

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**DƯƠNG THỊ MAI THƯƠNG**

**NGHIÊN CỨU VÀ PHÁT TRIỂN MÔ HÌNH HỌC SÂU  
CHO NHẬN DẠNG CẢM XÚC TỪ TÍN HIỆU ĐIỆN NÃO  
TRONG ĐIỀU KIỆN TÀI NGUYÊN HẠN CHẾ**

**LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH**

**THÁI NGUYÊN - NĂM 2025**

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**DƯƠNG THỊ MAI THƯƠNG**

**NGHIÊN CỨU VÀ PHÁT TRIỂN MÔ HÌNH HỌC SÂU  
CHO NHẬN DẠNG CẢM XÚC TỪ TÍN HIỆU ĐIỆN NÃO  
TRONG ĐIỀU KIỆN TÀI NGUYÊN HẠN CHẾ**

**Ngành: Khoa học máy tính  
Mã số: 9.48.01.01**

**LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**PGS.TS. PHÙNG TRUNG NGHĨA**

**THÁI NGUYÊN - NĂM 2025**

## LỜI CAM ĐOAN

Tôi xin cam đoan các kết quả trình bày trong luận án Tiến sĩ “*Nghiên cứu và phát triển mô hình học sâu cho nhận dạng cảm xúc từ tín hiệu điện não trong điều kiện tài nguyên hạn chế*” là các công trình nghiên cứu của tôi dưới sự hướng dẫn của PGS. TS. Phùng Trung Nghĩa, trừ những kiến thức tham khảo từ các tài liệu đã được tham chiếu rõ ràng.

Các kết quả nghiên cứu trong luận án là trung thực, một phần đã được công bố trên các Tạp chí, Hội thảo khoa học (danh sách các công trình được liệt kê tại cuối Luận án), phần còn lại chưa được công bố trong bất kỳ công trình nào khác.

Mọi nội dung dữ liệu được tham khảo trong luận án đều được trích dẫn đầy đủ và đúng quy định.

Thái Nguyên, ngày .... tháng .... năm 2025

Tác giả luận án

**Dương Thị Mai Thương**

# LỜI CẢM ƠN

Trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án, tôi luôn nhận được sự hướng dẫn tận tình, góp ý quý báu và động viên từ các Thầy Cô, các Nhà khoa học, các Cơ quan, Đoàn thể và Gia đình. Với lòng biết ơn sâu sắc, tôi xin gửi lời cảm ơn chân thành đến:

Ban lãnh đạo Trường Đại học Công nghệ Thông tin và Truyền thông Thái Nguyên, cùng Ban lãnh đạo Khoa Công nghệ Thông tin và các Thầy Cô, Cán bộ quản lý của trường. Sự quan tâm và hỗ trợ quý báu về thời gian và tài chính đã tạo điều kiện thuận lợi để tôi hoàn thành nhiệm vụ nghiên cứu.

Đặc biệt, tôi xin tri ân sâu sắc đến PGS.TS. Phùng Trung Nghĩa, người đã luôn tận tâm hướng dẫn và động viên tôi trong suốt hành trình học tập và nghiên cứu.

Tôi xin cảm ơn Khoa Công nghệ Thông tin, Bộ môn Hệ thống thông tin, cùng các thầy cô, đồng nghiệp đã tạo điều kiện thuận lợi để tôi hoàn thành quá trình học tập và nghiên cứu.

Cuối cùng, tôi dành lòng biết ơn sâu sắc đến gia đình, bố mẹ, chồng con và những người thân đã luôn đồng hành, ủng hộ và chia sẻ để tôi có thể tập trung thực hiện nhiệm vụ học tập và nghiên cứu.

Mặc dù đã nỗ lực hết mình, tôi nhận thấy luận án vẫn còn những hạn chế về chuyên môn và cách trình bày. Tôi mong nhận được những góp ý quý báu từ các Thầy Cô và các Nhà khoa học để tiếp tục hoàn thiện hơn.

Xin chân thành biết ơn!

**NCS. Dương Thị Mai Thương**



# Mục lục

|                                                                                       |             |
|---------------------------------------------------------------------------------------|-------------|
| <b>LỜI CAM ĐOAN</b>                                                                   | <b>i</b>    |
| <b>LỜI CẢM ƠN</b>                                                                     | <b>ii</b>   |
| <b>DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT</b>                                           | <b>iv</b>   |
| <b>DANH MỤC HÌNH ẢNH</b>                                                              | <b>vii</b>  |
| <b>DANH MỤC BẢNG BIỂU</b>                                                             | <b>viii</b> |
| <b>MỞ ĐẦU</b>                                                                         | <b>1</b>    |
| 1. Tính cấp thiết của đề tài . . . . .                                                | 1           |
| 2. Mục tiêu của luận án . . . . .                                                     | 4           |
| 3. Đối tượng, phạm vi và phương pháp nghiên cứu . . . . .                             | 5           |
| 4. Các đóng góp của luận án . . . . .                                                 | 5           |
| 5. Ý nghĩa khoa học và thực tiễn . . . . .                                            | 6           |
| 6. Bố cục của luận án . . . . .                                                       | 7           |
| <b>CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG CẢM XÚC DỰA TRÊN TÍN HIỆU ĐIỆN NÃO</b>   | <b>9</b>    |
| 1.1 Tổng quan về bài toán nhận dạng cảm xúc . . . . .                                 | 9           |
| 1.1.1 Cảm xúc là gì? . . . . .                                                        | 9           |
| 1.1.2 Tầm quan trọng của việc nhận dạng cảm xúc trong giao tiếp người - máy . . . . . | 9           |
| 1.1.3 Các mô hình cảm xúc . . . . .                                                   | 10          |
| 1.2 Nhận dạng cảm xúc dựa trên tín hiệu điện não đồ EEG . . . . .                     | 12          |
| 1.2.1 Giới thiệu về EEG . . . . .                                                     | 12          |
| 1.2.2 Hệ thống nhận dạng cảm xúc dựa trên tín hiệu điện não . . . . .                 | 15          |
| 1.3 Thu nhận và xử lý số tín hiệu EEG . . . . .                                       | 16          |
| 1.3.1 Kích thích cảm xúc . . . . .                                                    | 16          |
| 1.3.2 Thu nhận và xử lý tín hiệu . . . . .                                            | 18          |
| 1.3.3 Một số bộ dữ liệu cho bài toán <i>EEG_ER</i> . . . . .                          | 21          |
| 1.4 Trích xuất đặc trưng tín hiệu EEG . . . . .                                       | 24          |
| 1.4.1 Trích xuất miền thời gian . . . . .                                             | 24          |
| 1.4.2 Trích xuất miền tần số . . . . .                                                | 24          |
| 1.4.3 Trích xuất miền thời gian–tần số . . . . .                                      | 25          |
| 1.5 Phân lớp tín hiệu EEG trong bài toán nhận dạng cảm xúc . . . . .                  | 26          |
| 1.5.1 Các kỹ thuật học máy truyền thống . . . . .                                     | 26          |

|       |                                                                       |    |
|-------|-----------------------------------------------------------------------|----|
| 1.5.2 | Các kỹ thuật học sâu . . . . .                                        | 27 |
| 1.6   | Phát triển các hệ thống nhận dạng tín hiệu EEG theo định hướng AIoT . | 32 |
| 1.7   | Kết luận chương 1 . . . . .                                           | 35 |

## **CHƯƠNG 2. LỰA CHỌN KIẾN TRÚC HỌC SÂU NỀN TẢNG VÀ ĐẶC TRƯNG ĐẦU VÀO PHÙ HỢP CHO BÀI TOÁN EEG-ER** **37**

|       |                                                                                  |    |
|-------|----------------------------------------------------------------------------------|----|
| 2.1   | Biểu diễn và chuẩn hóa dữ liệu . . . . .                                         | 37 |
| 2.1.1 | Biểu diễn tín hiệu EEG dưới dạng ma trận . . . . .                               | 38 |
| 2.1.2 | Biểu diễn tín hiệu EEG theo batch và cửa sổ thời gian . . . . .                  | 39 |
| 2.2   | Lựa chọn các phương pháp trích chọn đặc trưng . . . . .                          | 40 |
| 2.2.1 | <b>Cơ sở lựa chọn các phương pháp trích chọn đặc trưng</b> . . . . .             | 40 |
| 2.2.2 | Trích xuất đặc trưng bằng biến đổi Fourier nhanh (FFT) . . . . .                 | 41 |
| 2.2.3 | Trích xuất năng lượng tín hiệu EEG bằng Mật độ phổ công suất (PSD) . . . . .     | 43 |
| 2.2.4 | Trích xuất đặc trưng thời gian - tần số EEG bằng Biến đổi Wavelet (WT) . . . . . | 44 |
| 2.3   | Lựa chọn kiến trúc mô hình nền tảng . . . . .                                    | 45 |
| 2.3.1 | Cơ sở lựa chọn mô hình học sâu . . . . .                                         | 45 |
| 2.3.2 | Mạng CNN . . . . .                                                               | 46 |
| 2.3.3 | Mạng BiLSTM . . . . .                                                            | 48 |
| 2.3.4 | Mô hình kết hợp CNN và BiLSTM . . . . .                                          | 49 |
| 2.4   | Tiến hành thực nghiệm và đánh giá kết quả . . . . .                              | 51 |
| 2.4.1 | Thiết lập thí nghiệm . . . . .                                                   | 51 |
| 2.4.2 | Thực nghiệm và đánh giá kết quả . . . . .                                        | 55 |
| 2.5   | Kết luận chương 2 . . . . .                                                      | 56 |

## **CHƯƠNG 3. PHÁT TRIỂN MÔ HÌNH HỌC SÂU NHẸ CHO HỆ THỐNG EEG\_ER** **58**

|       |                                                                                                         |    |
|-------|---------------------------------------------------------------------------------------------------------|----|
| 3.1   | Phân tích yêu cầu khi triển khai bài toán EEG_ER trên thiết bị có tài nguyên hạn chế . . . . .          | 58 |
| 3.1.1 | Giới hạn về tài nguyên tính toán trên phần cứng thực hiện tác vụ nhận dạng . . . . .                    | 59 |
| 3.1.2 | Hạn chế về số lượng điện cực EEG trên thiết bị thu thập tín hiệu EEG có khả năng di động được . . . . . | 60 |
| 3.2   | Các kỹ thuật tối ưu hóa mô hình học sâu . . . . .                                                       | 62 |
| 3.2.1 | Kỹ thuật giảm overfitting và cải thiện tổng quát hóa . . . . .                                          | 63 |
| 3.2.2 | Kỹ thuật cải thiện khả năng học sâu và hội tụ nhanh . . . . .                                           | 66 |
| 3.2.3 | Nâng cao khả năng học đặc trưng trong EEG_ER . . . . .                                                  | 67 |
| 3.3   | Đề xuất mô hình học sâu nhẹ cho bài toán EEG_ER . . . . .                                               | 71 |

|                                                                                      |                                                                                                                                |            |
|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.3.1                                                                                | Mô tả chung . . . . .                                                                                                          | 71         |
| 3.3.2                                                                                | Kiến trúc chi tiết . . . . .                                                                                                   | 71         |
| 3.3.3                                                                                | Đánh giá cấu trúc và hiệu suất thực nghiệm của EEG_MCISNet,<br>EEG_SICNet . . . . .                                            | 75         |
| 3.4                                                                                  | Kết luận chương 3 . . . . .                                                                                                    | 85         |
| <b>CHƯƠNG 4. TỐI ƯU HÓA MÔ HÌNH EEG_ER SỬ DỤNG KỸ THUẬT<br/>TỰ CHẮT LỌC TRI THỨC</b> |                                                                                                                                | <b>86</b>  |
| 4.1                                                                                  | Tổng quan về kỹ thuật chặt lọc tri thức và kỹ thuật tự chặt lọc tri thức . .                                                   | 87         |
| 4.1.1                                                                                | Chặt lọc tri thức . . . . .                                                                                                    | 87         |
| 4.1.2                                                                                | Tự chặt lọc tri thức . . . . .                                                                                                 | 88         |
| 4.2                                                                                  | Kỹ thuật tối ưu hóa mô hình EEG-ER sử dụng SKD và Scaled Weights .                                                             | 88         |
| 4.2.1                                                                                | Thiết lập bài toán . . . . .                                                                                                   | 89         |
| 4.2.2                                                                                | Đề xuất kiến trúc EEG-SKNet . . . . .                                                                                          | 89         |
| 4.2.3                                                                                | Trọng số có điều chỉnh (Scaled-Weights) . . . . .                                                                              | 92         |
| 4.2.4                                                                                | Chứng minh toán học về tác động của hệ số nhiệt và trọng số có<br>điều chỉnh lên phân phối xác suất đầu ra trong EEG-SKNet . . | 93         |
| 4.2.5                                                                                | Thiết lập kiến trúc mạng . . . . .                                                                                             | 95         |
| 4.3                                                                                  | Thực nghiệm . . . . .                                                                                                          | 97         |
| 4.3.1                                                                                | Bộ dữ liệu . . . . .                                                                                                           | 97         |
| 4.3.2                                                                                | Chi tiết triển khai . . . . .                                                                                                  | 97         |
| 4.3.3                                                                                | So sánh và đánh giá với phương pháp cơ sở . . . . .                                                                            | 98         |
| 4.3.4                                                                                | So sánh với các phương pháp tiên tiến (SOTA) . . . . .                                                                         | 100        |
| 4.3.5                                                                                | Nghiên cứu loại bỏ (Ablation Study) . . . . .                                                                                  | 103        |
| 4.4                                                                                  | Kết luận chương 4 . . . . .                                                                                                    | 104        |
| <b>KẾT LUẬN VÀ NHỮNG KIẾN NGHỊ NGHIÊN CỨU TIẾP THEO</b>                              |                                                                                                                                | <b>105</b> |
| <b>DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC</b>                                              |                                                                                                                                | <b>107</b> |
| <b>DANH MỤC TÀI LIỆU THAM KHẢO</b>                                                   |                                                                                                                                | <b>107</b> |

## DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT

| <b>Thuật ngữ</b> | <b>Diễn giải tiếng Anh</b>             | <b>Diễn giải tiếng Việt</b>                        |
|------------------|----------------------------------------|----------------------------------------------------|
| AI               | Artificial Intelligence                | Trí tuệ nhân tạo                                   |
| AC               | Affective Computing                    | Tính toán ảnh hưởng                                |
| BCI              | Brain-Computer Interface               | Giao diện não - máy tính                           |
| CNN              | Convolutional Neural Networks          | Mạng nơ-ron tích chập                              |
| EEG              | Electroencephalogram                   | Điện não đồ                                        |
| EEG-ER           | EEG based Emotion Recognition          | Nhận dạng cảm xúc dựa trên tín hiệu điện não       |
| ER               | Emotion Recognition                    | Nhận dạng cảm xúc                                  |
| FLOPs            | FLoating-point Operations Per Second   | Số phép tính dấu phẩy động trên giây               |
| GNN              | Graph Neural Networks                  | Mạng nơ ron đồ thị                                 |
| GRU              | Gated Recurrent Unit                   | Bộ nhớ hồi quy có cổng                             |
| GSR              | Galvanic Skin Response                 | Phản ứng điện da                                   |
| HANV             | High Arousal Negative Valence          | Trạng thái cảm xúc kích thích cao và valence âm    |
| HAPV             | High Arousal Positive Valence          | Trạng thái cảm xúc kích thích cao và valence dương |
| HCI              | Human-Computer Interaction             | Tương tác Người - Máy                              |
| HOC              | Higher Order Crossing                  | Kỹ thuật cắt trục không                            |
| HRV              | Heart Rate Variability                 | Biến thiên nhịp tim                                |
| IADS             | International Affective Digital Sounds | Thư viện âm thanh cảm xúc Quốc tế                  |
| IAPS             | International Affective Picture System | Hệ thống hình ảnh cảm xúc Quốc tế                  |
| ICA              | Independent Component Analysis         | Phân tích thành phần độc lập                       |

| <b>Thuật ngữ</b> | <b>Diễn giải tiếng Anh</b>              | <b>Diễn giải tiếng Việt</b>                         |
|------------------|-----------------------------------------|-----------------------------------------------------|
| KD               | Knowledge Distillation                  | Chất lọc tri thức                                   |
| KNN              | K-Nearest Neighbors                     | Phương pháp láng giềng gần nhất                     |
| LANV             | Low Arousal Negative Valence            | Trạng thái cảm xúc kích thích thấp và valence âm    |
| LAPV             | Low Arousal Positive Valence            | Trạng thái cảm xúc kích thích thấp và valence dương |
| LDA              | Linear Discriminant Analysis            | Phân tích phân biệt tuyến tính                      |
| LSTM             | Long Short-Term Memory                  | Bộ nhớ ngắn dài hạn                                 |
| MAHNOB-HCI       | Multimodal Emotion Recognition Dataset  | Bộ dữ liệu nhận dạng cảm xúc đa phương thức         |
| MLP              | Multilayer Perceptron                   | Mạng nơ-ron đa tầng                                 |
| MT-KD            | Multi-Teacher Knowledge Distillation    | Chất lọc tri thức đa giáo viên                      |
| NB               | Naïve Bayes                             | Thuật toán Naïve Bayes                              |
| P300             | Positive deflection at 300ms            | Sóng P300 trong ERP                                 |
| PCA              | Principal Component Analysis            | Phân tích thành phần chính                          |
| PPG              | Photoplethysmography                    | Quang phổ kế                                        |
| PS-KD            | Progressive Self-Knowledge Distillation | Chất lọc tri thức tự cải tiến                       |
| PSD              | Power Spectral Density                  | Mật độ phổ công suất                                |
| RASM             | Rational Asymmetry                      | Bất đối xứng tỷ lệ                                  |
| RF               | Random Forest                           | Rừng ngẫu nhiên                                     |
| RMS              | Root Mean Square                        | Giá trị trung bình bình phương                      |
| RNN              | Recurrent Neural Network                | Mạng nơ-ron hồi quy                                 |
| ReLU             | Rectified Linear Unit                   | Hàm kích hoạt ReLU                                  |
| SAM              | Self-Assessment Manikin                 | Thang đo cảm xúc tự đánh giá                        |
| SE               | Squeeze-and-Excitation                  | Cấu trúc SE Block                                   |
| SEED             | SJTU Emotion EEG Dataset                | Bộ dữ liệu EEG cảm xúc SJTU                         |

| <b>Thuật ngữ</b> | <b>Diễn giải tiếng Anh</b>       | <b>Diễn giải tiếng Việt</b>           |
|------------------|----------------------------------|---------------------------------------|
| SEED-IV          | Extended version of SEED dataset | Phiên bản mở rộng của bộ dữ liệu SEED |
| SKD              | Self-Knowledge Distillation      | Tự chắt lọc tri thức                  |
| STFT             | Short-Time Fourier Transform     | Biến đổi Fourier ngắn hạn             |
| SVM              | Support Vector Machine           | Máy vector hỗ trợ                     |
| SW               | Scaled Weights                   | Trọng số co giãn                      |
| VR               | Virtual Reality                  | Thực tế ảo                            |
| WT               | Wavelet Transform                | Biến đổi Wavelet                      |

# DANH MỤC HÌNH ẢNH

|          |                                                                                                                                       |     |
|----------|---------------------------------------------------------------------------------------------------------------------------------------|-----|
| Hình 0.1 | Các bước quan trọng của bài toán EEG_ER . . . . .                                                                                     | 2   |
| Hình 0.2 | Hệ thống EEG_ER theo tiếp cận AIoT . . . . .                                                                                          | 3   |
| Hình 1.1 | Mô hình cảm xúc 2D Russell . . . . .                                                                                                  | 11  |
| Hình 1.2 | Mô hình cảm xúc Schloberg . . . . .                                                                                                   | 11  |
| Hình 1.3 | Thiết lập ghi tín hiệu EEG . . . . .                                                                                                  | 12  |
| Hình 1.4 | Nhận dạng các dạng sóng theo tần số . . . . .                                                                                         | 13  |
| Hình 1.5 | Bản đồ phân bố điện cực . . . . .                                                                                                     | 20  |
| Hình 2.1 | Kiến trúc mạng 1D-CNN [51] . . . . .                                                                                                  | 46  |
| Hình 2.2 | Mạng BiLSTM - Bidirectional Long Short-Term Memory [26] . .                                                                           | 49  |
| Hình 2.3 | Cấu trúc của mô hình kết hợp CNN với BiLSTM [43] . . . . .                                                                            | 50  |
| Hình 2.4 | Bộ dữ liệu DEAP [60] . . . . .                                                                                                        | 51  |
| Hình 2.5 | Các bước thực hiện trích chọn đặc trưng . . . . .                                                                                     | 52  |
| Hình 2.6 | Kiến trúc mạng CNN . . . . .                                                                                                          | 53  |
| Hình 2.7 | Kiến trúc mạng Bi-LSTM . . . . .                                                                                                      | 54  |
| Hình 2.8 | Kiến trúc mạng kết hợp CNN-BiLSTM . . . . .                                                                                           | 54  |
| Hình 3.1 | Cấu trúc khối SE [45]. . . . .                                                                                                        | 68  |
| Hình 3.2 | Cấu trúc khối Inception [105]. . . . .                                                                                                | 69  |
| Hình 3.3 | Kiến trúc chung của hai mô hình đề xuất . . . . .                                                                                     | 72  |
| Hình 3.4 | Biểu đồ quá trình học của mạng EEG_SICNet . . . . .                                                                                   | 77  |
| Hình 3.5 | Biểu đồ quá trình học khi huấn luyện mạng EEG_MCISNet với các<br>nhãn Valence, Arousal và Dominance với bộ dữ liệu EEG gồm 5 điện cực | 84  |
| Hình 4.1 | Tổng quan về phương pháp được đề xuất . . . . .                                                                                       | 90  |
| Hình 4.2 | Mạng nhẹ được sử dụng trong EEG-SKDNet . . . . .                                                                                      | 96  |
| Hình 4.3 | Minh hoạ sự hội tụ của EEG-SKDNet và phương pháp cơ sở trên<br>tập dữ liệu DEAP(32). . . . .                                          | 100 |

## DANH MỤC BẢNG BIỂU

|          |                                                                                                                                                                   |     |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Bảng 1.1 | Một số thiết bị thu thập tín hiệu EEG thông dụng . . . . .                                                                                                        | 18  |
| Bảng 1.2 | Một số bộ dữ liệu mẫu được sử dụng cho bài toán EEG_ER . . . . .                                                                                                  | 23  |
| Bảng 1.3 | Độ chính xác phân lớp của một số thuật toán học sâu . . . . .                                                                                                     | 31  |
| Bảng 1.4 | Một số vi điều khiển được sử dụng làm thiết bị biên trong mạng<br>AIoT . . . . .                                                                                  | 34  |
| Bảng 2.1 | Bảng so sánh các phương pháp trích xuất đặc trưng EEG . . . . .                                                                                                   | 41  |
| Bảng 2.2 | Đánh giá hiệu quả hoạt động của các kiến trúc mạng . . . . .                                                                                                      | 55  |
| Bảng 2.3 | So sánh hiệu quả nhận dạng của mạng CNN đề xuất với một số<br>mạng CNN tiêu biểu . . . . .                                                                        | 56  |
| Bảng 3.1 | Hiệu suất mạng EEG_SICNet trên các bộ dữ liệu EEG . . . . .                                                                                                       | 76  |
| Bảng 3.2 | Kết quả hoạt động của mạng EEG_SICNet với đầu vào là 5 điện cực                                                                                                   | 78  |
| Bảng 3.3 | So sánh hiệu suất của mạng EEG_SICNet với các phương pháp<br>khác trên bộ dữ liệu DEAP (32 điện cực). . . . .                                                     | 79  |
| Bảng 3.4 | Kết quả nhận dạng của thực tế của mô hình thử nghiệm . . . . .                                                                                                    | 81  |
| Bảng 3.5 | Độ chính xác (%) khi đánh giá theo băng tần trên các bộ dữ liệu . . . . .                                                                                         | 82  |
| Bảng 3.6 | Độ chính xác của mô hình EEG_MCISNet trên các bộ dữ liệu . . . . .                                                                                                | 83  |
| Bảng 3.7 | Độ chính xác của các mô hình học sâu trên bộ dữ liệu DEAP . . . . .                                                                                               | 84  |
| Bảng 4.1 | Chi tiết kiến trúc mạng nhẹ trong phương pháp đề xuất . . . . .                                                                                                   | 96  |
| Bảng 4.2 | So sánh giữa các kịch bản KD khác nhau . . . . .                                                                                                                  | 98  |
| Bảng 4.3 | So sánh độ chính xác top-1 giữa mô hình cơ sở và EEG-SKDNet<br>trên các thành phần cảm xúc trong ba bộ dữ liệu: DEAP(32), DEAP(14),<br>và DREAMER. . . . .        | 100 |
| Bảng 4.4 | So sánh hiệu suất giữa các phương pháp tiên tiến trên các tập dữ<br>liệu phổ biến. . . . .                                                                        | 101 |
| Bảng 4.5 | Độ chính xác Top-1 với các cấu hình khác nhau của hai siêu tham<br>số $\alpha$ và $\tau$ trong mô hình EEG-SKDNet trên tập dữ liệu DEAP (32 điện<br>cực). . . . . | 103 |



# MỞ ĐẦU

## 1. Tính cấp thiết của đề tài

Ngày nay, với sự tiến bộ của công nghệ, rất nhiều hệ thống máy móc thông minh được ra đời với mục tiêu hỗ trợ con người nâng cao chất lượng cuộc sống. Điều này góp phần hình thành nên một lĩnh vực nghiên cứu quan trọng, đó là tương tác giữa người và máy (Human-Computer Interaction – HCI) [81, 72].

HCI là một lĩnh vực nghiên cứu liên ngành, đề cập đến việc trao đổi thông tin giữa con người và máy tính, bao gồm cả việc máy tính cung cấp thông tin cho con người thông qua các thiết bị đầu ra và việc con người nhập thông tin liên quan vào máy tính qua các thiết bị đầu vào [72]. Nhờ những đóng góp to lớn của trí tuệ nhân tạo (Artificial Intelligence - AI), rất nhiều nghiên cứu đã kết hợp được giọng nói, hình ảnh, văn bản và các loại thông tin khác để thúc đẩy sự phát triển của HCI, làm cho HCI được ứng dụng thành công trong rất nhiều lĩnh vực như chăm sóc sức khỏe, giáo dục và công nghiệp...[81].

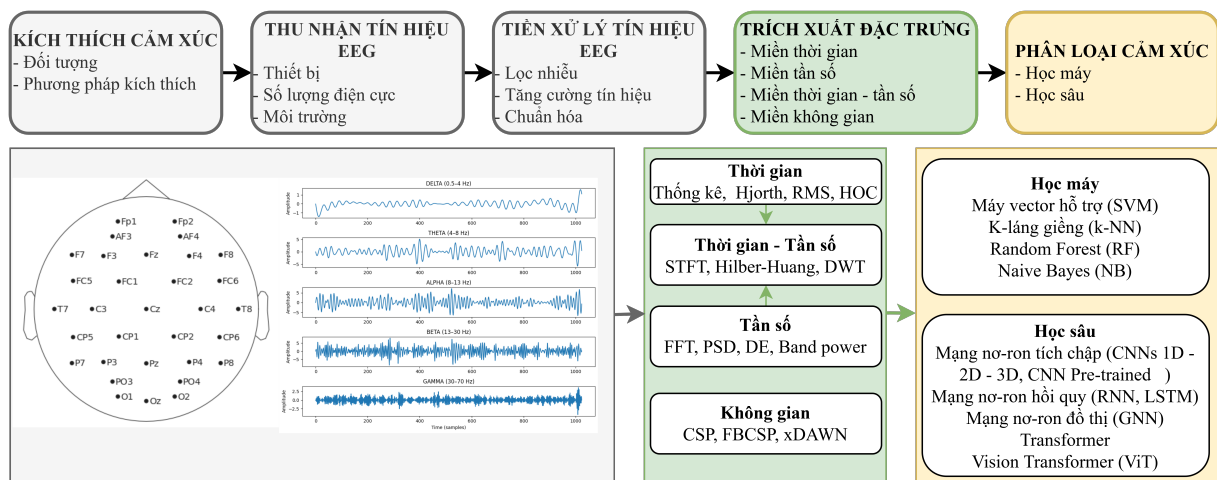
Tuy nhiên, hầu hết các hệ thống HCI hiện tại vẫn chưa có khả năng xử lý và hiểu được dữ liệu về cảm xúc, thiếu tính thông minh cảm xúc. Chúng không thể nhận biết được cảm xúc của con người và sử dụng dữ liệu cảm xúc đó để ra quyết định và hành động. Trong khi đó, việc giải quyết sự thiếu hụt mối quan hệ giữa con người và máy móc là rất quan trọng. Bất kỳ hệ thống HCI nào bỏ qua trạng thái cảm xúc của con người đều sẽ không thể phản ứng phù hợp với những cảm xúc đó. Để khắc phục điều này, rất nhiều nhà khoa học trên thế giới đã và đang quan tâm đến việc giải quyết bài toán nhận dạng cảm xúc (Emotion Recognition - ER) khi xây dựng các hệ thống HCI. Việc được trang bị thêm khả năng ER một cách chính xác có thể giúp các hệ thống HCI trở nên thông minh hơn, ra quyết định chính xác hơn, đáp ứng tốt hơn nhu cầu của con người trong đời sống thực tế. [107, 57].

Cũng như con người, cảm xúc có thể được các hệ thống HCI hiểu được thông qua nét mặt, tiếng nói, tư thế cơ thể (các tín hiệu phi sinh lý) [91, 38]. Tuy nhiên, có rất nhiều yếu tố ảnh hưởng đến độ chính xác của các phương pháp này như trạng thái tâm trí, giới tính, trình độ học vấn, độ tuổi và khả năng che giấu cảm xúc. Từ đó dẫn tới việc đánh giá cảm xúc khi sử dụng các phương pháp này trở nên khó khăn, thiếu chính xác do mang nặng yếu tố chủ quan.

Bên cạnh đó, một số nghiên cứu đã tìm thấy mối liên hệ chặt chẽ giữa các tín hiệu sinh lý và biểu hiện cảm xúc. Nhờ đó, một số phương pháp ER sử dụng tín hiệu sinh lý như nhịp tim, trở kháng của da, hô hấp và điện não đồ (Electroencephalogram – EEG)

đã được đề xuất [32]. Trong đó, việc nghiên cứu các cơ chế nhận thức cảm xúc sử dụng tín hiệu EEG đặc biệt quan trọng, nhất là đối với những người khuyết tật không có khả năng diễn đạt cảm xúc của mình thông qua lời nói, nét mặt hoặc tư thế cơ thể.

EEG có thể ghi lại các tín hiệu điện được tạo ra bởi các hoạt động bên trong của não bộ, đưa ra phép đo trực tiếp và không xâm lấn về chức năng não. Tín hiệu này phản ánh các quá trình nhận thức và tình cảm liên quan đến các trạng thái cảm xúc khác nhau. Điều này làm cho việc nhận dạng cảm xúc dựa trên tín hiệu điện não (EEG based Emotion Recognition – EEG\_ER) trở thành một con đường đầy hứa hẹn để phân tích cảm xúc một cách chính xác hơn [113].

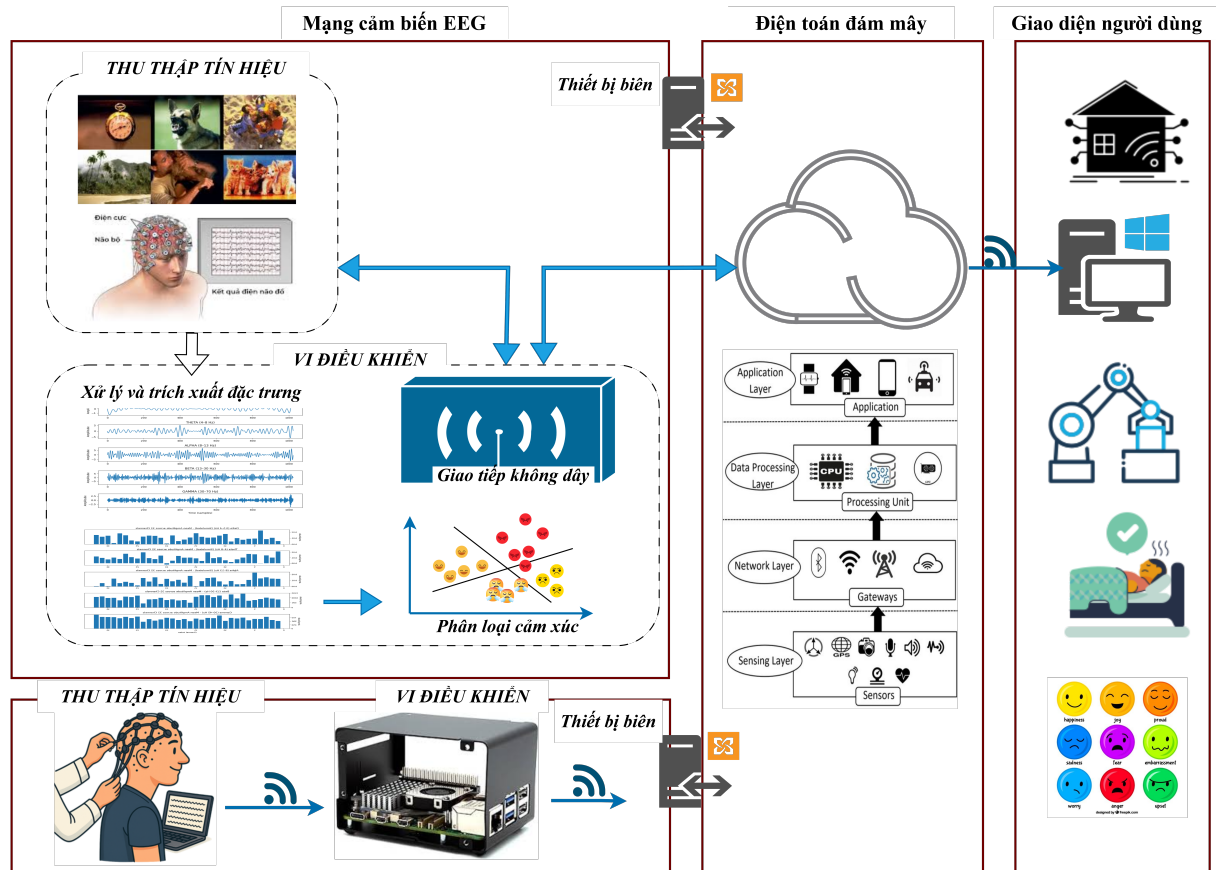


**Hình 0.1** Các bước quan trọng của bài toán EEG\_ER

Hình 0.1 mô tả một hệ thống EEG\_ER điển hình. Hệ thống này bao gồm năm khâu quan trọng bao gồm kích thích cảm xúc, thu nhận tín hiệu EEG, tiền xử lý tín hiệu EEG, trích xuất các đặc trưng EEG liên quan đến cảm xúc và cuối cùng là huấn luyện mô hình nhận dạng cảm xúc [109]. Qua đó có thể thấy rằng, hiệu quả hoạt động của một hệ thống EEG\_ER phụ thuộc vào rất nhiều yếu tố như thiết bị thu thập tín hiệu EEG, phương pháp trích xuất đặc trưng, kỹ thuật phân lớp cảm xúc.

Về thiết bị thu thập tín hiệu EEG, trên thực tế có thể phân ra hai loại thiết bị. Thiết bị có số lượng điện cực EEG lớn (Biosemi ActiveTwo, ESI NeuroScan System) và thiết bị có số lượng điện cực EEG nhỏ (Emotiv EPOC) [109, 66]. Các thiết bị có số lượng điện cực EEG lớn thường có độ chính xác cao, đắt tiền, sử dụng chủ yếu trong các phòng thí nghiệm và không thuận tiện cho việc di chuyển đi lại. Trái lại, các thiết bị có số lượng điện cực nhỏ có thể đeo được, di động, không dây và được thương mại nên chi phí phù hợp thì lại cho độ chính xác thấp hơn [66]. Do vậy, *khi triển khai ứng dụng thực tế, cần quan tâm xây dựng mô hình EEG\_ER sao cho mô hình này vẫn hoạt động tốt với số lượng điện cực EEG được sử dụng là thấp.*

Hiệu quả hoạt động của hệ thống EEG\_ER còn phụ thuộc rất nhiều vào các đặc trưng của tín hiệu EEG đầu vào. Từ dữ liệu EEG thô ban đầu, rất nhiều phương pháp trích, chọn đặc trưng được đề xuất bao gồm các đặc trưng về thời gian, đặc trưng về tần số hoặc là kết hợp cả đặc trưng về thời gian tần số [44, 114]. Do vậy, việc lựa chọn đặc trưng phù hợp sẽ góp phần nâng cao hiệu suất nhận dạng của các hệ thống EEG\_ER.



**Hình 0.2** Hệ thống EEG\_ER theo tiếp cận AIoT

Về kỹ thuật phân lớp cảm xúc, các phương pháp EEG\_ER trước đây chủ yếu được phát triển dựa trên học máy truyền thống [126]. Gần đây, lấy cảm hứng từ khả năng học của các kỹ thuật học sâu, nhiều mô hình học sâu tiên tiến khác nhau đã được sử dụng ngày càng nhiều trong nhận dạng cảm xúc EEG [119, 41, 67].

Với yêu cầu đòi hỏi ngày càng cao của cuộc sống, các ứng dụng EEG\_ER chỉ thực sự phù hợp và có ý nghĩa thực tế khi được triển khai theo định hướng trí tuệ nhân tạo vạn vật (Artificial Intelligence of Things – AIoT) [103]. Hình 0.2 mô tả kiến trúc thực tế của một hệ thống EEG\_ER theo định hướng AIoT. Trong đó, các hệ thống EEG\_ER thường được triển khai trên thiết bị biên của mạng AIoT. Điều này yêu cầu các thiết bị thu thập tín hiệu EEG ngoài tính di động và kết nối không dây thì thường có số lượng điện cực nhỏ nhằm đảm bảo tính cơ động, nhỏ gọn và tiết kiệm năng lượng. Trên thiết bị biên, các thuật toán xử lý tín hiệu, trích chọn đặc trưng và phân lớp để nhận dạng cảm xúc cũng phải được cài đặt trên các vi điều khiển nhỏ gọn, giới hạn về bộ nhớ và khả năng tính

toán (như Raspberry PI, STM32Cube.AI...). Kết quả nhận dạng sau đó sẽ được gửi lên đám mây rồi thực thi các tác vụ thông qua giao diện người dùng.

Từ các phân tích ở trên cho thấy vấn đề cấp thiết đặt ra là cần *ngiên cứu và phát triển một hệ thống EEG\_ER hoạt động hiệu quả trong điều kiện tài nguyên hạn chế*. Hạn chế thứ nhất ở đây là các thiết bị thu thập tín hiệu EEG đầu vào có số lượng điện cực nhỏ, có khả năng di động, giao tiếp không dây. Hạn chế thứ hai là các thuật toán xử lý tín hiệu, trích chọn đặc trưng và phân lớp phải được triển khai hiệu quả trên các vi điều khiển thông dụng (nhỏ gọn, giới hạn về bộ nhớ và khả năng tính toán). Để làm được điều này, bên cạnh việc lựa chọn đầu vào đặc trưng phù hợp, cần *ngiên cứu và phát triển một mô hình học sâu có kích thước nhỏ gọn, có thể hoạt động theo thời gian thực trên các vi điều khiển thông dụng*. Đồng thời, mô hình học sâu này cũng phải cho hiệu quả nhận dạng cảm xúc cao ngay cả khi các thiết bị EEG đầu vào có số lượng điện cực thấp. Đây cũng là mục đích chính của luận án.

## **2. Mục tiêu của luận án**

Để đạt được mục đích nêu trên, luận án cần phải thực hiện *bốn mục tiêu chính theo thứ tự lần lượt như sau*:

1. Cập nhật các kiến thức tổng quan về bài toán EEG\_ER; Nghiên cứu các kỹ thuật quan trọng cần triển khai trong các khâu thu thập, trích chọn đặc trưng tín hiệu EEG và phân lớp cảm xúc; Cập nhật các công trình nghiên cứu gần nhất ứng dụng kỹ thuật học sâu để giải quyết bài toán EEG\_ER; Xác định rõ các thách thức cần phải vượt qua khi xây dựng một hệ thống EEG\_ER theo định hướng AIoT (cài đặt trên các thiết bị biên có tài nguyên hạn chế).
2. Nghiên cứu khả năng ứng dụng của các kỹ thuật trích chọn đặc trưng, các kỹ thuật học sâu khác nhau trong bài toán EEG\_ER; Tiến hành các thử nghiệm kiểm chứng để từ đó xác định được đồng thời kiến trúc học sâu nền tảng cùng với dạng đầu vào đặc trưng phù hợp sao cho hiệu quả hoạt động là tốt nhất trong việc giải quyết bài toán EEG\_ER.
3. Nghiên cứu áp dụng các kỹ thuật mới kết hợp với kiến trúc học sâu nền tảng đã xác định được ở mục tiêu số 2 nhằm xây dựng một kiến trúc học sâu nhẹ có thể triển khai trên các thiết bị xử lý có tài nguyên hạn chế trong khi vẫn đảm bảo được hiệu quả nhận dạng cao đối với bài toán EEG\_ER.
4. Nghiên cứu áp dụng kỹ thuật học chuyển giao phù hợp nhằm nâng cao hơn nữa hiệu suất nhận dạng, tăng tốc độ suy luận mà không cần tăng kích thước mô hình. Chứng minh hiệu quả hoạt động thông qua toán học cũng như mô phỏng.

### **3. Đối tượng, phạm vi và phương pháp nghiên cứu**

#### **3.1. Đối tượng nghiên cứu**

Bài toán EEG\_ER; Các phương pháp trích chọn đặc trưng tín hiệu EEG; Các phương pháp học sâu; Các kỹ thuật nâng cao hiệu quả hoạt động của mạng học sâu; Các kỹ thuật học chuyển giao.

#### **3.2. Phạm vi nghiên cứu**

Nghiên cứu tìm kiếm mô hình học sâu hiệu quả cho bài toán EEG\_ER thông qua việc mô phỏng đánh giá hiệu quả hoạt động của kiến trúc trên các bộ mẫu dữ liệu EEG có sẵn, được cộng đồng khoa học trong lĩnh vực nghiên cứu công nhận (DEAP, AMIGOS và DREAMER). Các kịch bản mô phỏng được xây dựng trên công cụ Google Colab Pro (GPU T4).

#### **3.3. Phương pháp nghiên cứu**

Phương pháp nghiên cứu lý thuyết: Nghiên cứu lý thuyết về bài toán ER (đặc biệt là bài toán EEG\_ER); nghiên cứu lý thuyết về các mạng học sâu, các kỹ thuật nâng cao hiệu quả hoạt động của mạng học sâu; nghiên cứu lý thuyết về các kỹ thuật học chuyển giao nhằm tìm kiếm mô hình học sâu hiệu quả với kích thước nhỏ.

Phương pháp nghiên cứu dựa trên mô phỏng: Sử dụng công cụ mô phỏng số trên Google Colab Pro để hiệu chỉnh và đánh giá các kết quả nghiên cứu dựa trên lý thuyết.

Phương pháp trao đổi khoa học: thảo luận, seminar, lấy ý kiến chuyên gia, công bố các kết quả nghiên cứu trên các hội thảo, tạp chí khoa học trong và ngoài nước.

### **4. Các đóng góp của luận án**

NCS đã nghiên cứu đề xuất một số phương pháp học sâu nhằm xây dựng một hệ thống EEG\_ER hoạt động hiệu quả. Luận án có hai đóng góp chính sau đây:

1. Phát triển mô hình học sâu nhẹ: Trên cơ sở phân tích các công trình công bố gần nhất cũng như tiến hành các mô phỏng đánh giá cần thiết, NCS đã xác định được kiến trúc mạng học sâu 1D-CNN với đầu vào là các đặc trưng FFT của tín hiệu EEG là phù hợp nhất để triển khai bài toán EEG\_ER. Kiến trúc này sau đó được cải tiến bằng cách tích hợp các kỹ thuật Inception và Squeeze and Excitation, từ đó xây dựng mô hình học sâu có kích thước nhỏ (chỉ khoảng 40.14MB) nhưng vẫn duy trì hiệu suất nhận dạng cao, phù hợp triển khai trên thiết bị có tài nguyên hạn chế. Các nội dung chi tiết của đóng góp này được công bố trong các công trình CB [2],CB [3],CB [4],CB [5] và CB [6].

2. Áp dụng kỹ thuật Tự chất lọc tri thức (Self-Knowledge Distillation - SKD): NCS đã đề xuất mô hình học sâu EEG-SKNet kết hợp với kỹ thuật tăng cường dữ liệu, trọng số (Scaled weight) trong quá trình tính hàm Loss của SKD. Điều này giúp cho mô hình hội tụ nhanh hơn, giảm thời gian suy luận, giảm số lượng các phép toán dấu phẩy động (FLOPs) khi triển khai thực nghiệm, rất thích hợp triển khai trên thiết bị có tài nguyên hạn chế. Các nội dung chi tiết của đóng góp này được công bố trong công trìnhCB [7].

## **5. Ý nghĩa khoa học và thực tiễn**

### **5.1. Ý nghĩa khoa học**

Luận án là một trong những số ít công trình khoa học ở Việt Nam nghiên cứu vấn đề tìm kiếm mô hình học sâu hiệu quả cho bài toán EEG\_ER, nhằm định hướng triển khai trên các ứng dụng AIoT trong thực tế. Luận án đóng góp các giải pháp khoa học để xây dựng mô hình học sâu nhẹ hoạt động hiệu quả đối với bài toán EEG\_ER. Mô hình này vừa đảm bảo kích thước nhỏ (có thể triển khai được theo thời gian thực trên những vi điều khiển thông dụng) vừa có hiệu quả hoạt động cao. Hiệu quả hoạt động đã được đánh giá bằng mô phỏng trên các bộ mẫu dữ liệu: DEAP, AMIGOS và DREAMER (được đông đảo cộng đồng quốc tế công nhận). Kiến trúc này đạt được bằng hai biện pháp cải tiến. Trước tiên là việc áp dụng các kỹ thuật Inception và Squeeze and Excitation kết hợp với mạng học sâu 1D-CNN truyền thống để tạo nên một mô hình học sâu có thể triển khai trên các thiết bị tính toán có tài nguyên hạn chế. Tiếp đó là việc áp dụng kỹ thuật học chuyển giao SKD giúp tăng hiệu suất và độ chính xác mà không làm tăng kích thước mô hình.

### **5.2. Ý nghĩa thực tiễn**

Các kết quả nghiên cứu của luận án là cơ sở cho việc nghiên cứu áp dụng thử nghiệm, nâng cao chất lượng nhận dạng đối với các bài toán EEG\_ER theo định hướng AIoT trong đời sống xã hội. Một số ví dụ tiêu biểu cho các ứng dụng này là đánh giá hiệu quả của các video quảng cáo dựa trên EEG\_ER, đánh giá năng lực hành vi của người lái xe sau khi sử dụng nồng độ cồn, đánh giá mức độ tự kỷ của trẻ em... Đặc biệt, luận án rất có ý nghĩa trong lớp bài toán ứng dụng hỗ trợ những người mất khả năng vận động toàn thân tái hòa nhập với cộng đồng như: phát hiện nhu cầu của bệnh nhân trong sinh hoạt hằng ngày, điều khiển xe lăn điện, điều khiển thiết bị cho nhà thông minh dựa trên tín hiệu EEG.

## 6. Bố cục của luận án

Luận án đề xuất và đánh giá một chuỗi giải pháp tối ưu hoá mô hình học sâu cho nhận diện cảm xúc từ tín hiệu EEG nhằm đạt hiệu suất cao nhưng vẫn duy trì độ phức tạp tính toán thấp. Trọng tâm của nghiên cứu là phát triển mô hình nhẹ EEG-SKNet dựa trên *Self-Knowledge Distillation* và cơ chế trọng số điều chỉnh, phù hợp để có thể triển khai thực tế trên thiết bị biên của hệ thống AIoT (Hạn chế về tài nguyên). Vì vậy, ngoài phần mở đầu và kết luận, luận án được bố cục trong 4 chương (bám sát theo 4 mục tiêu đề ra) với nội dung chính như sau:

### ***Chương 1: Tổng quan về bài toán nhận dạng cảm xúc dựa trên tín hiệu điện não***

Chương này mô tả quá trình thực hiện mục tiêu thứ nhất của luận án. Nội dung chương tổng hợp các kiến thức cơ sở về bài toán nhận dạng cảm xúc, đặc biệt tập trung vào bài toán EEG\_ER; Phân tích các giải pháp kỹ thuật có thể triển khai khi xây dựng các khâu quan trọng của một hệ thống thu thập EEG\_ER như trích chọn đặc trưng tín hiệu EEG và phân lớp cảm xúc; Cập nhật các công bố gần nhất để xác định rõ các thách thức cần phải vượt qua khi xây dựng một hệ thống EEG\_ER theo định hướng AIoT.

### ***Chương 2: Tìm kiếm kiến trúc học sâu nền tảng và đặc trưng đầu vào phù hợp cho bài toán EEG-ER***

Chương 2 thực hiện khảo sát, phân tích và tìm kiếm kiến trúc học sâu nền tảng cho bài toán EEG\_ER. Trên cơ sở tổng hợp các nghiên cứu gần đây cũng như tiến hành các mô phỏng cần thiết, Nội dung chương sẽ chỉ ra rằng kiến trúc học sâu 1D-CNN kết hợp với đặc trưng đầu vào FFT cho hiệu suất nhận dạng tốt nhất. Kiến trúc này sẽ là nền tảng để tiến hành các cải tiến ở chương tiếp theo, đáp ứng các yêu cầu đặc thù cần thiết khi xây dựng một hệ thống EEG\_ER theo định hướng AIoT. Nội dung chương này gắn với mục tiêu thứ 2 của luận án.

### ***Chương 3: Phát triển mô hình học sâu nhẹ cho hệ thống EEG\_ER***

Trong chương 3, NCS đề xuất giải pháp cải tiến kiến trúc 1D-CNN đã thu được ở chương 2 nhằm xây dựng một mô hình học sâu có thể triển khai trên các thiết bị xử lý có tài nguyên hạn chế trong khi vẫn đảm bảo được hiệu quả nhận dạng cao đối với bài toán EEG\_ER. Giải pháp được đề xuất đó là kết hợp các kỹ thuật “Inception” và “Squeeze and Excitation” thêm vào kiến trúc 1D-CNN nhằm thu được một mô hình học sâu có kích thước nhỏ, có hiệu suất nhận dạng đủ lớn. Nội dung chương 3 cũng gắn liền với mục tiêu thứ 3 mà luận án cần đạt được.

### ***Chương 4: Tối ưu hoá mô hình EEG\_ER sử dụng kỹ thuật “Tự chất lọc tri thức”***

Chương 4 mô tả chi tiết việc thực hiện mục tiêu thứ 4 của luận án. Nội dung chương trình bày giải pháp nâng cao hiệu suất và khả năng khái quát hóa của mô hình bằng kỹ

thuật học chuyển giao SKD. Đồng thời, đề xuất cơ chế trọng số có điều chỉnh (Scaled-Weights) trong hàm mất mát giúp mô hình tập trung hơn vào các đoạn tín hiệu EEG quan trọng, góp phần cải thiện độ chính xác mà không làm tăng độ phức tạp. Hiệu quả giải pháp được đánh giá trên nhiều bộ dữ liệu EEG chuẩn.



# CHƯƠNG 1

## TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG CẢM XÚC DỰA TRÊN TÍN HIỆU ĐIỆN NÃO

Chương 1 được bố cục để thể hiện việc hoàn thành mục tiêu số 1 đã đề ra của luận án. Do vậy, phần đầu chương 1 sẽ trình bày các kiến thức tổng quan về bài toán nhận dạng cảm xúc. Trong đó, luận án nhấn mạnh tầm quan trọng của việc nhận dạng cảm xúc khi xây dựng các hệ thống HCI, đồng thời nêu rõ cơ sở của việc xác định cảm xúc thực tế thông qua tổ hợp các thành phần trong mô hình cảm xúc. Tiếp theo, luận án tập trung vào phân tích thể mạnh của định hướng EEG\_ER và mô tả chi tiết các khâu cơ bản trong một hệ thống EEG\_ER, cập nhật các nghiên cứu mới nhất về các kỹ thuật quan trọng cần triển khai trong các khâu thu thập, trích chọn đặc trưng tín hiệu EEG và phân lớp cảm xúc. Phần cuối chương tổng hợp các công trình nghiên cứu nhằm xác định rõ các thách thức cần phải vượt qua khi xây dựng một hệ thống EEG\_ER theo định hướng AIoT để từ đó làm nền tảng cho việc tìm kiếm một mô hình học sâu hiệu quả.

### 1.1 Tổng quan về bài toán nhận dạng cảm xúc

#### 1.1.1 Cảm xúc là gì?

Cảm xúc là một trạng thái tâm lý và sinh lý phức tạp, được hình thành như một phản ứng của con người đối với các kích thích bên ngoài và đóng vai trò quan trọng trong quá trình nhận thức cũng như hành vi xã hội [92]. Theo Russell, cảm xúc không chỉ là những trải nghiệm chủ quan mà còn có cơ sở sinh học rõ ràng, chịu sự điều chỉnh của hệ thần kinh trung ương và liên quan mật thiết đến các quá trình nhận thức, quyết định và tương tác giữa con người [31].

Ngoài ra, cảm xúc còn là phản ứng đối với các ý tưởng, ký ức, hoặc sự kiện xảy ra trong môi trường của chúng ta [74]. Cảm xúc rất quan trọng trong việc đưa ra quyết định và giao tiếp giữa các con người, do con người có xu hướng đưa ra quyết định dựa vào trạng thái cảm xúc của họ. Vì vậy, những cảm xúc xấu có thể dẫn đến không chỉ chối nhận về mặt tâm lý mà còn về mặt thể chất. Cảm xúc không tốt có thể góp phần vào sức khỏe kém, trong khi cảm xúc tích cực có thể dẫn đến mức sống cao hơn [41].

#### 1.1.2 Tầm quan trọng của việc nhận dạng cảm xúc trong giao tiếp người - máy

Trong suốt lịch sử tồn tại và phát triển của mình, con người luôn có xu hướng sử dụng máy móc thay mình để đáp ứng nhu cầu và giải phóng sức lao động. Điều này tạo

ra tính cấp thiết phải xây dựng được một hệ thống tương tác giữa người và máy sao cho con người có thể truyền đạt được mong muốn, yêu cầu của mình để máy hiểu được, hình thành nên các hệ thống HCI [81, 72].

Hệ thống tương tác người – máy (HCI) đóng vai trò quan trọng trong việc hỗ trợ con người trong nhiều lĩnh vực như trí tuệ nhân tạo, tự động hóa, chăm sóc sức khỏe, giáo dục và giao thông [44]. Tuy nhiên, hầu hết các hệ thống HCI hiện đại chưa thể nhận diện và phản hồi dựa trên trạng thái cảm xúc, khiến chúng thiếu đi sự thông minh cảm xúc, điều này làm giảm hiệu quả trong tương tác giữa con người và máy móc.

Một hệ thống HCI bỏ qua trạng thái cảm xúc của con người sẽ không thể phản ứng phù hợp, đầy đủ và chính xác. Do vậy, cần thiết phải có một hệ thống nhận diện cảm xúc đáng tin cậy, chính xác, linh hoạt và mạnh mẽ góp phần tạo nên hệ thống HCI thông minh [107, 57]. Điều này góp phần tạo nên một lĩnh vực nghiên cứu mới được gọi là tính toán ảnh hưởng (Affective computing - AC). AC là một lĩnh vực trí tuệ nhân tạo tập trung vào HCI thông qua việc phát hiện ảnh hưởng của người dùng. Một trong những mục tiêu chính của lĩnh vực AC là tạo ra các cách thức cho máy tính để giải thích cảm xúc của con người, từ đó cải thiện khả năng giao tiếp của chúng.

Các tín hiệu phi sinh lý (hành động, lời nói, biểu cảm khuôn mặt) và tín hiệu sinh lý (nhịp tim, huyết áp, chuyển động cơ...) đều có thể được sử dụng để xác định cảm xúc của con người. Tuy nhiên, phương pháp dựa trên tín hiệu phi sinh lý có tính chủ quan. Ví dụ, các chủ thể tham gia nghiên cứu có thể cố tình che giấu cảm xúc thật của họ, điều này có thể ảnh hưởng đến kết quả của nghiên cứu.

Nhận dạng cảm xúc dựa trên tín hiệu sinh lý có độ tin cậy và khách quan hơn [41, 67, 96]. Hệ thống này sẽ sử dụng các công nghệ cảm biến không xâm lấn, có thể mang theo được. Các cảm biến này thu thập tín hiệu sinh lý và sử dụng chúng làm đầu vào, giúp các hệ thống hiểu được sự tương quan giữa cảm xúc và trạng thái tâm lý để làm cho HCI trở nên thân thiện với con người.

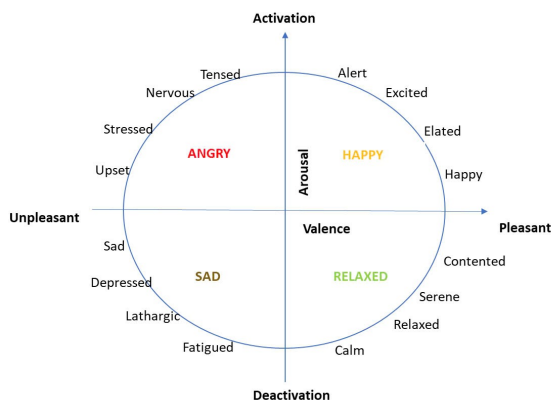
### **1.1.3 Các mô hình cảm xúc**

Cảm xúc là một trạng thái tâm lý phức tạp, có ảnh hưởng trực tiếp đến hành vi và nhận thức của con người. Trong các nghiên cứu khoa học, cảm xúc được mô hình hóa theo hai hướng chính: *mô hình cảm xúc rời rạc* và *mô hình cảm xúc theo chiều* [30, 93]. Mô hình cảm xúc rời rạc phân loại cảm xúc thành các nhóm riêng biệt, trong khi mô hình đa chiều biểu diễn cảm xúc theo không gian liên tục trên các chiều hay các thang đo, giúp mô tả chính xác hơn sự biến đổi trạng thái cảm xúc theo thời gian.

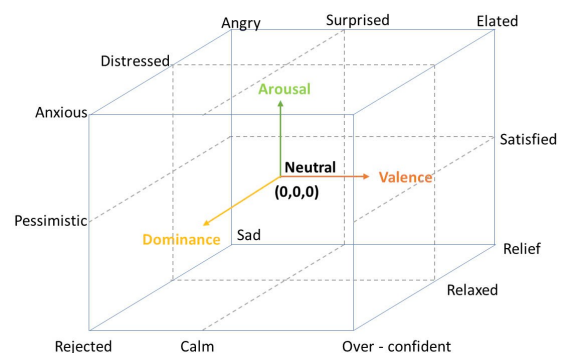
Đối với mô hình cảm xúc rời rạc, nhiều nhà lý thuyết đã thực hiện các thí nghiệm để xác định cảm xúc cơ bản và đã đề xuất một số mô hình được phân loại. Darwin đã đề

xuất một lý thuyết về cảm xúc bao gồm 9 cảm xúc cơ bản: *hứng thú - phấn khích, ngạc nhiên - giật mình, thích thú - vui mừng, đau khổ - thống khổ, mùi khó chịu, sợ hãi - kinh hoàng, tức giận - nổi điên, khinh miệt - ghê tởm, và xấu hổ - nhục nhã*. Bên cạnh đó, mô hình Ekman dựa trên một lý thuyết khác được chấp nhận rộng rãi [30]. Theo Ekman, các cảm xúc cơ bản phải bao gồm các đặc điểm sau: (1) cảm xúc là bẩm sinh; (2) nhiều người phát triển cùng một cảm xúc trong cùng một tình huống; (3) nhiều người biểu hiện cảm xúc cơ bản theo những cách tương tự; (4) các mô hình sinh lý của nhiều người là không thay đổi khi cảm xúc cơ bản được tạo ra. Ekman và các đồng nghiệp của ông xác định rằng có sáu cảm xúc chính được nhận biết rộng rãi qua biểu cảm khuôn mặt: *buồn, ngạc nhiên, hạnh phúc, ghê tởm, sợ hãi, và tức giận*. Các cảm xúc phụ (không cơ bản) khác, như *nhục nhã nhất, cảm giác tội lỗi, và khinh miệt*, có thể được tạo ra từ sáu cảm xúc cơ bản này. Nhiều nhà lý thuyết và tâm lý học đã bao gồm thêm các cảm xúc khác trong tập hợp cảm xúc cơ bản của họ mà không nằm trong sáu cảm xúc của Ekman. Một số người chia cảm xúc thành các nhóm nhỏ, tập trung vào cảm xúc chung, như sợ hãi hoặc tức giận (như các xúc tiêu cực) và hạnh phúc hoặc tình yêu (như cảm xúc tích cực). Người khác tập trung vào các sắc thái tinh tế hơn và chia cảm xúc thành các nhóm lớn hơn.

Tuy nhiên, một số nhà lý thuyết và nhà nghiên cứu tin rằng mô hình rời rạc có giới hạn về việc biểu diễn các cảm xúc cụ thể trên một phạm vi rộng lớn của trạng thái cảm xúc. Nói cách khác, trạng thái cảm xúc hàng ngày quá phức tạp để được biểu diễn tốt bằng một số lượng nhỏ các hạng mục rời rạc. Do đó, một phương pháp mới được biết đến như cảm xúc theo chiều đã được đề xuất. Trong mô hình này, cảm xúc được tổ chức theo một cách đa chiều, với mỗi chiều đại diện cho một đặc điểm cảm xúc. Mỗi cảm xúc có thể được biểu diễn như một điểm trong một không gian đa chiều. Thay vì chọn những nhãn rời rạc, người ta có thể biểu hiện cảm xúc của mình trên nhiều thang đo liên tục hoặc rời rạc, như *chú ý - từ chối hoặc dễ chịu - không dễ chịu*. Đến nay, các nhà nghiên cứu đã đề xuất nhiều phương pháp đa chiều để mô hình cảm xúc.



**Hình 1.1** Mô hình cảm xúc 2D Russell



**Hình 1.2** Mô hình cảm xúc Schloberg

Hình 1.1 minh họa mô hình 2D của Russell, có thể bao gồm đến 150 nhãn cảm xúc bằng cách sử dụng các chiều kích thích và giá trị. Trục dọc đại diện cho chiều kích thích (biểu thị độ mạnh của trải nghiệm cảm xúc, từ thấp đến phần khích), và trục ngang đại diện cho chiều giá trị (biểu thị mức độ vui vẻ hoặc niềm vui, từ tiêu cực đến tích cực). Có bốn loại cảm xúc trong hệ tọa độ kích thích - giá trị. Các cảm xúc tiêu cực được biểu diễn ở phía bên trái của hệ tọa độ và các cảm xúc tích cực được hiển thị ở phía bên phải. Trục giá trị biểu diễn cả cảm xúc tích cực và tiêu cực, và trục kích thích thay đổi từ cảm xúc không hoạt động đến cảm xúc hoạt động. Mô hình Russell thể hiện bốn góc phần tư tiêu biểu. Tương tự, Hình 1.2 thể hiện mô hình cảm xúc ba chiều của Schloberg, với đánh giá và kích hoạt là các chiều, thêm chiều chú ý - từ chối vào mô hình hai chiều [134].

## 1.2 Nhận dạng cảm xúc dựa trên tín hiệu điện não đồ EEG

### 1.2.1 Giới thiệu về EEG

EEG (Electroencephalography) là một kỹ thuật được sử dụng để ghi nhận hoạt động điện của não bộ. Đây là một phương pháp không xâm lấn, nghĩa là không gây tổn hại hay ảnh hưởng đến người tham gia đo. Trong kỹ thuật này, các điện cực sẽ được bố trí trên da đầu nhằm đo lường sự thay đổi điện áp ở các khu vực khác nhau (Hình 1.3). Phương pháp này lần đầu tiên được phát triển bởi nhà khoa học Berger vào năm 1924, khi ông sử dụng một điện cực đặt trên đầu con trai mình và ghi nhận được các dao động điện xuất hiện đều đặn. Ông phát hiện rằng trong bản ghi EEG, có thể phân biệt được một số loại sóng não dựa trên dải tần số – những loại sóng này về sau được xem là đặc trưng sinh lý quan trọng của hoạt động thần kinh.



**Hình 1.3** Thiết lập ghi tín hiệu EEG

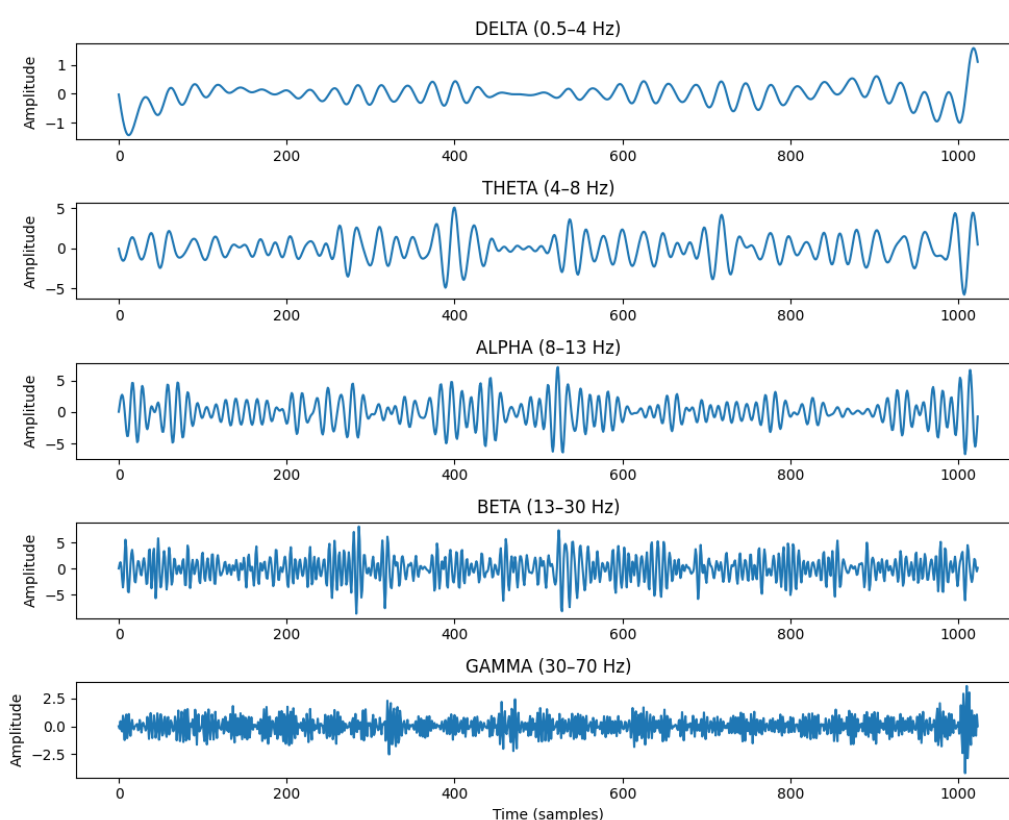
Như mô tả trong Hình 1.4, các tín hiệu EEG được phân loại thành năm loại dựa trên sự biến đổi trong các dải tần số: delta (0,5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), và gamma (>30 Hz).

Sóng Delta thường xuất hiện ở vùng vỏ não trước với biên độ từ 20–200  $\mu\text{V}$ . Chúng

thường được phát hiện trong trạng thái vô thức, giấc ngủ nông, không mộng mị hoặc bị gây mê. Sóng này sẽ biến mất ở người trưởng thành đang tỉnh táo và cảnh giác [106].

Sóng Theta thường xuất hiện ở các thùy đỉnh và thùy thái dương với biên độ 100–150  $\mu\text{V}$ . Chúng liên quan đến trạng thái thư giãn, trạng thái ngủ thiên, rất dễ xuất hiện khi cơ thể rơi vào trạng thái buồn ngủ hoặc bị thôi miên. Sóng Theta ở vùng trung tâm trước sẽ tăng khi cảm xúc tích cực được kích thích [78].

Sóng Alpha chủ yếu xuất hiện ở vùng thùy chẩm và thùy đỉnh với biên độ từ 20–100  $\mu\text{V}$ . Chúng có thể được phát hiện trong trạng thái nghỉ ngơi nhắm mắt. Các kích thích bên ngoài như kích thích thị giác hoặc thính giác, hoặc khi cá nhân tham gia vào hoạt động tư duy, có thể làm cho sóng Alpha biến mất. Chúng có năng lượng dao động nhiều hơn so với sóng beta và gamma trong cả cảm xúc tích cực và tiêu cực [78, 88].



**Hình 1.4** Nhận dạng các dạng sóng theo tần số

Sóng Beta thường chỉ được quan sát ở phần vỏ não ngay sau trán; tuy nhiên, khi một người đang suy nghĩ, sóng Beta xuất hiện ở nhiều vị trí khác nhau. Biên độ là 5–20  $\mu\text{V}$ . Chúng xuất hiện khi tâm trí của một người hoạt động cao và tập trung. Bộ não bị chi phối bởi sóng Alpha trong khi cơ thể con người đang thư giãn, và nhịp Beta dần mất đi khi hoạt động cảm xúc tăng lên. Khi hệ thần kinh trung ương chịu áp lực/stress/gắng sức, biên độ của sóng Alpha giảm trong khi tần số Beta tăng lên, và sóng Alpha dần dần chuyển thành sóng Beta. Khi vỏ não xuất hiện ở trạng thái Beta, điều này thường có nghĩa là não đang bị kích thích [11].

Sóng Gamma là loại sóng não có tần số cao nhất, nằm trong khoảng 30-100 Hz, và được liên kết với các hoạt động nhận thức phức tạp, tập trung cao độ, xử lý thông tin, và ý thức. Chúng liên quan đến các nhiệm vụ và chức năng nhận thức của não ở một cấp độ cao như việc nhận thông tin, xử lý, hòa nhập, truyền tải, và phản hồi trong hệ thống thần kinh cũng như các hoạt động đòi hỏi sự chú ý nhiều (tập trung). Chúng thường được quan sát trong quá trình xử lý giác quan đa chế độ [11, 54, 65].

Việc nhận dạng cảm xúc từ tín hiệu EEG mở ra tiềm năng phát triển to lớn khi ứng dụng vào nhiều lĩnh vực khác nhau, bao gồm y tế, giáo dục, giải trí, mua sắm, quân sự, xã hội, lái xe an toàn và nhiều lĩnh vực khác nữa [80]:

- Trong lĩnh vực y tế, thu nhận kịp thời các tín hiệu điện não đồ của bệnh nhân và phân tích nhanh trạng thái cảm xúc của họ có thể giúp các bác sĩ và y tá hiểu chính xác trạng thái tâm lý của bệnh nhân và sau đó đưa ra các quyết định y tế hợp lý, điều này có tác dụng quan trọng đối với việc phục hồi chức năng của một số người bị rối loạn tâm thần, chẳng hạn như bệnh tự kỷ, trầm cảm, bệnh Alzheimer và khuyết tật về thể chất.
- Về mặt giáo dục, công nghệ nhận dạng cảm xúc dựa trên tín hiệu điện não đồ có thể giúp đội ngũ giáo viên kịp thời điều chỉnh phương pháp giảng dạy và thái độ giảng dạy phù hợp với biểu hiện cảm xúc của các học viên khác nhau trong lớp, chẳng hạn như tăng hoặc giảm khối lượng công việc.
- Về mặt giải trí, chẳng hạn như trò chơi trên máy tính, các nhà nghiên cứu cố gắng phát hiện trạng thái cảm xúc của người chơi để thích nghi với độ khó, hình phạt và khuyến khích của trò chơi.
- Ở khía cạnh quân sự, có thể nắm bắt kịp thời, chính xác trạng thái cảm xúc của hạ sĩ quan, binh sĩ thông qua tín hiệu điện não đồ, từ đó kịp thời điều chỉnh bố trí chiến lược, nâng cao tỷ lệ thắng trận.
- Về mạng xã hội, chúng ta có thể tăng cường giao tiếp không rào cản trong hệ thống HCI, tăng cường sự hiểu biết và tương tác lẫn nhau trong kênh tương tác giữa người–máy–con người và tránh một số hiểu lầm và xích mích không cần thiết thông qua việc thu thập thông tin cảm xúc.
- Về mặt lái xe an toàn, việc phát hiện kịp thời các trạng thái cảm xúc bằng điện não đồ có thể giúp xe thực hiện khoá thông minh khi khởi động để chặn đánh lái hoặc chủ động mở chế độ lái tự động để can thiệp vào quỹ đạo chuyển động của xe cho đến khi đỗ xe ở vị trí an toàn, nhờ đó giảm thiểu đáng kể xảy ra tai nạn.

- Ứng dụng EEG trong truyền hình tương tác đang mở ra một hướng đi mới trong lĩnh vực giải trí và truyền thông. Hệ thống có thể phân tích phản ứng cảm xúc của người xem để cá nhân hoá nội dung, điều chỉnh độ sáng, âm thanh hoặc đề xuất chương trình phù hợp với tâm trạng. EEG cũng hỗ trợ “neuromarketing”, giúp tối ưu hoá quảng cáo dựa trên dữ liệu cảm xúc thực tế.
- Trong lĩnh vực điều khiển không tiếp xúc (hands-free control), EEG đóng vai trò quan trọng trong phát triển giao diện BCI, cho phép người dùng điều khiển thiết bị thông minh, máy tính hoặc robot chỉ bằng ý nghĩ. Công nghệ này đặc biệt hữu ích trong hỗ trợ người khuyết tật, giúp họ điều khiển xe lăn, cánh tay robot hoặc giao tiếp mà không cần thao tác vật lý. Ngoài ra, EEG còn được tích hợp vào các hệ thống công nghiệp và tự động hoá để điều khiển máy móc từ xa, tối ưu hoá hiệu suất làm việc.
- Trong môi trường thông minh, EEG được ứng dụng để tạo ra các không gian sống và làm việc có thể tự động thích nghi với cảm xúc của người dùng. Nhà thông minh có thể điều chỉnh ánh sáng, nhiệt độ, âm thanh hoặc hệ thống thông gió dựa trên trạng thái thần kinh, giúp nâng cao sự thoải mái và tiết kiệm năng lượng. Các hệ thống an ninh cũng có thể sử dụng EEG để phát hiện căng thẳng hoặc lo lắng, từ đó kích hoạt chế độ báo mật phù hợp.

Với sự kết hợp của EEG và các công nghệ tiên tiến như trí tuệ nhân tạo (AI), thực tế ảo (VR), thực tế tăng cường (AR) và Internet vạn vật (IoT), các ứng dụng này hứa hẹn sẽ tiếp tục phát triển mạnh mẽ, mang lại nhiều lợi ích vượt trội trong tương lai.

### ***1.2.2 Hệ thống nhận dạng cảm xúc dựa trên tín hiệu điện não***

Trong bối cảnh phát triển của trí tuệ nhân tạo và khoa học thần kinh, các hệ thống EEG\_ER đang trở thành một hướng nghiên cứu quan trọng. Không giống như các phương pháp nhận dạng cảm xúc dựa trên biểu cảm khuôn mặt, giọng nói hoặc hành vi, EEG có khả năng ghi nhận trực tiếp hoạt động điện sinh lý của não bộ, phản ánh trạng thái cảm xúc một cách khách quan và ít bị ảnh hưởng bởi các yếu tố môi trường. Điều này giúp EEG trở thành một công cụ hữu ích trong các hệ thống BCI, mang đến những ứng dụng quan trọng trong y tế, giáo dục, giải trí và công nghệ điều khiển [123, 20].

BCI là một hệ thống cho phép giao tiếp trực tiếp giữa não bộ và máy tính, giúp người dùng có thể điều khiển thiết bị hoặc tương tác với môi trường mà không cần sử dụng hệ thần kinh ngoại vi hay cơ bắp. Ban đầu, các hệ thống BCI chủ yếu được thiết kế để hỗ trợ người bị khuyết tật vận động hoặc các bệnh lý thần kinh nghiêm trọng như chấn thương tuỷ sống, đa xơ cứng. Tuy nhiên, với sự phát triển mạnh mẽ của khoa học

dữ liệu và trí tuệ nhân tạo, ứng dụng của BCI đã mở rộng sang nhiều lĩnh vực khác nhau, bao gồm chẩn đoán y tế, phục hồi chức năng, giám sát trạng thái tâm lý, nhà thông minh và điều khiển thiết bị điện tử.

Trong số các phương pháp thu nhận tín hiệu não phục vụ BCI, EEG được ưu tiên hơn so với các kỹ thuật khác như từ não đồ (MEG), điện cực nội sọ (ECoG) hay cận hồng ngoại chức năng (fNIRS) nhờ vào một số lợi thế đáng kể [20].

Thứ nhất, EEG là phương pháp không xâm lấn, chỉ sử dụng các điện cực gắn trên da đầu để ghi nhận tín hiệu não mà không cần can thiệp phẫu thuật như ECoG, giúp đảm bảo an toàn và dễ triển khai.

Thứ hai, EEG có chi phí thấp hơn và khả năng ứng dụng rộng rãi hơn so với MEG hoặc fMRI (chụp cộng hưởng từ chức năng), vốn đòi hỏi trang thiết bị đắt đỏ và chỉ có thể thực hiện trong môi trường phòng thí nghiệm. Từ đó dễ triển khai, giúp mở rộng khả năng ứng dụng trong thực tiễn, đặc biệt trên hệ thống có tài nguyên tính toán hạn chế.

Thứ ba, EEG có độ phân giải thời gian cao, có thể ghi nhận sự thay đổi của tín hiệu não theo từng mili-giây, trong khi các phương pháp như fNIRS hoặc fMRI thường có độ trễ lớn hơn, làm giảm tính khả dụng trong các hệ thống tương tác thời gian thực.

Cuối cùng, sự phát triển của thiết bị EEG di động, không dây, tiêu thụ ít năng lượng đã mở ra nhiều cơ hội để tích hợp EEG vào thiết bị đeo thông minh, điện thoại di động và các hệ thống IoT, tạo ra những nền tảng tương tác linh hoạt và dễ tiếp cận hơn.

Hoạt động của một hệ thống EEG\_ER điển hình được mô tả cụ thể trong Hình 0.1. Đầu tiên, hệ thống kích thích cảm xúc bằng cách trình bày các tác nhân kích thích (hình ảnh, âm thanh, video...) để tạo ra phản ứng cảm xúc tự nhiên. Sau đó, tín hiệu EEG được ghi nhận từ các điện cực, khuếch đại và xử lý để loại bỏ nhiễu, đảm bảo chất lượng dữ liệu. Tiếp theo, hệ thống trích xuất các đặc trưng điện sinh lý từ tín hiệu EEG, bao gồm năng lượng băng tần (Alpha, Beta, Theta, Gamma), độ đồng bộ hoá giữa các vùng não, hoặc các đặc trưng miền thời gian - tần số. Cuối cùng, hệ thống phân loại cảm xúc bằng các mô hình học máy (Machine Learning) hoặc học sâu (Deep Learning) để xác định trạng thái cảm xúc theo các mô hình như Arousal-Valence hoặc cảm xúc cơ bản (hạnh phúc, buồn bã, tức giận, sợ hãi, v.v.). Các phần tiếp theo của chương sẽ trình bày chi tiết các giải pháp đã được triển khai nhằm thực hiện tốt các khâu quan trọng trong hệ thống EEG\_ER.

## **1.3 Thu nhận và xử lý số tín hiệu EEG**

### **1.3.1 Kích thích cảm xúc**

Cảm xúc không dễ dàng bộc lộ, cần phải có những kích thích khiến cảm xúc được thể hiện. Vì vậy, khả năng kích thích/tạo ra trạng thái cảm xúc của đối tượng thí nghiệm



theo những cách phù hợp nhất định, tức là, kích thích cảm xúc, là một bước quan trọng trong việc phát hiện cảm xúc dựa trên tín hiệu sinh lý.

Hiện nay, có hai cách phổ biến để kích thích cảm xúc: gọi từ tưởng tượng và gọi từ sự kiện bên ngoài. Gọi từ tưởng tượng yêu cầu các đối tượng hồi tưởng lại các mảnh ký ức hoặc cảnh tượng cảm xúc cụ thể để kích thích họ tạo ra cảm xúc tương ứng. Một nghiên cứu đã đưa cho các đối tượng những mùi khác nhau trước khi thí nghiệm và sau đó yêu cầu họ nhớ lại mùi đó để gọi ra cảm xúc cụ thể [86]. Một nghiên cứu khác yêu cầu đối tượng tạo ra các cảnh tượng cảm xúc khác nhau thông qua các gợi ý để thu thập các tín hiệu EEG tương ứng [31]. Mặc dù phương pháp này có ưu điểm là cá nhân hóa trải nghiệm cảm xúc của từng đối tượng, nhưng nó cũng tồn tại một số hạn chế đáng kể. Trước hết, khả năng tưởng tượng và hồi tưởng cảm xúc của mỗi người là khác nhau, dẫn đến sự không đồng nhất trong dữ liệu EEG. Ngoài ra, trạng thái tinh thần của đối tượng tại thời điểm thí nghiệm cũng có thể ảnh hưởng đến mức độ phản ứng cảm xúc. Vì những lý do này, phương pháp gọi từ tưởng tượng ít được sử dụng trong các nghiên cứu EEG, đặc biệt là khi yêu cầu tính lặp lại cao và khả năng kiểm soát tốt giữa các lần thí nghiệm.

Khác với phương pháp gọi từ tưởng tượng, cách tiếp cận này sử dụng các kích thích bên ngoài như hình ảnh, âm thanh, video hoặc môi trường thực tế để khơi gợi cảm xúc một cách trực tiếp. Đây là phương pháp được sử dụng phổ biến nhất trong nghiên cứu EEG\_ER vì khả năng kiểm soát tốt hơn và tính lặp lại cao hơn giữa các lần thí nghiệm. Các phương pháp gợi cảm xúc từ sự kiện phổ biến bao gồm: hệ thống hình ảnh cảm xúc Quốc tế (International Emotional Picture System - IAPS) [15], thư viện âm thanh cảm xúc Quốc tế (International Emotional Audio Library - IADS) [14]. Ngoài ra, với sự phát triển của công nghệ thực tế ảo (VR) và thực tế tăng cường (AR), nhiều nghiên cứu đã tích hợp các công nghệ này vào quá trình kích thích cảm xúc. VR cho phép người tham gia trải nghiệm môi trường nhập vai, giúp tăng cường mức độ phản ứng cảm xúc so với các phương pháp truyền thống như xem video hoặc nghe nhạc. Điều này đặc biệt hữu ích trong các nghiên cứu cần tạo ra phản ứng cảm xúc mạnh mẽ và chân thực.

Trong bối cảnh hướng đến triển khai trên thiết bị di động và tài nguyên hạn chế, việc lựa chọn phương pháp kích thích cảm xúc cần xem xét đến hiệu quả tính toán và mục tiêu thụ tài nguyên. Các phương pháp sử dụng video hoặc VR đòi hỏi khả năng xử lý cao, có thể không phù hợp với các nền tảng di động hoặc hệ thống nhúng. Ngược lại, các phương pháp sử dụng nguồn kích thích bằng hình ảnh tĩnh (IAPS) hoặc âm thanh (IADS) tiêu tốn ít tài nguyên hơn, phù hợp hơn cho các ứng dụng trên thiết bị di động hoặc các hệ thống có khả năng tính toán hạn chế.

### 1.3.2 Thu nhận và xử lý tín hiệu

#### 1.3.2.1 Thiết bị thu thập

Trong quá trình thu thập tín hiệu EEG, việc lựa chọn thiết bị thu EEG và phân bố các điện cực đóng vai trò quan trọng trong việc trích xuất, phân tích và phân loại đặc trưng cảm xúc. Chất lượng của tín hiệu EEG thu được phụ thuộc nhiều vào thiết bị thu EEG, số lượng điện cực, tần số lấy mẫu, cũng như phương pháp thu thập. Việc lựa chọn thiết bị phù hợp không chỉ đảm bảo độ chính xác cao trong nghiên cứu mà còn quyết định khả năng triển khai thực tế, đặc biệt trên các hệ thống tài nguyên hạn chế và di động. Thống kê về thiết bị thu và các điện cực thường được sử dụng trong một số thí nghiệm EEG về cảm xúc được trình bày trong Bảng 1.1.

**Bảng 1.1** Một số thiết bị thu thập tín hiệu EEG thông dụng

| Thiết bị                    | Số điện cực      | Tần số lấy mẫu (Hz) |
|-----------------------------|------------------|---------------------|
| Emotiv EPOC FLEX (2018)     | 32 (+2)          | 128                 |
| Emotiv EPOC X (2020)        | 14 (+2)          | 128/256             |
| Emotiv EPOC + (2013)        | 14 (+2)          | 128/256             |
| Emotiv INSIGHT (2015)       | 5 (+2)           | 128                 |
| OpenBCI (2014)              | 4/8/16           | 250/125             |
| Biosemi Active Two (2011)   | 6/7/12/32/64/256 | 128/512/1000        |
| Neuroscan Quik-Cap          | 32               | 250                 |
| NeuroSky MindWave (2018)    | 1                | 512                 |
| ESI NeuroScan System (2018) | 62               | 1000                |

Các nghiên cứu thực nghiệm EEG về cảm xúc, thiết bị thu tín hiệu EEG khác nhau tùy thuộc vào các nhu cầu thực tế. Các thiết bị phổ biến bao gồm Biosemi Active Two, Emotiv EPOC và Neuroscan Quik-Cap.

Biosemi Active Two [80] là hệ thống phân tích EEG thế hệ thứ hai và có độ phân giải cao, được phát triển bởi Biosemi ở Hà Lan dựa trên Active One. Thiết bị này đứng đầu trong ngành về tần số lấy mẫu, băng thông và tỷ lệ khử chế độ chung. Độ trở kháng cao của điện cực dưới điện cực chủ động không ảnh hưởng đến chất lượng tín hiệu, vì vậy không cần chuẩn bị da, giúp thao tác thí nghiệm trở nên thuận tiện và nhanh chóng hơn. Tuy nhiên, Biosemi Active Two là hệ thống EEG cao cấp với giá thành cao, phù hợp cho các phòng thí nghiệm nghiên cứu chuyên sâu, việc thiết lập và vận hành đòi hỏi kiến thức chuyên môn và thời gian, do đó không phù hợp cho người mới bắt đầu.

Emotiv EPOC [123] là thiết bị điện cực không cấy ghép, được phát triển bởi Emotiv Systems để thu sóng não. Thiết kế không dây với 14 điện cực, dễ dàng thiết lập và sử dụng. Tuy nhiên, cần chú ý đến việc đặt vị trí cảm biến chính xác và hạn chế chuyển động đầu để đảm bảo chất lượng tín hiệu. Bên cạnh Emotiv EPOC, Emotiv còn có các phiên bản Emotiv X (2020), Emotiv plus (2022) phổ biến trong cộng đồng nghiên cứu và ứng dụng EEG di động, đặc biệt trong các nghiên cứu về giao diện não-máy tính và ứng dụng tiêu dùng. Emotiv là sản phẩm thương mại nên có giá thành phải chăng, phù hợp cho cá nhân và các nhóm nghiên cứu nhỏ với ngân sách hạn chế [122].

OpenBCI (Open Brain-Computer Interface) [16] là một nền tảng mã nguồn mở cho phép thu thập và phân tích các tín hiệu sinh học, bao gồm điện não đồ (EEG), điện cơ đồ (EMG), và điện tâm đồ (ECG). OpenBCI được thiết kế với mục tiêu mang lại một hệ thống giá thành thấp, dễ tiếp cận và linh hoạt cho các nhà nghiên cứu và người dùng cá nhân muốn khám phá các ứng dụng trong khoa học thần kinh, giao diện não-máy tính (BCI), và các lĩnh vực liên quan đến tín hiệu sinh lý. OpenBCI bao gồm các bộ điều khiển như Cyton Board (8 điện cực, có thể mở rộng lên 16 điện cực), Ganglion Board (4 điện cực) và mũ EEG Ultracortex, có thể tích hợp với các phần mềm phân tích như OpenBCI GUI, MATLAB, Python và các công cụ mã nguồn mở khác. Hệ thống sử dụng kết nối không dây Bluetooth và Wi-Fi, giúp thuận tiện trong quá trình ghi nhận và truyền tải dữ liệu.

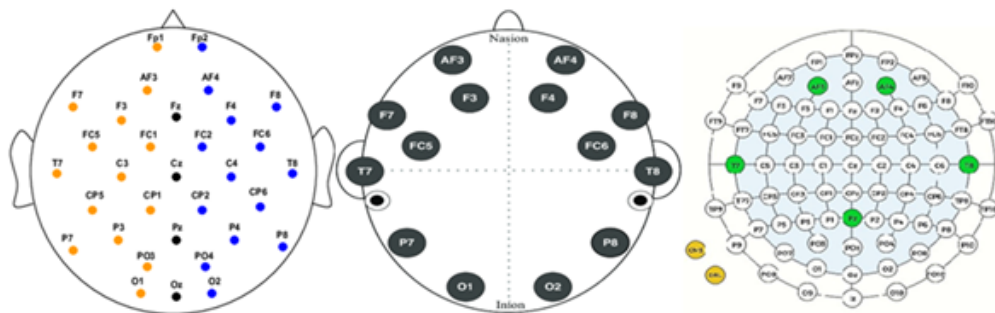
Quik-Cap của Neuroscan [61] được sử dụng rộng rãi trong nghiên cứu EEG lâm sàng và học thuật, đặc biệt trong các nghiên cứu về tiềm năng liên quan đến sự kiện (ERP). Thiết kế mũ chụp với các điện cực được đặt sẵn theo hệ thống 10–20 quốc tế, giúp giảm thời gian thiết lập. Tuy nhiên, việc chuẩn bị da đầu và áp dụng gel dẫn điện đòi hỏi có thể mất thời gian và yêu cầu kỹ năng, kinh nghiệm. Neuroscan Quik-Cap nằm ở mức giá trung bình đến cao, tùy thuộc vào số lượng điện cực và phụ kiện đi kèm, thích hợp cho các nghiên cứu yêu cầu độ chính xác cao và khả năng tái lập, nhưng việc thiết lập phức tạp hơn so với các hệ thống không dây, hạn chế trong việc di chuyển.

Đối với nghiên cứu hướng đến ứng dụng di động và thiết bị tài nguyên hạn chế, việc lựa chọn thiết bị cần đảm bảo cân bằng giữa hiệu suất và chi phí. Các hệ thống EEG như Emotiv EPOC và OpenBCI là những lựa chọn hợp lý do tính di động cao, tiêu thụ năng lượng thấp và dễ dàng triển khai. Ngược lại, các hệ thống như Biosemi Active Two tuy có độ chính xác cao nhưng đòi hỏi tài nguyên xử lý lớn và chi phí cao, không phù hợp với các thiết bị nhúng hoặc nền tảng di động.

#### *1.3.2.2 Thu nhận tín hiệu EEG*

Quy trình thu nhận và xử lý tín hiệu EEG đóng vai trò quan trọng trong hệ thống EEG\_ER vì nó ảnh hưởng trực tiếp đến chất lượng dữ liệu đầu vào. Tín hiệu EEG có thể được thu nhận thông qua hai phương pháp chính: xâm lấn và không xâm lấn. Phương

pháp xâm lấn sử dụng các điện cực cấy trực tiếp vào vỏ não để ghi nhận hoạt động điện sinh học của các tế bào thần kinh, mang lại tỷ lệ tín hiệu trên nhiễu (SNR) cao và độ phân giải tín hiệu tốt. Tuy nhiên, do yêu cầu phẫu thuật phức tạp và rủi ro y tế cao, phương pháp này chỉ được áp dụng trong các nghiên cứu lâm sàng và hỗ trợ bệnh nhân bị liệt. Ngược lại, phương pháp không xâm lấn sử dụng các điện cực gắn trên da đầu, có tính an toàn cao, dễ triển khai và là lựa chọn chính trong các nghiên cứu EEG\_ER hiện nay. Tuy nhiên, tín hiệu EEG thu được theo phương pháp này có thể bị ảnh hưởng bởi nhiễu từ chuyển động, tín hiệu điện lưới và các hoạt động sinh lý khác của cơ thể, do đó cần áp dụng các kỹ thuật tiền xử lý và lọc nhiễu phù hợp để đảm bảo chất lượng tín hiệu [85, 124].



**Hình 1.5** Bản đồ phân bố điện cực

Một yếu tố quan trọng trong thu nhận EEG là bố trí các điện cực trên da đầu. Trong đó, các vị trí phổ biến được đặt theo hệ thống quốc tế 10–20 [131]. Hình 1.5 minh họa bố trí các điện cực của các thiết bị EEG sử dụng số lượng điện cực khác nhau, bao gồm 32 điện cực (trái), 14 điện cực (giữa) và 5 điện cực (phải).

Các hệ thống EEG có thể có từ 1 đến 256 điện cực, nhưng trong nghiên cứu EEG\_ER, số lượng điện cực thường dao động trong khoảng 14 đến 64, tập trung vào các vùng não liên quan đến cảm xúc như thùy trán (FP, AF, F, FC), thùy đỉnh (CP, P), thùy chẩm (PO, O) và thùy thái dương (T) [126]. Việc sử dụng nhiều điện cực giúp cải thiện độ chính xác nhưng cũng làm tăng chi phí, độ phức tạp xử lý dữ liệu và yêu cầu tài nguyên tính toán cao hơn.

Một số nghiên cứu đã chỉ ra rằng việc sử dụng 14 điện cực EEG vẫn có thể đạt độ chính xác tương đương với mô hình 32 điện cực, trong khi giúp giảm chi phí và tối ưu hóa khả năng triển khai trên nền tảng tài nguyên hạn chế [108]. Điều này đặc biệt quan trọng đối với các ứng dụng EEG trên thiết bị di động, nơi mà tài nguyên tính toán bị giới hạn. Trong khi EEG 32 điện cực thường được sử dụng trong các nghiên cứu lâm sàng và khoa học thần kinh yêu cầu độ chính xác cao, EEG 14 điện cực là lựa chọn phù hợp cho các nghiên cứu ứng dụng trên các hệ thống di động. Đối với các ứng dụng tiêu dùng hoặc nghiên cứu trong môi trường thực tế, EEG 5 điện cực được xem là phương án khả thi, miễn là có các kỹ thuật tiền xử lý và trích xuất đặc trưng hiệu quả để bù đắp cho việc

giảm số lượng điện cực cần thiết khi thu thập.

### *1.3.2.3 Tiền xử lý tín hiệu EEG*

Sau khi tín hiệu EEG được thu thập, bước tiếp theo là xử lý tín hiệu nhằm cải thiện chất lượng dữ liệu. Quá trình này bao gồm tiền xử lý dữ liệu, lọc nhiễu.

Tiền xử lý tín hiệu EEG là giai đoạn quan trọng để làm sạch dữ liệu, loại bỏ các nhiễu không mong muốn phát sinh từ các chuyển động mắt, co bắp, nhiễu tim hoặc nhiễu từ môi trường xung quanh. Một số phương pháp phổ biến để tiền xử lý EEG bao gồm Phân tích thành phần độc lập (ICA), Phân tích thành phần chính (PCA), Tham chiếu trung bình chung (CAR) và Mẫu không gian chung (CSP). ICA và PCA thường được sử dụng để tách nhiễu do chuyển động mắt và co bắp, trong khi CAR giúp giảm nhiễu nền hiệu quả mà không làm mất thông tin quan trọng trong tín hiệu EEG. CSP chủ yếu được ứng dụng trong nhận diện vận động và giao diện não-máy, nhưng cũng có thể hỗ trợ tách biệt các tín hiệu liên quan đến cảm xúc [124].

Lọc nhiễu là một bước quan trọng trong xử lý tín hiệu EEG nhằm đảm bảo tín hiệu thu nhận có độ chính xác cao nhất. Một số phương pháp lọc nhiễu phổ biến bao gồm bộ lọc Butterworth (lọc thông cao, lọc thông thấp) và bộ lọc Notch (lọc giải hẹp). Bộ lọc Butterworth giúp giới hạn băng thông tín hiệu EEG trong khoảng 1–50 Hz, trong khi bộ lọc Notch được sử dụng để loại bỏ nhiễu điện lưới ở tần số 50–60 Hz tùy thuộc vào từng quốc gia. Việc sử dụng bộ lọc cần được thực hiện cẩn thận để tránh làm biến dạng tín hiệu gốc, ảnh hưởng đến quá trình phân tích cảm xúc [44].

Đối với các hệ thống EEG triển khai trên thiết bị tài nguyên hạn chế, cần cân nhắc lựa chọn số lượng điện cực EEG tối ưu để đảm bảo cân bằng giữa độ chính xác và khả năng xử lý dữ liệu. Việc sử dụng thuật toán lọc nhiễu nhẹ như Butterworth, Notch và ICA sẽ giúp tiết kiệm tài nguyên tính toán mà vẫn đảm bảo chất lượng tín hiệu. Ngoài ra, việc áp dụng mô hình học sâu nhẹ (lightweight deep learning) có thể giúp tối ưu hóa hiệu suất phân loại cảm xúc mà không yêu cầu tài nguyên xử lý quá lớn [17]. Trong bối cảnh nghiên cứu hướng tới ứng dụng di động và các nền tảng nhúng, việc tích hợp các phương pháp xử lý tín hiệu hiệu quả, kết hợp với các thiết bị EEG không dây như Emotiv EPOC hoặc OpenBCI, sẽ là hướng đi khả thi nhằm phát triển hệ thống EEG\_ER có tính ứng dụng cao trong thực tế.

### **1.3.3 Một số bộ dữ liệu cho bài toán EEG\_ER**

Việc phát triển các thuật toán EEG\_ER hiệu quả cao dựa trên các kỹ thuật học sâu đòi hỏi các tập dữ liệu lớn với số lượng đối tượng đáng kể. Do đó, các tập dữ liệu đóng vai trò quan trọng. Hiện tại, rất nhiều nhà nghiên cứu đã đưa ra một số tập dữ liệu có sẵn. Trong đó, DEAP là bộ dữ liệu phổ biến và có độ tin cậy cao thường được sử dụng nhất,

tiếp theo là AMIGOS, và DREAMER bộ dữ liệu được thu thập từ thiết bị có số lượng điện cực thấp. Các bộ dữ liệu này đều được thu thập trong môi trường thực nghiệm có kiểm soát, với thông tin EEG được đồng bộ hóa cùng các phản hồi cảm xúc tự đánh giá của người tham gia (Self-Assessment).

**DEAP:** DEAP là một bộ dữ liệu sinh lý nổi bật được thu thập từ 32 tình nguyện viên (gồm 16 nam và 16 nữ), trong đó mỗi người tham gia được ghi nhận hoạt động não thông qua 32 kênh EEG và 8 kênh tín hiệu sinh lý ngoại vi khác, tạo thành tổng cộng 40 kênh. Trong quá trình thu thập, các tình nguyện viên được yêu cầu xem 40 đoạn phim kích thích cảm xúc, mỗi đoạn kéo dài khoảng 60 giây. Sau mỗi đoạn phim, người tham gia đánh giá cảm xúc cá nhân theo bốn chiều: Valence, Arousal, Dominance và Liking, với thang điểm từ 1 đến 9. Nghiên cứu sinh sử dụng phiên bản dữ liệu đã được tiền xử lý bởi nhóm tác giả, trong đó đã bao gồm cả tín hiệu và nhãn. Về mặt tổ chức, dữ liệu được lưu dưới dạng ma trận có kích thước  $40 \times 40 \times 8064$  (với 40 video, 40 kênh, và 8064 điểm dữ liệu mỗi kênh), tương ứng với tín hiệu gốc. Riêng phần nhãn được biểu diễn bằng ma trận  $40 \times 4$ , thể hiện đánh giá cảm xúc ở bốn chiều chính. Các tín hiệu EEG trong DEAP được lấy mẫu ở tần số 128 Hz và đã qua bộ lọc bandpass từ 4 Hz đến 45 Hz để loại bỏ nhiễu không mong muốn. Ngoài ra, chỉ 32 kênh EEG trong số 40 kênh được sử dụng để phục vụ cho việc nhận dạng cảm xúc; 8 kênh còn lại (bao gồm EOG và các cảm biến đo huyết áp, nhiệt độ...) không được đưa vào phân tích do không liên quan trực tiếp đến mục tiêu nghiên cứu [60].

**DREAMER:** Tập dữ liệu DREAMER [56] là một ví dụ về tập dữ liệu đa phương thức, bao gồm các tín hiệu EEG và ECG được thu thập từ 23 đối tượng. Các tín hiệu này được ghi lại qua 14 điện cực EEG với tần số lấy mẫu 128 Hz. Dữ liệu EEG được thu thập trong khi các đối tượng xem 18 đoạn video, mỗi đoạn được thiết kế để gợi ra một trong chín cảm xúc cơ bản như bình tĩnh, vui vẻ, giận dữ, buồn bã, hạnh phúc, ghê tởm, phấn khích, ngạc nhiên và sợ hãi (Katsigiannis and Ramzan 2018). Các đoạn video được phát không gián đoạn trong khoảng thời gian từ 65–393 giây, được coi là đủ để gợi ra cảm xúc. Sau mỗi đoạn video, các đối tượng tự đánh giá mức độ kích thích, hứng thú, yêu thích và kiểm soát của họ trên thang điểm từ 1 đến 9. Trong quá trình tiền xử lý, tín hiệu EEG được giảm tần số xuống 128 Hz và áp dụng bộ lọc thông dải từ 4.0–45 Hz để loại bỏ nhiễu nền trước khi phân tích dữ liệu.

**AMIGOS:** AMIGOS [77] là bộ dữ liệu đa modal được thu thập từ 40 người tham gia trong hai bối cảnh: xem 16 video ngắn khi một mình và sau đó một số trong số đó xem thêm 4 video dài (cá nhân hoặc theo nhóm). Tín hiệu EEG gồm 14 kênh (128Hz) đã được lọc thông cao (high-pass filter 2 Hz) và căn giữa (average-referenced), ECG 2 kênh và GSR 1 kênh đều được lấy mẫu với tần số 60 Hz. Người tham gia tự đánh giá các cảm xúc như valence, arousal, dominance, liking và familiarity trên thang 1–9, cùng

đánh dấu cảm xúc cơ bản như vui, buồn với nhị phân 0/1. Thí nghiệm cũng bao gồm ghi hình HD phía trước và video RGB toàn thân để phân tích sâu hơn. Dữ liệu được tổ chức theo từng người (P01–P40), có sẵn định dạng .mat. Đây là một bộ dữ liệu đa mô thức phong phú, cho phép khai thác song song cả tín hiệu não và tín hiệu sinh lý ngoại vi để nâng cao độ chính xác của hệ thống nhận diện cảm xúc.

Bên cạnh đó, một số ít nghiên cứu sử dụng một số bộ dữ liệu khác như SEED, SEED – IV, GAMEEMO, MAHNOB–HCI. Thông tin tổng hợp về các bộ dữ liệu này được mô tả chi tiết trong Bảng 1.2

**Bảng 1.2** Một số bộ dữ liệu mẫu được sử dụng cho bài toán EEG\_ER

| Bộ dữ liệu                    | Nguồn kích thích   | Số người (M/F) | Thiết bị thu thập                      | Nhãn                                             | Tín hiệu thu  | Chú thích                                                                                                       |
|-------------------------------|--------------------|----------------|----------------------------------------|--------------------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------|
| <b>GAMEEMO (2020)</b> [4]     | Video qua trò chơi | 28 (20/8)      | EPOC+, 14 điện cực                     | Arousal, Valence (funny, boring, horror, calm)   | EEG           | 4 trạng thái cảm xúc; kích thích bằng trò chơi (game)                                                           |
| <b>SEED-IV (2018)</b> [135]   | Video              | 15             | ESI NeuroScan và kính theo dõi mắt SMI | Arousal, Valence                                 | EEG           | Phiên bản mở rộng của SEED, thêm cảm xúc “sợ hãi”.                                                              |
| <b>AMIGOS (2017)</b> [60]     | Video              | 40 (27/13)     | Emotiv EPOC neuro-headset              | Valence, arousal                                 | EEG, ECG, GSR | Dữ liệu đa phương thức cho hiệu suất, tính cách và tâm trạng.                                                   |
| <b>SEED (2017)</b> [71]       | Âm thanh, hình ảnh | 15 (7/8)       | N/A                                    | Arousal, Valence                                 | EEG           | Dữ liệu cho các mục đích khác nhau trong việc sử dụng tín hiệu EEG.                                             |
| <b>DREAMER (2015)</b> [56]    | Âm thanh, hình ảnh | 25 (14/11)     | EPOC X SHIMMER ECG sensor              | Valence, arousal, dominance                      | EEG, ECG      | Bộ dữ liệu không đắt; low-cost EEG và ECG.                                                                      |
| <b>DEAP (2014)</b> [60]       | Video              | 32 (16/16)     | Biosemi ActiveTwo                      | Valence, arousal, dominance, liking, familiarity | EEG           | Bộ dữ liệu đa phương thức bao gồm EEG và tín hiệu sinh lý ngoại vi.                                             |
| <b>MAHNOB–HCI (2012)</b> [99] | Video, hình ảnh    | 27 (14/11)     | Biosemi ActiveTwo                      | Valence, arousal                                 | EEG, ECG, GSR | Dữ liệu khuôn mặt, âm thanh và tín hiệu EEG, cùng với các tín hiệu sinh lý của hệ thần kinh ngoại vi/trung ương |

## 1.4 Trích xuất đặc trưng tín hiệu EEG

Trích xuất đặc trưng là một bước quan trọng trong EEG\_ER, đóng vai trò quyết định đối với hiệu suất của mô hình phân loại. Quá trình này nhằm chuyển đổi tín hiệu EEG thô thành các đặc trưng có ý nghĩa, phản ánh sự thay đổi trong hoạt động não bộ liên quan đến cảm xúc. Các phương pháp trích xuất đặc trưng EEG có thể được phân loại thành ba nhóm chính: miền thời gian, miền tần số, miền thời gian–tần số. Việc lựa chọn phương pháp phù hợp không chỉ dựa trên độ chính xác mà còn phải xét đến hiệu suất tính toán, đặc biệt là khi triển khai trên các thiết bị tài nguyên hạn chế.

### 1.4.1 Trích xuất miền thời gian

Đây là những đặc trưng tính toán trực tiếp trên tín hiệu EEG theo trục thời gian. Phổ biến nhất là các đặc trưng thống kê (mean, variance, skewness, kurtosis, v.v.), đo lường các thuộc tính cơ bản của biên độ sóng não. Ngoài ra, tham số Hjorth gồm activity (độ biến thiên năng lượng), mobility (độ phức tạp tần số) và complexity (độ phức tạp đường cong) cũng được dùng để đặc trưng hóa hình dạng tín hiệu theo thời gian [44].

Event-Related Potential (ERP) là một đặc trưng miền thời gian quan trọng, tập trung vào biên độ và độ trễ của các sóng EEG đặc thù xuất hiện đáp ứng với kích thích (ví dụ P300, N400). ERP phản ánh sự thay đổi điện thế theo thời gian liên quan đến sự kiện và thường được dùng trong nhận dạng trạng thái cảm xúc hoặc nhận thức. Ngoài ra, một số kỹ thuật khác như Higher Order Crossing (HOC) (đếm số lần tín hiệu cắt trục không), hoặc mô hình tự hồi quy (AR/AAR) cũng thuộc miền thời gian.

Các đặc trưng thời gian có thể dễ dàng kết hợp với mô hình học máy nhẹ (như SVM, LDA) nhờ vector đặc trưng nhỏ gọn. Tuy nhiên, nếu dùng mô hình học sâu, yêu cầu đặc trưng thời gian đơn giản có thể không đủ phong phú cho mô hình phức tạp, nên tín hiệu thời gian thô thường được sử dụng hơn để mạng tự học đặc trưng. Vậy có thể nhận thấy, đặc trưng miền thời gian dễ tính toán, chi phí thấp, nhưng có thể bỏ sót thông tin tần số quan trọng của tín hiệu EEG [79].

### 1.4.2 Trích xuất miền tần số

Các đặc trưng miền tần số được trích xuất sau khi chuyển đổi tín hiệu EEG sang miền tần số. Phổ biến nhất là Fast Fourier Transform (FFT) để lấy phổ tín hiệu, từ phổ này có thể tính mật độ phổ công suất (Power Spectral Density - PSD) theo các băng tần [2]. Các đặc trưng thường dùng gồm công suất hoặc năng lượng ở từng băng tần, tần số có công suất cực đại, hoặc tổng năng lượng phổ. Đặc biệt trong nhận dạng cảm xúc EEG, đặc trưng năng lượng/entropy vi phân (Differential Entropy - DE) được sử dụng nhiều — DE thực chất là entropy của tín hiệu giả định phân bố Gaussian, tỉ lệ với log của phương sai, tương đương với thông tin phổ năng lượng [44].



Ưu điểm của miền tần số là giúp trực quan hóa và phân tách được mức đóng góp của các dải sóng não khác nhau trong trạng thái cảm xúc. Phân tích miền tần số được đánh giá cao nhờ tính dễ tính toán, linh hoạt và có cơ sở sinh lý rõ ràng. Chẳng hạn, nghiên cứu của Zouridakis [41] đã chia EEG thành 5 băng tần sinh lý rồi tính PSD mỗi băng để nhận diện cảm xúc. Một nghiên cứu khác sử dụng FFT kết hợp lọc băng để trích xuất công suất theta, alpha, beta, gamma, sau đó chọn lọc đặc trưng bằng Relief-F, đã đạt độ chính xác trung bình 85–92% cho phân loại cảm xúc theo các thang Valence, Arousal, Dominance, Liking [28].

Bên cạnh đó, khả năng kết hợp với mô hình học sâu/nhẹ cũng rất linh hoạt: có thể đưa trực tiếp vector đặc trưng tần số vào các bộ phân lớp nhẹ (SVM, Random Forest) hoặc làm đầu vào cho mạng neural. Thực tế, việc sử dụng đặc trưng tần số giúp giảm kích thước đầu vào, cho phép mô hình học sâu nhẹ hơn mà vẫn đạt kết quả tốt [41].

Nhược điểm của phương pháp này là giả định tín hiệu tĩnh tại trong khoảng thời gian phân tích; nếu cảm xúc thay đổi đột ngột, phổ tần số có định có thể không phản ánh kịp thời sự biến đổi đó [119]. Tuy nhiên, các đặc trưng miền tần số (như công suất phổ theo băng tần) được sử dụng nhờ cân bằng tốt giữa thông tin hữu ích và chi phí tính toán.

#### ***1.4.3 Trích xuất miền thời gian–tần số***

Đây là nhóm đặc trưng kết hợp cả thông tin thời gian lẫn tần số nhằm mô tả tín hiệu EEG một cách toàn diện. Do EEG là tín hiệu không dừng, phân tích thời gian–tần số cho phép theo dõi sự biến đổi phổ theo thời gian, phù hợp để nhận dạng những thay đổi cảm xúc ngắn hạn [79]. Phương pháp phổ biến nhất là biến đổi wavelet, cho phép phân giải thời gian tốt ở tần số cao và phân giải tần số tốt ở tần số thấp [44]. Ngoài ra còn có STFT (Fourier ngắn hạn) – chia tín hiệu thành các cửa sổ nhỏ rồi áp dụng FFT, hay Hilbert-Huang (HHT với EMD) để trích xuất đặc trưng thời gian–tần số. Độ chính xác phân lớp cảm xúc với đặc trưng thời gian–tần số đạt cao nhất trong các miền, nhờ mô tả được đầy đủ động thái tín hiệu. Thống kê cho thấy phần lớn các nghiên cứu gần đây ưu tiên đặc trưng miền thời gian–tần số (chiếm ~35% công trình) so với miền tần số (27%) hay thời gian (20%) [127].

Khả năng tính toán là thách thức chính với phương pháp này, các phép biến đổi wavelet rời rạc (DWT) nhiều mức, hay EMD phân tích các mode nội tại, đều đòi hỏi tính toán nhiều hơn so với FFT thông thường. Điều này dẫn đến tốc độ chậm hơn – nhược điểm chung được ghi nhận khi dùng wavelet trên EEG. Do đó trên thiết bị di động yếu, việc tính toán liên tục các đặc trưng thời gian–tần số có thể quá sức nếu không tối ưu. Tuy nhiên, có thể giảm độ phức tạp bằng cách giảm số thang wavelet hoặc độ dài cửa sổ, hoặc tiền xử lý chọn đoạn tín hiệu chứa phản ứng cảm xúc mạnh. Một cách khác là áp dụng mô hình học sâu nhẹ để tự động trích xuất đặc trưng thời gian–tần số.

Đặc trưng thời gian–tần số kết hợp ưu điểm của hai miền, cho phép phản ánh biến đổi phổ theo thời gian thực. Đổi lại, chi phí tính toán cũng cao hơn, việc triển khai đòi hỏi tối ưu thuật toán hoặc phần cứng.

## 1.5 Phân lớp tín hiệu EEG trong bài toán nhận dạng cảm xúc

Trong những năm gần đây, học máy (machine learning) và đặc biệt là học sâu (deep learning) phát triển mạnh đã giúp máy tính tự động hiểu, nhận dạng và phân tích cảm xúc từ tín hiệu EEG hiệu quả hơn. Các phương pháp EEG\_ER có thể chia thành hai nhóm chính: (1) phương pháp học máy truyền thống (shallow learning) – thường yêu cầu trích xuất đặc trưng thủ công từ tín hiệu rồi dùng bộ phân loại truyền thống, và (2) phương pháp học sâu – sử dụng các mạng neural nhiều tầng để tự động học đặc trưng và phân loại trực tiếp từ dữ liệu thô hoặc đã tiền xử lý tối thiểu. Trong mục này, sẽ lần lượt điểm qua các kỹ thuật tiêu biểu của hai nhóm này trong bài toán phân loại EEG (đặc biệt cho nhận dạng cảm xúc), và đánh giá ưu nhược điểm của chúng.

### 1.5.1 Các kỹ thuật học máy truyền thống

Các phương pháp truyền thống thường tuân theo quy trình: tiền xử lý tín hiệu EEG (lọc nhiễu, artifact, v.v.), trích xuất đặc trưng thủ công (ví dụ: đặc trưng thời gian, tần số, các thống kê, entropy, v.v.), sau đó phân loại bằng mô hình học máy. Những kỹ thuật này dựa trên đặc trưng được chọn lọc và không mô hình hóa trực tiếp tính động thời gian của tín hiệu [44]. Một số thuật toán phổ biến gồm:

**Máy vector hỗ trợ (Support Vector Machine - SVM):** SVM là bộ phân loại mạnh trong các bài toán hai lớp và đã được dùng rộng rãi để phân loại cảm xúc từ EEG. Với lựa chọn hàm kernel phù hợp, SVM có thể đạt hiệu quả phân loại tốt trên dữ liệu EEG đã trích đặc trưng. Ưu điểm của SVM là ý tưởng phân loại đơn giản và tính toán khá hiệu quả. Tuy nhiên, SVM gặp khó khăn khi dữ liệu lớn hoặc bài toán nhiều lớp, đòi hỏi mở rộng one-vs-one hoặc one-vs-all và tốn chi phí tính toán cao hơn [127].

**k-láng giềng gần nhất (k-Nearest Neighbor – k-NN):** Đây là phương pháp phân loại dựa trên khoảng cách, đơn giản nhưng thường được dùng làm đường cơ sở. k-NN không cần huấn luyện mô hình phức tạp, chỉ yêu cầu lưu dữ liệu huấn luyện và tính khoảng cách tới k điểm lân cận để dự đoán nhãn. Trong EEG\_ER, k-NN có thể đạt độ chính xác khá nếu không gian đặc trưng được chọn tốt. Dù vậy, k-NN thường kém hơn các phương pháp khác về độ chính xác và độ ổn định, đặc biệt trên dữ liệu nhiễu [72].

**Rừng ngẫu nhiên (Random Forest - RF):** RF gồm tập hợp nhiều cây quyết định học ngẫu nhiên trên các phần dữ liệu, sau đó biểu quyết kết quả. Nhờ cơ chế ensemble, RF chống overfit tốt hơn cây quyết định đơn lẻ và xử lý tương tác phi tuyến giữa các đặc trưng. Trong EEG\_ER, RF thường cho kết quả ổn định, thích hợp khi số lượng đặc trưng

lớn và đa dạng, tuy nhiên số cây nhiều có thể làm tăng thời gian suy luận [44].

**Naive Bayes (NB):** Là bộ phân loại xác suất dựa trên định lý Bayes, với giả định các đặc trưng độc lập với nhau. NB có ưu điểm đơn giản, rất nhanh và yêu cầu ít mẫu để huấn luyện. Trong bối cảnh EEG\_ER, NB đôi khi được dùng làm mô hình nền tảng vì tốc độ và tính dễ triển khai. Tuy nhiên, do giả định độc lập không thật sự đúng với các đặc trưng EEG (vốn thường rất phức tạp), độ chính xác của NB thường thấp hơn các phương pháp khác [44].

Ngoài ra, một số thuật toán truyền thống khác cũng được áp dụng gồm Linear Discriminant Analysis (LDA), cây quyết định, Mạng neural truyền thống (MLP) hay Extreme Learning Machine (ELM). Nhìn chung, các mô hình học máy truyền thống phụ thuộc mạnh vào bước chọn đặc trưng thủ công. Việc trích xuất đặc trưng đòi hỏi kiến thức chuyên gia và đôi khi tạo ra tập đặc trưng dư thừa, không phổ quát, dẫn đến hiệu quả không như mong muốn trên bài toán EEG\_ER [72]. Mặc dù vậy, với đặc trưng phù hợp, các phương pháp này cũng đạt kết quả tương đối tốt trong các nghiên cứu trước đây.

### **1.5.2 Các kỹ thuật học sâu**

Không giống học máy truyền thống, các phương pháp học sâu tự học đặc trưng trực tiếp từ dữ liệu qua nhiều lớp mạng nơ-ron, giảm thiểu sự cần thiết của trích chọn đặc trưng thủ công. Với khả năng học phi tuyến mạnh, học sâu đang trở thành xu hướng chính trong EEG\_ER [127, 8]. Dưới đây giới thiệu một số mô hình sâu đã được nghiên cứu và áp dụng với EEG:

#### **1.5.2.1 Mạng nơ-ron tích chập**

Trong những năm gần đây, việc sử dụng các kiến trúc DL đã mang lại kết quả đáng kể trong các lĩnh vực khác nhau, chẳng hạn như nhận dạng cảm xúc. Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) là mô hình DL thành công nhất. Trong bối cảnh đó, các nhà nghiên cứu đặc biệt cố gắng giới thiệu các mô hình mới dựa trên các kiến trúc này. AlexNet là sự phát triển đầu tiên của kiến trúc CNN đã giành chiến thắng trong cuộc thi ImageNet. Có ba loại lớp chính trong một CNN: các lớp tích chập, các lớp pooling và các lớp fully connected. Lớp tích chập thực hiện việc trích xuất đặc trưng, và lớp pooling giảm độ chiều. Cuối cùng, lớp FC phân loại dữ liệu đầu vào hoặc các đặc trưng được phát hiện. Các mô hình CNN sử dụng các mẫu không gian để tạo ra một biểu diễn mạnh mẽ của dữ liệu hiện có. Do đó, các kiến trúc CNN có thể triển khai trên dữ liệu 1D, 2D và 3D [24].

CNN là mô hình sâu hiệu quả trong việc trích xuất đặc trưng không gian (spatial) và cấu trúc cục bộ từ dữ liệu. Đối với EEG, CNN thường được dùng để học đặc trưng từ nhiều điện cực đồng thời hoặc từ biểu diễn đặc trưng tần số, thời gian – tần số của tín

hiệu. Cấu trúc tích chập và pooling của CNN giúp tự động lọc nhiễu và chọn lọc thông tin quan trọng trong dữ liệu EEG. Nhờ đó, CNN xử lý tốt dữ liệu đầu vào có kích thước lớn, nhiều điện cực mà không gặp “áp lực” về chiều dữ liệu. Tuy nhiên, CNN cũng có nhược điểm: tầng pooling có thể làm mất một số thông tin cảm xúc hữu ích, và việc huấn luyện nhiều lớp tích chập phức tạp có nguy cơ rơi vào cực trị địa phương (local optimum) nếu không tối ưu tốt [20, 37].

CNN có thể được chia thành các mô hình 1D, 2D, 3D, và các mô hình huấn luyện sẵn. Dưới đây là giải thích về các kiến trúc CNN cho EEG\_ER.

**1D-CNN:** Các mô hình CNN có khả năng học các mẫu phức tạp từ dữ liệu y tế, làm cho chúng phù hợp với các ứng dụng đa dạng, với điều kiện chúng được huấn luyện trên các tập dữ liệu với số lượng đối tượng lớn. Tính năng độc đáo này đã làm cho CNNs trở thành lựa chọn phổ biến cho việc xử lý dữ liệu y tế 1D và 2D. Trong nghiên cứu EEG\_ER, tín hiệu EEG thường là tín hiệu 1D và có thể được xử lý bằng các lớp tích chập của kiến trúc 1D-CNN. Các lớp tích chập đóng vai trò quan trọng trong việc nhận dạng các mẫu trong tín hiệu EEG để nhận dạng cảm xúc. Ngoài ra, các lớp pooling và fully connected (FC) được sử dụng để giảm độ chiều và phân loại tín hiệu. So với các kiến trúc 2D-CNN, các kiến trúc 1D-CNN có ít siêu tham số hơn, làm cho chúng dễ dàng triển khai trên phần cứng có tài nguyên hạn chế. 1D-CNN có thể triển khai trên phần cứng có tài nguyên hạn chế. Tuy nhiên, một trong những nhược điểm chính của chúng là có thể không hoạt động tốt như các mô hình khác khi đối mặt với dữ liệu phức tạp và đa dạng cao [62].

**2D-CNN:** Các kiến trúc 2D-CNN rất phổ biến trong các ứng dụng y tế khác nhau [50, 36]. Chúng được thiết kế dựa trên các lớp tích chập, pooling, và FC, tương tự như kiến trúc 1D-CNN. Vì các kiến trúc 2D-CNN là kiến trúc mặc định cho CNNs, chúng được sử dụng rộng rãi để phân đoạn và phân loại dữ liệu y tế, cho phép trích xuất các đặc trưng không gian và hỗ trợ trong các quy trình chẩn đoán [117, 9]. Các kiến trúc 2D-CNN cũng đã được triển khai thành công trong một số nghiên cứu nhận dạng cảm xúc, nơi các tín hiệu 1D-CNN trước tiên được chuyển đổi thành hình ảnh 2D bằng các kỹ thuật tiền xử lý cấp cao. Trong bước tiếp theo, kiến trúc 2D-CNN trích xuất các đặc trưng và phân loại các hình ảnh 2D dựa trên tín hiệu EEG. 2D-CNN có nhiều ưu điểm trong việc nhận dạng cảm xúc từ tín hiệu EEG. Chúng xuất sắc trong việc nhận dạng các mẫu không gian và có thể hiệu quả trong việc nắm bắt các mối quan hệ phức tạp trong dữ liệu y tế, làm cho chúng rất phù hợp cho nhận dạng cảm xúc. Tuy nhiên, 2D-CNNs tốn kém về mặt tính toán và yêu cầu nhiều thời gian huấn luyện và sức mạnh xử lý hơn các phương pháp ML.

**3D-CNN:** Đối với các tín hiệu EEG có đặc điểm không gian và thời gian, 3D-CNN cung cấp một lựa chọn hiệu quả để xử lý dữ liệu này. Các mạng nơ-ron 3D có khả năng

xử lý dữ liệu ba chiều, cho phép chúng phân tích cả thông tin không gian và thời gian một cách đồng thời. Đối với tín hiệu EEG, điều này có thể có nghĩa là sử dụng một khối ba chiều của dữ liệu EEG, nơi mỗi chiều đại diện cho không gian, thời gian, và điện cực [6]. Các lớp tích chập 3D có thể trích xuất các mẫu thời gian-không gian từ tín hiệu EEG một cách hiệu quả, cung cấp một cách tiếp cận toàn diện để phân tích cảm xúc [120, 95]. Việc này đặc biệt hữu ích trong các tình huống mà các tín hiệu có đặc điểm phức tạp mà các mô hình 1D hoặc 2D có thể không hiệu quả trong việc xử lý, mặc dù 3D-CNN phức tạp, yêu cầu tài nguyên tính toán nhiều hơn so với 1D, 2D.

**CNN Pre-trained models:** Việc truy cập vào các tập dữ liệu y tế quy mô lớn là một thách thức đáng kể trong việc sử dụng các mô hình học sâu (Deep Learning - DL) [58]. Các nhà nghiên cứu thường làm việc với dữ liệu y tế hạn chế chứa số lượng lớn đối tượng, điều này làm cho việc sử dụng các kiến trúc DL cho các ứng dụng y tế trở nên thách thức [76]. Để vượt qua thách thức này, các nhà nghiên cứu đã phát triển nhiều mô hình huấn luyện sẵn. Những mạng này ban đầu được lấy cảm hứng từ các kiến trúc CNN và đầu tiên được huấn luyện trên tập dữ liệu ImageNet trước khi lưu trữ các trọng số mới của chúng [25].

Để thích ứng với ứng dụng mong muốn, các mô hình huấn luyện sẵn được tinh chỉnh bằng cách sử dụng các trọng số đã lưu trữ trên tập dữ liệu cụ thể. Trong số các kiến trúc huấn luyện sẵn được sử dụng rộng rãi nhất là GoogleNet, VGG, và AlexNet. Một số ít mô hình huấn luyện sẵn đã cho thấy kết quả hứa hẹn trong nhận dạng cảm xúc [24]. Các mô hình huấn luyện sẵn giúp vượt qua các thách thức của dữ liệu y tế hạn chế, cho phép các nhà nghiên cứu tinh chỉnh các mô hình trên các tập dữ liệu nhỏ hơn. Tuy nhiên, các mô hình huấn luyện sẵn có thể không luôn tổng quát hóa được cho các lĩnh vực y tế mới, và chúng có thể không hiệu quả khi nhiệm vụ nguồn và nhiệm vụ đích khác biệt đáng kể.

#### *1.5.2.2 Mạng nơ-ron hồi quy*

Mạng nơ-ron hồi quy (Recurrent Neural Network - RNN) được thiết kế để xử lý dữ liệu tuần tự theo thời gian, rất phù hợp với tín hiệu EEG (chuỗi thời gian nhiều điện cực). RNN có cơ chế nhớ trạng thái trước đó, cho phép mô hình hóa sự phụ thuộc thời gian trong tín hiệu cảm xúc. Nhờ tính chất này, nhiều nghiên cứu đã áp dụng RNN để phân loại cảm xúc dựa trên chuỗi EEG và thu được kết quả khả quan. Do đó, RNN thường được kết hợp với các cơ chế cải tiến (như LSTM, GRU) để cải thiện hiệu quả [127].

Đặc biệt, mạng bộ nhớ dài-ngắn hạn (Long Short-Term Memory Networks – LSTM) là một biến thể đặc biệt của RNN với các cổng nhớ có chủ đích (input, forget, output gates), giúp lưu giữ hoặc quên thông tin một cách chọn lọc. Nhờ kiến trúc này, LSTM học được các phụ thuộc dài hạn trong chuỗi EEG tốt hơn RNN thường, rất hữu ích để nhận biết xu hướng cảm xúc theo thời gian. LSTM phát huy hiệu quả đặc biệt

trên dữ liệu EEG tuân tự phức tạp. Nhược điểm chính của LSTM là mô hình phức tạp khó song song hóa, thời gian huấn luyện và suy luận lớn hơn so với các mô hình thông dụng khác [97].

#### 1.5.2.3 Mạng nơ ron đồ thị

Một kỹ thuật mới đang được quan tâm là mạng nơ-ron đồ thị (Graph Neural Networks -GNN), cho phép biểu diễn tín hiệu EEG dưới dạng đồ thị với các nút là điện cực và cạnh thể hiện mối liên kết giữa các vùng não. GNN khai thác hiệu quả sự phụ thuộc giữa các vùng não để trích xuất đặc trưng, giúp phân biệt rõ hơn các trạng thái cảm xúc và cải thiện độ chính xác phân loại so với các phương pháp truyền thống. Nhờ mô hình hóa trực tiếp cấu trúc kết nối não bộ, GNN không chỉ nâng cao hiệu quả nhận dạng cảm xúc mà còn mở ra tiềm năng khám phá hoạt động não trong các trạng thái cảm xúc. [59, 89].

Tuy nhiên, các phương pháp GNN hiện tại thường xây dựng đồ thị trên từng đoạn thời gian một cách độc lập, do đó thiếu sự tương quan giữa các điện cực khác nhau qua các lát thời gian. Nhiều mô hình GNN cho EEG hiện nay sử dụng đồ thị đầy đủ (fully-connected graph) giữa tất cả các điện cực, dẫn đến mỗi nút (kênh) có quá nhiều kết nối lân cận. GNN là mô hình phức tạp, đòi hỏi tính toán trên cấu trúc đồ thị. Số lượng tham số có thể lớn, đặc biệt khi học ma trận trọng số cạnh, cùng với trọng số mạng. Do đó, nguy cơ *overfitting* tăng nếu dữ liệu không đủ lớn hoặc không đa dạng. Hơn nữa, xây dựng đồ thị tối ưu cho EEG không đơn giản – nếu dùng kiến thức chuyên gia để xác định cạnh (dựa trên vị trí điện cực hoặc kết nối não bộ đã biết) thì có thể bỏ sót quan hệ ẩn. Ngược lại, nếu để mô hình học hoàn toàn cấu trúc đồ thị thì cần nhiều dữ liệu huấn luyện và chi phí tính toán cao. Việc cân bằng giữa đồ thị cố định dựa trên tri thức và đồ thị học được một cách thích nghi là một bài toán khó đang được nghiên cứu [59].

#### 1.5.2.4 Cập nhật một số kết quả nghiên cứu về ứng dụng các kỹ thuật học sâu trong bài toán EEG\_ER

Trong những năm gần đây, các kỹ thuật học sâu đã đạt được bước tiến lớn, ngày càng nhiều nhà nghiên cứu coi học sâu là một phương pháp nghiên cứu quan trọng cho việc EEG\_ER. Bảng 1.3 tóm tắt các nghiên cứu trong lĩnh vực này từ năm 2020 đến 2024. Kết quả cho thấy, độ chính xác phân lớp của các thuật toán học sâu dao động từ 61.25% đến 97.56%, độ chính xác này phụ thuộc lớn vào việc số lớp được nhận dạng. Có thể thấy rằng nhận dạng nhị phân (2 lớp) có độ chính xác rất cao – tiệm cận 100%, trong khi việc nhận dạng nhiều lớp hơn (ví dụ như 3, 4 hay 9 lớp) có độ chính xác thấp hơn hẳn, dao động quanh ngưỡng 90%. Ngoài ra, bộ dữ liệu cũng quyết định phần lớn độ chính xác của nghiên cứu, có thể thấy rằng nếu sử dụng 2 bộ dữ liệu cho 1 tác vụ thì độ chính xác sẽ giảm so với chỉ sử dụng 1 bộ dữ liệu.

**Bảng 1.3** Độ chính xác phân lớp của một số thuật toán học sâu

| Nghiên cứu | Bộ dữ liệu                   | Số điện cực | Phương pháp học                   | Cảm xúc                              | Acc (%)                                                                                        |
|------------|------------------------------|-------------|-----------------------------------|--------------------------------------|------------------------------------------------------------------------------------------------|
| [96] 2020  | SEED, DREAMER, Dữ liệu riêng | 32, 14      | DGCNN                             | 2 class/ Arousal; Valence; Dominance | SEED: 90.4; DREAMER: Arousal: 84.5, Valence: 86.23, Dominance: 85.02                           |
| [134] 2021 | Dữ liệu riêng                | 32          | BiLSTM                            | Positive, neutral, negative          | 72.83                                                                                          |
| [108] 2022 | DEAP, DREAMER, SEED, AMIGOS  | 32, 14      | CNN + SVM                         | Valence, Arousal, Positive, Negative | Arousal: 77.1, Valence: 76.6; Arousal: 90.4, Valence: 88.2; Arousal: 90.5, Valence: 78.4; 88.5 |
| [3] 2022   | DEAP                         | 32          | M1M2 (CNN)                        | Valence, Arousal                     | 96.8                                                                                           |
| [104] 2022 | DEAP                         | 32          | DECNN, BiLSTM                     | Valence, Arousal                     | 94.0                                                                                           |
| [25] 2023  | DEAP                         | 32          | RA2-3DCNN                         | Valence, Arousal                     | 97.19, 97.58                                                                                   |
| [46] 2023  | DEAP                         | 32          | CNN_1D                            | Valence, Arousal                     | 98.4, 97.7                                                                                     |
| [69] 2023  | DEAP                         | 32          | Pre-train Convolution Capsule     | Valence, Arousal                     | 93.89                                                                                          |
| [49] 2023  | SEED                         | 32          | CNN and LSTM Based Ensemble       | Valence, Arousal                     | 97.16                                                                                          |
| [40] 2024  | DEAP, SEED                   | 32          | 3D-CNN + Vision Transformer (ViT) | Valence, Arousal                     | DEAP: 92.44, SEED: 92.85, 98.69                                                                |
| [34] 2024  | DREAMER                      | 14          | ICaps-ResLSTM                     | 2 class (Arousal Valence Dominance)  | 94-95                                                                                          |
| [10] 2024  | AMIGOS                       | 14          | CNN-Vision Transformer (ViT) 31   | Valence, Arousal                     | 98.2                                                                                           |

Từ Bảng 1.3 cho thấy, các mô hình đạt độ chính xác rất cao ở thiết lập nhị phân nhưng suy giảm đáng kể khi tăng số lớp cảm xúc, phản ánh ranh giới quyết định tinh vi hơn đòi hỏi đặc trưng giàu ngữ cảnh. Về kiến trúc, mạng đồ thị/DGCNN khai thác tương quan không gian giữa các điện cực nên ổn định ở cấu hình kênh trung bình–cao [134]; BiLSTM và CNN+BiLSTM giúp nắm phụ thuộc dài hạn nhưng chi phí suy luận/độ trễ khiến tính phù hợp triển khai biên hạn chế [134, 104]; CNN-1D trên đặc trưng FFT/PSD cân bằng tốt giữa hiệu quả và chi phí, phù hợp vai trò baseline nhẹ [69]. Các biến thể đa thang như Inception và chú ý theo kênh như SE (khi tích hợp vào CNN) giúp bù đắp thiếu hụt đa thang và trọng số kênh, trong khi 3D-CNN hoặc hybrid CNN-ViT có thể nâng kết quả khi đầu vào là bản đồ thời–tần/không–thời nhưng tham số/MFLOPs thường cao hơn [25, 40, 10]. Một số hướng khác (Capsule/ICaps-ReLSTM, ensemble) ghi nhận kết quả tốt ở 2 lớp song đánh đổi bằng độ phức tạp mô hình và latency [28, 49, 104]. Từ các quan sát này, cho thấy việc lựa chọn kiến trúc cho EEG\_ER cần cân bằng giữa năng lực biểu đạt (đa thang, quan hệ không gian/thời gian) và chi phí suy luận, được quyết định bởi điều kiện dữ liệu (số điện cực, dạng đặc trưng, số lớp) cùng ngân sách, tài nguyên tính toán.

## 1.6 Phát triển các hệ thống nhận dạng tín hiệu EEG theo định hướng AIoT

AIoT là sự kết hợp giữa AI và IoT, tạo ra các hệ thống thông minh có khả năng thu thập, phân tích và đưa ra quyết định dựa trên dữ liệu từ các thiết bị kết nối. Sự hội tụ này cho phép các thiết bị IoT không chỉ thu thập và truyền tải dữ liệu từ xa mà còn sử dụng AI để xử lý và phản hồi một cách tự động và hiệu quả hơn [103].

Việc phát triển các hệ thống nhận dạng tín hiệu EEG theo định hướng AIoT mang lại nhiều ưu điểm nổi bật. Trước hết, AIoT cho phép phân tích dữ liệu EEG theo thời gian thực, giúp nhận diện và phản hồi nhanh chóng đối với trạng thái cảm xúc của người dùng. Điều này đặc biệt quan trọng trong các ứng dụng như chăm sóc sức khỏe tâm lý và giao diện não–máy tính, nơi mà phản hồi kịp thời có thể cải thiện chất lượng cuộc sống và hiệu suất tương tác. Hơn nữa, với khả năng xử lý và lưu trữ dữ liệu phân tán thông qua điện toán biên và điện toán đám mây, AIoT giúp giảm tải cho hệ thống trung tâm và tăng cường tính linh hoạt trong triển khai các ứng dụng nhận dạng cảm xúc [70].

Xuất phát từ các ưu điểm đó, gần đây, nhiều nghiên cứu đã tập trung vào việc ứng dụng AIoT trong nhận dạng tín hiệu EEG. Chẳng hạn, như mô tả trong Hình 0.2, việc tích hợp AI vào các thiết bị biên AIoT giúp theo dõi và phân tích tín hiệu EEG của người dùng trong môi trường thực tế, cung cấp dữ liệu liên tục cho các ứng dụng chăm sóc sức khỏe và cải thiện tương tác người–máy. Ngoài ra, sự phát triển của các thuật toán học máy và học sâu đã nâng cao độ chính xác trong việc giải mã tín hiệu EEG, mở ra cơ hội cho các ứng dụng như phát hiện sớm các rối loạn tâm lý và tối ưu hóa trải nghiệm người



dùng trong các hệ thống thông minh [100, 63].

Tuy nhiên, như thể hiện trong Hình 0.2, việc triển khai các mô hình học sâu trên các thiết bị biên của mạng AIoT đặt ra nhiều thách thức.

Thách thức thứ nhất là để tiện lợi khi sử dụng, thiết bị thu thập tín hiệu EEG phải đảm bảo tính di động được. Do đó, các thiết bị này thường có kích thước nhỏ, số lượng điện cực thấp. Điều này gây ra các hạn chế về lượng thông tin thông tin đầu vào đưa đến các mô hình học sâu. Kéo theo đó là ảnh hưởng đến độ chính xác nhận dạng.

Thách thức thứ hai nằm ở sự giới hạn về tài nguyên tính toán trên phần cứng (vi điều khiển) thực hiện tác vụ nhận dạng. Do là thiết bị biên nên phần cứng này bị giới hạn về CPU, bộ nhớ và tốc độ suy luận. Vì vậy, khi cài đặt các mô hình học sâu trên phần cứng này, kích thước mô hình phải được tối ưu hóa để phù hợp với khả năng lưu trữ và xử lý hạn chế của thiết bị. Việc sử dụng các mô hình quá phức tạp có thể dẫn đến tiêu tốn tài nguyên và năng lượng, ảnh hưởng đến hiệu suất và tuổi thọ của thiết bị. Do đó, cần áp dụng các kỹ thuật tối ưu hóa mô hình đặc thù cho thiết bị biên như: lượng tử hóa mô hình, cắt tỉa tham số và nén mô hình để giảm kích thước và yêu cầu tính toán, đồng thời duy trì độ chính xác chấp nhận được. Ngoài ra, việc huấn luyện mô hình trên nền tảng đám mây và sau đó triển khai các phiên bản nhẹ hơn trên thiết bị biên cũng là một giải pháp khả thi, giúp cân bằng giữa hiệu suất và tài nguyên [103, 63]. Bảng 1.4 thống kê một số vi điều khiển tiêu biểu thường được sử dụng làm các thiết bị biên trong hệ thống AIoT với các giới hạn về CPU, bộ nhớ và tốc độ suy luận. Đây chính là các điều kiện quan trọng cần phải được thỏa mãn.

**Bảng 1.4** Một số vi điều khiển được sử dụng làm thiết bị biên trong mạng AIoT

| Thiết bị biên                  | CPU / SoC AI                                               | Thông số phần cứng                                       | Hỗ trợ AI                            | Công suất<br>Thời gian suy luận | Ứng dụng phù hợp                      |
|--------------------------------|------------------------------------------------------------|----------------------------------------------------------|--------------------------------------|---------------------------------|---------------------------------------|
| <b>NVIDIA Jetson Nano</b>      | ARM Cortex-A57 (4 lõi) + GPU Maxwell 128 lõi               | 4GB LPDDR4<br>microSD 16GB (mở rộng)<br>500 GFLOPs       | TensorFlow, PyTorch, ONNX            | 5–10W<br>15–30ms                | Thị giác máy tính, robotics           |
| <b>NVIDIA Jetson Xavier NX</b> | 6 lõi Carmel ARMv8.2 + GPU Volta 384 lõi + 48 Tensor Cores | 8GB LPDDR4x<br>16/32GB eMMC<br>21,000 GFLOPs             | TensorRT, PyTorch, ONNX              | 10–15W<br>5–15ms                | AI hiệu năng cao tại biên             |
| <b>Google Coral Dev Board</b>  | Cortex-A53 (4 lõi) + Edge TPU                              | 1GB LPDDR4<br>8GB eMMC<br>4,000 GFLOPs                   | TensorFlow Lite                      | 2–4W<br>5–10ms                  | Phân loại ảnh, AI nhúng               |
| <b>Raspberry Pi 4 Model B</b>  | Cortex-A72 (4 lõi)                                         | 2–8GB LPDDR4<br>microSD<br>13 GFLOPs (tốc tính)          | TensorFlow Lite, ONNX                | 5–7W<br>50–100ms                | AI mức thấp, điện toán biên tổng quát |
| <b>Khadas VIM3</b>             | Amlogic A311D (4xA73 + 2xA53) + NPU                        | 4GB LPDDR4X<br>16GB eMMC<br>5,000 GFLOPs                 | TensorFlow Lite, Android NNAPI       | 5W<br>5–20ms                    | Camera thông minh, AI nhẹ             |
| <b>Arduino Portenta H7</b>     | STM32H747 lõi kép (Cortex-M7 + M4)                         | 1MB + SRAM mở rộng<br>Flash 16MB + microSD<br>1.5 GFLOPs | TensorFlow Lite for Microcontrollers | <1W<br>100–200ms                | TinyML, cảm biến IoT                  |
| <b>ESP32-S3</b>                | Xtensa LX7 lõi kép + Vector Extensions                     | 512KB + 8MB PSRAM<br>Flash (4–16MB)<br>0.2 GFLOPs        | TFLM, ESP-DL                         | <1W<br>200–500ms                | TinyML, nhận diện giọng nói/cử chỉ    |

Như vậy, việc phát triển các hệ thống EEG\_ER theo định hướng AIoT mang lại nhiều lợi ích tiềm năng, nhưng cũng đặt ra những yêu cầu khắt khe về độ chính xác và kích thước mô hình khi triển khai trên các thiết bị có tài nguyên hạn chế. Trong khuôn khổ luận án, “tài nguyên hạn chế” được hiểu theo hai pha khác nhau của vòng đời mô hình: (i) ở pha huấn luyện, các ràng buộc chủ yếu liên quan đến GPU/CPU, bộ nhớ và thời gian đào tạo; (ii) ở pha suy luận tại thiết bị biên (AIoT), các ràng buộc tập trung vào bộ nhớ khả dụng, công suất tiêu thụ, MFLOPs và độ trễ thời gian thực. Luận án tập trung tối ưu cho pha (ii), nơi mô hình phải hoạt động ổn định với độ trễ thấp trên phần cứng giới hạn, trong khi huấn luyện được thực hiện ngoại tuyến ở phía máy chủ. Cách tiếp cận này nhất quán với kiến trúc AIoT/biên, nơi xử lý cục bộ và phản hồi thời gian

thực là yêu cầu trọng yếu. Giải quyết những thách thức này đòi hỏi sự kết hợp giữa thiết kế mô hình hiệu quả và ứng dụng các kỹ thuật tối ưu hóa phù hợp. Từ đó NCS xác định các bước thực hiện như sau:

Bước 1. Tìm kiếm mô hình học sâu và nền tảng cùng với dạng đầu vào đặc trưng phù hợp để đạt hiệu quả nhận dạng cho bài toán EEG\_ER trong bối cảnh AIoT. Trọng tâm là lựa chọn kiến trúc gọn nhẹ cùng đặc trưng trích xuất nhanh nhằm bảo đảm cân bằng giữa độ chính xác và chi phí tính toán (số tham số, FLOPs, độ trễ suy luận), hướng tới triển khai trên thiết bị tài nguyên hạn chế. Kết quả thăm dò ở bước này sẽ xác lập baseline nhẹ cho các bước tối ưu tiếp theo.

Bước 2. Nghiên cứu áp dụng các kỹ thuật mới để cải tiến mô hình học sâu nền tảng này thành một mô hình nhẹ để có thể triển khai trên các thiết bị xử lý có tài nguyên hạn chế trong khi vẫn đảm bảo được hiệu quả nhận dạng cao đối với bài toán EEG\_ER

Bước 3. Nghiên cứu áp dụng kỹ thuật học chuyển giao phù hợp nhằm nâng cao hơn nữa hiệu suất nhận dạng, tăng tốc độ suy luận mà không cần tăng kích thước mô hình.

Nội dung các chương tiếp theo của luận án sẽ tập trung mô tả chi tiết quá trình thực hiện ba bước trên nhằm nghiên cứu và phát triển một mô hình học sâu có kích thước nhỏ gọn, hoạt động theo thời gian thực trên các vi điều khiển thông dụng, thường được sử dụng tại biên của mạng AIoT. Đồng thời, mô hình học sâu này cũng phải cho hiệu quả nhận dạng cảm xúc cao ngay cả khi các thiết bị EEG đầu vào có số lượng điện cực thấp.

## **1.7 Kết luận chương 1**

Qua nội dung chương 1, có thể thấy rằng việc giải quyết tốt các bài toán nhận dạng cảm xúc sẽ đem lại một giải pháp hiệu quả trong việc xây dựng các hệ thống HCI thông minh tiên tiến, ứng dụng trong nhiều lĩnh vực của cuộc sống như y tế, giáo dục, giải trí, quân sự, xã hội... Để nhận dạng đúng được cảm xúc của con người, chúng ta cần phải dựa trên các mô hình kích thích cảm xúc (dựa trên âm thanh, hình ảnh tác động) và đặc biệt là các mô hình cảm xúc phù hợp. Trong đó, được sử dụng nhiều nhất chính là mô hình cảm xúc 2D, 3D của Russell.

Bên cạnh đó, việc thu thập thông tin của não bộ có ý nghĩa rất to lớn. Hiện tại có rất nhiều cách thức như chụp cộng hưởng từ, chụp quang phổ cận hồng ngoại, chụp cắt lớp chức năng... Trong đó, thu thập thông tin thông qua tín hiệu EEG là một phương pháp có nhiều ưu điểm (không xâm lấn, giá thành rẻ). Do vậy, so với một số phương pháp như sử dụng tiếng nói, tư thế cơ thể hay nét mặt (các tín hiệu phi sinh lý), phương pháp nhận dạng cảm xúc từ tín hiệu EEG là một hướng tiếp cận có nhiều ưu điểm (nhỏ gọn, có thể di chuyển linh hoạt, giá thành thấp, không xâm lấn).

Tuy nhiên, để thực hiện tốt việc nhận dạng tín hiệu EEG cần áp dụng rất nhiều kỹ

thuật tiên tiến trong hai khâu quan trọng là trích chọn đặc trưng tín hiệu EEG và phân lớp cảm xúc.

Về trích chọn đặc trưng, có thể thấy ba phương pháp thông thường được sử dụng. Phương pháp ứng dụng biến đổi Fourier nhanh (FFT) cung cấp cách tiếp cận nhanh và hiệu quả để phân tích phổ tín hiệu. Phương pháp đánh giá mật độ phổ công suất (PSD) giúp giảm nhiễu và tăng độ ổn định của đặc trưng tần số. Phương pháp sử dụng biến đổi Wavelet (WT) giữ lại cả thông tin thời gian và tần số, giúp mô hình có khả năng học sâu tốt hơn, nhưng có độ phức tạp tính toán cao hơn.

Về phân lớp cảm xúc, các kỹ thuật học sâu đã thể hiện sự vượt trội so với các phương pháp học máy truyền thống. Tuy nhiên, hiện tại có rất nhiều kỹ thuật học sâu khác nhau, mỗi kỹ thuật lại phù hợp với một kỹ thuật trích chọn đặc trưng nhất định. Do đó, cần thiết phải tìm ra sự kết hợp hài hòa giữa các kỹ thuật này nhằm đem lại hiệu quả nhận dạng cao nhất. Vấn đề này đã được NCS tổng kết và đánh giá thông qua công trình công bố CB [1].

Bên cạnh đó, xu thế tất yếu của các hệ thống *EEG\_ER* sẽ là được phát triển tại các thiết bị biên của mạng AIoT. Các tác vụ nhận dạng phải được thực hiện trực tiếp ở thiết bị biên và gửi kết quả nhận dạng về trung tâm qua kết nối IoT. Điều này có nghĩa là các kỹ thuật học sâu phải được tích hợp trên các thiết bị nhỏ gọn, dễ dàng di chuyển (điều kiện tài nguyên hạn chế). Từ đó đặt ra thách thức phải tìm kiếm một mô hình học sâu với số lượng tham số nhỏ để có thể triển khai được theo thời gian thực trong khi vẫn phải duy trì hiệu quả nhận dạng ở mức chấp nhận được. Việc giải quyết vấn đề này cũng sẽ được trình bày trong các chương tiếp theo của luận án.

## **CHƯƠNG 2**

# **TÌM KIẾM KIẾN TRÚC HỌC SÂU NỀN TẢNG VÀ ĐẶC TRƯNG ĐẦU VÀO PHÙ HỢP CHO BÀI TOÁN EEG-ER**

Tín hiệu EEG là một trong những phương thức phổ biến để thu thập dữ liệu não bộ, đóng vai trò quan trọng trong nghiên cứu về nhận diện cảm xúc. Tuy nhiên, do bản chất phi tuyến, không tĩnh và nhạy cảm với nhiễu, việc xử lý và trích xuất đặc trưng từ tín hiệu EEG là một thách thức lớn. Để có thể áp dụng thành công các mô hình học sâu vào bài toán EEG-ER, dữ liệu đầu vào cần phải được chuẩn hóa, biểu diễn hợp lý và trích xuất đặc trưng hiệu quả trước khi đưa vào mô hình huấn luyện.

Do vậy, chương 2 của luận án sẽ tập trung vào phân tích chi tiết các kỹ thuật thường được sử dụng trong việc xây dựng hệ thống EEG\_ER dựa trên học sâu. Phần đầu chương mô tả các kỹ thuật biểu diễn và chuẩn hóa dữ liệu EEG về khuôn dạng phù hợp. Phần tiếp theo, đi sâu vào phân tích một số phương pháp trích chọn đặc trưng thường được sử dụng như FFT, PSD, WT. Tiếp đó, luận án đi sâu vào phân tích hoạt động của ba kiến trúc học sâu thường được sử dụng trong bài toán EEG-ER đó là mô hình CNN, mô hình BiLSTM và mô hình kết hợp giữa CNN và BiLSTM.

Hiệu quả nhận dạng của một mô hình học sâu cho bài toán EEG\_ER phụ thuộc rất lớn vào việc lựa chọn kỹ thuật trích xuất đặc trưng đầu vào và kiến trúc học sâu được sử dụng. Chính vì vậy, phần cuối chương 2 sẽ mô tả các bước tiến hành thực nghiệm để từ đó có được kết luận về mô hình học sâu và phương pháp trích xuất đặc trưng được lựa chọn sao cho hiệu quả nhận dạng là tốt nhất.

Các nội dung trong chương 2 sẽ đặt nền tảng cho chương tiếp theo, nơi nghiên cứu sẽ tập trung vào việc xây dựng và tối ưu hóa mô hình, hướng tới khả năng triển khai thực tế trong các ứng dụng tại thiết bị biên của mạng AIoT.

### **2.1 Biểu diễn và chuẩn hóa dữ liệu**

Để triển khai các mô hình học sâu hiệu quả trên tín hiệu EEG, việc biểu diễn dữ liệu đầu vào cần được đảm bảo:

- Giảm thiểu độ phức tạp tính toán, đặc biệt quan trọng khi triển khai trên các thiết bị tài nguyên hạn chế.
- Tận dụng được thông tin theo nhiều chiều (thời gian, không gian và điện cực).

- Hỗ trợ tốt cho các phương pháp trích xuất đặc trưng.

Có nhiều cách để biểu diễn tín hiệu EEG tùy vào mục đích sử dụng. Trong phạm vi luận án, NCS biểu diễn theo các cách phổ biến gồm:

- Biểu diễn dưới dạng ma trận, phản ánh dữ liệu thô theo từng điện cực và thời gian.
- Biểu diễn theo batch và cửa sổ thời gian, giúp tối ưu hóa quá trình huấn luyện mô hình học sâu.
- Biểu diễn trong miền tần số, khai thác thông tin quan trọng từ phổ tín hiệu.

Dưới đây, NCS trình bày chi tiết từng phương pháp biểu diễn tín hiệu EEG và đánh giá ưu điểm của chúng trong bài toán nhận diện cảm xúc.

### 2.1.1 Biểu diễn tín hiệu EEG dưới dạng ma trận

Tín hiệu EEG thu được từ  $C$  điện cực, với mỗi điện cực cung cấp một chuỗi thời gian rời rạc. Giả sử tín hiệu EEG được thu thập trong khoảng thời gian  $T$  giây với tần số lấy mẫu  $f_s$ , dữ liệu EEG có thể được biểu diễn dưới dạng ma trận hai chiều theo điện cực và thời gian, được định nghĩa như sau:

$$X \in \mathbb{R}^{C \times N} \quad (2.1)$$

Trong đó:  $X$  là ma trận tín hiệu thô, biểu diễn toàn bộ tín hiệu EEG thu được từ  $C$  điện cực trong thời gian  $T$ ;  $N$  là số điểm mẫu theo thời gian, được tính bằng  $N = f_s \times T$ . Mỗi hàng của ma trận  $X$  đại diện cho tín hiệu từ một điện cực EEG cụ thể, còn mỗi cột biểu diễn giá trị tín hiệu tại một thời điểm nhất định.

Biểu diễn tín hiệu EEG dưới dạng ma trận điện cực-thời gian đã được sử dụng rộng rãi trong nhiều nghiên cứu về phân tích tín hiệu não bộ. Theo [33], ma trận  $X$  giữ nguyên cấu trúc tín hiệu EEG thô, giúp bảo toàn thông tin liên quan đến sự thay đổi động của tín hiệu theo thời gian. Các mạng học sâu, đặc biệt là 1D-CNN, có thể dễ dàng áp dụng trên ma trận EEG để trích xuất đặc trưng cục bộ dọc theo trục thời gian [129]. Biểu diễn dạng ma trận phù hợp để áp dụng các phương pháp phân tích tín hiệu như Biến đổi Fourier nhanh (FFT), Mật độ phổ công suất (PSD), và Biến đổi Wavelet (WT) [102].

Tuy nhiên, phương pháp biểu diễn này cũng tồn tại một số hạn chế. Khi số điện cực EEG lớn (ví dụ,  $C = 64$ ), ma trận có kích thước lớn, gây khó khăn trong xử lý và lưu trữ trên các hệ thống có tài nguyên giới hạn. Bên cạnh đó, việc biểu diễn trong miền thời gian có thể làm mất đi một số thông tin quan trọng về sự thay đổi động của tín hiệu theo tần số.

### 2.1.2 Biểu diễn tín hiệu EEG theo batch và cửa sổ thời gian

Trong học sâu, dữ liệu EEG không chỉ cần được biểu diễn dưới dạng ma trận hai chiều mà còn cần được tổ chức thành batch để tối ưu hóa quá trình huấn luyện. Khi đó, dữ liệu EEG được biểu diễn dưới dạng tensor ba chiều:

$$X_{\text{batch}} \in \mathbb{R}^{B \times C \times N} \quad (2.2)$$

trong đó:  $B$  là kích thước batch (số lượng mẫu EEG được xử lý đồng thời);  $C$  là số điện cực EEG;  $N$  là số điểm mẫu trong một khoảng thời gian nhất định.

Ví dụ, nếu có một tập dữ liệu EEG gồm 5120 mẫu, với batch size  $B = 32$ , ta có tổng cộng:  $5120/32 = 160$  batch và mỗi batch sẽ chứa dữ liệu của 32 mẫu EEG khác nhau, giúp cải thiện hiệu suất tính toán trong quá trình huấn luyện mô hình.

Mặc dù batch giúp tối ưu hóa tính toán, nhưng mô hình học sâu vẫn cần khai thác đặc trưng động của tín hiệu EEG theo từng khoảng thời gian nhỏ để đảm bảo không mất đi thông tin quan trọng. Vì vậy, tín hiệu EEG thường được chia thành các cửa sổ thời gian nhỏ hơn thay vì xử lý toàn bộ tín hiệu trong một lần. Khi đó, mỗi cửa sổ có kích thước:

$$X_{\text{windowed}} \in \mathbb{R}^{C \times W} \quad (2.3)$$

trong đó:  $W$  là số điểm mẫu trong cửa sổ thời gian,  $W = f_s \times t$  ( $t$  là độ dài cửa sổ thời gian); Các cửa sổ có thể chồng lấn lên nhau với bước nhảy  $S$  (stride) để đảm bảo không mất thông tin quan trọng.

Ví dụ, nếu  $f_s = 128$  và  $t = 2$  giây, khi đó  $W = f_s \times t = 256$ . Và giả sử tín hiệu EEG được thu thập từ 14 điện cực, mỗi cửa sổ có kích thước:  $(14 \times 256)$ .

Tín hiệu EEG sau khi chia thành các cửa sổ thời gian có thể được kết hợp với batch để tạo thành tensor dữ liệu phục vụ huấn luyện:

$$X_{\text{process}} \in \mathbb{R}^{B \times C \times W} \quad (2.4)$$

Ví dụ, với batch size = 32, ta có tensor dữ liệu:  $X_{\text{process}} \in \mathbb{R}^{32 \times 14 \times 256}$ .

Khi đó, mỗi batch sẽ chứa dữ liệu từ 32 mẫu EEG khác nhau, mỗi mẫu được chia thành nhiều cửa sổ thời gian, đảm bảo rằng mô hình có thể học được các biến đổi theo thời gian trong tín hiệu EEG. Với việc tổ chức tín hiệu EEG theo batch và cửa sổ thời gian, dữ liệu được xây dựng cách biểu diễn phù hợp cho các mô hình học sâu.

Tuy nhiên, dữ liệu EEG chứa thông tin không chỉ trong miền thời gian mà còn trong

miền tần số. Do bản chất dao động của tín hiệu EEG, [33] đã chỉ ra rằng nhiều đặc trưng có ý nghĩa hơn nằm trong phổ tần số của tín hiệu thay vì tín hiệu gốc theo thời gian. Do đó, một trong những bước quan trọng trong xử lý tín hiệu EEG là tìm cách biểu diễn đặc trưng dữ liệu trong đó tận dụng được cả miền tần số. Vấn đề này sẽ được trình bày trong phần tiếp theo của chương.

## **2.2 Lựa chọn các phương pháp trích chọn đặc trưng**

### **2.2.1 Cơ sở lựa chọn các phương pháp trích chọn đặc trưng**

Để khai thác tốt hơn các đặc trưng có ý nghĩa từ tín hiệu EEG, một bước quan trọng trong tiền xử lý dữ liệu là chuyển đổi tín hiệu từ miền thời gian sang miền tần số. Việc này giúp xác định các dao động đặc trưng trong từng dải tần EEG, qua đó phản ánh chính xác hơn hoạt động thần kinh của não bộ. Các nghiên cứu đã chỉ ra rằng tín hiệu EEG chứa thông tin quan trọng không chỉ trong miền thời gian mà còn trong miền tần số, nơi các dao động EEG được phân tách thành các dải tần khác nhau. Cụ thể, các dải tần Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), Gamma (>30 Hz) có liên quan chặt chẽ đến trạng thái nhận thức và cảm xúc của con người [33, 102]. Việc trích xuất đặc trưng từ các dải tần này đã được chứng minh là một bước quan trọng trong các bài toán nhận diện cảm xúc EEG [57, 7].

Có nhiều phương pháp được sử dụng để trích xuất đặc trưng từ tín hiệu EEG trong miền tần số, bao gồm: Biến đổi Fourier nhanh (FFT), Mật độ phổ công suất (PSD), Biến đổi Wavelet (WT), Biến đổi Fourier ngắn hạn (STFT), Hilbert-Huang Transform (HHT), và Empirical Mode Decomposition (EMD). Tuy nhiên, không phải phương pháp nào cũng phù hợp cho các ứng dụng yêu cầu hiệu suất tính toán cao và triển khai trên thiết bị có tài nguyên hạn chế. Do đó, việc lựa chọn phương pháp trích xuất đặc trưng trong nghiên cứu này tuân theo ba tiêu chí sau:

1. Các phương pháp được lựa chọn phải được sử dụng rộng rãi trong nghiên cứu liên quan đến nhận dạng cảm xúc từ EEG nhằm đảm bảo độ tin cậy.
2. Do mục tiêu nghiên cứu là triển khai mô hình trên thiết bị giới hạn về tài nguyên, các phương pháp trích xuất đặc trưng cần có độ phức tạp tính toán hợp lý để đảm bảo tốc độ xử lý gần thời gian thực.
3. Các phương pháp phải đại diện tốt cho phổ tần số của tín hiệu EEG, và có thể giữ lại thông tin thời gian quan trọng.

Bảng 2.1 tổng hợp so sánh các phương pháp trích xuất đặc trưng chính. Dựa trên ba tiêu chí đã đặt ra, có thể thấy rằng FFT là phương pháp phù hợp nhất về hiệu suất tính toán, PSD giúp giảm nhiễu và hỗ trợ nhận dạng tốt hơn, trong khi Wavelet có khả năng



giữ thông tin thời gian nhưng có chi phí tính toán cao hơn. Những kết quả này sẽ là nền tảng cho các bước xử lý tín hiệu EEG trong nghiên cứu tiếp theo.

**Bảng 2.1** Bảng so sánh các phương pháp trích xuất đặc trưng EEG

| Phương pháp | Đặc trưng chính               | Giữ thông tin tần số | Giữ thông tin thời gian | Khả năng lọc nhiễu                    | Tốc độ tính toán | Khả năng triển khai   | Nguồn     |
|-------------|-------------------------------|----------------------|-------------------------|---------------------------------------|------------------|-----------------------|-----------|
| FFT         | Phân tích phổ tín hiệu        | Xuất sắc             | Không                   | Không hiệu quả với tín hiệu phi tuyến | Rất nhanh        | Phù hợp               | [33, 102] |
| PSD         | Phân tích năng lượng tín hiệu | Xuất sắc             | Không                   | Tốt (giảm nhiễu bằng trung bình hóa)  | Nhanh            | Phù hợp               | [33, 102] |
| Wavelet     | Phân tích tần số và thời gian | Tốt                  | Xuất sắc                | Tốt                                   | Chậm hơn         | Khó triển khai mobile | [33, 102] |
| STFT        | Phân tích thời gian - tần số  | Tốt                  | Tốt                     | Trung bình                            | Chậm hơn         | Khó triển khai mobile | [33, 102] |
| EMD         | Phân tích chế độ dao động     | Tốt                  | Xuất sắc                | Tốt (tách nhiễu nội tại)              | Rất chậm         | Không phù hợp         | [129]     |
| HHT         | Phân tích tần số tức thời     | Xuất sắc             | Xuất sắc                | Tốt                                   | Rất chậm         | Không phù hợp         | [129]     |

### 2.2.2 Trích xuất đặc trưng bằng biến đổi Fourier nhanh (FFT)

Tín hiệu EEG là một chuỗi dữ liệu biến đổi theo thời gian, nhưng phần lớn các thông tin quan trọng về hoạt động não bộ lại được biểu hiện trong miền tần số. Do đó, một trong những phương pháp phổ biến nhất để trích xuất đặc trưng từ tín hiệu EEG là sử dụng Biến đổi Fourier nhanh (FFT - *Fast Fourier Transform*), cho phép chuyển đổi tín hiệu từ miền thời gian sang miền tần số một cách hiệu quả.

Biến đổi Fourier nhanh là một phiên bản tối ưu hoá của Biến đổi Fourier rời rạc (DFT - *Discrete Fourier Transform*), giúp tính toán phổ tín hiệu với độ phức tạp  $O(N \log N)$  thay vì  $O(N^2)$  của DFT truyền thống [7]. Điều này làm cho FFT trở thành lựa chọn phù hợp trong các hệ thống phân tích EEG có dung lượng dữ liệu lớn, đặc biệt khi triển khai trên các thiết bị tài nguyên hạn chế. Phương pháp tính toán FFT cho tín

hiệu EEG rời rạc  $x(n)$  từ biến đổi Fourier rời rạc (DFT) được định nghĩa như sau:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}}, \quad k = 0, 1, 2, \dots, N-1 \quad (2.5)$$

Trong đó,  $X(k)$  là phổ tần số của tín hiệu tại chỉ số  $k$ ,  $N$  là số điểm mẫu trong cửa sổ thời gian,  $x(n)$  là giá trị tín hiệu EEG tại thời điểm  $n$ . Mặc dù DFT có thể cung cấp thông tin đầy đủ về thành phần tần số của tín hiệu EEG, nhưng nó có độ phức tạp tính toán cao khiến cho việc áp dụng trên dữ liệu EEG lớn trở nên không khả thi. Để giảm chi phí tính toán, FFT được sử dụng nhằm giảm đáng kể độ phức tạp tính toán xuống còn  $O(N \log N)$ . FFT vẫn giữ nguyên lý hoạt động của DFT, nhưng tối ưu hóa thuật toán tính toán, từ đó giúp trích xuất đặc trưng tần số một cách hiệu quả.

Việc áp dụng FFT trên toàn bộ tín hiệu có thể làm mất đi thông tin quan trọng về sự thay đổi đặc trưng EEG theo thời gian. Để khắc phục vấn đề này, một phương pháp phổ biến là áp dụng phân tích theo cửa sổ thời gian (*windowing*), trong đó tín hiệu EEG được chia thành các đoạn nhỏ trước khi thực hiện FFT. Cụ thể, nếu ta chia tín hiệu EEG thành các cửa sổ có độ dài  $W$  điểm mẫu với bước dịch  $S$ , thì mỗi cửa sổ được xử lý riêng biệt để trích xuất đặc trưng tần số tại từng khoảng thời gian cụ thể. Biểu thức FFT cho một cửa sổ có độ dài  $W$  là:

$$X_w(k) = \sum_{n=0}^{W-1} x_w(n)e^{-j\frac{2\pi kn}{W}}, \quad k = 0, 1, 2, \dots, W-1 \quad (2.6)$$

Khi đó,  $X_w(k)$  là phổ tần số tại cửa sổ thứ  $w$  và  $x_w(n)$  là giá trị tín hiệu EEG sau khi được phân đoạn. Do FFT có tính đối xứng khi áp dụng lên tín hiệu thực, ta chỉ giữ lại 1/2 giá trị phổ thực có ích. Để trích xuất đặc trưng có ý nghĩa từ phổ FFT, ta chỉ tập trung vào các băng tần EEG chính, vì mỗi băng tần có liên quan đến một trạng thái thần kinh cụ thể. Năng lượng của mỗi băng tần được tính bằng trung bình phương biên độ FFT trong phạm vi tần số tương ứng:

$$P_{\text{band}} = \sum_{f=f_{\min}}^{f_{\max}} |X(f)|^2 \quad (2.7)$$

Trong đó,  $f_1$  và  $f_2$  là giới hạn dưới và giới hạn trên của băng tần.

Sau khi thực hiện FFT trên từng cửa sổ tín hiệu và tính toán năng lượng của các băng tần EEG, dữ liệu được biểu diễn dưới dạng tensor ba chiều ( $B \times C \times F$ ). Với  $B$  là số lượng cửa sổ trong batch,  $C$  là số điện cực EEG và  $F$  là số băng tần trích xuất từ FFT. Dữ liệu này sau đó có thể được đưa vào mô hình học sâu để nhận diện cảm xúc dựa trên các đặc trưng tần số của EEG.

Việc sử dụng FFT trong phân tích EEG mang lại nhiều lợi ích, giúp trích xuất thông tin tần số một cách hiệu quả, từ đó cải thiện hiệu suất của mô hình học sâu. Tuy nhiên, bên cạnh FFT, một số nghiên cứu gần đây cũng đã nêu ra hạn chế của FFT đó là khả năng biểu diễn ở nhiều mức độ phân giải khác nhau, hay nắm bắt sự thay đổi cục bộ của tín hiệu EEG theo thời gian [33].

### 2.2.3 Trích xuất năng lượng tín hiệu EEG bằng Mật độ phổ công suất (PSD)

Mật độ phổ công suất (Power Spectral Density - PSD) là một phương pháp quan trọng trong phân tích tín hiệu EEG, giúp đo lường sự phân bố năng lượng của tín hiệu theo tần số. Không giống như FFT, phương pháp này tập trung vào việc ước lượng năng lượng trung bình của tín hiệu trong từng băng tần EEG, từ đó phản ánh mức độ hoạt động nào bộ trong các dải tần quan trọng [2, 7].

PSD là một kỹ thuật hữu ích vì nó cung cấp đặc trưng tần số ổn định hơn so với FFT. Thay vì chỉ phân tích thành phần tần số tại một thời điểm, PSD tính toán mật độ năng lượng trung bình trong khoảng thời gian nhất định, giúp giảm nhiễu và cải thiện độ chính xác trong phân tích tín hiệu EEG. Có nhiều phương pháp để tính toán PSD từ tín hiệu EEG, trong đó phương pháp Welch là một trong những phương pháp phổ biến nhất nhờ khả năng giảm nhiễu và cải thiện độ mượt của phổ tín hiệu. Welch hoạt động bằng cách chia tín hiệu EEG thành nhiều cửa sổ có chồng lấn để giảm thiểu sự biến động của phổ tín hiệu trong từng cửa sổ. Công thức tính PSD dựa trên biến đổi Fourier của từng cửa sổ tín hiệu như sau:

$$P_{\text{PSD}}(f) = \frac{1}{M} \sum_{m=1}^M \frac{|X_m(f)|^2}{T} \quad (2.8)$$

Trong đó,  $P_{\text{PSD}}(f)$  là mật độ phổ công suất tại tần số  $f$ ,  $X_m(f)$  là biến đổi Fourier của tín hiệu tại tần số  $f$  trong cửa sổ thứ  $m$ ;  $M$  là số cửa sổ được sử dụng để tính trung bình và  $T$  là độ dài mỗi cửa sổ.

Tương tự như FFT, PSD cũng được áp dụng trên các cửa sổ tín hiệu EEG có độ dài  $W$ , với mỗi cửa sổ được xử lý độc lập để tính toán năng lượng của từng băng tần EEG. Nhờ khả năng giảm nhiễu và làm mượt phổ tín hiệu, PSD thường được sử dụng để tăng độ chính xác của mô hình học sâu trong nhận diện cảm xúc EEG. Mặc dù PSD có hiệu suất thấp hơn FFT, nhưng nó lại cung cấp thông tin ổn định hơn, đặc biệt khi phân tích tín hiệu EEG trong điều kiện nhiễu cao.

Trong phần tiếp theo, nghiên cứu sẽ tập trung vào biến đổi Wavelet (WT), một phương pháp mạnh mẽ hơn giúp khai thác đồng thời đặc trưng thời gian - tần số từ tín hiệu EEG.

#### 2.2.4 Trích xuất đặc trưng thời gian - tần số EEG bằng Biến đổi Wavelet (WT)

Biến đổi Wavelet (Wavelet Transform - WT) là một phương pháp mạnh mẽ trong phân tích tín hiệu EEG, giúp trích xuất đặc trưng thời gian - tần số, cho phép mô hình học sâu khai thác đồng thời cả hai miền thông tin này. Không giống như FFT hay PSD, vốn chỉ cung cấp thông tin về thành phần tần số của tín hiệu, WT có khả năng phân tích tín hiệu EEG ở nhiều mức độ phân giải khác nhau, giúp nhận diện các đặc trưng quan trọng mà không làm mất thông tin thời gian.

Một trong những ưu điểm quan trọng của WT là khả năng phân tích tín hiệu phi tuyến (non-stationary), đặc biệt phù hợp với tín hiệu EEG do tính chất biến đổi liên tục của nó. WT có thể phát hiện các biến đổi trong phổ tín hiệu theo thời gian, giúp cải thiện độ chính xác của các hệ thống nhận diện cảm xúc từ EEG.

$$WT(a, b) = \frac{1}{\sqrt{|a|}} \sum_{n=0}^{N-1} X[n] \psi^* \left( \frac{n-b}{a} \right) \quad (2.9)$$

Trong thực tế, để áp dụng trên tín hiệu EEG, thường sử dụng Biến đổi Wavelet rời rạc (DWT), trong đó tín hiệu EEG được phân tách thành nhiều mức độ phân giải khác nhau, giúp mô hình học sâu có thể học được cả đặc trưng tần số cao và thấp. Sau khi áp dụng WT lên từng cửa sổ tín hiệu EEG, ta có thể trích xuất các đặc trưng liên quan đến từng băng tần EEG, tương tự như FFT và PSD. Dữ liệu EEG sau trích xuất đặc trưng từ Wavelet được biểu diễn dưới dạng tensor 3 chiều  $B \times C \times L$  với  $L$  là số mức phân giải.

Biến đổi Wavelet mang lại lợi thế lớn trong việc phân tích tín hiệu EEG vì nó giữ lại được cả thông tin thời gian và tần số, điều này đặc biệt quan trọng đối với tín hiệu não bộ có tính chất động và phi tuyến. Tuy nhiên, WT có chi phí tính toán cao hơn so với FFT và PSD, khiến việc triển khai trên các thiết bị tài nguyên hạn chế trở nên khó khăn hơn. Vì vậy, trong nghiên cứu này, WT được sử dụng như một phương pháp bổ sung để kiểm chứng hiệu quả của FFT và PSD.

Để đánh giá mức độ ảnh hưởng của các phương pháp trích xuất đặc trưng đến hiệu suất nhận dạng cảm xúc từ tín hiệu EEG, NCS đã thực hiện các thử nghiệm thực nghiệm trên tập dữ liệu EEG chuẩn DEAP. Kết quả thực nghiệm cho thấy rằng hiệu suất nhận dạng không chỉ phụ thuộc vào phương pháp trích xuất đặc trưng mà còn liên quan mật thiết đến cấu trúc của mô hình học sâu được sử dụng. Do đó, để có thể đưa ra kết luận chính xác về phương pháp tối ưu nhất cho bài toán EEG\_ER, phần tiếp theo của luận án sẽ tiếp tục phân tích, thử nghiệm và xác định sự kết hợp tối ưu giữa phương pháp trích xuất đặc trưng và mô hình học sâu.

## 2.3 Lựa chọn kiến trúc mô hình nền tảng

### 2.3.1 Cơ sở lựa chọn mô hình học sâu

Tín hiệu EEG là một dạng dữ liệu sinh học có tính thời gian, phản ánh hoạt động não bộ trong các trạng thái cảm xúc khác nhau. Do đó, mô hình học sâu cần có khả năng khai thác hiệu quả đặc trưng của tín hiệu này. Các nghiên cứu trước đây đã chỉ ra rằng việc lựa chọn mô hình học sâu phù hợp phải dựa trên cách tín hiệu EEG được trích xuất và đặc trưng của dữ liệu đầu vào [67, 29, 52].

Sau khi trích xuất đặc trưng từ tín hiệu EEG, dữ liệu có thể được biểu diễn theo ba dạng chính: (1) FFT và PSD, chủ yếu chứa thông tin tần số và được biểu diễn dưới dạng ma trận phổ, (2) Wavelet Transform, chứa cả thông tin thời gian và tần số, tạo thành dữ liệu có cấu trúc phức tạp hơn. Điều này dẫn đến yêu cầu cao đối với lựa chọn mô hình học sao cho khả năng học được cả đặc trưng không gian từ dữ liệu phổ tần số và đặc trưng thời gian từ dữ liệu có quan hệ tuần tự. Các nghiên cứu tổng quan đã chỉ ra rằng CNN là mô hình phù hợp nhất để xử lý dữ liệu dạng phổ tần số, trong khi các mô hình hồi tiếp như BiLSTM hoạt động hiệu quả hơn với dữ liệu có đặc trưng tuần tự [44, 5, 48].

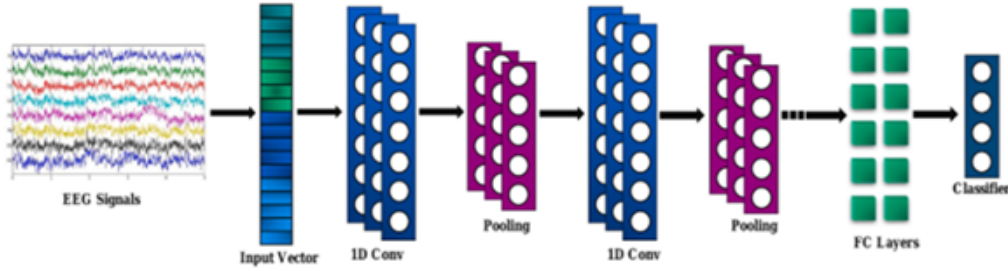
Trước khi lựa chọn CNN, BiLSTM và CNN+BiLSTM, nghiên cứu sinh cũng đã xem xét một số mô hình học sâu phổ biến: Mô hình MLP (Multi-Layer Perceptron) tuy có khả năng học đặc trưng tổng quát nhưng không tận dụng được đặc trưng không gian hoặc thời gian của tín hiệu EEG, do đó chỉ hoạt động tốt trên dữ liệu có cấu trúc đơn giản [126]. Mô hình RNN thuần túy (Vanilla RNN) không có khả năng học được quan hệ dài hạn trong dữ liệu EEG và dễ gặp vấn đề vanishing gradient, dẫn đến hiệu suất kém hơn [50]. Trong khi đó, các kiến trúc Transformer có khả năng học quan hệ dài hạn, song phiên bản self-attention chuẩn có độ phức tạp  $\mathcal{O}(T^2)$  theo độ dài của chuỗi, kéo theo chi phí bộ nhớ và độ trễ suy luận cao [52]. Tuy đã có các biến thể Transformer “nhẹ”, nút thắt vẫn nằm ở chi phí self-attention theo chuỗi. Bên cạnh đó, phần lớn các mô hình Transformer “nhẹ” được thiết kế cho ảnh/ngôn ngữ và cần tiền huấn luyện lớn, các bộ dữ liệu EEG quy mô vừa–nhỏ, chúng dễ kém ổn định/khái quát hơn so với CNN-1D vốn đã mã hoá sẵn tính cục bộ đa thang theo thời gian [75].

Từ các tài liệu và phân tích trên, nghiên cứu sinh lựa chọn ba mô hình vì khả năng khai thác hiệu quả đặc trưng EEG. CNN là mô hình lý tưởng để học đặc trưng không gian của tín hiệu EEG do khả năng phát hiện các mẫu quan trọng trong phổ tần số của EEG. Mô hình BiLSTM giúp khai thác tốt mối quan hệ thời gian trong tín hiệu EEG, đặc biệt với các phương pháp trích xuất có chứa thông tin tuần tự mạnh như Wavelet.

Trong các phần tiếp theo, nghiên cứu sinh sẽ trình bày chi tiết về từng mô hình này và đánh giá hiệu suất thực tế của chúng.

### 2.3.2 Mạng CNN

CNN là một trong những kiến trúc mạng học sâu phổ biến nhất, đặc biệt phù hợp với dữ liệu có cấu trúc không gian như hình ảnh và tín hiệu EEG. CNN đã được áp dụng rộng rãi trong các nghiên cứu nhận dạng cảm xúc từ EEG nhờ khả năng tự động trích xuất đặc trưng từ dữ liệu mà không cần quá trình thiết kế đặc trưng thủ công. Thông qua các phép toán tích chập và gộp, CNN giúp phát hiện các mẫu đặc trưng trong tín hiệu EEG, hỗ trợ cải thiện độ chính xác của các hệ thống nhận dạng cảm xúc [44].



**Hình 2.1** Kiến trúc mạng 1D-CNN [51]

Trong bài toán EEG, CNN có thể hoạt động trên dữ liệu 1D (tín hiệu theo thời gian), 2D (biểu diễn không gian-điện cực EEG) hoặc 3D (kết hợp không gian-thời gian-tần số). 1D-CNN thường được sử dụng khi làm việc với tín hiệu EEG theo chuỗi thời gian, trong khi 2D-CNN có thể được áp dụng khi EEG được biểu diễn dưới dạng bản đồ topo điện cực (topographic maps) nhằm khai thác đặc trưng không gian giữa các điện cực. Gần đây, 3D-CNN đã được đề xuất để tận dụng đồng thời thông tin không gian, thời gian và tần số, giúp mô hình học tốt hơn các đặc trưng phức hợp của tín hiệu EEG [48, 12].

CNN sử dụng một chuỗi các phép toán để xử lý tín hiệu EEG, bao gồm phép tích chập, bước nhảy, đệm, gộp và kích hoạt phi tuyến. Các phép toán này giúp CNN học các đặc trưng quan trọng trong tín hiệu EEG mà không cần thiết kế đặc trưng thủ công.

**Phép tích chập (Convolutional Operation):** Là nền tảng của CNN, giúp mô hình học các mẫu đặc trưng trong tín hiệu EEG bằng cách sử dụng các bộ lọc (kernels) [62]. Trong 1D-CNN, một bộ lọc có kích thước  $k$  sẽ trượt dọc theo chuỗi thời gian và tính toán một giá trị mới tại mỗi vị trí bằng cách lấy tích vô hướng giữa bộ lọc và một vùng con của tín hiệu. Công thức toán học của phép tích chập 1D được biểu diễn như sau:

$$y[i] = \sum_{j=0}^k w_j x[i+j] \quad (2.10)$$

Trong đó:  $x$  là tín hiệu đầu vào,  $w$  là trọng số của bộ lọc tích chập với  $k$  kích thước bộ lọc và  $y(i)$  là đầu ra tại vị trí  $i$ .

**Bước nhảy (Stride):** Là số bước mà bộ lọc di chuyển trên dữ liệu đầu vào. Khi bước nhảy tăng, kích thước đầu ra giảm. Công thức tính kích thước đầu ra khi sử dụng stride  $S$  là:

$$L_{\text{out}} = \left\lfloor \frac{L_{\text{in}} - k}{S} \right\rfloor + 1 \quad (2.11)$$

Với  $L_{\text{in}}$  và  $L_{\text{out}}$  là chiều dài tín hiệu vào và ra.

**Đệm (Padding):** Để kiểm soát kích thước của đầu ra, có thể thêm giá trị 0 xung quanh biên của tín hiệu EEG. Các phương pháp padding phổ biến gồm:

*Valid padding:* Không thêm giá trị nào, dẫn đến đầu ra nhỏ hơn đầu vào.

*Same padding:* Thêm đủ số 0 để đầu ra có cùng kích thước với đầu vào.

**Lớp gộp (Pooling Layer):** Lớp gộp giúp giảm kích thước đầu ra, giảm số lượng tham số và ngăn chặn overfitting. Hai phương pháp phổ biến gồm:

*Max Pooling:* Chọn giá trị lớn nhất trong một cửa sổ con.

*Average Pooling:* Tính giá trị trung bình trong cửa sổ.

Công thức tổng quát của Max Pooling là:

$$y[i] = \max(x[i], x[i+1], \dots, x[i+k-1]) \quad (2.12)$$

**Hàm kích hoạt (Activation Function):** Sau mỗi lớp tích chập, một hàm kích hoạt phi tuyến được áp dụng để tăng khả năng học phi tuyến của mô hình. Hàm kích hoạt phổ biến nhất là ReLU:

$$f(x) = \max(0, x) \quad (2.13)$$

Hàm này giúp loại bỏ giá trị âm, cải thiện hiệu suất tính toán và giảm vấn đề vanishing gradient.

**Lớp kết nối đầy đủ (Fully Connected Layer):** Sau khi tín hiệu EEG được xử lý qua các lớp tích chập và gộp, dữ liệu đầu ra thường có dạng ma trận nhiều chiều. Để chuẩn bị cho quá trình phân loại hoặc hồi quy, dữ liệu này cần được chuyển đổi thành một vector một chiều thông qua quá trình làm phẳng (flattening). Vector này sau đó được đưa vào các lớp kết nối đầy đủ, nơi mỗi nơ-ron được kết nối với tất cả các nơ-ron trong lớp trước đó.

Lớp Fully Connected thực hiện phép toán:

$$h^{(l)} = f \left( W^{(l)} h^{(l-1)} + b^{(l)} \right) \quad (2.14)$$

Trong đó:  $h^{(l)}$ : đầu ra của lớp kết nối đầy đủ thứ  $l$ ;  $W^{(l)}$  là ma trận trọng số kết nối giữa lớp  $(l-1)$  và  $l$ ;  $b^{(l)}$ : vector bias của lớp  $l$ ;  $f$ : hàm kích hoạt, thường dùng ReLU, Softmax tùy thuộc vào bài toán cụ thể.

Trong ngữ cảnh phân loại cảm xúc từ tín hiệu EEG, lớp Fully Connected cuối cùng thường sử dụng hàm Softmax để chuyển đổi đầu ra thành xác suất cho các lớp cảm xúc khác nhau.

CNN đã được chứng minh là hiệu quả trong việc xử lý tín hiệu EEG nhờ khả năng trích xuất đặc trưng không gian từ dữ liệu tần số. NCS đã nghiên cứu mô hình CNN và được thử nghiệm với dữ liệu vào là FFT, PSD và Wavelet để so sánh hiệu quả trích xuất đặc trưng tương ứng.

### 2.3.3 Mạng BiLSTM

Mạng RNN là một lớp mô hình học sâu được thiết kế để xử lý dữ liệu chuỗi, trong đó đầu ra tại mỗi thời điểm phụ thuộc vào trạng thái trước đó. Tuy nhiên, các mô hình RNN truyền thống thường gặp phải vấn đề tiêu biến hoặc bùng nổ gradient khi xử lý các chuỗi dài, dẫn đến khả năng ghi nhớ ngữ cảnh dài hạn bị suy giảm [62, 97, 12]. Để khắc phục hạn chế này, mạng bộ nhớ dài ngắn (Long Short-Term Memory – LSTM) được đề xuất, cho phép duy trì và điều khiển luồng thông tin thông qua các cổng: cổng vào, cổng quên và cổng đầu ra. Nhờ cơ chế lưu giữ trạng thái bộ nhớ có kiểm soát, LSTM có thể học được các phụ thuộc dài hạn hiệu quả hơn trong chuỗi tín hiệu.

Mở rộng từ LSTM, mạng LSTM hai chiều (BiLSTM) cho phép truy xuất ngữ cảnh cả từ quá khứ và tương lai bằng cách đồng thời sử dụng hai chuỗi LSTM: một theo chiều thời gian thuận và một theo chiều ngược. Tại mỗi thời điểm  $t$ , đầu ra của BiLSTM là kết hợp của hai trạng thái ẩn [44]:

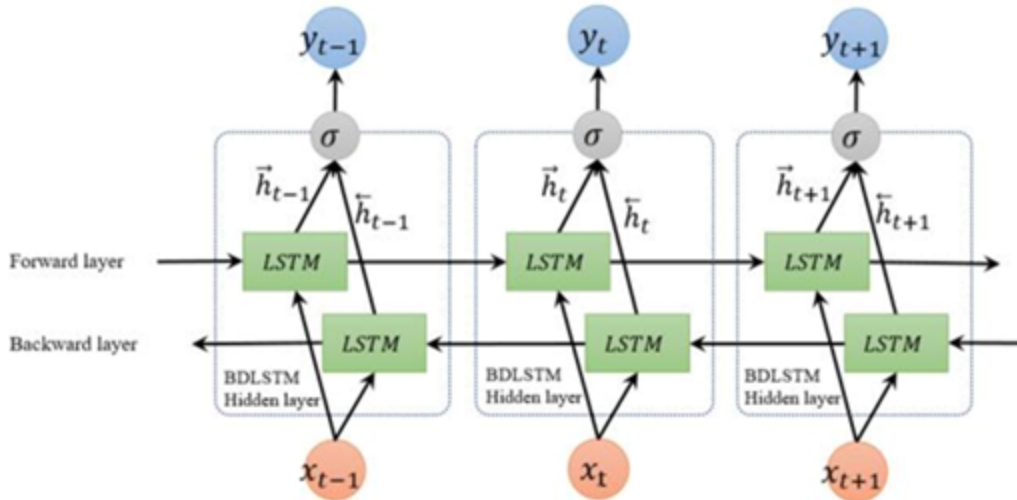
$$h_t = \left( \overrightarrow{h}_t; \overleftarrow{h}_t \right) \quad (2.15)$$

Trong đó,  $\overrightarrow{h}_t, \overleftarrow{h}_t$  lần lượt là đầu ra của LSTM theo hướng thuận và ngược thời gian. BiLSTM mang lại nhiều ưu điểm vượt trội trong việc xử lý tín hiệu EEG so với các mô hình hồi tiếp truyền thống. Nhờ khả năng học cả hai hướng thời gian, BiLSTM có thể nắm bắt các quan hệ dài hạn trong dữ liệu EEG một cách hiệu quả hơn so với LSTM một chiều, vốn chỉ học thông tin từ quá khứ đến hiện tại. Điều này giúp BiLSTM tổng hợp thông tin tốt hơn và cải thiện độ chính xác trong các bài toán nhận diện cảm xúc.

Tuy nhiên, so với LSTM một chiều, BiLSTM có chi phí tính toán cao hơn do phải



xử lý hai mạng LSTM đồng thời. Việc lựa chọn số lượng đơn vị ẩn hợp lý là yếu tố quan trọng để tránh hiện tượng quá khớp khi xử lý dữ liệu EEG có nhiễu cao [72, 51, 87].



**Hình 2.2** Mạng BiLSTM - Bidirectional Long Short-Term Memory [26]

Trong các nghiên cứu EEG gần đây, BiLSTM được sử dụng nhằm khai thác các mối quan hệ ngữ cảnh phức tạp giữa các đoạn tín hiệu EEG theo thời gian. Tuy nhiên, nhược điểm chính của BiLSTM là số lượng tham số lớn, thời gian huấn luyện kéo dài và độ phức tạp tính toán cao, đặc biệt khi kết hợp với đầu vào có nhiều chiều như dải tần  $\times$  điện cực  $\times$  thời gian.

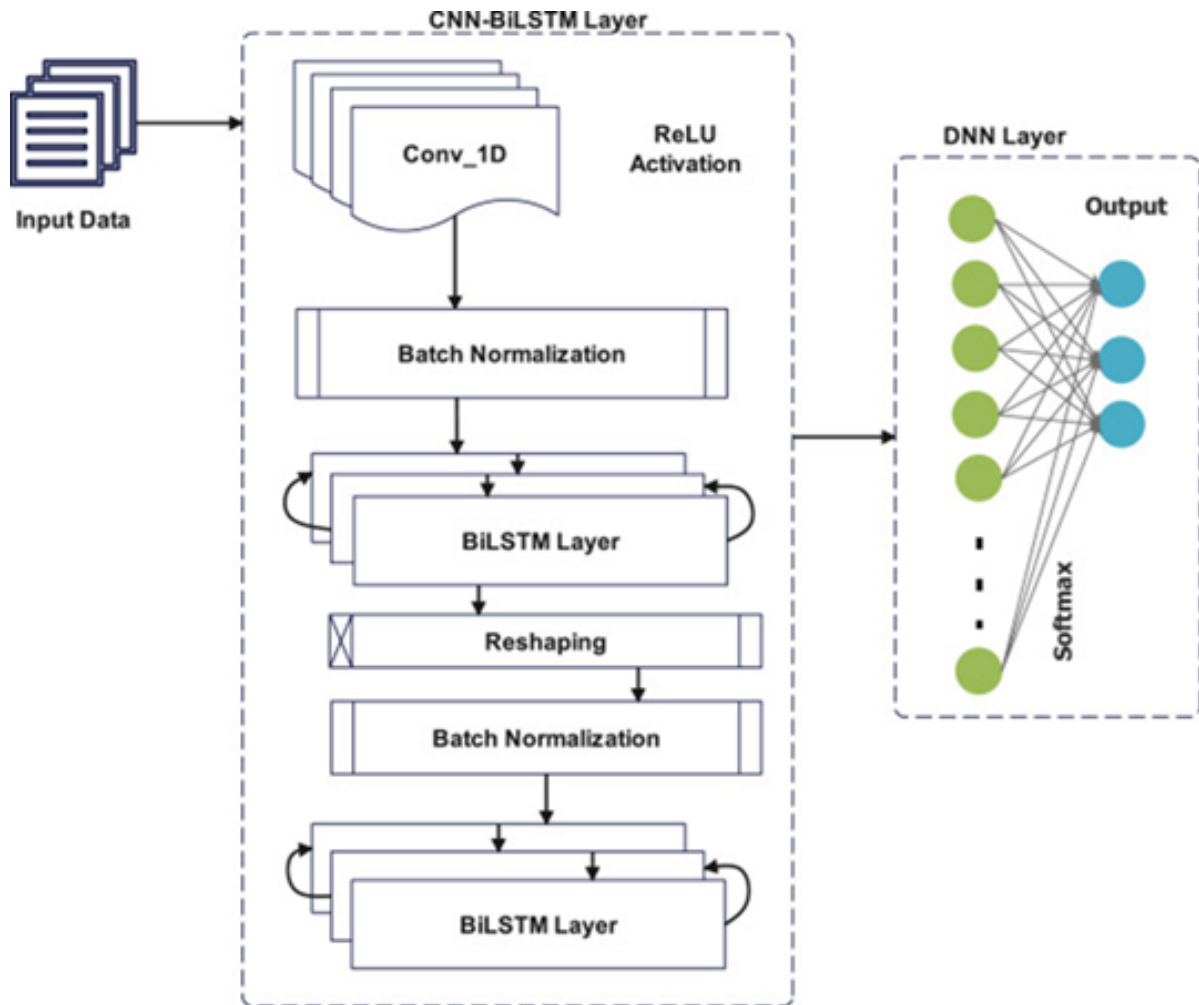
#### 2.3.4 Mô hình kết hợp CNN và BiLSTM

Mô hình kết hợp CNN và BiLSTM là một trong những hướng tiếp cận hiệu quả nhằm tận dụng cả đặc trưng không gian và thời gian của tín hiệu EEG. CNN có khả năng trích xuất đặc trưng không gian từ các phổ tần số của tín hiệu EEG, trong khi BiLSTM giúp mô hình học được các mối quan hệ tuần tự dài hạn trong dữ liệu. Sự kết hợp giữa hai mô hình này tạo ra một hệ thống mạnh mẽ hơn so với việc sử dụng riêng lẻ từng mô hình, giúp cải thiện độ chính xác trong bài toán nhận diện cảm xúc từ tín hiệu EEG [44, 51]. Như mô tả trong Hình 2.3, kiến trúc mô hình bao gồm hai giai đoạn chính:

**Giai đoạn 1:** Trích xuất đặc trưng không gian bằng mạng tích chập (CNN), mô hình thực hiện trích xuất đặc trưng không gian từ dữ liệu EEG đã được chuyển đổi sang miền tần số. Các phương pháp như Biến đổi Fourier nhanh (FFT), Mật độ phổ công suất (PSD) hoặc Biến đổi Wavelet (WT) được áp dụng để chuyển đổi tín hiệu EEG từ miền thời gian sang miền tần số, giúp mô hình học sâu có thể khai thác tốt hơn các đặc trưng quan trọng trong dữ liệu.

**Giai đoạn 2:** Đầu ra từ CNN được đưa vào mạng BiLSTM, nơi thông tin tín hiệu

EEG được xử lý theo cả hai hướng thời gian, giúp mô hình học được mối quan hệ động của tín hiệu EEG. Đầu ra của BiLSTM sau đó được kết nối với một lớp Dense để thực hiện phân loại cảm xúc. Kiến trúc này tận dụng khả năng của CNN trong phát hiện đặc trưng không gian và BiLSTM trong phân tích chuỗi thời gian, tạo nên một mô hình hiệu quả cho nhận diện cảm xúc từ EEG.



**Hình 2.3** Cấu trúc của mô hình kết hợp CNN với BiLSTM [43]

Bằng sự kết hợp này, mô hình CNN-BiLSTM mang lại nhiều ưu điểm vượt trội so với việc sử dụng riêng CNN hoặc BiLSTM. CNN có khả năng trích xuất đặc trưng không gian hiệu quả từ phổ tín hiệu EEG, trong khi BiLSTM có khả năng học tốt hơn mối quan hệ thời gian trong dữ liệu EEG. Khi kết hợp hai mô hình này, hệ thống có thể:

- Tận dụng tối đa thông tin từ cả đặc trưng không gian và thời gian, giúp mô hình học tốt hơn sự biến đổi phức tạp của tín hiệu EEG.
- Cải thiện khả năng tổng quát hóa, nhờ CNN giúp giảm nhiễu và tập trung vào các đặc trưng quan trọng, trong khi BiLSTM giúp hiểu rõ mối quan hệ thời gian trong dữ liệu EEG.

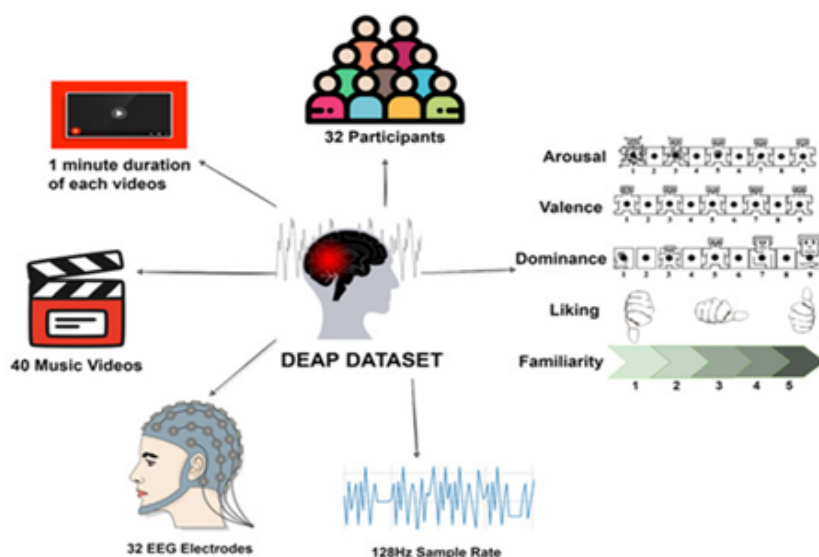
- Nâng cao độ chính xác trong phân loại cảm xúc EEG, do tận dụng được ưu điểm của cả hai mô hình.

## 2.4 Tiến hành thực nghiệm và đánh giá kết quả

### 2.4.1 Thiết lập thí nghiệm

#### 2.4.1.1 Lựa chọn bộ dữ liệu chuẩn

Để có cơ sở so sánh, đánh giá hiệu quả của các kiến trúc học sâu và phương pháp trích xuất đặc trưng trong bài toán EEG\_ER, các nghiên cứu cần phải được tiến hành trên các bộ dữ liệu chuẩn. Ba bộ dữ liệu phổ biến và có độ tin cậy thường được sử dụng là DEAP, AMIGOS, và DREAMER. Trong đó, DEAP là bộ dữ liệu được sử dụng nhiều nhất.

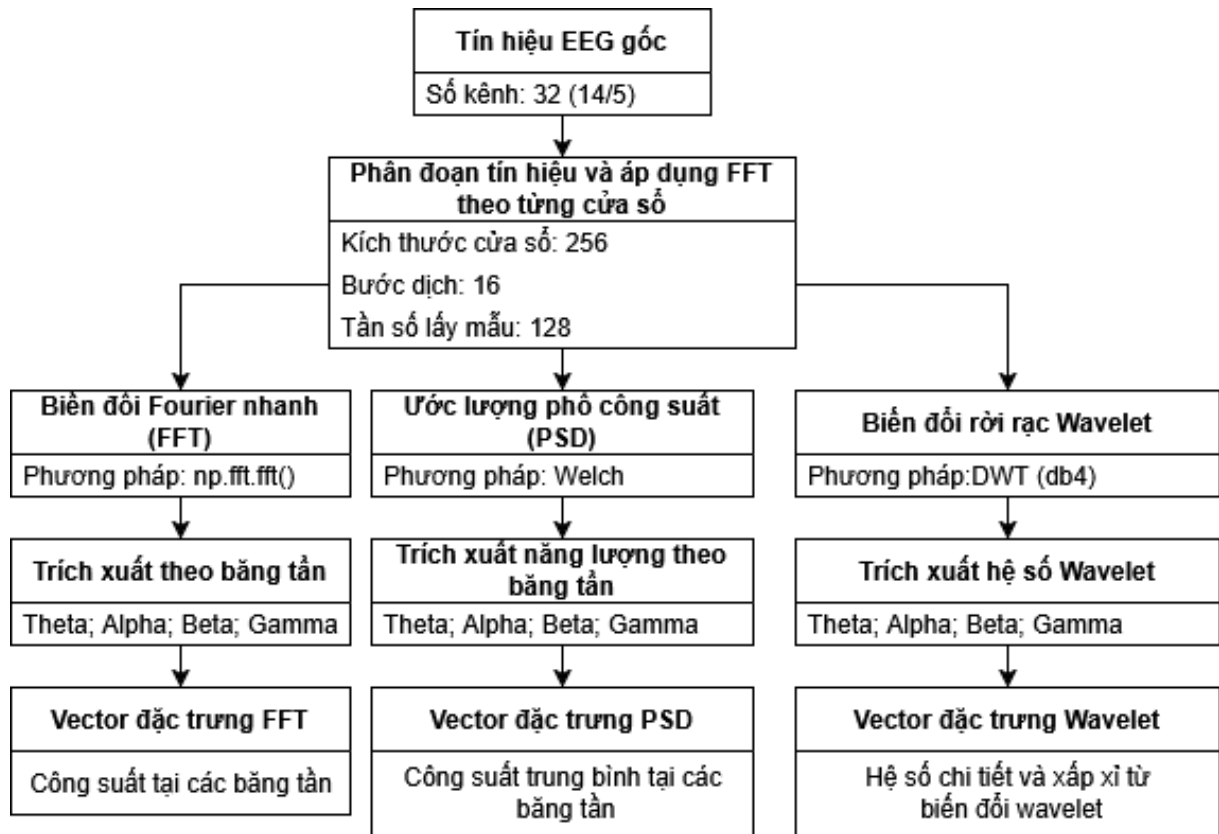


**Hình 2.4** Bộ dữ liệu DEAP [60]

Do vậy, bộ dữ liệu được NCS lựa chọn khảo sát cũng là bộ dữ liệu DEAP [60]. Chi tiết về bộ dữ liệu này đã được mô tả trong phần 1.3.3. Ở đây, NCS sử dụng phiên bản đã được tiền xử lý trước các dữ liệu thô được cung cấp bởi các tác giả của bộ dữ liệu DEAP.

#### 2.4.1.2 Cài đặt các thuật toán trích chọn đặc trưng

Ba thuật toán phổ biến được sử dụng để trích xuất đặc trưng tần số từ tín hiệu EEG gồm: FFT, Welch, và Biến đổi Wavelet rời rạc (DWT). Các thuật toán này giúp chuyển tín hiệu EEG từ miền thời gian sang miền tần số hoặc thời gian–tần số, từ đó làm nổi bật các đặc trưng dao động não bộ liên quan đến cảm xúc. Các bước thực hiện được mô tả ở Hình 2.5. Trong phạm vi luận án, NCS thực nghiệm với cả 3 thuật toán, mục đích tìm ra thuật toán phù hợp nhất với bài toán EEG\_ER.



**Hình 2.5** Các bước thực hiện trích chọn đặc trưng

#### 2.4.1.3 Thiết lập kiến trúc mạng

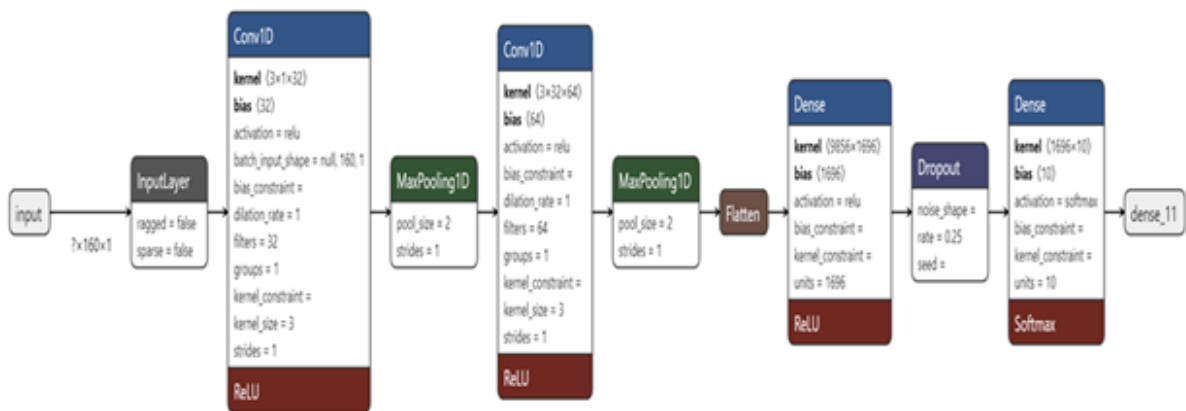
Hiệu quả hoạt động của một mô hình học sâu phụ thuộc vào đồng thời ba yếu tố: kiến trúc mạng, loại dữ liệu đầu vào, các thành phần trong vec tơ đặc trưng. Chính vì vậy, chúng ta cần phải lựa chọn kiến trúc mạng phù hợp với bộ dữ liệu EEG được sử dụng và dạng đặc trưng đầu vào đưa tới mô hình học sâu như đặc trưng thời gian, đặc trưng tần số hoặc là sự kết hợp của cả thời gian và tần số.

Bộ dữ liệu DEAP được xem là tập dữ liệu đa chiều, trong đó mỗi tín hiệu EEG đại diện cho một chiều dữ liệu (được ghi từ 32 điện cực), phản ánh hoạt động của các vùng não khác nhau. Đặc điểm này mang lại giá trị lớn trong việc phân tích không gian, khiến mạng CNN trở thành một lựa chọn phù hợp để khai thác các đặc trưng không gian từ dữ liệu EEG.

Ngoài ra, tín hiệu EEG còn bao gồm năm dải tần cơ bản: Alpha, Beta, Gamma, Delta và Theta, vốn có liên hệ mật thiết với cảm xúc của con người. Việc chuyển đổi tín hiệu từ miền thời gian sang miền tần số giúp làm nổi bật các đặc trưng theo tần số. Khi chia tín hiệu EEG thành các đoạn nhỏ, mỗi đoạn có thể phản ánh mối liên hệ thời gian riêng biệt. Mạng BiLSTM có thể học được mối quan hệ tuần tự này, đồng thời linh hoạt trong việc thích ứng với sự thay đổi theo thời gian. Đặc biệt, cấu trúc "hai chiều" của BiLSTM cho phép mạng học được cả thông tin từ quá khứ lẫn tương lai, nâng cao hiệu quả trong các bài toán nhận dạng tín hiệu EEG.

Sau tiền xử lý và trích chọn (FFT/PSD), mỗi cửa sổ huấn luyện được tổ chức thành tensor  $X \in \mathbb{R}^{(B,T,C)}$ . Các đặc trưng từ 32 (14, 5) điện cực và 5 dải tần được xếp nối theo trục thời gian để tạo thành một chuỗi duy nhất với  $T = 32 \times 5 = 160$  và  $C = 1$  (input\_shape = (160, 1)). Cách sắp xếp này cho phép lớp Conv1D đầu tiên tích chập dọc trục  $T$  trên một kênh duy nhất, học các mẫu biến thiên theo thời gian của tín hiệu EEG đã được gộp; đồng thời giảm số tham số so với việc giữ nhiều kênh song song, thuận lợi cho triển khai trên thiết bị hạn chế tài nguyên. Đối với BiLSTM, đầu vào vẫn là chuỗi độ dài  $T$  với véc-tơ đặc trưng bậc một ( $C = 1$ ) tại mỗi bước; trong kiến trúc CNN và BiLSTM, Conv1D trích xuất bản đồ đặc trưng theo thời gian rồi BiLSTM xử lý chuỗi này, hình dạng đầu vào tuân theo  $(T, 1)$  (một phương án thay thế là giữ cấu trúc đa kênh  $((T, C) = (5, 32))$  khi coi 5 dải là trục thời gian và 32 điện cực là kênh). Trong các thực nghiệm của luận án, phương án xếp nối thành  $(160, 1)$  được ưu tiên vì gọn nhẹ và phù hợp với mục tiêu triển khai.

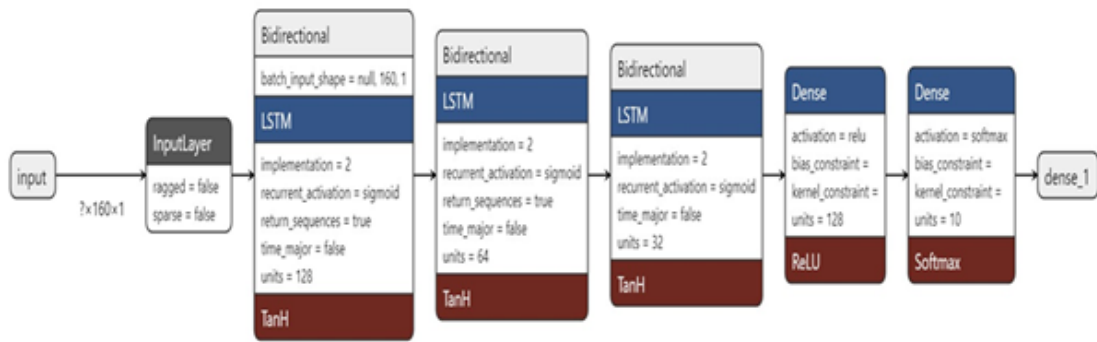
Từ các trên các phân tích trên, nghiên cứu sinh tiến hành khảo sát khả năng học đặc trưng của các kiến trúc học sâu như CNN, BiLSTM và mô hình kết hợp CNN–BiLSTM. Các mô hình này được huấn luyện và đánh giá thực nghiệm với nhiều kỹ thuật tối ưu như tinh chỉnh siêu tham số (hyperparameter tuning), trong môi trường Colab Pro của Google.



**Hình 2.6** Kiến trúc mạng CNN

Hình 2.6 trình bày chi tiết kiến trúc của mạng CNN cùng với các siêu tham số được sử dụng. Cụ thể, kiến trúc này bao gồm hai lớp Conv1D đầu tiên dùng để học các đặc trưng từ tín hiệu EEG đầu vào. Giữa hai lớp này, hai lớp MaxPooling1D được chèn vào nhằm giảm kích thước đầu ra, hạn chế hiện tượng overfitting và đồng thời tăng tốc độ xử lý. Sau đó, tín hiệu đi qua một lớp Dense gồm 1696 đơn vị kết hợp với hàm kích hoạt ReLU. Tiếp theo là một lớp Dropout (tỷ lệ 0.2) nhằm loại bỏ ngẫu nhiên một số kết nối để tránh học quá khớp. Phần cuối của mạng là một lớp Fully Connected gồm 10 đầu ra, tương ứng với thang điểm đánh giá cảm xúc, sử dụng hàm Softmax để tính xác suất dự

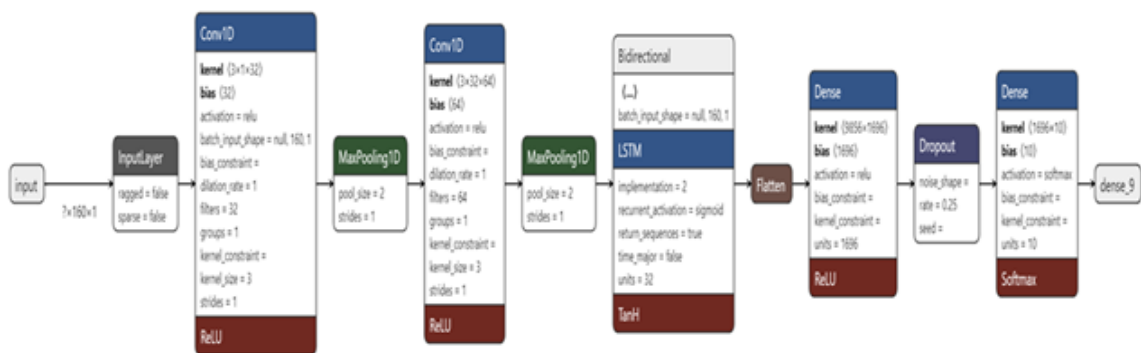
đoán cho từng lớp.



**Hình 2.7** Kiến trúc mạng Bi-LSTM

Do tín hiệu EEG có tính chất phức tạp và thường ẩn chứa các mối liên hệ theo thời gian giữa các đoạn dữ liệu, nên nghiên cứu sinh lựa chọn thêm mạng BiLSTM nhằm khai thác đặc trưng theo chuỗi. Hình 2.7 trình bày kiến trúc mạng BiLSTM được đề xuất, trong đó bao gồm một lớp BiLSTM và hai lớp Dense. Chi tiết cấu trúc của mạng này sẽ được trình bày sau khi thực hiện bước tinh chỉnh siêu tham số.

Tương tự, kiến trúc mạng kết hợp giữa CNN và BiLSTM cùng các tham số được mô tả chi tiết trong Hình 2.8. Trong mạng này, lớp tích chập đầu tiên có nhiệm vụ học các đặc trưng cơ bản của dữ liệu, lớp tích chập thứ hai sẽ học những đặc trưng phức tạp hơn, 2 lớp MaxPooling vẫn nhằm giảm không gian và ngăn ngừa quá khớp. Lớp BiLSTM được sử dụng để xử lý thông tin theo dạng chuỗi thời gian nhằm nắm bắt được mối liên hệ dài hơn trong dữ liệu.



**Hình 2.8** Kiến trúc mạng kết hợp CNN-BiLSTM

#### 2.4.2 Thực nghiệm và đánh giá kết quả

**Bảng 2.2** Đánh giá hiệu quả hoạt động của các kiến trúc mạng

| Đặc trưng đầu vào | Kiến trúc mạng | Độ chính xác trung bình (%) | Độ tổn thất trung bình |
|-------------------|----------------|-----------------------------|------------------------|
| FFT               | CNN            | <b>95.17</b>                | <b>0.1886</b>          |
| FFT               | BiLSTM         | 83.65                       | 0.2257                 |
| FFT               | CNN+BiLSTM     | 93.37                       | 0.2419                 |
| Welch             | CNN            | 94.54                       | 0.1790                 |
| Welch             | BiLSTM         | 83.75                       | 0.2495                 |
| Welch             | CNN+BiLSTM     | 93.81                       | 0.2265                 |
| Wavelet           | CNN            | 84.47                       | 0.4851                 |
| Wavelet           | BiLSTM         | 81.86                       | 0.5409                 |
| Wavelet           | CNN+BiLSTM     | 80.82                       | 0.5347                 |

Bằng việc tiến hành các thực nghiệm trên môi trường Google Colab Pro, sử dụng GPU NVIDIA T4 (16GB), cho phép tăng tốc quá trình huấn luyện và thử nghiệm mô hình. Các mô hình được xây dựng bằng thư viện TensorFlow/Keras, với các siêu tham số được điều chỉnh thủ công dựa trên hiệu suất quan sát được sau mỗi lần huấn luyện.

Dữ liệu EEG sau khi trích đặc trưng được chia thành hai tập với tỷ lệ 75% cho huấn luyện và 25% cho kiểm thử, áp dụng cho từng đối tượng riêng biệt nhằm đảm bảo sự đồng đều trong phân bố nhãn. Ngoài ra, để tăng tính khách quan và kiểm định khả năng tổng quát của mô hình, kỹ thuật chia dữ liệu theo K-Fold chéo ( $K = 5$ ) cũng được áp dụng. Kết quả được tổng hợp từ trung bình các fold, giúp đánh giá độ ổn định của mô hình trong nhiều kịch bản phân chia dữ liệu khác nhau.

Từ các số liệu chỉ ra trong Bảng 2.2 cho thấy mô hình CNN 1D đạt cao nhất với FFT (95.17%, loss 0.1886) và Welch (94.54%, loss 0.1790), vượt trội so với BiLSTM và CNN+BiLSTM trên cùng đặc trưng. Lý do chính là *inductive bias* cục bộ–đa thang của tích chập 1D phù hợp với cấu trúc phổ tần trong FFT/PSD, trong khi phụ thuộc chuỗi còn lại sau tổng hợp theo cửa sổ không đủ lớn để bù cho chi phí/độ trễ của BiLSTM; do đó thêm BiLSTM không mang lại cải thiện tương xứng. Biểu diễn Wavelet tạo không gian đặc trưng có chiều cao và dư thừa lớn trong thiết lập nhãn mức–trial, dẫn đến hiệu năng thấp hơn một cách nhất quán ở cả ba kiến trúc.

**Bảng 2.3** So sánh hiệu quả nhận dạng của mạng CNN đề xuất với một số mạng CNN tiêu biểu

| Mô hình                 | Độ chính xác trung bình (%) | Độ tổn thất trung bình | Số lượng tham số (Parameters) |
|-------------------------|-----------------------------|------------------------|-------------------------------|
| Alexnet                 | 90.95                       | 0.3211                 | 22.178.186                    |
| Resnet                  | 92.25                       | 0.3088                 | 4.323.394                     |
| GoogleNet               | 90.21                       | 0.3689                 | 11.647.042                    |
| VGG16                   | 92.39                       | 0.3484                 | 32.219.466                    |
| MobileNet               | 22.34                       | 0.2080                 | 3.216.714                     |
| DenseNet                | 91.09                       | 0.3463                 | 3.727.562                     |
| <b>Mạng CNN đề xuất</b> | <b>95.17</b>                | <b>0.1886</b>          | <b>16.740.778</b>             |

NCS cũng thực hiện đánh giá hiệu quả hoạt động của mạng CNN đề xuất so với một số mạng CNN tiêu biểu trong tác vụ EEG\_ER từ dạng dữ liệu DEAP [49, 73]. So với các CNN 2D tiêu biểu cho ảnh (AlexNet, VGG16, ResNet, GoogleNet, DenseNet), mạng CNN 1D đề xuất đạt 95.17% với 16.74M tham số—cao hơn ResNet/VGG16 dù số tham số ở mức trung bình. Sự khác biệt đến từ việc mạng đề xuất điều chỉnh kernel 1D, độ sâu và đầu phân loại theo cấu trúc FFT/PSD của EEG; trong khi các kiến trúc 2D tối ưu cho biên/cấu trúc không gian ảnh không phù hợp với chuỗi 1D phổ nên hiệu năng thấp hơn. Trường hợp MobileNet cho thấy rõ hiệu ứng *mismatch* giữa depthwise–separable 2D và đặc trưng 1D, dẫn tới độ chính xác rất thấp.

## 2.5 Kết luận chương 2

Chương 2 đã phân tích chi tiết một số kỹ thuật quan trọng cần phải được thực hiện khi xây dựng một mô hình EEG\_ER. Từ việc biểu diễn tín hiệu EEG, trích xuất đặc trưng đến thử nghiệm các mô hình học sâu phổ biến. Các nội dung chính có thể được tóm lược như sau:

- **Phương pháp biểu diễn tín hiệu EEG:** Đã phân tích và đánh giá ba phương pháp biểu diễn dữ liệu EEG, bao gồm miền thời gian, miền tần số và miền thời gian - tần số. Kết quả cho thấy biểu diễn miền tần số cung cấp thông tin quan trọng hơn cho quá trình học sâu, do tín hiệu EEG chứa nhiều đặc trưng hữu ích trong các dải tần khác nhau.
- **Trích xuất đặc trưng EEG:** Các phương pháp Biến đổi Fourier nhanh (FFT), Mật độ phổ công suất (PSD) và Biến đổi Wavelet (WT) đã được thử nghiệm và so sánh. Kết quả thực nghiệm cho thấy FFT là phương pháp phù hợp nhất cho bài toán nhận diện cảm xúc EEG do tốc độ xử lý nhanh và khả năng biểu diễn đặc trưng hiệu quả



trên các thiết bị tài nguyên hạn chế.

- **Đánh giá các mô hình học sâu:** Các thử nghiệm trên mô hình 1D-CNN, BiLSTM và CNN+BiLSTM cùng với các đầu vào đặc trưng khác nhau đã được thực hiện để xác định sự kết hợp giữa kiến trúc mạng và đầu vào đặc trưng phù hợp nhất cho bài toán EEG\_ER. Kết quả cho thấy 1D-CNN với đặc trưng FFT đạt độ chính xác cao nhất, trong khi mô hình kết hợp CNN-BiLSTM không mang lại lợi thế đáng kể về hiệu suất so với CNN đơn thuần.

Kết quả của chương này đã xác định được phương pháp tiền xử lý, biểu diễn tín hiệu, trích xuất đặc trưng và mô hình học sâu nền tảng phù hợp nhất cho bài toán nhận diện cảm xúc từ EEG. Các kết quả này cũng được NCS công bố trong các công trình CB [2], CB [3], CB [4]. Đây sẽ là nền tảng quan trọng cho Chương 3, nơi nghiên cứu tập trung vào việc tối ưu hóa kiến trúc nhằm tìm ra mô hình học sâu nhẹ, giảm thiểu chi phí tính toán và nâng cao khả năng triển khai trên thiết bị tài nguyên hạn chế.

# CHƯƠNG 3

## PHÁT TRIỂN MÔ HÌNH HỌC SÂU NHẸ CHO HỆ THỐNG EEG\_ER

EEG\_ER là một lĩnh vực nghiên cứu quan trọng với nhiều ứng dụng trong giao diện não – máy (BCI), y tế và công nghệ cảm xúc. Các mô hình học sâu hiện đại, đặc biệt là CNN, đã chứng minh được khả năng trích xuất đặc trưng mạnh mẽ từ dữ liệu EEG. Tuy nhiên, hầu hết các nghiên cứu trước đây chủ yếu tập trung vào các mô hình có độ phức tạp cao, yêu cầu tài nguyên tính toán lớn, và ít chú trọng đến khả năng triển khai thực tế trên các thiết bị có tài nguyên hạn chế.

Do đó, mục tiêu của chương này là thiết kế và tối ưu hóa mô hình học sâu cho EEG\_ER nhằm đáp ứng các yêu cầu về hiệu suất cao, tài nguyên hạn chế, và khả năng triển khai thực tế trên các thiết bị biên của mạng AIoT. Cụ thể, phần đầu chương sẽ phân tích các thách thức khi triển khai mô hình EEG\_ER trên thiết bị tài nguyên hạn chế, bao gồm giới hạn tính toán, số lượng điện cực EEG ít, và khả năng tổng quát hóa của mô hình. Sau đó, nội dung chương tập trung sâu vào phân tích các kỹ thuật tối ưu hóa mô hình học sâu bao gồm kỹ thuật giảm *overfitting* và cải thiện tổng quát hóa, kỹ thuật cải thiện khả năng học sâu và hội tụ nhanh, kỹ thuật nâng cao khả năng học đặc trưng trong EEG. Ở phần kế tiếp, luận án đề xuất hai mô hình kiến trúc EEG\_MCISNet (cho bài toán phân lớp nhị phân) và EEG\_SICNet (cho bài toán phân lớp đa lớp). Hai kiến trúc này được thiết kế dựa trên kiến trúc nền tảng 1D-CNN (đã xác định ở chương 2) kết hợp thêm các khối Squeeze-and-Excitation (SE) và Inception Module nhằm tối ưu hóa việc học đặc trưng không gian – thời gian từ tín hiệu EEG. Phần cuối chương sẽ mô tả kết quả đánh giá hiệu suất thực nghiệm của hai mô hình trên các bộ dữ liệu tiêu chuẩn (DEAP, DREAMER, AMIGOS).

Kết quả đạt được từ chương này sẽ góp phần thu hẹp khoảng cách giữa mô hình học sâu EEG\_ER hiệu suất cao và các yêu cầu thực tế khi triển khai trên thiết bị có tài nguyên hạn chế, đồng thời cung cấp cơ sở khoa học cho các nghiên cứu tiếp theo trong lĩnh vực EEG\_ER.

### 3.1 Phân tích yêu cầu khi triển khai bài toán EEG\_ER trên thiết bị có tài nguyên hạn chế

Việc triển khai bài toán EEG\_ER trên các thiết bị có tài nguyên hạn chế đặt ra những yêu cầu nghiêm ngặt về khả năng tính toán, bộ nhớ, năng lượng tiêu thụ và khả

năng tổng quát hóa. Phần này sẽ làm rõ các thách thức chính khi thiết kế một mô hình học sâu có thể hoạt động hiệu quả trên thiết bị tài nguyên hạn chế, thông qua việc phân tích hai yếu tố quan trọng:

- Giới hạn về tài nguyên tính toán trên phần cứng thực hiện tác vụ nhận dạng, bao gồm giới hạn về CPU, bộ nhớ và tốc độ suy luận.
- Hạn chế về số lượng điện cực EEG của thiết bị thu thập, ảnh hưởng đến khả năng trích xuất đặc trưng và độ chính xác của mô hình.

Các phân tích này giúp xác định những yếu tố cần tối ưu hóa trong kiến trúc mô hình EEG\_ER, làm cơ sở cho các phương pháp cải tiến được đề xuất trong phần tiếp theo.

### ***3.1.1 Giới hạn về tài nguyên tính toán trên phần cứng thực hiện tác vụ nhận dạng***

Việc triển khai mô hình học sâu phần cứng thực hiện tác vụ nhận dạng đặt ra nhiều thách thức liên quan đến bộ nhớ, hiệu suất tính toán, tốc độ suy luận và mức tiêu thụ năng lượng. Mặc dù các mô hình như 2D-CNN, RNN, và Transformer đã đạt hiệu suất cao trong các ứng dụng phân tích tín hiệu EEG, nhưng việc triển khai các mô hình này trên nền tảng có tài nguyên hạn chế vẫn gặp phải nhiều trở ngại [67, 58, 29].

Bộ nhớ hệ thống là một trong những yếu tố quan trọng ảnh hưởng đến khả năng triển khai mô hình trên thiết bị di động. Raspberry Pi 4 – một trong những nền tảng phổ biến – chỉ có các phiên bản RAM từ 2GB đến 8GB, trong khi các mô hình học sâu hiện đại có thể yêu cầu từ 16GB đến 32GB RAM để hoạt động hiệu quả [23, 83]. Ví dụ, một Transformer tiêu chuẩn có thể tiêu thụ hơn 16GB bộ nhớ trong quá trình suy luận, điều này khiến việc triển khai trên các thiết bị bị những rào cản lớn nếu không có các biện pháp tối ưu hóa. Bên cạnh đó, băng thông bộ nhớ và khả năng truy xuất dữ liệu cũng bị giới hạn. Raspberry Pi 4 chỉ hỗ trợ LPDDR4-3200 SDRAM, với tốc độ thấp hơn đáng kể so với các hệ thống có GPU mạnh được thiết kế để tối ưu hóa tính toán song song. Điều này làm ảnh hưởng đến tốc độ xử lý của mô hình và gây ra hiện tượng nghẽn băng thông, đặc biệt khi thực hiện suy luận trên dữ liệu EEG có độ phân giải cao [103].

Các thiết bị tính toán tại các nút biên trong hệ thống AIoT thường phải di động được. Vì vậy, chúng không được trang bị GPU mạnh như trên các hệ thống máy tính cao cấp. Do đó, hiệu suất tính toán phụ thuộc vào CPU, vốn có hiệu suất tính toán thấp hơn nhiều. Ví dụ, Raspberry Pi 4 sử dụng bộ vi xử lý Broadcom BCM2711, Quad-core Cortex-A72 (ARM v8) @ 1.5GHz có hiệu suất tính toán nhỏ hơn nhiều so với một GPU NVIDIA RTX 3090 (có thể đạt hiệu suất lên đến 35,600 GFLOPS) [103]. Sự chênh lệch lớn về hiệu suất tính toán này làm cho các mô hình phức tạp như ResNet, Transformer

hoặc LSTM không thể chạy hiệu quả trên các thiết bị có tài nguyên hạn chế nếu không có các kỹ thuật tối ưu hóa đặc biệt. Chẳng hạn, một nghiên cứu chỉ ra rằng 2D-CNN tiêu chuẩn có thể yêu cầu hàng trăm GFLOPS để thực hiện suy luận trong thời gian thực, trong khi Raspberry Pi chỉ có thể cung cấp hiệu suất tính toán 13 GFLOPS, thấp hơn ít nhất 10 lần so với yêu cầu [23].

Tiêu thụ năng lượng là một yếu tố quan trọng khi triển khai mô hình trên thiết bị có khả năng di động. Do các hệ thống nhúng thường hoạt động bằng pin, việc triển khai các mô hình học sâu có thể dẫn đến mức tiêu thụ năng lượng cao, làm giảm tuổi thọ pin và ảnh hưởng đến khả năng sử dụng liên tục của thiết bị. Một số nghiên cứu cho thấy các mô hình CNN truyền thống có thể tiêu thụ gấp 5-10 lần năng lượng so với các mô hình được tối ưu hóa cho thiết bị có khả năng di động. Điều này đòi hỏi các mô hình triển khai trên thiết bị tài nguyên hạn chế phải có mức tiêu thụ năng lượng thấp, giảm tải tính toán không cần thiết để đảm bảo hiệu suất cao nhưng vẫn duy trì thời gian hoạt động lâu dài [23].

Một yếu tố quan trọng khác cần xem xét là tốc độ suy luận khi triển khai mô hình trên thiết bị di động. Trong các ứng dụng EEG\_ER thực tế, mô hình phải đưa ra kết quả trong thời gian ngắn để đảm bảo phản hồi thời gian thực. Tuy nhiên, trên các hệ thống nhúng, thời gian suy luận có thể bị kéo dài đáng kể do hiệu suất phần cứng yếu và băng thông bộ nhớ thấp. Chẳng hạn, một nghiên cứu chỉ ra rằng Transformer tiêu chuẩn có thể mất 200ms để xử lý một mẫu EEG, trong khi các ứng dụng thực tế đối với các ứng dụng nhận diện cảm xúc là dưới 100ms. Điều này cho thấy các mô hình cần phải tối ưu hóa để giảm thời gian suy luận trên thiết bị tài nguyên hạn chế [67].

Những hạn chế về bộ nhớ, hiệu suất tính toán, tiêu thụ năng lượng và tốc độ suy luận đặt ra thách thức lớn trong việc phát triển các mô hình nhẹ, hiệu quả và có khả năng hoạt động trên thiết bị tài nguyên hạn chế.

### ***3.1.2 Hạn chế về số lượng điện cực EEG trên thiết bị thu thập tín hiệu EEG có khả năng di động được***

Việc thu thập tín hiệu EEG với số lượng điện cực lớn chủ yếu được thực hiện trong môi trường phòng thí nghiệm, nơi có điều kiện kiểm soát nghiêm ngặt nhằm đảm bảo chất lượng tín hiệu tối ưu. Các thiết bị EEG sử dụng trong nghiên cứu, chẳng hạn như BioSemi, NeuroScan... hoặc các thiết bị tiêu chuẩn để thu các tập dữ liệu như DEAP, AMIGOS, DREAMER, thường bao gồm 32 đến 64 điện cực, giúp ghi nhận tín hiệu từ nhiều vùng não khác nhau, đảm bảo khả năng phân tích chính xác mối quan hệ giữa các hoạt động thần kinh và trạng thái cảm xúc [72, 127]. Tuy nhiên, khi triển khai EEG\_ER trong các ứng dụng thực tế, chẳng hạn như trên các thiết bị đeo, hệ thống theo dõi cảm xúc di động hoặc ứng dụng y tế cá nhân, số lượng điện cực EEG thường bị giới hạn

xuống còn 14, 5 hoặc 2 điện cực do các yếu tố sau:

- **Yêu cầu về tính di động và tiện lợi khi sử dụng:** Các thiết bị EEG thương mại như Emotiv Epoc (14 điện cực), Muse S (5 điện cực) hoặc NeuroSky MindWave (2 điện cực) được thiết kế để tối ưu hoá trải nghiệm người dùng, giúp dễ dàng triển khai mà không cần kỹ thuật viên hỗ trợ. Trong khi đó, các thiết bị EEG sử dụng trong nghiên cứu yêu cầu một quy trình thiết lập phức tạp, bao gồm lắp đặt nhiều điện cực và sử dụng gel dẫn điện để tăng chất lượng tiếp xúc giữa điện cực và da đầu [52].
- **Hạn chế về phần cứng và chi phí sản xuất:** Việc trang bị số lượng điện cực lớn không chỉ làm tăng kích thước và khối lượng của thiết bị, mà còn đòi hỏi vi xử lý mạnh mẽ hơn để xử lý tín hiệu đồng thời từ nhiều điện cực. Điều này dẫn đến chi phí sản xuất cao hơn, không phù hợp với các thiết bị EEG thương mại hướng đến thị trường phổ thông [127, 5].
- **Tác động của môi trường đo lường thực tế:** Khi triển khai EEG\_ER trong các môi trường thực tế như văn phòng, phương tiện giao thông hoặc không gian mở, tín hiệu EEG có thể bị ảnh hưởng đáng kể bởi chuyển động của người dùng, nhiễu điện từ từ các thiết bị điện tử xung quanh và điều kiện tiếp xúc không ổn định giữa điện cực và da đầu.

Trong phòng thí nghiệm, những yếu tố này được kiểm soát nghiêm ngặt để đảm bảo độ chính xác cao, nhưng trong môi trường thực tế, nhiều có thể làm suy giảm đáng kể chất lượng tín hiệu thu nhận, đặc biệt là khi số lượng điện cực EEG bị giới hạn. Sự khác biệt giữa thiết bị thu thập EEG trong phòng thí nghiệm và thiết bị EEG có khả năng di động đặt ra thách thức lớn trong việc thiết kế các mô hình phù hợp với điều kiện triển khai thực tế. Một số nghiên cứu đã chỉ ra rằng, khi sử dụng thiết bị EEG di động có số lượng điện cực hạn chế, hiệu suất của mô hình nhận diện cảm xúc giảm đáng kể [53, 13].

Một trong những nguyên nhân chính dẫn đến suy giảm hiệu suất là tín hiệu EEG có số lượng điện cực thấp dễ bị ảnh hưởng bởi nhiễu môi trường và chuyển động của người dùng. Trong phòng thí nghiệm, các thiết bị EEG thường được cố định cẩn thận trên da đầu với gel dẫn điện, giúp giảm trở kháng và tăng độ chính xác của tín hiệu. Ngược lại, các thiết bị di động, điện cực khô hoặc không xâm lấn có thể dẫn đến tín hiệu không ổn định, bị ảnh hưởng bởi cử động cơ bắp (EMG) và nhiễu điện từ (EMI) từ các thiết bị điện tử khác trong môi trường xung quanh.

Bên cạnh ảnh hưởng đến độ chính xác, việc thu thập tín hiệu EEG với số lượng điện cực thấp còn làm tăng độ trễ suy luận. Một giả định phổ biến là số điện cực ít hơn sẽ giúp mô hình xử lý nhanh hơn, nhưng trên thực tế, điều này không hoàn toàn đúng.

Khi số điện cực ít, mô hình cần nhiều bước tiền xử lý hơn để loại bỏ nhiễu, do đó có thể làm tăng thời gian suy luận thay vì giảm.

Trong triển khai thực tế, mô hình được huấn luyện ngoại tuyến trên hạ tầng trung tâm, sau đó đóng gói và đưa xuống thiết bị biên để suy luận thời gian thực. Cách tổ chức này tách bạch chi phí huấn luyện khỏi các ràng buộc phần cứng tại thiết bị, đồng thời cho phép tối ưu mô hình theo các thước đo triển khai (kích thước tham số, MFLOPs, bộ nhớ, độ trễ).

Từ các phân tích trên, có thể thấy rằng việc triển khai mô hình EEG\_ER tại các nút biên trong hệ thống AIoT gặp phải hai thách thức lớn:

1. **Giới hạn tài nguyên tính toán:** Phần cứng thực hiện tác vụ nhận dạng có bộ nhớ thấp, hiệu suất CPU hạn chế và tiêu thụ năng lượng cao, dẫn đến khó khăn trong việc triển khai các mô hình học sâu phức tạp.
2. **Số lượng điện cực EEG trên thiết bị thu thập có khả năng di động được là hạn chế:** Các thiết bị EEG thương mại thường chỉ có từ 2 đến 14 điện cực, gây mất thông tin không gian giữa vùng não, ảnh hưởng đến hiệu suất của mô hình.

Những vấn đề này đòi hỏi một kiến trúc mô hình nhẹ, hiệu quả, có khả năng học tốt trên số lượng điện cực EEG thấp mà vẫn đảm bảo độ chính xác cao. Trong phần tiếp theo, luận án sẽ đề xuất các phương pháp giúp khắc phục hạn chế, bao gồm các chiến lược tối ưu hóa mô hình, kỹ thuật học tăng cường, và các phương pháp kết hợp dữ liệu EEG với các nguồn dữ liệu khác nhằm cải thiện hiệu suất phân loại.

### 3.2 Các kỹ thuật tối ưu hóa mô hình học sâu

Như đã phân tích trong phần trước đó, việc triển khai mô hình EEG\_ER trên thiết bị có tính di động được đặt ra nhiều thách thức do giới hạn tài nguyên tính toán, số lượng điện cực EEG thấp và khả năng tổng quát hóa của mô hình. Nếu không có các phương pháp tối ưu hóa phù hợp, mô hình sẽ gặp phải các vấn đề nghiêm trọng như hiệu suất suy giảm, thời gian suy luận kéo dài và tiêu tốn tài nguyên vượt mức cho phép, làm giảm tính khả thi khi triển khai trên các hệ thống thực tế.

Do đó, để đảm bảo mô hình có thể hoạt động hiệu quả trên các nền tảng tài nguyên hạn chế, việc tối ưu hóa kiến trúc mạng là một yêu cầu bắt buộc. Các kỹ thuật tối ưu hóa cần được thiết kế nhằm giảm thiểu độ phức tạp của mô hình, nâng cao khả năng học đặc trưng từ số lượng điện cực EEG hạn chế, và tối ưu hóa hiệu suất suy luận trên thiết bị có thể di động được. Mục tiêu chính của các kỹ thuật tối ưu hóa được đề xuất trong phần này là:

- Giảm *overfitting* và cải thiện khả năng tổng quát hóa, giúp mô hình duy trì hiệu suất cao ngay cả khi dữ liệu huấn luyện bị hạn chế.
- Cải thiện quá trình huấn luyện và suy luận trên thiết bị có khả năng di động, đảm bảo mô hình hội tụ nhanh, tiêu thụ ít tài nguyên mà vẫn duy trì độ chính xác.
- Tối ưu hóa việc học đặc trưng EEG, giúp mô hình tận dụng tối đa số lượng điện cực EEG hạn chế và giảm thiểu ảnh hưởng của nhiễu bằng cách sử dụng các cơ chế như *SE Block*, *Inception Module* và *Attention Mechanism*.

Các phương pháp tối ưu hóa này không chỉ giúp cải thiện hiệu suất nhận dạng cảm xúc, mà còn đảm bảo khả năng triển khai thực tế trên các thiết bị biên của mạng AIoT, đáp ứng yêu cầu về tốc độ suy luận nhanh và mức tiêu thụ tài nguyên tối thiểu.

### 3.2.1 Kỹ thuật giảm *overfitting* và cải thiện tổng quát hóa

Một trong những thách thức lớn khi triển khai mô hình EEG\_ER là *overfitting*, tức mô hình học quá mức vào dữ liệu huấn luyện, làm giảm khả năng tổng quát hóa khi áp dụng trên dữ liệu thực tế. Hiện tượng này xảy ra phổ biến do tập dữ liệu EEG thường có số lượng mẫu hạn chế, đặc biệt khi số lượng điện cực EEG thấp, dẫn đến nguy cơ mô hình chỉ ghi nhớ dữ liệu huấn luyện thay vì học các đặc trưng thực sự có ý nghĩa. Để khắc phục vấn đề này, nghiên cứu áp dụng một số kỹ thuật giảm *overfitting* và tăng cường khả năng tổng quát hóa sau đây.

#### 3.2.1.1 Regularization ( $L1$ , $L2$ )

Regularization là một phương pháp được sử dụng phổ biến để kiểm soát độ phức tạp của mô hình bằng cách thêm một điều khoản phạt vào hàm mất mát, nhằm hạn chế sự phát triển quá mức của trọng số [62]. Trong học sâu, hai dạng phổ biến nhất là  $L1$  Regularization (Lasso) và  $L2$  Regularization (Weight Decay).

$L1$  Regularization thực hiện phạt trọng số theo chuẩn  $L_1$ , tức là tổng giá trị tuyệt đối của các trọng số, giúp mô hình tự động loại bỏ các trọng số không quan trọng bằng cách đưa chúng về 0. Công thức:

$$L(W) = L_{\text{original}} + \lambda \sum_i |W_i| \quad (3.1)$$

Trong đó:  $L_{\text{original}}$  là hàm mất mát ban đầu;  $W$  là trọng số của mô hình;  $\lambda$  là hệ số điều chỉnh mức độ *regularization*;  $\sum_i |W_i|$  là tổng giá trị tuyệt đối của các trọng số, giúp mô hình loại bỏ trọng số nhỏ và tạo ra kiến trúc thưa hơn.

Phương pháp này có ưu điểm là giúp mô hình trở nên thưa hơn (*sparse*), tức là chỉ giữ lại các trọng số quan trọng, giúp loại bỏ những đặc trưng không cần thiết. Điều này rất hữu ích trong phân tích EEG, khi một số điện cực EEG có thể không đóng góp đáng

kể vào việc nhận dạng cảm xúc. Tuy nhiên, nhược điểm của phương pháp này là có thể làm mất đi một số thông tin quan trọng nếu hệ số regularization quá lớn, dẫn đến hiệu suất mô hình giảm.

L2 Regularization phạt trọng số theo chuẩn  $L_2$ , tức là tổng bình phương của các trọng số. Điều này giúp làm giảm giá trị của các trọng số lớn, khiến mô hình trở nên ổn định hơn mà không triệt tiêu hoàn toàn trọng số. Công thức:

$$L(W) = L_{\text{original}} + \lambda \sum_i W_i^2 \quad (3.2)$$

So với L1, phương pháp L2 giúp làm mượt trọng số và giảm biến động cao của mô hình, giúp mô hình ổn định hơn và tránh hiện tượng quá khớp khi làm việc với dữ liệu có nhiều cỡ nhỏ như EEG. Tuy nhiên, do không làm thưa trọng số như L1, phương pháp này không thể loại bỏ hoàn toàn các điện cực EEG kém quan trọng, điều này có thể không tăng được tối ưu trên các mô hình nhỏ mà không cải thiện đáng kể hiệu suất.

#### 3.2.1.2 Dropout

Dropout là một phương pháp regularization được giới thiệu bởi Srivastava [101] là một kỹ thuật đơn giản nhưng hiệu quả để giảm *overfitting* bằng cách vô hiệu hóa ngẫu nhiên một số nơ ron trong quá trình huấn luyện, giúp mô hình không phụ thuộc quá mức vào một số đặc trưng cụ thể. Cơ chế Dropout được mô tả như sau:

- Trong quá trình huấn luyện: Mỗi nơ ron (bao gồm cả nơ ron ẩn và nơ ron đầu vào) có xác suất  $p$  bị loại bỏ tại mỗi bước huấn luyện. Điều này tạo ra một mạng con ngẫu nhiên tại mỗi lần cập nhật, giúp mô hình tránh việc dựa quá mức vào bất kỳ đặc trưng nào.
- Trong quá trình kiểm tra (*testing*): Toàn bộ mạng được sử dụng mà không áp dụng Dropout. Tuy nhiên, để duy trì nhất quán với giai đoạn huấn luyện, các trọng số được nhân với hệ số  $1 - p$ . Điều này tương đương với việc lấy trung bình dự đoán của tất cả các mạng con đã được huấn luyện.

Dropout có ưu điểm lớn trong việc ngăn chặn mô hình học quá mức vào một số đặc trưng nhất định, đồng thời giúp mô hình trở nên linh hoạt hơn bằng cách tạo ra các phiên bản ngẫu nhiên của mô hình trong quá trình huấn luyện. Điều này đặc biệt quan trọng trong xử lý tín hiệu EEG, khi mà dữ liệu có sự biến động lớn giữa các cá nhân.

Tuy nhiên, việc áp dụng Dropout có thể làm chậm tốc độ huấn luyện, do mỗi *batch* có một kiến trúc mạng hơi khác nhau. Ngoài ra, khi sử dụng với *Batch Normalization*, Dropout có thể gây suy giảm hiệu suất do làm thay đổi phân phối dữ liệu đầu vào.

#### 3.2.1.3 Batch Normalization (BN)



Batch Normalization (BN) là một phương pháp giúp cải thiện sự ổn định trong quá trình huấn luyện bằng cách chuẩn hóa đầu ra của mỗi lớp ẩn, giúp giảm sự thay đổi phân phối dữ liệu giữa các batch [47]. Cách thức hoạt động của Batch Normalization:

- **Chuẩn hóa đầu vào:** Với mỗi mini-batch, tính toán giá trị trung bình ( $\mu$ ) và phương sai ( $\sigma^2$ ) của đầu vào.
- **Chuẩn hóa:** Sử dụng các giá trị này để chuẩn hóa đầu vào về phân phối với trung bình bằng 0 và phương sai bằng 1.
- **Tham số hóa:** Áp dụng các tham số học được ( $\gamma$  và  $\beta$ ) để duy trì khả năng biểu diễn của mô hình.

Công thức Batch Normalization được biểu diễn như sau:

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu^{(k)}}{\sqrt{(\sigma^{(k)})^2 + \epsilon}}, \quad y^{(k)} = \gamma^{(k)}\hat{x}^{(k)} + \beta^{(k)} \quad (3.3)$$

Trong đó:  $x^{(k)}$ : Đầu vào của lớp  $k$ ;  $\mu^{(k)}, (\sigma^{(k)})^2$ : Giá trị trung bình và phương sai của  $x^{(k)}$  trong mini-batch;  $\hat{x}^{(k)}$ : Đầu vào đã được chuẩn hóa;  $\gamma^{(k)}, \beta^{(k)}$ : Tham số học được để điều chỉnh phân phối của  $y^{(k)}$ ;  $\epsilon$ : Hằng số nhỏ để tránh chia cho 0.

Ưu điểm chính của Batch Normalization là giúp mạng học sâu hội tụ nhanh hơn và ổn định hơn, giảm sự phụ thuộc vào việc khởi tạo trọng số ban đầu. Điều này đặc biệt hữu ích với các mô hình như CNN hoặc GNN, vốn có nhiều lớp ẩn và thường gặp khó khăn trong việc tối ưu hóa. Tuy nhiên, BN có thể làm tăng chi phí tính toán do phải tính toán giá trị trung bình và phương sai trong mỗi batch, đồng thời không phải lúc nào cũng cải thiện hiệu suất, đặc biệt đối với các mô hình nông.

#### 3.2.1.4. Early Stopping

Early Stopping [115] là một phương pháp đơn giản nhưng hiệu quả để tránh *overfitting* bằng cách dừng quá trình huấn luyện sớm nếu hiệu suất trên tập kiểm tra không còn cải thiện. Mô hình theo dõi hàm mất mát trên tập kiểm tra sau mỗi epoch, nếu không có cải thiện sau một số epoch nhất định, quá trình huấn luyện sẽ dừng lại.

$$L_{val}^t \geq L_{val}^{t-k}, \quad \forall t \geq patience \quad (3.4)$$

Với  $L_{val}^t$  là hàm mất mát trên tập kiểm tra tại epoch thứ  $t$ , *patience* là số epoch cho phép trước khi dừng huấn luyện.

Early Stopping có ưu điểm là giúp giảm chi phí tính toán bằng cách dừng sớm khi mô hình không còn

### 3.2.2 Kỹ thuật cải thiện khả năng học sâu và hội tụ nhanh

Trong quá trình huấn luyện mô hình EEG\_ER trên các thiết bị tài nguyên hạn chế, nhiều thách thức có thể ảnh hưởng đến hiệu suất của mô hình, bao gồm tốc độ hội tụ chậm, tiêu thụ tài nguyên cao và khả năng tối ưu hóa kém đối với dữ liệu EEG có biến động lớn. Việc cải thiện hiệu suất của mô hình không chỉ liên quan đến thiết kế kiến trúc mạng mà còn cần đến các chiến lược tối ưu hóa huấn luyện, giúp mô hình học nhanh hơn, giảm tiêu hao tài nguyên và vẫn duy trì độ chính xác cao. Một số kỹ thuật quan trọng nhằm tối ưu hóa quá trình học sâu và tăng tốc hội tụ, bao gồm:

Tốc độ học ( $\eta$ ) là một trong những tham số quan trọng quyết định sự hội tụ của mô hình. Nếu  $\eta$  quá lớn, mô hình có thể dao động mạnh và không đạt đến điểm hội tụ. Ngược lại, nếu  $\eta$  quá nhỏ, quá trình huấn luyện diễn ra chậm, có nguy cơ mắc kẹt tại cực tiểu cục bộ. Để tối ưu hóa tốc độ học, nhiều chiến lược thích ứng đã được đề xuất, trong đó các phương pháp phổ biến bao gồm Learning Rate Decay, ReduceLROnPlateau, và đặc biệt là Cyclical Learning Rates (CLR) – một phương pháp tiên tiến giúp mô hình hội tụ hiệu quả hơn [98].

Trong phương pháp *Learning Rate Decay*, tốc độ học được giảm dần theo số lượng epoch, giúp mô hình học nhanh trong giai đoạn đầu và tinh chỉnh trong số ít các giai đoạn sau. Một phương pháp thường được sử dụng là *Exponential Decay*, trong đó tốc độ học giảm theo công thức:

$$\eta_t = \eta_0 \cdot e^{-kt} \quad (3.5)$$

Với  $\eta_0$  là tốc độ học ban đầu,  $k$  là hệ số giảm tốc độ học,  $t$  là số epoch. Việc giảm tốc độ học theo cấp số mũ giúp mô hình hội tụ mượt mà hơn và giảm nguy cơ dao động khi đến gần điểm hội tụ tối ưu.

Một chiến lược khác là *ReduceLROnPlateau*, trong đó tốc độ học chỉ giảm khi hiệu suất trên tập kiểm tra không cải thiện sau một số epoch nhất định. Phương pháp này giúp tối ưu hóa quá trình cập nhật trọng số bằng cách chỉ giảm tốc độ học khi cần thiết, thay vì giảm đều theo lịch trình cố định. Nếu hàm mất mát trên tập kiểm tra được cải thiện trong quá trình huấn luyện, tốc độ học  $\eta$  sẽ giữ nguyên ở mức hiện tại. Ngược lại sau một số lượng epoch nhất định (được xác định bởi tham số *patience*), tốc độ học sẽ giảm theo một hệ số  $\gamma$  (thường nằm trong khoảng 0.1 đến 0.5). Quá trình này tiếp tục lặp lại trong suốt quá trình huấn luyện để đảm bảo mô hình không sử dụng tốc độ học quá lớn khi đã gần hội tụ, giúp tăng độ ổn định trong quá trình tối ưu.

Phương pháp *ReduceLROnPlateau* thường được sử dụng trong các bài toán nhận diện cảm xúc từ EEG, giúp tự động điều chỉnh tốc độ học theo hiệu suất thực tế của mô hình.

Một phương pháp mới được đề xuất bởi Leslie N. Smith [98] là CLR. Thay vì giảm tốc độ học một cách liên tục, CLR điều chỉnh tốc độ học theo chu kỳ, dao động giữa một giá trị tối thiểu và tối đa trong suốt quá trình huấn luyện. Công thức của CLR được xác định như sau:

$$\eta_t = \eta_{\min} + (\eta_{\max} - \eta_{\min}) \cdot \frac{1 + \cos\left(\frac{\pi t}{T}\right)}{2} \quad (3.6)$$

Trong đó:  $\eta_{\min}$  và  $\eta_{\max}$  là biên dưới và biên trên của tốc độ học,  $T$  là số epoch trong mỗi chu kỳ.

CLR giúp mô hình tránh bẫy cực tiểu cục bộ bằng cách thay đổi tốc độ học tuần hoàn. Giảm số epoch cần thiết để hội tụ, giúp mô hình đạt hiệu suất cao hơn mà không cần điều chỉnh thủ công tốc độ học. Hơn nữa, có thể áp dụng trực tiếp vào các bộ tối ưu phổ biến như SGD, Adam, RMSprop.

Một trong những yếu tố quyết định đến hiệu suất huấn luyện của mô hình là thuật toán tối ưu hóa. Các thuật toán như *Momentum*, *RMSProp* và *Adam* giúp cải thiện quá trình cập nhật trọng số, đảm bảo hội tụ nhanh hơn so với phương pháp *Gradient Descent* truyền thống. Trong *Momentum*, thay vì cập nhật trọng số chỉ dựa trên gradient tại mỗi bước, một thành phần “động lực” ( $v_t$ ) được sử dụng để lưu trữ hướng di chuyển trước đó, giúp tránh dao động mạnh.

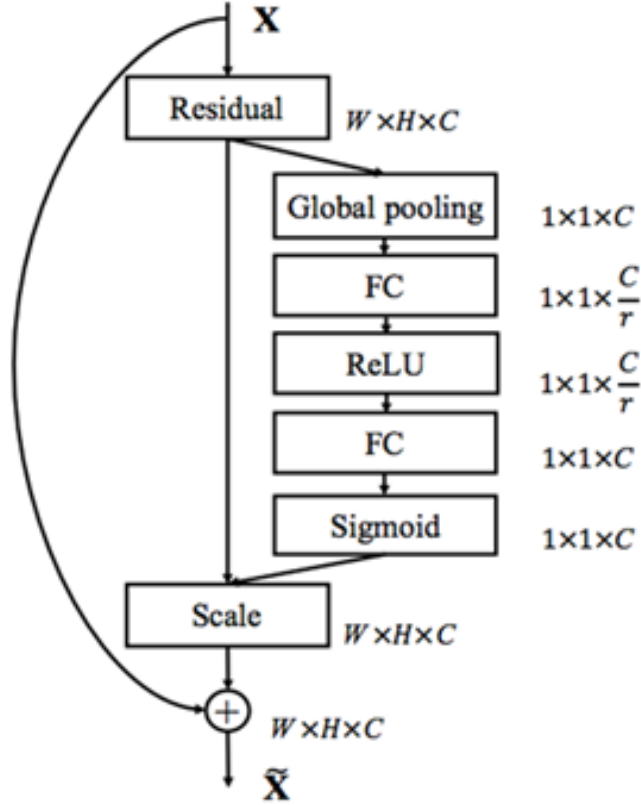
### 3.2.3 Nâng cao khả năng học đặc trưng trong EEG\_ER

Khi triển khai các mô hình học sâu, khả năng học đặc trưng (*Feature Learning*) đóng vai trò quan trọng trong việc trích xuất thông tin hữu ích từ dữ liệu đầu vào, đặc biệt đối với tín hiệu EEG – một loại tín hiệu có đặc tính phi tuyến, đa chiều và nhiễu cao. Nếu mô hình không được thiết kế để khai thác đúng đặc trưng quan trọng, hiệu suất phân loại có thể bị ảnh hưởng đáng kể. Do đó, việc phát triển các kỹ thuật học đặc trưng hiệu quả giúp cải thiện độ chính xác và khả năng tổng quát hóa của mô hình là một bước không thể thiếu trong nghiên cứu EEG.

Trong phần này, các phương pháp chính giúp nâng cao khả năng học đặc trưng từ tín hiệu EEG sẽ được phân tích, bao gồm cơ chế tập trung vào điện cực EEG quan trọng (*Channel Attention*), học đặc trưng không gian-thời gian (*Spatial-Temporal Feature Learning*), học đa tầng (*Multi-Scale Feature Learning*) và kết hợp nhiều nguồn đặc trưng (*Feature Fusion*). Những phương pháp này không chỉ giúp tối ưu hóa quá trình trích xuất đặc trưng, mà còn giúp mô hình học sâu có thể tận dụng tốt hơn các thông tin từ EEG để cải thiện hiệu suất phân loại.

#### 3.2.3.1 Tập trung vào các điện cực EEG quan trọng (*Channel Attention*)

Dữ liệu EEG có tính chất không đồng nhất giữa các điện cực, nghĩa là không phải tất cả các điện cực đều có mức độ quan trọng như nhau trong việc nhận diện cảm xúc, một số vị trí mang tín hiệu liên quan mạnh hơn, số khác dễ lẫn nhiễu. Các mô hình học sâu thông thường xử lý đồng đều mọi kênh, từ đó chưa khai thác hết các điện cực giàu thông tin và dễ bị suy giảm bởi nhiễu. Để khắc phục, luận án sử dụng *Squeeze-and-Excitation* (SE) [45] nhằm *học trọng số theo kênh* một cách dữ liệu-dẫn dắt.



**Hình 3.1** Cấu trúc khối SE [45].

Nguyên lý hoạt động của khối SE là: cho đặc trưng trung gian  $\mathbf{X} \in \mathbb{R}^{C \times L}$  (với  $C$  điện cực và  $L$  bước thời gian). Khối SE gồm ba giai đoạn:

(i) **Squeeze.** Tóm tắt thông tin theo trục thời gian của *mỗi điện cực* bằng Global Average Pooling (GAP):

$$z_c = \frac{1}{L} \sum_{t=1}^L X_{c,t}, \quad c = 1, \dots, C. \quad (3.7)$$

Vector  $\mathbf{z} = [z_1, \dots, z_C]^T \in \mathbb{R}^C$  phản ánh “mức hoạt hoá trung bình” của từng điện cực trong cửa sổ đang xét.

(ii) **Excitation.** Học *trọng số kênh* phi tuyến từ  $\mathbf{z}$  qua hai lớp FC với tỉ lệ giảm chiều  $r$  và kích hoạt ReLU-sigmoid:

$$\mathbf{s} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{z})), \quad \mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}, \mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}. \quad (3.8)$$

Kết quả  $\mathbf{s} \in (0, 1)^C$  là các hệ số nhân/giảm theo *điện cực* được suy ra từ dữ liệu.

(iii) **Scale.** Áp trọng số lên toàn bộ chuỗi theo kênh:

$$\hat{X}_{c,t} = s_c X_{c,t}, \quad \forall c, t. \quad (3.9)$$

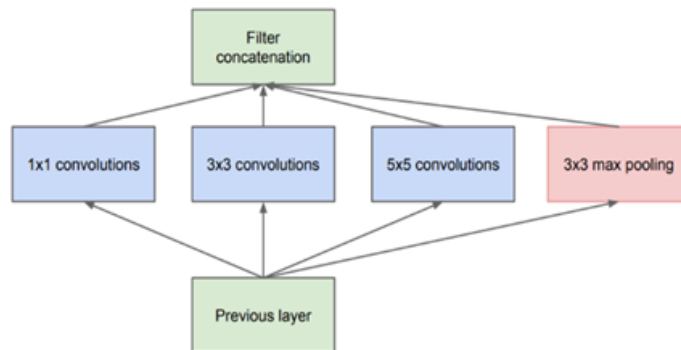
Thu được  $\hat{\mathbf{X}} \in \mathbb{R}^{C \times L}$ , trong đó kênh quan trọng được khuếch đại, kênh nhiều/ít liên quan bị suy giảm.

Cơ chế SE *không áp đặt* vị trí điện cực quan trọng theo kinh nghiệm, mà *học* phân bổ trọng số từ dữ liệu qua  $\mathbf{z} \rightarrow \mathbf{s}$ . Khi số điện cực giảm (14 hoặc 5), trọng số kênh giúp tập trung vào các vị trí còn lại có đóng góp cao (thường ở vùng trán/thái dương nếu dữ liệu cho thấy), nhờ đó duy trì hiệu suất trong ràng buộc phần cứng. So với cách xử lý đồng đều, SE vừa *giảm tác động nhiễu* ở kênh ít liên quan, vừa *tăng độ tương phản thông tin* ở kênh hữu ích, làm rõ ràng hơn các đặc trưng phục vụ phân loại cảm xúc.

### 3.2.3.2 Học đa tầng (Multi-Scale Feature Learning)

Học đa tầng (*Multi-Scale Feature Learning*) là kỹ thuật quan trọng giúp mô hình học các đặc trưng EEG ở nhiều cấp độ khác nhau. Tín hiệu EEG có cấu trúc phức tạp, đồng thời chứa các thành phần tần số chậm–nhanh phản ánh nhiều khía cạnh hoạt động thần kinh; các dải như *alpha*, *theta*, *beta*, *gamma* liên hệ tới các trạng thái nhận thức/cảm xúc khác nhau. Vì vậy, trích xuất đặc trưng từ nhiều thang thời gian giúp cải thiện hiệu suất nhận diện cảm xúc. Tuy nhiên, các mô hình CNN truyền thống thường dùng một kích thước hạt nhân cố định, dễ bỏ sót thông tin ở các thang khác nhau.

Để khắc phục, luận án áp dụng *khối Inception* [105], một kiến trúc cho phép mô hình học đồng thời các đặc trưng ở nhiều phạm vi thời gian. Thay vì một nhánh đơn, Inception kết hợp nhiều nhánh tích chập song song với các kích thước hạt nhân khác nhau (ví dụ:  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ), từ đó thu nhận cả đặc trưng cục bộ lẫn tổng quát của tín hiệu EEG và tối ưu khả năng phân loại cảm xúc.



**Hình 3.2** Cấu trúc khối *Inception* [105].

Như minh hoạ ở Hình 3.2, mỗi nhánh tích chập 1D với kích thước hạt nhân có độ giãn khác nhau sẽ “quan sát” một độ phủ theo thời gian khác nhau: nhánh nhỏ nhạy với

biến thiên nhanh, nhánh lớn nhảy với biến thiên chậm. Khi hoạt động song song và được nối đặc trưng (concatenation), khối Inception hình thành năng lực học đa thang một cách tự nhiên mà không cần tăng sâu mạng hay cố định ở một thang duy nhất; điều này phù hợp với EEG vốn đồng thời chứa các thành phần chậm–nhanh.

Trong triển khai cụ thể, các lớp tích chập  $1 \times 1$  còn giúp giảm số kênh/điện cực đầu vào và học tương tác cục bộ theo kênh, trong khi các lớp  $3 \times 3$  và  $5 \times 5$  nhắm tới các mẫu ở phạm vi trung bình và rộng hơn. Các lớp *pooling* (*Max/Average*) góp phần giảm kích thước và tổng hợp thông tin theo vùng. Kết quả từ các nhánh được nối lại thành biểu diễn đặc trưng phong phú nhưng vẫn gọn nhẹ về tính toán. Khi kết hợp tiếp với khối *Squeeze-and-Excitation* (trình bày ở mục sau), kiến trúc EEG\_SICNet vừa học được đặc trưng đa thang, vừa tập trung vào các điện cực quan trọng, qua đó nâng cao hiệu suất nhận diện và độ ổn định khi tài nguyên hạn chế.

### 3.2.3.3 Kết hợp nhiều nguồn đặc trưng (*Feature Fusion*)

Tín hiệu EEG có thể được trích xuất đặc trưng từ nhiều miền khác nhau, bao gồm đặc trưng miền thời gian (*Time-Domain*), miền tần số (*Frequency-Domain*) và miền thời gian–tần số (*Time-Frequency Features*). Việc kết hợp nhiều loại đặc trưng có thể giúp mô hình học được thông tin toàn diện hơn về tín hiệu EEG.

Một cách tiếp cận phổ biến là sử dụng mạng song song (*Dual-Stream Network*), trong đó một nhánh trích xuất đặc trưng từ miền thời gian và một nhánh trích xuất đặc trưng từ miền tần số. Đặc trưng cuối cùng được kết hợp thông qua:

$$Y_{\text{final}} = \alpha Y_{\text{time}} + (1 - \alpha) Y_{\text{freq}} \quad (3.10)$$

Trong đó,  $\alpha$  là trọng số điều chỉnh mức độ quan trọng của từng loại đặc trưng.

Tổng kết lại, việc triển khai mô hình nhận diện cảm xúc EEG trên thiết bị biên của mạng AIoT đòi hỏi những giải pháp giảm tải tài nguyên nhưng vẫn duy trì hiệu suất cao. Trong mục 3.2, luận án đã phân tích hai hướng tối ưu quan trọng:

- Khối *Squeeze-and-Excitation* (SE), giúp mô hình học cách chọn lọc và tăng cường các điện cực EEG quan trọng nhất.
- Khối *Inception*, cho phép trích xuất đặc trưng EEG trên nhiều thang đo thời gian, giúp cải thiện khả năng nhận diện cảm xúc với số điện cực thấp.

Những kỹ thuật này đóng vai trò cốt lõi trong việc xây dựng mô hình EEG nhẹ, phù hợp với việc triển khai bài toán EEG\_ER theo định hướng AIoT. Dựa trên các tối ưu này, luận án đề xuất giới thiệu hai kiến trúc EEG\_MCSNet và EEG\_SICNet, được thiết kế để đảm bảo vừa thu gọn mô hình vừa tối ưu hóa hiệu quả toàn danh nhận dạng.

### 3.3 Đề xuất mô hình học sâu nhẹ cho bài toán EEG\_ER

#### 3.3.1 Mô tả chung

Trong bối cảnh các thiết bị EEG có khả năng di động ngày càng phổ biến, yêu cầu đặt ra đối với các mô hình nhận dạng cảm xúc là không chỉ đảm bảo độ chính xác cao mà còn phải có kích thước gọn nhẹ, chi phí tính toán thấp và khả năng triển khai thực tế trên phần cứng giới hạn về tài nguyên. Để đáp ứng những yêu cầu này, NCS đề xuất hai mô hình mạng học sâu riêng biệt, được tối ưu hóa cho hai bài toán EEG\_ER khác nhau: bài toán phân loại nhị phân và bài toán phân lớp đa cảm xúc. Cả hai mô hình đều kế thừa nền tảng kiến trúc 1D-CNN với đặc trưng đầu vào là tín hiệu EEG đã được biến đổi miền tần số (sử dụng FFT), được tích hợp thêm các khối SE và Inception nhằm tăng cường khả năng học đặc trưng EEG hiệu quả hơn. Hai mô hình đề xuất là hai hướng triển khai song song, độc lập nhưng bổ sung cho nhau, phản ánh sự phù hợp với các kịch bản ứng dụng khác nhau trong thực tế. Các phần tiếp theo sẽ trình bày chi tiết thiết lập bài toán, kiến trúc mô hình, kết quả thực nghiệm và phân tích đánh giá đối với từng mô hình.

#### 3.3.2 Kiến trúc chi tiết

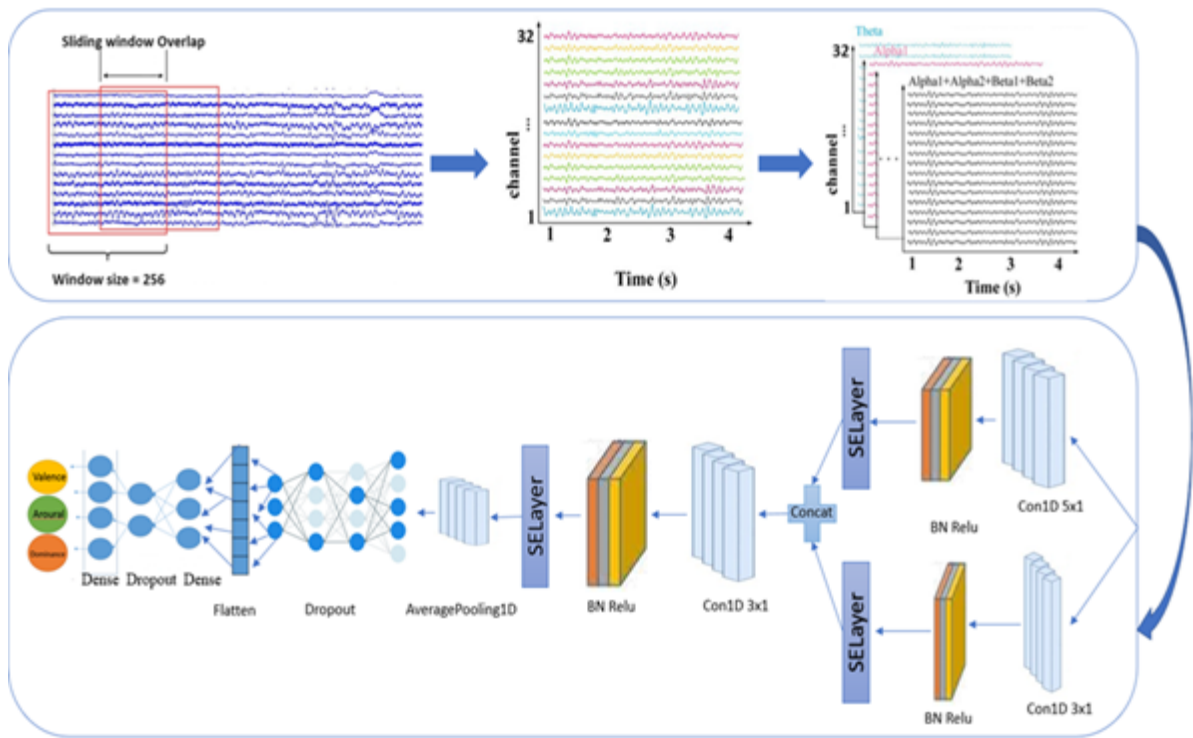
Kiến trúc chung của hai mô hình đề xuất được thể hiện trong Hình 3.3. Kiến trúc này gồm ba giai đoạn chính.

**Giai đoạn 1 - Trích xuất đặc trưng song song:** Tín hiệu EEG sau khi được biến đổi về miền tần số (FFT) được đưa vào hai nhánh xử lý song song nhằm khai thác đặc trưng từ nhiều góc độ.

Nhánh 1 sử dụng Conv1D với kernel size nhỏ (3), kết hợp BatchNormalization để thu nhận các đặc trưng cục bộ. Khối SE được chèn vào cuối nhánh nhằm học trọng số động theo từng điện cực, từ đó làm nổi bật các điện cực quan trọng và giảm thiểu nhiễu.

Nhánh 2 áp dụng Conv1D với kernel size lớn hơn (5) và dilation rate để khai thác thông tin từ các khoảng thời gian rộng hơn, phù hợp với các đặc trưng EEG có tính liên kết dài. Khối SE tiếp tục được tích hợp ở cuối nhánh. Kết quả từ hai nhánh được nối với nhau (concatenation), tạo nên một biểu diễn đặc trưng đa dạng, cân bằng giữa thông tin chi tiết và tổng quát.

**Giai đoạn 2 – Tăng cường toàn cục với SE:** Biểu diễn đặc trưng hợp nhất được đưa vào một khối SE thứ hai. Giai đoạn này có vai trò tái cân bằng toàn cục, giúp mô hình tập trung mạnh hơn vào các vùng tín hiệu có giá trị phân loại cao, đồng thời giảm thiểu ảnh hưởng của nhiễu từ các điện cực ít liên quan. Điều này đặc biệt quan trọng khi xử lý dữ liệu EEG từ thiết bị có số điện cực giới hạn (ví dụ: 14 điện cực hoặc 5 điện cực).



**Hình 3.3** Kiến trúc chung của hai mô hình đề xuất

**Giai đoạn 3 – Phân loại:** Biểu diễn đặc trưng sau SE được đưa qua một khối Conv1D cuối, kết hợp với AveragePooling1D, Dropout và Dense để giảm nhiễu, trích xuất đặc trưng cuối cùng và thực hiện phân loại. Tùy thuộc vào mục tiêu của bài toán, lớp đầu ra sử dụng: Hàm kích hoạt Sigmoid cho bài toán phân loại nhị phân, hoặc Softmax cho bài toán phân loại đa lớp.

Sự kết hợp giữa khối *SE* và khối *Inception* trong mô hình giúp cân bằng giữa tối ưu hóa trọng số điện cực EEG và học đặc trưng đa tỉ lệ. Mỗi tầng 1D-CNN có một khối *SE* đi kèm, giúp mô hình học cách tập trung vào các điện cực EEG quan trọng trước khi đưa tín hiệu vào khối *Inception* để trích xuất đặc trưng ở nhiều khoảng thời gian khác nhau. Điều này giúp mô hình vừa tận dụng tối đa thông tin EEG, vừa đảm bảo hiệu suất tối ưu trên thiết bị di động.

Trong các ứng dụng EEG\_ER, mô hình triển khai không chỉ cần đảm bảo độ chính xác cao mà còn phải phù hợp với điều kiện hoạt động thực tế và tài nguyên tính toán của thiết bị. Tùy theo ứng dụng cụ thể, bài toán có thể được xây dựng dưới hai dạng chính:

- **Phân loại nhị phân (Binary Classification):** Ứng dụng trong các hệ thống dò trạng thái tinh thần (ví dụ: nhận diện căng thẳng - thư giãn, tỉnh táo - buồn ngủ). Đây là yêu cầu phổ biến trong các hệ thống cảnh báo nguy hiểm, thiết bị hỗ trợ y tế, nơi thời gian suy luận cần nhanh chóng và tiêu thụ tài nguyên hạn chế.
- **Phân loại đa lớp (Multi-Class Classification):** Phục vụ các ứng dụng phân tích



cảm xúc chuyên sâu, như nhận diện nhiều trạng thái cảm xúc khác nhau (hạnh phúc, buồn bã, tức giận, sợ hãi, v.v.). Những bài toán này thường yêu cầu mô hình phức tạp hơn để trích xuất đặc trưng từ tín hiệu EEG một cách tối ưu.

---

### Thuật toán 3.1: EEG-SICNet

---

**Input:**  $X_{\text{raw}}$ : dữ liệu EEG;

$y_{\text{raw}}$ : nhãn cảm xúc tương ứng;

Các siêu tham số: kích thước cửa sổ FFT  $w_{\text{FFT}}$ , bước trượt  $s_{\text{FFT}}$ , số điện cực;

**Output:**  $X$ : tín hiệu sau xử lý với kích thước  $(N, T, C)$

```

1: // Bước 1: Tiền xử lý
2: for each  $x \in X_{\text{raw}}$  do
3:     Chia tín hiệu bằng cửa sổ trượt kích thước  $w_{\text{FFT}}$ , bước  $s_{\text{FFT}}$ ;
4:     Áp dụng FFT và trích xuất đặc trưng phổ theo các dải:  $\delta, \theta, \alpha, \beta, \gamma$ ;
5:     Tạo  $X$  với kích thước  $(N, T, C)$ ;
6: end for
7: // Bước 2: Chia dữ liệu
8: Chia  $X$  và  $y_{\text{raw}}$  thành tập huấn luyện  $(X_{\text{train}}, y_{\text{train}})$  và kiểm tra  $(X_{\text{test}}, y_{\text{test}})$  theo tỷ lệ 75/25;
9: // Bước 3: Khởi tạo mô hình
10: Khởi tạo mô hình EEG-SICNet tích hợp khối SE và Inception;
11: // Bước 4: Huấn luyện mô hình
12: for  $epoch = 1$  to  $N_{\text{epoch}}$  do
13:     for mỗi batch  $(X, y)$  trong  $X_{\text{train}}$  do
14:         Trích xuất đặc trưng:
15:             Nhánh 1:  $X_1 = \text{Conv1D}(X) \rightarrow \text{BatchNorm}(X_1) \rightarrow \text{SELayer}(X_1)$ ;
16:             Nhánh 2:  $X_2 = \text{Conv1D}(X) \rightarrow \text{BatchNorm}(X_2) \rightarrow \text{SELayer}(X_2)$ ;
17:             Nhánh tổng hợp:  $X_{\text{Concat}} = \text{Concatenate}([X_1, X_2])$ ;
18:             Dự đoán:  $\hat{y} = \text{Softmax}(X_{\text{Concat}})$ ;
19:             Tính hàm mất mát  $\mathcal{L}_{\text{CE}}$  và cập nhật gradient;
20:     end for
21: end for
22: return mô hình EEG-SICNet đã huấn luyện

```

---

Thuật toán được thực hiện qua 4 bước:

**Bước 1 — Tiền xử lý và biểu diễn.** Mỗi bản ghi  $x \in X_{\text{raw}}$  được lát bằng cửa sổ kích thước  $w_{\text{FFT}}$  với bước trượt  $s_{\text{FFT}}$  nhằm vừa bảo đảm đủ ngữ cảnh thời gian, vừa tăng số mẫu huấn luyện. Trên từng cửa sổ, tín hiệu được biến đổi FFT và gom *năng lượng*

theo bằng  $(\delta, \theta, \alpha, \beta, \gamma)$  để thu được đặc trưng ổn định hơn trước nhiễu và phù hợp với các hạt nhân tích chập 1D. Đặc trưng được chuẩn hoá theo kênh (z-score hoặc min-max) nhằm cân bằng thang đo giữa các điện cực, từ đó thuận lợi cho tối ưu hoá. Dữ liệu đầu ra được tổ chức theo quy ước  $(N, T, C)$ , với  $N$  là số cửa sổ sau tiền xử lý,  $T$  là số bước theo thời gian (hoặc chiều đặc trưng theo trục thời gian), và  $C$  là số điện cực.

**Bước 2 — Chia dữ liệu và gán nhãn.** Việc phân chia train/test được thực hiện trước khi tạo batch và ở cấp *trial/subject* (tuỳ giao thức) để tránh rò rỉ thông kê giữa hai tập. Nhãn của từng cửa sổ kế thừa nhãn của trial tương ứng; với các thiết lập gán nhãn theo khung thời gian, sử dụng quy tắc đa số hoặc quy tắc dựa trên thời điểm kích thích để đảm bảo nhất quán.

**Bước 3 — Khởi tạo mô hình EEG\_SICNet.** Mô hình có hai nhánh tích chập song song: nhánh *hạt nhân nhỏ* (ví dụ kernel 3) nhắm tới đặc trưng cục bộ; nhánh *hạt nhân lớn hơn* (ví dụ kernel 5 kèm *dilation*) bao phủ biến thiên dài hơn. Mỗi nhánh gắn SE ở cuối để học *trọng số theo điện cực* trên cơ sở dữ liệu, nhấn mạnh kênh hữu ích và giảm ảnh hưởng của nhiễu. Đặc trưng hai nhánh được Concatenate theo trục kênh/đặc trưng để hình thành biểu diễn hợp nhất; một SE **toàn cục** (nếu sử dụng) giúp tái cân bằng liên kênh ở mức hợp nhất mà không làm tăng đáng kể chi phí suy luận.

**Bước 4 — Huấn luyện và tối ưu.** Trong mỗi mini-batch  $(X, y)$ , đặc trưng được trích xuất qua hai nhánh  $\rightarrow$  nối đặc trưng  $\rightarrow$  (tuỳ chọn) SE toàn cục  $\rightarrow$  *đầu phân loại gọn* gồm Conv1D cuối + GlobalAveragePooling1D (hoặc AveragePooling1D) + Dropout + Dense. Hàm kích hoạt Sigmoid/Softmax được chọn tương ứng với bài toán nhị phân/đa lớp. Tham số được tối ưu bằng Cross-Entropy; có thể áp dụng lịch học và/hoặc early stopping để tránh quá khớp. Cách bố trí này duy trì cân bằng giữa năng lực biểu đạt (đa thang + chú ý theo kênh) và chi phí suy luận, phù hợp triển khai trên thiết bị tài nguyên hạn chế.

Từ hai hướng triển khai trên, NCS đề xuất:

**EEG\_MCISNet** hướng đến bài toán phân loại nhị phân, giúp triển khai trên các hệ thống tài nguyên hạn chế như thiết bị di động hoặc AIoT. Mô hình được tinh giản để tối ưu hóa tốc độ suy luận trong khi vẫn giữ được khả năng nhận diện chính xác.

**EEG\_SICNet** được tối ưu hóa cho bài toán phân loại đa lớp, nhằm nhận diện nhiều trạng thái cảm xúc với độ chính xác cao nhất.

Để hiện thực hóa kiến trúc đã đề xuất, một quy trình huấn luyện được thiết kế nhằm tối ưu hóa các tham số trong mô hình *EEG\_SICNet*. Thuật toán 3.1 chi tiết hóa từng bước trong quá trình tiền xử lý dữ liệu, trích xuất đặc trưng và tối ưu hóa mô hình thông qua việc điều chỉnh trọng số của *khối SE* và *khối Inception*. Phần tiếp theo sẽ đi sâu vào so sánh kết quả thực nghiệm của hai mô hình EEG\_SICNet và EEG\_MCISNet trên các

tập dữ liệu EEG tiêu chuẩn, đánh giá hiệu suất thực tế của mô hình trên các nền tảng tính toán khác nhau.

### 3.3.3 *Đánh giá cấu trúc và hiệu suất thực nghiệm của EEG\_MCISNet, EEG\_SICNet*

Nhằm đánh giá hiệu quả của hai mô hình EEG\_SICNet và EEG\_MCISNet trong bài toán EEG\_ER, các thử nghiệm thực nghiệm đã được tiến hành trên các bộ dữ liệu tiêu chuẩn và môi trường triển khai khác nhau. Việc đánh giá được thực hiện trên hai tiêu chí chính:

**Hiệu suất nhận dạng cảm xúc EEG:** So sánh kết quả của hai mô hình với các phương pháp hiện có trên tập dữ liệu tiêu chuẩn. EEG\_SICNet được đánh giá trong bài toán phân loại nhiều lớp cảm xúc, trong khi EEG\_MCISNet được thử nghiệm trên bài toán phân loại nhị phân.

**Khả năng triển khai thực tế:** Đánh giá tốc độ suy luận, mức tiêu thụ bộ nhớ, và khả năng hoạt động của mô hình trên Raspberry Pi.

Các thử nghiệm được tiến hành trên các bộ dữ liệu EEG: DEAP [60], AMIGOS [77], DREAMER [56] (được mô tả trong phần 1.3.3), với số lượng điện cực EEG khác nhau (32, 14, 5 điện cực). Môi trường huấn luyện chính được thực hiện trên Google Colab GPU, trong khi hiệu suất triển khai được kiểm tra trên thiết bị tài nguyên hạn chế nhằm đảm bảo tính ứng dụng thực tế của mô hình.

#### 3.3.3.1 *Đánh giá thực nghiệm EEG\_SICNet*

EEG\_SICNet được đánh giá trên các tập dữ liệu tiêu chuẩn DEAP, AMIGOS và DREAMER nhằm kiểm tra khả năng tổng quát hóa của mô hình đối với các tập dữ liệu EEG khác nhau. Mô hình EEG\_SICNet được huấn luyện trong môi trường Google Colab Pro, sử dụng GPU NVIDIA T100, với các tham số cài đặt như sau: Các bộ dữ liệu được chia với tỉ lệ 75% cho huấn luyện và 25% cho kiểm tra, batch size 256, learning rate  $1.25 \times 10^{-4}$ , và tối đa 200 epoch, với cơ chế dừng sớm dựa trên hiệu suất tập kiểm định. Mô hình sử dụng thuật toán tối ưu hóa Adam với các tham số mặc định và hàm mất mát cross-entropy để phân loại đa lớp. Đầu ra của mạng bao gồm ba nhãn cảm xúc tương ứng với ba chiều valence, arousal, và dominance.

Bên cạnh môi trường huấn luyện, nhóm nghiên cứu cũng triển khai EEG\_SICNet trên thiết bị Raspberry Pi 4 nhằm đánh giá khả năng ứng dụng trong thực tế. Các chỉ số được ghi nhận bao gồm: độ chính xác (Accuracy), F1-score, kích thước mô hình (Model size), và thời gian suy luận (Inference time).

Bảng 3.1 cho thấy rằng mạng EEG\_SICNet đạt hiệu suất cao nhất trên bộ dữ liệu DEAP với 32 điện cực. Cụ thể, đối với nhãn *Arousal*, độ chính xác đạt  $95.56 \pm 0.78\%$ , F1 Score là  $95.41 \pm 0.97\%$ . Khi số lượng điện cực giảm xuống còn 14, hiệu suất trên

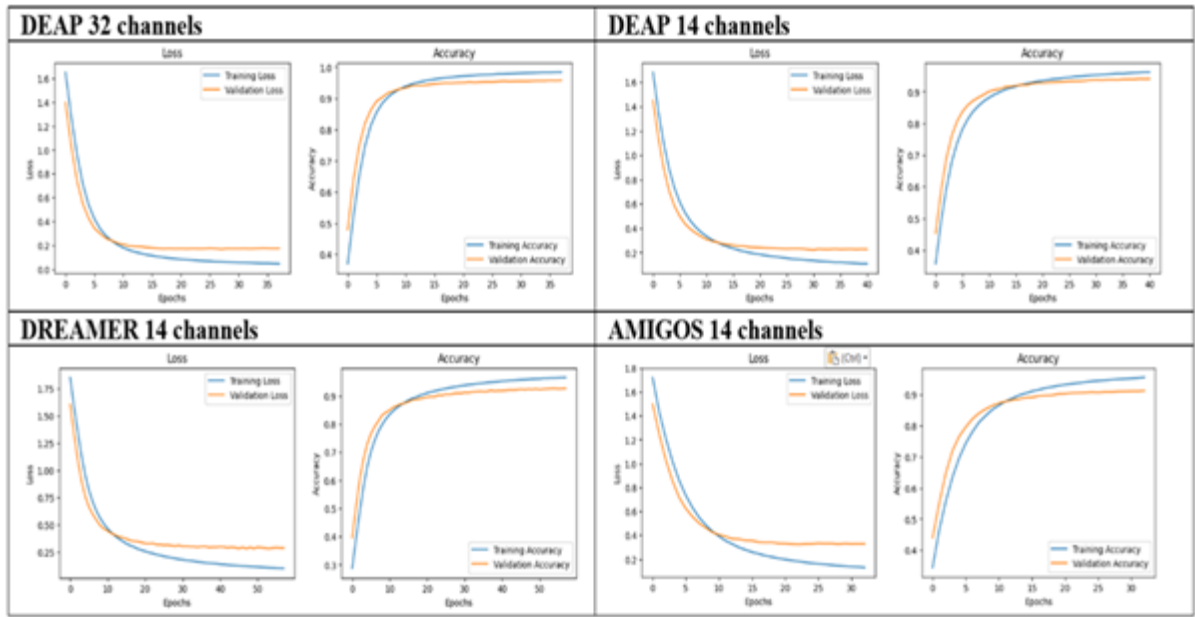
DEAP vẫn duy trì ở mức cao với Accuracy là  $93.87 \pm 1.02\%$ , cho thấy tính tổng quát hóa tốt của mô hình.

**Bảng 3.1** Hiệu suất mạng EEG\_SICNet trên các bộ dữ liệu EEG

| Bộ dữ liệu          | Nhãn      | Accuracy (%)                       | F1 Score (%)     | Recall (%)       | Precision (%)    |
|---------------------|-----------|------------------------------------|------------------|------------------|------------------|
| <b>DEAP (32)</b>    | Arousal   | <b><math>95.56 \pm 0.78</math></b> | $95.41 \pm 0.97$ | $95.44 \pm 0.72$ | $95.42 \pm 0.86$ |
|                     | Valence   | <b><math>95.92 \pm 0.81</math></b> | $95.81 \pm 0.88$ | $95.88 \pm 0.62$ | $95.92 \pm 0.64$ |
|                     | Dominance | <b><math>95.76 \pm 0.69</math></b> | $95.80 \pm 0.77$ | $95.73 \pm 0.67$ | $95.74 \pm 0.63$ |
| <b>DEAP (14)</b>    | Arousal   | $93.87 \pm 1.02$                   | $93.87 \pm 1.11$ | $93.81 \pm 1.04$ | $93.82 \pm 1.07$ |
|                     | Valence   | $93.79 \pm 0.85$                   | $93.81 \pm 0.93$ | $93.95 \pm 0.87$ | $93.91 \pm 0.99$ |
|                     | Dominance | $93.81 \pm 0.90$                   | $93.88 \pm 0.93$ | $93.85 \pm 0.85$ | $93.81 \pm 0.91$ |
| <b>DREAMER (14)</b> | Arousal   | $92.41 \pm 0.44$                   | $92.43 \pm 0.50$ | $92.41 \pm 0.51$ | $92.41 \pm 0.51$ |
|                     | Valence   | $92.56 \pm 0.54$                   | $92.55 \pm 0.53$ | $92.56 \pm 0.55$ | $92.56 \pm 0.54$ |
|                     | Dominance | $92.28 \pm 0.55$                   | $92.25 \pm 0.56$ | $92.27 \pm 0.57$ | $92.29 \pm 0.59$ |
| <b>AMIGOS (14)</b>  | Arousal   | $90.53 \pm 0.66$                   | $90.54 \pm 0.67$ | $90.53 \pm 0.65$ | $90.54 \pm 0.65$ |
|                     | Valence   | $90.74 \pm 1.01$                   | $90.77 \pm 0.74$ | $90.75 \pm 0.70$ | $90.75 \pm 0.70$ |
|                     | Dominance | $90.50 \pm 0.77$                   | $90.48 \pm 0.70$ | $90.47 \pm 0.65$ | $90.51 \pm 0.65$ |

Khi thử nghiệm trên các bộ dữ liệu DREAMER và AMIGOS, mạng EEG\_SICNet cũng cho thấy sự ổn định với độ chính xác lần lượt là  $92.41 \pm 0.44\%$  và  $90.53 \pm 0.66\%$  cho nhãn Arousal. Điều này có thể được giải thích bởi tính phức tạp của các tín hiệu cảm xúc trong bộ dữ liệu này và sự khác biệt trong phương pháp thu thập.

Để phân tích chi tiết hơn hiệu suất của mạng EEG\_SICNet, NCS đánh giá các đường cong học (Learning Curves) của nhãn Arousal trên các bộ dữ liệu với đầu vào 32 và 14 điện cực. Hình 3.4 minh họa sự thay đổi của *Loss* và *Accuracy* trong quá trình huấn luyện và kiểm tra trên từng bộ dữ liệu. Đối với bộ dữ liệu DEAP 32 điện cực, cả đường *Training Loss* và *Validation Loss* đều giảm ổn định qua các epoch, với độ sai số hội tụ ở mức thấp ( $\sim 0.17$ ). Tương tự, độ chính xác (*Accuracy*) trong huấn luyện và kiểm tra đều nhanh chóng đạt mức bão hòa trên 95%, cho thấy mạng đã học tốt các đặc trưng cảm xúc từ dữ liệu. Điều này khẳng định rằng cấu hình đầu vào đầy đủ 32 điện cực cung cấp thông tin phong phú, giúp mạng khai thác tối đa các đặc trưng của tín hiệu EEG.



**Hình 3.4** Biểu đồ quá trình học của mạng EEG\_SICNet

Trong khi đó, với cấu hình 14 điện cực, đường cong học cũng thể hiện xu hướng hội tụ tương tự, tuy nhiên, mức hội tụ của *Loss* cao hơn một chút và độ chính xác thấp hơn so với cấu hình 32 điện cực. Cụ thể, đối với bộ dữ liệu DREAMER và AMIGOS, sự chênh lệch giữa *Training Accuracy* và *Validation Accuracy* là nhỏ, chứng tỏ mạng không bị hiện tượng *overfitting* đáng kể dù số lượng điện cực đầu vào giảm. Điều này là kết quả từ thiết kế linh hoạt của mạng EEG\_SICNet, vốn được tối ưu hóa để hoạt động tốt trên các cấu hình dữ liệu hạn chế. Tuy nhiên, tốc độ hội tụ trên các bộ dữ liệu như DREAMER và AMIGOS có phần chậm hơn so với DEAP, điều này do tính không đồng nhất trong dữ liệu hoặc số lượng mẫu hạn chế hơn. Mặc dù vậy, hiệu suất của mạng EEG\_SICNet vẫn duy trì ở mức ổn định trên tất cả các bộ dữ liệu, khẳng định tính tổng quát hóa và hiệu quả của mạng khi áp dụng trong các môi trường khác nhau.

Bảng 3.2 cung cấp một cái nhìn tổng quan về hiệu năng của mạng EEG\_SICNet khi giảm số lượng điện cực đầu vào xuống còn 5 điện cực, nhằm kiểm chứng tính khả dụng của mô hình trong bối cảnh sử dụng các thiết bị EEG di động hoặc giá rẻ với số lượng điện cực hạn chế. Kết quả thu được cho thấy rằng dù số lượng điện cực giảm, các chỉ số Accuracy và F1-score vẫn duy trì ở mức chấp nhận được trên tất cả các bộ dữ liệu, cụ thể:

- **Bộ dữ liệu DEAP:** Mạng EEG\_SICNet đạt Accuracy cao nhất ở nhãn Valence với  $87.25 \pm 0.22$ , và F1-score đạt  $87.24 \pm 0.32$ . Điều này cho thấy ngay cả khi giảm số điện cực, mạng vẫn giữ được hiệu năng khá tốt trong việc phân biệt cảm xúc, đặc biệt là ở nhãn Valence.

- **Bộ dữ liệu DREAMER:** Mạng đạt Accuracy  $85.71 \pm 0.68$  và F1-score  $85.67 \pm 0.70$  cho nhãn Valence, trong khi các nhãn khác cũng giữ được hiệu năng tương tự với độ lệch chuẩn thấp hơn so với DEAP và AMIGOS. Điều này phản ánh rằng DREAMER có cấu trúc dữ liệu đồng nhất hơn, hỗ trợ mô hình duy trì hiệu năng ổn định.
- **Bộ dữ liệu AMIGOS:** Dữ liệu AMIGOS với độ phức tạp cao hơn vẫn đạt được Accuracy  $83.64 \pm 0.51$  và F1-score  $83.65 \pm 0.56$  ở nhãn Valence, mặc dù thấp hơn DEAP và DREAMER. Sai số có xu hướng lớn hơn trên bộ dữ liệu này, đặc biệt ở nhãn Arousal và Dominance, phản ánh sự đa dạng và biến động trong dữ liệu.

**Bảng 3.2** Kết quả hoạt động của mạng EEG\_SICNet với đầu vào là 5 điện cực

| Chỉ số           | DEAP                               | DREAMER                            | AMIGOS           |
|------------------|------------------------------------|------------------------------------|------------------|
| <b>Arousal</b>   |                                    |                                    |                  |
| Accuracy (%)     | $85.95 \pm 0.47$                   | $84.67 \pm 0.13$                   | $83.45 \pm 0.23$ |
| F1 Score (%)     | $85.96 \pm 0.20$                   | $84.66 \pm 0.44$                   | $83.45 \pm 0.31$ |
| <b>Valence</b>   |                                    |                                    |                  |
| Accuracy (%)     | <b><math>87.25 \pm 0.22</math></b> | $85.71 \pm 0.68$                   | $83.64 \pm 0.21$ |
| F1 Score (%)     | <b><math>87.24 \pm 0.32</math></b> | $85.67 \pm 0.70$                   | $83.65 \pm 0.56$ |
| <b>Dominance</b> |                                    |                                    |                  |
| Accuracy (%)     | $86.35 \pm 0.33$                   | <b><math>84.96 \pm 0.30</math></b> | $83.40 \pm 0.29$ |
| F1 Score (%)     | $86.33 \pm 0.34$                   | <b><math>84.90 \pm 0.30</math></b> | $83.42 \pm 0.32$ |

Sai số (độ lệch chuẩn) trên các chỉ số dao động từ 0.13 đến 0.55, điều này chứng tỏ rằng mạng EEG\_SICNet không chỉ duy trì hiệu năng cao mà còn đảm bảo độ ổn định trong các lần thử nghiệm khác nhau. Đặc biệt, sai số thấp trên bộ dữ liệu DREAMER cho thấy dữ liệu này ít biến động hơn, nhưng điều đó không làm giảm giá trị của DEAP và AMIGOS trong việc cung cấp nhiều tình huống đa dạng hơn.

Kết quả thực nghiệm từ Bảng 3.5 khi đánh giá theo băng tần cho thấy, các băng tần riêng lẻ, đặc biệt là Beta, đóng vai trò quan trọng trong việc nhận diện cảm xúc. Băng Beta đạt độ chính xác cao nhất trên cả ba bộ dữ liệu, đặc biệt với các nhãn Arousal và Valence trên DEAP ( $90.21 \pm 0.67$  và  $90.31 \pm 0.51$ ). Điều này phản ánh rằng các hoạt động trong dải tần Beta có sự liên hệ mật thiết với các trạng thái kích thích và cảm xúc tích cực. Ngoài ra, băng Alpha và Theta cũng thể hiện hiệu quả cao, đặc biệt trên nhãn Arousal và Valence, với độ chính xác trung bình dao động từ 83% đến 84% trên DREAMER và AMIGOS, gợi ý rằng chúng bổ sung thông tin quan trọng trong việc phân loại cảm xúc. Mặc dù các băng tần thấp như Delta và cao như Gamma có hiệu suất

thấp hơn, chúng vẫn cung cấp một mức độ thông tin nhất định, với độ chính xác ổn định trên cả ba bộ dữ liệu.

Khi kết hợp các băng tần, đặc biệt là Beta, Alpha và Theta, mang lại hiệu quả vượt trội so với việc sử dụng từng băng tần riêng lẻ. Trên DEAP, sự kết hợp này đạt hiệu suất tối ưu, với độ chính xác lên đến  **$90.70 \pm 0.60$  (Arousal)** và  **$90.60 \pm 0.55$  (Valence)**, chứng minh rằng các băng tần có khả năng bổ trợ thông tin lẫn nhau để cải thiện hiệu quả. Điều này đặc biệt có ý nghĩa khi áp dụng vào các bộ dữ liệu hạn chế tài nguyên như DREAMER và AMIGOS, nơi các tổ hợp băng tần vẫn duy trì hiệu suất ổn định khoảng **83%**. Những phát hiện này mở ra tiềm năng tối ưu hóa việc lựa chọn băng tần dựa trên mục tiêu ứng dụng, giúp giảm thiểu chi phí tính toán mà vẫn đảm bảo hiệu suất cao.

**Bảng 3.3** So sánh hiệu suất của mạng EEG\_SICNet với các phương pháp khác trên bộ dữ liệu DEAP (32 điện cực).

| Mô hình                     | Valence      | Arousal      | Dominance    | Nguồn tài liệu |
|-----------------------------|--------------|--------------|--------------|----------------|
| 1D-CNN, 2D-CNN              | 95.18        | 95.58        | -            | [84]           |
| STCN                        | 95.02        | 95.29        | -            | [125]          |
| AP-CapsNet                  | 93.89        | 95.04        | 95.08        | [69]           |
| FPN-LSTM                    | 90.05        | 90.84        | -            | [133]          |
| 2D-CNN                      | 90.76        | 92.92        | 92.97        | [108]          |
| <b>EEG_SICNet (Đề xuất)</b> | <b>95.56</b> | <b>95.92</b> | <b>95.76</b> | -              |

Kết quả trong Bảng 3.3 cho thấy mạng EEG\_SICNet đạt được hiệu suất tốt hơn so với các phương pháp khác đã được công bố trên cùng bộ dữ liệu DEAP sử dụng 32 điện cực EEG. Cụ thể, EEG\_SICNet đạt **95.56%** ở nhãn Valence, **95.92%** ở nhãn Arousal, và **95.76%** ở nhãn Dominance, đều cao hơn so với các mô hình 1D-CNN, 2D-CNN, STCN, AP-CapsNet, FPN-LSTM và 2D-CNN.

So với mô hình 1D-CNN và 2D-CNN (Acc Val = 95.18%, Acc Aro = 95.58%), EEG\_SICNet mang lại sự cải thiện rõ rệt ở cả ba nhãn, nhờ khả năng khai thác thông tin hiệu quả từ module SE và Inception được tích hợp trong kiến trúc mạng. Trong khi đó, các phương pháp như STCN (Acc Val = 95.02%, Acc Aro = 95.29%) và AP-CapsNet (Acc Val = 93.89%, Acc Aro = 95.04%, Acc Dom = 95.08%) tuy đạt hiệu năng tốt ở các nhãn Arousal và Dominance, nhưng vẫn thấp hơn EEG\_SICNet.

Đặc biệt, so với các phương pháp dựa trên 2D-CNN như FPN-LSTM (Acc Val = 90.05%) và 2D-CNN (Acc Val = 90.76%), sự cải thiện của EEG\_SICNet trở nên rõ rệt. Điều này nhấn mạnh vai trò quan trọng của việc sử dụng các khối SE và kiến trúc đa nhánh (Inception) trong việc khai thác đặc trưng không gian và tần số từ tín hiệu EEG.

Từ kết quả này không chỉ khẳng định hiệu suất của mạng EEG\_SICNet mà còn chứng minh khả năng áp dụng thực tế trong các ứng dụng EEG\_ER. Sự cải thiện hiệu năng so với các phương pháp trước đó cho thấy mạng EEG\_SICNet đã giải quyết tốt bài toán phân loại cảm xúc, đặc biệt ở các nhãn *Valence* và *Dominance*, vốn thường gặp khó khăn trong các nghiên cứu trước đây.

Trong lĩnh vực EEG\_ER, các thử nghiệm lý thuyết thường được tiến hành trên các bộ dữ liệu chuẩn hóa, như DEAP, DREAMER hay AMIGOS, nơi mà tín hiệu được thu thập trong điều kiện lý tưởng, môi trường được kiểm soát chặt chẽ, thiết bị đo lường chất lượng cao và người tham gia được hướng dẫn cụ thể. Tuy nhiên, để đánh giá tính ứng dụng thực tế của một hệ thống nhận dạng cảm xúc, cần thiết phải kiểm tra nó trong các điều kiện thực tế, nơi các yếu tố bất định và hạn chế của thiết bị có thể ảnh hưởng đến hiệu suất.

NCS và nhóm nghiên cứu đã thiết kế một hệ thống EEG\_ER với các thành phần và quy trình tương tự kịch bản thử nghiệm của bộ dữ liệu DEAP. Tuy nhiên, thay vì sử dụng tín hiệu từ các bộ dữ liệu chuẩn, tín hiệu EEG được thu thập từ 4 tình nguyện viên (2 nam, 2 nữ) trong môi trường thực tế. Tín hiệu được ghi bằng thiết bị không dây Emotiv Epoc (14 điện cực, tần số lấy mẫu 128 Hz) — một thiết bị di động, chi phí thấp, phù hợp cho các tình huống sử dụng hàng ngày. Quy trình thu thập diễn ra trong môi trường không kiểm soát nghiêm ngặt; người tham gia xem các đoạn video gợi cảm xúc tương tự thiết lập của DEAP. Dữ liệu được thu qua phần mềm EmotivPro và xử lý bằng EEGLAB để trích xuất các đặc trưng FFT (năng lượng theo các dải delta, theta, alpha, beta, gamma) như đã mô tả trước đó. Để hiện thực hoá hệ thống trên thiết bị biên, Emotiv Epoc được kết nối không dây với Raspberry Pi 4 (RAM 8 GB), nơi mô hình EEG\_SICNet được triển khai suy luận thời gian thực với cấu hình batch = 1 và độ chính xác số học FP32, không sử dụng GPU/NPU. Lưu ý rằng các kết quả thực nghiệm phản ánh hiệu năng trên nền tảng tham chiếu Raspberry Pi 4; việc so sánh độ trễ với phần cứng khác không được thực hiện nhằm đảm bảo tính công bằng và nhất quán thí nghiệm.

Bảng 3.4 trình bày kết quả thực nghiệm của mô hình EEG\_SICNet trên thiết bị thực tế, bao gồm độ chính xác (Accuracy), độ mất mát (Loss) và điểm F1 tương ứng với ba nhãn *Valence*, *Arousal*, và *Dominance*. Kết quả cho thấy mô hình đạt độ chính xác lần lượt là 73.32%, 72.34%, và 72.03% đối với các nhãn *Valence*, *Arousal*, và *Dominance*. Trong khi đó, điểm F1 cũng đạt các giá trị khá tương đồng, dao động từ 75.88% đến 78.45%. Mặc dù các kết quả này thấp hơn đáng kể so với các kết quả thu được trên bộ dữ liệu chuẩn như *DEAP*, *AMIGOS*, và *DREAMER*, chúng vẫn cho thấy mô hình EEG\_SICNet có tiềm năng ứng dụng thực tế.

Sự suy giảm hiệu suất này có thể được lý giải bởi một số nguyên nhân chính. Đầu tiên, việc thu thập dữ liệu từ thiết bị thực tế nhiều yếu tố gây nhiễu, như môi trường thu



**Bảng 3.4** Kết quả nhận dạng của thực tế của mô hình thử nghiệm

| Label        | Valence          | Arousal          | Dominance        |
|--------------|------------------|------------------|------------------|
| Accuracy (%) | $73.32 \pm 0.57$ | $72.34 \pm 0.51$ | $72.03 \pm 0.52$ |
| F1 Score     | $78.45 \pm 0.65$ | $76.52 \pm 0.60$ | $75.88 \pm 0.62$ |

tín hiệu không được kiểm soát chặt chẽ, chuyển động của người tham gia, và tín hiệu bị nhiễu từ các thiết bị điện tử xung quanh. Thứ hai, số lượng điện cực của thiết bị thực tế (14 điện cực) ít hơn đáng kể so với các bộ dữ liệu chuẩn (thường sử dụng 32 hoặc 64 điện cực), dẫn đến mất mát thông tin quan trọng trong quá trình phân tích tín hiệu EEG. Thứ ba, số lượng mẫu thử nghiệm hạn chế (chỉ gồm 4 người tình nguyện) cũng làm tăng sự không đồng nhất và độ biến động trong dữ liệu, khiến mô hình khó đạt được hiệu suất tối ưu.

Tuy nhiên, các kết quả này vẫn đáng khích lệ với nhóm nghiên cứu, chứng minh rằng mô hình EEG\_SICNet có thể hoạt động trên các thiết bị thực tế, với tiềm năng ứng dụng trong các hệ thống nhận diện cảm xúc sử dụng tín hiệu EEG.

#### 3.3.3.2 Đánh giá thực nghiệm EEG\_MCISNet

EEG\_MCISNet được thiết kế hướng đến các ứng dụng nhận diện trạng thái cảm xúc nhị phân, chẳng hạn như *stress detection* (phát hiện căng thẳng) hoặc *emotion detection* (phát hiện cảm xúc tích cực - tiêu cực). Tương tự như các điều kiện thí nghiệm của bộ dữ liệu của EEG\_SICNet, NCS tiến hành thử nghiệm và thu được các kết quả được mô tả trong Bảng 3.5

Bảng 3.6 trình bày độ chính xác của mô hình EEG\_MCISNet khi áp dụng trên ba bộ dữ liệu EEG phổ biến: DEAP, DREAMER và AMIGOS với các cấu hình số lượng điện cực khác nhau. Kết quả cho thấy mô hình đạt độ chính xác rất cao trên bộ dữ liệu DEAP với 32 điện cực, vượt ngưỡng 99% ở cả ba nhãn *Arousal*, *Valence* và *Dominance*. Khi giảm số lượng điện cực xuống còn 14 và 5, mô hình vẫn duy trì hiệu suất ổn định, với độ chính xác trên 92% cho mọi nhãn trong bộ dữ liệu DEAP, cho thấy khả năng thích ứng tốt với dữ liệu thu thập từ thiết bị EEG có cấu hình giới hạn. Trên các bộ dữ liệu DREAMER và AMIGOS, mô hình EEG\_MCISNet cũng thể hiện độ chính xác đáng tin cậy, đặc biệt đạt kết quả tốt hơn khi sử dụng 14 điện cực so với 5 điện cực, điều này cho thấy mối quan hệ tỷ lệ thuận giữa số lượng kênh và hiệu quả phân loại trong hầu hết các trường hợp.

Mặc dù số điện cực và tham số giảm mạnh, EEG\_MCISNet vẫn duy trì độ chính xác cao nhờ (i) *SE* học trọng số theo kênh từ dữ liệu, tự động nhấn mạnh các điện cực còn lại có đóng góp lớn và triệt bớt kênh nhiễu khi rút gọn điện cực; (ii) hai nhánh *Conv1D*

**Bảng 3.5 Độ chính xác (%) khi đánh giá theo băng tần trên các bộ dữ liệu**

| Dữ liệu           |     | Alpha               | Beta                | Gamma               | Theta               | Delta                |
|-------------------|-----|---------------------|---------------------|---------------------|---------------------|----------------------|
| DEAP              | Aro | 90.63 ± 0.38        | <b>90.21 ± 0.67</b> | 89.78 ± 0.45        | 89.91 ± 1.03        | 89.72 ± 0.49         |
|                   | Val | 89.91 ± 0.23        | <b>90.31 ± 0.51</b> | 89.77 ± 0.71        | 90.40 ± 1.32        | <b>90.01 ± 1.40</b>  |
|                   | Dom | <b>90.02 ± 0.31</b> | 90.03 ± 0.50        | 89.77 ± 0.60        | 89.88 ± 1.15        | 89.89 ± 1.12         |
| DREAMER           | Aro | <b>83.29 ± 0.15</b> | <b>83.42 ± 0.37</b> | 82.94 ± 0.35        | <b>83.29 ± 0.72</b> | <b>83.27 ± 0.72</b>  |
|                   | Val | <b>83.32 ± 0.42</b> | <b>83.56 ± 0.26</b> | 83.10 ± 0.54        | <b>83.47 ± 0.45</b> | <b>83.28 ± 0.76</b>  |
|                   | Dom | <b>83.25 ± 0.28</b> | <b>83.60 ± 0.26</b> | 82.96 ± 0.38        | <b>83.31 ± 0.71</b> | <b>83.20 ± 0.80</b>  |
| AMIGOS            | Aro | <b>82.51 ± 0.13</b> | <b>83.14 ± 0.40</b> | <b>83.07 ± 0.36</b> | <b>82.90 ± 0.40</b> | 81.66 ± 0.20         |
|                   | Val | <b>82.24 ± 0.47</b> | 81.35 ± 0.47        | <b>83.07 ± 0.37</b> | <b>82.95 ± 0.20</b> | <b>83.00 ± 0.23</b>  |
|                   | Dom | <b>82.87 ± 0.23</b> | <b>81.96 ± 0.40</b> | <b>81.77 ± 0.37</b> | <b>82.90 ± 0.20</b> | <b>83.20 ± 0.32</b>  |
| Tổng hợp băng tần |     | Beta + Alpha        | Beta + Theta        | Beta + Gamma        | Alpha + Theta       | Beta + Alpha + Theta |
| DEAP              | Aro | 90.50 ± 0.55        | 90.60 ± 0.52        | 90.40 ± 0.53        | 90.00 ± 0.60        | <b>90.70 ± 0.60</b>  |
|                   | Val | 90.40 ± 0.44        | 90.50 ± 0.47        | 90.30 ± 0.48        | 90.10 ± 0.55        | <b>90.60 ± 0.55</b>  |
|                   | Dom | 90.20 ± 0.41        | 90.40 ± 0.42        | 90.10 ± 0.43        | 90.00 ± 0.50        | <b>90.50 ± 0.50</b>  |
| DREAMER           | Aro | 83.50 ± 0.36        | 83.60 ± 0.24        | 83.40 ± 0.22        | 83.50 ± 0.32        | <b>83.70 ± 0.30</b>  |
|                   | Val | 83.60 ± 0.22        | 83.70 ± 0.25        | 83.50 ± 0.30        | 83.40 ± 0.30        | <b>83.80 ± 0.32</b>  |
|                   | Dom | 83.60 ± 0.25        | <b>83.80 ± 0.24</b> | 83.50 ± 0.27        | 83.50 ± 0.34        | <b>83.80 ± 0.33</b>  |
| AMIGOS            | Aro | 83.00 ± 0.33        | 83.20 ± 0.35        | 83.10 ± 0.27        | 83.00 ± 0.40        | <b>83.30 ± 0.32</b>  |
|                   | Val | 83.00 ± 0.29        | 83.00 ± 0.33        | 82.95 ± 0.32        | 82.85 ± 0.35        | <b>83.10 ± 0.34</b>  |
|                   | Dom | 83.00 ± 0.30        | 83.10 ± 0.36        | 83.00 ± 0.32        | 83.00 ± 0.30        | <b>83.20 ± 0.33</b>  |

**Bảng 3.6** Độ chính xác của mô hình EEG\_MCISNet trên các bộ dữ liệu

| Dữ liệu        | Số điện cực | Tham số (Mb) | Arousal | Valence | Dominance |
|----------------|-------------|--------------|---------|---------|-----------|
| <b>DEAP</b>    | 32          | 12.700       | 99.60   | 99.62   | 99.70     |
|                | 14          | 6.067        | 99.55   | 99.62   | 99.67     |
|                | 5           | 2.966        | 99.32   | 98.57   | 98.83     |
| <b>DREAMER</b> | 14          | 3.412        | 97.50   | 96.87   | 96.11     |
|                | 5           | 3.349        | 85.12   | 85.93   | 84.67     |
| <b>AMIGOS</b>  | 14          | 5.611        | 84.14   | 83.20   | 83.02     |
|                | 5           | 4.057        | 85.29   | 82.91   | 82.03     |

đa thang giúp nắm bắt đồng thời mẫu ngắn–dài ngay trên từng kênh, nên ít phụ thuộc vào mật độ điện cực; (iii) đặc trưng miền tần số (FFT gộp theo băng) đã cô đọng thông tin, giúp giảm nhu cầu số kênh mà vẫn giữ tín hiệu phân biệt; (iv) mô hình nhỏ hơn đóng vai trò *regularization*, hạn chế quá khớp. Vì vậy trên DEAP, độ chính xác gần như giữ nguyên kể cả với 5 kênh; trên DREAMER/AMIGOS, khi rút còn 5 điện cực, độ chính xác giảm hơn do khác biệt bố trí điện cực và sự khác biệt về dữ liệu; việc loại bỏ kênh làm suy hao thông tin không gian cần thiết cho phân biệt, nên mô hình bị ảnh hưởng rõ rệt hơn.

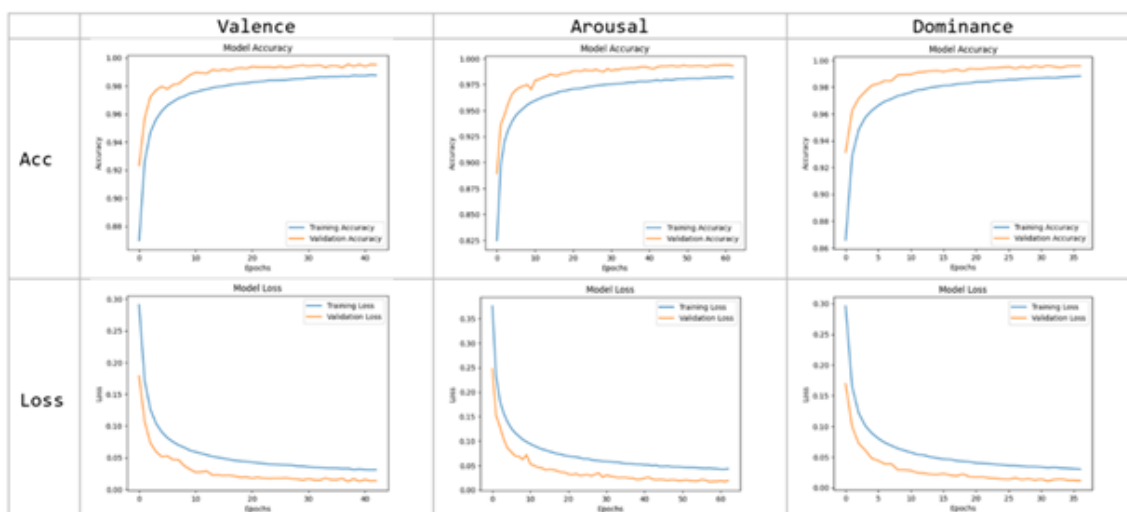
Hiệu quả hoạt động của mạng EEG\_MCISNet đã được NCS so sánh với một số công trình công bố có liên quan gần đây. Trong [3], các tác giả đề xuất một mô hình mạng CNN cải tiến được đặt tên là M2. Hiệu quả hoạt động của mô hình này được đánh giá trên tập dữ liệu DEAP sử dụng 14 điện cực. Với kịch bản tương tự, trong [1], các tác giả cũng đưa ra một kiến trúc mạng CNN cho phân lớp cảm xúc trên bộ dữ liệu DEAP với 14 điện cực đầu vào. Từ mã nguồn do các tác giả của hai công trình trên cung cấp, NCS đã tiến hành đánh giá thêm với trường hợp đầu vào là 32 và 5 điện cực. Kết quả hoạt động các mô hình được thể hiện trên Bảng 3.7 với các tiêu chí đánh giá là số lượng tham số của mô hình, độ tổn thất, độ chính xác và số epoch khi dừng huấn luyện (sử dụng kỹ thuật Early stop).

**Bảng 3.7** Độ chính xác của các mô hình học sâu trên bộ dữ liệu DEAP

| Số điện cực | Mô hình     | Tham số (Mb) | Arousal      | Valence      | Dominance    |
|-------------|-------------|--------------|--------------|--------------|--------------|
| 32          | EEG_MCISNet | 12.701       | <b>99.60</b> | <b>99.62</b> | <b>99.70</b> |
|             | M2          | 2.494        | 99.10        | 96.16        | 99.25        |
|             | CNN         | 6.253        | 94.33        | 93.25        | 99.12        |
| 14          | EEG_MCISNet | 6.067        | <b>99.55</b> | <b>99.62</b> | <b>99.67</b> |
|             | M2          | 1.822        | 97.80        | 92.00        | 98.51        |
|             | CNN         | 3.413        | 92.25        | 89.34        | 95.82        |
| 5           | EEG_MCISNet | 2.966        | <b>99.32</b> | <b>98.57</b> | <b>98.83</b> |
|             | M2          | 4.940        | 92.71        | 90.75        | 96.01        |
|             | CNN         | 3.553        | 86.95        | 85.61        | 91.64        |

Các số liệu thể hiện trên Bảng 3.7 cũng cho thấy mô hình mà NCS đề xuất có số lượng tham số đủ nhỏ, luôn hội tụ nhanh hơn (số Epoch luôn nhỏ hơn) với độ chính xác nhận dạng lớn hơn so với các mô hình còn lại.

Hình 3.5 mô tả quá trình học khi huấn luyện mạng EEG\_MCISNet với các nhãn Valence, Arousal và Dominance khi dữ liệu vào là bộ mẫu được lấy từ 5 điện cực (số điện cực nhỏ nhất). Kết quả này cho thấy kiến trúc mạng hội tụ nhanh và cho kết quả phân loại tốt. Sau 60 epoch, chỉ số acc đạt tới 98% trong khi loss giảm xuống dưới 2%.



**Hình 3.5** Biểu đồ quá trình học khi huấn luyện mạng EEG\_MCISNet với các nhãn Valence, Arousal và Dominance với bộ dữ liệu EEG gồm 5 điện cực

Kết quả thực nghiệm cho thấy EEG\_SICNet và EEG\_MCISNet đều đáp ứng tốt mục tiêu tối ưu hóa mô hình nhận diện cảm xúc từ tín hiệu EEG, với khả năng duy trì độ

chính xác cao ngay cả khi số lượng điện cực EEG giảm. Cả hai mô hình đều thể hiện khả năng tổng quát hóa tốt trên các bộ dữ liệu tiêu chuẩn (DEAP, DREAMER, AMIGOS) và vượt trội so với các phương pháp trước đây nhờ vào sự kết hợp của khối SE và khối Inception.

### 3.4 Kết luận chương 3

Trong chương này, luận án đã tập trung giải quyết mục tiêu thứ ba, đó là đề xuất một kiến trúc mạng học sâu phù hợp để triển khai trên các thiết bị có tài nguyên hạn chế, trong khi vẫn đảm bảo hiệu suất EEG\_ER. Trên cơ sở kiến trúc nền tảng 1D-CNN xác định được ở chương 2, hai kỹ thuật tối ưu cấu trúc là “khối Inception” và “khối Squeeze-and-Excitation” đã được tích hợp nhằm nâng cao khả năng trích xuất đặc trưng trong không gian tần số và giữa các điện cực EEG. Mô hình đề xuất – EEG\_MCISNet – cho bài toán phân loại nhị phân, và mô hình mở rộng – EEG\_SICNet – cho bài toán phân loại đa lớp, đã cho thấy hiệu quả rõ rệt về cả độ chính xác và khả năng triển khai trên nền tảng giới hạn tài nguyên, như được trình bày trong các kết quả thực nghiệm.

Tuy nhiên, trong bối cảnh tín hiệu EEG mang tính nhiễu cao và dữ liệu thu thập thực tế có thể không đồng nhất, mô hình học sâu vẫn có thể gặp khó khăn trong việc tổng quát hóa. Do đó, để tiếp tục nâng cao hiệu suất mô hình mà không làm gia tăng đáng kể số lượng tham số, chương tiếp theo sẽ kế thừa kiến trúc đã được đề xuất, và tiếp tục tích hợp các kỹ thuật như “Tự chất lọc tri thức” (Self-Knowledge Distillation – SKD) và Scaled Weights. Những cải tiến này hứa hẹn sẽ tối ưu hóa quá trình học, tăng độ chính xác và độ ổn định của mô hình trong điều kiện dữ liệu thực tế phức tạp hơn.

Các kết quả nghiên cứu chính của chương 3 đã được NCS công bố trong các công trình CB [5] và CB [6]. Những kết quả này đóng góp quan trọng trong việc phát triển các hệ thống EEG\_ER hiệu quả, tối ưu tài nguyên, mở ra tiềm năng ứng dụng trong các hệ thống theo dõi cảm xúc trong điều kiện tài nguyên hạn chế.

## CHƯƠNG 4

# TỐI ƯU HÓA MÔ HÌNH EEG\_ER SỬ DỤNG KỸ THUẬT “TỰ CHẮT LỌC TRI THỨC”

Trong chương 3, luận án đã đề xuất các kiến trúc mạng học sâu hiệu quả, có khả năng EEG\_ER với độ chính xác cao và kích thước mô hình phù hợp để triển khai trên các thiết bị có tài nguyên hạn chế. Trong thực tế, trước khi áp dụng mô hình trên thiết bị biên, cần phải làm nhẹ mô hình bằng các phương pháp: Cắt tỉa mô hình (Model Pruning), chia sẻ trọng số (Parameter Sharing), lượng tử hóa mô hình (Model Quantization), chắt lọc tri thức (KD), phân rã hạng thấp (Low-rank Factorization), điểm chung của các phương pháp này đều làm giảm độ phức tạp mô hình, đánh đổi bằng hiệu năng. Rẻ theKD ban đầu được xây dựng với mục đích nén mô hình, tuy nhiên với nhiều biến thể, KD không chỉ đơn thuần là phương pháp nén nữa, mà còn được mở rộng ra ứng dụng giải quyết vấn đề khó khăn trong việc khái quát hóa, tăng hiệu năng mà vẫn đảm bảo mô hình gọn nhẹ.

Trong chương này, luận án đề xuất một mô hình học sâu được tối ưu hóa mạnh mẽ mang tên EEG-SKDNet, kết hợp giữa kiến trúc mạng học sâu nhẹ được xây dựng ở chương 3 với kỹ thuật Tự chắt lọc tri thức (Self-Knowledge Distillation – SKD) và cơ chế trọng số có điều chỉnh (Scaled Weights). Đây là một trong những cách tiếp cận đầu tiên áp dụng SKD cho bài toán EEG\_ER. Khác với các phương pháp chắt lọc tri thức (Knowledge Distillation – KD) truyền thống vốn yêu cầu một mô hình giáo viên (Teacher) mạnh được huấn luyện trước, kỹ thuật SKD cho phép mô hình tự học tri thức từ chính nó, thông qua các biểu diễn khác nhau của cùng một đầu vào. Cách tiếp cận này không chỉ giảm đáng kể chi phí huấn luyện mà còn giúp tăng cường khả năng tổng quát hóa, giảm hiện tượng quá khớp và đặc biệt phù hợp với các hệ thống theo định hướng AIoT. Bên cạnh đó, luận án còn đề xuất cơ chế Scaled Weights – một hàm trọng số động được điều chỉnh theo mức năng lượng trung bình của từng đoạn tín hiệu EEG, giúp mô hình tập trung vào các vùng tín hiệu có giá trị cảm xúc cao, qua đó tối ưu hiệu quả học tri thức nội bộ.

Nội dung chính trong chương này bao gồm: Trình bày cơ sở lý thuyết và tổng quan các kỹ thuật chắt lọc tri thức (KD) và tự chắt lọc tri thức (SKD); Đề xuất mô hình EEG-SKDNet kết hợp SKD và Scaled Weights trên kiến trúc 1D-CNN + SE + Inception; Thực hiện các thí nghiệm so sánh hiệu quả với mô hình gốc và các phương pháp KD khác để chứng minh tính ưu việt của phương pháp đề xuất.

## 4.1 Tổng quan về kỹ thuật chắt lọc tri thức và kỹ thuật tự chắt lọc tri thức

### 4.1.1 Chắt lọc tri thức

KD là một kỹ thuật nén mô hình hiệu quả được giới thiệu lần đầu tiên bởi [42]. Khác với học chuyển giao (transfer learning), trong phương pháp KD, chúng ta sử dụng hai mạng: một mạng lớn gọi là mô hình teacher và một mạng nhỏ hơn gọi là mô hình student. KD được đề xuất nhằm truyền tri thức từ mô hình teacher lớn sang mô hình student nhỏ. Với cách tiếp cận này, mô hình student có thể cải thiện khả năng khái quát hóa và hiệu suất nhờ vào tri thức được chắt lọc từ mô hình teacher, trong khi vẫn duy trì chi phí tính toán thấp. KD đã trở thành một hướng nghiên cứu quan trọng trong học máy và học sâu, đặc biệt phù hợp để triển khai các mô hình trong các hệ thống hạn chế tài nguyên như thiết bị di động hoặc nhúng [111].

Trong những năm gần đây, có rất nhiều phương pháp được đề xuất nhằm cải tiến KD truyền thống thông qua các cơ chế truyền tri thức khác nhau. Cụ thể, các phương pháp KD được công bố có thể được phân thành hai lớp chính bao gồm KD dựa trên đầu ra (response-based KD) và KD dựa trên đặc trưng (feature-based KD). Đối với KD dựa trên đầu ra, các phương pháp này được đề xuất để chắt lọc tri thức ở tầng đầu ra, tức là tầng kết nối đầy đủ. Nói cách khác, mô hình học sinh (student) được huấn luyện để mô phỏng đầu ra từ mô hình giáo viên (teacher) [39]. Đối với KD dựa trên đặc trưng, các phương pháp này giới thiệu một cách tiếp cận hiệu quả để truyền tri thức từ các đặc trưng trung gian của mô hình giáo viên sang mô hình học sinh. Cụ thể, mô hình học sinh được yêu cầu học các đặc trưng từ các tầng trung gian của mô hình giáo viên. Điều này được chứng minh là có hiệu quả trong việc giám sát toàn diện, như bản đồ đặc trưng và thông tin điều chỉnh từ các tầng trung gian. Các phương pháp nêu trên đã được áp dụng rộng rãi vào nhiều tác vụ chuẩn như phân loại hình ảnh [55], phát hiện đối tượng [18] và xử lý ngôn ngữ tự nhiên (NLP) [68].

Trong lĩnh vực EEG\_ER, một số nghiên cứu đã áp dụng KD. Ví dụ, [116] sử dụng ResNet-34 làm mô hình giáo viên để hướng dẫn huấn luyện ResNet-8 làm mô hình sinh viên thông qua phương pháp KD. Kết quả thực nghiệm trên hai bộ dữ liệu tiêu chuẩn, bao gồm SEED-IV và DEAP, cho thấy phương pháp được đề xuất đạt hiệu suất tiên tiến hàng đầu.

Một cách để tăng khả năng khái quát hóa của mô hình là sử dụng đa phương thức đầu vào, tuy nhiên điều này thường đòi hỏi nhiều mạng, tức là nhiều luồng xử lý, và mỗi luồng xử lý một kiểu đầu vào khác nhau. Để giảm chi phí tính toán, trong [132], các tác giả đề xuất một phương pháp chắt lọc tri thức xuyên phương thức (cross-modal) cho bài toán nhận diện cảm xúc liên tục. Cụ thể, mô hình teacher trong được huấn luyện trên tác vụ thị giác, ví dụ như nhận diện khuôn mặt, sau đó được sử dụng để hướng dẫn việc

huấn luyện mô hình sinh viên với đầu vào là tín hiệu EEG.

#### **4.1.2 Tự chất lọc tri thức**

Mặc dù các phương pháp KD nêu trên đã đạt được hiệu suất tiên tiến trên nhiều bộ dữ liệu cho bài toán phân loại cảm xúc, chúng vẫn tồn tại một số nhược điểm như:

- (i) Việc tồn tại của mô hình giáo viên làm tăng độ phức tạp của quá trình huấn luyện và đòi hỏi nhiều thời gian cũng như lượng lớn dữ liệu được gán nhãn để huấn luyện một mô hình giáo viên có hiệu suất cao, trong khi tài nguyên tính toán và các bộ dữ liệu lớn không phải lúc nào cũng sẵn có [39].
- (ii) Khoảng cách về năng lực giữa mô hình giáo viên và học sinh khiến mô hình học sinh khó có thể học được đầy đủ thông tin mà giáo viên truyền đạt. Mặc dù Mirzadeh et al. [27] đã đề xuất sử dụng một mô hình trợ giảng (teacher assistant - TA) để giảm khoảng cách về năng lực giữa mô hình giáo và học sinh, khoảng cách này vẫn còn tồn tại, và độ phức tạp của phương pháp tăng đáng kể khi trợ giảng được thêm vào giữa giáo viên và học sinh. Để giải quyết tất cả các hạn chế nêu trên, phương pháp SKD (Self-Knowledge Distillation) đã được đưa ra. Cụ thể, với SKD, chúng ta không cần sử dụng mô hình giáo viên, do đó thời gian huấn luyện được giảm đáng kể.

Hơn nữa, vì không có mô hình giáo viên, nên không tồn tại khoảng cách về năng lực giữa giáo viên và học sinh trong SKD. Kết quả là, toàn bộ quá trình chưng cất tri thức bên trong được thực hiện chỉ bởi mô hình học sinh [111]. Nhiều phương pháp SKD khác nhau đã được đề xuất cho nhiều tác vụ khác nhau, ví dụ như phân loại, phát hiện đối tượng [121, 35], và xử lý ngôn ngữ tự nhiên (NLP) [130].

Lấy cảm hứng từ SKD trong các lĩnh vực khác, trong nghiên cứu này, NCS giới thiệu một phương pháp SKD mạnh mẽ có tên là EEG-SKDNet. Các chi tiết về khung phương án được khai sẽ được trình bày ở các phần tiếp theo.

## **4.2 Kỹ thuật tối ưu hóa mô hình EEG-ER sử dụng SKD và Scaled Weights**

Trong phần này, NCS đề xuất mô hình học sâu EEG-SKDNet hiệu quả cho bài toán EEG\_ER. Trước tiên, NCS giới thiệu ngắn gọn thiết lập bài toán, sau đó trình bày khung EEG-SKDNet được đề xuất, bao gồm chi tiết về phương pháp đề xuất, cơ chế trọng số có điều chỉnh (scaled-weights), và kiến trúc mạng nhẹ cho bài toán EEG\_ER.



### 4.2.1 Thiết lập bài toán

Giả sử có một tập gồm  $N$  mẫu có nhãn, được ký hiệu là  $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^N$ , trong đó,  $X_i$  là tín hiệu EEG thứ  $i$  và  $y_i$  là nhãn tương ứng của nó (tức là trạng thái cảm xúc). Lưu ý rằng  $y_i$  là một vector one-hot có chiều là  $K$  – tổng số lớp.

Bên cạnh đó, tín hiệu  $X_i$  thường rất dài – khoảng 40 đến 100 giây (sau khi đã loại bỏ phần đầu và phần kết của bản ghi). Điều này gây khó khăn khi đưa trực tiếp vào mô hình, do đó thường được chia nhỏ thành các đoạn tín hiệu ngắn hơn gọi là  $x_{ij}$ , trong đó  $x_{ij}$  là đoạn thứ  $j$  của  $X_i$ . Để thiết lập cửa sổ trượt có độ dài  $L = 256$  (tương đương khoảng 2 giây), tức là kích thước của mỗi  $x_{ij}$  là  $L$ , bước trượt là 15, suy ra tần số lấy mẫu là 128Hz, thực hiện biến đổi FFT cho mỗi  $x_{ij}$  thu được, công thức:  $\mathcal{B} = \{(x_{ij}, y_i)\}_{j=1}^n$  là một mini-batch được lấy từ  $\mathcal{D}$  với  $n \ll N$ .

Hàm mất mát cross-entropy giữa nhãn thực  $y$  và phân phối xác suất đầu ra  $p$  được ký hiệu là:  $\text{CE}(y, p)$

$$\text{CE}(y, p) = - \sum_{k=1}^K y_k \log p_k \quad (4.1)$$

Gọi độ đo mất mát Kullback-Leibler (KL) giữa hai phân phối xác suất  $p$  và  $q$  là:

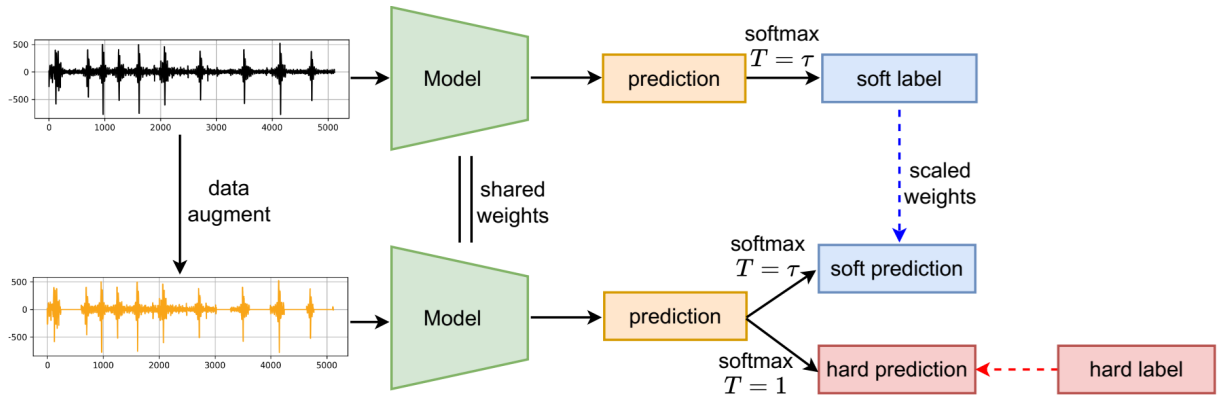
$$\text{KL}(p, q) = \sum_{k=1}^K p_k \log \frac{p_k}{q_k} \quad (4.2)$$

### 4.2.2 Đề xuất kiến trúc EEG-SKDNet

Tổng quan về phương pháp đề xuất của NCS được minh họa trong Hình 4.1. Trong đó: đường gạch đỏ thể hiện hàm mất mát cross-entropy tiêu chuẩn, đường gạch xanh thể hiện hàm mất mát knowledge distillation,  $T$  biểu thị giá trị của hệ số nhiệt độ (temperature value).

Cụ thể, với tín hiệu EEG đầu vào là  $x_{ij}$ , NCS áp dụng các chiến lược tăng cường dữ liệu để tạo ra một tín hiệu tổng hợp mới ký hiệu là  $x'_{ij}$ . Gọi mô hình là  $F_\theta$  với  $\theta$  là tập các trọng số có thể huấn luyện của mô hình  $F$ . Các dự đoán của  $F_\theta$  với hai đầu vào khác nhau là  $P_i$  và  $P'_i$ , tức là:

$$P_i = F_\theta(x_{ij}), \quad P'_i = F_\theta(x'_{ij})$$



**Hình 4.1** Tổng quan về phương pháp được đề xuất

Hàm mất mát cho mô hình EEG-SKDNet bao gồm hai thành phần: hàm mất mát cross-entropy tiêu chuẩn và hàm mất mát distillation. Hàm mất mát cross-entropy tiêu chuẩn được trình bày như sau:

$$\mathcal{L}_{CE} = CE(y_i, \text{softmax}(P_j)) + CE(y_i, \text{softmax}(P'_j)) \quad (4.3)$$

Công thức (4.3) nhằm tính sai lệch giữa nhãn thực  $y_i$  (nhãn cứng) và các dự đoán  $P_j$  và  $P'_j$  từ mô hình  $F_\theta$ . Hàm mất mát distillation được trình bày như sau:

$$\mathcal{L}_{\text{distill}} = w_{\text{scaled}} \cdot KL\left(\text{softmax}\left(\frac{P_j}{\tau}\right), \text{softmax}\left(\frac{P'_j}{\tau}\right)\right) \quad (4.4)$$

Trong đó,  $\tau$  là hệ số nhiệt độ của hàm softmax giúp làm mềm xác suất. Phương trình này nhằm truyền tri thức từ mô hình với đầu vào là tín hiệu EEG gốc sang chính nó với đầu vào là tín hiệu tổng hợp (gọi là chính – student). Dữ liệu từ tín hiệu được đổi chiều nhãn để huấn luyện chính nó.

Tuy nhiên, lượng tri thức được truyền từ mô hình đóng vai trò giáo viên sang mô hình học sinh sẽ được điều tiết chọn lọc thay vì truyền toàn bộ. Quá trình điều tiết này được kiểm soát bởi tham số  $w_{\text{scaled}}$ , tham số này điều chỉnh mức độ ảnh hưởng của mỗi mẫu EEG trong quá trình chưng cất (distillation). Cụ thể,  $w_{\text{scaled}}$  xác định mức độ đóng góp của các xác suất dự đoán liên quan đến từng tín hiệu EEG, hiệu quả trong việc xác định tín hiệu nào nên được giữ lại hoặc loại bỏ trong quá trình distillation. Việc tính toán  $w_{\text{scaled}}$  dựa trên cường độ tín hiệu của từng đầu vào EEG, với các chi tiết thực hiện được trình bày trong phần tiếp theo. Hàm mất mát cuối cùng cho phương pháp được đề xuất của NCS được trình bày như sau:

$$\mathcal{L}_{\text{EEG-SKD}} = \alpha \times \mathcal{L}_{CE} + (1 - \alpha) \times \mathcal{L}_{\text{distill}} \quad (4.5)$$

Trong đó,  $\alpha$  là siêu tham số điều chỉnh mức cân bằng giữa hai hàm mất mát  $\mathcal{L}_{\text{CE}}$  và  $\mathcal{L}_{\text{distill}}$ . NCS sẽ trình bày chi tiết ý tưởng tổng thể và trực giác của phương pháp được đề xuất trong phần dưới đây. Mã giả (pseudo-code) của khung mô hình EEG-SKNet được trình bày trong Thuật toán 4.1.

---

**Thuật toán 4.1: EEG-SKNet**


---

**Input:**  $F_\theta$ : mô hình với trọng số huấn luyện được  $\theta$ ;

$\mathcal{D}$ : tập huấn luyện;

$\tau$ : hệ số nhiệt độ;

$\mathcal{T}$ : tập các phép biến đổi tăng cường dữ liệu;

$E$ : tổng số epoch;

$\alpha$ : siêu tham số như trong Công thức (4.5);

**Output:** Trả về  $F_\theta$  tối ưu;

- 1: Khởi tạo các tham số  $\theta$  cho mạng  $F$ ;
  - 2: **for**  $epoch = 1$  **to**  $E$  **do**
  - 3:     **for** mỗi  $batch$  trong  $tập\ huấn\ luyện\ \mathcal{D}$  **do**
  - 4:         Lấy một batch  $(x, y)$  từ  $\mathcal{D}$ ;
  - 5:         Tính  $w_{\text{scaled}}$ ;
  - 6:          $x' = \mathcal{T}(x)$ ;
  - 7:          $p = F_\theta(x)$ ;    $p' = F_\theta(x')$ ;
  - 8:         Tính  $\mathcal{L}_{\text{CE}}$  theo Công thức (4.3);
  - 9:         Tính  $\mathcal{L}_{\text{distill}}$  theo Công thức (4.4);
  - 10:        Tính  $\mathcal{L}_{\text{EEG-SKD}}$  theo Công thức (4.5);
  - 11:        Cập nhật các tham số  $\theta$  bằng gradient của  $\mathcal{L}_{\text{EEG-SKD}}$ ;
  - 12:     **end for**
  - 13:     **if**  $F_\theta$  hội tụ **then**
  - 14:         **break**;   // Dừng sớm
  - 15:     **end if**
  - 16: **end for**
  - 17: **return**  $F_\theta$
- 

Thuật toán 4.1 (EEG-SKNet) huấn luyện mô hình theo cơ chế *self-knowledge distillation*, giúp mô hình duy trì tính nhất quán dự đoán giữa hai quan sát tương đương của cùng một mẫu EEG để tăng khả năng khái quát hoá và giảm quá khớp. Dữ liệu vào gồm mô hình  $F_\theta$  cần tối ưu, tập huấn luyện  $\mathcal{D} = \{(x, y)\}$ , tập các phép tăng cường  $\mathcal{T}$  bảo toàn nhãn, cùng các siêu tham số  $\tau$  (nhiệt độ làm mềm phân phối),  $\alpha$  (hệ số cân bằng giữa mất mát nhãn cứng và mất mát nhất quán) và số epoch  $E$ . Đầu ra là mô hình  $F_\theta$  sau huấn luyện, tối ưu theo hàm mất mát tổng  $\mathcal{L}_{\text{EEG-SKD}} = \mathcal{L}_{\text{CE}} + \alpha \mathcal{L}_{\text{distill}}$ .

Quy trình thực hiện như sau: ở mỗi batch, từ mẫu  $x$  sinh một quan sát tăng cường  $x' = \mathcal{T}(x)$ ; mô hình suy luận hai lần thu được  $p = F_{\theta}(x)$  và  $p' = F_{\theta}(x')$ . Mất mát nhãn cứng  $\mathcal{L}_{\text{CE}}$  được tính giữa dự đoán và nhãn  $y$ ; song song, mất mát nhất quán  $\mathcal{L}_{\text{distill}}$  đo khoảng cách giữa hai dự đoán đã làm mềm bởi  $\tau$  để ép mô hình ổn định trước các biến thiên hợp lệ của tín hiệu. Hai thành phần được kết hợp bằng  $\alpha$ , lan truyền ngược cập nhật  $\theta$ , và lặp lại đến khi đạt  $E$  epoch hoặc thỏa điều kiện dừng sớm. Cách thiết kế này tạo “nhân mềm” nội sinh mà không cần mô hình giáo viên, giúp mô hình học đặc trưng EEG bền vững hơn với nhiễu và biến thiên, từ đó nâng cao hiệu năng tổng quát.

Khác với mô hình học một chiều giữa giáo viên và học sinh trong distillation truyền thống [42], trong phương pháp của NCS, NCS giới thiệu SKD tức là chỉ sử dụng một mô hình học sinh duy nhất, và mô hình này tự học, tự chia sẻ, và tự học lại tri thức được triết xuất từ chính nó trong quá trình huấn luyện. Cụ thể, mô hình học sinh trong khung NCS đề xuất sẽ nhận hai phiên bản khác nhau của tín hiệu EEG làm đầu vào. Sau đó, các dự đoán từ mô hình với hai đầu vào này sẽ được sử dụng để tính toán các hàm mất mát, bao gồm cả hàm mất mát distillation. Bằng cách này, phương pháp đề xuất đã được chứng minh là: (i) tăng khả năng khái quát hóa (generalization) của mô hình, và (ii) tránh hiện tượng quá khớp (overfitting).

Để tăng khả năng khái quát hóa, cơ chế scaled-weight cung cấp cho mô hình học sinh các đầu vào đa dạng hơn, từ đó giúp cải thiện hiệu năng tổng thể của mô hình so với mạng chỉ sử dụng tín hiệu EEG gốc. Cơ chế scaled-weights là một kỹ thuật tăng cường dữ liệu hiệu quả đối với tín hiệu EEG. Nó giúp tăng số lượng mẫu huấn luyện mà không cần gán nhãn. Do đó, khi có thêm nhiều mẫu huấn luyện hơn, hiện tượng quá khớp sẽ diễn ra chậm hơn.

### 4.2.3 Trọng số có điều chỉnh (Scaled-Weights)

Trong các tác vụ nhận diện cảm xúc dựa trên tín hiệu EEG, việc học và trích xuất đặc trưng từ dữ liệu nhiễu, có cấu trúc phức tạp theo thời gian là một thách thức lớn. Tín hiệu EEG bao gồm nhiều thành phần dao động với các tần số khác nhau, thường được chia thành năm dải tần số chính [51]: Delta (0.1 - 4 Hz), Theta (4 - 8 Hz), Alpha (8 - 13 Hz), Beta (13 - 30 Hz), và Gamma (30 - 170 Hz) (xem Hình 1.4). Mỗi dải tần phản ánh một trạng thái sinh lý và nhận thức khác nhau của não bộ.

Do các đặc điểm thời gian – không gian đặc thù của tín hiệu EEG, các chiến lược tăng cường dữ liệu cần được thiết kế phù hợp cho tín hiệu này, khác biệt đáng kể so với dữ liệu hình ảnh hoặc giọng nói. Trong nghiên cứu này, NCS áp dụng khung tăng cường dữ liệu được đề xuất trong [64] để sinh ra các mẫu EEG tổng hợp, từ đó nâng cao tính đa dạng và khả năng kháng nhiễu của tập dữ liệu huấn luyện:

$$x'_{ij} = \mathcal{T}(x_{ij}) \quad (4.6)$$

Trong đó  $\mathcal{T}$  là tập các phép biến đổi tăng cường dữ liệu. Cần lưu ý rằng cả dữ liệu EEG gốc và dữ liệu đã được tăng cường đều được đưa vào mô hình học sinh  $F$ . Tuy nhiên, chỉ các đầu ra dự đoán từ dữ liệu gốc mới được sử dụng như các nhãn mềm trong khuôn khổ học chung cất tri thức (knowledge distillation). Để cải thiện chất lượng của các nhãn mềm này, NCS đề xuất tính toán tham số trọng số  $w_{\text{scaled}}$  một cách linh hoạt theo từng batch. Việc tính toán  $w_{\text{scaled}}$  bao gồm hai bước chính:

1. Tính phổ công suất trung bình  $P_{\text{avg}}$  cho mỗi cửa sổ thời gian, biểu thị mức năng lượng chứa trong tín hiệu ở các thời điểm khác nhau.
2. Tính  $w_{\text{scaled}}$  dựa trên  $P_{\text{avg}}$ .

Đối với bước đầu tiên, cần tính phổ công suất  $P(\cdot)$  từ bình phương biên độ FFT như sau:

$$P(x_{ij}) = (\text{FFT}(x_{ij}))^2 \quad (4.7)$$

trong đó FFT là phép biến đổi Fourier nhanh. Để giảm nhiễu và tăng tính đại diện, phổ công suất trung bình được tính theo:

$$P_{\text{avg}}(x_{ij}) = \frac{1}{L} \sum_{k=1}^L P(x_{ij})[k] \quad (4.8)$$

Ở bước thứ hai, trọng số đã chuẩn hóa  $w_{ij}^{\text{scaled}}$  được tính, phản ánh tầm quan trọng của từng đoạn tín hiệu EEG như sau:

$$w_{ij}^{\text{scaled}} = \frac{P_{\text{avg}}(x_{ij}) - \min(P_{\text{avg}}(x_{ij}))}{\max(P_{\text{avg}}(x_{ij})) - \min(P_{\text{avg}}(x_{ij}))} \quad (4.9)$$

trong đó  $\min(\cdot)$  và  $\max(\cdot)$  lần lượt là toán tử lấy giá trị nhỏ nhất và lớn nhất trong một batch. Việc tích hợp cơ chế Scaled-Weights cho phép mô hình tập trung vào các đoạn tín hiệu EEG tại những thời điểm mang tín hiệu cảm xúc mạnh. Như được định nghĩa trong công thức (4.9), trọng số  $w^{\text{scaled}}$  bị ràng buộc trong khoảng  $[0, 1]$ , trong đó các tín hiệu EEG có biên độ lớn hơn – thường mang thông tin cảm xúc nổi bật hơn – sẽ được gán giá trị gần 1, trong khi các tín hiệu ít thông tin nhận giá trị nhỏ hơn. Cơ chế này giúp tận dụng hiệu quả các vùng tín hiệu cường độ cao mà không ảnh hưởng đến độ ổn định của quá trình học từ các nhãn mềm. Việc kết hợp SKD và Scaled-Weights giúp mô hình học sâu hơn vào các thông tin có liên quan, từ đó tăng độ chính xác và khả năng khái quát của mô hình.

#### 4.2.4 Chứng minh toán học về tác động của hệ số nhiệt và trọng số có điều chỉnh lên phân phối xác suất đầu ra trong EEG-SKDNet

NCS đã áp dụng các kỹ thuật chuẩn hóa xác suất như hệ số nhiệt (temperature scaling) và trọng số có điều chỉnh (scaled weights) trong mô hình EEG-SKDNet nhằm

điều chỉnh phân phối xác suất đầu ra để làm giảm hiện tượng quá khớp vào các lớp cụ thể (overconfident) và tăng tính tổng quát của mô hình. Tuy nhiên, để khẳng định chắc chắn và rõ ràng hơn về lợi ích của kỹ thuật này, NCS tiến hành chứng minh toán học nhằm chỉ ra tác động tổng thể của các kỹ thuật này đối với phương sai của phân phối xác suất đầu ra.

Giả sử đầu ra của mô hình EEG-SKDNet được biểu diễn dưới dạng vector logit với  $K$  lớp như sau:

$$\mathbf{z}(x) = [\bar{z}_1(x), \bar{z}_2(x), \dots, \bar{z}_K(x)] \quad (4.10)$$

Phân phối xác suất đầu ra sau khi áp dụng đồng thời temperature scaling ( $\tau$ ) và scaled weights ( $w_k$ ) được định nghĩa như sau:

$$p_{w,\tau}^k(x) = \frac{e^{w_k \cdot \bar{z}_k(x)/\tau}}{\sum_{j=1}^K e^{w_j \cdot \bar{z}_j(x)/\tau}} \quad (4.11)$$

trong đó  $\tau \geq 1$  là hệ số nhiệt độ và  $w_k \in (0, 1]$  là trọng số co giãn điều chỉnh mức ảnh hưởng của từng logit. Phương sai của phân phối xác suất này được biểu diễn theo giá trị kỳ vọng như sau:

$$\text{Var}(p_{w,\tau}(x)) = \mathbb{E}(p_{w,\tau}(x)^2) - (\mathbb{E}(p_{w,\tau}(x)))^2 = \mathbb{E}(p_{w,\tau}(x)^2) - \left(\frac{1}{K}\right)^2 \quad (4.12)$$

Vì softmax tạo ra phân phối chuẩn hóa, trung bình xác suất mỗi lớp luôn là  $1/K$ , biến số thay đổi duy nhất ảnh hưởng đến phương sai là  $\mathbb{E}(p_{w,\tau}(x)^2)$ , vậy việc chứng minh phân phối mượt hơn tương đương với chứng minh:

$$\mathbb{E}(p_{w,\tau}(x)^2) \text{ giảm khi } \tau > 1, \quad w_k < 1 \quad (4.13)$$

Xét hàm mục tiêu sau để thuận tiện phân tích, lưu ý rằng nhân thêm hệ số  $K$  giúp loại bỏ mẫu số để thuận tiện cho việc tính đạo hàm mà không làm thay đổi bản chất của phép chứng minh. Ta vẫn giữ được chiều biến thiên của hàm mục tiêu, vì  $K$  là hằng số dương không ảnh hưởng đến dấu đạo hàm:

$$f(\mathbf{z}, \mathbf{w}, \tau) = K \cdot \mathbb{E}(p_{w,\tau}(x)^2) = K \sum_{k=1}^K \left( \frac{e^{w_k \bar{z}_k / \tau}}{\sum_{j=1}^K e^{w_j \bar{z}_j / \tau}} \right)^2 \quad (4.14)$$

Đặt  $u_k = \frac{w_k \bar{z}_k}{\tau}$ , khi đó hàm mục tiêu viết thành:

$$f(\mathbf{z}, \mathbf{w}, \tau) = \sum_{k=1}^K \sigma(u_k)^2 \quad \text{với} \quad \sigma(u_k) = \frac{e^{u_k}}{\sum_{j=1}^K e^{u_j}} \quad (4.15)$$

Tính đạo hàm theo  $\tau$ :

$$\frac{\partial f(\mathbf{z}, \mathbf{w}, \tau)}{\partial \tau} = -\frac{2}{\tau^2} \sum_{k=1}^K w_k \bar{z}_k \cdot \sigma(u_k)^2 (1 - \sigma(u_k)) \leq 0, \quad \forall \tau \geq 1 \quad (4.16)$$

Dấu âm cho thấy rằng hàm  $f(\mathbf{z}, \mathbf{w}, \tau)$  giảm khi nhiệt độ  $\tau$  tăng, tức là tăng  $\tau$  sẽ làm giảm phương sai và làm mượt hơn phân phối xác suất.

Tiếp theo, xem xét ảnh hưởng khi thay đổi trọng số  $w_k^*$ . Tính đạo hàm theo  $w_k^*$ :

$$\frac{\partial f(z, w, \tau)}{\partial w_k^*} = 2 \cdot \sigma(u_k)^2 (1 - \sigma(u_k)) \cdot \frac{\bar{z}_k}{\tau} \quad \text{với } u_k = \frac{w_k^* \bar{z}_k}{\tau} \quad (4.17)$$

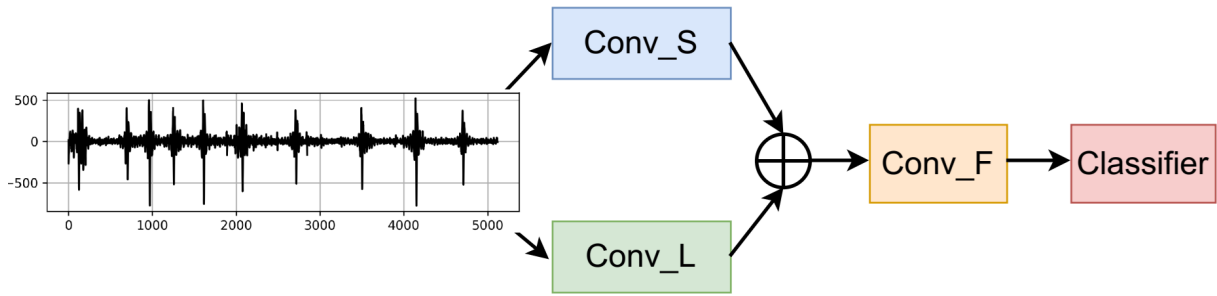
Vì  $\sigma(u_k) \in (0, 1)$  và  $\tau > 0, \bar{z}_k > 0$  nên đạo hàm luôn dương. Kết quả này cho thấy rằng khi giảm trọng số  $w_k^*$  (tức  $w_k^* < 1$ ), giá trị của hàm  $f(z, w, \tau)$  giảm theo. Điều này đồng nghĩa với việc làm giảm phương sai, từ đó phân phối xác suất đầu ra trở nên mượt hơn.

Thông qua các bước lập luận trên, NCS đã chứng minh được rằng việc áp dụng hiệu chỉnh nhiệt độ  $\tau > 1$  và trọng số có điều chỉnh  $w_k^* < 1$  đều góp phần làm giảm phương sai của phân phối xác suất đầu ra. Điều này đồng nghĩa với việc phân phối xác suất trở nên mượt hơn (smooth), phản ánh đúng hơn độ không chắc chắn trong dự đoán, đồng thời hỗ trợ tốt hơn cho quá trình truyền tri thức giữa các nhánh trong mô hình EEG-SKDNet. Chứng minh trên đóng vai trò là một cơ sở toán học quan trọng nhằm củng cố cho thiết kế mô hình và các thành phần được đề xuất trong luận án này.

#### 4.2.5 Thiết lập kiến trúc mạng

Để chứng minh tính hiệu quả của phương pháp EEG-SKDNet, trong phần này, NCS đề xuất một mạng như được minh họa trong Hình 4.2. Trong đó, Conv\_S biểu thị bộ trích xuất đặc trưng ngắn hạn, Conv\_L là bộ trích xuất đặc trưng dài hạn, Conv\_F là bộ trích xuất cuối cùng, Dấu  $\oplus$  là phép nối đặc trưng (concatenation).

Cụ thể, mạng đề xuất gồm hai nhánh xử lý song song. Nhánh thứ nhất được thiết kế để khai thác đặc trưng gần theo thời gian—tức các biến thiên cục bộ diễn ra trong khoảng ngắn của cửa sổ tín hiệu (ví dụ biên độ/đột biến, dao động nhanh); nhánh này sử dụng hạt nhân tích chập nhỏ (độ giãn thấp) để có độ phủ thời gian ngắn. Nhánh thứ hai tập trung vào đặc trưng xa theo thời gian—các phụ thuộc dài và nhịp điệu chậm trong cùng cửa sổ (xu hướng dài, điều biến biên độ), được học bằng hạt nhân lớn hơn hoặc độ giãn cao hơn để mở rộng độ phủ thời gian. Hai bản đồ đặc trưng từ hai nhánh sau đó được nối và đưa vào khối trích xuất cuối cùng cùng bộ phân loại.



**Hình 4.2** Mạng nhẹ được sử dụng trong EEG-SKDNet

Thông tin chi tiết về kiến trúc mạng được sử dụng trong EEG-SKDNet được trình bày trong Bảng 4.1. Ở đây,  $k$  là kích thước kernel,  $s$  là bước nhảy (stride),  $d$  là hệ số giãn (dilation rate). Trong bảng này, Khối Conv\_S là bộ trích xuất đặc trưng ngắn hạn; Khối Conv\_L là bộ trích xuất đặc trưng dài hạn; Khối Conv\_F là khối trích xuất cuối cùng, nhận đầu vào từ cả Conv\_S và Conv\_L thông qua phép nối (concatenate). Mỗi khối tích chập bao gồm một lớp convolution, tiếp theo là một lớp chuẩn hóa theo batch (batch normalization) và một lớp Squeeze-and-Excitation (SE). Lưu ý rằng lớp SE [45] được áp dụng cho cả Conv\_S và Conv\_L để làm nổi bật các đặc trưng quan trọng từ đầu vào.

**Bảng 4.1** Chi tiết kiến trúc mạng nhẹ trong phương pháp đề xuất

| Tầng (Layer)        | Thông số (Specification)                      | Đầu vào (Input) |
|---------------------|-----------------------------------------------|-----------------|
| Input               | -                                             | -               |
| Khối Conv_S         | num_filters = 32; $k = 3$ ; $s = 1$           | Input           |
| Khối Conv_L         | num_filters = 32; $k = 5$ ; $s = 1$ ; $d = 2$ | Input           |
| Nối (Concat)        | -                                             | Conv_S; Conv_L  |
| Khối Conv_F         | num_filters = 64; $k = 3$ ; $s = 1$           | Concat          |
| Flatten             | -                                             | Conv_F          |
| FC (kết nối đầy đủ) | Unit = 512                                    | Flatten         |
| Phân loại           | Unit = num_classes                            | FC              |

Trong khối Conv\_F, NCS thêm hai lớp bao gồm hàm kích hoạt ReLU và pooling trung bình 1D, sau đó là lớp làm phẳng (flatten). Đối với khối Fully Connected (FC), NCS sử dụng dropout với tỷ lệ 0.2 để tránh overfitting trong quá trình huấn luyện. Trong khối phân loại (classifier), NCS sử dụng hàm softmax để tạo ra xác suất cho từng lớp dựa trên đầu ra từ lớp FC.



### 4.3 Thực nghiệm

Trong phần này, trước tiên, NCS lựa chọn một số bộ dữ liệu tiêu chuẩn dùng cho đánh giá hiệu quả của mô hình học sâu đề xuất cho bài toán EEG\_ER. Tiếp đó NCS trình bày chi tiết về cách triển khai khung phương pháp được đề xuất. Cuối cùng, NCS thực hiện các thí nghiệm với các kịch bản khác nhau để đánh giá hiệu quả của EEG-SKDNet.

#### 4.3.1 Bộ dữ liệu

Để đánh giá hiệu suất của EEG-SKDNet, NCS tiến hành các thí nghiệm trên ba bộ dữ liệu chuẩn thường được sử dụng trong nhận diện cảm xúc dựa trên EEG, bao gồm DEAP và DREAMER.

**DEAP:** Bộ dữ liệu DEAP [60] được mô tả trong mục 2.4.1.1 được tiếp tục sử dụng, NCS thử nghiệm trên hai phiên bản, bao gồm DEAP(32) và DEAP(14):

- **DEAP(32):** Phiên bản đầy đủ sử dụng toàn bộ 32 điện cực EEG, cho phép khai thác tối đa thông tin.
- **DEAP(14):** Phiên bản rút gọn từ DEAP(32) bằng cách chọn 14 điện cực, tương ứng với sơ đồ điện cực của thiết bị Emotiv ("hệ thống giống Emotiv 10-20") – một thiết bị EEG thương mại phổ biến. Phiên bản này mô phỏng các tình huống triển khai thực tế trên thiết bị di động hoặc thiết bị đeo với số lượng điện cực hạn chế.

**DREAMER:** Bộ dữ liệu DREAMER [56] chứa các tín hiệu EEG từ 14 điện cực, thu thập trong khi người tham gia xem các video kích thích cảm xúc. Các nhãn cảm xúc bao gồm: Valence, Arousal và Dominance, được đánh giá theo thang điểm từ 1 đến 5. Quy trình xử lý tín hiệu EEG bao gồm áp dụng FFT để trích xuất đặc trưng và chuẩn hóa tín hiệu nhằm đảm bảo tương thích với các mô hình.

#### 4.3.2 Chi tiết triển khai

Trong các thí nghiệm, mỗi mô hình được huấn luyện và kiểm tra ba lần trên từng bộ dữ liệu bằng cách sử dụng cross-validation 5-fold ( $K=5$ ) để chia bộ dữ liệu thành các tập huấn luyện và kiểm thử khác nhau. Các kết quả được báo cáo bao gồm độ chính xác cao nhất (Accuracy - Acc) và chỉ số F1. Tất cả kết quả được trình bày dưới dạng giá trị trung bình  $\pm$  độ lệch chuẩn ( $\pm$  std), nhằm đánh giá tính ổn định của mô hình.

EEG-SKDNet cũng được so sánh với nhiều phương pháp KD khác nhau, bao gồm: KD dựa trên đầu ra (Response-Based KD); KD dựa trên đặc trưng (Feature-Based KD); Progressive Self-Knowledge Distillation (PS-KD); Multi-Teacher Knowledge Distillation (MT-KD). Tất cả các thí nghiệm được thực hiện trên Google Colab Pro với GPU NVIDIA Tesla T4. Quá trình huấn luyện tuân theo các thiết lập cụ thể như sau:

Batch size: 256; Learning rate khởi đầu: 0.01 (suy giảm theo hàm cosine); Tối ưu hóa bằng Adam; Tổng số epoch tối đa: 200; Áp dụng early stopping sau 10 epoch; Tham số nhiệt độ  $\tau$ : 10; Trọng số  $\alpha$ : 0.7

### 4.3.3 So sánh và đánh giá với phương pháp cơ sở

Để kiểm chứng hiệu quả của mô hình được đề xuất, NCS so sánh EEG-SKDNet với phương pháp cơ sở (huấn luyện từ đầu với cùng kiến trúc mạng được mô tả trong Phần 4.2.5), cùng với một số phương pháp KD và SKD khác nhau.

Bảng 4.2 liệt kê hiệu quả hoạt động của các kịch bản KD bao gồm phương pháp cơ sở, KD dựa trên đầu ra (Response-Based KD), KD dựa trên đặc trưng (Feature-Based KD), Progressive Self-Knowledge Distillation (PS-KD), Multi-Teacher Knowledge Distillation (MT-KD), và phương pháp được đề xuất với các cấu hình khác nhau. Tất cả kết quả được báo cáo trên tập DEAP(32), với các siêu tham số được giữ nguyên.

Như thể hiện trong Bảng 4.2, KD dựa trên đầu ra (Response-based KD) chỉ cải thiện nhẹ phương sai nhưng suy giảm nhẹ về hiệu suất trung bình so với phương pháp cơ sở. KD dựa trên đặc trưng (Feature-based KD) cho hiệu suất thấp đáng kể (90.87%), cho thấy rằng việc truyền trực tiếp đặc trưng có thể không tối ưu cho biểu diễn EEG. PS-KD và MT-KD cải thiện nhẹ so với mô hình cơ sở, với PS-KD đạt 94.02%, cho thấy các phương pháp KD tiên tiến hơn có thể mang lại lợi ích nhiều hơn so với huấn luyện từ đầu.

**Bảng 4.2** So sánh giữa các kịch bản KD khác nhau

| Method                      | Acc                                | F1-score                           |
|-----------------------------|------------------------------------|------------------------------------|
| Baseline                    | 93.44 $\pm$ 0.78                   | 93.43 $\pm$ 0.65                   |
| Response-based KD           | 93.15 $\pm$ 1.14                   | 93.18 $\pm$ 1.17                   |
| Feature-based KD            | 90.87 $\pm$ 0.98                   | 90.87 $\pm$ 0.93                   |
| PS-KD                       | 94.02 $\pm$ 0.77                   | 94.02 $\pm$ 0.79                   |
| MT-KD                       | 93.99 $\pm$ 1.28                   | 93.99 $\pm$ 1.30                   |
| EEG-SKDNet w/o $w_{scaled}$ | 95.83 $\pm$ 0.53                   | 95.87 $\pm$ 0.45                   |
| <b>EEG-SKDNet</b>           | <b>96.48 <math>\pm</math> 0.35</b> | <b>96.48 <math>\pm</math> 0.36</b> |

EEG-SKDNet không dùng  $w_{scaled}$  (tức không áp dụng trọng số điều chỉnh) đã vượt qua tất cả các phương pháp KD, đạt 95.88% accuracy và 95.87% F1-score, khẳng định hiệu quả của kiến trúc mạng được đề xuất. Cuối cùng, EEG-SKDNet đầy đủ, với cơ chế trọng số có điều chỉnh, đạt độ chính xác 96.48% và F1-score 96.48%, với độ lệch chuẩn

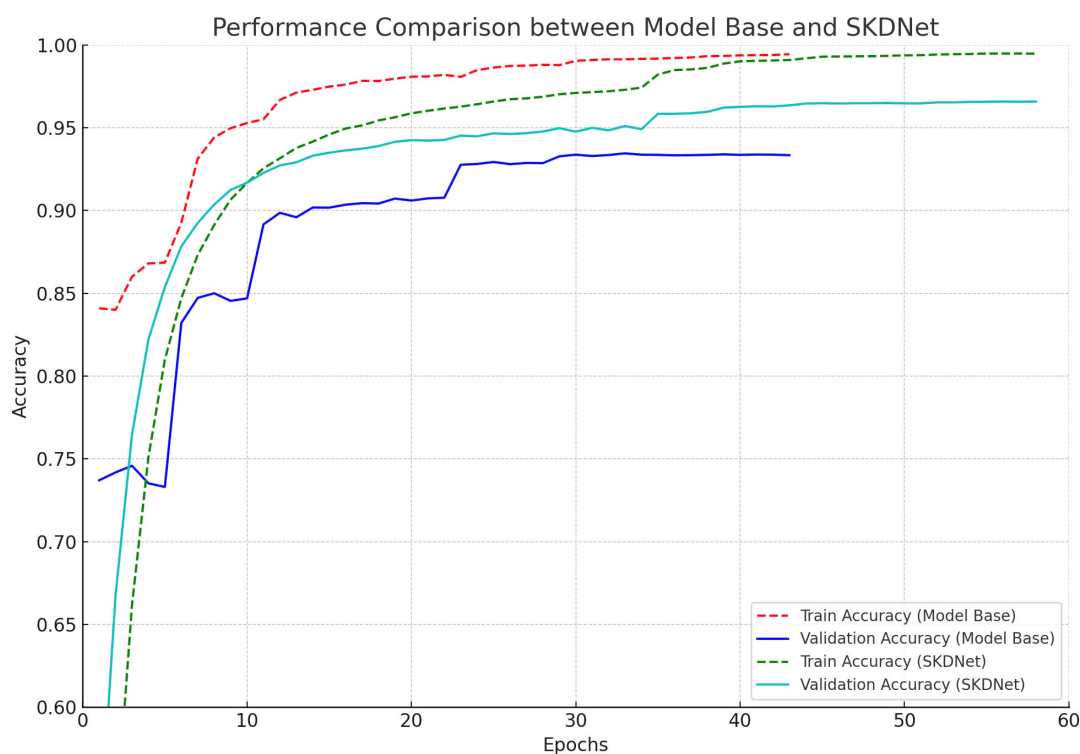
thấp nhất ( $\pm 0.35$  và  $\pm 0.36$ ). Điều này không chỉ thể hiện hiệu suất phân loại vượt trội mà còn chứng minh khả năng khái quát hóa và độ tin cậy của mô hình.

EEG\_SKDNet đạt được kết quả như vậy nhờ ba yếu tố bổ trợ nhau. Thứ nhất, khung self-KD ràng buộc tính nhất quán giữa hai quan sát tương đương của cùng một mẫu EEG ( $x$  và  $x' = \mathcal{T}(x)$ ), kết hợp làm mềm dự đoán bằng nhiệt độ  $\tau$ , giúp mô hình học đặc trưng bền vững trước nhiễu, lệch pha và biến đổi biên độ; Baseline không khai thác điều này, còn PS-KD/MT-KD không dùng soft targets làm mềm bởi  $\tau$  để truyền thông tin quan hệ giữa các lớp, nên hiệu lực ràng buộc yếu hơn. Thứ hai, do không dùng mạng giáo viên, self-KD tránh được sai lệch thầy-trò thường gặp ở KD dựa response/feature (khác kiến trúc hoặc khác miền đặc trưng), nhờ đó tối ưu ổn định hơn trên đặc trưng phổ EEG. Thứ ba, việc trọng số  $w_{\text{scaled}}$  tăng dần cho nhánh nhất quán giúp  $\mathcal{L}_{\text{CE}}$  dẫn dắt giai đoạn đầu, rồi SKD củng cố ổn định về sau; bằng chứng là biến thể “w/o  $w_{\text{scaled}}$ ” kém hơn (95.83% so với 96.48%).

Hình 4.3 thể hiện rằng phương pháp EEG-SKDNet có thể cải thiện khả năng khái quát hóa và giảm hiện tượng overfitting thông qua việc phân tích sự hội tụ của mô hình EEG-SKDNet và mô hình cơ sở. Cụ thể: Phương pháp cơ sở hội tụ sau 23 epoch; EEG-SKDNet hội tụ sau 35 epoch. Mặc dù EEG-SKDNet chia sẻ cùng kiến trúc với mô hình cơ sở, nhưng được huấn luyện trên tập dữ liệu bao gồm cả dữ liệu gốc và dữ liệu tổng hợp.

Cách tiếp cận này giúp tăng cường khả năng khái quát hóa trong khi làm chậm quá trình hội tụ, giúp mô hình đạt hiệu suất cao hơn. Do đó, hiện tượng overfitting được giảm thiểu đáng kể, thể hiện qua sự thu hẹp khoảng cách giữa đường huấn luyện và đường kiểm thử so với mô hình cơ sở. Kết quả này làm nổi bật hiệu quả của EEG-SKDNet trong việc cải thiện khả năng khái quát hóa và giảm overfitting.

Bảng 4.3 mô tả chi tiết đánh giá toàn diện về EEG-SKDNet trên ba thành phần cảm xúc: Valence, Arousal và Dominance trong EEG-SKDNet được so sánh với mô hình cơ sở để đánh giá hiệu quả của mô hình trên từng thành phần cụ thể. Nhìn chung, EEG-SKDNet liên tục vượt trội so với mô hình cơ sở, với mức cải thiện từ 0.43% đến 1.67% trên cả ba tập dữ liệu. Ngoài ra, EEG-SKDNet cho thấy độ ổn định cao hơn với sai số chuẩn thấp hơn rõ rệt so với mô hình cơ sở. Cụ thể, sai số chuẩn của EEG-SKDNet khoảng  $\pm 0.3$ , trong khi mô hình cơ sở có sai số lớn hơn, lên tới  $\pm 0.8$ , và trong một số trường hợp đạt tới  $\pm 1.02$ . Những phát hiện này cho thấy rõ tính ổn định và độ tin cậy của EEG-SKDNet.



**Hình 4.3** Minh họa sự hội tụ của EEG-SKDNet và phương pháp cơ sở trên tập dữ liệu DEAP(32).

**Bảng 4.3** So sánh độ chính xác top-1 giữa mô hình cơ sở và EEG-SKDNet trên các thành phần cảm xúc trong ba bộ dữ liệu: DEAP(32), DEAP(14), và DREAMER.

| Component | Method     | DEAP (32)           | DEAP (14)           | DREAMER             |
|-----------|------------|---------------------|---------------------|---------------------|
| Valence   | Baseline   | 95.92 ± 0.81        | 93.29 ± 0.85        | 91.16 ± 0.54        |
|           | EEG-SKDNet | <b>96.67 ± 0.34</b> | <b>93.72 ± 0.27</b> | <b>92.41 ± 0.32</b> |
| Arousal   | Baseline   | 95.56 ± 0.78        | 91.03 ± 0.92        | 91.21 ± 0.44        |
|           | EEG-SKDNet | <b>96.32 ± 0.36</b> | <b>94.00 ± 0.36</b> | <b>91.90 ± 0.41</b> |
| Dominance | Baseline   | 95.76 ± 0.69        | 93.21 ± 1.01        | 90.26 ± 0.55        |
|           | EEG-SKDNet | <b>96.45 ± 0.36</b> | <b>93.79 ± 0.28</b> | <b>92.34 ± 0.31</b> |
| All       | Baseline   | 95.75 ± 0.76        | 93.26 ± 0.89        | 91.22 ± 0.51        |
|           | EEG-SKDNet | <b>96.48 ± 0.35</b> | <b>93.84 ± 0.27</b> | <b>92.56 ± 0.31</b> |

#### 4.3.4 So sánh với các phương pháp tiên tiến (SOTA)

Phần này trình bày phân tích so sánh giữa EEG-SKDNet với các phương pháp tiên tiến gần đây trên các bộ dữ liệu chuẩn, bao gồm DEAP(32) và DREAMER. Ngoài độ chính xác Top-1, phép so sánh còn xét đến nhiều yếu tố khác như: Kích thước đầu

vào; Số lượng lớp đầu ra (#classes); Tổng số tham số (#params); Độ phức tạp tính toán (MFLOPs); Thời gian suy luận (tính bằng mili giây). Để đảm bảo so sánh công bằng, toàn bộ thời gian suy luận đều được đo trên cùng một nền tảng phần cứng gồm CPU 2.2 GHz, RAM 16 GB, và hệ điều hành Linux 64-bit (Ubuntu).

**Bảng 4.4** So sánh hiệu suất giữa các phương pháp tiên tiến trên các tập dữ liệu phổ biến.

| Method                          | Acc          | Input shape   | #classes | #params (M) | MFLOPs | Inference time (ms) |
|---------------------------------|--------------|---------------|----------|-------------|--------|---------------------|
| <b>DEAP (32)</b>                |              |               |          |             |        |                     |
| MT-KD [118]                     | 65.00        | (128, 64)     | 2        | –           | –      | –                   |
| SDC-GAT [112]                   | 88.17        | (32×6)        | 2        | –           | –      | –                   |
| M1 [3]                          | 99.22        | (160, 1)      | 2        | 9.49        | 27.44  | 35.66               |
| CNN & LSTM [128]                | 99.22        | (256, 1)      | 2        | –           | –      | 475.14              |
| TC_Net [82]                     | 98.82        | (48, 128, 32) | 2        | 1.45        | 7.22   | 122.97              |
| ResNet34 [116]                  | 83.90        | (32, 28)      | 2        | 7.29        | 17.55  | 75.04               |
| MSDA-SFE [94]                   | 70.10        | (48, 128, 14) | 2        | 17.02       | 242.02 | 3445.77             |
| TDMNN [19]                      | 99.00        | (9,8,32)      | 2        | 15.25       | –      | –                   |
| DSE-Mixer [22]                  | 95.19        | (9,8,32)      | 2        | 0.32        | 90.00  | –                   |
| STILN [21]                      | 68.31        | –             | 2        | –           | 67.27  | –                   |
| <b>EEG-SKDNet (ours)</b>        | <b>99.18</b> | (160, 1)      | 2        | 2.63        | 9.33   | 22.12               |
| <b>Rhythm-specific CNN [90]</b> | 95.00        | (1280, 32)    | 9        | 0.19        | 153.09 | 40.524              |
| MTL_MSRL [110]                  | 71.92        | (32, 32)      | 9        | 1.00        | 350.00 | –                   |
| <b>EEG-SKDNet (ours)</b>        | <b>96.48</b> | (160, 1)      | 9        | 2.64        | 9.34   | 22.46               |
| <b>DREAMER</b>                  |              |               |          |             |        |                     |
| Rhythm-specific CNN [90]        | 94.45        | (1280, 14)    | 2        | 0.18        | 132.45 | 1280.3              |
| TC_Net [82]                     | 98.67        | (48, 128, 14) | 2        | 1.44        | 7.26   | 122.07              |
| TDMNN [19]                      | 95.00        | (9,8,14)      | 2        | 15.25       | 31.92  | –                   |
| <b>EEG-SKDNet (ours)</b>        | <b>98.76</b> | (105, 1)      | 2        | 1.72        | 6.10   | 20.62               |

Bảng 4.4 trình bày sự so sánh toàn diện giữa EEG-SKDNet với các phương pháp tiên tiến gần đây trên các bộ dữ liệu DEAP(32) và DREAMER. Cột “#classes” biểu thị số lượng lớp được sử dụng trong mỗi phương pháp. Cột “#params” biểu thị số lượng tham số trong mỗi phương pháp (bộ nhớ lưu trọng số), đại diện cho độ phức tạp về không gian. MFLOPs (số phép tính dấu chấm động mỗi giây – triệu FLOPs) biểu thị độ phức tạp tính toán, tức là độ phức tạp của từng mô hình (gắn trực tiếp với độ trễ/công suất). Tất cả kết quả trong cột “thời gian suy luận” đều được thực hiện trên cùng một máy tính.

Việc đánh giá không chỉ dựa trên độ chính xác phân loại mà còn xét đến các chỉ số hiệu quả của mô hình như số tham số, độ phức tạp tính toán và thời gian suy luận – cung

cấp cái nhìn toàn diện về mức độ phù hợp của từng mô hình trong ứng dụng thực tế.

Trên bộ dữ liệu **DEAP** (2 lớp – nhị phân): EEG-SKDNet đạt độ chính xác top-1 là **99.18%**, so sánh với các mô hình hiệu suất cao như MI (99.22%) và CNN & LSTM (99.35%). Tuy nhiên, khác với các mô hình kia, EEG-SKDNet duy trì độ phức tạp tính toán thấp hơn đáng kể (9.33 MFLOPs) và thời gian suy luận nhanh hơn (22.12 ms). Ngược lại, CNN & LSTM mặc dù có độ chính xác cao, nhưng yêu cầu thời gian suy luận hơn 20 lần (475.14 ms) và độ phức tạp MFLOP cao hơn, khiến mô hình khó áp dụng trong các môi trường nhạy cảm với độ trễ.

Trên nhiệm vụ phân loại nhiều lớp (9 lớp) của DEAP: EEG-SKDNet tiếp tục cho thấy hiệu suất vượt trội với độ chính xác **96.48%**, vượt qua các phương pháp như: Rhythm-specific CNN (95.00%); MTL\_MSRLN (71.92%). Mặc dù phải xử lý số lượng lớp lớn hơn, EEG-SKDNet vẫn giữ được kiến trúc nhẹ và khả năng suy luận nhanh – cho thấy tính mở rộng tốt (scalability) của mô hình.

Trên bộ dữ liệu **DREAMER**: EEG-SKDNet tiếp tục thể hiện hiệu suất vượt trội với độ chính xác **98.76%**, vượt qua các phương pháp gần đây như: TC\_Net (98.67%); TDMNN (95.00%); Rhythm-specific CNN (94.45%). Đáng chú ý, EEG-SKDNet đạt được điều này với MFLOPs thấp nhất (6.10) và thời gian suy luận ngắn nhất (20.62 ms), thể hiện hiệu quả tính toán vượt trội. Trong khi đó: TC\_Net: 77.66 MFLOPs, 122.07 ms; Rhythm-specific CNN: 132.45 MFLOPs, 1280.3 ms.

Để làm rõ mối liên hệ giữa số tham số (params) và chi phí tính toán (FLOPs) trong nghiên cứu, NCS phân tích riêng hai đại lượng này. Với Conv1D, số tham số chủ yếu do tích  $C_{in} \times C_{out} \times k$ , trong khi FLOPs còn nhân thêm chiều dài đặc trưng theo thời gian qua các tầng (độ phân giải đầu ra  $T_{out}$ ), nên cùng #params nhưng nếu giữ  $T_{out}$  lớn (ít pooling) thì MFLOPs vẫn tăng mạnh. Quan sát Bảng 4.4: EEG\_SKDNet đạt 2.63M tham số và 9.33 MFLOPs với 22.12 ms nhờ hạt nhân Conv1D nhỏ kèm giãn (mở rộng vùng bao phủ mà không tăng nhiều tham số/phép tính) và cấu hình pooling giảm  $T_{out}$ ; khối SE chỉ bổ sung chi phí rất nhỏ nhưng cải thiện chọn lọc kênh. Ngược lại, ResNet34 có 7.29M tham số song 17.55 MFLOPs và 75.04 ms vì mạng sâu và duy trì độ phân giải đặc trưng cao hơn; MSDA-SFE 17.02M tham số nhưng tới 242.02 MFLOPs và 3445.77 ms do pipeline/độ phân giải đặc trưng lớn làm “phình” FLOPs; M1 (9.49M, 27.44 MFLOPs, 35.66 ms) cũng cho thấy FLOPs phụ thuộc mạnh vào cấu trúc và độ dài đặc trưng chứ không chỉ #params. Như vậy, kết luận rằng tham số params và Flops không tuyến tính, trong bối cảnh dữ liệu phổ EEG, tiêu chí triển khai hiệu quả là giảm FLOPs ở cùng mức #params; EEG\_SKDNet đạt cân bằng tốt giữa kích thước nhỏ, chi phí tính toán thấp và độ chính xác cao.

Tổng thể, EEG-SKDNet thể hiện sự cân bằng hấp dẫn giữa độ chính xác và hiệu quả. Mô hình duy trì chi phí tính toán thấp và thời gian suy luận nhanh mà không ảnh

hưởng đến hiệu suất phân loại, cho thấy lợi ích rõ ràng của việc tích hợp Self-Knowledge Distillation trong các mô hình nhận diện cảm xúc dựa trên EEG. Những đặc điểm này khiến EEG-SKDNet đặc biệt phù hợp để triển khai trong các ứng dụng thời gian thực hoặc tài nguyên hạn chế như: Giao diện não–máy tính (BCIs); Hệ thống theo dõi sức khỏe di động.

#### 4.3.5 Nghiên cứu loại bỏ (Ablation Study)

Bảng 4.5 trình bày kết quả của một nghiên cứu loại bỏ có hệ thống nhằm đánh giá ảnh hưởng của hai siêu tham số quan trọng: Nhiệt độ (temperature, ký hiệu  $\tau$ ), và hệ số cân bằng mất mát (loss balancing coefficient, ký hiệu  $\alpha$ ) đến độ chính xác top-1 của EEG-SKDNet trên tập dữ liệu DEAP(32). Hai siêu tham số này đóng vai trò trung tâm trong quá trình tự học chung cất tri thức (self-knowledge distillation), bằng cách điều chỉnh hành vi của các nhãn mềm (soft targets) và mức độ đóng góp tương đối của thành phần mất mát chung cất (distillation loss).

Ảnh hưởng của  $\alpha$  (hệ số cân bằng mất mát): Với mọi giá trị  $\tau$ , việc tăng  $\alpha$  từ 0.2 đến 0.7 đều dẫn đến sự cải thiện ổn định về độ chính xác. Xu hướng này cho thấy rằng việc đặt trọng số nhiều hơn cho tín hiệu học có giám sát (tức là mất mát cross-entropy với nhãn cứng) sẽ có lợi khi được cân bằng một cách hợp lý với mất mát chung cất. Tuy nhiên, giá trị  $\alpha$  quá nhỏ (ví dụ  $\alpha=0.2$ ) sẽ làm suy giảm hiệu suất, cho thấy rằng nếu quá phụ thuộc vào tín hiệu tự chung cất (distillation) trong các giai đoạn đầu của quá trình huấn luyện có thể làm cản trở quá trình hội tụ.

**Bảng 4.5** Độ chính xác Top-1 với các cấu hình khác nhau của hai siêu tham số  $\alpha$  và  $\tau$  trong mô hình EEG-SKDNet trên tập dữ liệu DEAP (32 điện cực).

| $\tau$ | $\alpha$ |       |       |              |
|--------|----------|-------|-------|--------------|
|        | 0.2      | 0.3   | 0.5   | 0.7          |
| 3      | 90.06    | 92.12 | 94.72 | 95.71        |
| 5      | 90.15    | 92.34 | 94.05 | 96.02        |
| 7      | 91.32    | 92.50 | 94.62 | 96.19        |
| 10     | 91.45    | 92.68 | 95.00 | <b>96.48</b> |

Ảnh hưởng của  $\tau$  (nhiệt độ): Khi  $\tau$  tăng từ 3 đến 10, hiệu suất phân loại tăng lên rõ rệt ở mọi giá trị  $\alpha$ . Giá trị  $\tau$  cao hơn tạo ra phân phối xác suất mịn hơn (softer), giúp mô hình thu được các mối quan hệ giữa các lớp chi tiết hơn (informative inter-class relationships), từ đó cải thiện chất lượng tín hiệu tự chung cất. Hiệu suất tốt nhất đạt được tại  $\tau = 10$  và  $\alpha = 0.7$ , với độ chính xác top-1 đạt **96.48%**.

Các kết quả này minh chứng rõ ràng vai trò quan trọng của cả  $\alpha$  và  $\tau$  trong việc tinh chỉnh hiệu suất của EEG-SKDNet. Việc cân bằng tối ưu giữa mất mát có giám sát và tín hiệu tự chung cất là cần thiết để tối đa hóa khả năng khái quát hóa của mô hình. Đặc biệt, tham số nhiệt độ  $\tau$  đóng vai trò then chốt trong việc nâng cao chất lượng của quá trình chung cất tri thức, bằng cách làm nổi bật các mối quan hệ giữa các lớp.

#### 4.4 Kết luận chương 4

Chương 4 trình bày một giải pháp mới cho bài toán EEG\_ER bằng cách giải quyết sự đánh đổi giữa độ chính xác của mô hình và hiệu quả tính toán. Mặc dù các mô hình học sâu đã đạt được hiệu suất cao trong lĩnh vực này, nhưng yêu cầu tính toán lớn của chúng lại giới hạn khả năng ứng dụng trong các môi trường thời gian thực và có tài nguyên hạn chế. Để khắc phục điều đó, NCS đề xuất một khung học nhẹ dựa trên *Self-Knowledge Distillation* (SKD), không phụ thuộc vào mô hình giáo viên được huấn luyện trước. Phương pháp SKD sử dụng hai biểu diễn tăng cường của cùng một tín hiệu EEG đầu vào, truyền qua cùng một mô hình học sinh để điều chỉnh sự dự đoán, cho phép mô hình tự tinh chỉnh kiến thức và cải thiện khả năng khái quát hóa. Bên cạnh đó, cơ chế trọng số có điều chỉnh (scaled-weights) cũng được đưa vào nhằm ưu tiên các đoạn EEG mang nhiều thông tin bằng cách điều chỉnh ảnh hưởng của các dự đoán mềm trong quá trình huấn luyện.

Kết quả thực nghiệm trên nhiều bộ dữ liệu chuẩn cho thấy phương pháp đề xuất vượt trội so với các mô hình nhẹ truyền thống và có khả năng cạnh tranh với các mô hình phức tạp hơn, trong khi vẫn giảm đáng kể chi phí tính toán và thời gian huấn luyện. Điều này cho thấy khung mô hình rất phù hợp cho các ứng dụng thời gian thực và nhúng.

Đồng thời, nghiên cứu cũng mở ra nhiều hướng phát triển mới, chẳng hạn như áp dụng SKD cho nhận diện cảm xúc đa phương thức và nâng cao khả năng khái quát hóa giữa các cá nhân, góp phần thúc đẩy sự phát triển của các hệ thống nhận biết cảm xúc hiệu quả và có thể triển khai thực tế.

Nhờ những tối ưu hóa này, EEG-SKDNet có thể duy trì hiệu suất nhận diện cảm xúc cao mà không làm tăng đáng kể mức tiêu thụ tài nguyên. Cấu trúc mô hình được thiết kế để phù hợp với các hệ thống EEG\_ER theo định hướng AIoT, giúp đảm bảo khả năng triển khai thực tế mà vẫn đạt được độ chính xác vượt trội so với các phương pháp truyền thống. Các kết quả đạt được của chương 4 cũng được NCS công bố trong CB [7].



# KẾT LUẬN VÀ NHỮNG KIẾN NGHỊ NGHIÊN CỨU TIẾP THEO

## 1. Một số kết luận

Xây dựng kiến trúc mạng học sâu hiệu quả cho việc triển khai bài toán EEG\_ER cho các ứng dụng HCI thực tế, triển khai tại các thiết bị biên của mạng AIoT, nơi mà điều kiện tài nguyên bị hạn chế, là một xu hướng tất yếu. Tuy nhiên, khó khăn gặp phải là cần tìm kiếm một kiến trúc học sâu gọn nhẹ sao cho có thể cài đặt trên các thiết bị bị giới hạn bởi khả năng lưu trữ và xử lý trong khi vẫn phải duy trì được hiệu quả nhận dạng cao. Để giải quyết vấn đề này, trước tiên chúng ta cần phải xác định được kiến trúc học sâu nền tảng. Tiếp đó, cần áp dụng một số biện pháp kỹ thuật để cải tiến kiến trúc này sao cho đạt được mục tiêu về kích thước mô hình cũng như hiệu quả nhận dạng. So với mục tiêu đề ra, luận án đã thực hiện được các nội dung chủ yếu sau:

- Cập nhật các kiến thức tổng quan về bài toán EEG\_ER, tập trung sâu vào nghiên cứu các kỹ thuật quan trọng cần triển khai trong các khâu thu thập, trích chọn đặc trưng tín hiệu EEG và phân lớp cảm xúc; xác định rõ các thách thức cần phải vượt qua khi xây dựng một hệ thống EEG\_ER theo định hướng AIoT. Một phần của nội dung này cũng đã được công bố trong CB [1].
- Nghiên cứu khả năng ứng dụng của các kỹ thuật học sâu khác nhau trong bài toán EEG\_ER; cập nhật các công trình nghiên cứu gần nhất ứng dụng kỹ thuật học sâu để giải quyết bài toán EEG\_ER; tiến hành các thử nghiệm kiểm chứng để từ đó xác định được *kiến trúc học sâu 1D-CNN cùng với dạng đầu vào đặc trưng FFT là phù hợp nhất đối với định hướng triển khai bài toán EEG\_ER trong điều kiện hạn chế về tài nguyên*. Công bố các kết quả nghiên cứu trong các bài báo CB [2], CB [3], CB [4].
- Nghiên cứu *áp dụng hai kỹ thuật “Inception” và “Squeeze and Excitation” để cải tiến kiến trúc học sâu 1D-CNN truyền thống sao cho phù hợp hơn với các ứng dụng EEG\_ER trong điều kiện hạn chế về tài nguyên*. Đề xuất hai mô hình học sâu có kích thước nhỏ nhưng hiệu quả nhận dạng cao. Chứng minh được hiệu quả hoạt động của hai mô hình này thông qua các công cụ mô phỏng. Công bố kết quả nghiên cứu trên các công trình CB [5] (mô hình EEG\_MCISNet cho bài toán phân loại nhị phân) và CB [6] (mô hình EEG\_SICNet cho bài toán phân loại đa lớp).
- Dựa trên mô hình học sâu EEG\_SICNet đã thu được, nghiên cứu *áp dụng kỹ thuật học chuyển giao SKD nhằm tăng hiệu suất nhận dạng, tăng tốc độ suy luận mà không cần tăng kích thước mô hình*. Trong đó, kỹ thuật Scaled weight cho student

loss giúp mô hình tập trung hơn vào các tín hiệu quan trọng sau khi trích xuất FFT, giúp tăng hiệu suất và độ chính xác mà không cần tăng kích thước mô hình. Hiệu quả hoạt động của kiến trúc đạt được (mô hình EEG\_SKDNet) đã được chứng minh thông qua các công cụ mô phỏng. Công bố kết quả nghiên cứu trên công trình CB [7].

## 2. Những kiến nghị nghiên cứu tiếp theo

Có thể thấy rằng bài toán EEG\_ER là một lĩnh vực nghiên cứu hấp dẫn, đầy tiềm năng, có thể áp dụng trong rất nhiều lĩnh vực trong đời sống thực tế. Đây là một bài toán phức tạp nhưng sẽ được giải quyết nếu ta biết ứng dụng các thành tựu nghiên cứu trong các lĩnh vực như xử lý tín hiệu số, trí tuệ nhân tạo... Trong đó, việc áp dụng các giải pháp cải tiến kiến trúc học sâu 1D-CNN truyền thống với đặc trưng đầu vào là FFT (như đã trình bày chi tiết trong luận án) đem lại các kết quả đáng khích lệ khi triển khai các ứng dụng trong điều kiện hạn chế về tài nguyên. Vì vậy, theo quan điểm của học viên, *luận án còn có một số hướng phát triển sau:*

- Nghiên cứu triển khai thử nghiệm cho một số bài toán hỗ trợ quyết định dựa trên thông tin từ tín hiệu EEG như điều khiển xe lăn điện, điều khiển nhà thông minh...
- Nghiên cứu áp dụng các giải pháp kỹ thuật mới nhằm nâng cao hơn nữa hiệu quả hoạt động của kiến trúc học sâu cho bài toán EEG\_ER khi triển khai trên các ứng dụng AIoT thực tế.
- Đặc biệt đối với bài toán EEG\_ER, nội dung của luận án mới chỉ tập trung vào việc tìm kiếm kiến trúc học sâu hiệu quả. Bên cạnh đó, dữ liệu cũng đóng vai trò vô cùng quan trọng. Việc thiếu hụt dữ liệu cũng như sự không đồng nhất và mất cân bằng dữ liệu là những vấn đề lớn cần phải giải quyết. Cần tìm kiếm các giải pháp mở rộng và làm giàu dữ liệu hiện có, sử dụng kỹ thuật bán giám sát để gán nhãn tự động, phát triển bộ dữ liệu kiểm tra chuẩn nhằm nâng cao độ chính xác và khả năng tổng quát hóa của các mô hình nhận dạng cảm xúc.

Do giới hạn về thời gian nghiên cứu và kiến thức của bản thân, luận án khó có thể tránh khỏi một số sai sót nhất định. NCS rất mong nhận được sự đóng góp ý kiến của các thầy cô, các nhà khoa học quan tâm để luận án được hoàn thiện hơn.

# DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC

- [CB1] Dương Thị Mai Thương, Phùng Trung Nghĩa (2021), “Ứng dụng học sâu trong các hệ thống giao tiếp não - máy, hiện trạng, thách thức và định hướng nghiên cứu”, *Kỷ yếu Hội thảo quốc gia lần thứ XXIV: Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông – Thái Nguyên*, 13–14/12/2021.
- [CB2] Dương Thị Mai Thương, Phùng Trung Nghĩa, Phùng Thị Hồng Dung (2022), “Ứng dụng mạng nơ-ron tích chập trong nhận dạng tín hiệu điện não”, *Tạp chí Khoa học và Công nghệ Đại học Thái Nguyên*, 227(18): 212–221.
- [CB3] Dương Thị Mai Thương, Phùng Trung Nghĩa (2023), “Phân tích hiệu quả hoạt động của một số kiến trúc học sâu cho bài toán nhận dạng cảm xúc dựa trên tín hiệu điện não”, *Kỷ yếu Hội nghị Khoa học công nghệ Quốc gia lần thứ XVI về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR)*, Đà Nẵng, 28–29/9/2023, tr. 88–94. <https://DOI:10.15625/vap.2023.0025>
- [CB4] Mai, T.D.T., Phung, T.N. (2023), “Evaluating the Performance of Some Deep Learning Model for the Problem of Emotion Recognition Based on EEG Signal”. In: Nghia, P.T., Thai, V.D., Thuy, N.T., Son, L.H., Huynh, V.N. (eds), *Advances in Information and Communication Technology. ICTA 2023. Lecture Notes in Networks and Systems*, vol. 847. Springer, Cham. [https://doi.org/10.1007/978-3-031-49529-8\\_19](https://doi.org/10.1007/978-3-031-49529-8_19). (Q4)
- [CB5] Mai, T.D.T., Trung, N.P., Phuong, H.N. (2024), “A CNN Approach in Building EEG-Based Emotion Recognition System for AIoT Applications”. In: Dao, N.N., Pham, Q.D., Cho, S., Nguyen, N.T. (eds), *Intelligence of Things: Technologies and Applications. ICIT 2024. Lecture Notes on Data Engineering and Communications Technologies*, vol. 230. Springer, Cham. [https://doi.org/10.1007/978-3-031-75596-5\\_12](https://doi.org/10.1007/978-3-031-75596-5_12). (Q3)
- [CB6] Thuong, D.T.M., Nguyen, P.H., Phung, T.N., Dang, N.C. (2024), “A Lightweight Deep Learning Network for Emotion Recognition Applications on Portable Devices”, *International Journal of Knowledge and Systems Science*. <https://DOI:10.4018/IJKSS.373712> (Q3)
- [CB7] Thuong, D.T.M., Phung, T.N., Nguyen, P.H. (2025), “EEG-SKNet: A Self-Knowledge Distillation Model with Scaled Weights for Emotion Recognition from EEG Signals”, *IEEE Access*, **Accepted**. (Q1)

# DANH MỤC TÀI LIỆU THAM KHẢO

- [1] D. Acharya **and others**. “Multi-class Emotion Classification Using EEG Signals”. *in Proceedings of the International Conference on Intelligent Computing and Communication*: [https://doi.org/10.1007/978-981-16-0401-0\\_38](https://doi.org/10.1007/978-981-16-0401-0_38). Springer, 2021, **pages** 474–491.
- [2] P. Ackermann **and others**. “EEG-based automatic emotion recognition: Feature extraction, selection and classification methods”. *in 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*: <https://doi.org/10.1109/HealthCom.2016.7749447>. IEEE, **september** 2016, **pages** 1–6.
- [3] S. Akter **and others**. “M1M2: Deep-Learning-Based Real-Time Emotion Recognition from Neural Activity”. *in Sensors*: 22.21 (**november** 2022). <https://doi.org/10.3390/s22218467>, **page** 8467.
- [4] T. B. Alakus, M. Gonen **and** I. Turkoglu. “Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO”. *in Biomedical Signal Processing and Control*: 60 (2020). <https://doi.org/10.1016/j.bspc.2020.101951>, **page** 101951.
- [5] S. M. Alarcao **and** M. J. Fonseca. “Emotions Recognition Using EEG Signals: A Survey”. *in IEEE Transactions on Affective Computing*: 10.3 (**july** 2019). <https://doi.org/10.1109/TAFFC.2017.2714671>, **pages** 374–393.
- [6] M. Alnaanah, M. Wahdow **and** M. Alrashdan. “CNN models for EEG motor imagery signal classification”. *in Signal Image Video Process*: 17 (**july** 2022). <https://doi.org/10.1007/s11760-022-02293-1>.
- [7] H. U. Amin **and others**. “EEG Feature Extraction with Fast Fourier Transform for Investigating different Brain regions in Cognitive and Reasoning Activity”. *in 2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA)*: <https://doi.org/10.1109/ROMA55875.2022.9915664>. IEEE, **august** 2022, **pages** 1–4.
- [8] A. Apicella **and others**. “A Survey on EEG-Based Solutions for Emotion Recognition With a Low Number of Channels”. *in IEEE Access*: 10 (2022). <https://doi.org/10.1109/ACCESS.2022.3219844>, **pages** 117411–117428.
- [9] B. Ari **and others**. “Wavelet ELM-AE Based Data Augmentation and Deep Learning for Efficient Emotion Recognition Using EEG Recordings”. *in IEEE Access*: 10 (2022). <https://doi.org/10.1109/ACCESS.2022.3181887>, **pages** 72171–72181.

- [10] A. Waheed Awan **and others**. “Advancing Emotional Health Assessments: A Hybrid Deep Learning Approach Using Physiological Signals for Robust Emotion Recognition”. **in***IEEE Access*: 12 (2024). <https://doi.org/10.1109/ACCESS.2024.3463746>, **pages** 141890–141904.
- [11] M. Bandarabadi **and others**. “Dynamic modulation of theta–gamma coupling during rapid eye movement sleep”. **in***Sleep*: 42.12 (**december** 2019). <https://doi.org/10.1093/sleep/zsz182>.
- [12] Leonid Berlyand **and** Pierre-Emmanuel Jabin. “A Review of the chain rule”. **in***Mathematics of Deep Learning*: <https://doi.org/10.1515/9783111025551-011>. De Gruyter, 2023, **pages** 119–120.
- [13] P. J. Bota **and others**. “A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals”. **in***IEEE Access*: (2019). <https://doi.org/10.1109/ACCESS.2019.2944001>.
- [14] M. Bradley **and** P. Lang. *The International Affective Digitized Sounds: Affective Ratings of Sounds and Instruction Manual*. techreport. NIMH Center for the Study of Emotion **and** Attention.
- [15] M. M. Bradley **and** P. J. Lang. “International Affective Picture System”. **in***Encyclopedia of Personality and Individual Differences*: [https://doi.org/10.1007/978-3-319-28099-8\\_42-1](https://doi.org/10.1007/978-3-319-28099-8_42-1). Cham: Springer International Publishing, 2017, **pages** 1–4.
- [16] C. Cannard, H. Wahbeh **and** A. Delorme. “Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry”. **in**(**november** 2021): <https://doi.org/10.1101/2021.11.02.466989>.
- [17] A. J. Casson. “Wearable EEG and beyond”. **in***Biomedical Engineering Letters*: 9.1 (**february** 2019). <https://doi.org/10.1007/s13534-018-00093-6>, **pages** 53–71.
- [18] A. Chawla **and others**. “Data-free Knowledge Distillation for Object Detection”. **in***2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*: <https://doi.org/10.1109/WACV48630.2021.00333>. IEEE, **january** 2021, **pages** 3288–3297.
- [19] T. Chen **and others**. “A Simple Framework for Contrastive Learning of Visual Representations”. **in***International Conference on Machine Learning (ICML)*: <https://proceedings.mlr.press/v119/chen20j.html>. 2020, **pages** 1597–1607.

- [20] W. Chen **and others**. “EEG-based emotion recognition using graph convolutional neural network with dual attention mechanism”. *in Frontiers in Computational Neuroscience*: 18 (**july** 2024). <https://doi.org/10.3389/fncom.2024.1416494>.
- [21] X. Chen **and** K. He. “Exploring Simple Siamese Representation Learning”. *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*: 2021, **pages** 15750–15758. DOI: 10.1109/CVPR46437.2021.01549.
- [22] X. Chen **and others**. “Improved Baselines with Momentum Contrastive Learning”. *in Proceedings of the European Conference on Computer Vision (ECCV)*: 2020, **pages** 682–699. DOI: 10.1109/CVPR46437.2021.01549.
- [23] Y. Chen **and others**. “Deep Learning on Mobile and Embedded Devices”. *in ACM Computing Surveys*: 53.4 (**july** 2021). <https://doi.org/10.1145/3398209>, **pages** 1–37.
- [24] Y. Cimtay **and** E. Ekmekcioglu. “Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition”. *in Sensors (Switzerland)*: 20.7 (**april** 2020). <https://doi.org/10.3390/s20072034>.
- [25] D. Cui **and others**. “Emotion Recognition on EEG Signal Using ResNeXt Attention 2D-3D Convolution Neural Networks”. *in Neural Processing Letters*: 55.5 (2023). <https://doi.org/10.1007/s11063-022-11120-0>, **pages** 5943–5957.
- [26] Z. Cui **and others**. “Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values”. *in Transportation Research Part C: Emerging Technologies*: 118 (**september** 2020). <https://doi.org/10.1016/j.trc.2020.102674>, **page** 102674.
- [27] X. Dong **and others**. “Improved Knowledge Distillation via Teacher Assistants for Sentiment Analysis”. *in* (**december** 2023): <https://doi.org/10.1109/SSCI52147.2023.10371965>, **pages** 300–305.
- [28] R.-N. Duan, J.-Y. Zhu **and** B.-L. Lu. *Differential entropy feature for EEG-based emotion classification*. <https://doi.org/10.1109/NER.2013.6695876>. 2013.
- [29] M. Egger, M. Ley **and** S. Hanke. “Emotion Recognition from Physiological Signal Analysis: A Review”. *in Electronic Notes in Theoretical Computer Science*: 343 (2019). <https://doi.org/10.1016/j.entcs.2019.04.009>, **pages** 35–55.

- [30] P. Ekman. “Are there basic emotions?” *in**Psychological Review*: 99.3 (1992). <https://doi.org/10.1037/0033-295X.99.3.550>, **pages** 550–553.
- [31] P. Ekman, R. W. Levenson **and** W. V. Friesen. “Autonomic Nervous System Activity Distinguishes Among Emotions”. *in**Science* (1979): 221.4616 (**september** 1983). <https://doi.org/10.1126/science.6612338>, **pages** 1208–1210.
- [32] K. Erat **and others**. “Emotion recognition with EEG-based brain-computer interfaces: a systematic literature review”. *in**Multimedia Tools and Applications*: 83.33 (2024). <https://doi.org/10.1007/s11042-024-18259-z>, **pages** 79647–79694.
- [33] A. S. Al-Fahoum **and** A. A. Al-Fraihat. “Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains”. *in**ISRN Neuroscience*: 2014 (**february** 2014). <https://doi.org/10.1155/2014/730218>, **pages** 1–7.
- [34] J. V. M. R. Fernandes **and others**. “Emotion Detection from EEG Signals Using Machine Deep Learning Models”. *in**Bioengineering*: 11.8 (**august** 2024). <https://doi.org/10.3390/bioengineering11080782>, **page** 782.
- [35] X. Fu **and others**. “Meta-Knowledge Distillation: Towards Fast Adaptation in Deep Networks”. *in**Proceedings of the International Conference on Machine Learning (ICML)*: /url<https://doi.org/10.48550/arXiv.2202.07940>. 2022, **pages** 6942–6951.
- [36] N. Ghassemi **and others**. “Automatic diagnosis of COVID-19 from CT images using CycleGAN and transfer learning”. *in**Applied Soft Computing*: 144 (2023). <https://doi.org/10.1016/j.asoc.2023.110511>, **page** 110511.
- [37] D. Ghimire, D. Kil **and** S. H. Kim. “A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration”. *in**Electronics*: 11.6 (**march** 2022). <https://doi.org/10.3390/electronics11060945>.
- [38] L. Gong **and others**. “Emotion recognition from multiple physiological signals using intra- and inter-modality attention fusion network”. *in**Digital Signal Processing*: 144 (**october** 2023). <https://doi.org/10.1016/j.dsp.2023.104278>, **page** 104278.
- [39] J. Gou **and others**. “Knowledge Distillation: A Survey”. *in**International Journal of Computer Vision*: 129.6 (**june** 2021). <https://doi.org/10.1007/s11263-021-01453-z>, **pages** 1789–1819.
- [40] Y. Gu **and others**. “A Domain Generative Graph Network for EEG-Based Emotion Recognition”. *in**IEEE Journal of Biomedical and Health Informatics*: 27.5 (**may** 2023). <https://doi.org/10.1109/JBHI.2023.3242090>, **pages** 2377–2386.

- [41] H. A. Hamzah **and** K. K. Abdalla. “EEG-based emotion recognition systems; comprehensive study”. **in***Heliyon*: 10.10 (**may** 2024). <https://doi.org/10.1016/j.heliyon.2024.e31485>, e31485.
- [42] G. Hinton, O. Vinyals **and** J. Dean. *Distilling the Knowledge in a Neural Network*. <http://arxiv.org/abs/1503.02531>. **march** 2015.
- [43] V. Hnamte **and** J. Hussain. “DCNNBiLSTM: An Efficient Hybrid Deep Learning-Based Intrusion Detection System”. **in***Telematics and Informatics Reports*: 10 (**june** 2023). <https://doi.org/10.1016/j.teler.2023.100053>, **page** 100053.
- [44] E. H. Houssein, A. Hammad **and** A. A. Ali. “Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review”. **in***Neural Computing and Applications*: 34.15 (**august** 2022). <https://doi.org/10.1007/s00521-022-07292-4>, **pages** 12527–12557.
- [45] J. Hu **and others**. “Squeeze-and-Excitation Networks”. **in***arXiv preprint*: (2017). <http://arxiv.org/abs/1709.01507>.
- [46] M. Hussain **and others**. “Emotion Recognition System Based on Two-Level Ensemble of Deep-Convolutional Neural Network Models”. **in***IEEE Access*: 11 (2023). <https://doi.org/10.1109/ACCESS.2023.3245830>, **pages** 16875–16895.
- [47] S. Ioffe **and** C. Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. **in***Proceedings of the 32nd International Conference on Machine Learning (ICML)*: <https://arxiv.org/abs/1502.03167>. 2015.
- [48] Md. R. Islam **and others**. “Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques”. **in***IEEE Access*: 9 (2021). <https://doi.org/10.1109/ACCESS.2021.3091487>, **pages** 94601–94624.
- [49] A. Iyer **and others**. “CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings”. **in***Multimedia Tools and Applications*: 82.4 (**february** 2023). <https://doi.org/10.1007/s11042-022-12310-7>, **pages** 4883–4896.
- [50] M. Jafari **and others**. “Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review”. **in***Computers in Biology and Medicine*: 160 (2023). <https://doi.org/10.1016/j.combiomed.2023.106998>, **page** 106998.



- [51] M. Jafari **and others**. “Emotion recognition in EEG signals using deep learning methods: A review”. **in***Elsevier Ltd.*: (october 2023). <https://doi.org/10.1016/j.combiomed.2023.107450>.
- [52] M. Jafari **and others**. “Emotion recognition in EEG signals using deep learning methods: A review”. **in***Computer Biology and Medicine*: 165 (october 2023). <https://doi.org/10.1016/j.combiomed.2023.107450>, **page** 107450.
- [53] M. Jafari **and others**. “Emotion recognition in EEG signals using deep learning methods: A review”. **in***Computer Biology and Medicine*: 165 (october 2023). <https://doi.org/10.1016/j.combiomed.2023.107450>, **page** 107450.
- [54] X. Jia **and** A. Kohn. “Gamma Rhythms in the Brain”. **in***PLoS Biology*: 9.4 (april 2011). <https://doi.org/10.1371/journal.pbio.1001045>, e1001045.
- [55] Y. Jin, J. Wang **and** D. Lin. “Multi-Level Logit Distillation”. **in***2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*: <https://doi.org/10.1109/CVPR52729.2023.02325>. IEEE, **june** 2023, **pages** 24276–24285.
- [56] S. Katsigiannis **and** N. Ramzan. “DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices”. **in***IEEE Journal of Biomedical and Health Informatics*: 22.1 (2018). <https://doi.org/10.1109/JBHI.2017.2688239>, **pages** 98–107.
- [57] S. Khare **and others**. “Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations”. **in***Information Fusion*: 102 (september 2023). <https://doi.org/10.1016/j.inffus.2023.102019>, **page** 102019.
- [58] I. Ashraf Kiyani **and** A. Razaq. “A Comprehensive Review on Sentiment Perception Using Electroencephalography (EEG)”. **in***SN Computer Science*: 3.3 (2022). <https://doi.org/10.1007/s42979-022-01155-4>, **page** 245.
- [59] D. Klepl, M. Wu **and** F. He. “Graph Neural Network-Based EEG Classification: A Survey”. **in***IEEE Transactions on Neural Systems and Rehabilitation Engineering*: 32 (2024). <https://doi.org/10.1109/TNSRE.2024.3355750>, **pages** 493–503.
- [60] S. Koelstra **and others**. “DEAP: A Database for Emotion Analysis Using Physiological Signals”. **in***IEEE Transactions on Affective Computing*: 3.1 (january 2012). <https://doi.org/10.1109/T-AFFC.2011.15>, **pages** 18–31.

- [61] J. LaRocco, M. D. Le **and** D.-G. Paeng. “A Systemic Review of Available Low-Cost EEG Headsets Used for Drowsiness Detection”. **in***Frontiers in Neuroinformatics*: 14 (**october** 2020). <https://doi.org/10.3389/fninf.2020.553352>.
- [62] Y. LeCun, Y. Bengio **and** G. Hinton. “Deep learning”. **in***Nature*: 521.7553 (2015). <https://doi.org/10.1038/nature14539>, **pages** 436–444.
- [63] J.-Y. Li **and** W.-C. Fang. “An Edge AI Accelerator Design Based on HDC Model for Real-time EEG-based Emotion Recognition System with RISC-V FPGA Platform”. **in***2024 IEEE International Symposium on Circuits and Systems (ISCAS)*: <https://doi.org/10.1109/ISCAS58744.2024.10558319>. IEEE, **may** 2024, **pages** 1–5.
- [64] L. Li **and others**. “Knowledge distillation via knowledge translation for efficient EEG emotion recognition”. **in***Information Sciences*: 608 (2022), **pages** 659–673. DOI: 10.1016/j.ins.2022.05.072.
- [65] Mu Li **and** Bao-Liang Lu. “Emotion classification based on gamma-band EEG”. **in***2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*: <https://doi.org/10.1109/IEMBS.2009.5334139>. IEEE, **september** 2009, **pages** 1223–1226.
- [66] W. Lin **and** C. Li. “Review of Studies on Emotion Recognition and Judgment Based on Physiological Signals”. **in***Applied Sciences*: 13.4 (2023). <https://doi.org/10.3390/app13042573>.
- [67] H. Liu **and others**. “Review on Emotion Recognition Based on Electroencephalography”. **in***Frontiers in Computational Neuroscience*: 15 (**october** 2021). <https://doi.org/10.3389/fncom.2021.758212>.
- [68] P. Liu **and others**. “Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing”. **in***ACM Computing Surveys*: 55.9 (**september** 2023). <https://doi.org/10.1145/3560815>, **pages** 1–35.
- [69] S. Liu **and others**. “EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network”. **in***Knowledge-Based Systems*: 265 (**april** 2023). <https://doi.org/10.1016/j.knosys.2023.110372>, **page** 110372.
- [70] X. Liu **and others**. “Federated Neural Architecture Search for Medical Data Security”. **in***IEEE Transactions on Industrial Informatics*: 18.8 (**august** 2022). <https://doi.org/10.1109/TII.2022.3144016>, **pages** 5628–5636.

- [71] Prof. Bao-Liang Lu **and** Prof. Wei-Long Zheng. *SEED Dataset: A dataset collection for various purposes using EEG signals*. <https://bcmi.sjtu.edu.cn/home/seed/index.html>. **july** 2024.
- [72] Z. Lyu **and others**. “Deep Learning for Intelligent Human–Computer Interaction”. **in** *Applied Sciences*: 12 (**november** 2022). <https://doi.org/10.3390/app122211457>, **page** 11457.
- [73] A. Mahmoud **and others**. “A CNN Approach for Emotion Recognition via EEG”. **in** *Symmetry (Basel)*: 15.10 (**september** 2023). <https://doi.org/10.3390/sym15101822>, **page** 1822.
- [74] I. B. Mauss **and** M. D. Robinson. “Measures of emotion: A review”. **in** *Cognition and Emotion*: 23.2 (**february** 2009). <https://doi.org/10.1080/02699930802204677>, **pages** 209–237.
- [75] Sachin Mehta **and others**. “DeLighT: Very Deep and Light-weight Transformer”. **in** *Proceedings of the International Conference on Learning Representations (ICLR)*: <https://arxiv.org/abs/2008.00623>. 2021.
- [76] M. Meng **and others**. “A deep subdomain associate adaptation network for cross-session and cross-subject EEG emotion recognition”. **in** *Biomedical Signal Processing and Control*: 78 (2022). <https://doi.org/10.1016/j.bspc.2022.103873>, **page** 103873.
- [77] J. A. Miranda-Correa **and others**. *AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups*. <http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/>. 2017.
- [78] M. F. Mridha **and others**. “Brain-Computer Interface: Advancement and Challenges.” **in** *Sensors (Basel)*: 21.17 (**august** 2021). <https://doi.org/10.3390/s21175746>.
- [79] M. Murugappan **and** S. Murugappan. “Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT)”. **in** *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*: <https://doi.org/10.1109/CSPA.2013.6530058>. IEEE, **march** 2013, **pages** 289–294.
- [80] F. Mushtaq **and others**. “One hundred years of EEG for brain and behaviour research”. **in** *Nature Human Behaviour*: 8.8 (**august** 2024). <https://doi.org/10.1038/s41562-024-01941-5>, **pages** 1437–1443.

- [81] M. Nazar **and others**. “A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques”. **in***IEEE Access*: PP.1 (**november** 2021). <https://doi.org/10.1109/ACCESS.2021.3127881>, **page** 1.
- [82] H. Nguyen **and** T. Tran. “Self-distillation with weak supervision for EEG classification”. **in***2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*: 2023, **pages** 1556–1561. DOI: 10.1109/BIBM59053.2023.10359581.
- [83] A. Paleyes, R.-G. Urma **and** N. D. Lawrence. “Challenges in Deploying Machine Learning: A Survey of Case Studies”. **in***ACM Computing Surveys*: 55.6 (**july** 2023). <https://doi.org/10.1145/3533378>, **pages** 1–29.
- [84] G. Peng **and others**. “Temporal relative transformer encoding cooperating with channel attention for EEG emotion analysis”. **in***Computer Biology and Medicine*: 154 (**march** 2023). <https://doi.org/10.1016/j.combiomed.2023.106537>, **page** 106537.
- [85] V. Peterson **and others**. “A feasibility study of a complete low-cost consumer-grade brain-computer interface system”. **in***Heliyon*: 6.3 (**march** 2020). <https://doi.org/10.1016/j.heliyon.2020.e03425>, e03425.
- [86] G. Placidi **and others**. “Basis for the implementation of an EEG-based single-trial binary brain computer interface through the disgust produced by remembering unpleasant odors”. **in***Neurocomputing*: 160 (**july** 2015). <https://doi.org/10.1016/j.neucom.2015.02.034>, **pages** 308–318.
- [87] D. Merlin Praveena, D. Angelin Sarah **and** S. Thomas George. “Deep Learning Techniques for EEG Signal Applications – A Review”. **in***IETE Journal of Research*: 68.4 (**july** 2022). <https://doi.org/10.1080/03772063.2020.1749143>, **pages** 3030–3037.
- [88] S. Qiu **and others**. “Event-Related Beta EEG Changes During Active, Passive Movement and Functional Electrical Stimulation of the Lower Limb”. **in***IEEE Transactions on Neural Systems and Rehabilitation Engineering*: 24.2 (**february** 2016). <https://doi.org/10.1109/TNSRE.2015.2476481>, **pages** 283–290.
- [89] S. Qiu **and others**. “A review on semi-supervised learning for EEG-based emotion recognition”. **in***Information Fusion*: 104 (**april** 2024). <https://doi.org/10.1016/j.inffus.2023.102190>.
- [90] S. Isola R. Zhang **and** A. A. Efros. “Colorful Image Colorization”. **in***ECCV*: (2016). <https://arxiv.org/abs/1603.08511>, **pages** 649–666.

- [91] R. Rastogi **and others**. “Emotion Detection via Voice and Speech Recognition”. **in***International Journal of Cyber Behavior, Psychology and Learning*: 13.1 (**november** 2023). <https://doi.org/10.4018/IJCBPL.333473>, **pages** 1–24.
- [92] J. Russell. “A Circumplex Model of Affect”. **in***Journal of Personality and Social Psychology*: 39 (**december** 1980). <https://doi.org/10.1037/h0077714>, **pages** 1161–1178.
- [93] J. A. Russell. “A circumplex model of affect.” **in***Journal of Personality and Social Psychology*: 39.6 (**december** 1980). <https://doi.org/10.1037/h0077714>, **pages** 1161–1178.
- [94] J. Park S. Lee **and** K. Kim. “Confidence-aware self-knowledge distillation for EEG-based emotion classification”. **in***Sensors*: 22.4 (2022), **page** 1583. DOI: 10.3390/s22041583.
- [95] E. S. Salama **and others**. “EEG-Based Emotion Recognition using 3D Convolutional Neural Networks”. **in***International Journal of Advanced Computer Science and Applications*: 9 (2018). <https://api.semanticscholar.org/CorpusID:52846885>.
- [96] Y.-S. Seo **and** J.-H. Huh. “Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications”. **in***Electronics (Basel)*: 8.2 (**february** 2019). <https://doi.org/10.3390/electronics8020164>, **page** 164.
- [97] S. Siامي-ناميني, N. Tavakoli **and** A. S. Namin. “The Performance of LSTM and BiLSTM in Forecasting Time Series”. **in***2019 IEEE International Conference on Big Data (Big Data)*: <https://doi.org/10.1109/BigData47090.2019.9005997>. IEEE, **december** 2019, **pages** 3285–3292.
- [98] L. N. Smith. “Cyclical Learning Rates for Training Neural Networks”. **in***2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*: <https://doi.org/10.1109/WACV.2017.58>. IEEE, **march** 2017, **pages** 464–472.
- [99] M. Soleymani **and others**. “A Multimodal Database for Affect Recognition and Implicit Tagging”. **in***IEEE Transactions on Affective Computing*: 3.1 (2012). <https://doi.org/10.1109/T-AFFC.2011.25>, **pages** 42–55.
- [100] K. Song **and others**. “An Intelligent Epileptic Prediction System Based on Synchrosqueezed Wavelet Transform and Multi-Level Feature CNN for Smart Healthcare IoT”. **in***Sensors*: 22.17 (**august** 2022). <https://doi.org/10.3390/s22176458>, **page** 6458.
- [101] N. Srivastava **and others**. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. **in***Journal of Machine Learning Research*: 15 (2014), **pages** 1929–1958.

- [102] A. Subasi **and** E. Erçelebi. “Classification of EEG signals using neural network and logistic regression”. **in***Computer Methods and Programs in Biomedicine*: 78.2 (**may** 2005). <https://doi.org/10.1016/j.cmpb.2004.10.009>, **pages** 87–99.
- [103] K. Sun, X. Wang **and** Q. Zhao. *A Review of AIoT-based Edge Devices and Lightweight Deployment*. No DOI available. 2023.
- [104] M. Sun **and others**. “A Dual-Branch Dynamic Graph Convolution Based Adaptive TransFormer Feature Fusion Network for EEG Emotion Recognition”. **in***IEEE Trans Affect Comput*: 13.4 (2022). <https://doi.org/10.1109/TAFFC.2022.3199075>, **pages** 2218–2228.
- [105] C. Szegedy **and others**. “Going deeper with convolutions”. **in***2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: <https://doi.org/10.1109/CVPR.2015.7298594>. IEEE, 2015, **pages** 1–9.
- [106] K. Takahashi **and others**. “Propagating Waves in Human Motor Cortex”. **in***Frontiers in Human Neuroscience*: 5 (2011). <https://doi.org/10.3389/fnhum.2011.00040>.
- [107] S. Tippannavar **and others**. “Advances and Challenges in Human Emotion Recognition Systems: A Comprehensive Review”. **in***Journal of Trends in Computer Science and Smart Technology*: 5 (**november** 2023). <https://doi.org/10.36548/jtcsst.2023.4.001>, **pages** 367–387.
- [108] A. Topic **and others**. “Emotion Recognition Using a Reduced Set of EEG Channels Based on Holographic Feature Maps”. **in***Sensors*: 22.9 (**april** 2022). <https://doi.org/10.3390/s22093248>, **page** 3248.
- [109] R. Vempati **and** L. D. Sharma. “A systematic review on automated human emotion recognition using electroencephalogram signals and artificial intelligence”. **in***Results in Engineering*: 18 (2023). <https://doi.org/10.1016/j.rineng.2023.101027>, **page** 101027.
- [110] P. Vincent **and others**. “Extracting and composing robust features with denoising autoencoders”. **in***Proceedings of the 25th International Conference on Machine Learning*: (2008), **pages** 1096–1103. DOI: <https://doi.org/10.1145/1390156.1390294>.
- [111] D. Q. Vu, N. Le **and** J. C. Wang. “Teaching Yourself: A Self-Knowledge Distillation Approach to Action Recognition”. **in***IEEE Access*: 9 (2021). <https://doi.org/10.1109/ACCESS.2021.3099856>, **pages** 105711–105723.

- [112] Y. Li W. Zhao **and** H. Chen. “Improving generalization in EEG classification via multi-source knowledge distillation”. *in 2022 International Joint Conference on Neural Networks (IJCNN)*: 2022, **pages** 1–8. DOI: 10.1109/IJCNN55064.2022.9891944.
- [113] J. Wang **and** M. Wang. “Review of the emotional feature extraction and classification using EEG signals”. *in Cognitive Robotics*: 1 (2021). <https://doi.org/10.1016/j.cogr.2021.04.001>, **pages** 29–40.
- [114] J. Wang **and** M. Wang. “Review of the emotional feature extraction and classification using EEG signals”. *in Cognitive Robotics*: 1 (2021). <https://doi.org/10.1016/j.cogr.2021.04.001>, **pages** 29–40.
- [115] L. Wang **and others**. “The Implicit Regularization of Momentum Gradient Descent in Overparametrized Models”. *in Proceedings of the AAAI Conference on Artificial Intelligence*: 37.8 (june 2023). <https://doi.org/10.1609/aaai.v37i8.26209>, **pages** 10149–10156.
- [116] P. Wang **and others**. “EEG emotion recognition based on knowledge distillation optimized residual networks”. *in 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*: <https://doi.org/10.1109/IAEAC54830.2022.9929699>. IEEE, **october** 2022, **pages** 574–581.
- [117] Q. Wang **and others**. “Multi-modal emotion recognition using EEG and speech signals”. *in Computers in Biology and Medicine*: 149 (2022). <https://doi.org/10.1016/j.combiomed.2022.105907>, **page** 105907.
- [118] T. Wang **and others**. “Graph-based knowledge distillation for EEG emotion recognition”. *in IEEE Transactions on Affective Computing*: (2023). DOI: 10.1109/TAFFC.2023.3242432.
- [119] X. Wang **and others**. “Deep learning-based EEG emotion recognition: Current trends and future perspectives”. *in Frontiers in Psychology*: 14 (february 2023). <https://doi.org/10.3389/fpsyg.2023.1126994>.
- [120] Y. Wang **and others**. “EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition”. *in 2018 International Joint Conference on Neural Networks (IJCNN)*: <https://doi.org/10.1109/IJCNN.2018.8489715>. 2018.
- [121] Z. Wang **and others**. *Meta-Knowledge Distillation for Few-shot Video Classification*. <https://arxiv.org/abs/2103.15671>. 2021.
- [122] N. S. Williams, G. M. McArthur **and** N. A. Badcock. “10 years of EPOC: A scoping review of Emotiv’s portable EEG device”. *in* (july 2020): <https://doi.org/10.1101/2020.07.14.202085>.

- [123] J. R. Wolpaw **and others**. “Brain–computer interfaces for communication and control”. **in***Clinical Neurophysiology*: 113.6 (**june** 2002). [https://doi.org/10.1016/S1388-2457\(02\)00057-3](https://doi.org/10.1016/S1388-2457(02)00057-3), **pages** 767–791.
- [124] D. Wu, Y. Xu **and** B.-L. Lu. “Transfer Learning for EEG-Based Brain–Computer Interfaces: A Review of Progress Made Since 2016”. **in***IEEE Transactions on Cognitive and Developmental Systems*: 14.1 (**march** 2022). <https://doi.org/10.1109/TCDS.2020.3007453>, **pages** 4–19.
- [125] L. Yang **and others**. “Stochastic weight averaging enhanced temporal convolution network for EEG-based emotion recognition”. **in***Biomedical Signal Processing and Control*: 83 (**may** 2023). <https://doi.org/10.1016/j.bspc.2023.104661>, **page** 104661.
- [126] C. Yu **and** M. Wang. “Survey of emotion recognition methods using EEG information”. **in***Cognitive Robotics*: 2 (2022). <https://doi.org/10.1016/j.cogr.2022.06.001>, **pages** 132–146.
- [127] C. Yu **and** M. Wang. “Survey of emotion recognition methods using EEG information”. **in***Cognitive Robotics*: 2 (**november** 2022). <https://doi.org/10.1016/J.COGR.2022.06.001>, **pages** 132–146.
- [128] L. Ma Z. Xu **and** Q. Liu. “Cross-subject EEG emotion recognition via spatial-temporal self-knowledge distillation”. **in***Pattern Recognition*: 141 (2023), **page** 109507. DOI: 10.1016/j.patcog.2023.109507.
- [129] H. Zhang **and others**. “The applied principles of EEG analysis methods in neuroscience and clinical neurology”. **in***Military Medical Research*: 10.1 (**december** 2023). <https://doi.org/10.1186/s40779-023-00502-7>, **page** 67.
- [130] J. Zhang **and** K. Yang. “Multi-teacher Knowledge Distillation for EEG-based Emotion Recognition”. **in***IEEE Access*: 10 (2022), **pages** 109657–109667. DOI: 10.1109/ACCESS.2022.3211781.
- [131] J. Zhang **and others**. “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review”. **in***Information Fusion*: 59 (**july** 2020). <https://doi.org/10.1016/j.inffus.2020.01.011>, **pages** 103–126.
- [132] S. Zhang, C. Tang **and** C. Guan. “Visual-to-EEG cross-modal knowledge distillation for continuous emotion recognition”. **in***Pattern Recognition*: 130 (**october** 2022). <https://doi.org/10.1016/j.patcog.2022.108833>, **page** 108833.



- [133] X. Zhang **and others**. “Feature Pyramid Networks and Long Short-Term Memory for EEG Feature Map-Based Emotion Recognition”. **in***Sensors*: 23.3 (**february** 2023). <https://doi.org/10.3390/s23031622>, **page** 1622.
- [134] Y. Zhang **and others**. “An Investigation of Deep Learning Models for EEG-Based Emotion Recognition”. **in***Frontiers in Neuroscience*: 14 (**december** 2020). <https://doi.org/10.3389/fnins.2020.622759>.
- [135] W.-L. Zheng **and others**. “EmotionMeter: A Multimodal Framework for Recognizing Human Emotions”. **in***IEEE Transactions on Cybernetics*: 49.3 (**march** 2019). <https://doi.org/10.1109/TCYB.2018.2797176>, **pages** 1110–1122.