



Bài báo nghiên cứu

MÔ HÌNH ECL-tEEG ƯỚC LƯỢNG CẢM XÚC CỦA NGƯỜI HỌC DỰA TRÊN SÓNG NÃO

Phan Thị Quý Thịnh, Phùng Ngọc Nhân, Mai Yến Khoa,

Trần Thanh Nhã, Võ Lê Phúc Hậu, Nguyễn Việt Hưng*

Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam

**Tác giả liên hệ: Võ Lê Phúc Hậu – Email: tg.hauvlp@lecturer.hcmue.edu.vn*

Ngày nhận bài: 09-4-2025; Ngày nhận bài sửa: 21-4-2025; Ngày duyệt đăng: 24-4-2025

TÓM TẮT

Nhận diện cảm xúc và đánh giá mức độ tham gia của người học đóng vai trò quan trọng trong việc nâng cao hiệu quả giảng dạy và tối ưu hóa phương pháp giảng dạy cá nhân hóa. Các nghiên cứu trước đây chủ yếu sử dụng video hoặc hình ảnh kết hợp với mô hình học sâu để phân tích hành vi và cảm xúc. Tuy nhiên, phương pháp này dễ bị ảnh hưởng bởi góc quay, điều kiện ánh sáng và biểu cảm khuôn mặt không rõ ràng, dẫn đến sai lệch trong đánh giá cảm xúc. Để khắc phục những hạn chế đó, nghiên cứu này đề xuất mô hình ECL-tEEG, một mô hình kết hợp CNN, LSTM-GRU và Transformer nhằm phân tích tín hiệu điện não (EEG) để nhận diện cảm xúc và đánh giá mức độ tham gia của người học. Mô hình đề xuất tận dụng CNN để trích xuất đặc trưng không gian, LSTM-GRU để học đặc trưng tuần tự theo thời gian và Transformer để khai thác các mối quan hệ dài hạn theo cơ chế học song song. Để đánh giá hiệu suất, nghiên cứu tiến hành thử nghiệm trên ba mô hình: CNN-LSTM-GRU, CNN-Transformer và ECL-tEEG. Kết quả thực nghiệm cho thấy ECL-tEEG đạt độ chính xác 71%, vượt trội hơn hai mô hình còn lại trong phân loại cảm xúc tích cực, trung tính và tiêu cực. Nghiên cứu này không chỉ khẳng định tiềm năng ứng dụng của EEG trong nhận diện cảm xúc người học, mà còn đặt nền tảng cho hệ thống hỗ trợ giảng dạy thông minh, giúp cá nhân hóa trải nghiệm học tập và hỗ trợ giáo viên trong việc điều chỉnh phương pháp giảng dạy dựa trên tín hiệu não bộ thay vì chỉ dựa vào quan sát hành vi.

Từ khóa: nhận dạng cảm xúc; điện não đồ (EEG); giáo dục; học máy; CNN; LSTM-GRU; Transformer

1. Giới thiệu

Trong những năm gần đây, nhận diện cảm xúc là vấn đề được nhiều ngành nghề, nhiều nhà khoa học quan tâm. Phần lớn các mô hình trích xuất đặc trưng tín hiệu cảm xúc của người học sử dụng biểu cảm khuôn mặt, giọng nói, cử chỉ hành vi cơ thể và các tín hiệu sinh

Cite this article as: Phan, T. Q., Phung, N. N., Mai, Y. K., Tran, T. N., Vo, L. P. H., & Nguyen, V. H. (2026). ECL-tEEG: A hybrid deep learning model for estimating learners' emotions from EEG signals. *Ho Chi Minh City University of Education Journal of Science*, 23(6), 1467-1478. [https://doi.org/10.54607/hcmue.js.23.6.4873\(2026\)](https://doi.org/10.54607/hcmue.js.23.6.4873(2026))

lí. Biểu cảm khuôn mặt và giọng nói được sử dụng rộng rãi trong việc đánh giá cảm xúc của người học (Florestiyanto, 2024). Tuy nhiên, vào những hoàn cảnh đặc biệt, những cảm xúc này có thể không được thể hiện chính xác nhằm che giấu cảm xúc thật của chính bản thân mình. Do đó, biểu cảm gương mặt và giọng nói chưa hẳn đã diễn tả chân thực và chính xác nhất cảm xúc của con người (Song et al., 2019). Bên cạnh đó, các tín hiệu từ sinh lí của con người được các nhà nghiên cứu đánh giá cao về độ tin cậy và cho rằng chúng khó có thể ngụy tạo như các biểu cảm khác (Yu & Wang, 2022). Một số tín hiệu sinh lí có thể nhắc đến như điện não đồ, điện cơ đồ, điện tim đồ, huyết áp, nhiệt độ cơ thể... (Hari Krishna et al., 2024).

Cảm xúc của người học trong lĩnh vực giáo dục là yếu tố then chốt để đảm bảo chất lượng giảng dạy, góp phần xây dựng một môi trường giáo dục lành mạnh, hiện đại, lấy trải nghiệm cảm xúc trong quá trình tiếp thu tri thức của từng học sinh làm trọng tâm. Trong đó, ước lượng cảm xúc của người học dựa trên sóng não hứa hẹn mở ra một hướng đi mới, mang đến một giải pháp đóng góp hỗ trợ người dạy trong môi trường giáo dục hiện đại. Sự tham gia về mặt cảm xúc lẫn nhận thức trong giờ học nếu như được nhà giáo nắm bắt theo thời gian thực sẽ giúp tối ưu hóa, cũng như kịp thời điều chỉnh chương trình dạy học phù hợp với tình hình tham gia của lớp học (Apicella et al., 2022). Do đó, việc xây dựng một mô hình giúp phỏng đoán cảm xúc của người học, hỗ trợ nhà giáo trong thích ứng với sự chuyển đổi giữa các giai đoạn giáo dục là thật sự cần thiết.

Hiện nay, các mô hình học sâu, đặc biệt là mạng nơ-ron tích chập (CNN) kết hợp với bộ nhớ dài – ngắn hạn (LSTM) – mạng nơ-ron hồi tiếp với nút có cổng (GRU), đã chứng minh hiệu quả vượt trội trong việc kết hợp các loại thông tin. Sử dụng cơ chế self-attention của phương pháp bộ chuyển đổi (Transformer) giúp nắm bắt được toàn bộ thông tin của đoạn tín hiệu, mang lại những bước tiến lớn trong việc nhận diện cảm xúc và trải nghiệm của người học dựa trên sóng não. Sự kết hợp này không chỉ giúp khai thác tối đa các đặc điểm tín hiệu quan trọng, mà còn tối ưu hóa tài nguyên, góp phần phát triển các ứng dụng trong lĩnh vực giáo dục thông minh. Trong nghiên cứu này, chúng tôi đề xuất mô hình ECL-tEEG kết hợp giữa CNN, LSTM-GRU và Transformer, nhằm tối ưu hóa trích xuất đặc trưng không gian, thời gian và mối quan hệ dài hạn từ tín hiệu EEG. Đánh giá hiệu suất của mô hình đề xuất thông qua thực nghiệm trên ba mô hình khác nhau: CNN-LSTM-GRU, CNN-Transformer và ECL-tEEG, nhằm xác định mức độ cải thiện trong nhận diện cảm xúc người học.

2. Nội dung nghiên cứu

2.1. Công trình nghiên cứu liên quan

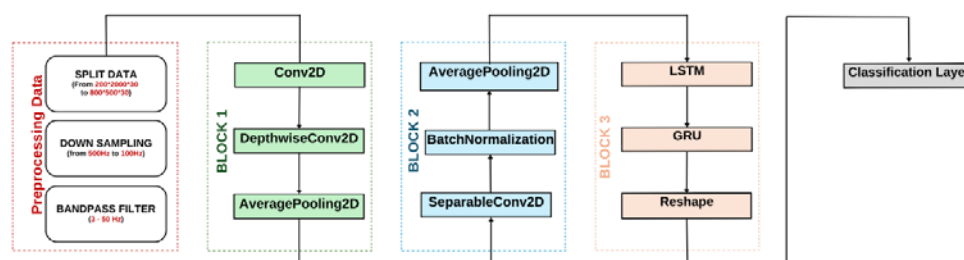
Các phương pháp truyền thống trong việc nhận diện cảm xúc thường dựa vào phân tích biểu cảm khuôn mặt, giọng nói và cử chỉ của con người. Tuy nhiên, những phương pháp này gặp phải nhiều hạn chế, như dễ bị tác động, không phản ánh chính xác trạng thái nội tâm và dễ chịu ảnh hưởng bởi yếu tố chủ quan (Taware & Thakare, 2023). Trong khi đó, với tín hiệu điện não (EEG) – phương pháp ghi nhận trực tiếp hoạt động của não bộ, mang lại một cách tiếp cận khách quan hơn trong việc đo lường và nhận diện cảm xúc. Một hướng nghiên

cứ đáng chú ý là phân tích các dữ liệu sóng não bằng các mô hình học sâu (Deep Learning) nhằm đề ra một số phương pháp kết hợp các mô hình tăng hiệu quả chính xác trong thực tiễn (Mahmoud et al., 2023). Theo nghiên cứu của Chaofei Yu và Mei Wang, họ cho rằng nhận diện cảm xúc dựa trên tín hiệu EEG trải qua nhiều bước quan trọng: kích thích cảm xúc (emotion induction), tiền xử lý tín hiệu EEG, trích xuất đặc trưng và phân loại cảm xúc (Yu & Wang, 2022). Ngoài ra, mô hình kết hợp các lớp tích chập để học các đặc trưng cục bộ (theo thời gian và không gian) và sau đó áp dụng cơ chế tự chú ý (self-attention) để nắm bắt các mối quan hệ dài hạn trong dữ liệu EEG (Song et al., 2023). Trong giai đoạn phân loại cảm xúc, các mô hình học sâu như mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi quy (RNN) và bộ nhớ ngắn - dài hạn (LSTM) đã được áp dụng để cải thiện độ chính xác. Trong các nghiên cứu gần đây của Zhu và các cộng sự, đã thử nghiệm kết hợp CNN và LSTM trong mô hình nhận diện cảm xúc dựa trên sóng não, đã đạt được độ chính xác lên đến 95,82% trên bộ dữ liệu DEAP (Zhu et al., 2024).

Trong nghiên cứu của Mahmoud và các cộng sự (2023) đã sử dụng lớp Convolutional Layer để phát hiện các đặc trưng cơ bản của tín hiệu EEG như các mẫu tần số hoặc các đặc trưng thời gian của tín hiệu, lớp Pooling để giảm kích thước của các đặc trưng sau lớp tích chập, giúp giảm số lượng tham số và giúp mô hình học các đặc trưng tổng quát hơn, nghiên cứu này sử dụng bộ dữ liệu SEED và DEAP để so sánh hiệu quả của mô hình CNN với các phương pháp khác, chẳng hạn như các mô hình học máy truyền thống như SVM (Support Vector Machine) hoặc Random Forest, nhằm chứng minh tính hiệu quả của CNN trong việc nhận diện cảm xúc từ tín hiệu EEG. Tiếp đến về mô hình LSTM – GRU, Joshi và các cộng sự đã sử dụng BiLSTM và MD-DE để phân tích đặc trưng của chuỗi dữ liệu EEG nhằm nắm bắt được đặc trưng không gian của các vùng khác nhau trên não bộ, đồng thời kết hợp cùng chuỗi dữ liệu thời gian dài hạn, nhưng bộ dữ liệu IDEA chưa được sử dụng rộng rãi (Joshi & Ghongade, 2022). Những mô hình này không chỉ giúp tăng cường độ chính xác mà còn cải thiện khả năng phát hiện cảm xúc nhanh chóng của con người, đóng góp phần lớn vào lĩnh vực giáo dục và y khoa.

2.2. Mô hình ước lượng cảm xúc dựa trên sóng não

2.2.1. Mô hình CNN-LSTM-GRU



Hình 1. Kiến trúc mô hình CNN-LSTM-GRU

Kiến trúc mô hình CNN-LSTM-GRU thể hiện trong Hình 1 được đề xuất để xử lý tín hiệu điện não thông qua việc sử dụng các lớp tích chập CNN kết hợp cùng các loại mạng

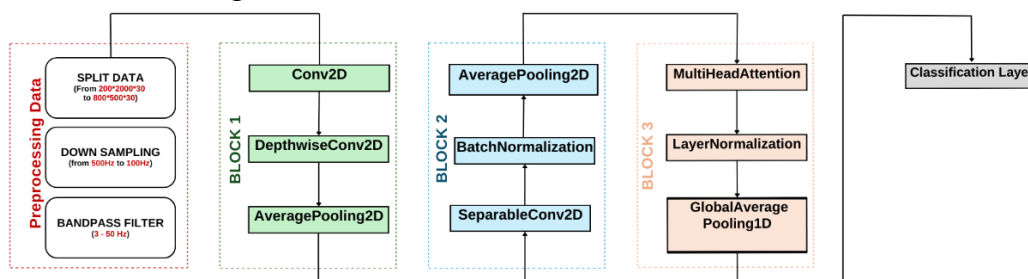
neuron hồi quy LSTM và GRU. CNN hiệu quả trong trích xuất các đặc trưng về không gian, LSTM và GRU giúp mô hình học tuần tự các đặc trưng về thời gian, cách kết hợp trên giúp mô hình tận dụng được tối đa tài nguyên về không gian và thời gian của tín hiệu sóng não. Sau khi dữ liệu trải qua các giai đoạn tiền xử lý như lọc nhiễu, giảm tốc độ lấy mẫu và chia lại đoạn dữ liệu, dữ liệu đầu vào có dạng 3D Tensor với kích thước (800x500x30), dữ liệu được đưa vào Block 1 là giai đoạn để trích xuất các đặc trưng không gian của EEG bằng cách sử dụng các lớp tích chập được thể hiện trong block 1 như: 1 lớp Convolutional 2D (Conv2D) có 8 lớp filters với kích thước kernel là (1x100), 1 lớp Depthwise Convolutional 2D (DepthwiseConv2D) có 8 lớp depth multiplier với kích thước (30x1) đan xen với các lớp Batch Normalization nhằm giúp mô hình ổn định và cải thiện tốc độ học. Sau đó, qua lớp Activation và lớp Average Pooling 2D với kích thước 1x4. Để tránh hiện tượng overfitting, chúng tôi thêm vào một lớp Dropout có tỉ lệ 0.5 để tránh sự phụ thuộc của mô hình vào một số đặc trưng nhất định. Sau block 1, mô hình học được mối quan hệ giữa các vùng não khác nhau tại từng thời điểm, đây là đặc trưng không gian của tín hiệu EEG. Mục tiêu của block 2 là trích xuất được các đặc trưng EEG tiếp nối sau block 1. Kiến trúc của Block 2 gồm các lớp Separable Convolutional 2D (SeparableConv2D), Batch Normalization, Activation, Average Pooling 2D và Dropout. Mô hình trích xuất các đặc trưng bằng SeparableConv2D với 64 filters, kernel (1x16) nhằm giảm số lượng tham số nhưng vẫn giữ nguyên các đặc trưng cần thiết của tín hiệu. Dữ liệu tiếp tục thông qua các lớp Activation có hàm kích hoạt Elu và AveragePooling2D. Sau khi qua Block 2, mô hình đã học được các đặc trưng hiệu quả giúp tăng tính ổn định của mô hình. Dữ liệu tiếp tục đi đến Block 3, đây là nơi để mô hình học đặc trưng theo thời gian. Tiếp theo, các đặc trưng từ LSTM được chuyển qua lớp GRU nhằm tối ưu hóa quá trình học chuỗi thời gian. Cả lớp LSTM và GRU trên đều sử dụng 64 units. GRU hoạt động tương tự như LSTM nhưng với cấu trúc đơn giản hơn, giúp giảm bớt độ phức tạp tính toán và tăng tốc độ huấn luyện. Cuối cùng là lớp phân loại đầu ra Classification với 1 lớp Fully Connected Layer với 3 units đi kèm với hàm kích hoạt Softmax để phân loại cảm xúc thành ba nhóm chính: tích cực, tiêu cực và trung tính.

2.2.2. Mô hình CNN-Transformer

Kiến trúc mô hình CNN-Transformer được mô tả trong Hình 2. Mô hình sử dụng các lớp tích chập trong block 1 và 2 tương tự như phương pháp CNN-LSTM-GRU Hình 1. Tuy nhiên, thay vì sử dụng LSTM-GRU để xử lý mối quan hệ theo thời gian một cách tuần tự, chúng tôi sử dụng Transformer để học song song các đặc trưng và khai thác mối quan hệ giữa các phần của tín hiệu EEG một cách hiệu quả hơn khi việc học dữ liệu dựa trên quan hệ toàn cục.

Tương tự mô hình CNN-LSTM-GRU, dữ liệu sau khi qua block 1 và block 2, mô hình sẽ trích xuất đặc trưng không gian ban đầu và học được các đặc trưng hiệu quả. Dữ liệu sẽ đi vào block 3 để học các mối quan hệ toàn cục bằng Attention bằng việc tận dụng cơ chế Multi-Head Self-Attention để mô hình có thể học đa đặc trưng của EEG đồng thời tối ưu hóa

hiệu suất. Multi-Head Self-Attention giúp mô hình nắm bắt được sự tương tác giữa các vùng não xuyên suốt thời gian phù hợp cho nắm bắt mối liên hệ phức tạp, không tuần tự. Một MultiHeadAttention layer với 4 đầu attention (multi-heads) và key dimension = 16 nhằm làm nổi bật các mối quan hệ phụ thuộc dài hạn trong chuỗi tín hiệu. Lớp Multi-Head Self-Attention có Layer Normalization giúp chuẩn hóa dữ liệu, trong khi GlobalAveragePooling1D tổng hợp tất cả đặc trưng quan trọng dưới dạng một vector duy nhất. Cuối cùng là lớp phân loại đầu ra Classification với 1 lớp Fully Connected Layer với 3 units đi kèm với hàm kích hoạt Softmax để phân loại cảm xúc thành ba nhóm chính: tích cực, tiêu cực và trung tính.



Hình 2. Kiến trúc mô hình CNN-Transformer

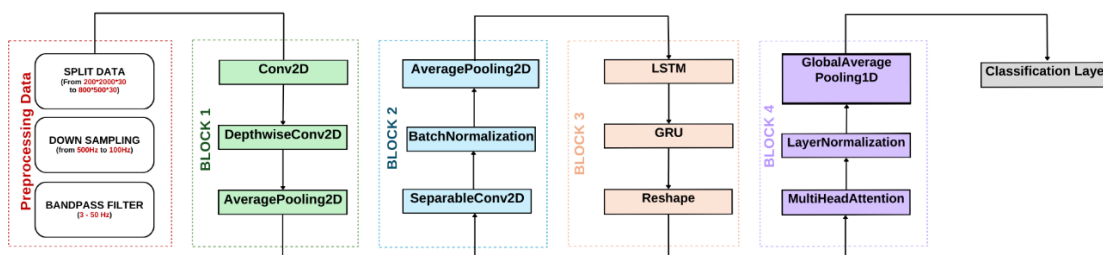
2.2.3. ECL-tEEG

Chúng tôi đề xuất mô hình ECL-tEEG, kết hợp CNN, LSTM-GRU và Transformer, có kiến trúc được mô tả trong Hình 3. Nhận thấy ưu điểm vượt trội của từng mô hình riêng lẻ, chúng tôi kết hợp cả ba phương pháp để giải quyết các vấn đề trích xuất đặc trưng không gian, thời gian tuần tự và mối quan hệ dài hạn của tín hiệu EEG. Mục tiêu của mô hình là học được đặc trưng không gian từ CNN, khai thác đặc trưng thời gian theo hai hướng tuần tự (LSTM-GRU) và song song (Transformer), giúp tối ưu hóa khả năng phân tích tín hiệu EEG.

Do tín hiệu EEG là một chuỗi tín hiệu thời gian, việc nắm bắt đặc trưng dài hạn là rất quan trọng. Mạng nơ-ron hồi quy LSTM-GRU có khả năng ghi nhớ thông tin theo trình tự, giúp mô hình phát hiện sự thay đổi về biên độ sóng và học đặc trưng thời gian hiệu quả. Tuy nhiên, khi chuỗi tín hiệu quá dài, LSTM-GRU gặp khó khăn trong việc ghi nhớ thông tin xa do hiện tượng suy giảm gradient (vanishing gradient problem). Do đó, trước khi sử dụng LSTM-GRU, việc sử dụng CNN nhằm trích xuất được các đặc trưng quan trọng đồng thời giảm độ phức tạp của mô hình là hết sức cần thiết. Đồng thời, cơ chế Multi-Head Self-Attention với nhiều head, mỗi head có khả năng học một khía cạnh đặc trưng khác nhau của tín hiệu, từ đó tăng tính đa dạng và chiều sâu biểu diễn. Điều này giúp mô hình hiểu rõ hơn về các biểu hiện cảm xúc phức tạp, vốn có thể phân bố không đồng đều theo thời gian và không gian.

Theo nghiên cứu của (Andayani et al., 2022) đặc trưng của mạng nơ-ron hồi quy chính là việc xử lý tuần tự dữ liệu nhưng phương pháp này lại gặp trở ngại trong việc xử lý những dữ liệu có thời gian quá dài. Tín hiệu sóng não là dạng tín hiệu thay đổi theo thời gian. Do đó, việc nắm bắt được các đặc trưng theo thứ tự tuần tự giúp mô hình nắm bắt tốt được các

thay đổi trong tín hiệu. Tuy nhiên, số lượng đặc trưng quá lớn có thể khiến LSTM-GRU gặp khó khăn trong việc ghi nhớ và liên kết thông tin trong khoảng thời gian dài. Để khắc phục điều này, Transformer với cơ chế Multi-Head Attention cho phép mô hình tập trung vào các đặc trưng quan trọng, giúp kết nối thông tin và khai thác triệt để các mối quan hệ gần, xa của tín hiệu. Sự kết hợp cả ba phương pháp trên giúp khai thác triệt để tín hiệu theo các hướng không gian, thời gian, tuần tự và song song.



Hình 3. Kiến trúc mô hình ECL – tEEG

2.2.4. Bộ dữ liệu

Trong nghiên cứu này, chúng tôi sử dụng dữ liệu về tín hiệu sóng não trong bộ dữ liệu đa phương thức EAV gồm EEG, Audio và Visual trong một nghiên cứu về nhận dạng cảm xúc dựa trên bối cảnh đàm thoại (Lee et al., 2024). Mặc dù các bộ dữ liệu như DEAP và SEED hiện đang được sử dụng phổ biến trong nghiên cứu nhận diện cảm xúc, nhưng bối cảnh thu thập dữ liệu tập trung vào việc xem video phim hoặc nghe nhạc. Do đó, các bộ dữ liệu trên chưa thực sự phù hợp với môi trường giáo dục. Chính vì vậy, trong nghiên cứu này, chúng tôi lựa chọn sử dụng bộ dữ liệu đa phương thức EAV, được xây dựng dựa trên các tình huống đàm thoại mang tính giáo dục – nhằm đảm bảo tính thực tiễn và phù hợp hơn với mục tiêu ứng dụng trong môi trường giáo dục hiện nay. Bộ dữ liệu được xây dựng nhằm xác định 5 cảm xúc của con người gồm vui vẻ, bình tĩnh, trung tính, tức giận, buồn bã trong bối cảnh đàm thoại với sự tham gia hỗ trợ của 42 tình nguyện viên với độ tuổi khoảng 20-30 tuổi. Dữ liệu về sóng não được thu tại 30 kênh sau: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, PO10. Dữ liệu đầu vào có dạng 3D Tensor với kích thước (200x10000x30) ứng với (số mẫu x (thời gian lấy mẫu x tốc độ lấy mẫu) x số kênh). Mỗi đoạn dữ liệu được thu trong 20s với tốc độ lấy mẫu 500Hz. Sau khi trải qua giai đoạn tiền xử lý gồm các bước như lọc nhiễu, giảm tốc độ lấy mẫu từ 500Hz xuống 100Hz nhưng vẫn đảm bảo giữ nguyên được các đặc trưng của tín hiệu và cuối cùng thực hiện chia lại kiến trúc bộ dữ liệu. Cách thức chia lại bộ dữ liệu như sau: một mẫu dữ liệu 20s được chia làm 4 với độ dài mỗi mẫu là 5s. Do đó, số lượng mẫu dữ liệu tăng lên gấp 4 lần, từ (200x2000x30) sang (800x500x30). Chúng tôi thực hiện thay đổi nhãn dữ liệu sao cho phù hợp với mục tiêu nghiên cứu như sau: tích cực (vui vẻ, bình tĩnh), trung tính (trung tính) và tiêu cực (tức giận, buồn bã). Dữ liệu được chia theo tỉ lệ 70% dùng cho huấn luyện và 30% dành cho thực nghiệm.

Bảng 1. Số mẫu dữ liệu

	Huấn luyện	Thực nghiệm	Tổng
Tích cực	1411	605	2016
Trung tính	705	303	1008
Tiêu cực	1411	605	2016

2.2.5. Phương pháp đánh giá

Dựa trên các nghiên cứu trước đó của (Hsu et al., 2020; Jaworska et al., 2019) trong việc đánh giá hiệu suất của mô hình có liên quan đến dữ liệu EEG, chúng tôi sẽ đánh giá mô hình qua hai độ đo là Accuracy và F1 Score. Các số liệu này được tính toán bằng các công thức (1), (4); với TP (True Positive) biểu thị cho số lượng trường hợp mô hình dự đoán chính xác cảm xúc tích cực hoặc trung lập hoặc tiêu cực, FP (False Positive) biểu thị cho số lượng trường hợp dự đoán không chính xác cảm xúc tích cực hoặc trung lập hoặc tiêu cực, TN (True Negative) biểu thị cho số lượng trường hợp mô hình phân loại chính xác một mẫu không thuộc về một nhãn cảm xúc tích cực hoặc trung lập hoặc tiêu cực, FN (False Negative) biểu thị cho số lượng trường hợp phân loại sai cảm xúc tích cực hoặc trung lập hoặc tiêu cực

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Recall + Precision} \quad (4)$$

3. Kết quả và thảo luận

Chúng tôi thực hiện so sánh 3 mô hình trên với mô hình trong nghiên cứu của Lee cùng các cộng sự (2024) và một số phương pháp truyền thống như SVM và Random Forest hay phương pháp sử dụng BiLSTM trong Bảng 2. Tất cả mô hình đều sử dụng chung các tham số huấn luyện như sau: batch_size = 32, số epoch = 100, optimizer: Adam, learning rate: 0.001, hàm mất mát: categorical_crossentropy.

Bảng 2. Bảng so sánh hiệu suất các mô hình

STT	Nghiên cứu	Phương pháp	Accuracy	F1-Score	Precision	Recall
1	(Lee et al., 2024)	CNN	0.70	0.68	0.68	0.68
2		SVM	0.45	0.42	0.47	0.45
3		Random Forest	0.47	0.42	0.52	0.57
4		BiLSTM	0.43	0.43	0.42	0.43
5	Mô hình đề xuất	CNN-LSTM-GRU	0.69	0.68	0.68	0.68
6	Mô hình đề xuất	CNN-Transformer	0.68	0.66	0.67	0.65
7	Mô hình đề xuất	ECL-tEEG	0.71	0.69	0.70	0.69

Kết quả thực nghiệm của các mô hình được thể hiện trong Hình 4. Để đánh giá chi tiết hơn về hiệu suất của các mô hình, chúng tôi tiến hành so sánh hiệu suất của các mô hình trên cùng bộ dữ liệu. Các so sánh về các thông số đánh giá được trình bày trong các Bảng 3 và so sánh về độ chính xác của các mô hình được trực quan hóa trong Hình 5.

Bảng 3. Bảng đánh giá hiệu suất giữa các lớp

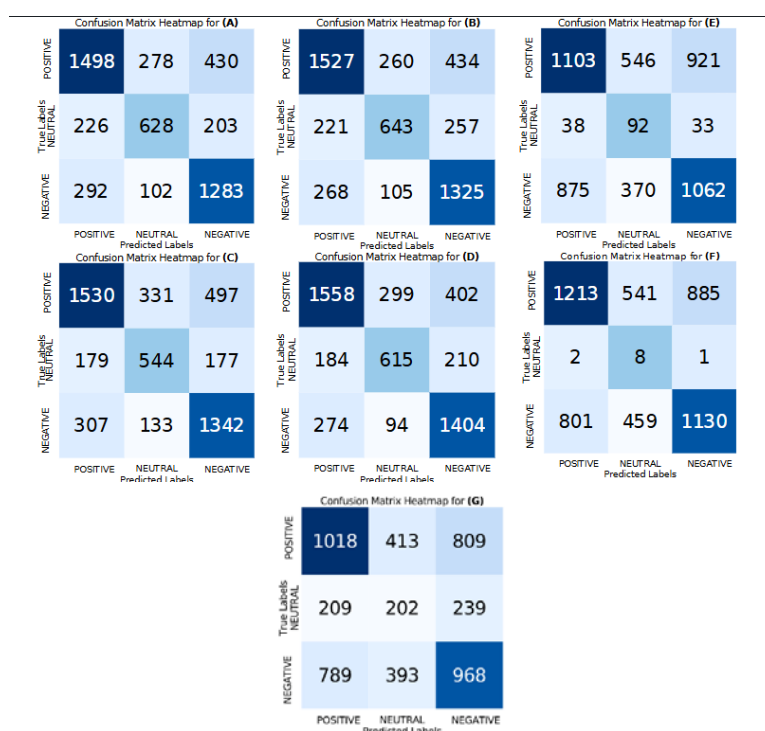
Mô hình	Cảm xúc	Presicion	Recall	F1-score	Accuracy
(Lee et al., 2024)	TÍCH CỰC	0.68	0.74	0.71	0.74
	TRUNG TÍNH	0.59	0.62	0.61	0.62
	TIÊU CỰC	0.78	0.69	0.73	0.69
SVM	TÍCH CỰC	0.43	0.55	0.48	0.55
	TRUNG TÍNH	0.56	0.09	0.16	0.09
	TIÊU CỰC	0.46	0.53	0.49	0.53
RANDOM FOREST	TÍCH CỰC	0.46	0.60	0.52	0.60
	TRUNG TÍNH	0.73	0.01	0.03	0.07
	TIÊU CỰC	0.47	0.56	0.51	0.56
BiLSTM	TÍCH CỰC	0.45	0.50	0.48	0.50
	TRUNG TÍNH	0.31	0.20	0.24	0.20
	TIÊU CỰC	0.45	0.48	0.46	0.48
CNN-LSTM-GRU	TÍCH CỰC	0.60	0.67	0.64	0.67
	TRUNG TÍNH	0.53	0.44	0.48	0.44
	TIÊU CỰC	0.65	0.63	0.64	0.63
CNN-Transformer	TÍCH CỰC	0.64	0.74	0.69	0.74
	TRUNG TÍNH	0.59	0.51	0.55	0.51
	TIÊU CỰC	0.74	0.68	0.71	0.68
ECL-tEEG	TÍCH CỰC	0.69	0.77	0.73	0.77
	TRUNG TÍNH	0.61	0.61	0.61	0.61
	TIÊU CỰC	0.79	0.70	0.74	0.70

Bảng 4. So sánh độ chính xác của các mô hình

Mô hình	Độ chính xác
(Lee et al., 2024)	0.70
SVM	0.45
RANDOM FOREST	0.47
BiLSTM	0.43
CNN-LSTM-GRU	0.69
CNN-TRANSFORMER	0.68
ECL-tEEG	0.71

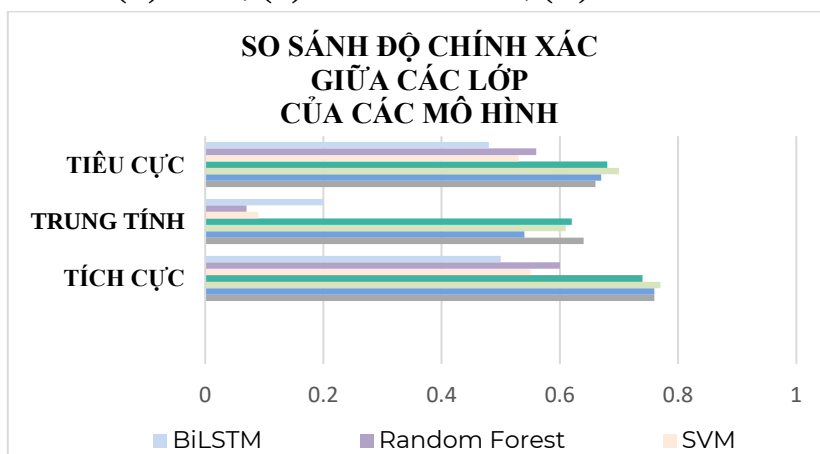
Bảng 5. So sánh độ phức tạp của các mô hình

STT	Nghiên cứu	Phương pháp	Tổng tham số	Thời gian huấn luyện / epoch (ms)
1	(Lee et al., 2024)	CNN	8,579	36
2		SVM	42,004	-
3		Random Forest	30,004	-
4		BiLSTM	179,459	64
5	Mô hình đề xuất	CNN-LSTM-GRU	58,371	36
6	Mô hình đề xuất	CNN-Transformer	199,683	48
7	Mô hình đề xuất	ECL-tEEG	75,139	39



Hình 4. Ma trận nhầm lẫn của các mô hình

(A) (Lee et al., 2024), (B) CNN-LSTM-GRU, (C) CNN-Transformer, (D) ECL-tEEG
(E) SVM, (F) Random Forest, (G) BiLSTM



Hình 5. So sánh độ chính xác giữa các lớp của các mô hình

Nhìn chung, mô hình ECL-tEEG đạt được độ ổn định cao hơn so với các mô hình còn lại. Trong quá trình huấn luyện và kiểm thử, chúng tôi nhận thấy mô hình đạt được kết quả chính xác vượt trội trong việc phân loại cảm xúc tích cực, tiêu cực và nổi bật trong việc xác định trạng thái tích cực (Hình 5). Kết quả thực nghiệm cho thấy mô hình có tiềm năng ứng dụng trong giáo dục, giúp các nhà giáo dục điều chỉnh phương pháp giảng dạy sao cho phù hợp với trạng thái cảm xúc của người học. Việc duy trì cảm xúc tích cực và tạo hứng thú học tập là một trong những mục tiêu quan trọng của giáo dục hiện nay. Mô hình ECL-tEEG đạt được độ chính xác cao hơn các mô hình còn lại trong việc so sánh độ chính xác trong các

lớp tích cực, trung tính, tiêu cực. Cụ thể, mô hình ECL-tEEG nhận được kết quả nhận diện 3 cảm xúc tích cực, trung tính, tiêu cực lần lượt là 77%, 61%, 70%. Mô hình CNN-LSTM-GRU, mô hình CNN-Transformer và mô hình BiLSTM nhận kết quả lần lượt là 76%, 64%, 66% và 76%, 54%, 67%. Các mô hình truyền thống như SVM, Random Forest có kết quả lần lượt là 55%, 09%, 53% và 60%, 07% và 56%. Các mô hình trên đều có những ưu điểm phù hợp với các dạng dữ liệu khác nhau. CNN-TRANSFORMER phù hợp với các đoạn tín hiệu với thời gian dài như EEG nhưng lại kém hiệu quả hơn khi xử lý các bộ dữ liệu với các đặc trưng quan hệ ngắn hạn. Mô hình CNN-LSTM-GRU có khả năng học tốt được các phụ thuộc tuần tự ngắn hạn nhưng lại gặp khó khăn và dễ bị bão hòa gradient khi chuỗi thời gian dữ liệu quá dài. Kết hợp giữa hai mô hình, mô hình ECL-tEEG mang lại hiệu suất vượt trội khi kết hợp được lợi thế của ba phương pháp xử lý dữ liệu trên khi xử lý các đoạn dữ liệu với nhiều đặc trưng, xử lý được các mối quan hệ dài, ngắn hạn. Tuy nhiên, mô hình này có độ phức tạp cao hơn các mô hình SVM, Random Forest, CNN và CNN-LSTM-GRU về mặt tham số, nhưng thời gian huấn luyện trên từng epoch lại nhanh hơn hai mô hình BiLSTM và CNN-Transformer. Do đó, tùy vào đặc trưng dữ liệu đầu vào, ta cần tinh chỉnh sao cho phù hợp.

Mô hình ECL-tEEG không chỉ cho thấy hiệu quả vượt trội trong phân loại tín hiệu EEG mà còn sở hữu tiềm năng ứng dụng cao trong các tình huống thực tế. Cụ thể, ECL-tEEG có thể được tích hợp vào hệ thống hỗ trợ giảng dạy thông minh (Smart Education Support Systems) nhằm theo dõi trạng thái cảm xúc và mức độ tập trung của học sinh theo thời gian thực. Thông qua đó, người dạy có thể phát hiện sớm các biểu hiện căng thẳng hoặc mất tập trung, từ đó có phương hướng điều chỉnh phương pháp giảng dạy và đưa ra thời gian nghỉ phù hợp với thực tế. Ngoài ra, mô hình có thể ứng dụng trong các nền tảng E-learning để tùy chỉnh nhịp độ bài giảng và lựa chọn linh hoạt các hoạt động học phù hợp với từng người học. Điều này không chỉ tăng mức độ tương tác mà còn hỗ trợ tiếp thu kiến thức hiệu quả hơn, ngay cả khi không có người giảng dạy trực tiếp. Trong tương lai, mô hình này còn có thể đóng vai trò trong các hệ thống chăm sóc sức khỏe tinh thần học đường, hỗ trợ theo dõi và phát hiện sớm các vấn đề tâm lý như lo âu hoặc stress học tập, từ đó can thiệp kịp thời và hiệu quả. Mô hình ECL-tEEG góp phần cá nhân hóa trải nghiệm học tập và nâng cao chất lượng giảng dạy trong môi trường giáo dục hiện đại.

4. Kết luận

Trong nghiên cứu này, chúng tôi đã đề xuất mô hình ECL-tEEG, kết hợp ba phương pháp CNN, LSTM-GRU và Transformer để tận dụng ưu điểm của từng mô hình trong phân tích tín hiệu EEG. Mô hình đề xuất có khả năng trích xuất hiệu quả đặc trưng không gian, học đặc trưng tuần tự và phát hiện các mối quan hệ dài hạn trong tín hiệu EEG, giúp nâng cao độ chính xác trong nhận diện cảm xúc của người học. Kết quả thực nghiệm cho thấy mô hình ECL-tEEG đạt độ chính xác cao hơn so với hai mô hình riêng lẻ CNN-LSTM-GRU và CNN-Transformer, với độ chính xác tổng thể đạt 71%. Đặc biệt, mô hình hoạt động tốt trong việc phân loại cảm xúc tích cực và tiêu cực, góp phần hỗ trợ giáo viên trong việc theo dõi trạng thái cảm xúc của người học và cải thiện phương pháp giảng dạy. Tuy nhiên, mô hình

vẫn chưa đạt được độ chính xác lí tưởng, đòi hỏi tinh chỉnh tham số phù hợp và có thể cần thêm dữ liệu huấn luyện từ các tình huống thực tế để cải thiện khả năng tổng quát hóa. Trong tương lai, chúng tôi đề xuất mở rộng nghiên cứu bằng cách thu thập dữ liệu thực nghiệm trên người học trong môi trường giáo dục, từ đó phát triển hệ thống hỗ trợ giảng dạy thông minh dựa trên phân tích cảm xúc EEG.

- ❖ *Tuyên bố về quyền lợi: Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.*
- ❖ *Lời cảm ơn: Nghiên cứu này được tài trợ bởi Nguồn ngân sách khoa học và công nghệ Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong đề tài NCKH của sinh viên năm học 2024-2025*

TÀI LIỆU THAM KHẢO

- Andayani, F., Theng, L. B., Tsun, M. T., & Chua, C. (2022). Hybrid LSTM-Transformer Model for Emotion Recognition From Speech Audio Files. *IEEE Access*, 10, 36018–36027. <https://doi.org/10.1109/access.2022.3163856>
- Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., & Pollastro, A. (2022). EEG-based measurement system for monitoring student engagement in learning 4.0. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-09578-y>
- Florestiyanto, M. Y. (2024). Emotion Recognition for Improving Online Learning Environments: A Systematic Review of the Literature. *Deleted Journal*, 20(4s), 1860–1873. <https://doi.org/10.52783/jes.2255>
- Hari Krishna, B., Sharon Rose Victor, J., Srinivasa Rao, G., Raja Kishore Babu, Ch., Srujan Raju, K., Ghouse Basha, T. S., & Bharath Simha Reddy, V. (2024). Emotion-net: Automatic emotion recognition system using optimal feature selection-based hidden markov CNN model. *Ain Shams Engineering Journal*, 103038. <https://doi.org/10.1016/j.asej.2024.103038>
- Hsu, C. F., Chao, H.-H., Yang, A. C., Yeh, C.-W., Hsu, L., & Chi, S. (2020). Discrimination of Severity of Alzheimer's Disease with Multiscale Entropy Analysis of EEG Dynamics. *Applied Sciences*, 10(4), 1244–1244. <https://doi.org/10.3390/app10041244>
- Jaworska, N., de la Salle, S., Ibrahim, M.-H., Blier, P., & Knott, V. (2019). Leveraging Machine Learning Approaches for Predicting Antidepressant Treatment Response Using Electroencephalography (EEG) and Clinical Data. *Frontiers in Psychiatry*, 9. <https://doi.org/10.3389/fpsy.2018.00768>
- Joshi, V. M., & Ghongade, R. B. (2022). IDEA: Intellect database for emotion analysis using EEG signal. *Journal of King Saud University - Computer and Information Sciences*, 34(7), 4433–4447. <https://doi.org/10.1016/j.jksuci.2020.10.007>
- Lee, M.-H., Shomanov, A., Begim, B., Kabidenova, Z., Nyssanbay, A., Yazici, A., & Lee, S.-W. (2024). EAV: EEG-Audio-Video Dataset for Emotion Recognition in Conversational Contexts. *Scientific Data*, 11(1). <https://doi.org/10.1038/s41597-024-03838-4>
- Mahmoud, A., Amin, K., Al Rahhal, M. M., Elkilani, W. S., Mekhalfi, M. L., & Ibrahim, M. (2023). A CNN Approach for Emotion Recognition via EEG. *Symmetry*, 15(10), 1822. <https://doi.org/10.3390/sym15101822>

- Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., & Cui, Z. (2019). MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition. *IEEE Access*, 7, 12177–12191. <https://doi.org/10.1109/ACCESS.2019.2891579>
- Song, Y., Zheng, Q., Liu, B., & Gao, X. (2023). EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 710–719. <https://doi.org/10.1109/tnsre.2022.3230250>
- Taware, S., & Thakare, A. (2023). Critical Analysis on Multimodal Emotion Recognition in Meeting the Requirements for Next Generation Human Computer Interactions. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(9s), 523–534. <https://doi.org/10.17762/ijritcc.v11i9s.7464>
- Yu, C., & Wang, M. (2022). Survey of emotion recognition methods using EEG information. *Cognitive Robotics*, 2, 132–146. <https://doi.org/10.1016/j.cogr.2022.06.001>
- Zhu, X., Song, Y., & Li, D. (2024). EEG Emotion Recognition Based on CNN+LSTM. *2024 IEEE 25th China Conference on System Simulation Technology and Its Application (CCSSTA)*, 144–148. <https://doi.org/10.1109/ccssta62096.2024.10691696>

**ECL-tEEG: A HYBRID DEEP LEARNING MODEL
FOR ESTIMATING LEARNERS' EMOTIONS FROM EEG SIGNALS**

**Phan Thi Quy Thinh, Phung Ngoc Nhan, Mai Yen Khoa,
Tran Thanh Nha, Vo Le Phuc Hau*, Nguyen Viet Hung**

Ho Chi Minh City University of Education, Vietnam

*Corresponding Author: Vo Le Phuc Hau – Email: tg.hauvlp@lecturer.hcmue.edu.vn

Received: April 09, 2025; Revised: April 21, 2025 ; Accepted: April 24, 2025

ABSTRACT

Recognizing emotions and assessing student engagement are important for improving instructional effectiveness and personalizing teaching. Previous studies have primarily used video or image data with deep learning models to analyze learners' behavior and emotions. However, these approaches are sensitive to camera angles, lighting conditions, and ambiguous facial expressions, which may reduce the accuracy of emotion assessment. To address these limitations, this study proposes ECL-tEEG, a model that combines a convolutional neural network (CNN), an LSTM–GRU architecture, and a Transformer to analyze electroencephalogram (EEG) signals for emotion recognition and student engagement assessment. The model uses the CNN to extract spatial features, the LSTM–GRU architecture to capture temporal dependencies, and the Transformer to model long-range relationships through parallel processing. Its performance was evaluated against two comparison models: CNN-LSTM-GRU and CNN-Transformer. ECL-tEEG achieved an accuracy of 71%, outperforming both comparison models in classifying positive, neutral, and negative emotions. The findings demonstrate the potential of EEG-based emotion recognition and provide a basis for intelligent instructional support systems that can personalize learning and help teachers adjust their instruction using neural signals alongside behavioral observations.

Keywords: emotion recognition; electroencephalogram (EEG); education; machine learning; CNN; LSTM-GRU; Transformer