

그래프 강화 시공간 EEG-Deformer를 이용한 감정 인식 모델

김형국*, 김민서**

Emotion Recognition Model using a Graph-Enhanced
Spatiotemporal EEG-Deformer

Hyoung-Gook Kim*, Min-Seo Kim**

이 논문은 2026년도 광운대학교 우수연구자 지원 사업에 의해 연구되었음

요 약

본 논문에서는 EEG 기반 감정 인식의 성능 향상을 위해 그래프 강화 시공간 EEG-Deformer 모델을 제안한다. 제안 모델은 그래프 기반 시공간 특징 인코더를 통해 초기 단계에서 시간적 특징과 공간적 의존성을 통합한 시공간 표현을 효과적으로 학습한다. 또한, 계층적 거친-정밀 트랜스포머로 다양한 시간 해상도에서의 특징을 모델링하고, 적응형 밀집 정보 정제 모듈을 활용하여 불필요한 정보를 억제하고 중요한 시공간 특징을 선택적으로 강조한다. 이후 각 모듈에서 추출된 특징을 통합하여 다층 퍼셉트론으로 최종 감정 상태를 분류한다. 3가지의 인지 상태 및 4가지의 감정 인식 데이터셋 실험에서, 제안 모델은 기존 대비 정확도와 F1 점수 모두에서 일관된 향상을 보였으며, 이는 그래프 기반 시공간 특징 학습과 계층적 시간 모델링, 정보 정제 및 통합의 결합이 성능 향상에 기여함을 보여준다.

Abstract

This paper proposes a graph-enhanced spatiotemporal EEG-Deformer model to improve EEG-based emotion recognition performance. The proposed model employs a graph-based spatiotemporal feature encoder to effectively learn unified representations of temporal dynamics and spatial dependencies at the early stage. In addition, a hierarchical coarse-to-fine transformer is introduced to capture features at multiple temporal resolutions, while an adaptive dense information purification module suppresses irrelevant information and selectively emphasizes important spatiotemporal features. The extracted features from each module are subsequently integrated and fed into a multilayer perceptron for final emotion classification. Experimental results on three cognitive state datasets and four emotion recognition datasets demonstrate that the proposed model consistently improves both accuracy and F1-score compared to existing methods. These results suggest that effectively integrating graph-based spatiotemporal feature learning, hierarchical temporal modeling, and information purification and fusion improves performance.

Keywords

electroencephalography, EEG, emotion recognition, EEG-deformer, graph attention, transformer

* 광운대학교 전자융합공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-1518-4100>
** 광운대학교 전자융합공학과 학사과정
- ORCID: <https://orcid.org/0009-0008-5473-9392>

· Received: Apr. 22, 2026, Revised: May 22, 2026, Accepted: May 25, 2026
· Corresponding Author: Hyoung-Gook Kim
Dept. of Electronic Convergence Engineering, Kwangwoon University, 20
Gwangun-ro, Nowon-gu, Seoul 01897, Korea
Tel.: +82-2-940-5574, Email: hkim@kw.ac.kr

1. 서 론

감정은 인간의 인지 과정과 의사결정에 핵심적인 영향을 미치며, 이를 정량적으로 인식하는 기술은 HCI, 정신건강 모니터링, 맞춤형 헬스케어, 몰입형 콘텐츠 시스템 등 다양한 분야에서 중요한 기반 기술로 활용되고 있다. 이를 위해 심전도, 피부전도반응, 안면 표정 등 여러 생리·행동 신호가 활용되어 왔으며, 그중 뇌전도(EEG, Electroencephalography)는 비침습적으로 뇌의 전기적 활동을 측정하고 높은 시간 해상도로 감정 변화의 동적 특성을 포착할 수 있다는 장점이 있다. 이러한 이유로 EEG 기반 감정 인식은 인간의 내적 상태를 실시간으로 추정하는 효과적인 접근법으로 연구되고 있다.

그러나 EEG 신호는 시공간적 비선형성, 개인 간 변동성, 그리고 낮은 신호 대 잡음비 등 복잡한 특성을 가지므로, 이를 효과적으로 반영할 수 있는 정교한 모델링이 요구된다. 특히, 시공간적 정보를 통합적으로 학습하면서 잡음에 강인한 특징 표현을 도출하는 것은 여전히 중요한 과제이다. 기존 머신러닝 방법들은 이러한 복잡한 구조를 충분히 반영하지 못해 미세한 시간적 변화와 채널 간 공간적 상호 관계를 동시에 포착하는 데 한계를 보인다. 이를 극복하기 위해 딥러닝 기반 방법이 도입되었으며, 합성곱 신경망(CNN, Convolutional Neural Network)은 국소적 특징 추출에, 순환 신경망(RNN, Recurrent Neural Network)은 시간적 의존성 모델링에 효과적인 성능을 보여왔다. 최근 Transformer 기반 모델은 SA(Self-Attention) 메커니즘을 통해 전역적 의존성을 학습하는 데 강점을 제공하지만, 장거리 의존성 학습의 한계, 계산 효율성 문제, 그리고 채널 간 구조적 관계 반영의 제약 등은 여전히 남아 있다.

한편, EEG-Deformer[1]는 CNN 기반 특징 추출과 계층적 Transformer 구조를 결합하여 다중 시간 해상도 기반 시간적 표현을 학습함으로써, 인지주의, 운전 피로도, 작업부하 등 다양한 인지 상태 인식 문제에서 우수한 성능을 보였다. 그러나 해당 모델은 EEG 채널 간의 공간적 상호작용을 명시적으로 모델링하지 못하며, 채널 간 관계를 고정된 방식으

로 처리한다는 한계가 존재한다. 또한 기존 밀집 정보 정제(DIP, Dense Information Purification) 모듈은 서로 다른 계층에서 추출된 특징을 동일한 비중으로 통합하기 때문에 특징 중요도를 적응적으로 반영하지 못한다는 문제가 존재한다. EEG 기반 감정 인식에서는 특정 뇌 영역 간의 동적 상호작용과 시공간적 의존성을 함께 고려하는 표현 학습이 중요하므로, 이를 효과적으로 반영할 수 있는 모델 구조가 필요하다.

이에 본 연구에서는 기존 EEG 기반 감정 인식 모델의 한계를 보완하고 시공간 특징 표현의 학습 능력을 향상시키기 위해 그래프 강화 시공간 EEG-Deformer(Graph-Enhanced Spatiotemporal EEG-Deformer, GST-ED)를 제안한다. 제안하는 모델은 EEG 신호의 공간적 상관관계와 시간적 동역학을 효과적으로 반영하기 위한 세 가지 핵심 모듈로 구성된다.

첫째, EEG 채널 간의 구조적 관계와 국소 시공간 의존성을 효과적으로 반영하기 위해 그래프 어텐션 기반 그래프 강화 시공간 특징 인코더(GSTFE, Graph-Enhanced Spatiotemporal Feature Encoder)를 도입하였다. GSTFE는 채널 간 연결 관계를 그래프 형태로 모델링하여 특정 뇌 영역 간의 동적 상호작용을 반영하고, 공간적·시간적 특징을 통합적으로 추출함으로써 기존 EEG-Deformer의 제한적인 채널 관계 표현 한계를 개선한다.

둘째, 다양한 시간 해상도에서 EEG의 장·단기 시간 의존성을 효과적으로 학습하기 위해 계층적 거친-정밀 Transformer(HCFT, Hierarchical Coarse-to-Fine Transformer)를 활용하였다. 특히 기존 MSA(Multi-Head Self-Attention)의 정보 희석 및 높은 계산 비용 문제를 완화하기 위해 다중 스케일 게이트 선형 어텐션(MA, Multiscale Gated Linear Attention)을 적용하여 중요한 시간 구간을 강조하면서 전역적인 시간 의존성을 효율적으로 학습하도록 설계하였다.

셋째, 기존 DIP 모듈을 확장한 적응형 밀집 정보 정제(ADIP, Adaptive Dense Information Purification) 모듈을 제안하였다. ADIP는 계층별 및 특징 차원별 중요도를 적응적으로 반영하여 중요한 시공간 특징

은 강조하고 불필요한 정보와 잡음은 억제함으로써, 서로 다른 계층에서 추출된 특징들의 활용도와 최종 특징 표현의 판별력을 향상시킨다.

마지막으로, GSTFE와 HCFT 및 ADIP 모듈에서 추출된 시공간 특징들을 통합적으로 학습함으로써 EEG 기반 감정 인식의 정확도를 향상시키고자 하였다. 이러한 구성 모듈들은 제안 모델이 EEG 신호의 복잡한 공간적 의존성과 시간적 동역학을 효과적으로 반영하도록 하며, 잡음에 강인한 특징 표현과 우수한 일반화 성능을 달성할 수 있도록 한다.

본 논문의 구성은 다음과 같다. 2장에서는 EEG 기반 감정 인식 관련 연구를 소개하고 3장에서는 제안 방법의 구조와 각 구성 요소를 설명한다. 4장에서는 실험을 통해 모델의 성능을 평가하며, 5장에서는 결론과 함께 향후 연구 방향을 제시한다.

II. 관련 연구

EEG 기반 감정 인식은 인간의 내적 상태를 비침습적 접근으로 반영할 수 있어 활발히 연구되어 왔다. 초기 연구에서는 파워 스펙트럼, 차분 엔트로피, 로그 분산 등의 수작업 특징을 활용하여 서포트 벡터 머신(SVM, Support Vector Machine), K-최근접 이웃(KNN, K-Nearest Neighbors), 랜덤 포레스트와 같은 전통적인 기계학습 모델을 적용하였다. 이러한 접근은 구현이 간단하고 해석 가능하다는 장점이 있으나, EEG 신호의 복잡한 시공간적 의존성을 충분히 반영하지 못하는 한계를 가진다.

이후 딥러닝 기반 방법론이 도입되면서 EEG 신호로부터 특징을 자동으로 학습하는 접근이 가능해졌으며, 다양한 신경망 구조가 제안되었다. CNN 기반 모델은 공간적 패턴과 국소적 주파수 정보를 효과적으로 추출할 수 있어 널리 활용되었으며, 대표적으로 EEGNet[2]은 경량 구조를 통해 효율적인 특징 학습을 가능하게 하였다. RNN 계열 모델은 시계열 특성을 모델링하기 위해 적용되었으며, LSTM 및 GRU 기반 구조는 시간적 의존성을 모델링하는데 활용되었다. 특히 BGRU[3]는 양방향 시계열 정보를 반영하여 성능을 향상시켰다. 그러나 CNN 기반 모델은 장거리 의존성 모델링에 한계를 가지며,

RNN 계열은 순차적 계산 구조로 인해 학습 효율성과 병렬 처리 측면에서 제약이 존재한다. 최근에는 SA 메커니즘을 기반으로 하는 Transformer가 도입되면서 EEG 신호 내 장거리 의존성 학습이 가능해졌다. EEG-Conformer[4]는 CNN과 Transformer를 결합하여 국소적 특징 추출과 전역적 관계 모델링을 동시에 수행하는 구조를 제안하였다. 그러나 이러한 Transformer 기반 방법은 입력 토큰 간 관계를 완전 연결 방식으로 처리하기 때문에 EEG 채널 간의 구조적 특성을 명시적으로 반영하는 데 한계가 있다.

이러한 한계를 해결하기 위해 EEG 채널 간 관계를 명시적으로 모델링하는 그래프 신경망(GNN, Graph Neural Network) 기반 접근이 활발히 연구되고 있다. GNN 기반 방법은 EEG 전극을 노드로, 채널 간 상관관계를 엣지로 정의함으로써 뇌의 공간적 구조를 효과적으로 반영할 수 있다. 최근에는 GNN과 Transformer를 결합한 Graph Transformer 구조가 제안되어, 구조적 관계 학습과 장거리 의존성 모델링을 동시에 수행하는 방향으로 발전되고 있다.

이러한 연구의 흐름 속에서 EEG-Deformer는 얇은 CNN 기반 인코더(SFE, Shallow Feature Encoder)로 국소적 특징을 추출한 후, HCFT 구조를 통해 다중 해상도의 시간적 표현을 학습하는 방법을 제안하였다. 또한, DIP 모듈을 통해 중간 표현을 점진적으로 정제하고 통합함으로써 표현 학습의 효율성을 향상시켰다. 그러나 해당 모델은 채널 간 연결 구조가 고정되어 있어 뇌 영역 간의 동적 상호작용을 충분히 반영하지 못하며, 채널 간 공간적 관계와 시간적 의존성을 보다 유연하게 결합하는 데에는 제한이 존재한다. 더불어 특징 정제 과정에서 모듈 간 기여도를 동적으로 조절하는 메커니즘이 충분히 고려되지 않아, 표현 학습 최적화 측면에서 개선 여지가 존재한다.

III. 제안 모델 : GST-ED

본 연구에서는 그림 1에 나타난 바와 같이 기존 EEG-Deformer 구조를 확장하여, SFE를 GSTFE로 대체하고, HCFT 및 ADIP 모듈을 통합한 GST-ED를 제안한다.

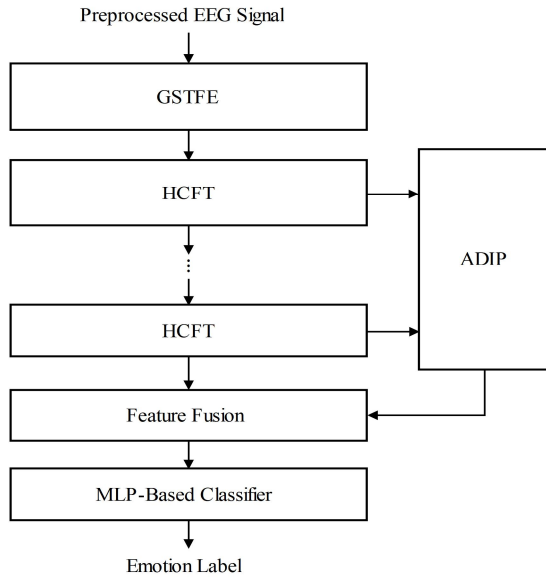


그림 1. GST-ED의 구조

Fig. 1. Architecture of GST-ED

입력 EEG 신호는 아티팩트 제거 및 대역통과 필터링 등의 전처리 과정을 거친 후, GSTFE를 통해 Transformer 입력에 적합한 토큰 형태로 변환된다. GSTFE는 채널 간 그래프 구조를 반영하여 저차원의 시공간 특징 표현을 추출한다. 이후 HCFT는 생성된 토큰 시퀀스를 기반으로 다중 시간 해상도에서 특징을 계층적으로 학습하여 단기 및 장기 시간 의존성을 동시에 포착한다. 이어서 ADIP 모듈은 특징 차원별 적응적 게이팅 메커니즘을 통해 불필요한 정보를 억제하고 중요한 시공간 특징을 선택적으로 강조한다. 최종적으로 각 단계에서 추출된 특징을 특징 융합(Feature Fusion)을 통해 통합한 후, 다층 퍼셉트론 기반 분류기(MLP-Based Classifier)를 이용하여 EEG 기반 감정 상태를 예측한다.

3.1 그래프 강화 시공간 특징 인코더(GSTFE)

그림 2에 나타난 바와 같이, GSTFE는 깊이별 시계열 합성곱층(DTC, Depthwise Temporal Convolution), 배치 정규화(BN, Batch Normalization), ELU(Exponential Linear Unit), 그래프 어텐션 네트워크(GAT, Graph Attention Network), BN, ELU, 평균 풀링(MP, Mean Pooling), 채널 통합 기반 토큰화(CAT, Channel Aggregation-Based Tokenization)로 구성된다.

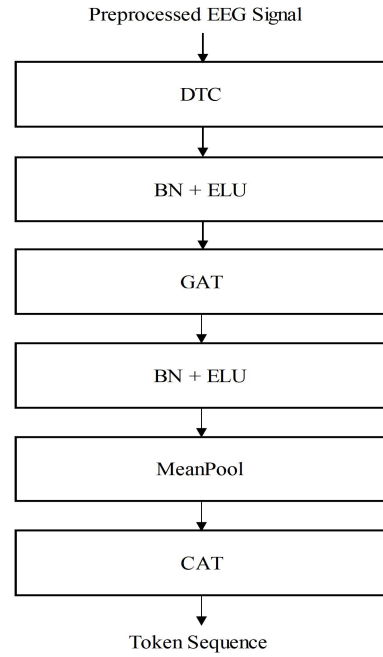


그림 2. GSTFE의 구조

Fig. 2. Architecture of the GSTFE

먼저, 각 EEG 채널의 시간적 패턴을 독립적으로 추출하기 위해 DTC를 적용한다. 여기서 DTC는 채널 간 혼합 없이 각 채널의 국소 시간 특징을 추출하는 역할을 수행한다. 이후, 추출된 특징에 대해 BN을 적용하여 특징 분포를 정규화하고 학습 안정성을 향상시키며, EEG 신호의 양·음 진폭 정보를 모두 보존하기 위해 ELU 활성화 함수를 사용한다.

다음으로, EEG 채널 간 관계를 그래프 구조로 구성하고 GAT를 적용하여 채널 간 동적 상호작용을 학습한다. 여기서 EEG 채널은 각 전극 위치에서 측정된 신호를 의미하며, 전극 위치에 따른 뇌 영역별 공간 정보를 반영한다. 이를 통해 시간적 특징과 공간적 의존성이 통합된 시공간 표현을 생성한다. GAT는 고정된 공간 합성곱과 달리 데이터에 따라 채널 간 관계를 적응적으로 학습할 수 있다는 장점을 가진다. 각 채널을 그래프의 노드로 정의하면, 노드 간 어텐션 점수는 식 (1)과 같이 계산된다.

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_i \parallel \mathbf{W} \mathbf{h}_j]) \quad (1)$$

여기서 $\mathbf{h}_i$ 는 노드 i 의 특징 벡터, $\mathbf{W}$ 는 학습 가능한 가중치 행렬을 나타낸다. 이때 $\mathbf{W} \mathbf{h}_i$ 와 $\mathbf{W} \mathbf{h}_j$ 는

각각 가중치 행렬 $\mathbf{W}$ 에 의해 변환된 노드 i 와 노드 j 의 특징 벡터이며, $\parallel$ 는 두 벡터를 이어 붙이는 연결 연산을 의미한다. 또한, $\mathbf{a}$ 는 연결된 특징 벡터와 곱셈이 가능하도록 동일한 길이를 갖는 학습 가능한 어텐션 벡터이다.

이후 어텐션 점수는 식 (2)와 같이 소프트맥스 함수를 통해 정규화되어 어텐션 가중치로 변환된다.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^C \exp(e_{ik})} \quad (2)$$

여기서 C 는 채널 수를 나타낸다.

정규화된 가중치를 바탕으로 각 채널의 표현은 이웃 채널의 정보를 반영하여 식 (3)과 같이 갱신된다.

$$\mathbf{h}_i^{\text{GAT}} = \sigma \left(\sum_{j=1}^C \alpha_{ij} \mathbf{W} \mathbf{h}_j \right) \quad (3)$$

이후 전체 GAT 출력은 식 (4)와 같이 갱신된 모든 채널 표현을 결합한 형태로 구성된다.

$$\mathbf{H}^{\text{GAT}} = \{\mathbf{h}_i^{\text{GAT}}\}_{i=1}^C \quad (4)$$

다음으로, 식 (5)와 같이 GAT를 통해 강화된 특징에 대해 BN과 ELU를 적용하여 특징 분포를 안정화하고 비선형 표현력을 향상시킨다. 이후 시간 차원에 대해 평균 풀링을 적용하여 국소적인 시간 정보를 요약하고 잡음을 감소시키면서 중요한 패턴을 유지한다. 특히 GAT를 통해 중요한 특징이 상대적으로 강조된 상태이므로, 평균 기반 집계는 과도한 변동성을 줄이면서 비교적 안정적인 특징 표현을 제공한다.

$$\tilde{\mathbf{H}}^{\text{GAT}} = \text{MeanPool}(\text{ELU}(\text{BN}(\mathbf{H}^{\text{GAT}}))) \quad (5)$$

이후 CAT를 통해 식 (6)과 같이 채널 차원의 정보를 통합하여 각 시간 지점에 대응하는 토큰 표현을 생성한다.

$$\mathbf{z}_t = \text{CAT}(\tilde{\mathbf{H}}_t^{\text{GAT}}), t = 1, 2, \dots, T \quad (6)$$

최종적으로 생성된 토큰들을 시간 순서에 따라 배열하여 식 (7)과 같은 토큰 시퀀스를 구성한다.

$$\mathbf{Z} = \{\mathbf{z}_t\}_{t=1}^T \quad (7)$$

여기서 $\tilde{\mathbf{H}}_t^{\text{GAT}}$ 는 시점 t 에서의 GAT 기반 채널 특징을 의미하며, $\text{CAT}(\cdot)$ 는 채널 통합 기반 토큰화 연산을 나타낸다. 또한 t 는 시간 인덱스, T 는 전체 시간 길이(또는 토큰의 개수)를 의미한다.

이와 같이 구성된 특징 표현은 이후 HCFT에 입력되어 시공간 통합 특징 학습에 활용된다.

3.2 계층적 거친-정밀 트랜스포머(HCFT)

HCFT는 GSTFE를 통해 생성된 특징을 입력으로 받아 시간 축에 대한 계층적 특징을 학습하는 모듈이다. EEG 신호는 다양한 시간 스케일에서 의미 있는 패턴이 나타나므로, EEG 시간 정보에 대한 정밀하고 세분화된 표현을 형성하기 위해 HCFT는 CGB(Coarse-grained Branch)와 FGB(Fine-grained Branch)로 구성된 병렬 구조를 사용한다. 이러한 구조는 여러 계층으로 스택되어 적용되며, 각 계층 l 에서는 입력 토큰 $\mathbf{z}_t^{(l)}$ 를 기반으로 다중 시간 해상도의 특징 표현 $\mathbf{z}_t^{(l+1)}$ 을 생성한다. 전체 구조는 그림 3에 제시되어 있으며, 본 논문에서는 3개의 계층을 사용하였다.

CGB는 장기적인 시간 의존성과 거친 수준의 EEG 시간 동역학을 학습하기 위해 설계되었다. 이를 위해 어텐션 연산 이전에 최대 풀링(Max Pooling)을 적용하여 입력 시퀀스의 시간 해상도를 감소시킨다. 이 과정은 실제 EEG 획득 과정의 샘플링 주파수를 변경하는 것이 아니라 모델 내부에서 수행되는 시간축 다운샘플링(Temporal down-sampling)에 해당한다. 입력 시퀀스 길이는 풀링의 스트라이드(Stride)에 비례하여 감소하며, 이를 통해 시간적 수용 영역(Receptive field)이 확대되고 고주파의 국소적 변동이 완화된다. 이후 다운샘플링된 시퀀스는 선형 투영(LP, Linear Projection)을 통해 Query(Q), Key(K), Value(V)로 변환된다.

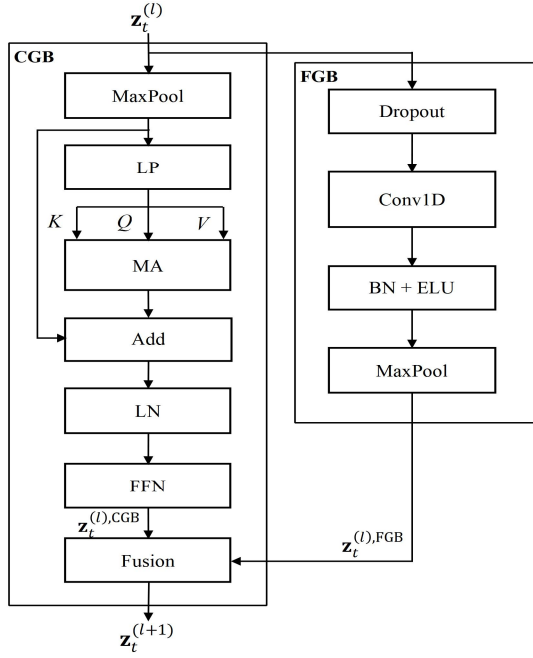


그림 3. HCFT의 구조

Fig. 3. Architecture of the HCFT

이후 전역적인 시간 의존성을 효율적으로 학습하기 위해 MA를 적용하였다. 기존 MSA는 모든 토큰 간 상호작용을 동일한 방식으로 처리하기 때문에 중요한 시간 구간의 정보가 희석될 수 있으며 계산 비용 또한 증가하는 한계가 있다. 반면 MA는 중요한 시간 영역의 특징을 선택적으로 강조하면서 선형 복잡도 기반의 어텐션 연산을 수행함으로써 계산 효율성과 표현 능력을 동시에 향상시킨다. MA의 출력은 최대 풀링을 거친 입력 시퀀스와 잔차 연결(Residual Connection)을 통해 결합되며, 안정적인 학습을 위해 레이어 정규화(LN, Layer Normalization)가 적용된다. 이후 단순 선형 변환만으로는 포착하기 어려운 복잡한 비선형 패턴을 학습하기 위해 피드포워드 신경망(FFN, Feed-Forward Network)을 추가로 적용한다. 이를 통해 식 (8)과 같이 CGB는 장기적인 시간 의존성을 효과적으로 반영하는 특징 표현 $z_t^{(l),CGB}$ 을 생성한다.

이때 CGB에서 전역 시간 의존성을 학습하기 위해 사용되는 MA 연산은 식 (9)와 같이 정의된다.

$$z_t^{(l),CGB} = \text{FFN}(\text{LN}(z_t^{(l),MA} + \text{MaxPool}(z_t^{(l)}))) \quad (8)$$

$$\text{MA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sum_{s=1}^S \mathbf{G}_s \odot (\Phi(\mathbf{Q}_s) (\Phi(\mathbf{K}_s)^T \mathbf{V}_s)) \quad (9)$$

여기서 S 는 다중 스케일의 개수를 나타내며, $\mathbf{Q}_s, \mathbf{K}_s, \mathbf{V}_s$ 는 최대 풀링을 거친 입력 토큰 $z_t^{(l)}$ 을 각 스케일 s 에 대응하는 선형 투영을 통해 변환한 Query, Key, Value 표현을 나타낸다. $\Phi(\cdot)$ 는 기존의 내적 주의집중을 선형화하기 위해 특징 사상을 도입하는 것을 의미하며, 이를 통해 주의집중 연산의 계산 복잡도를 선형 시간으로 줄일 수 있다. $\mathbf{G}_s$ 는 각 시간 스케일에서 중요한 정보 구간을 강조하기 위해 학습되는 게이팅 벡터이며, $\odot$ 는 원소별 곱셈 연산을 나타낸다.

반면, FGB는 높은 시간 해상도를 유지한 상태에서 단기적이고 세밀한 EEG 변화를 학습하기 위해 설계되었다. 이를 위해 먼저 낮은 비율의 드롭아웃(Dropout)을 적용하여 정보 손실을 최소화하면서 과적합을 방지하고, 이후 커널 크기 3의 1차원 합성곱(Conv1D)을 사용하여 짧은 시간 구간 내의 국소적 패턴을 추출한다. 이어서 BN과 ELU 활성화 함수를 순차적으로 적용함으로써 학습 안정성을 향상시키고 미세한 진폭 변화에 대한 민감도를 높인다. 이후 스트라이드 1 기반의 풀링 연산을 적용하여 시간축 정보를 크게 손상시키지 않으면서 정제된 특징 표현 $z_t^{(l),FGB}$ 을 생성하며, 이는 식 (10)과 같이 정의된다. 이를 통해 FGB는 고주파 기반의 세밀한 시간 표현을 효과적으로 학습한다.

$$z_t^{(l),FGB} = \text{MP}(\text{ELU}(\text{BN}(\text{Conv1D}(\text{DP}(z_t^{(l)})))))) \quad (10)$$

각 경로에서 생성된 특징 $z_t^{(l),CGB}$ 와 $z_t^{(l),FGB}$ 는 융합 단계(Fusion)에서 $z_t^{(l+1)}$ 로 통합되어 다음 계층의 입력으로 전달된다. 이러한 계층적 구조가 반복적으로 적용됨에 따라 EEG 시계열 전반에 걸쳐 다양한 시간 스케일의 정보가 누적되며, 전역적 맥락과 세밀한 시간 패턴을 동시에 학습하는 계층적 특징 표현을 형성한다. 최종적으로, 마지막 계층 L 의 출력은 식 (11)과 같이 표현된다.

$$\mathbf{Z}^{\text{HCFT}} = \{\mathbf{z}_t^{(L)}\}_{t=1}^T \quad (11)$$

결과적으로 제안한 이중 경로(Branch) 구조는 서로 다른 시간 해상도에서 EEG 신호를 동시에 분석함으로써, 저주파 기반의 전역적 동역학과 고주파 기반의 국소적 순간 변화(Transient) 패턴을 상보적으로 학습할 수 있도록 한다.

3.3 적응형 밀집 정보 정제(ADIP)

HCFT에서 추출한 다중 시간 해상도 특징 $\mathbf{Z}^{\text{HCFT}}$ 은 시계열의 전역적 맥락과 국소적 변화를 모두 포괄하지만, 기존 DIP 기법에서는 HCFT의 $\mathbf{z}_t^{(l),\text{FGB}}$ 에 동일한 가중치를 부여함으로써 정보의 중요도를 충분히 반영하지 못하는 한계가 있다. 이를 보완하기 위해, 본 연구에서는 DIP 구조를 확장한 ADIP 모듈을 제안한다. 해당 모듈은 특징 차원별 적응적 게이팅 메커니즘을 통해 입력 특징의 각 차원에 대한 중요도를 학습하고, 이를 기반으로 불필요한 정보는 억제하고 중요한 시공간 특징을 선택적으로 강조한다.

먼저, ADIP는 입력 특징 $\mathbf{z}_t^{(l),\text{FGB}}$ 를 잠재 표현(Latent Representation)으로 변환한다. 이를 위해 DIP와 동일하게 로그 파워 연산을 수행하여 입력 토큰의 세부 정보를 강조하며, 이는 식 (12)와 같이 정의된다.

$$\mathbf{z}_t^{(l),\text{latent}} = \log((\mathbf{z}_t^{(l),\text{FGB}})^2 + \epsilon) \quad (12)$$

그 후, 식 (13)과 같이 잠재 표현과 원본 특징을 결합하여 기본 정제된 특징을 생성한다.

$$\mathbf{z}_t^{(l),\text{refined}} = \mathbf{z}_t^{(l),\text{FGB}} + \mathbf{z}_t^{(l),\text{latent}} \quad (13)$$

여기서 ϵ 은 로그 연산 안정화를 위한 작은 상수이며, $\mathbf{z}_t^{(l),\text{refined}}$ 는 각 계층 l 의 시점 t 에서 FGB 특징이 정제되어 얻어진 잠재 표현을 의미한다.

다음 단계로, 식 (14)와 같이 각 $\mathbf{z}_t^{(l),\text{refined}}$ 로부터 학습 가능한 적응적 가중치 β_t 를 산출하고, 이를 식 (15)와 같이 잠재 표현과 결합함으로써 중요한 정보

는 강조하고 불필요한 정보는 억제하는 선택적 정제를 수행한다.

$$\beta_t = \sigma(f_\theta(\mathbf{z}_t^{(l),\text{refined}})) \quad (14)$$

$$\mathbf{z}_t^{(l),\text{ADIP}} = \mathbf{z}_t^{(l),\text{refined}} \odot \beta_t \quad (15)$$

이후 모든 시점에 대해 이를 확장하면 식 (16)과 같이 표현된다.

$$\mathbf{Z}^{(l),\text{ADIP}} = \{\mathbf{z}_t^{(l),\text{ADIP}}\}_{t=1}^T \quad (16)$$

여기서 $\sigma(\cdot)$ 는 시그모이드 함수로 각 요소의 중요도를 0~1 범위로 정규화하고, f_θ 는 각 특징의 중요도를 학습하는 함수이며, 입력 표현에 대한 중요도 점수를 생성한다. 그리고 $\mathbf{z}_t^{(l),\text{ADIP}}$ 는 중요도가 반영된 정제된 특징을 나타낸다.

최종적으로, 각 계층에서 얻어진 특징들을 결합하여 식 (17)과 같이 전체 표현을 구성한다.

$$\mathbf{Z}^{\text{ADIP}} = [\mathbf{Z}^{(1),\text{ADIP}}, \mathbf{Z}^{(2),\text{ADIP}}, \dots, \mathbf{Z}^{(L),\text{ADIP}}] \quad (17)$$

이때 밀집(Dense) 연결을 통해 다중 레벨의 특징 정보를 통합함으로써 보다 풍부한 표현을 형성한다.

기존 DIP에서는 $\beta_t = 1$ 로 고정되어 있어 특징이 동일한 비중으로 반영되는 반면, ADIP는 데이터로부터 가중치를 학습하여 각 특징의 중요도를 적응적으로 조절한다. 이를 통해 중요한 정보는 강조하고 불필요한 정보는 억제함으로써, 입력 데이터의 핵심적인 특성을 보다 효과적으로 반영할 수 있다.

3.4 특징 융합과 다층 퍼셉트론 기반 분류기

특징 융합 단계에서는 식 (18)과 같이 HCFT의 최종 출력 특징 $\mathbf{Z}^{\text{HCFT}}$ 와 ADIP를 통해 정제된 특징 $\mathbf{Z}^{\text{ADIP}}$ 를 결합하여 통합 특징 $\mathbf{Z}^{\text{concat}}$ 을 구성한다.

$$\mathbf{Z}^{\text{concat}} = \text{Concat}(\mathbf{Z}^{\text{HCFT}}, \mathbf{Z}^{\text{ADIP}}) \quad (18)$$

두 특징을 결합함으로써 전역 정보와 정제된 세부 정보가 동시에 반영된 통합 특징을 얻을 수 있다. 생성된 통합 특징 Z^{concat} 은 다층 퍼셉트론 기반 분류기에 입력되어 최종 감정 클래스를 예측한다. 다층 퍼셉트론 기반 분류기는 완전 연결층으로 구성되며, 입력 특징을 바탕으로 감정 상태에 대한 최종 분류 결과를 생성한다.

IV. 실험 설계 및 실험 결과

4.1 실험 설계

본 연구에서는 제안한 모델의 성능 평가를 위해 총 7개의 EEG 데이터셋을 활용하였다. 인지 작업 일반화 성능 검증을 위해 EEG-Deformer 연구에서 사용된 세 가지 데이터셋(인지주의, 운전 피로도, 정신적 작업부하)을 포함하였고, 실제 환경에서의 감정 인식 성능을 평가하기 위해 DEAP[5], SEED-V[6], MAHNOB-HCI[7], DEBHA의 네 개 감정 관련 EEG 데이터셋을 추가로 활용하였다.

인지 관련 데이터셋 중 인지주의(Attention) 데이터셋[8]은 26명의 피험자가 변별/선택 반응 과제를 수행하는 동안 수집된 28채널 EEG(1kHz)로 구성되며, 집중(Attentive) 및 비집중(Inattentive)의 두 가지 상태로 분류된다. 운전 피로도(Driving fatigue) 데이터셋[9]은 27명의 피험자가 90분간 가상현실 운전 과제를 수행하며 수집된 32채널 EEG(500 Hz) 데이터로, 피로 수준에 따라 저피로(Low fatigue)와 고평로(High fatigue)의 두 가지 상태로 이진 라벨링된다. 정신적 작업부하(Workload) 데이터셋[10]은 36명의 피험자가 연속 감산 과제를 수행하는 과정에서 수집된 19채널 EEG 데이터로 구성되며, 작업 수행 조건에 따라 낮은 작업부하(Low workload)와 높은 작업부하(High workload)의 두 가지 상태를 분류 대상으로 활용하였다. 해당 데이터셋은 ICA 기반 아티팩트 제거를 수행한 후, 4초 슬라이딩 윈도우(2초 중첩)를 적용하여 분할하였다.

감정 인식 데이터셋의 경우, DEAP은 32채널 EEG(512Hz)를 기반으로 음악 자극에 대한 정서 반응을 수집한 데이터셋으로, 감정가(Valence)와 각성(Arousal)을 기준으로 고감정가-고각성(HVHA), 저감

정가-고각성(LVHA), 저감정가-저각성(LVLA), 고감정가-저각성(HVLA)의 네 가지 감정 상태로 분류하였다. SEED-V는 16명의 피험자가 세 개의 세션 동안 감정 유발 영상을 시청하는 과정에서 기록된 62채널 EEG 데이터셋으로, 행복(Happy), 슬픔(Sad), 중립(Neutral), 공포(Fear), 혐오(Disgust)의 다섯 가지 감정 범주를 포함한다. MAHNOB-HCI는 다중모달 감정 데이터셋 중 EEG 신호만을 활용하였으며, 감정가와 각성도를 기준으로 네 가지 감정 상태를 분류 대상으로 활용하였다. 마지막으로 DEBHA는 자체 구축된 고해상도 EEG 데이터셋으로, 행복(Happy), 슬픔(Sad), 분노(Angry), 이완(Relaxed)의 네 가지 감정 상태를 포함한다.

그림 4는 DEAP 데이터셋의 하나의 실험(Trial) 구간에서 일부 채널의 신호를 시간 축에 따라 시각화한 예시이다. 각 서브플롯은 개별 EEG 채널을 의미하며, 가로축은 시간, 세로축은 해당 채널에서 측정된 전압 신호의 진폭을 나타낸다. 시각화된 신호는 채널별로 서로 다른 진폭 변화와 시간적 패턴을 보이며, 이는 EEG 데이터가 시간적 변화와 채널별 공간 정보를 함께 포함하는 다채널 시계열 데이터임을 보여준다.

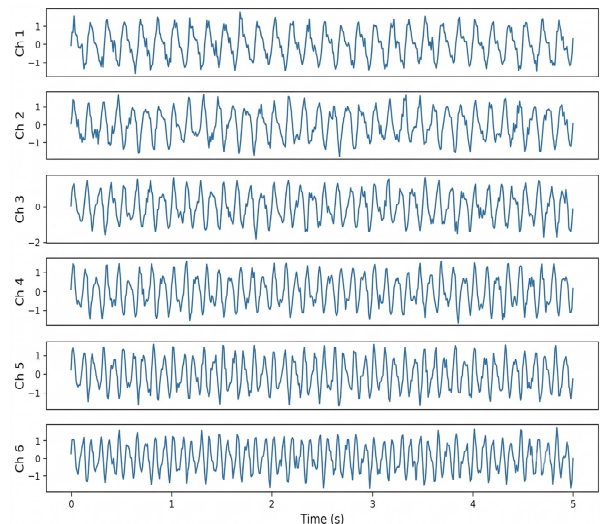


그림 4. DEAP 데이터셋 기반 채널별 EEG 학습 데이터 샘플 예시

Fig. 4. Example of channel-wise EEG training data samples from the DEAP dataset

모든 EEG 신호는 0.5-50 Hz 대역통과 필터링과 Z-score 정규화를 수행한 후, 4초 길이로 분할하였

다. 또한 모델의 일반화 및 강건성 향상을 위해 가우시안 잡음의 추가(표준편차 0.1 - 0.2), 윈도우 이동(1 - 4초, 30~50% 중첩), 시간 왜곡($\pm 10\%$) 등의 데이터 증강 기법을 적용하였다.

성능 비교를 위해 BGRU, EEGNet, EEG-Conformer, EEG-Deformer를 비교 모델로 선정하였으며, 각각 순환 신경망, 경량 합성곱, 하이브리드, 계층적 Transformer 기반 모델을 대표한다.

모든 실험은 동일한 환경에서 Adam 옵티마이저(배치 크기 64, 가중치 감소 0.01)를 사용해 1000 에폭 동안 학습하였고, 클래스 불균형 완화를 위해 가중치가 적용된 교차 엔트로피 손실을 사용하였다. 실험은 NVIDIA RTX 4090 GPU에서 수행되었으며, 성능은 5-fold 교차 검증을 통해 평균 정확도와 F1 점수로 평가하였다.

4.2 실험 결과

표 1은 세 가지 인지 데이터셋에 대해 제안한 GST-ED와 기존 EEG 기반 분류 모델들의 성능을 비교한 결과를 나타낸다. 표의 결과는 평균 $\pm$ 표준편차 형태로 제시하였다. 제안한 GST-ED는 모든 데이터셋에서 정확도와 F1 점수 기준으로 가장 높은 평균 성능을 기록하였으며, 기존 EEG-Deformer 대비 일관된 성능 향상을 보였다.

표 1. 인지 데이터셋에 대한 분류 성능 비교

Table 1. Comparison of classification performance on cognitive datasets

Model	Attention (ACC/F1)	Driving fatigue (ACC/F1)	Workload (ACC/F1)
BGRU	71.3 $\pm$ 10.6/ 70.5 $\pm$ 10.8	70.4 $\pm$ 9.8/ 69.7 $\pm$ 10.4	63.8 $\pm$ 10.3/ 60.3 $\pm$ 10.7
EEGNet	73.9 $\pm$ 10.2/ 72.5 $\pm$ 10.5	72.3 $\pm$ 10.0/ 70.5 $\pm$ 10.6	65.1 $\pm$ 10.6/ 61.5 $\pm$ 10.8
EEG-Conformer	78.1 $\pm$ 11.1/ 78.6 $\pm$ 10.7	74.4 $\pm$ 10.2/ 72.5 $\pm$ 10.7	68.6 $\pm$ 11.2/ 65.5 $\pm$ 10.6
EEG-Deformer	81.4 $\pm$ 9.8/ 80.8 $\pm$ 9.5	77.9 $\pm$ 10.1/ 76.0 $\pm$ 9.9	72.2 $\pm$ 9.7/ 69.2 $\pm$ 10.2
GST-ED	85.7$\pm$9.3/ 84.3$\pm$9.6	83.5$\pm$9.4/ 81.2$\pm$9.8	75.6$\pm$9.7/ 72.3$\pm$9.4

인지주의(Cognitive attention) 데이터셋에서는 제안

모델이 정확도 85.7%, F1 점수 84.3%를 달성하여 EEG-Deformer 대비 각각 4.3%, 3.5% 향상된 성능을 보였다. 이는 여러 EEG 채널에서 짧은 시간 내 나타나는 미세 변화가 복합적으로 발생하는 특성으로 인해, GSTFE의 DTC 층에서 추출된 저수준 특징을 기반으로 그래프 관계 학습을 수행해 채널 간 상호 의존성을 모델링함으로써 공간적 구조 정보를 강화한 결과로 판단된다. 특히 이러한 채널 관계 모델링은 주의 상태의 분산된 신경 활성화 패턴을 보다 일관된 표현 공간으로 정렬하는 데 기여했다. 또한 F1 점수 향상은 클래스 간 균형적인 분류 성능과 예측 안정성 개선으로 이어졌음을 의미한다.

운전 피로도(Driving fatigue) 데이터셋에서는 정확도 83.5%, F1 81.2%를 기록하여, EEG-Deformer 대비 각각 5.6%p, 5.2%p의 성능 향상을 보였다. 이는 장시간 운전 환경에서 수집된 비정상 시계열 특성과 다양한 생리적·환경적 잡음과 관련이 있을 수 있다. 모델 구성 측면에서 HCFT는 계층적 시계열 표현을 통해 장기 의존성을 학습하여 피로 상태의 점진적 변화를 안정적으로 모델링하고, ADIP는 잡음 성분의 영향을 완화하고 유효 특징의 중요도를 강조해 표현 신뢰성을 높인 것으로 해석된다. 또한 GSTFE는 채널 간 상호작용을 구조적으로 모델링함으로써 기능적 연결성 패턴을 포착하는 데 도움을 주어 분류 성능 향상에 기여했을 것으로 추정된다.

반면 작업부하(Cognitive Workload) 데이터셋에서는 정확도 75.6%, F1 점수 72.3%로 비교 모델 중 가장 높은 성능을 달성했으나, EEG-Deformer 대비 성능 향상 폭은 각각 3.4%, 3.1%로 다른 데이터셋 대비 제한적인 수준이었다. 이는 작업부하 상태가 다양한 인지 전략과 피험자 간 이질성을 포함하고 EEG의 개인 의존성이 강하기 때문이라고 볼 수 있다. 또한 해당 데이터셋은 피로 데이터셋 대비 장기 비정상성이나 드리프트가 약해 HCFT의 장기 시계열 모델링 효과가 제한되었을 가능성이 있으며, ADIP 역시 낮은 잡음 환경에서는 기여도가 감소했을 수 있다.

한편, BGRU, EEGNet, EEG-Conformer, EEG-Deformer로 이어지는 기존 모델들은 구조가 고도화됨에 따라 점진적인 성능 향상을 보이며, EEG 신호

의 시공간적 특성을 반영하는 모델 설계의 중요성을 보여준다. 그럼에도 제안 모델은 이들보다 일관되게 높은 성능을 달성했으며, 이는 단순한 모델 확장이 아니라 그래프 기반 시공간 특징 학습, 계층적 시간 표현 학습과 정보 정제 기반 구조 설계의 효과로 판단된다.

표 2는 DEAP, SEED-V, MAHNOB-HCI, DEBHA의 감정 인식 데이터셋에 대한 성능 비교 결과를 나타낸다. 제안 모델은 모든 데이터셋에서 비교된 기존 모델 대비 전반적으로 우수한 성능을 보였으며, 다양한 데이터 수집 환경과 피험자 특성에도 불구하고 일관된 성능 향상을 나타냈다.

표 2. EEG 기반 감정 인식 모델 성능 비교(정확도/F1 점수)

Table 2. Comparison of EEG-based emotion recognition models(Accuracy/F1 score)

Model	DEAP	SEED-V	MAHNOB	DEBHA
BGRU	68.5±10.6/ 67.9±10.4	74.5±10.2/ 74.7±9.9	70.3±11.2/ 68.9±10.6	76.3±10.4/ 75.6±9.8
EEGNet	70.8±11.3/ 70.1±11.0	78.5±10.6/ 78.2±10.2	72.4±11.7/ 71.9±11.1	78.1±11.2/ 77.5±10.7
EEG-Conformer	74.4±9.8/ 74.2±10.1	81.3±11.6/ 81.7±10.7	74.1±10.2/ 73.6±10.3	81.4±11.2/ 81.8±10.7
EEG-Deformer	77.5±8.7/ 76.8±9.1	86.4±9.2/ 84.7±9.8	78.1±8.9/ 77.9±9.8	85.8±11.2/ 85.2±10.8
GST-ED	81.3±8.3/ 81.0±8.1	91.4±8.2/ 90.8±7.9	81.5±10.3/ 81.6±9.9	91.4±8.4/ 90.8±8.9

구체적으로 DEAP 데이터셋에서는 정확도 81.3%, F1 점수 81.0%를 기록하여, 기존 최고 성능(EEG-Deformer 대비) 기준으로 각각 3.8%, 4.2% 향상을 보였다. DEAP 데이터셋은 개인 간 편차와 잡음이 비교적 큰 환경으로 알려져 있으며, 이러한 조건에서 GSTFE를 통한 채널 간 관계 학습과 ADIP 기반 특징 정제가 EEG 표현 학습에 효과적으로 기여한 것으로 해석된다. 특히 F1 점수의 향상이 더 크게 나타난 점은 클래스 간 균형 측면에서도 개선이 이루어졌음을 시사한다.

SEED-V 데이터셋에서는 정확도 91.4%, F1 점수 90.8%를 기록하여, EEG-Deformer 대비 각각 5%, 6.1% 향상을 보였다. 이는 감정 상태가 비교적 명확하게 구분되는 데이터 특성과 제안 모델의 시공

간 패턴 학습 능력이 효과적으로 결합된 결과로 해석될 수 있다.

MAHNOB-HCI 데이터셋에서는 정확도 81.5%, F1 점수 81.6%를 기록하여, EEG-Deformer 대비 각각 3.4%, 3.7% 향상을 보였다. 다른 데이터셋 대비 향상 폭은 다소 제한적이지만, 다양한 실험 환경과 피험자 간 변동성이 존재하는 조건에서도 안정적인 성능 향상을 유지한 것으로 관찰된다.

DEBHA 데이터셋에서는 정확도 91.4%, F1 점수 90.8%를 기록하여, EEG-Deformer 대비 각각 5.6%, 5.6%의 성능 향상을 보였다. 정확도와 F1 점수의 동반 향상은 전체적인 분류 성능뿐 아니라 클래스 간 균형 측면에서도 개선이 이루어졌음을 의미하며, 이는 HCFT와 ADIP의 결합이 감정 관련 특징 표현을 효과적으로 강화했음을 시사한다.

표 3은 DEAP 데이터셋에 대한 요소별 제거 실험(Ablation study) 결과를 나타낸다. 제안 모델(GST-ED)은 정확도 81.3%, F1 점수 81.0%를 기록하였다. 구성 요소별 기여도를 분석하기 위해 각 모델을 제거한 실험을 수행하였다.

표 3. DEAP 데이터셋 기반 GST-ED의 요소별 제거 실험 결과

Table 3. Results of the GST-ED ablation study based on the DEAP dataset

Ablation model	Accuracy	F1 score
GST-ED	81.3±8.3	81.0±8.1
w/o GSTFE	79.1±8.6	78.2±8.4
w/o ADIP	79.7±9.8	79.6±9.2

GSTFE를 제거한 경우(w/o GSTFE)에는 정확도 79.1%, F1 점수 78.2%로 감소하여, 기존 모델 GST-ED 대비 각각 2.2%, 2.8%의 성능 저하가 나타났다. 이는 EEG 채널 간 구조적 관계를 학습하는 GSTFE가 전체 성능 향상에 중요한 역할을 수행함을 보여준다.

또한 ADIP를 제거하고 중간 특징 맵에 대해 동적 정제를 수행하지 않는 DIP만 적용한 경우(w/o ADIP)에는 정확도 79.7%, F1 점수 79.6%로 기존 모델 GST-ED 대비 각각 성능 폭이 1.6%, 1.4%로 감소하였으며, 이는 ADIP가 중간 특징 맵에서 감정 관련 핵심 정보를 분리·강화하고 잡음을 억제하여

표현 학습 및 분류 성능 향상에 기여함을 나타낸다.

이러한 결과는 제안 모델의 성능 향상이 특정 단일 구성 요소보다는 GSTFE, HCFT, ADIP 간 상호 보완적인 구조적 결합에 의한 것임을 시사하며, 이는 EEG 신호의 시공간적 특성 표현을 더 효과적으로 학습하는 방향으로 작용했음을 보여준다.

표 4는 EEG-Deformer와 제안한 GST-ED의 모델 복잡도를 비교한 결과를 나타낸다. 본 연구에서는 제안 모델의 계산 복잡도를 파라미터 수(Params), 연산량(FLOPs), 추론 시간(Inference time) 측면에서 분석하였다. 파라미터 수는 모델에 포함된 학습 가능한 가중치의 총 개수를 의미하며, 모델의 메모리 요구량과 학습 복잡도를 나타낸다. FLOPs(Floating Point Operations)는 단일 입력에 대해 수행되는 부동 소수점 연산의 총 횟수로, 모델의 연산 비용과 추론 효율성을 평가하는 지표이다. 계산 복잡도는 실제 데이터 값과 무관하며, 모델 구조와 입력 크기에 의해 결정된다. 본 연구에서는 입력 EEG 신호의 크기를 채널 수 $C=32$, 시간 길이 $T=256$ 으로 설정하여 모든 모델의 복잡도를 동일한 조건에서 비교하였다.

표 4. 모델 복잡도 비교

Table 4. Model complexity comparison

Model	Params (M)	FLOPs (G)	Inference time (ms)
EEG-Deformer	5.6	3.1	5.2
GST-ED	6.1	3.6	6.3

EEG-Deformer의 주요 연산 비용은 셀프 어텐션 기반의 계층적 시간 모델링(HCFT) 구조에서 발생한다. 제안한 GST-ED는 기존 구조에 그래프 기반 채널 관계 학습과 ADIP 모듈을 추가하여 EEG 채널 간 공간적 상호작용과 특징 중요도를 함께 반영하도록 설계되었다. 이로 인해 추가적인 연산이 발생하지만, EEG 채널 수가 상대적으로 제한적이기 때문에 전체 계산량 증가는 비교적 제한적인 수준에 머문다.

표 4에서 확인할 수 있듯이, GST-ED의 전체 파라미터 수와 FLOPs는 EEG-Deformer 대비 각각 5.6M에서 6.1M, 3.1 GFLOPs에서 3.6 GFLOPs로 증가하였다. 그러나 이러한 계산 복잡도 증가에도 불

구하고, 추론 시간은 EEG-Deformer의 약 5.2 ms에서 GST-ED의 약 6.3 ms로 소폭 증가하는 데 그쳤다. 이는 약 1 ms 수준의 제한적인 지연 증가에 해당하며, 제안 모델이 성능 향상 대비 효율적인 계산 비용을 유지하고 있음을 보여준다. 또한, 이러한 결과는 GST-ED가 실시간 EEG 기반 감정 인식 환경에서도 충분한 적용 가능성을 가질 수 있음을 시사한다. 다만, 추론 시간은 사용된 하드웨어 환경 및 구현 방식에 따라 달라질 수 있다.

V. 결론 및 향후 연구 방향

본 연구에서는 EEG 기반 감정 인식 성능 향상을 위해 EEG-Deformer를 확장한 GST-ED를 제안하였다. 제안 모델은 채널 간 구조적 관계를 효과적으로 모델링하기 위해 GSTFE를 도입하고, HCFT를 통해 다중 스케일 시간 특징을 학습하도록 설계되었다. 또한 ADIP 모듈을 활용해 중요한 표현을 선택적으로 강조함으로써 특징 표현의 효율성과 판별력을 향상시켰다.

실험 결과, 제안한 GST-ED 모델은 EEGNet, BGRU, EEG-Conformer, 그리고 EEG-Deformer와 비교하여 정확도와 F1 점수 측면에서 일관된 성능 향상을 보였다. 이는 기존 방법들의 한계였던 EEG 채널 간 동적 공간 관계 학습 부족, 장·단기 시간 의존성의 통합적 모델링 한계, 그리고 특징 중요도에 대한 적응적 정제 부족 문제를 효과적으로 보완하였기 때문으로 판단된다. 특히 GSTFE는 그래프 기반 시공간 특징 학습을 통해 채널 간 상호작용을 효과적으로 반영하였으며, HCFT는 다양한 시간 해상도에서의 시간적 동역학을 계층적으로 모델링하였다. 또한 ADIP는 특징 차원별 중요도를 적응적으로 반영하여 불필요한 정보를 억제하고 핵심적인 감정 관련 특징을 강조함으로써 최종 분류 성능 향상에 기여하였다.

구성 요소 제거 실험을 통해 GSTFE와 ADIP 모듈이 성능 향상에 유의미하게 기여함을 확인하였으며, 이를 통해 GST-ED의 성능 향상이 특정 단일 모듈의 효과가 아니라 시공간 특징 학습, 계층적 시간 모델링, 그리고 적응적 특징 정제가 유기적으로 결합된 결과임을 검증하였다. 추가적으로 모델 복잡도

분석 결과, GST-ED는 EEG-Deformer 대비 파라미터 수와 FLOPs가 일부 증가하였으나 추론 시간 증가는 제한적인 수준에 머물렀다. 이는 제안 모델이 성능 향상과 계산 비용 사이에서 비교적 효율적인 균형을 유지하고 있음을 보여준다.

이러한 결과는 EEG 기반 감정 인식에서 채널 간 동적 관계 모델링과 계층별 특징 중요도를 반영한 정제 및 통합 학습이 핵심적인 역할을 함을 시사하며, 제안한 GST-ED가 기존 EEG-Deformer의 한계를 효과적으로 보완할 수 있음을 보여준다.

향후 연구에서는 도메인 적응 및 개인화 학습 기법을 통해 피험자 간 분포 차이를 완화하고 일반화 성능을 더욱 향상시킬 필요가 있다. 또한 멀티모달 정보 융합과 모델 경량화를 통해 실시간 EEG 기반 감정 인식 시스템으로의 확장을 고려할 수 있다. 나아가 연속적인 정서 변화 추적, 스트레스 및 인지 상태 분석 등 다양한 응용 분야에 적용함으로써 EEG 기반 감정 인식 기술의 활용 범위를 더욱 확대하고자 한다.

References

- [1] Y. Ding, Y. Li, H. Sun, R. Liu, C. Tong, C. Liu, X. Zhou, and C. Guan, "EEG-Deformer: A dense convolutional transformer for brain-computer interfaces", *IEEE Journal of Biomedical and Health Informatics*, Vol. 29, No. 3, pp. 1909-1918, Mar. 2025. <https://doi.org/10.1109/JBHI.2024.3504604>.
- [2] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces", *Journal of Neural Engineering*, Vol. 15, No. 5, Art No. 056013, Jul. 2018. <https://doi.org/10.1088/1741-2552/aace8c>.
- [3] J. X. Chen, D. M. Jiang, and Y. N. Zhang, "A hierarchical bidirectional GRU model with attention for EEG-based emotion classification", *IEEE Access*, Vol. 7, pp. 118530-118540, Aug. 2019. <https://doi.org/10.1109/ACCESS.2019.2936817>.
- [4] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG Conformer: Convolutional transformer for EEG decoding and visualization", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 31, pp. 710-719, Dec. 2022. <https://doi.org/10.1109/TNSRE.2022.3230250>.
- [5] S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals", *IEEE Transactions on Affective Computing*, Vol. 3, No. 1, pp. 18-31, Jan.-Mar. 2012. <https://doi.org/10.1109/T-AFFC.2011.15>.
- [6] T.-H. Li, W. Liu, W.-L. Zheng, and B.-L. Lu, "Classification of five emotions from EEG and eye movement signals: Discrimination ability and stability over time", *Proc. 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Francisco, CA, USA, pp. 607-610, Mar. 2019. <https://doi.org/10.1109/NER.2019.8716943>.
- [7] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging", *IEEE Transactions on Affective Computing*, Vol. 3, No. 1, pp. 42-55, Jan.-Mar. 2012. <https://doi.org/10.1109/T-AFFC.2011.25>.
- [8] J. Shin, A. von Lümann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset", *Scientific Data*, Vol. 5, Art. no. 180003, Feb. 2018. <https://doi.org/10.1038/sdata.2018.3>.
- [9] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task", *Scientific Data*, Vol. 6, Art. no. 19, Apr. 2019. <https://doi.org/10.1038/s41597-019-0027-4>.
- [10] I. Zyma, S. Tukaiev, I. Seleznev, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance", *Data*, Vol. 4, No. 1, Art. no. 14, Jan. 2019. <https://doi.org/10.3390/data4010014>.

저자소개

김 형 국 (Hyoung-Gook Kim)



2002년 2월 :

독일베를린공과대학교

전자공학과 컴퓨터과학(공학박사)

2007년 3월 ~ 현재 : 광운대학교

전자융합공학과 교수

관심분야 : 신호처리, 인공지능

김 민 서 (Min-Seo Kim)



2020년 3월 ~ 현재 : 광운대학교

전자융합공학과 학사과정

관심분야 : 신호처리, 인공지능