Họ và tên: Hồng Anh Khoa

MSSV: 22110351

**Báo cáo về tập dữ liệu: CapnoBase IEEE TBME Respiratory Rate Benchmark**

# **Giới thiệu chung**

Bộ dữ liệu CapnoBase TBME RR benchmark chứa 42 trường hợp ghi lại trong 8 phút. Ngoài các dạng sóng CO₂ (capnogram), các trường hợp này còn có cả tín hiệu Photoplethysmogram (PPG) từ đo độ bão hòa oxy trong máu (pulse oximetry). Nhãn từ chuyên gia có sẵn để đánh dấu các đỉnh xung từ PPG và các nhịp thở từ CO₂.

Ngoài ra, bộ dữ liệu benchmark này còn bao gồm kết quả từ bài báo của Karlen et al. trên IEEE TBME về tính toán nhịp thở đa tham số. Điều này giúp so sánh trực tiếp hiệu suất của các thuật toán.

Bộ dữ liệu này được các nhà nghiên cứu sử dụng để kiểm tra và so sánh các thuật toán. Tuy nhiên, nó không nên được sử dụng để huấn luyện hoặc điều chỉnh thuật toán, vì điều đó có thể làm sai lệch kết quả hiệu suất.

# **Quyền truy cập**

* Bộ dữ liệu **không phải là dữ liệu mở** và có thể yêu cầu đăng ký hoặc sự cho phép từ bên cung cấp.
* Người dùng cần tuân thủ điều khoản sử dụng, không được chia sẻ hoặc sử dụng ngoài phạm vi cho phép.
* Dữ liệu này **không nên được sử dụng để huấn luyện thuật toán** vì có thể gây thiên lệch khi đánh giá hiệu suất.

# **Giấy phép sử dụng**

* Thông thường, bộ dữ liệu này chỉ được sử dụng cho **mục đích nghiên cứu** và so sánh thuật toán.
* **Không được sử dụng cho mục đích thương mại** hoặc ứng dụng lâm sàng mà không có sự chấp thuận từ nhà cung cấp dữ liệu.
* Nếu công bố nghiên cứu sử dụng dữ liệu này, cần **trích dẫn nguồn** theo quy định.

# **Phân tích bộ dữ liệu:**

## **1. Cấu trúc thư mục và tên file trong bộ dữ liệu**

Bộ dữ liệu **CapnoBase IEEE TBME Respiratory Rate Benchmark** chứa nhiều tệp dữ liệu khác nhau, mỗi tệp cung cấp một loại thông tin riêng biệt. Các tệp dữ liệu thường được đặt tên theo quy tắc:

**[caseID]\_8min\_[datatype].csv**

Ví dụ:

* 0009\_8min\_labels.csv
* 0009\_8min\_meta.csv
* 0009\_8min\_param.csv
* 0009\_8min\_reference.csv
* 0009\_8min\_SFresults.csv
* 0009\_8min\_signal.csv

Trong đó:

* **0009**: ID của trường hợp bệnh nhân (mỗi file tương ứng với một phiên đo dữ liệu kéo dài 8 phút).
* **8min**: Chỉ ra rằng mỗi file chứa dữ liệu trong 8 phút đo.
* **[datatype]**: Loại dữ liệu có trong file.

## **2. Ý nghĩa và nội dung của từng file**

|  |  |
| --- | --- |
| **Tên file** | **Ý nghĩa** |
| 0009\_8min\_meta.csv | Chứa **thông tin meta** như nhân khẩu học (tuổi, cân nặng, giới tính) và chế độ thở của bệnh nhân. |
| 0009\_8min\_param.csv | Chứa **các thông số hệ thống**, như tần số lấy mẫu và phương pháp thông khí. |
| 0009\_8min\_signal.csv | Chứa **tín hiệu thô** của CO₂ (capnogram) và Photoplethysmogram (PPG) được ghi nhận trong 8 phút. |
| 0009\_8min\_labels.csv | Chứa **nhãn được chuyên gia đánh dấu**, bao gồm vị trí nhịp tim (pulse peaks) từ PPG và nhịp thở từ CO₂. |
| 0009\_8min\_reference.csv | Chứa **dữ liệu tham chiếu** được tính toán từ nhãn, cung cấp giá trị chuẩn của tần số hô hấp. |
| 0009\_8min\_SFresults.csv | Chứa **kết quả thuật toán Smart Fusion (SF)** của Karlen et al., dùng để ước lượng nhịp thở từ PPG. |

## **3. Cách sử dụng dữ liệu**

Dữ liệu này thường được sử dụng trong nghiên cứu về nhịp thở và nhịp tim, với các ứng dụng như:  
✅ Kiểm tra và so sánh thuật toán ước lượng tần số hô hấp.  
✅ Nghiên cứu về nhịp tim và nhịp thở từ PPG và CO₂.  
✅ Phân tích ảnh hưởng của chế độ thông khí lên nhịp thở.  
✅ Đánh giá độ chính xác của các cảm biến y tế trong đo hô hấp.

## **4. Chi tiết các tập dữ liệu:**

* **0009\_8min\_signal.csv: Chứa dữ liệu tín hiệu gồm:**
* co2\_y: Dữ liệu CO2
* pleth\_y: Dữ liệu PPG (nhịp tim)
* ecg\_y: Dữ liệu ECG
* **0009\_8min\_labels.csv: Chứa các nhãn từ chuyên gia, bao gồm:**
* co2\_startexp\_x, co2\_startinsp\_x: Các điểm bắt đầu thở ra/hít vào dựa trên CO2
* pleth\_peak\_x: Các đỉnh của tín hiệu PPG
* ecg\_peak\_x: Các đỉnh của tín hiệu ECG
* units\_x: Đơn vị đo
* **0009\_8min\_meta.csv: Thông tin mô tả bệnh nhân và điều kiện đo, gồm:**
* subject\_age: Tuổi bệnh nhân
* subject\_weight: Cân nặng
* treatment\_ventilation: Điều kiện thông khí
* **0009\_8min\_param.csv: Thông tin về tần số lấy mẫu:**
* samplingrate\_pleth, samplingrate\_co2, samplingrate\_ecg: Tần số lấy mẫu của từng loại tín hiệu
* case\_ventilation: Loại thông khí
* **0009\_8min\_reference.csv: Dữ liệu tham chiếu về nhịp thở (rr\_co2\_y) và nhịp tim (hr\_pleth\_y, hr\_ecg\_y).**
* **0009\_8min\_SFresults.csv: Kết quả từ thuật toán phân tích tín hiệu, có thể chứa dữ liệu về nhịp thở/tim ước tính.**

# **Đọc dữ liệu:**

## **1. Kết nối đến Drive cá nhân:**

|  |
| --- |
| from google.colab import drive  drive.mount('/content/drive') |

Kết quả:

|  |
| --- |
| Mounted at /content/drive |

## **2. Đọc các tệp trong thư mục:**

|  |
| --- |
| import os  path = "/content/drive/MyDrive/dataverse\_files"  # Liệt kê file trong thư mục sau khi giải nén  files = os.listdir(path)  print(files) |

Kết quả:

|  |
| --- |
| ['MANIFEST.TXT', 'README.txt', 'preview', 'data'] |

## **3. Cài đặt thư viện:**

|  |
| --- |
| import pandas as pd  import numpy as np  import matplotlib.pyplot as plt  import seaborn as sns |

## **4. Đọc các file tín hiệu của 1 người**

|  |
| --- |
| # Đọc file tín hiệu chính  signal\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_signal.csv")  # Đọc file nhãn (label)  labels\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_labels.csv")  # Đọc file meta (thông tin chung)  meta\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_meta.csv")  # Đọc file tham chiếu  reference\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_reference.csv")  # Đọc file thông số  param\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_param.csv")  # Đọc kết quả thuật toán SF  sf\_results\_df = pd.read\_csv("/content/drive/MyDrive/dataverse\_files/data/csv/0009\_8min\_SFresults.csv") |

## **5. Kiểm tra các file:**

|  |
| --- |
| print("Signal Data:\n", signal\_df.head())  print("\n\n\nLabels Data:\n", labels\_df.head())  print("\n\n\nMeta Data:\n", meta\_df.head())  print("\n\n\nReference Data:\n", reference\_df.head())  print("\n\n\nParameters Data:\n", param\_df.head())  print("\n\n\nSF Results Data:\n", sf\_results\_df.head()) |

Kết quả:

|  |
| --- |
| Signal Data:  co2\_y pleth\_y ecg\_y  0 1.50 -0.64 -0.31  1 1.54 -0.64 -0.34  2 1.59 -0.67 -0.31  3 1.64 -0.69 -0.26  4 1.68 -0.72 -0.18  Labels Data:  co2\_artif\_x co2\_startexp\_x \  0 NaN 974 1990 2974 3966 4990 5953 6998 7988 8995 9...  co2\_startinsp\_x pleth\_artif\_x \  0 659 1660 2661 3647 4658 5636 6670 7655 8676 9... NaN  pleth\_peak\_x \  0 59 242 422 611 800 983 1157 1334 1517 1703 18...  ecg\_peak\_x ecg\_artif\_x units\_x  0 164 346 532 718 898 1078 1258 1440 1621 1800 ... NaN samples  Meta Data:  subject\_gender subject\_age subject\_weight treatment\_ventilation \  0 NaN 4 18 spontaneous  persistentID  0 <https://doi.org/10.5683/SP2/NLB8IT>  Reference Data:  rr\_co2\_x \  0 21.5333 24.87 28.1567 31.5267 34.7867 38.2333...  rr\_co2\_y \  0 17.982 17.982 18.2556 17.8042 18.4049 17.4081...  hr\_pleth\_x \  0 0.806667 1.40667 2.03667 2.66667 3.27667 3.85...  hr\_pleth\_y \  0 98.3607 100 95.2381 95.2381 98.3607 103.448 1...  hr\_ecg\_x \  0 1.15333 1.77333 2.39333 2.99333 3.59333 4.193...  hr\_ecg\_y units\_x units\_hr\_y \  0 98.9011 96.7742 96.7742 100 100 100 98.9011 9... s beats/min  units\_rr\_y  0 breaths/min  Parameters Data:  samplingrate\_pleth samplingrate\_co2 samplingrate\_ecg case\_id \  0 300 300 300 0009\_8min  case\_ventilation  0 spontaneous  SF Results Data:  x RIFV\_y RIAV\_y RIIV\_y \  0 32 35 38 41 44 47 50 53 56 59 62 65 68 71 74 ... NaN NaN NaN  Fusion\_y SmartFusion\_y  0 NaN NaN |

## **6. Xử lý dữ liệu:**

|  |
| --- |
| # Đổi tên cột để dễ hiểu hơn  signal\_df.columns = ["co2\_y", "pleth\_y", "ecg\_y"]  # Thêm cột thời gian (giả sử mỗi mẫu cách nhau 0.01 giây)  signal\_df["time"] = np.arange(0, len(signal\_df) \* 0.01, 0.01) |

## **7. Vẽ biểu đồ tín hiệu:**

|  |
| --- |
| plt.figure(figsize=(15, 10))  # Vẽ tín hiệu CO2  plt.subplot(3, 1, 1)  plt.plot(signal\_df["time"], signal\_df["co2\_y"], color="blue")  plt.ylabel("CO2 Signal")  plt.title("CO2, PPG, and ECG Signals")  plt.grid()  # Vẽ tín hiệu PPG  plt.subplot(3, 1, 2)  plt.plot(signal\_df["time"], signal\_df["pleth\_y"], color="green")  plt.ylabel("PPG Signal")  plt.grid()  # Vẽ tín hiệu ECG  plt.subplot(3, 1, 3)  plt.plot(signal\_df["time"], signal\_df["ecg\_y"], color="red")  plt.xlabel("Time (seconds)")  plt.ylabel("ECG Signal")  plt.grid()  plt.tight\_layout()  plt.show() |

Kết quả:

## **8. Phân tích thống kê số liệu:**

|  |
| --- |
| print("Statistical Summary for CO2 Signal:")  print(signal\_df["co2\_y"].describe())  print("\nStatistical Summary for PPG Signal:")  print(signal\_df["pleth\_y"].describe())  print("\nStatistical Summary for ECG Signal:")  print(signal\_df["ecg\_y"].describe()) |

Kết quả:

|  |
| --- |
| Statistical Summary for CO2 Signal:  count 144001.000000  mean 5.165647  std 2.666212  min 0.460000  25% 2.310000  50% 6.780000  75% 7.330000  max 7.960000  Name: co2\_y, dtype: float64  Statistical Summary for PPG Signal:  count 144001.000000  mean -1.477106  std 4.615148  min -10.160000  25% -4.350000  50% -1.360000  75% 1.650000  max 10.240000  Name: pleth\_y, dtype: float64  Statistical Summary for ECG Signal:  count 144001.000000  mean 0.410500  std 2.757249  min -13.630000  25% -0.290000  50% 0.200000  75% 1.240000  max 14.810000  Name: ecg\_y, dtype: float64 |

## **9. Phân tích nhịp thở:**

|  |
| --- |
| # Tìm đỉnh trong tín hiệu CO2 (nhịp thở)  peaks\_co2, \_ = find\_peaks(signal\_df["co2\_y"], distance=100)  # Số lượng đỉnh trong 8 phút  breaths\_per\_min = len(peaks\_co2) / 8  print(f"Estimated Respiratory Rate (RR): {breaths\_per\_min:.2f} breaths per minute") |

Kết quả:

|  |
| --- |
| Estimated Respiratory Rate (RR): 49.75 breaths per minute |

# **Tổng kết:**

## **1. Thông tin về người 0009:**

Giới tính: không rõ

Tuổi: 4

Cân nặng: 14

Hình thức hô hấp: tự nhiên

## **2. Thông tin cài đặt và thông số thiết bị:**

Tần số lẫy mẫu PPG (samplingrate\_pleth): 300

Tần số lẫy mẫu CO2 (samplingrate\_co2): 300

Tần số lẫy mẫu ECG (samplingrate\_ecg): 300

Mã người bệnh (case\_id): 0009\_8min

## **3. Dữ liệu tham chiếu nhịp tim và thở:**

Nhịp thở từ CO2 (rr\_co2\_x, rr\_co2\_y):

Nhịp tim từ PPG (hr\_pleth\_x, hr\_pleth\_y)

Nhịp tim từ ECG (hr\_ecg\_x, hr\_ecg\_y)

Đơn vị đo (units\_x, units\_hr\_y, units\_rr\_y)

## **4. Tín hiệu sinh lý theo thời gian:**

CO2 (co2\_y)

Tín hiệu PPG (pleth\_y)

ECG (ecg\_y)

Tổng cộng có 144.001 điểm dữ liệu