

Ứng dụng thuật toán mạng bộ nhớ dài ngắn hạn trong phân loại tín hiệu sóng não

Huỳnh Quốc Việt^{1,2}, Nguyễn Thị Như Quỳnh^{1,2}, Trần Đức Minh^{1,2}, Lê Ngọc Ánh^{1,2}, Nguyễn Thanh Phước^{1,2}, Huỳnh Văn Tuấn^{1,2,*}



Use your smartphone to scan this QR code and download this article

TÓM TẮT

Cảm xúc con người đóng vai trò quan trọng trong việc giao tiếp khi không sử dụng ngôn ngữ, và cảm xúc cũng hỗ trợ cho các nghiên cứu về hành vi của con người. Ngoài ra, tín hiệu điện não đã được các nhà nghiên cứu đánh giá cao về độ tin cậy cũng như dễ dàng lưu trữ và ghi nhận. Do đó phương pháp sử dụng điện não đồ để nhận dạng cảm xúc của con người hiện đang là một lĩnh vực tương đối mới. Các ý tưởng chính như tiến xử lý tín hiệu, trích xuất tính năng và tối ưu hóa thuật toán là những phương pháp đang được nhiều nhà nghiên cứu nhắm tới trong lĩnh vực này. Trong bài nghiên cứu này, chúng tôi hướng đến nhận diện các trạng thái cảm xúc thông qua tín hiệu điện não bằng cách sử dụng các thuật toán bộ nhớ dài ngắn hạn. Tập dữ liệu tín hiệu cảm xúc được lấy từ cơ sở dữ liệu DEAP của tác giả Koelstra và các cộng sự để phục vụ cho quá trình nghiên cứu. Việc nghiên cứu sẽ tập trung vào độ chính xác phân loại và thời gian huấn luyện, đồng thời kiểm tra các dạng kiến trúc khác nhau cũng như các thông số khởi tạo của mạng bộ nhớ dài ngắn hạn (LSTM). Các kết quả đạt được chỉ ra rằng cấu trúc bộ dữ liệu 3 chiều có hiệu suất tốt hơn so với cấu trúc bộ dữ liệu 2 chiều và mạng bộ nhớ dài ngắn hạn là phù hợp cho nhiệm vụ này. Ngoài ra, nghiên cứu của chúng tôi cũng được đối sánh với các nghiên cứu của các tác giả khác nhằm chứng minh sự hiệu quả của thuật toán phân loại.

Từ khoá: dữ liệu hai chiều, dữ liệu ba chiều, điện não đồ, cảm xúc, mạng bộ nhớ dài ngắn hạn

GIỚI THIỆU

Trong những năm gần đây, đã có rất nhiều các phương pháp được sử dụng để trích xuất đặc trưng của tín hiệu cảm xúc con người như là biểu hiện khuôn mặt¹, giọng nói^{2,3}, nháy mắt⁴, hoặc sử dụng các tín hiệu sinh lý. Khi so sánh các biểu hiện đã đề cập, chỉ có biểu hiện sử dụng các tín hiệu sinh lý từ con người là được các nhà nghiên cứu đánh giá cao về độ tin cậy⁵, do việc sử dụng tín hiệu sinh lý đã được các nhà nghiên cứu khẳng định rằng là khó có thể làm giả⁶. Ngoài ra, đây là dạng tín hiệu cực kỳ đa dạng như tín hiệu điện não đồ (EEG)^{7,8}, điện cơ (EMG)⁹, huyết áp (BVP)¹⁰, điện tim (ECG)^{11,12}... Các tín hiệu này có thể dễ dàng ghi nhận và lưu trữ. Trong bài báo này, chúng tôi sử dụng tín hiệu EEG nhằm phục vụ quá trình phân loại cảm xúc của con người.

Để phân loại các trạng thái cảm xúc thông qua tín hiệu EEG, thì quá trình phân loại cần phải khắc phục các hạn chế ở các bước như tiền xử lý, trích xuất các đặc trưng của tín hiệu, và tối ưu hóa mô hình phân loại. Nhiều nhà nghiên cứu đã đề xuất các phương pháp để giải quyết những hạn chế kể trên. Nhà nghiên cứu Zheng và Lu đã sử dụng mạng nơron học sâu nhằm kiểm tra các dải tần số quan trọng để

nhận dạng cảm xúc thông qua tín hiệu EEG và các tác giả đã đưa ra kết luận rằng dải beta và gamma phù hợp nhất¹³. Tác giả Lakshmi và các cộng sự đã khảo sát các phương pháp xử lý và trích xuất tín hiệu EEG, bao gồm phương pháp phân tích thành phần độc lập (ICA), phương pháp phân tích thành phần cơ bản (PCA), phương pháp biến đổi Wavelet, và phương pháp biến đổi Fourier¹⁴. Tác giả Hatamikia và Nasrabadi đã đề xuất sử dụng bốn phương pháp trích xuất đặc trưng bao gồm entropy gần đúng, entropy phổ, phân cụm khoảng cách của Katz và phân cụm khoảng cách của Petrosian, từ đó kết hợp các phương pháp với nhau, sau đó họ sử dụng chỉ số *Dunn* nhằm chọn những tính năng nhiều thông tin nhất và phân loại cảm xúc bằng cách sử dụng bản đồ tự tổ chức (SOM)¹⁵.

Trong tất cả các nghiên cứu được đề cập trước đó, các tác giả đã sử dụng rất nhiều phương pháp để trích xuất các tính năng đặc trưng của tín hiệu EEG, cũng như sử dụng nhiều các thuật toán phân loại khác nhau. Chúng tôi nhận thấy việc trích xuất tính năng của các tác giả chỉ tập trung vào các chiều không gian từ một chiều (1D) đến hai chiều (2D) của tín hiệu. Tuy nhiên, nếu nghiên cứu thêm tín hiệu ba chiều (3D) thì chúng ta có thể mô tả tổng quát hơn về độ sâu của dữ liệu.

Trích dẫn bài báo này: Việt H Q, Quỳnh N T N, Minh T D, Ánh L N, Phước N T, Tuấn H V. **Ứng dụng thuật toán mạng bộ nhớ dài ngắn hạn trong phân loại tín hiệu sóng não.** *Sci. Tech. Dev. J. - Nat. Sci.*; 5(2):1167-1178.

¹Khoa Vật lý - Vật lý Kỹ thuật, Trường Đại học Khoa học Tự Nhiên, Việt Nam

²Đại học Quốc gia TP. HCM, Việt Nam

Liên hệ

Huỳnh Văn Tuấn, Khoa Vật lý - Vật lý Kỹ thuật, Trường Đại học Khoa học Tự Nhiên, Việt Nam

Đại học Quốc gia TP. HCM, Việt Nam

Email: hvtuan@hcmus.edu.vn

Lịch sử

- Ngày nhận: 30-12-2020
- Ngày chấp nhận: 25-3-2021
- Ngày đăng: 30-4-2021

DOI: 10.32508/stdjns.v5i2.1006



Bản quyền

© ĐHQG TP.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



Ngoài ra, các tín hiệu EEG là tín hiệu liên tục theo thời gian¹⁶, chính vì thế các phân đoạn tại một thời điểm nhất định cũng sẽ có ảnh hưởng tới các phân đoạn trước hoặc sau nó. Và việc sử dụng các thuật toán chỉ có khả năng phân loại tín hiệu ở những thời điểm nhất định như các nghiên cứu trước sẽ khó khăn khi các tín hiệu biến thiên liên tục. Các thuật toán phân loại như vậy có thể bỏ sót một số thông tin của tín hiệu.

Trong bài nghiên cứu này, chúng tôi phân loại các trạng thái cảm xúc của con người thông qua tín hiệu điện não bằng cách sử dụng các thuật toán bộ nhớ dài ngắn hạn (LSTM). Tập dữ liệu tín hiệu cảm xúc được lấy từ cơ sở dữ liệu DEAP để phục vụ cho quá trình nghiên cứu. Chúng tôi chia tập dữ liệu thành các phân đoạn có độ dài thời gian bằng nhau, các phân đoạn này được trích xuất thành bốn dải tần số (δ , α , β , γ).

- Dải δ (0.5-4 Hz) là dải tần số có biên độ khoảng 20 mV thường được xuất hiện khi con người trong trạng thái ngủ. Khi dải δ thay đổi biên độ và tần số là dấu hiệu của thiếu oxy não và có nguy cơ đột quỵ.

- Dải α (7.5-12.5 Hz) là dải tần số có biên độ trung bình từ 50-70 mV và thường xuất hiện khi con người thư giãn, tỉnh táo hoặc khi nhắm mắt, tổn thương võng mạc.

- Dải β (13-40 Hz) là dải tần số có biên độ nhỏ hơn 15mV và thường xuất hiện khi con người trong trạng thái suy nghĩ, tập trung. Khi dải β thay đổi biểu hiện cho dấu hiệu của sự căng thẳng, lo âu.

- Dải γ (40-60 Hz) là dải tần số có biên độ nhỏ và thường xuất hiện khi con người thiếu định, tập trung chú ý, và quản lý bộ nhớ.

Thông qua bộ lọc và sau đó chúng tôi kết hợp các phân đoạn ở từng dải tần số thành một hình khối 3D. Việc nghiên cứu sẽ tập trung vào độ chính xác phân loại và thời gian huấn luyện thông qua việc khảo sát sự ảnh hưởng của các thông số bên trong kiến trúc mạng LSTM. Ngoài ra, việc nghiên cứu của chúng tôi cũng được so sánh với các phương pháp của những tác giả trước đây để đánh giá sự hiệu quả của thuật toán.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Bộ dữ liệu DEAP

Bộ dữ liệu DEAP đã được sử dụng rộng rãi trong các nghiên cứu nhận dạng cảm xúc tín hiệu EEG. Trong tập dữ liệu này, tín hiệu EEG của 32 người tình nguyện tham gia được ghi lại ở hai phòng nghiên cứu khác nhau, những người tham gia được đánh số từ 1 đến 22 sẽ được thu tín hiệu ở Twente (Mỹ) và những người còn lại sẽ được thu ở Geneva (Thụy Sĩ). Quá trình thu tín hiệu điện não đã được diễn ra khi các tình nguyện viên ngồi xem 40 đoạn phim khác nhau

với mỗi đoạn phim có độ dài 60 giây. Với mỗi đoạn phim sau khi được xem xong, mỗi người tham gia sẽ phải đánh giá ngay lập tức cảm xúc theo các mức từ một đến chín ở bốn loại tín hiệu cảm xúc bao gồm: *Arousal*, *Valence*, *Ấn tượng* và *Yêu thích*. Các kết quả đánh giá của những người tham gia sẽ được sử dụng như là *nhãn* tương ứng cho quá trình thu đó. Sau khi đánh giá và tín hiệu được kiểm tra là phù hợp, thì quá trình thu tín hiệu sẽ lặp lại ở đoạn phim tiếp theo¹⁷.

Ngoài ra, các tác giả của bộ dữ liệu DEAP cũng cung cấp một phiên bản đã được tiền xử lý trước các dữ liệu thô nhằm phục vụ cho các quá trình nghiên cứu. Phiên bản này được chia làm hai phần: dữ liệu và nhãn. Trong phần dữ liệu sẽ chứa một ma trận với kích thước $40 \times 40 \times 8064$ (số mẫu \times video \times kênh), và đối với phần nhãn sẽ có một ma trận 40×4 (mức độ \times cảm xúc) với mỗi cột là một dạng cảm xúc (*Arousal*, *Valence*, *Ấn tượng* và *Yêu thích*). Trong phiên bản tiền xử lý này, các dữ liệu được lấy mẫu xuống còn 128 Hz và các tín hiệu được lọc thông qua bộ lọc thông dải có tần số từ 4Hz - 40,5Hz. Các tín hiệu đo thử trong 3 giây đầu được xóa¹⁷. Bảng 1 trình bày tóm tắt về phiên bản tiền xử lý của tập dữ liệu DEAP. Trong nghiên cứu này, chúng tôi chỉ sử dụng 32 kênh trong bộ dữ liệu được tiền xử lý bởi vì các kênh còn lại (kênh 33 đến kênh 40) được sử dụng nhằm phục vụ đo các tín hiệu EOG và các thông số khác như nhiệt độ, huyết áp ... Chính vì thế các kênh này không liên quan đến nghiên cứu của bài báo này. Ngoài ra đối với nhãn chỉ sử dụng cảm xúc *Arousal* để phân loại do đây là cảm xúc đặc trưng có kết quả cao nhất và đặc trưng nhất cho quá trình phân loại¹⁸.

Tách dải tần số và trích xuất các tính năng đặc trưng

Mỗi tín hiệu EEG của những lần thu được phân đoạn theo ba độ dài lần lượt là 0,5 giây, 1 giây và 1,5 giây. Mỗi độ dài phân đoạn được phân cụm chung nhằm tạo thành ba tập dữ liệu mới; Các tập dữ liệu sẽ được đặt tên lần lượt là “*input_1*” (tập dữ liệu với phân đoạn dài 0,5 giây), “*input_2*” (tập dữ liệu với phân đoạn dài 1 giây) và “*input_3*” (tập dữ liệu với phân đoạn dài 1,5 giây).

Các nghiên cứu trước đã chỉ ra rằng các dải tần số với tần suất xuất hiện khác nhau sẽ đại diện cho từng trạng thái cảm xúc của con người¹⁹. Do đó, các phân đoạn được tách thành bốn dải tần số (δ , α , β , γ) thông qua bộ lọc Butterworth.

Chúng tôi sử dụng công thức entropy vi sai (DE) liên tục để tính toán được độ phức tạp của từng phân đoạn tín hiệu EEG trong từng bộ dữ liệu²⁰. Đối với một chuỗi tín hiệu tuân theo phân phối Gauss N (μ , δ^2)

Bảng 1: Thông số tổng quan của bộ dữ liệu DEAP

Tên thành phần	Cấu trúc	Các thành phần
Dữ liệu	40×40×8064	Đoạn phim × Kênh × Số mẫu
Nhân	40×4	Số lần thử × Nhân (Arousal, Valence, Ấn tượng và Yêu thích)

14, gọi X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất khi đó công thức của DE của $h(X)$ được mô tả như sau:

$$h(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \times \log \left(\frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \right) dx \pi \quad (1)$$

$$h(X) = \frac{1}{2} \log(2\pi e \delta^2) \quad (2)$$

Dựa vào kết quả khảo sát của Laskhmi¹⁴, các tín hiệu EEG đã được chứng minh rằng chúng hoạt động theo phân bố Gaussian đối với các dải tần số từ 2Hz - 44Hz. Do đó, công thức entropy vi sai cho từng tần số được mô tả như sau²⁰:

$$h_i(X) = \frac{1}{2} \log(2\pi e \delta_i^2) \quad (3)$$

trong đó h_i là entropy vi sai của các tín hiệu EEG tương ứng trong dải tần số i và δ_i^2 là phương sai của tín hiệu.

Dựa vào công thức (3), kết quả tính toán của mỗi phân đoạn EEG sẽ là một vec-tơ 1D. Toàn bộ quá trình trên được mô tả tổng quát như Hình 1.

Xây dựng cấu trúc 3 chiều và kết hợp các phân đoạn

Trong mục này, bốn vec-tơ đại diện cho bốn dải tần số ($\delta, \alpha, \beta, \gamma$) xây dựng thành các ma trận 2D nhằm mô phỏng cấu trúc của thiết bị đo trong tập dữ liệu DEAP. Từ đó chúng tôi có thể duy trì thông tin khoảng cách giữa các kênh trong thiết bị thu. Một ma trận 2D với kích thước $h \times w$ sẽ được sử dụng, trong đó h và w là số lượng tối đa các điện cực khi xét theo chiều dọc và ngang. Với bộ dữ liệu DEAP, giá trị lớn nhất cho h và w bằng chín¹⁸. Ngoài ra, các giá trị bằng không được thêm vào những điểm không có điện cực như Hình 2. Dựa vào Hình 2, mỗi dải tần số sẽ được biến đổi thành những ma trận 2D. Các ma trận này được kết hợp với nhau để tạo thành một khối 3D, và chúng chính là một khung (frame) của chuỗi tín hiệu EEG Hình 3.

Mạng bộ nhớ dài ngắn hạn (LSTM)

Mạng LSTM là mô hình được sử dụng để giải quyết các bài toán liên quan đến chuỗi tín hiệu. Điểm đặc biệt của mạng LSTM là số lượng hàm hoạt động có trong một đơn vị kiến trúc. Nói cách khác, một đơn vị có thể chứa nhiều hơn một hàm hoạt động²¹. Trong đó, các hàm hoạt động được kết hợp với nhau để cùng thực hiện một chức năng được gọi là một “cổng”. Kiến trúc của mỗi đơn vị trong mạng LSTM gồm ba cổng và được mô tả tổng quan như Hình 4.

Cổng đầu tiên là “*cổng quên*”, nhiệm vụ của cổng nhằm xác định những thông tin từ các đơn vị ở phía trước và quyết định chúng có nên bị loại bỏ hay không. Nếu giá trị đầu ra ở “*cổng quên*” bằng không thì thông tin tương ứng sẽ bị xóa. Ngược lại, nếu giá trị đầu ra là một thì nó cho phép thông tin tương ứng đi vào cổng tiếp theo²². Đầu ra của cổng được tính toán theo (4):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

Cổng thứ hai là “*Cổng đầu vào*”, nhiệm vụ của cổng là tổng hợp thông tin từ tín hiệu đầu vào và kết quả đầu ra của “*cổng quên*”. Nói cách khác, kết quả tính toán tại “*Cổng đầu vào*” là sự kết hợp giữa thông tin từ đơn vị trước và thông tin đầu vào tại thời điểm tính toán²², công thức được mô tả theo (5):

$$C_t = C_{t-1} * f_t + \tilde{C}_t * i_t \quad (5)$$

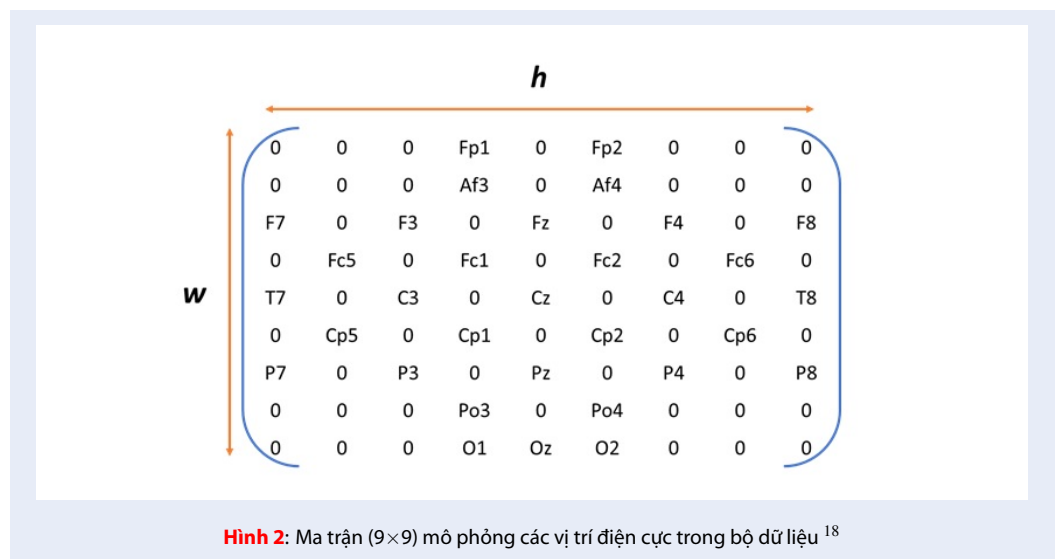
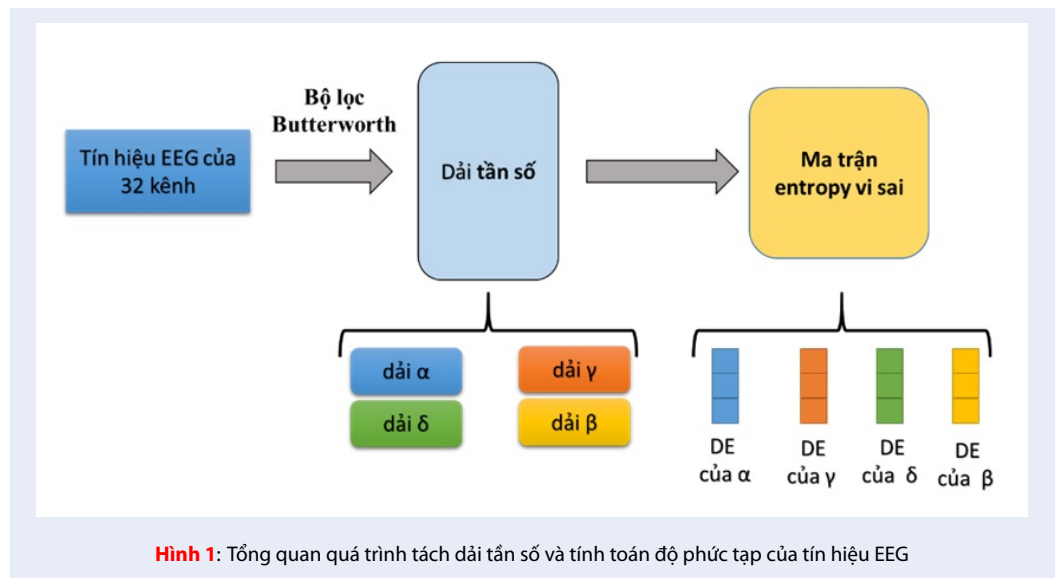
Cổng cuối cùng là “*Cổng đầu ra*”. Cổng này quyết định các giá trị được đưa vào đơn vị tiếp theo hoặc giá trị đầu ra cuối cùng tại thời điểm tính toán²². Công thức được tính theo (6) và (7):

$$o_t = \sigma(W_o \cdot [h_t - 1, x_t] + b_o) \quad (6)$$

$$h_t = \tanh(C_t) * o_t \quad (7)$$

KẾT QUẢ VÀ THẢO LUẬN

Trong mục này, chúng tôi thiết kế mạng LSTM hai lớp, với lớp thứ nhất có 512 đơn vị và lớp thứ hai có 64 đơn vị. Kiến trúc bên trong mỗi đơn vị gồm hai hàm hoạt động là hàm *Tanh* và hàm *Sigmoid*. Chúng tôi tạo một lớp kết nối đầy đủ với 1024 nút để nhận các kết quả đầu ra cuối cùng của lớp thứ hai nhằm phân loại các



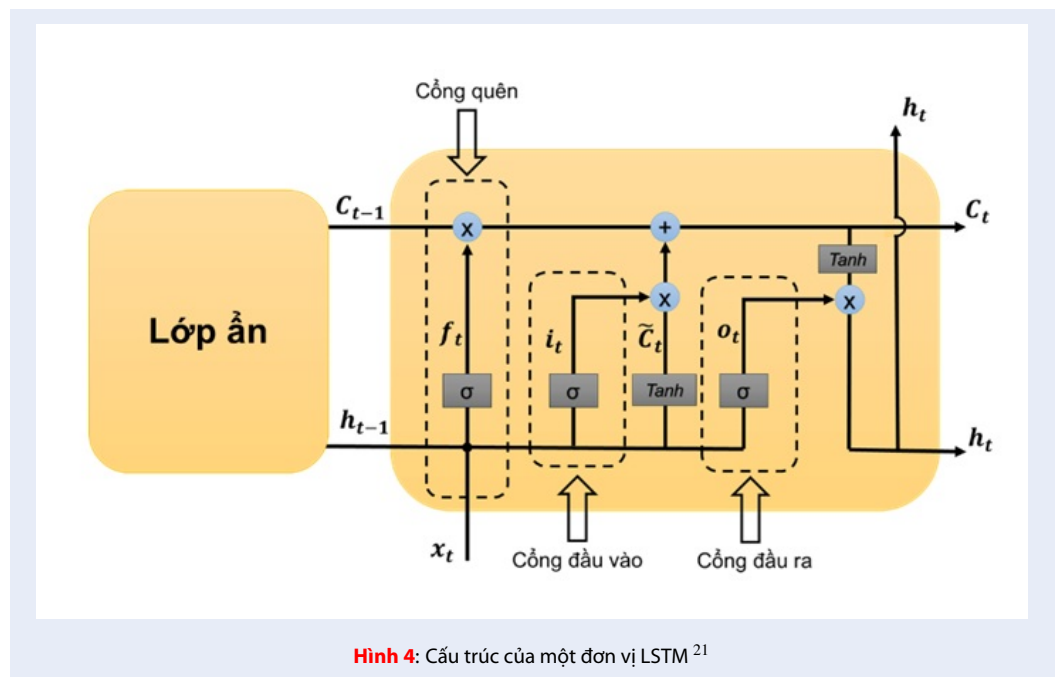
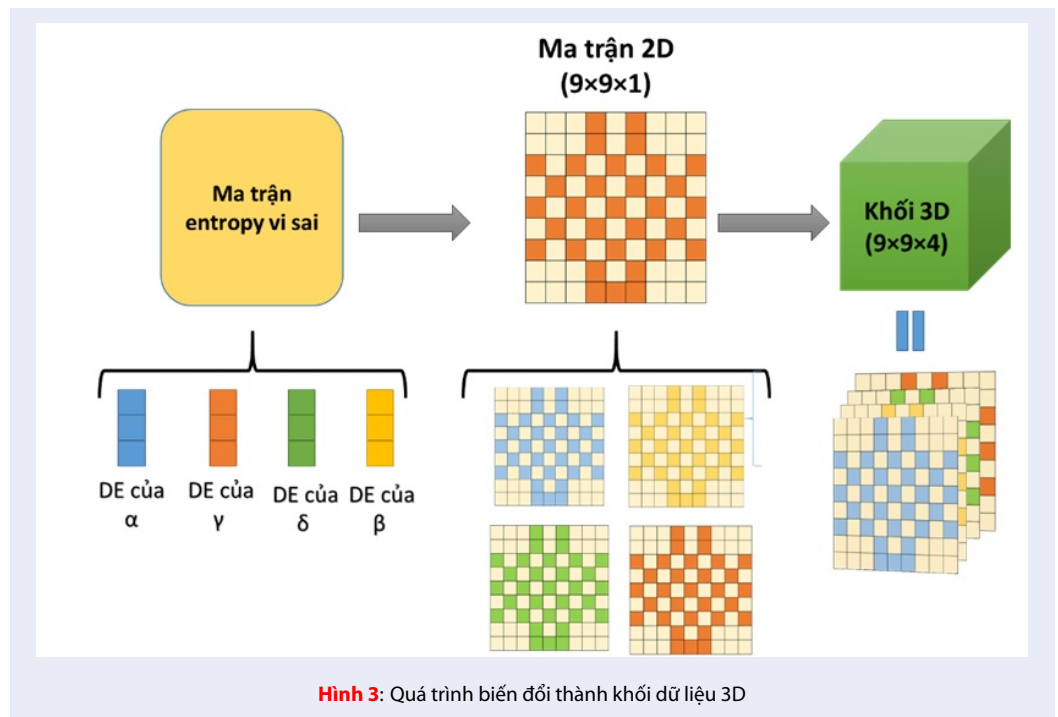
giá trị được tính toán tại đó. Các giá trị kết quả được đưa vào mỗi nút, từ đó lớp kết nối đầy đủ bỏ phiếu để đưa ra trạng thái cảm xúc nào là tương ứng với tín hiệu EEG được huấn luyện.

Đối với tập nhân của dữ liệu thì được phân chia thành hai mức cao và thấp của mức độ cảm xúc. Chúng tôi đặt ngưỡng với giá trị bằng năm để chia các giá trị nhân thành hai tập dữ liệu cao và thấp. Mức cao có giá trị lớn hơn hoặc bằng năm và mức thấp sẽ có giá trị nhỏ hơn năm. Để đảm bảo sự cân bằng giữa mức cao và thấp của trạng thái cảm xúc, chúng tôi sử dụng phương pháp entropy chéo (cross-entropy) để chia dữ liệu thành 10 phần bằng nhau (chín phần cho quá trình huấn luyện và một phần cho quá trình kiểm tra). Độ chính xác cuối cùng là kết quả trung bình của

10 lần thay đổi giữa phần huấn luyện và kiểm tra. Giá trị của tốc độ học được chọn ngẫu nhiên bằng 10^{-4} và giá trị chuẩn hóa $L2$ ($L2$ regularization) được khởi tạo bằng 0,005 nhằm tránh tình trạng “overfitting” (khi trong trạng thái này, mạng sẽ không thể hoặc gặp khó khăn trong quá trình huấn luyện các dữ liệu mới).

Khảo sát sự ảnh hưởng của từng giai đoạn huấn luyện tới mô hình phân loại

Chúng tôi sử dụng “input_1” làm tập dữ liệu cho quá trình khảo sát. Đầu tiên, chúng tôi khảo sát mạng LSTM với số lần huấn luyện khác nhau (epoch), từ đó đánh giá sự ảnh hưởng của số epoch đến độ chính xác phân loại và thời gian huấn luyện. Kết quả được



mô tả như Hình 5, với epoch bằng 10 thì độ chính xác phân loại đạt giá trị cao nhất (65,6%). Ngoài ra, chúng tôi nhận thấy với epoch lớn hơn 10 thì độ chính xác giảm đáng kể (xấp xỉ 7,8%), nguyên nhân là do mạng không thể học thêm được các dữ liệu mới nữa nhưng vẫn kéo dài huấn luyện. Bởi vì khi các dữ liệu mới được huấn luyện thì mạng sẽ phải tiếp tục thay đổi kiến trúc để thích ứng, quá trình thay đổi này có thể tạo ra một kiến trúc mạng không phù hợp với các dữ liệu đã được huấn luyện.

Về mặt thời gian huấn luyện, thời gian tăng dần theo số epoch và dao động từ 81 giây đến 434 giây, chúng tôi có thể kết luận rằng thời gian tỉ lệ thuận với số epoch. Tuy nhiên, đối với epoch bằng 3 thì thời gian không còn tăng tuyến tính nữa mà tăng lên một cách đột biến. Do đó, chúng tôi rút ra được mạng LSTM không thể huấn luyện với epoch bằng 3.

Khảo sát các giá trị tốc độ học ảnh hưởng tới mô hình phân loại

Trong mục này, để xác định rõ hơn mối quan hệ giữa độ chính xác phân loại và thời gian huấn luyện với tốc độ học. Chúng tôi sẽ huấn luyện mạng LSTM với tốc độ học thay đổi từ 10^{-1} - 10^{-5} ; Kết quả được trình bày trong Bảng 2.

Trong Bảng 2, độ chính xác phân loại hiệu quả nhất (65,6%) khi độ học bằng 10^{-4} và đạt hiệu quả thấp nhất (51,6%) cho độ học bằng 10^{-5} . Theo kết quả nghiên cứu của Jing²³, nếu tốc độ học quá lớn thì mạng sẽ học các dữ liệu mới hiệu quả hơn nhưng quá trình hội tụ diễn ra nhanh hơn, điều này dẫn đến mạng sẽ kết thúc quá trình học trước khi hoàn thành số epoch được chỉ định. Ngược lại, khi tốc độ học quá nhỏ thì quá trình hội tụ sẽ diễn ra chậm hơn nhưng khả năng học những tính năng mới sẽ bị hạn chế. Dựa vào kết quả ở Bảng 2, chúng tôi kết luận rằng tốc độ học bằng 10^{-4} là giá trị phù hợp cho mô hình mạng LSTM, do với tốc độ học bằng 10^{-5} thì mạng đã bị chậm lại trong quá trình học các tín hiệu mới, điều này dẫn đến độ chính xác phân loại bị giảm.

Đối với thời gian huấn luyện, các kết quả chỉ ra rằng thời gian huấn luyện tỉ lệ thuận với độ học, và khi tốc độ học thay đổi sẽ ít dẫn đến tình trạng “overfitting” hơn so với quá trình thay đổi số epoch.

Khảo sát ảnh hưởng của số lớp trong mô hình đến độ chính xác phân loại

Trong mục này, chúng tôi muốn khảo sát sự ảnh hưởng của số lớp trong mạng LSTM với độ chính xác phân loại. Chúng tôi khởi tạo bốn kiến trúc khác nhau của mạng LSTM, bao gồm một lớp, hai lớp, ba lớp và bốn lớp với cùng một đầu vào và cùng các thông số kiến trúc. Kết quả được trình bày như Bảng 3.

Trong Bảng 3, mạng hai lớp cho kết quả có độ chính xác phân loại cao nhất, mạng bốn lớp có độ chính xác thấp nhất (50,5%). Kết quả khảo sát chi tiết hơn trình bày trong Hình 6, chúng tôi nhận thấy mạng LSTM ba lớp và bốn lớp không thể phân loại được các trạng thái ở mức cao và thấp của tín hiệu cảm xúc, do mạng không huấn luyện được và chỉ nhận ra một kiểu nhân (nhân không). Nguyên nhân dẫn đến việc không huấn luyện được là do có quá nhiều đơn vị giữa các lớp trong mạng LSTM, dẫn đến sự mất mát thông tin trong quá trình truyền tín hiệu (do “cồng quèn” loại bỏ quá nhiều thông tin).

Khảo sát mối quan hệ giữa đặc điểm dữ liệu với độ chính xác phân loại và thời gian huấn luyện

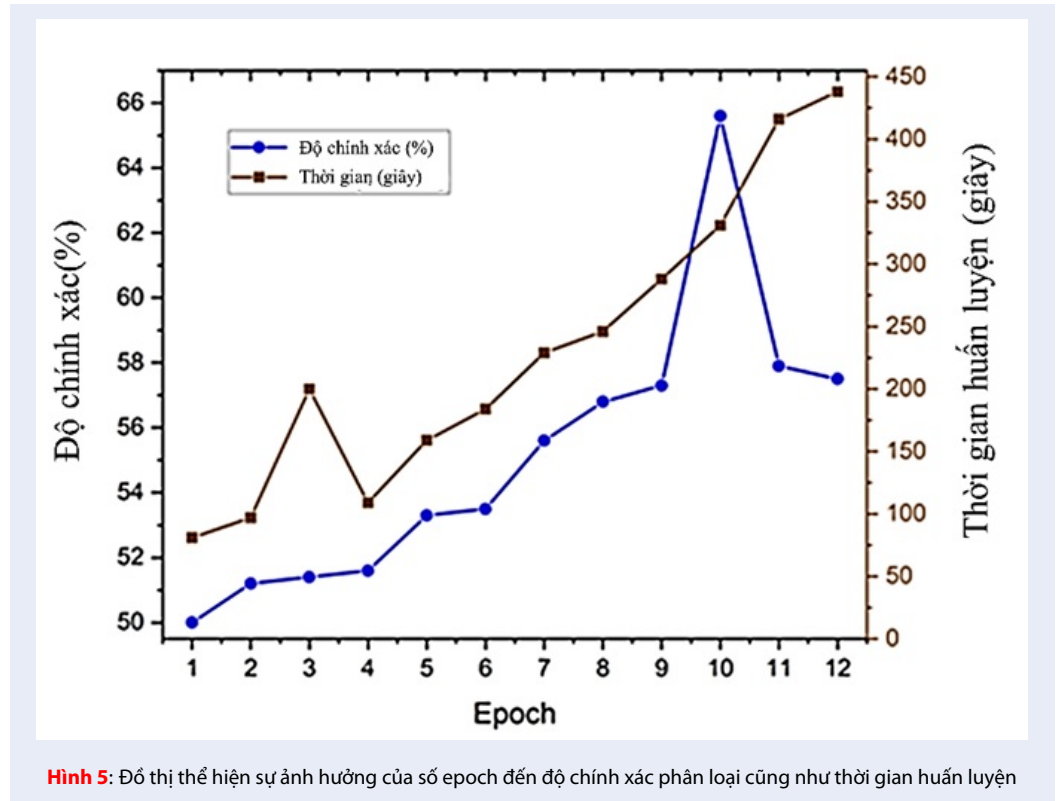
Tiếp theo, chúng tôi khảo sát sự ảnh hưởng của độ lớn dữ liệu đến độ chính xác phân loại của mạng. Mạng LSTM được cung cấp với các dữ liệu có độ lớn khác nhau, chúng tôi lấy n người trong số 32 người tham gia một cách ngẫu nhiên, và huấn luyện mạng với số lần bằng với kết quả làm tròn của $\frac{32}{n}$, độ chính xác phân loại sẽ là trung bình của $\frac{32}{n}$ lần lấy ngẫu nhiên. Kết quả được mô tả như Bảng 4.

Kết quả cho thấy độ chính xác phân loại cao nhất là với 32 người tham gia. Với 5 người tham gia và 10 người tham gia, mạng không thể học và phân loại các tín hiệu EEG, vấn đề xảy ra tương tự như mạng ba lớp (Hình 6). Dựa vào kết quả, chúng tôi kết luận rằng, mạng LSTM không phù hợp để huấn luyện với các bộ dữ liệu nhỏ, do đặc tính xử lý tín hiệu chuỗi của mạng cần phải có nhiều thông tin từ các tín hiệu phía trước hỗ trợ. Đối với thời gian huấn luyện, chúng tôi kết luận rằng thời gian huấn luyện tỉ lệ thuận với kích thước của tập dữ liệu. Nguyên nhân xuất phát từ việc có nhiều dữ liệu thì mạng cần phải học tuần tự toàn bộ số dữ liệu ấy dẫn đến thời gian sẽ lâu hơn.

Trong quá trình khảo sát tiếp theo, các tập dữ liệu “input_1”, “input_2” và “input_3” được sử dụng với cùng một kiến trúc mạng LSTM tốt nhất, từ đó khảo sát sự ảnh hưởng của độ dài phân đoạn đến độ chính xác phân loại.

Trong Bảng 5, chúng tôi thấy rằng độ dài dữ liệu càng lớn thì độ chính xác phân loại càng thấp. Quá trình xử lý với dữ liệu chuỗi tuần tự thì chúng ta phải cần sử dụng các phân đoạn có độ dài phù hợp, nếu phân đoạn có độ dài quá lớn thì nó có nhiều thông tin tổng quát và ít có các thông tin chi tiết. Do đó, thông tin này không cung cấp nhiều tính năng mới cho quá trình huấn luyện.

Ngoài ra khi xét về mặt thời gian, dữ liệu có độ dài phân đoạn càng ngắn thì sẽ có thời gian huấn luyện



Bảng 2: Mối quan hệ giữa độ chính xác phân loại và thời gian huấn luyện tương ứng với các tốc độ học khác nhau

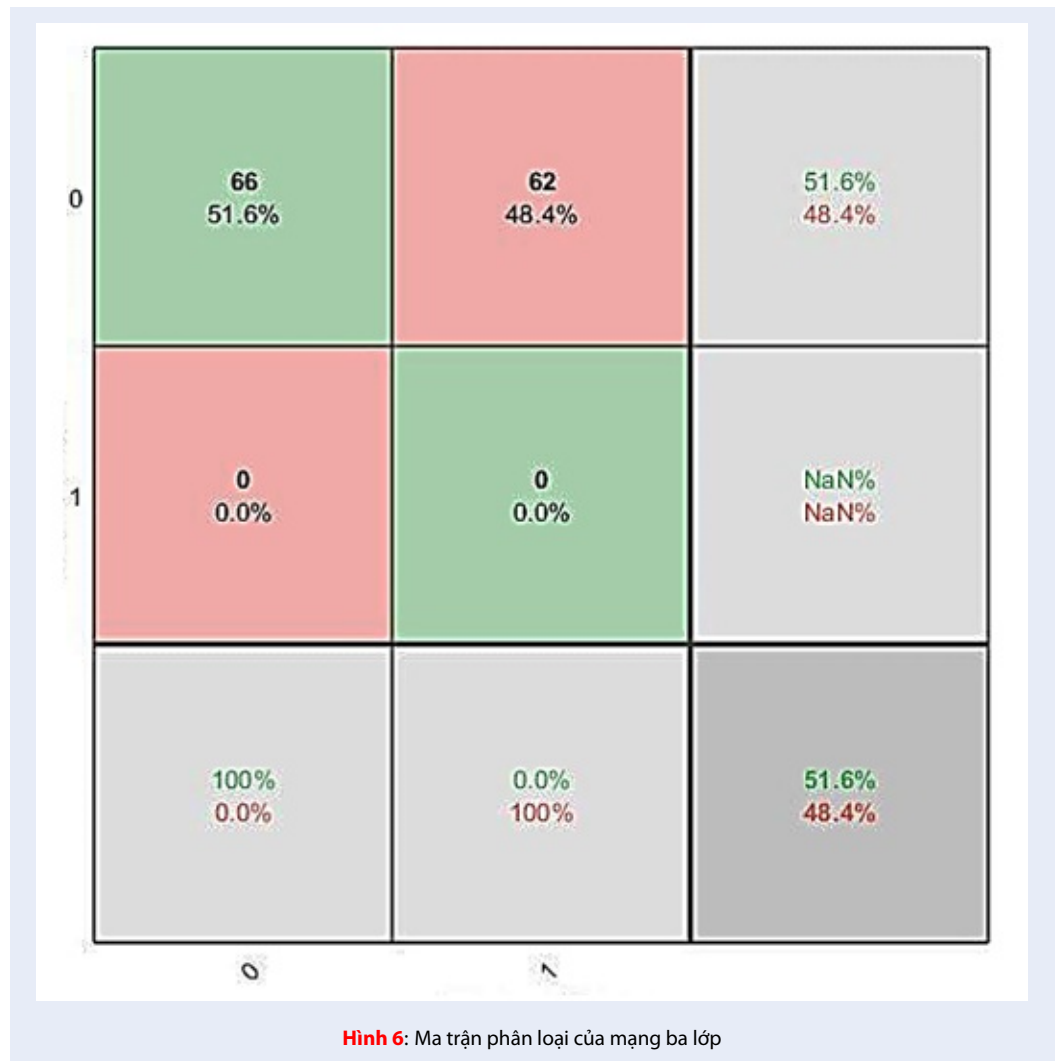
Tốc độ học	Độ chính xác (%)	Thời gian huấn luyện (giây)
10^{-1}	60,3	563
10^{-2}	62,5	535
10^{-3}	62,3	478
10^{-4}	65,6	347
10^{-5}	51,6	331

Bảng 3: Mối quan hệ giữa số lớp và độ chính xác phân loại của mạng LSTM

Số lớp	Một lớp	Hai lớp	Ba lớp	Bốn lớp
Độ chính xác (%)	60,9	65,6	51,6	50,5

Bảng 4: Mối quan hệ giữa độ lớn dữ liệu với độ chính xác phân loại và thời gian huấn luyện

Số người tham gia	5	10	15	20	25	30	32
Độ chính xác (%)	50,0	50,0	56,6	57,5	60,0	63,0	65,6
Thời gian huấn luyện (giây)	66	136	259	355	401	426	437



Bảng 5: Sự ảnh hưởng của độ dài phân đoạn và số chiều dữ liệu đến độ chính xác phân loại

Độ dài phân đoạn (giây)	Độ chính xác phân loại (%)		Thời gian huấn luyện (giây)	
	Dữ liệu 2D	Dữ liệu 3D	Dữ liệu 2D	Dữ liệu 3D
0,5	60,4	65,6	437	331
1	57,8	61,7	84	76
1,5	50	58,2	55	44

càng lâu, nguyên nhân là do sự thay đổi số lượng dữ liệu đầu vào, do tín hiệu được thu chỉ trong 60 giây, do đó khi chia dữ liệu độ dài là 0,5 giây thì sẽ có tới 120 mẫu và khi chia một giây sẽ có 60 mẫu. Khi số lượng mẫu tăng lên thì việc tính toán cũng sẽ diễn ra lâu hơn, điều này đã được chứng minh Bảng 4.

Ngoài ra, chúng tôi còn so sánh dữ liệu dạng 2D và 3D với độ dài phân đoạn khác nhau (Bảng 5). Các kết quả cho thấy, các dữ liệu 3D cho ra kết quả tốt hơn các dữ

liệu ở 2D trong tất cả độ dài phân đoạn cả về độ chính xác phân loại và thời gian huấn luyện. Đối với dữ liệu ở 2D được đưa vào huấn luyện trong mạng LSTM sẽ là một tấm hình bao gồm trục x và y tương đương với kích thước của độ dài và rộng của ma trận và các thông số thể hiện độ phức tạp của mỗi kênh thu của bộ dữ liệu. Ngoài ra dữ liệu 2D không cung cấp các thông tin liên hệ giữa các dải tần số, do trong quá trình huấn luyện mạng sẽ phải cắt dữ liệu thành những lát cắt

riêng biệt cho từng dải tần số và học một cách riêng biệt. Ngược lại, dữ liệu 3D khi đưa vào huấn luyện cũng sẽ bao gồm toàn bộ các thông số như trong dữ liệu 2D (kích thước ma trận và các thông số thể hiện độ phức tạp), tuy nhiên dữ liệu 3D sẽ cung cấp thêm một trục z, trục này sẽ cung cấp thêm các thông số như mối liên hệ giữa các dải tần số khi được ghép với nhau. Trong quá trình huấn luyện, mạng LSTM sẽ xử lý dữ liệu đồng loạt cho 4 dải tần số mà không cần phải chia thành lát cắt nhỏ hơn để xử lý từng thành phần. Chúng tôi kết luận rằng, việc sử dụng tín hiệu dạng 3D sẽ giúp mô tả được chi tiết từng thành phần kết hợp trong tín hiệu, từ đó có thể mô tả chi tiết hơn về các dải tần số hoạt động và các mối quan hệ giữa chúng, đây là điều mà tín hiệu 2D khó có thể làm được (do các thành phần trong tín hiệu khi kết hợp chỉ được tính toán để cho ra một giá trị duy nhất).

So sánh phương pháp đề xuất với các nghiên cứu trước

Trong mục này, chúng tôi so sánh phương pháp đề xuất với các nghiên cứu có sử dụng bộ dữ liệu DEAP như nghiên cứu của Evi Septiana Pane và các cộng sự sử dụng phương pháp vec-tơ học máy (SVM) để phân loại trích xuất các tính năng đặc trưng²⁴. Tác giả Katarzyna Mrozik, và Bożena Kostek sử dụng phép biến đổi wavelet và phân loại theo phương pháp *k-nearest neighbors* và mạng nơ ron nhân tạo (ANN)²⁵. Nhà nghiên cứu Xiang Li đã sử dụng chín đặc trưng tuyến tính và ba đặc trưng phi tuyến được trích xuất từ các tín hiệu và dùng phương pháp *Random Forest* (RF) để phân loại trạng thái cảm xúc²⁶. Các kết quả được mô tả như Bảng 6.

Dựa vào Bảng 6, các kết quả cho thấy phương pháp đề xuất sử dụng tín hiệu ở dạng 3D và xử lý các tín hiệu trong chiều thời gian đã góp phần làm tăng hiệu quả của quá trình phân loại. Cụ thể hơn, phương pháp của chúng tôi thu được kết quả cao hơn phương pháp ANN khoảng 0,7%, kết quả này không thể xem là vượt trội, do cả hai phương pháp đều có cùng đặc tính huấn luyện giống nhau. Tuy nhiên, với thuật toán SVM thì phương pháp của chúng tôi cho ra hiệu quả cao hơn là 3,9%, và khi so với phương pháp *k-nearest neighbors* là 6,0%, tiếp đó là hiệu quả hơn 3,9% đối với phương pháp SVM, cuối cùng là cao hơn 3,8% so với phương pháp *Random Forest*.

Các kết quả so sánh thu được có thể được chứng minh thông qua các phương pháp xử lý trích xuất tín hiệu của các phương pháp, tác giả Evi Septiana Pane sử dụng 11 tính năng trong chiều thời gian để làm tính năng cho thuật toán SVM 24 và chỉ sử dụng một dải tần số duy nhất như đầu vào của thuật toán phân loại,

việc thiếu hụt việc sử dụng nhiều dải tần số đã làm giảm hiệu quả thực hiện do không thể mô tả chi tiết của các trạng thái cảm xúc. Đối với tác giả Xiang Li, tác giả sử dụng thuật toán *Random Forest* với năm phần dữ liệu chia bằng nhau cho việc học và kiểm tra do đó dữ liệu sẽ thay đổi khi so với bộ dữ liệu của chúng tôi (dữ liệu huấn luyện nhỏ hơn nhưng dữ liệu kiểm tra lại lớn hơn) điều này làm quá trình phân loại sẽ không hiệu quả do không đủ dữ liệu cho quá trình huấn luyện mô hình. Ngoài ra việc trích xuất tính năng chỉ có 12 giá trị cho mỗi kênh khi so với số tính năng của chúng tôi là thấp hơn, 32 tính năng cho mỗi kênh, do đó việc huấn luyện sẽ có ít các thông tin hơn để học. Tác giả Katarzyna Mrozik đã sử dụng dữ liệu có độ dài mười giây, độ dài lớn hơn rất nhiều so với dữ liệu của chúng tôi (0,5 giây), việc sử dụng dữ liệu có độ dài lớn khi trích xuất tính năng sẽ không thể lấy các giá trị chi tiết để có thể phân loại các trạng thái cảm xúc. Ngoài ra trong nghiên cứu của tác giả chỉ tập trung vào khảo sát việc trích xuất dữ liệu bằng phương pháp Wavelet và đưa vào các thuật toán phân loại khác nhau để kiểm tra, do đó, các tác giả không chú trọng vào việc thiết kế chi tiết các mô hình phân loại mà chỉ dùng các thông số mặc định. Ngoài ra, việc chia bộ dữ liệu thành 60% cho quá trình huấn luyện và 40% cho quá trình kiểm tra cũng sẽ gặp sự khó khăn như tác giả Xiang Li. Những đặc điểm đó đã dẫn đến việc phân loại sẽ không được tối ưu nhất.

KẾT LUẬN

Nghiên cứu này đã khẳng định được mạng LSTM có thể được sử dụng cho việc phân loại các trạng thái cảm xúc của con người. Trong bài nghiên cứu này, chúng tôi đã khảo sát được sự ảnh hưởng của các thông số có trong mạng bao gồm số epoch, tốc độ học cũng như số lớp, các thông số này có sự ảnh hưởng đáng kể đến quá trình phân loại của mô hình. Chúng tôi kết luận rằng với epoch bằng 10, tốc độ học bằng 10^{-4} và số lớp bằng hai là các thông số phù hợp cho mạng LSTM để phân loại được tín hiệu trong phương pháp của chúng tôi. Thêm vào đó, các kết quả nghiên cứu cũng chỉ ra rằng các đặc điểm dữ liệu khác nhau như độ dài phân đoạn cũng có ảnh hưởng lớn đến quá trình huấn luyện. Kết quả của tập dữ liệu với độ dài phân đoạn là 0,5 giây ("*input_1*") cho ra kết quả tốt nhất khi so sánh các tập dữ liệu với độ dài phân đoạn lớn hơn. Ngoài ra, mạng LSTM sử dụng dữ liệu ở dạng 3D cho ra kết quả tốt hơn so với dữ liệu ở dạng 1D và 2D. Các kết quả nghiên cứu còn được so sánh với các nghiên cứu khác, phương pháp đề xuất của chúng tôi cho thấy sự hiệu quả cao trong quá trình tín hiệu cảm xúc, đồng thời khẳng định việc sử dụng mối liên hệ giữa các tín hiệu EEG trong chiều thời gian sẽ cải thiện được quá trình phân loại.

Bảng 6: Bảng so sánh giữa phương pháp đề xuất với các nghiên cứu khác

Tác giả	Phương pháp	Độ chính xác (%)
Xiang Li	Random Forest	61,8
Evi Septiana Pane	SVM	61,6
Katarzyna Mrozik	k-nearest neighbors	59,4
Katarzyna Mrozik	ANN	64,7
Phương pháp đề xuất	LSTM	65,6

DANH MỤC CÁC TỪ VIẾT TẮT

1D: 1-dimension (1 chiều)

2D: 2-dimensions (2 chiều)

3D: 3-dimensions (3 chiều)

ANN: Artificial Neural Network (Mạng nơ-ron nhân tạo)

BVP: Blood Volume Pulse (Huyết áp)

DE: Differential entropy (Entropy vi sai)

EEG: Electroencephalogram (Điện não đồ)

EMG: Electromyography (Điện cơ)

EOG: ElectroOculogram (Điện cơ mắt)

ICA: Independent component analysis (Phân tích các thành phần độc lập)

LSTM: Long short term memory (Mạng bộ nhớ dài ngắn hạn)

RF: Random Forest (Rừng quyết định ngẫu nhiên)

RNN: Recurrent neural network (Mạng hồi quy)

SOM: Self-organizing map (Mạng tự tổ chức)

SVM: Support vector machine (Vec-tơ hỗ trợ học máy)

PCA: Principal component analysis (Phân tích các thành phần chính)

XUNG ĐỘT LỢI ÍCH

Các tác giả đồng ý không có bất kì xung đột lợi ích nào liên quan đến các kết quả công bố.

ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Tác giả Huỳnh Quốc Việt viết chương trình, tổng hợp bản thảo.

Các tác giả Nguyễn Thị Như Quỳnh, Trần Đức Minh, Lê Ngọc Ánh, Nguyễn Thanh Phước tham gia đo đạc và xử lý dữ liệu.

Tác giả Huỳnh Văn Tuấn tham gia viết và chỉnh sửa bản thảo.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP. Hồ Chí Minh thông qua đề tài mã số T2020-03.

TÀI LIỆU THAM KHẢO

- Anderson KMP. A real-time automated system for the recognition of human facial expressions. *IEEE Trans Syst Man Cybern Part B Cybern*;36(1):96–105. PMID: 16468569. Available from: <https://doi.org/10.1109/TSMCB.2005.854502>.
- El Ayadi M, Mohamed S. Kamel, and Fakhri Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*. 2017;4(3):575-587; Available from: <https://doi.org/10.1016/j.patcog.2010.09.020>.
- Petrushin V. Emotion in speech: recognition and application to call centers. *Proceedings of Artificial Neural Networks in Engineering*. 1999;710; Available from: https://doi.org/10.1007/978-3-319-68600-4_25.
- Soleymani MPMP. Multimodal emotion recognition in response to videos *IEEE Trans Affect Comput*. 2012;3(2):211-223; Available from: <https://doi.org/10.1109/T-AFFC.2011.37>.
- Sahayadhas KSMMP. Physiological signal based detection of driver hypovigilance using higher order spectra. *Expert Systems with Applications*. 2015;42(22):8669-8677; Available from: <https://doi.org/10.1016/j.eswa.2015.07.021>.
- Alex M, et al. Discrimination of Genuine and Acted Emotional Expressions Using EEG Signal and Machine Learning. *IEEE Access*. 2020;8:191080-191089; Available from: [10.1109/ACCESS.2020.3032380](https://doi.org/10.1109/ACCESS.2020.3032380).
- MA SUA, et al. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generation Computer Systems*. 2019;101:542-554; Available from: <https://doi.org/10.1016/j.future.2019.06.027>.
- Bhatti MH, et al. Soft Computing-Based EEG Classification by Optimal Feature Selection and Neural Networks. *IEEE Transactions on Industrial Informatics*. 2019;15(10):5747-5754; Available from: [10.1109/TII.2019.2925624](https://doi.org/10.1109/TII.2019.2925624).
- Rahman MM, et al. Sleep stage classification using single-channel EOG. *Computers in Biology and Medicine*. 2018;102; Available from: <https://doi.org/10.1016/j.compbiomed.2018.08.022>.
- Nakisa B, et al. Automatic Emotion Recognition Using Temporal Multimodal Deep Learning. *IEEE Access*. 2020; Available from: [10.1109/ACCESS.2020.3027026](https://doi.org/10.1109/ACCESS.2020.3027026).
- Wagner JK, Jonghwa; ANDRÉ, Elisabeth. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. 2005 IEEE international conference on multimedia and expo IEEE2005;p. 940–943. Available from: [10.1109/ICME.2005.1521579](https://doi.org/10.1109/ICME.2005.1521579).
- Gouizi K. Negative emotion detection using EMG signal. *International Conference on Control, Decision and Information Technologies (CoDIT)*; Metz. 2014;p. 690–695. Available from: <https://doi.org/10.1109/CoDIT.2014.6996980>.
- Zheng WLB. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans Auton Mental Dev*. 2015;7(3):162-175; Available from: <https://doi.org/10.1109/TAMD.2015.2431497>.
- Lakshmi MR, et al. Survey on EEG Signal. *International Journal of Advanced Research*. 2014;4(1):84-91;

15. Hatamikia S, and Ali MN. Recognition of emotional states induced by music videos based on nonlinear feature extraction and some classification 2014 21th Iranian Conference on Biomedical Engineering (ICBME). 2014;p. 333–337. Available from: <https://doi.org/10.1109/ICBME.2014.7043946>.
16. Kannathal N, et al. Characterization of EEG-a comparative study. Computer methods and Programs in Biomedicine. 2005;80(1):17-23;PMID: 16099533. Available from: <https://doi.org/10.1016/j.cmpb.2005.06.005>.
17. Koelstra S, et al. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing. 2011;3(1):18-31;Available from: <https://doi.org/10.1109/T-AFFC.2011.15>.
18. Yang Y, et al. Continuous Convolutional Neural Network with 3D Input for EEG-Based Emotion Recognition. The 25th International Conference on Neural Information Processing; Cambodia. 2018;p. 433–443. Available from: https://doi.org/10.1007/978-3-030-04239-4_39.
19. Elisabeth EA. Reasons, years and frequency of yoga practice: Effect on emotion response reactivity. Frontiers in human neuroscience. 2018;12(264);PMID: 30022932. Available from: <https://doi.org/10.3389/fnhum.2018.00264>.
20. Shi L-CJ, et al. Differential entropy feature for EEG-based vigilance estimation. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Osaka. 2013;p. 6627–6630. Available from: [10.1109/EMBC.2013.6611075](https://doi.org/10.1109/EMBC.2013.6611075).
21. Yong EA. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation. 2019;31(7):1235-1270;Available from: https://doi.org/10.1162/neco_a_01199.
22. Zhu LMaXX. A Recurrent Convolutional Neural Network for Land Cover Change Detection in Multispectral Images. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; Valencia. 2018;p. 4363–4366. Available from: [10.1109/IGARSS.2018.8517375](https://doi.org/10.1109/IGARSS.2018.8517375).
23. LI GS, et al. On comparing three artificial neural networks for wind speed forecasting. Applied Energy. 2010;87(7):2313-2320;Available from: <https://doi.org/10.1016/j.apenergy.2009.12.013>.
24. Evi S. Identifying rules for electroencephalograph (EEG) emotion recognition and classification. 2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME); Bandung. 2017;p. 167–172. Available from: [10.1109/ICICI-BME.2017.8537731](https://doi.org/10.1109/ICICI-BME.2017.8537731).
25. Mrozik K, et al. Comparison of selected electroencephalographic signal classification methods. 2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA); Poznan. 2017;p. 36–41. Available from: <https://doi.org/10.23919/SPA.2017.8166834>.
26. Li X, et al. Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Shenzhen. 2016;p. 352–359. Available from: [10.1109/BIBM.2016.7822545](https://doi.org/10.1109/BIBM.2016.7822545).

Application of long short term memory algorithm in classification electroencephalogram

Viet Quoc Huynh^{1,2}, Quynh Nguyen-Thi-Nhu^{1,2}, Minh Duc Tran^{1,2}, Anh Ngoc Le^{1,2}, Phuoc Thanh Nguyen^{1,2}, Tuan Van Huynh^{1,2,*}



Use your smartphone to scan this QR code and download this article

ABSTRACT

Human emotion plays an important role in communication without language, and it also supports research on human behavior. In addition, electroencephalogram signals have been highly confirmed by researchers for reliability as well as ease of storage and recognition. So, the use of electroencephalogram to identify emotion signals are currently a relatively new field. Many researchers are targeting the key ideas in this research field such as signal preprocessing, feature extraction and algorithm optimization. In this paper, we aim to recognize emotion signals using Long Short Term Memory (LSTM) algorithms. Emotional signals dataset was taken from DEAP database of koelstra authors and associates to serve this research. The research will focus on accuracy and training time, and it will test different architectural types as well as the initials of LSTM. The obtained results show the 3-dimensional cubes's structure has better performance than the 2-dimensional cubes's structure. In addition, our research is also compared with other authors' studies to prove the effectiveness of the classification algorithm.

Key words: electroencephalogram, emotion, LSTM, 2D data, 3D data

¹Faculty of Physics and Engineering Physics, University of Science, VNU-HCM, Vietnam

²Viet Nam National University Ho Chi Minh city, Vietnam

Correspondence

Tuan Van Huynh, Faculty of Physics and Engineering Physics, University of Science, VNU-HCM, Vietnam

Viet Nam National University Ho Chi Minh city, Vietnam

Email: hvtuan@hcmus.edu.vn

History

- Received: 30-12-2020
- Accepted: 25-3-2021
- Published: 30-4-2021

DOI : 10.32508/stdjns.v5i2.1006



Copyright

© VNU-HCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



Cite this article : Huynh V Q, Nguyen-Thi-Nhu Q, Tran M D, Le A N, Nguyen P T, Huynh T V. **Application of long short term memory algorithm in classification electroencephalogram.** *Sci. Tech. Dev. J. - Nat. Sci.*; 5(2):1167-1178.