

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ - ĐẠI HỌC QUỐC
GIA HÀ NỘI
VIỆN TRÍ TUỆ NHÂN TẠO**

-----***-----

**BÁO CÁO MINI-PROJECT BCI VỀ BÀI TOÁN EMOTION
RECOGNITION SỬ DỤNG CÁC THUẬT TOÁN HỌC MÁY VÀ HỌC
SÂU TRÊN BỘ DỮ LIỆU SEED-IV**

**Họ và tên sinh viên thực hiện: Lèng Hữu Phúc
Mã sinh viên: 22022570**

HÀ NỘI, 12/2024

MỞ ĐẦU

Nhận diện cảm xúc (Emotion Recognition) là một lĩnh vực nghiên cứu quan trọng trong trí tuệ nhân tạo và học máy, với mục tiêu xác định và phân loại trạng thái cảm xúc của con người dựa trên các tín hiệu đầu vào như hình ảnh khuôn mặt, giọng nói, văn bản, hoặc dữ liệu sinh lý. Cảm xúc đóng vai trò quan trọng trong giao tiếp và tương tác xã hội, ảnh hưởng sâu sắc đến hành vi, suy nghĩ và quyết định của con người. Do đó, việc phát triển các hệ thống có khả năng nhận diện cảm xúc không chỉ hỗ trợ cải thiện giao tiếp giữa người và máy mà còn mở ra nhiều ứng dụng trong các lĩnh vực như chăm sóc sức khỏe, giáo dục, thương mại và giải trí.

Bài toán nhận diện cảm xúc thường đối mặt với nhiều thách thức, bao gồm sự đa dạng trong cách thể hiện cảm xúc của mỗi cá nhân, sự khác biệt văn hóa và ngữ cảnh, cũng như hạn chế về chất lượng và kích thước của dữ liệu huấn luyện. Các phương pháp hiện đại thường dựa trên các mô hình học sâu, sử dụng mạng nơ-ron tích chập (CNN) hoặc mạng nơ-ron tuần hoàn (RNN) để xử lý dữ liệu hình ảnh và chuỗi thời gian. Bên cạnh đó, việc kết hợp đa nguồn dữ liệu (hình ảnh, âm thanh, và sinh lý) cũng đang là một hướng nghiên cứu tiềm năng nhằm nâng cao độ chính xác của hệ thống.

Với sự tiến bộ của công nghệ, đặc biệt là sự phát triển của các máy BCI không xâm lấn đã đem một luồng gió mới đến cho bài toán emotion recognition. Với các tín hiệu EEG thì bài toán trở nên được uy tín hơn và nó cũng đặt nền móng cho công cuộc khai phá trí thức của loài người.

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN VỀ TÁC VỤ BCI VÀ BÀI TOÁN EMOTION RECOGNITION	3
1.1 Tín hiệu EEG	3
1.2 Giới thiệu về tác vụ BCI	4
1.3 Các dạng sóng trong điện não đồ	5
1.4 Bài toán Emotion Recognition	5
CHƯƠNG 2: TỔNG QUAN VỀ BỘ DỮ LIỆU SEED-IV	5
2.1 Giới thiệu về bộ dữ liệu SEED-IV	5
CHƯƠNG 3: GIỚI THIỆU CÁC THUẬT TOÁN HỌC MÁY - KIẾN TRÚC HỌC SÂU ĐƯỢC SỬ DỤNG	8
3.1 Giới thiệu các thuật toán học máy	9
3.2 Các kiến trúc mạng học sâu	15
CHƯƠNG 4: GIỚI THIỆU VỀ CẤU TRÚC DỰ ÁN	16
5.1 Kết quả thực nghiệm	16
5.2 Thảo luận	17
5.3 Phương hướng tương lai	17

CHƯƠNG 1: TỔNG QUAN VỀ TÁC VỤ BCI VÀ BÀI TOÁN EMOTION RECOGNITION

1.1 Tín hiệu EEG

Điện não đồ (Electroencephalogram, viết tắt là EEG) là một phương pháp y học dùng để ghi lại hoạt động điện của não bộ con người hoặc động vật. Điều này thường được thực hiện bằng cách đặt các điện cực nhỏ lên da đầu của người hoặc động vật và ghi lại các tín hiệu điện tử phát ra từ não.

EEG giúp tạo ra một biểu đồ hoặc bản đồ của hoạt động điện của não, cho phép các chuyên gia y tế và nghiên cứu viên theo dõi và phân tích sự thay đổi trong hoạt động não bộ. Các sóng điện não đồ được phân thành nhiều loại khác nhau, bao gồm sóng alpha, beta, theta, delta, mỗi loại sóng có tính chất riêng biệt. EEG thường được sử dụng trong nhiều lĩnh vực y học và nghiên cứu, bao gồm:

- Chẩn đoán rối loạn động kinh: EEG là công cụ quan trọng trong việc chẩn đoán và phân loại các loại động kinh. Nó giúp xác định sự thay đổi trong hoạt động sóng não của bệnh nhân trong khi họ trải qua cơn động kinh.
- Nghiên cứu về tâm trạng và tâm lý: Điện não đồ có thể được sử dụng để nghiên cứu sự thay đổi trong hoạt động não bộ liên quan đến tâm trạng, tâm lý và các rối loạn tâm thần khác.
- Theo dõi sự thay đổi trong giấc ngủ: EEG có thể giúp theo dõi các chu kỳ giấc ngủ và xác định sự thay đổi trong hoạt động não bộ trong suốt quá trình ngủ.
- Nghiên cứu về sự tập trung và tư duy: Điện não đồ có thể được sử dụng để nghiên cứu hoạt động não bộ trong khi người tham gia thực hiện các nhiệm vụ tập trung hoặc tư duy đặc biệt.
- Điều khiển thiết bị bằng suy nghĩ: Trong lĩnh vực trí tuệ nhân tạo và giao tiếp bằng suy nghĩ, EEG được sử dụng để đọc tín hiệu não và điều khiển các thiết bị như máy tính hoặc xe lăn bằng suy nghĩ.

1.2 Giới thiệu về tác vụ BCI

Tác vụ trong Giao diện não-máy (Brain-Computer Interface, BCI) liên quan đến việc sử dụng tín hiệu thần kinh để điều khiển các thiết bị hoặc thực hiện các hành động mà không cần đến sự can thiệp cơ học truyền thống. Các tác vụ phổ biến bao gồm điều khiển con trỏ máy tính, vận hành robot, đánh máy bằng ý nghĩ, hoặc phục hồi chức năng cho người bị liệt thông qua các thiết bị hỗ trợ. Một số tác vụ phức tạp hơn, như nhận diện cảm xúc, dự đoán trạng thái tập trung, hoặc giao tiếp trực tiếp qua tín hiệu não, cũng đang được nghiên cứu sâu rộng.

Những tác vụ này thường dựa vào việc giải mã các tín hiệu điện não đồ (EEG) hoặc các dạng tín hiệu thần kinh khác thông qua các thuật toán học máy. Tác vụ BCI không chỉ đòi hỏi độ chính xác cao mà còn phải đảm bảo tính tương tác thời gian thực và sự thoải mái cho người dùng. Tiềm năng của BCI trong cải thiện cuộc sống con người là rất lớn, đặc biệt là trong y tế, giáo dục và ứng dụng công nghệ tiên tiến.

1.3 Các dạng sóng trong điện não đồ

Điện não đồ (EEG) ghi lại các tín hiệu điện phát ra từ hoạt động của não bộ, được biểu diễn dưới dạng các dải tần số khác nhau gọi là **các dạng sóng não**. Các dạng sóng chính bao gồm:

- **Delta (0.5 - 4 Hz):** Xuất hiện trong trạng thái ngủ sâu, liên quan đến quá trình phục hồi và tái tạo cơ thể.
- **Theta (4 - 8 Hz):** Liên kết với trạng thái thư giãn, suy nghĩ sáng tạo, hoặc giấc ngủ nông.

- **Alpha (8 - 13 Hz):** Phản ánh trạng thái thư giãn tỉnh táo, thường xuất hiện khi mắt nhắm và cơ thể nghỉ ngơi.
- **Beta (13 - 30 Hz):** Liên quan đến sự tập trung, tư duy tích cực, hoặc các hoạt động tâm lý căng thẳng.
- **Gamma (30 - 100 Hz):** Được cho là liên quan đến quá trình nhận thức cao cấp như ghi nhớ, học tập, và xử lý thông tin.

1.4 Bài toán Emotion Recognition trong hệ thống cải thiện người dùng mua sắm

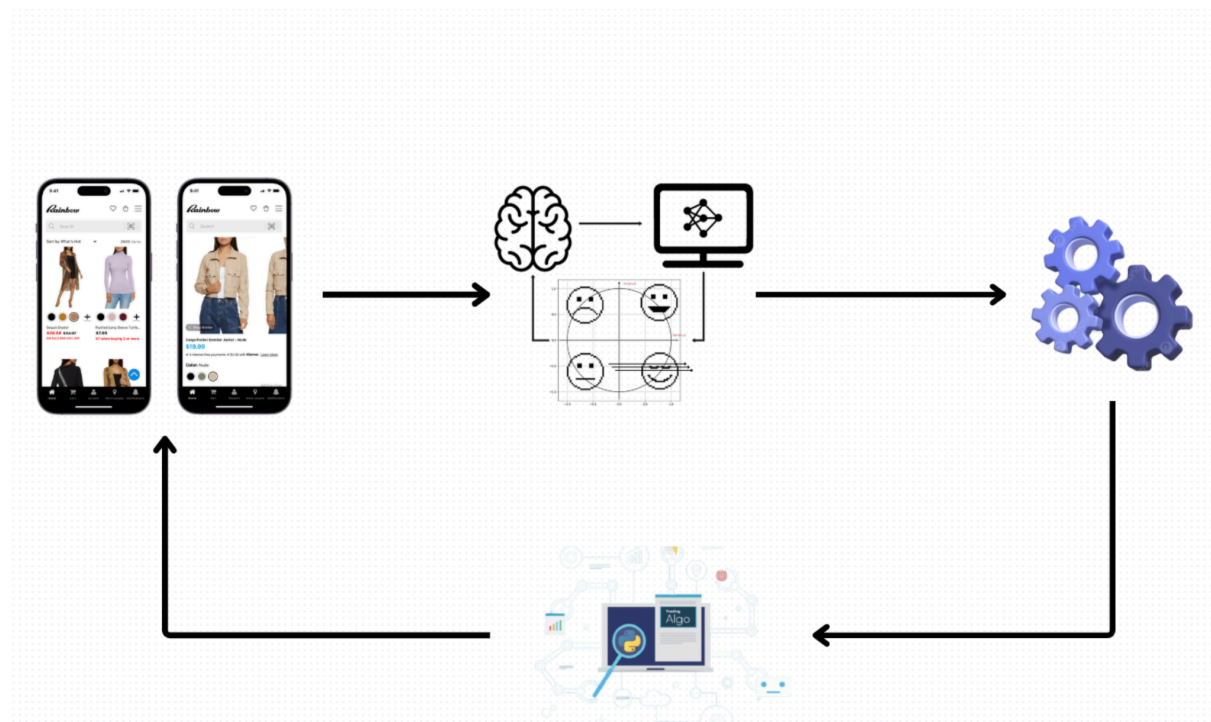
Bài toán Emotion Recognition thuộc vào bài toán Classification trong lĩnh vực trí tuệ nhân tạo. Trong bài báo cáo này thì bài toán được mô tả như sau:

Input: Tín hiệu EEG

Output: Các cảm xúc như vui buồn ..

Sử dụng các thuật toán phân loại trong học máy hay các mạng học sâu nhằm giải quyết bài toán này là mục tiêu của bài báo cáo này.

Giới thiệu sơ qua về hệ thống cải thiện người dùng mua sắm sử dụng emotion Recognition:



Người sử dụng app được mời ngẫu nhiên thành các cụm khác nhau để lấy mẫu tín hiệu EEG khi sử dụng app - Trong khi sử dụng thì các thuật toán chạy theo thời gian thực nhằm phân tích xem khách hàng sử dụng sẽ cảm thấy thú vị với các quảng cáo nào hay cũng như họ tỏ ra khó chịu khi mà sử dụng ứng dụng ở chỗ nào -> để từ đó cải thiện lại các thuật toán theo cụm người được lấy mẫu nhằm cho mỗi khách hàng theo cụm sẽ được tối ưu riêng, tăng trải nghiệm người dùng tối đa cho các sản phẩm thương mại.

CHƯƠNG 2: TỔNG QUAN VỀ BỘ DỮ LIỆU SEED-IV

2.1 Giới thiệu về bộ dữ liệu SEED-IV

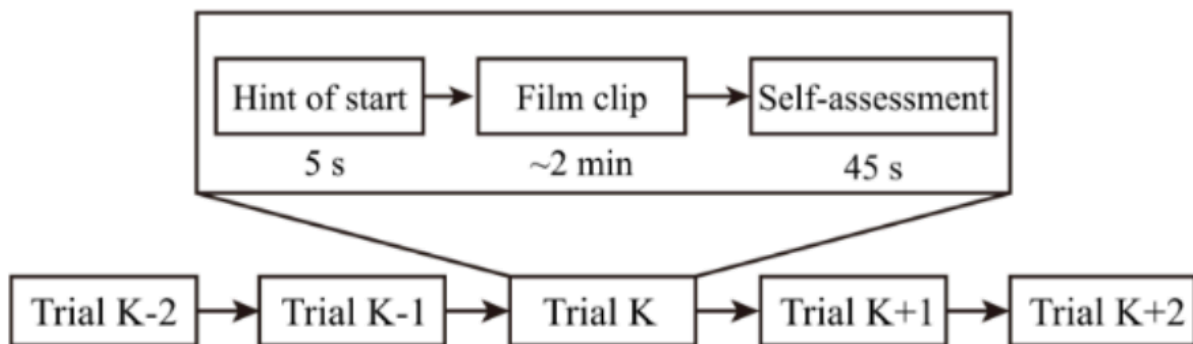
SEED-IV là một bộ dữ liệu nổi tiếng trong lĩnh vực nghiên cứu nhận diện cảm xúc sử dụng tín hiệu não, được phát triển bởi nhóm nghiên cứu của Đại học Khoa học và Công nghệ Trung Quốc (USTC). Bộ dữ liệu này mở rộng từ bộ SEED gốc, tập trung vào bốn trạng thái cảm xúc: **vui vẻ, buồn bã, tức giận, và bình thản**.

Dữ liệu SEED-IV được thu thập thông qua điện não đồ (EEG) sử dụng 62 kênh điện cực từ các đối tượng đang xem các đoạn video được chọn lọc để kích thích cảm xúc. Mỗi đối tượng tham gia thực hiện nhiều phiên ghi, đảm bảo tính đa dạng và độ chính xác của dữ liệu. Bộ dữ liệu này cũng bao gồm thông tin như tín hiệu EOG (điện nhãn đồ) để hỗ trợ xử lý nhiễu và cải thiện phân tích.

a, Phương pháp thu nhập tín hiệu

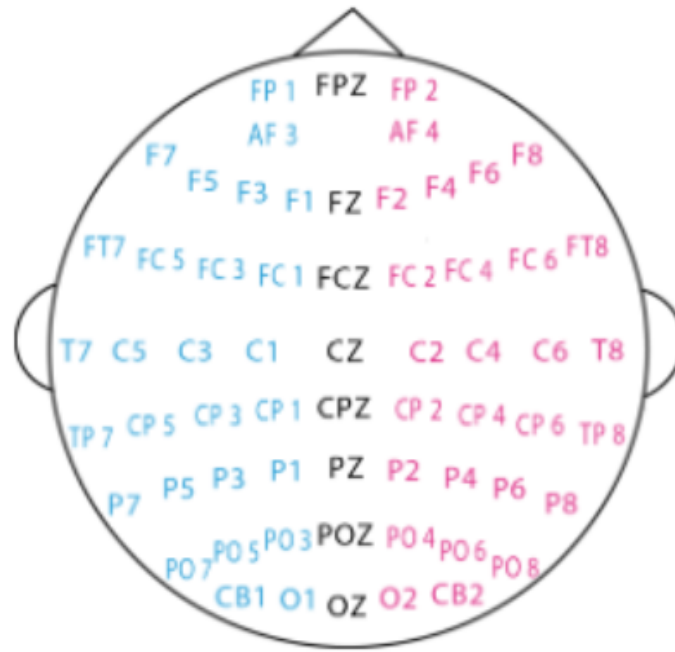
Các nhà nghiên cứu thuộc Đại học Khoa học và Công nghệ Trung Quốc (USTC) sử dụng 72 đoạn phim có xu hướng tạo ra cảm xúc **vui vẻ, buồn bã, tức giận, và bình thản**

Tổng cộng có 15 đối tượng qua 3 phiên, mỗi phiên 24 lần thử nghiệm, sử dụng hệ thống ESI NeuroScan và kính SMI, sơ đồ thí nghiệm như sau:



Sơ đồ thí nghiệm của bộ dataset SeedIV.

64 Kênh EEG được thu nhập thông qua hệ thống BCI không xâm lấn được mô tả theo sơ đồ sau:



Hình mô tả các điện cực trên đầu .

b, Phương pháp trích suất đặc trưng

Đầu tiên, tín hiệu EEG thô được downsample xuống tần số 200 Hz. Bộ dữ liệu này đã được làm sạch bằng cách sử dụng một band-pass filter trong khoảng 1 tới 75 Hz để lọc bớt nhiễu.

Sau đó, tiếp tục trích suất đặc trưng PSD (Power spectral density) và đặc trưng DE (differential entropy) với 5 dải tần số:

1 delta: 1 - 4 Hz;

2 theta: 4 - 8 Hz;

3 alpha: 8 - 14 Hz;

4 beta: 14 - 31 Hz;

5 gamma: 31 - 50 Hz.

Công thức tính PSD và DE của 1 biến x là:

$$\text{PSD} = \mathbb{E}[x^2]$$

$$DE = - \int_{-\infty}^{\infty} P(x) \ln(P(x)) dx$$

Giả sử tín hiệu EEG tuân theo một phân phối chuẩn $x \sim \mathcal{N}(\mu, \sigma^2)$. Có thể tính DE như sau:

$$DE = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx = \frac{1}{2} \ln(2\pi e \sigma^2)$$

Mỗi tệp ".mat" lưu trữ một cấu trúc với các trường có tên "X_YZ"

- "X" biểu thị loại đặc trưng, có thể là "psd" hoặc "de".
- "Y" biểu thị phương pháp làm mượt, có thể là "movingAve" hoặc "LDS". Hệ thống động tuyến tính (LDS) và trung bình động (moving averages) là hai phương pháp khác nhau để loại bỏ nhiễu và các tín hiệu không liên quan đến các đặc trưng EEG.
- "Z" biểu thị số thử nghiệm (trial number)

c, Mô tả tập dữ liệu

Bộ dữ liệu được chia thành 4 folder:

- "eeg_raw_data" đây là dữ liệu EEG raw
- "eeg_feature_smooth" dữ liệu EEG đã được tiền xử lý
- "eye_raw_data" dữ liệu mắt raw
- "eye_feature_smooth" dữ liệu mắt đã tiền xử lý.

CHƯƠNG 3: GIỚI THIỆU CÁC THUẬT TOÁN HỌC MÁY - KIẾN TRÚC HỌC SÂU ĐƯỢC SỬ DỤNG

3.1 Giới thiệu các thuật toán học máy

a, Support vector machine - SVM

Support vector machine là thuật toán tìm siêu phẳng tối ưu hyperland trong không gian đặc trưng hyperland nhằm phân loại các điểm dữ liệu sao cho khoảng cách giữa siêu phẳng và các điểm dữ liệu gần nhất của mỗi lớp là lớn nhất.

Phương trình siêu phẳng trong không gian R^d có phương trình tổng quát là:

$$w^t x + b = 0$$

Trong đó:

- $w \in R^d$ là vector trọng số
- b là hệ số điều chỉnh (bias)

Khoảng cách giữa siêu phẳng và các điểm gần nhất là:

$$\text{Margin} = \frac{1}{\|w\|}$$

Điều kiện phân loại chính xác: mỗi điểm dữ liệu thỏa mãn điều kiện phân loại:

$$y_i(w^T x_i + b) \geq 1, \forall i$$

Hàm mục tiêu:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

Ràng buộc:

$$\min_{w,b} \text{ subject to } y_i(w^T x_i + b) \geq 1, \forall i$$

b, Thuật toán XGBoost

XGBoost (Extreme Gradient Boosting) là một thuật toán học máy mạnh mẽ, thuộc loại ensemble learning và sử dụng phương pháp gradient boosting để xây dựng các mô hình dự đoán. Mục tiêu của XGBoost là cải thiện hiệu suất của mô hình thông qua việc kết hợp các cây quyết định yếu thành một mô hình mạnh mẽ. XGBoost xây dựng mô hình bằng cách thêm các cây quyết định vào mô hình hiện tại theo cách mà mỗi cây mới sẽ sửa lỗi (residuals) của cây trước đó. Quá trình này được thực hiện theo từng bước, mỗi bước cải thiện dần mô hình dựa trên gradient của hàm mất mát.

Hàm mất mát:

$$L(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \Omega(f)$$

Trong đó:

- $L(y_i, \hat{y}_i)$ là sai số giữa giá trị thực và giá trị dự đoán
- $\Omega(f)$ là điều chỉnh độ phức tạp của mô hình để tránh overfitting

Phương pháp:

- Gradient Descent: XGBoost sử dụng gradient descent nhằm tối ưu hóa các tham số của mô hình với mỗi decision tree được thêm vào để điều chỉnh các sai số của mô hình dự đoán trước đó.
- Tối ưu hóa với Regularization: XGBoost áp dụng kỹ thuật regularizations (L1 và L2) nhằm ngăn ngừa overfitting cũng như cải thiện tổng quát mô hình

c, Thuật toán Logistic Regression

Định nghĩa: là một biến thể của hồi quy tuyến tính phục vụ cho bài toán phân loại,

1. Mô hình tuyến tính:

Giống như hồi quy tuyến tính, logistic regression bắt đầu bằng cách tính giá trị

z (còn gọi là logit) dựa trên các biến đầu vào x :

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p = \theta^T x$$

Trong đó θ là vector tham số còn x là vector đặc trưng.

2. Hàm logistic:

Để đưa z vào đoạn $[0, 1]$ để biểu diễn xác suất xuất hiện của nhãn, logistic regression sử dụng hàm logistic (hay sigmoid) để chuyển đổi giá trị tuyến tính z :

$$h(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

Kết quả của hàm logistic này, $g(z)$, biểu diễn xác suất mà một điểm dữ liệu thuộc về lớp 1 (so với lớp 0).

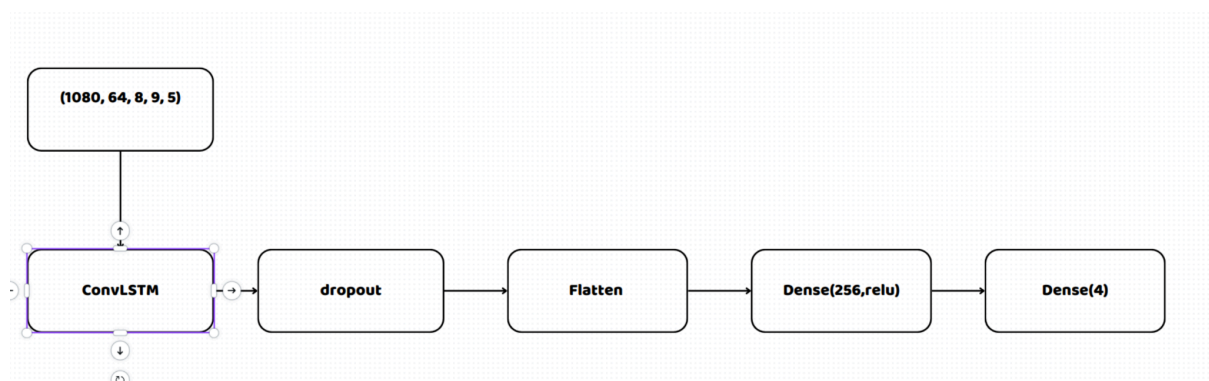
3. Phân loại:

Sau khi tính toán xác suất $g(z)$, mô hình logistic regression sẽ phân loại điểm dữ liệu vào lớp 1 nếu $g(z) \geq \text{threshold}$ và vào lớp 0 nếu $g(z) < \text{threshold}$.

3.2 Các kiến trúc mạng học sâu

a, LSTM

Long Short-Term Memory (LSTM) là một loại mạng nơ-ron hồi tiếp (Recurrent Neural Network - RNN) được thiết kế để xử lý và học từ dữ liệu tuần tự. Khác với RNN thông thường, LSTM vượt qua vấn đề "vanishing gradient" nhờ vào cấu trúc đặc biệt của các đơn vị tế bào (cell), gồm ba cổng: cổng đầu vào (input gate), cổng quên (forget gate), và cổng đầu ra (output gate). Những cổng này cho phép LSTM duy trì thông tin quan trọng trong thời gian dài và loại bỏ thông tin không cần thiết.



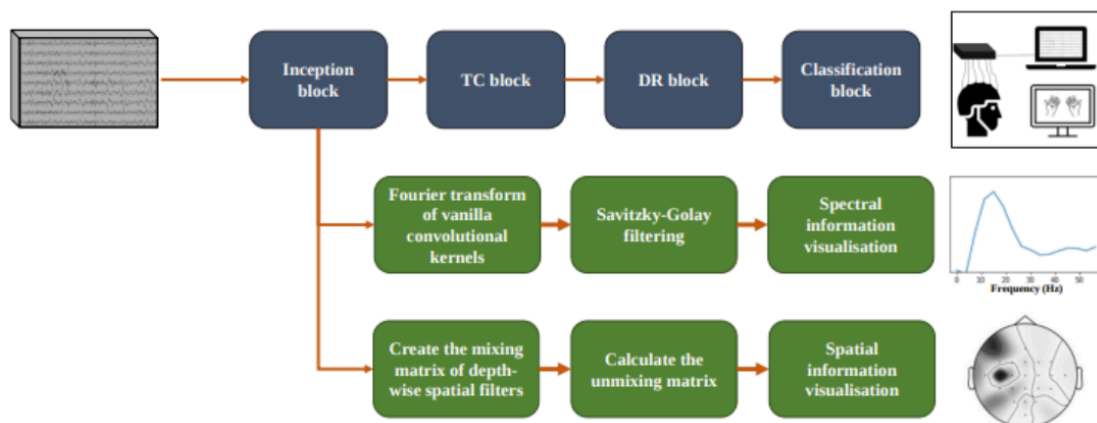
Kiến trúc mạng ConvLSTM2D mà em xây dựng có mục tiêu xử lý dữ liệu dạng tuần tự với sự kết hợp giữa đặc trưng không gian (spatial) và thời gian (temporal). Cách thiết kế này xuất phát từ nhu cầu phân tích những dữ liệu phức tạp như video hoặc chuỗi tín hiệu đa chiều, nơi thông tin không chỉ thay đổi theo thời gian mà còn tồn tại các mối quan hệ không gian.

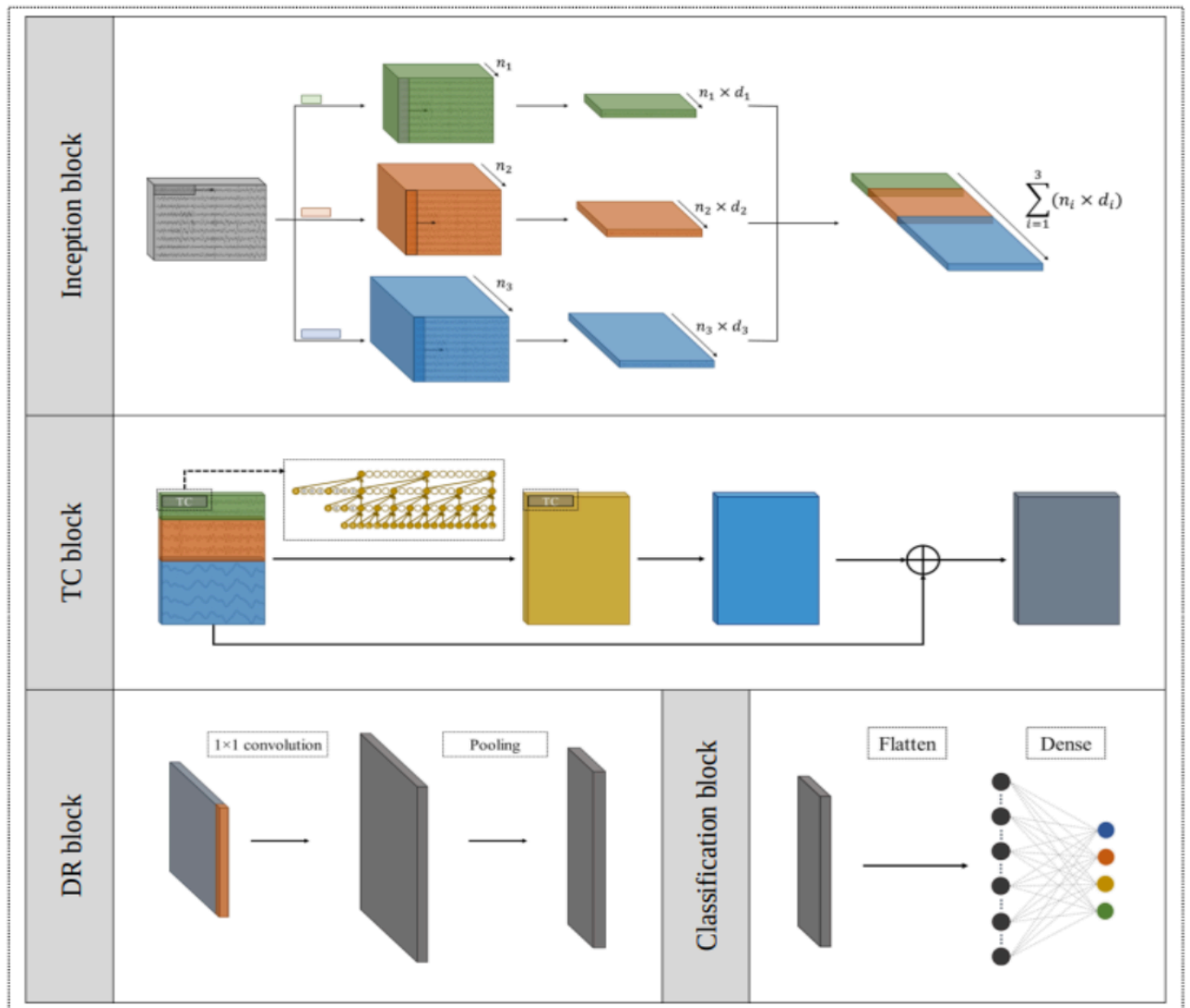
Trong kiến trúc này, lớp ConvLSTM2D đóng vai trò rất quan trọng. Nó vừa có khả năng học được đặc trưng không gian từ từng khung dữ liệu, vừa duy trì và phân tích thông tin tuần tự thông qua cơ chế của LSTM. Đầu vào của mô hình gồm các chuỗi dữ liệu với kích thước (64, 8, 9, 5), tương ứng với 64 bước thời gian, trong đó mỗi bước là một ma trận 8x9 với 5 kênh đặc trưng.

Sau đó, lớp Dropout giúp làm giảm nguy cơ overfitting – một vấn đề thường gặp khi làm việc với dữ liệu phức tạp. Em thêm Dropout để mạng học được những đặc trưng cốt lõi hơn thay vì quá khớp vào dữ liệu huấn luyện. Tiếp theo, em sử dụng lớp Flatten để chuyển đổi đầu ra của lớp ConvLSTM2D thành dạng 1D, tạo điều kiện cho các lớp dense phía sau học thêm các đặc trưng cao cấp hơn.

Cuối cùng, mạng có hai lớp fully connected: lớp đầu với 256 neuron dùng ReLU để nắm bắt các đặc trưng phi tuyến, và lớp cuối với 4 neuron để tạo đầu ra, phù hợp với bài toán em đang hướng đến, chẳng hạn như phân loại hoặc dự đoán. Đây là một cách tiếp cận mạnh mẽ và có tiềm năng cao cho các ứng dụng xử lý dữ liệu phức tạp.

b, EEGITN





Hình mô tả các lớp của EEGITNet.

Cấu trúc các lớp:

1, Inception Block:

- Khối này bắt đầu với ba nhóm lớp song song. Mỗi nhóm gồm một lớp tích chập 2D dọc theo trục thời gian (bộ lọc tần số) và một lớp tích chập chiều sâu 2D (bộ lọc không gian). Việc sử dụng các mô-đun inception với các kích thước kernel khác nhau giúp mạng học được các bộ lọc cho các dải tần số con khác nhau, thay vì chỉ dùng một kích thước kernel cố định.
- Kích thước kernel của lớp tích chập chiều sâu bằng số điện cực trong bộ dữ liệu để kết hợp tất cả các điện cực và tìm ra nguồn gốc hoạt động của não.
- Cuối cùng, khối này kết thúc bằng một hàm kích hoạt phi tuyến và dropout để học các đặc trưng không gian phức tạp và tránh overfitting.

Temporal Convolution (TC) Block:

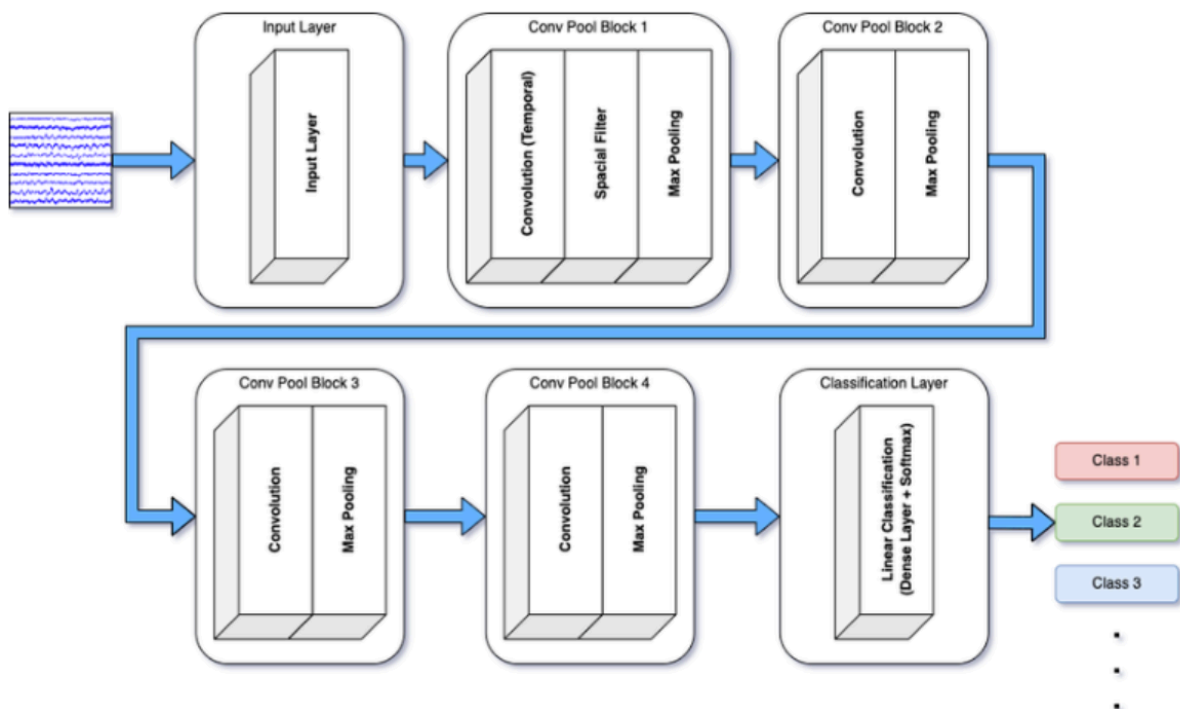
- Sau khi trích xuất các nguồn tín hiệu từ các dải tần số, khối TC được sử dụng để trích xuất các đặc trưng theo thời gian từ tín hiệu EEG.
- Khối này bao gồm các lớp tích chập chiều sâu causal (theo thời gian), giúp duy trì thông tin về thời gian trong chuỗi tín hiệu. Lớp này cũng sử dụng batch normalization thay vì chuẩn hóa trọng số để cải thiện quá trình học.
- Cuối cùng, khối này kết thúc bằng một lớp pooling trung bình để giảm kích thước của tensor.

Classification Block:

- Khối cuối cùng của mô hình, gồm một lớp kết nối đầy đủ và hàm kích hoạt "softmax" sau một lớp flatten.
- Mặc dù đây được gọi là khối phân loại, nó có thể được sửa đổi tùy theo mục tiêu bài toán và loại đầu ra mong muốn.

c, Deep4Net

Deep4Net là một mô hình học sâu khác được thiết kế để trích xuất các đặc trưng phân cấp trong dữ liệu EEG. Mô hình này sử dụng nhiều lớp tích chập để trích xuất các mẫu trừu tượng từ tín hiệu EEG, làm cho nó rất hiệu quả trong việc phân loại các cảm xúc thể hiện sự khác biệt tinh tế trong hoạt động của não. Deep4Net vượt trội trong việc nhận dạng cảm xúc nhờ khả năng xử lý dữ liệu phức tạp, phi tuyến tính..



Hình mô tả kiến trúc của Deep4Net.

Tầng tích chập đầu tiên (First Convolutional Layer):

- Áp dụng tích chập thời gian (temporal convolution) trên các tín hiệu EEG thô.
- Có một lớp kế tiếp thực hiện tích chập không gian (spatial convolution) trên các kênh EEG (channels). Điều này giúp mạng học được các mối quan hệ không gian giữa các điện cực.
- Sử dụng hàm kích hoạt ReLU..

Các tầng tích chập kế tiếp:

- Tầng thứ hai, thứ ba và thứ tư tiếp tục áp dụng các kernel lớn hơn để học đặc trưng ngày càng phức tạp hơn.
- Các đặc trưng học được từ các tầng sau sẽ mang tính phân biệt cao hơn.

Tầng Dropout:

- Sau mỗi tầng tích chập thường có tầng Dropout để giảm overfitting. Đây là một kỹ thuật làm "tắt" ngẫu nhiên một số nơ-ron trong quá trình huấn luyện, giúp mạng học tốt hơn.

Tầng gộp (Pooling Layers):

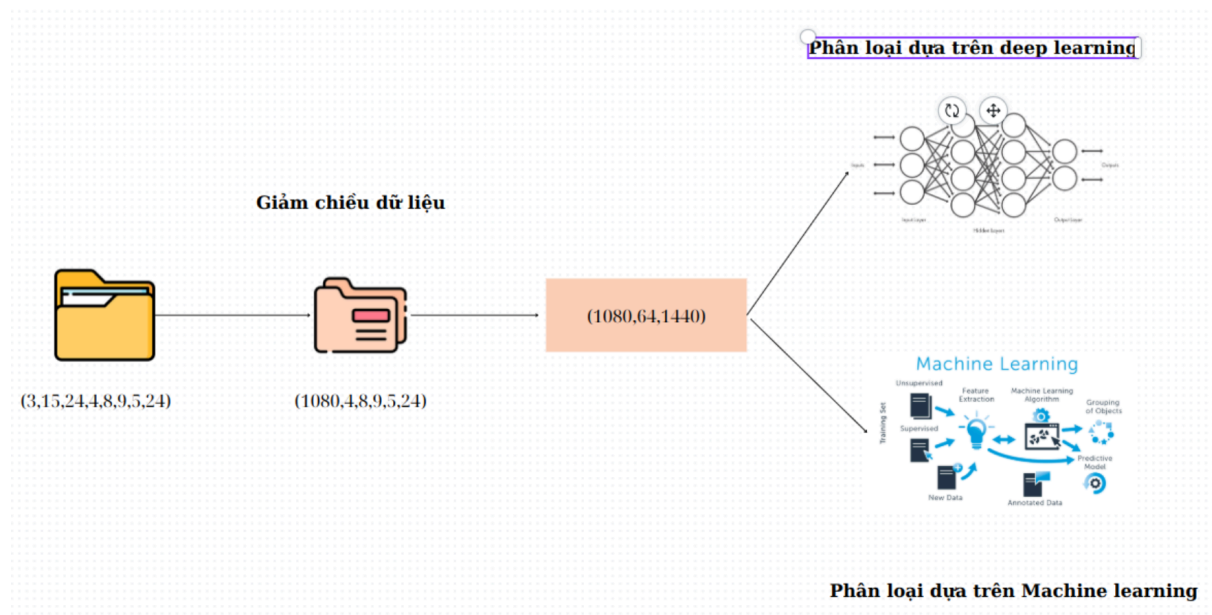
- Tầng gộp (Max Pooling) được sử dụng sau các tầng tích chập nhằm giảm kích thước của đầu ra, đồng thời làm nổi bật các đặc trưng quan trọng.

Tầng Fully Connected (Dense Layer):

- Đầu ra từ các tầng tích chập được làm phẳng và đưa vào một hoặc nhiều lớp fully connected.
- Lớp cuối cùng là lớp Softmax hoặc Sigmoid, đưa ra xác suất cho từng nhãn.

CHƯƠNG 4: GIỚI THIỆU VỀ CẤU TRÚC DỰ ÁN

4.1 Cấu trúc chương trình



Dữ liệu sau khi được giảm chiều sẽ được đưa qua 2 nhánh phân loại là machine learning và deep learning và được đánh giá dựa trên K-fold

CHƯƠNG 5: KẾT QUẢ, THẢO LUẬN VÀ PHƯƠNG HƯỚNG TƯƠNG LAI

5.1 Kết quả thực nghiệm

Thực nghiệm với GPU T4 x 2 trên môi trường máy ảo của kaggle với 2 thuật toán học máy là SVM, Logistic Regression, XGBoost cùng với 3 mạng học sâu tiên tiến là LSTM, Deep4Net, EEGITNet ta được bảng so sánh hiệu suất sau

	Accuracy	Precision	F1-Score
SVM	0.80	0.8016	0.7977
Logistic Regression	0.78	0.7794	0.7773
LSTM	0.70	/	/
EEGITNet	0.56	/	/
Deep4Net	0.57	/	/

Bảng so sánh hiệu suất các thuật toán sử dụng.

Nhận xét:

SVM đạt kết quả tốt nhất với 80% và tiếp sau đó là thuật toán Logistic Regression là 78%. Trong thực nghiệm này ta thấy được các thuật toán học sâu tiến bộ thực sự đạt kết quả kém hơn rất nhiều so với các thuật toán học máy lần lượt LSTM là 70%, EEGITNet là 56% và Deep4Net là 57%

5.2 Thảo luận

Trong thực nghiệm trên thì các thuật toán học máy đã tỏ ra vượt trội hơn các thuật toán học sâu tiên tiến. Với bộ dữ liệu chỉ có kích thước 1080 sample thì không khó để ta khẳng định độ ưu việt của các thuật toán học máy cũng như là sự tối ưu của các thuật toán deep learning đặc biệt là các mạng học sâu tiên tiến như EEGITNet hay Deep4Net. Nhìn chung với một kích thước dữ liệu tương đối nhỏ, khi tôi thực nghiệm với LSTM nông thì đạt kết quả tốt hơn khi thực nghiệm với LSTM sâu có thể cho thấy rằng khi dữ liệu hạn chế về số lượng, các mô hình học sâu với kiến trúc phức tạp như LSTM sâu dễ dẫn đến hiện tượng overfitting, khiến kết quả không tốt bằng các mô hình đơn giản hơn. Điều này là do các mạng sâu yêu cầu lượng dữ liệu lớn để học được các đặc trưng phức tạp, trong khi với dữ liệu nhỏ, chúng có xu hướng ghi nhớ dữ liệu huấn luyện thay vì tổng quát hóa. Ngược lại, các mô hình học máy và LSTM nông thường có ít tham số hơn, giảm nguy cơ overfitting và phù hợp hơn cho các bộ dữ liệu nhỏ, nhờ khả năng học hiệu quả từ thông tin sẵn có.

5.3 Phương hướng tương lai

Trong tương lai, việc phát triển mô hình có thể tập trung vào một số hướng chính để nâng cao hiệu suất và khả năng ứng dụng. Đầu tiên, tối ưu hóa kiến trúc mạng là một ưu tiên

quan trọng. Các nghiên cứu có thể thử nghiệm với các mô hình phức tạp hơn như hybrid models, kết hợp ConvLSTM2D với các mô hình khác như Transformer hoặc Attention Mechanisms, để cải thiện khả năng học đặc trưng không gian-thời gian.

Thứ hai, mở rộng và đa dạng hóa tập dữ liệu cũng là một hướng đi cần thiết. Bổ sung các dữ liệu thực tế từ nhiều nguồn khác nhau sẽ giúp mô hình tổng quát hóa tốt hơn và ứng dụng được vào các bài toán đa dạng như y tế, dự đoán môi trường, hoặc phân tích hành vi người dùng.

Thứ ba, các phương pháp tối ưu hóa như kỹ thuật tăng cường dữ liệu (data augmentation), điều chỉnh siêu tham số (hyperparameter tuning), và tích hợp các chiến lược chống overfitting như regularization cũng cần được nghiên cứu thêm.

Cuối cùng, việc triển khai mô hình vào thực tế trên các nền tảng edge computing hoặc cloud computing có thể giúp cải thiện hiệu quả tính toán và mở rộng ứng dụng của mô hình ra quy mô lớn. Những hướng này hứa hẹn sẽ mở ra nhiều tiềm năng ứng dụng mới và đóng góp đáng kể cho lĩnh vực học sâu.

Tài liệu tham khảo

- [1] A. B. Schwartz, X. T. Cui, D. Weber, and D. W. Moran, “Brain-controlled interfaces: Movement restoration with neural prosthetics,” *Neuron*, vol. 52, no. 1, pp. 205 – 220, 2006.
- [2] Salami, J. Andreu-Perez and H. Gillmeister, “EEG-ITNet: An Explainable Inception Temporal Convolutional Network for motor imagery classification,” in *IEEE Access*, doi: 10.1109/ACCESS.2022.3161489.
- [3] Schirrmester, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggersperger, K., Tangermann, M., Hutter, F. & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* , Aug. 2017. Online: <http://dx.doi.org/10.1002/hbm.23730>