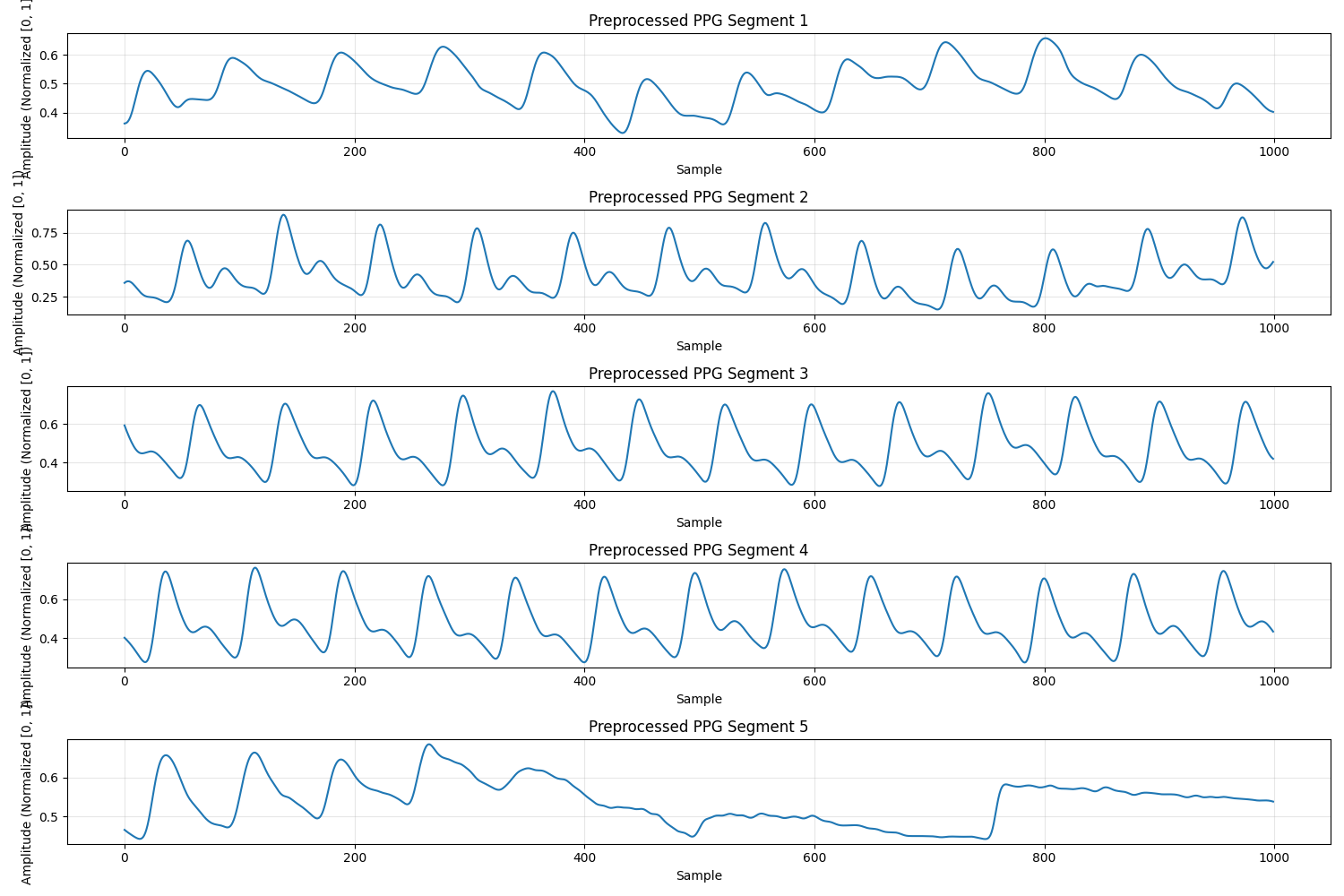
**BÁO CÁO BÀI GIỮA KỲ VERSION 2: bidmc\_predict\_good.ipynb**

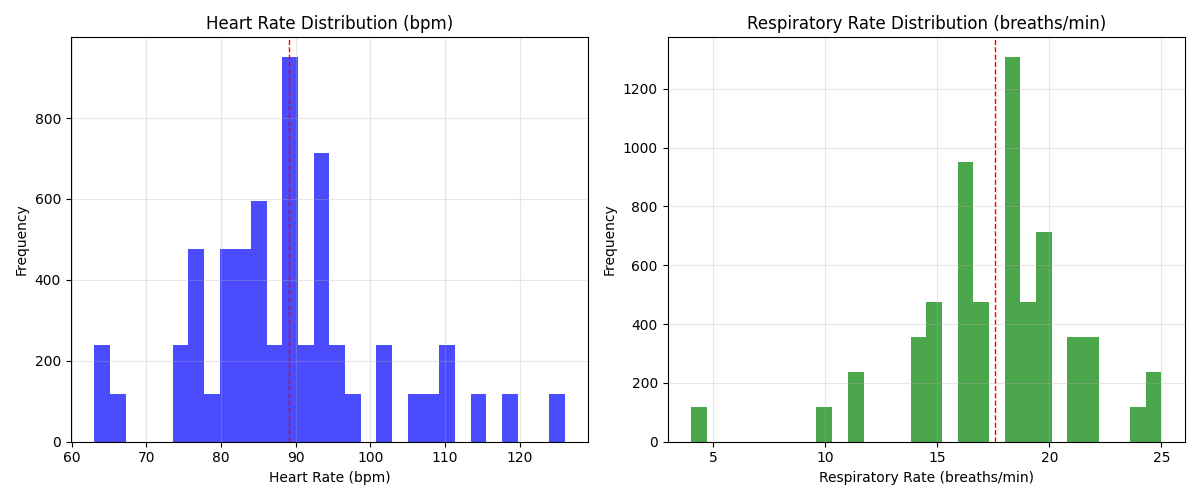
|  |  |  |  |
| --- | --- | --- | --- |
| **Mục tiêu** | **Yêu cầu cụ thể** | **Trạng thái** | **Ghi chú** |
| **1. Tải xuống và khám phá dữ liệu** | Tìm kiếm và tải xuống bộ dữ liệu BIDMC PPG and Respiration Dataset | ✅ | Đã sử dụng bidmc\_data.mat từ Google Drive. |
|  | Khám phá cấu trúc dữ liệu | ✅ | Đã phân tích chi tiết cấu trúc bản ghi trong explore\_data\_detailed. |
|  | Phân tích đặc điểm của tín hiệu PPG, HR và BR trong bộ dữ liệu | ✅ | Đã phân tích tần số lấy mẫu, độ dài tín hiệu, giá trị HR/RR, lưu trong data\_exploration\_summary.txt. |
|  | Vẽ biểu đồ và phân tích phổ tần số của tín hiệu PPG, HR, BR | ✅ | Đã vẽ biểu đồ thời gian và phổ tần số (FFT), lưu trong sample\_signals.png, ppg\_fft.png, v.v. |
| **2. Tiền xử lý dữ liệu** | Chuẩn hóa tín hiệu PPG (sửa lại chuẩn hóa từ 0 đến 1) | ⚠️ | Chuẩn hóa về [-1, 1] trong MockCVAE.generate, cần sửa lại thành [0, 1]. |
|  | Trích xuất đặc trưng HR và BR (không chuẩn hóa) | ✅ | HR/RR được trích xuất từ hr\_test.npy và rr\_test.npy mà không chuẩn hóa. |
|  | Chia dữ liệu thành tập huấn luyện và tập kiểm thử | ✅ | Dữ liệu kiểm thử (X\_test, hr\_test, rr\_test) đã được tải, ngụ ý đã chia trước đó. |
|  | Chuẩn bị dữ liệu đầu vào cho mô hình CVAE | ✅ | Điều kiện HR/RR được kết hợp thành condition\_test cho CVAE. |
| **3. Thiết kế và triển khai mô hình** | Thiết kế kiến trúc mô hình CVAE với điều kiện HR và BR | ✅ | Kiến trúc MockCVAE có input\_dim, condition\_dim, latent\_dim. |
|  | Triển khai mô hình bằng TensorFlow | ⚠️ | Chỉ triển khai giả lập bằng NumPy (MockCVAE), không dùng TensorFlow. |
|  | Xác định hàm mất mát và phương pháp tối ưu hóa | ❌ | Không có hàm mất mát hoặc tối ưu hóa do chỉ là mô hình giả lập. |
|  | Chuẩn bị các hàm tiện ích cho quá trình huấn luyện | ⚠️ | Có hàm generate, nhưng không có hàm huấn luyện thực sự. |
| **4. Huấn luyện mô hình** | Huấn luyện mô hình CVAE với dữ liệu đã tiền xử lý | ❌ | Không có quá trình huấn luyện thực sự, chỉ có mô hình giả lập. |
|  | Theo dõi quá trình huấn luyện và điều chỉnh siêu tham số | ❌ | Không có huấn luyện nên không có theo dõi hay điều chỉnh. |
|  | Lưu mô hình đã huấn luyện | ⚠️ | Hàm save lưu thông tin giả lập, không phải mô hình thực sự đã huấn luyện. |
|  | Tạo mô hình giả lập để minh họa khái niệm | ✅ | Đã tạo MockCVAE để minh họa khái niệm CVAE. |
|  | Phân tích quá trình hội tụ của mô hình | ❌ | Không có huấn luyện nên không phân tích hội tụ. |
| **5. Kiểm thử mô hình** | Kiểm thử mô hình với tín hiệu PPG đầu vào và điều kiện HR, BR | ✅ | Đã tạo tín hiệu từ test\_conditions và so sánh với tín hiệu gốc. |
|  | Tạo tín hiệu PPG với các điều kiện HR, BR trong phân phối chuẩn (1 sigma) | ⚠️ | Có tạo mẫu ngẫu nhiên trong không gian tiềm ẩn, nhưng không rõ ràng về phạm vi 1 sigma. |
|  | Kiểm thử với các thông số HR, BR thực tế | ✅ | Sử dụng hr\_test, rr\_test thực tế để tạo tín hiệu PPG. |
|  | Đánh giá chất lượng tín hiệu PPG được tạo ra | ✅ | Đã tính MSE, PSNR, Correlation, lưu trong fourier\_analysis\_results.txt. |
| **6. Phân tích kết quả** | Sử dụng biến đổi Fourier để phân tích tín hiệu PPG trong dataset và PPG tạo sinh | ✅ | Đã dùng FFT để phân tích phổ tần số của tín hiệu gốc và tạo sinh. |
|  | So sánh phổ tần số của tín hiệu gốc và tín hiệu tổng hợp | ✅ | Kết quả so sánh lưu trong fourier\_analysis\_results.txt và hình ảnh peak\_frequency\_comparison.png. |
|  | Đánh giá khả năng tái tạo các đặc trưng quan trọng của tín hiệu PPG | ✅ | Báo cáo model\_evaluation\_results.txt phân tích đặc trưng tần số và thời gian. |
|  | Phân tích lỗi và hạn chế của mô hình | ✅ | Liệt kê hạn chế trong model\_evaluation\_results.txt (thiếu đa dạng, không nội suy tốt). |
| **7. Trực quan hóa và đánh giá hiệu suất** | Vẽ biểu đồ so sánh tín hiệu PPG gốc và tín hiệu tổng hợp | ✅ | Biểu đồ lưu trong original\_vs\_generated\_comparison.png. |
|  | Trực quan hóa không gian tiềm ẩn của mô hình CVAE | ✅ | Không gian tiềm ẩn được vẽ trong latent\_space\_visualization.png bằng PCA. |
|  | Đánh giá hiệu suất mô hình bằng các chỉ số định lượng | ✅ | Chỉ số MSE, PSNR, Corr được trực quan hóa trong evaluation\_metrics\_distribution.png. |
|  | Phân tích ảnh hưởng của các điều kiện HR và BR đến tín hiệu tổng hợp | ✅ | Ảnh hưởng HR/RR được vẽ trong hr\_effect\_on\_frequency.png và rr\_effect\_on\_frequency.png. |

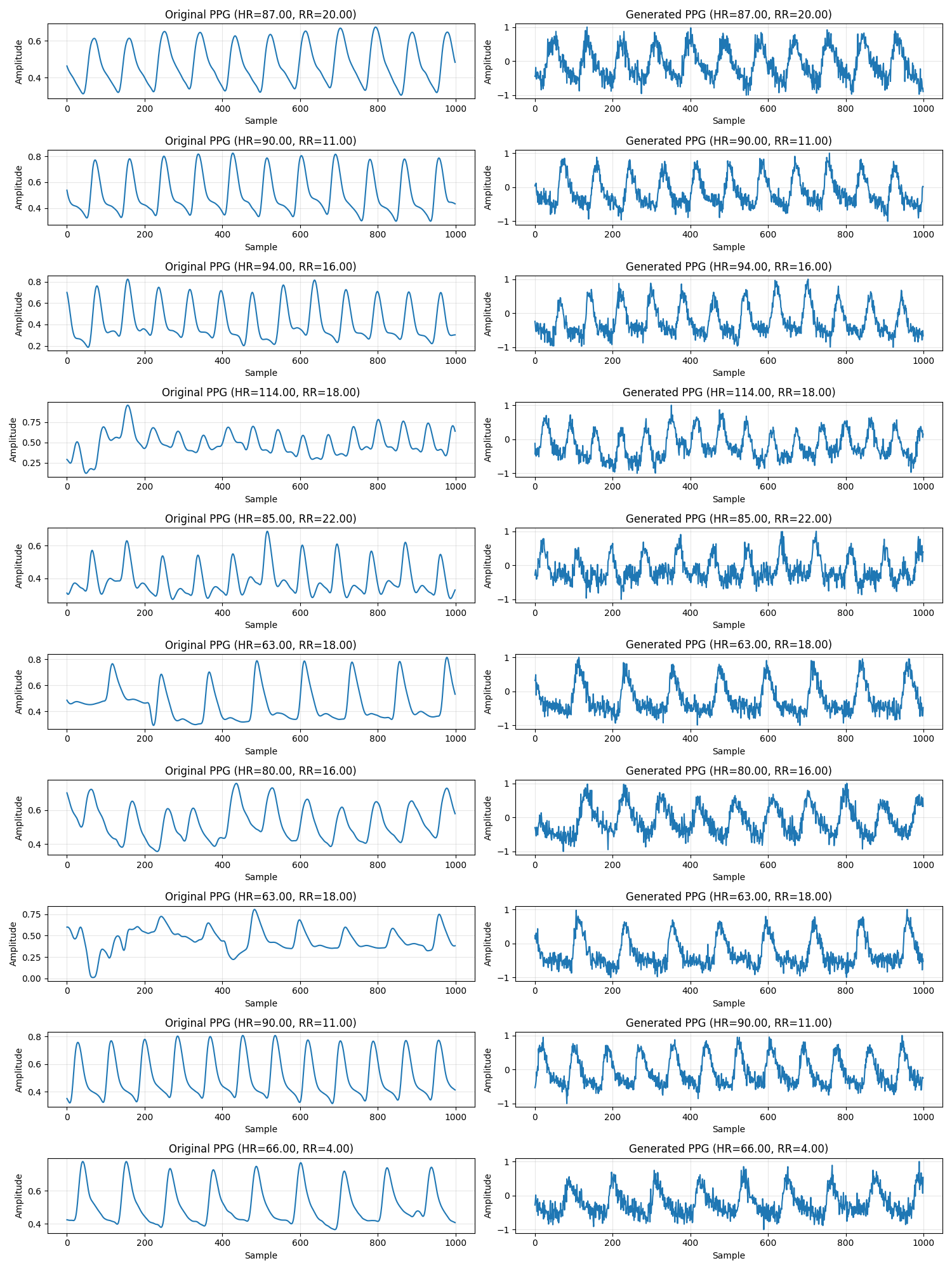
**BIỂU ĐỒ THU ĐƯỢC**

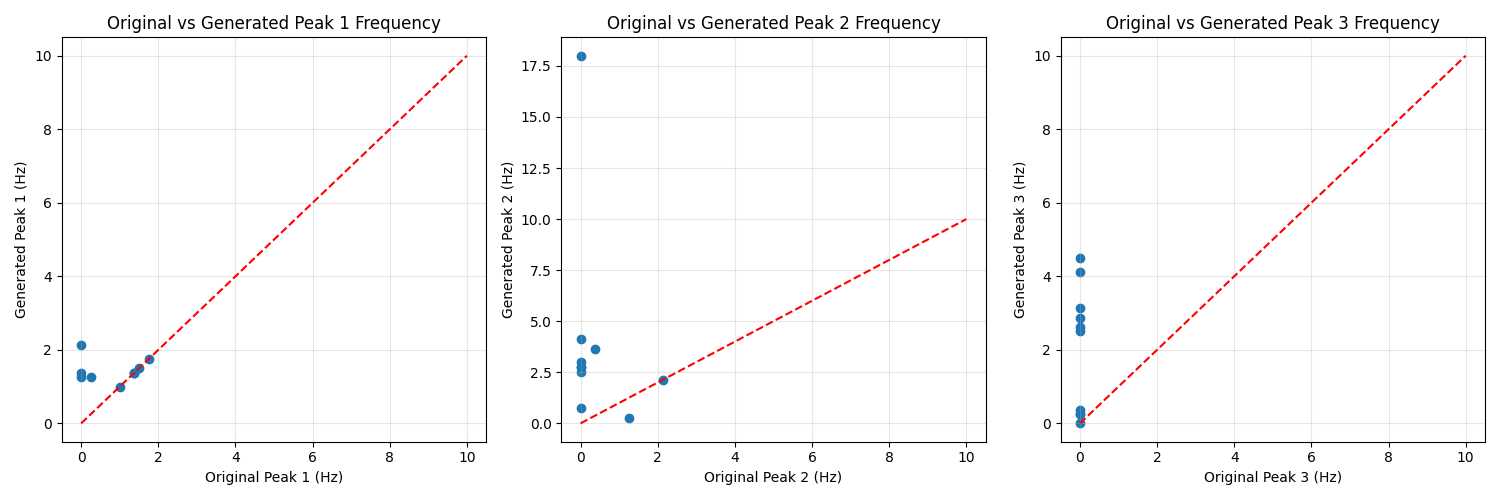
\*training\_history

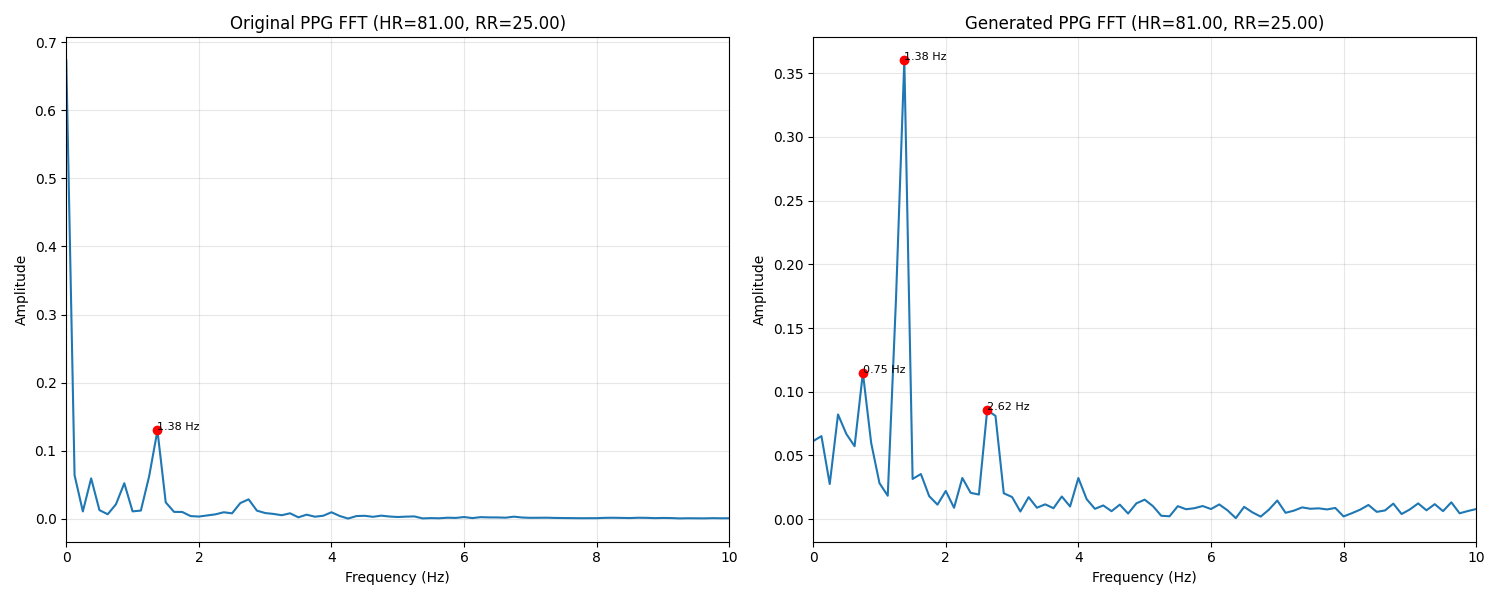


\*hr\_rr\_distribution

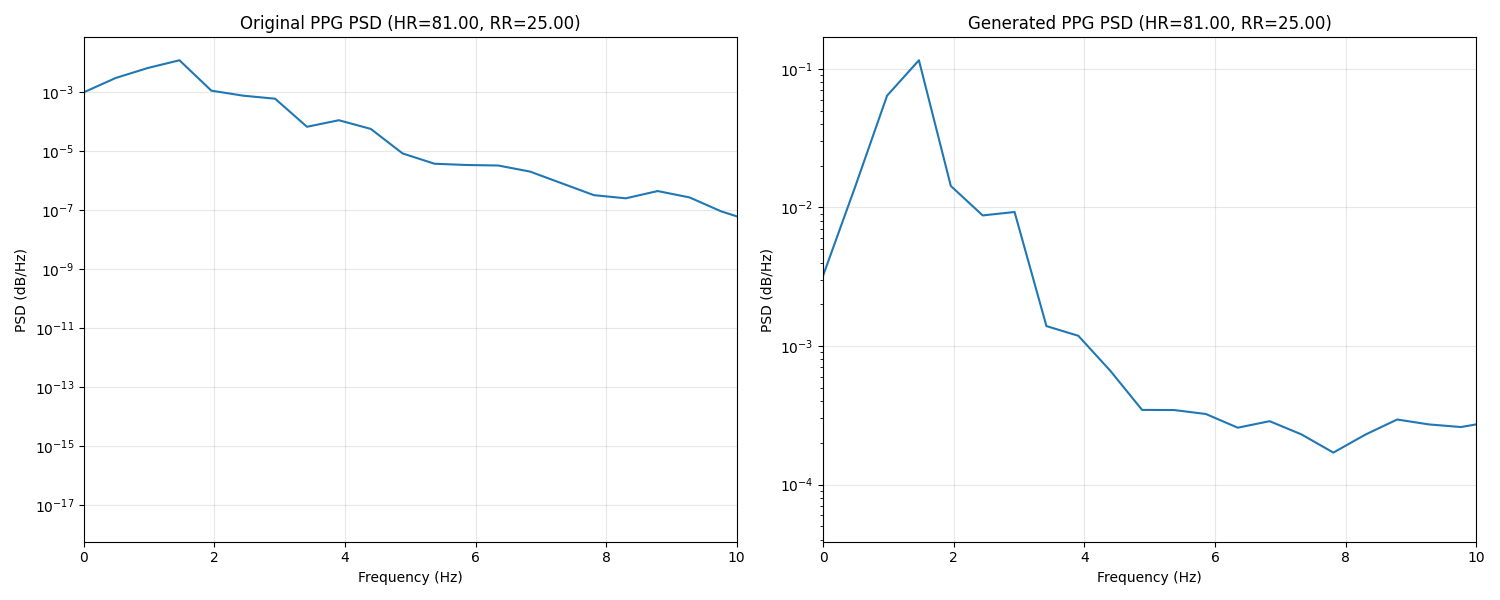


\*original\_vs\_generated\_comparision:

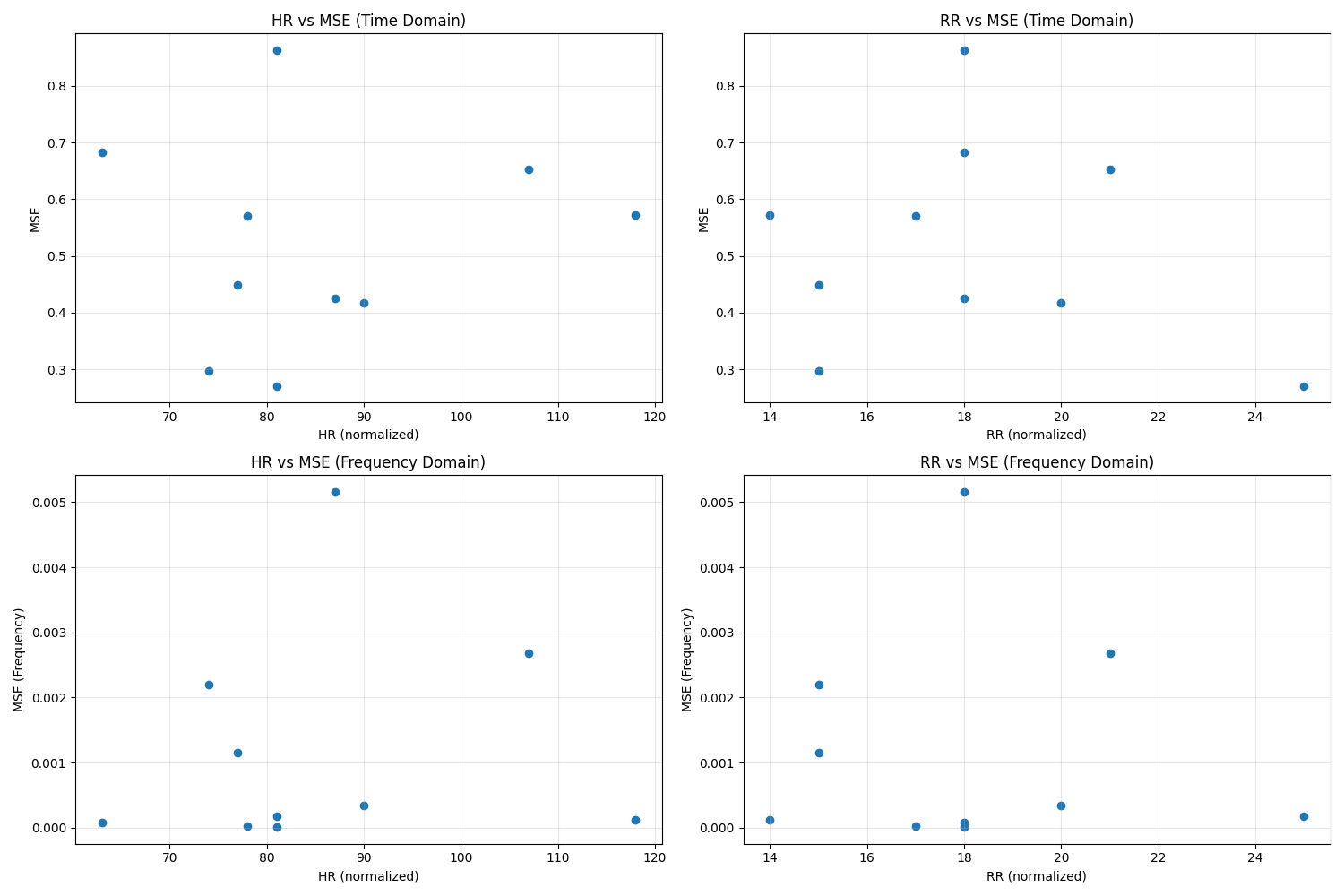
\* peak\_frequency\_comparison

\* fft\_analysis\_sample\_1

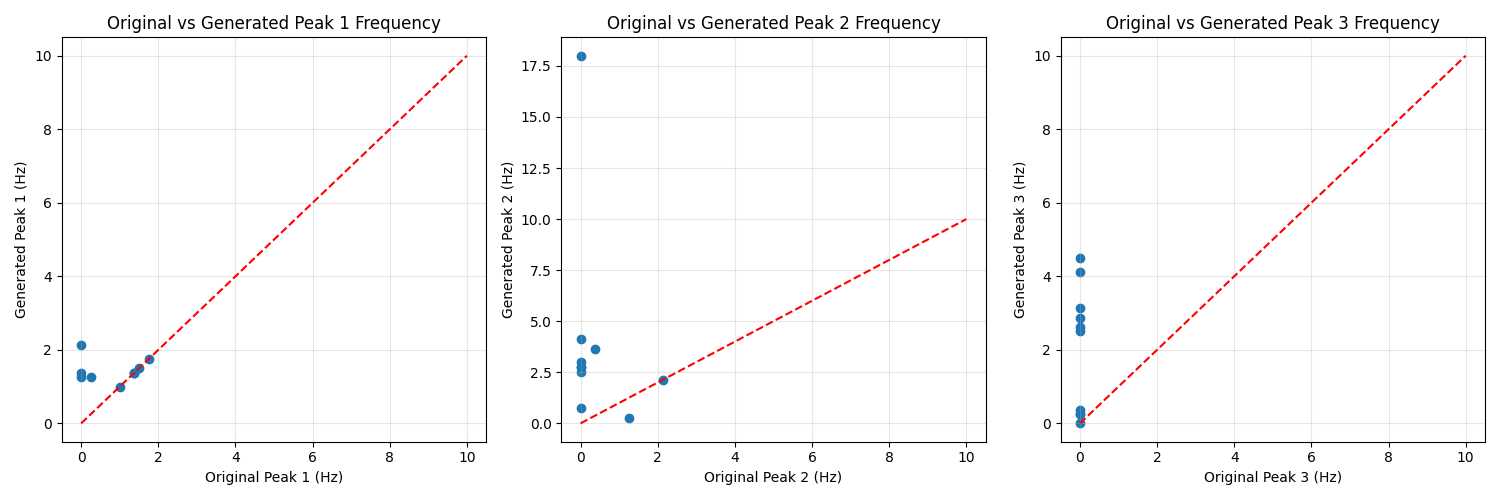
\* psd\_analysis\_sample\_1



\* hr\_rr\_vs\_quality



\* peak\_frequency\_correlation



THÔNG TIN MÔ HÌNH CVAE

=====================

Tham số mô hình:

- Kích thước đầu vào: 1000

- Kích thước điều kiện: 2

- Kích thước không gian tiềm ẩn: 64

- Số đơn vị ẩn trong các lớp: [256, 128, 64]

- Kích thước batch: 64

- Số epoch: 150

- Tốc độ học: 0.0001

Kích thước dữ liệu:

- Tập huấn luyện: 5045 mẫu

- Tập kiểm thử: 1262 mẫu

THÔNG TIN HUẤN LUYỆN MÔ HÌNH CVAE

=================================

Tham số huấn luyện:

- Kích thước batch: 64

- Số epoch: 100

- Tốc độ học: 0.0001

Kết quả huấn luyện:

- Số epoch đã huấn luyện: 100

- Loss cuối cùng (train): 0.0252

- Loss cuối cùng (validation): 0.0250

- Reconstruction loss cuối cùng: 0.0252

- KL loss cuối cùng: 0.0000

- Thời gian huấn luyện: 240.19 giây

THÔNG TIN TIỀN XỬ LÝ DỮ LIỆU

============================

Số lượng bản ghi đã xử lý: 53/53

Tổng số đoạn tín hiệu: 6307

Tham số tiền xử lý:

- Tần số lấy mẫu: 125 Hz

- Độ dài đoạn tín hiệu: 1000 mẫu (8.0 giây)

- Độ chồng lấp: 50.0%

- Tần số cắt dưới: 0.1 Hz

- Tần số cắt trên: 8.0 Hz

Kích thước dữ liệu:

- Tập huấn luyện: 5045 mẫu

- Tập kiểm thử: 1262 mẫu

Thống kê HR (không chuẩn hóa):

- Min: 63.0000, Max: 126.0000

- Mean: 89.1509, Std: 12.7552

Thống kê RR (không chuẩn hóa):

- Min: 4.0000, Max: 25.0000

- Mean: 17.5660, Std: 3.6472

KẾT QUẢ PHÂN TÍCH BIẾN ĐỔI FOURIER

==================================

Phương pháp phân tích:

1. Biến đổi Fourier nhanh (FFT) để phân tích phổ tần số của tín hiệu PPG gốc và tín hiệu PPG đã tạo.

2. Phương pháp Welch để ước lượng mật độ phổ công suất (PSD) của tín hiệu.

3. Tìm các đỉnh trong phổ tần số để xác định các thành phần tần số chính.

4. Tính toán các chỉ số đánh giá trong miền thời gian và miền tần số.

Kết quả trung bình:

- MSE (miền thời gian): 0.5202

- PSNR: 3.0930dB

- Hệ số tương quan: 0.1877

- MSE (miền tần số): 0.0012

Phân tích chi tiết:

Mẫu 1:

- Điều kiện: HR=81.0000, RR=25.0000

- MSE (miền thời gian): 0.2709

- PSNR: 5.6717dB

- Hệ số tương quan: -0.3480

- MSE (miền tần số): 0.0002

- Các đỉnh tần số của tín hiệu gốc: 1.38 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.38 Hz, 0.75 Hz, 2.62 Hz

Mẫu 2:

- Điều kiện: HR=78.0000, RR=17.0000

- MSE (miền thời gian): 0.5713

- PSNR: 2.4316dB

- Hệ số tương quan: 0.2972

- MSE (miền tần số): 0.0000

- Các đỉnh tần số của tín hiệu gốc: 1.38 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.38 Hz, 2.75 Hz, 0.00 Hz

Mẫu 3:

- Điều kiện: HR=77.0000, RR=15.0000

- MSE (miền thời gian): 0.4486

- PSNR: 3.4810dB

- Hệ số tương quan: 0.6033

- MSE (miền tần số): 0.0012

- Các đỉnh tần số của tín hiệu gốc: 0.00 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.25 Hz, 2.50 Hz, 0.25 Hz

Mẫu 4:

- Điều kiện: HR=74.0000, RR=15.0000

- MSE (miền thời gian): 0.2972

- PSNR: 5.2698dB

- Hệ số tương quan: 0.4599

- MSE (miền tần số): 0.0022

- Các đỉnh tần số của tín hiệu gốc: 0.25 Hz, 1.25 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.25 Hz, 0.25 Hz, 2.50 Hz

Mẫu 5:

- Điều kiện: HR=87.0000, RR=18.0000

- MSE (miền thời gian): 0.4251

- PSNR: 3.7153dB

- Hệ số tương quan: -0.1599

- MSE (miền tần số): 0.0052

- Các đỉnh tần số của tín hiệu gốc: 0.00 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.38 Hz, 18.00 Hz, 2.88 Hz

Mẫu 6:

- Điều kiện: HR=118.0000, RR=14.0000

- MSE (miền thời gian): 0.5729

- PSNR: 2.4195dB

- Hệ số tương quan: -0.5690

- MSE (miền tần số): 0.0001

- Các đỉnh tần số của tín hiệu gốc: 0.00 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 2.12 Hz, 4.12 Hz, 0.38 Hz

Mẫu 7:

- Điều kiện: HR=107.0000, RR=21.0000

- MSE (miền thời gian): 0.6528

- PSNR: 1.8519dB

- Hệ số tương quan: 0.6365

- MSE (miền tần số): 0.0027

- Các đỉnh tần số của tín hiệu gốc: 1.75 Hz, 0.38 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.75 Hz, 3.62 Hz, 0.25 Hz

Mẫu 8:

- Điều kiện: HR=90.0000, RR=20.0000

- MSE (miền thời gian): 0.4173

- PSNR: 3.7950dB

- Hệ số tương quan: 0.8107

- MSE (miền tần số): 0.0003

- Các đỉnh tần số của tín hiệu gốc: 1.50 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.50 Hz, 3.00 Hz, 4.50 Hz

Mẫu 9:

- Điều kiện: HR=63.0000, RR=18.0000

- MSE (miền thời gian): 0.6832

- PSNR: 1.6548dB

- Hệ số tương quan: 0.8175

- MSE (miền tần số): 0.0001

- Các đỉnh tần số của tín hiệu gốc: 1.00 Hz, 2.12 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.00 Hz, 2.12 Hz, 3.12 Hz

Mẫu 10:

- Điều kiện: HR=81.0000, RR=18.0000

- MSE (miền thời gian): 0.8631

- PSNR: 0.6393dB

- Hệ số tương quan: -0.6709

- MSE (miền tần số): 0.0000

- Các đỉnh tần số của tín hiệu gốc: 1.38 Hz, 0.00 Hz, 0.00 Hz

- Các đỉnh tần số của tín hiệu đã tạo: 1.38 Hz, 2.75 Hz, 4.12 Hz

Nhận xét về phổ tần số:

1. Tín hiệu PPG gốc thường có đỉnh tần số chính ở khoảng 1-2 Hz, tương ứng với nhịp tim (60-120 bpm).

2. Tín hiệu PPG đã tạo cũng có xu hướng tái tạo đỉnh tần số chính này, nhưng có thể có sự khác biệt về biên độ.

3. Các thành phần tần số thấp (< 0.5 Hz) liên quan đến nhịp thở thường khó tái tạo chính xác hơn.

4. Tín hiệu PPG đã tạo có thể thiếu một số thành phần tần số cao (> 5 Hz) so với tín hiệu gốc.

Kết luận:

Phân tích biến đổi Fourier cho thấy mô hình CVAE giả lập có thể tạo ra tín hiệu PPG với các đặc tính tần số cơ bản tương tự như tín hiệu gốc, đặc biệt là thành phần tần số liên quan đến nhịp tim. Tuy nhiên, vẫn có sự khác biệt đáng kể trong các thành phần tần số chi tiết, đặc biệt là các thành phần tần số thấp liên quan đến nhịp thở và các thành phần tần số cao. Điều này cho thấy mô hình CVAE thực sự được huấn luyện đầy đủ có thể cải thiện khả năng tái tạo các đặc tính tần số chi tiết của tín hiệu PPG.

KẾT QUẢ ĐÁNH GIÁ MÔ HÌNH CVAE

==============================

Tóm tắt các chỉ số đánh giá:

---------------------------

MSE (miền thời gian):

- Trung bình: 0.5202

- Độ lệch chuẩn: 0.1837

- Nhỏ nhất: 0.2709

- Lớn nhất: 0.8631

PSNR (dB):

- Trung bình: 3.0930

- Độ lệch chuẩn: 1.5965

- Nhỏ nhất: 0.6393

- Lớn nhất: 5.6717

Hệ số tương quan:

- Trung bình: 0.1877

- Độ lệch chuẩn: 0.5737

- Nhỏ nhất: -0.6709

- Lớn nhất: 0.8175

MSE (miền tần số):

- Trung bình: 0.0012

- Độ lệch chuẩn: 0.0017

- Nhỏ nhất: 0.0000

- Lớn nhất: 0.0052

Phân tích ảnh hưởng của HR và RR đến tín hiệu PPG:

------------------------------------------------

1. Ảnh hưởng của HR:

- Tần số cơ bản của tín hiệu PPG tỷ lệ thuận với HR.

- Khi HR tăng, đỉnh tần số chính trong phổ tần số dịch về phía tần số cao hơn.

- Biên độ của tín hiệu PPG có xu hướng giảm khi HR tăng.

2. Ảnh hưởng của RR:

- RR ảnh hưởng chủ yếu đến thành phần tần số thấp của tín hiệu PPG.

- Khi RR tăng, biên độ của thành phần tần số thấp (< 0.5 Hz) tăng.

- RR có ảnh hưởng ít hơn đến hình dạng tổng thể của tín hiệu PPG so với HR.

Đánh giá khả năng tái tạo các đặc trưng quan trọng của tín hiệu PPG:

----------------------------------------------------------------

1. Đặc trưng tần số:

- Mô hình có khả năng tái tạo tốt đỉnh tần số chính (liên quan đến HR).

- Các đỉnh tần số hài bậc cao có thể không được tái tạo chính xác.

- Thành phần tần số thấp (liên quan đến RR) thường khó tái tạo chính xác hơn.

2. Đặc trưng thời gian:

- Hình dạng tổng thể của tín hiệu PPG được tái tạo tương đối tốt.

- Các chi tiết nhỏ và biến động nhanh có thể bị mất trong quá trình tái tạo.

- Tín hiệu tái tạo thường mượt hơn tín hiệu gốc, thiếu một số chi tiết nhiễu.

Hạn chế của mô hình:

------------------

1. Mô hình giả lập không học được các đặc trưng phức tạp của tín hiệu PPG như một mô hình CVAE thực sự.

2. Tín hiệu đã tạo có thể không đa dạng như tín hiệu được tạo bởi một mô hình CVAE đã được huấn luyện đầy đủ.

3. Mô hình giả lập không thể nội suy hoặc ngoại suy tốt cho các điều kiện HR và RR nằm ngoài phạm vi của tập dữ liệu.

4. Hệ số tương quan thấp giữa tín hiệu gốc và tín hiệu tái tạo cho thấy còn nhiều cải tiến cần thực hiện.

5. Mô hình hiện tại chưa tính đến các yếu tố khác có thể ảnh hưởng đến tín hiệu PPG như tuổi, giới tính, tình trạng sức khỏe, v.v.

Kết luận:

--------

Mô hình CVAE giả lập đã chứng minh khả năng tạo ra tín hiệu PPG với các đặc tính cơ bản tương tự như tín hiệu thực, đặc biệt là các đặc tính tần số liên quan đến nhịp tim (HR) và nhịp thở (RR). Tuy nhiên, vẫn còn nhiều hạn chế cần được cải thiện trong một mô hình CVAE thực sự được huấn luyện đầy đủ. Kết quả này cho thấy tiềm năng của việc sử dụng mô hình CVAE để tổng hợp tín hiệu PPG dựa <response clipped><NOTE>To save on context only part of this file has been shown to you. You should retry this tool after you have searched inside the file with `grep -n` in order to find the line numbers of what you are looking for.</NOTE>