMI được lọc riêng để kiểm chứng tính giải thích của mô hình.
- Phân tách Stratified: Sử dụng 10 nếp gấp (folds). Folds 1-8 dành cho huấn luyện, Fold 9 để kiểm định (validation) và Fold 10 dùng cho kiểm thử độc lập (test).
- Hàm mất mát (Loss Function): Sử dụng Binary Cross-entropy để xử lý bài toán phân loại đa nhãn:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

### 5. Cơ chế Chú ý cấp độ đạo trình (Channel-level Attention)

Do các đặc trưng bệnh lý thường chỉ xuất hiện trên một số đạo trình nhất định, mô hình sử dụng tầng chú ý để ưu tiên thông tin từ các kênh quan trọng.

- Trọng số chú ý: Mô hình học tham số để gán trọng số  $\gamma$  cho từng kênh thông qua hàm Softmax:

$$\gamma = \text{softmax}(V_c \tanh(W_c R^c + b_c))$$

- Vector ngũ cành: Vector  $C$  đại diện cho toàn bộ tín hiệu 12 kênh được tổng hợp bằng tổng có trọng số:

$$C = \sum_{i=1}^M \gamma_i R_i^c$$

Điều này cho phép mô hình tự động "tập trung" vào các đạo trình mang dấu hiệu chẩn đoán quan trọng nhất (như II, III, aVF đối với bệnh IMI).