# **Báo cáo tuần 8**

Các công việc đã làm được trong tuần trước

* Nghiên cứu về dataset sẽ sử dụng
* Đọc các technique đã được sử dụng với EEG
* Triển khai trên 1 dataset

Các công việc làm được trong tuần này:

* Xây dựng được bộ Feature cho dataset sử dụng
* So sánh việc sử dụng bộc lọc Band Pass Filter
* Giảm chiều dữ liệu, tăng accuracy

1. **DATASET**

Các dataset về EEG là tín hiệu tin sinh, nên việc public các dataset đó rất hạn chế, và số lượng dataset cũng là rất ít. Ở đây em có nghiên cứu về 2 bộ dataset chính

|  |  |
| --- | --- |
| **MODMA (Cai – 2020)** | **Mumtaz - 2016** |
| HC/MDD: 29/24 | HC/MDD: 28/30 |
| Age: 31.5 9.2 | Age: 38.2 15.6 |
| Location: Lanzhou, China | Location: HUSM, Malaysia |
| Electrode: 128 | Electrode: 19 |
| Fs = 250 Hz | Fs = 256 Hz |
| Info:   * Resting: trong 5 phút, giữ tỉnh táo, nhắm mắt và không tạo ra bất kỳ hoạt động cơ thể nào (.mat) * Act: tập trung sự chú ý vào hình ảnh của các khuôn mặt cảm xúc (sad, happy, fear) (.raw) | Info:   * Nhắm mắt trong 5 phút, mở mắt trong 5 phút và có một kích thích nhất định trong vòng 10 phút (.edf) |
| **Feature extraction:**   * Trích xuất các đặc trưng tuyến tính và phi tuyến của các sóng đặc trưng: alpha, beta, gama, theta * Sử dụng các phương pháp giảm chiều dữ liệu | **Feature Extraction:**   * Tạo ra các ma trận bất đối xứng nhịp điệu khác nhau ở bán cầu trái và bán cầu phải * Sử dụng ICA để loại bỏ nhiễu (EC, OC) |

1. **TIẾN HÀNH TRIỂN KHAI TRÊN DATASET MODMA**

A diagram of a process

Description automatically generated

*Flow Chart prediction depression using EEG signal*

* Preprocessing:

Vì dữ liệu EEG đầu vào là gồm 128 kênh, và để đơn giản bài toán, ta sẽ rút gọn đi còn 16 kênh chủ đạo cho việc phân tích.

A diagram of a person's hand

Description automatically generated

A diagram of a diagram of a electrode placement

Description automatically generated

Với dataset MODMA này, ta chỉ xét các đặc tính tuyến tính và phi tuyến của các sóng cơ bản delta (0,5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) và dải beta (13–40 Hz), nên ta sẽ dùng bộ lọc Band Pass filter để chỉ lấy trong vùng từ 0 - 40Hz

A graph showing a number of colors

Description automatically generated

Sóng EEG được đặc trưng bởi 4 loại sóng cơ bản: Alpha, Beta, Delta, Theta, với mỗi sóng cơ bản sẽ được xuất hiện ở những tần số nhất định. Những sóng này sẽ cung cấp cho chúng ta thông tin về tình trạng sức khỏe của người bệnh thông qua những đặc điểm được thể hiện ở từng loại sóng. Do đó, ta sẽ xem xét tất cả những đặc trưng mà nó thể hiện.

**Đặc trưng được chia ra làm 2 loại:**

**Linear Feature:** là các đặc điểm tuyến tính bao gồm công suất ở 4 sóng cơ bản. Biên độ của tín hiệu công suất bao gồm giá trị trung bình, trung vị, cực đại và cực tiểu được sử dụng làm đặc điểm tuyến tính.

Với đặc điểm tuyến tính này, ta sẽ trích xuất **max, min, mean, median công suất** của cả đoạn (từ 0 – 40Hz) và giá trị trung bình biên độ của 4 sóng cơ bản.

A table with numbers and symbols

Description automatically generated

*Ví dụ minh họa đặc trưng tuyến tính*

**Non-Linear Feature**: Các đặc điểm phi tuyến được sử dụng trong phân tích là entropy phổ và entropy lắng đọng giá trị đơn (Singular – value, Spectral entropy, Permenone entropy…)

A table with numbers and letters

Description automatically generated

*Ví dụ minh họa đặc trưng phi tuyến*

Có 21 kênh kích thích đầu vào, đặc trưng cho từng cảm xúc, trong đó phân chia thành 3 kích thích cơ bản: **Fear, Happy, Sad**, với mỗi cảm xúc được theo dõi bởi chỉ số thời gian của các mẫu dữ liệu, nên ta có phép phân tích cụ thể các phản ứng của não đối với từng loại kích thích, và nó sẽ thể hiện đặc trưng trên từng sóng cơ bản. Do đó, ta sẽ trích xuất kích thích của **Fear, Happy, Sad** lên từng sóng cơ bản một xem những đặc trưng của nó được thể hiện như thế nào.

1. **TRAINING MODEL**

A screenshot of a computer program

Description automatically generated

*Hình ảnh tính toán tổng số Features*

Có tất cả 53 bệnh nhân, với mỗi bệnh nhân ta sẽ trích xuất đặc điểm của từng kênh sóng một (cụ thể là ta sẽ trích xuất trên 16 channels trước). Ở đây, với mỗi channel sẽ có các đặc điểm về biên độ tuyến tính và công suất phi tuyến. Và đồng thời với nó là các kích thích đầu vào ảnh hưởng lên các đặc điểm đó. Tổng cộng lại ta sẽ có 512 features tuyến tính và 192 features phi tuyến.

A table with numbers and letters

Description automatically generated

*Hình ảnh ví dụ 1 số feature ở kênh 104*

Ta sẽ sử dụng kiến thức Machine Learning cơ bản để train model. Dưới đây ta sẽ đi xem xét và so sánh trước và sau khi sử dụng các bộ lọc

A graph of a number of blue rectangular bars

Description automatically generated with medium confidence

Trước và sau khi lọc band pass filter (chỉ ảnh hưởng đến feature phi tuyến) (sau – trước)

A screenshot of a black screen

Description automatically generated A screenshot of a black screen

Description automatically generated

*Hình ảnh feature phi tuyến trong 2 trường hợp*

A graph of blue and black bars

Description automatically generated with medium confidence

Ta nhận thấy rằng, việc sử dụng bộ lọc Band Pass filter chỉ ảnh hưởng đến tính toán công suất (192 features), và sự ảnh hưởng đó đến độ chính xác của model là không nhiều. Cụ thể, so sánh giữa việc sử dụng và không sử dụng thì không có sự thay đổi rõ rệt nào cả.

Vì độ chính xác chưa cao, ta sẽ đi giảm chiều dữ liệu bằng cách tính toán các tương quan các kênh với nhau

A colorful squares with black text

Description automatically generated A colorful squares with black text

Description automatically generated with medium confidence

Ta sẽ sử dụng tương quan Pearson trên các kênh và các kích thích đầu vào để có thể loại bớt các kênh không có nhiều ảnh hưởng tới mô hình.

Chi tiết về kết quả sau khi tính toán tương quan & PCA em xin được báo cáo trong tuần sau.