JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 10, pp. 9-27, October 2026
https://doi.org/10.9708/jksci.2026.31.10.009

A SVM-LSTM Hybrid Model for Improving EEG-Based Emotion Recognition Performance

Hyun-Ju Lee*, Dong-Il Shin**

*Lecturer, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea
**Professor, Department of Computer Engineering, Sejong University, Seoul, Korea

[Abstract]

This study proposes an SVM+PUK-LSTM hybrid model for EEG-based emotion recognition and evaluation of generalization to unseen subjects. Raw EEG signals from the DEAP dataset were resampled to 256Hz and processed into 3-s non-overlapping epochs, from which a 160-dimensional PSD feature vector was constructed using 32 EEG channels and five frequency bands; emotional states were defined into four Arousal–Valence-based classes: HAHV, HALV, LAHV, and LALV. The proposed model achieved an Accuracy of 98.42±3.82% and a Macro F1-score of 97.55±6.31% under 10-fold cross-validation, and an Accuracy of 90.11±7.91% and a Macro F1-score of 88.38±11.04% under 32-fold subject-independent LOSO evaluation. Under the same LOSO protocol and random seed of 1, comparisons with standalone LSTM and SVM+RBF-LSTM showed statistically significant differences in Accuracy using the Wilcoxon signed-rank test with Holm–Bonferroni correction. These results demonstrate the proposed model's 4-class EEG emotion classification performance and potential for generalization to unseen subjects.

▸ **Key words**: EEG-based Emotion Recognition, SVM-LSTM Hybrid Model, DEAP Dataset, Subject-Independent Evaluation, Leave-One-Subject-Out(LOSO)

[요 약]

본 연구는 뇌파 기반 감정인식과 새로운 피험자에 대한 일반화 성능을 평가하기 위해 서포트 벡터 머신(SVM)과 장단기 메모리(LSTM)를 결합한 SVM+PUK-LSTM 하이브리드 모델을 제안하였다. DEAP 데이터 세트의 원본 EEG 신호를 256Hz로 리샘플링하고 5개 주파수 대역으로 전처리한 후, 3초 단위의 서로 겹치지 않는 구간에서 32개 EEG 채널과 5개 주파수 대역의 스펙트럼 밀도를 이용하여 160차원 특징 벡터를 구성하였으며, 감정 상태는 Arousal과 Valence를 기준으로 HAHV, HALV, LAHV 및 LALV의 Arousal–Valence 기반 4개 클래스로 정의하였다. 제안 모델은 기존의 10-fold 교차검증에서 평균 정확도는 98.42±3.82%, 매크로 평균 F1 스코어는 97.55±6.31%를 나타냈으며, 피험자 간 데이터 중복을 배제한 32회 단일 피험자 제외(LOSO) 평가에서는 평균 정확도는 90.11±7.91%, 매크로 평균 F1 스코어는 88.38±11.04%를 나타냈다. 동일한 LOSO 및 랜덤 시드가 1인 조건에서 단일 LSTM 및 SVM+RBF-LSTM과 비교하였으며, Wilcoxon 부호순위 검정와 Holm–Bonferroni 보정을 적용한 결과 제안 모델과 두 비교모델 간 정확도 차이가 통계적으로 유의하게 나타났다. 이를 통해 제안 모델 4-class의 뇌파 감정 분류 성능과 새로운 피험자에 대한 일반화 가능성을 확인하였다.

▸ **주제어**: 뇌파 기반 감정인식, 서포트 벡터 머신-장단기 메모리 하이브리드 모델, DEAP 데이터 세트, 피험자 독립 평가, 단일 피험자 제외 교차검증

• First Author: Hyun-Ju Lee, Corresponding Author: Dong-Il Shin
*Hyun-Ju Lee (itgirldoc@gmail.com), College of Agriculture and Life Sciences, Seoul National University
**Dong-Il Shin (dshin@sejong.ac.kr), Department of Computer Engineering, Sejong University
• Received: 2026. 07. 21, Revised: 2026. 09. 10, Accepted: 2026. 09. 28.

Copyright © 2026 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945
www.kci.go.kr

I. Introduction

EEG(Electroencephalography) 기반 감정인식(Emotion Recognition) 연구는 감정 유형(type)의 판별과 유형 내 유사 응집력의 정도를 파악하기 위하여 기계학습(Machine Learning)과 딥러닝(Deep Learning) 분야에서 꾸준히 진행되어 왔다[1,2]. 1990년대 초기 연구에서는 주로 주파수 대역별 특징(Theta, Alpha, Beta, Gamma)을 추출하여 서포트 벡터 머신(Support Vector Machine, SVM), k-NN(k-Nearest Neighbor) 등의 전통적인 기계학습 기법을 적용하는 방식이 주류를 이루었으며[3,4], 이후 딥러닝 기술의 발전과 함께 합성곱 신경망(Convolutional Neural Network, CNN), 장단기 메모리(Long Short-Term Memory, LSTM) 기반 모델이 EEG 감정인식의 핵심 방법론으로 자리 잡았다[5,6]. 또한 Russell[7]의 Valence-Arousal 순환 모델을 기반으로 한 감정 표현 체계와 감정인식을 분별하는 4분면(그림 2)의 x축과 y축을 나타내는 Arousal-Valence 평면[8,9], 그리고 혼동행렬(Confusion Matrix)을 도입하여 분류 정확도를 높이려는 연구[10,11]들이 활발히 이루어졌다.

EEG 신호는 비정상적(Non-stationary)이고 개인차가 크며 고차원의 특징과 다양한 잡음을 포함하고 있어, 정확한 감정 분류를 위해서는 효율적인 특징 추출과 분류 기법이 요구된다. 이에 Dataset(데이터 세트)의 참여자 뇌파 표본들의 정확률을 산출하고 개선하기 위한 방법으로 1) 수식의 다양화, 2) 분석대상의 사분면화, 3) 결과해석의 다중화를 유도하는 방법들이 제시되어 왔다[8,10]. 첫 번째는 SVM, CNN과 같은 정확률 산출 방법의 창안과 이를 활용해 정확률을 상승시키는 방식이다. 두 번째는 EEG 신호의 주파수 특성을 효과적으로 추출하고 감정 상태를 명확하게 정의하는 과정이 중요하다. DEAP Dataset[11] 에서는 Valence, Arousal, Dominance 및 Liking 등의 자기평가 점수를 제공하며, 이 중 Valence와 Arousal은 감정 상태를 2차원 공간에서 표현하는 데 널리 활용된다. 또한 분류모델의 성능은 정확도(Accuracy)뿐만 아니라 정밀도(Precision), 재현율(Recall) 및 F1 스코어(F1-score) 등의 지표를 이용하여 평가할 수 있으며, 다중 클래스 분류에서는 클래스별 성능과 매크로 평균(Macro average), 마이크로 평균(Micro average) 및 가중치 평균(Weighted average)을 함께 고려함으로써 모델의 분류 특성을 분석할 수 있다.

그러나 EEG 신호는 높은 개인 간 변동성을 가지므로, 동일 피험자의 데이터가 학습과 평가에 함께 포함되는 평가 방식에서는 새로운 피험자에 대한 일반화 성능을 충분히 확인하기 어렵다.

또한 기존의 EEG 기반 감정인식 연구는 감정 클래스 정의, 특징 구성 및 평가 프로토콜(Protocol)이 서로 달라 보고된 분류 성능을 동일한 조건에서 직접 비교하는 데 한계가 있다. 이에 본 연구에서는 PUK 커널(Pearson VII function-based Universal Kernel) 기반 SVM과 LSTM을 결합한 SVM+PUK-LSTM(SupportVectorMachine+PUKkernel-LongShort-TermMemory) 하이브리드 모델(Hybrid Model)을 구성하고, Arousal-Valence 조합에 따른 4-class(4개 클래스) 감정 분류에서 피험자 독립(Subject- dependent) 일반화 성능을 검증하고자 한다.

DEAP Dataset[11]의 32명 피험자에 대한 원본 EEG 데이터를 사용하였으며, EEGLAB[12] 기반 전처리를 통해 512Hz의 원시 신호를 256Hz로 리샘플링한 후 FIR(Finite impulse response) filtering, CleanLine 및 독립성분분석(Independent component analysis, ICA) 기반의 인공산물(Artifact) 처리 과정을 수행하였다. 이후 32개 EEG 채널과 5개 주파수 대역으로부터 160개의 특징값(32 channels × 5 frequency bands)을 구성하여 분류 모델의 입력으로 사용하였다. 감정 클래스는 Arousal과 Valence를 기준으로 HAHV(High Arousal-High Valence), HALV(High Arousal-Low Valence), LAHV(Low Arousal-High Valence) 및 LALV(Low Arousal-Low Valence)의 네 가지 상호 배타적인 상태로 정의하였다. 본 연구의 주요 기여는 다음과 같다. 첫째, SVM-PUK과 LSTM을 결합한 하이브리드 구조를 Arousal-Valence 기반 4-class EEG 감정분류에 적용하여 비선형 분류기와 신경망 기반 분류기의 결합 가능성을 검토하였다. 둘째, 기존 10-fold 교차검증(10-fold cross-validation) 결과와 함께 DEAP Dataset의 전체 32명 피험자를 대상으로 32-fold LOSO(Leave-One-Subject-Out) 기반의 피험자 독립 평가(subject-independent evaluation)를 수행하여 피험자 간 데이터 중복을 배제한 새로운 피험자에 대한 일반화 성능을 검증하였다. 셋째, 동일한 LOSO(Leave-One-Subject-Out) 분할과 랜덤 시드(Random Seed) 조건에서 단일 LSTM, SVM+RBF-LSTM(SupportVectorMachine+RBFkernel-LongShort-Term Memory) 및 제안한 SVM+PUK-LSTM을 비교하고, 피험자별 성능에 대한 대응표본 통계 비교(Paired statistical comparison)을 수행하여 제안 모델과 비교모델 간 성능 차이를 통계적으로 검증하였다. 이를 통해 본 연구는 명확한 4-class 정의,

www.kci.go.kr

피험자 독립 평가 및 통계적 검증을 함께 적용함으로써 제안 모델의 성능과 새로운 피험자에 대한 일반화 가능성을 보다 체계적으로 확인하였다.

II. Related works

1. DEAP Dataset

DEAP Dataset[11]은 EEG 기반 감정인식 연구에서 널리 활용되는 공개 데이터 세트이다. Koelstra 등[13]이 구축한 DEAP은 총 32명의 피험자가 각각 40개의 음악 비디오 자극을 시청하는 동안 측정한 32채널 EEG와 주변 생체신호(Peripheral Physiological Signals)로 구성된다. 원시 EEG 신호는 512Hz로 수집되었으며, DEAP Dataset에서는 별도로 128Hz로 다운샘플링하고, 안구 전위 신호(Electrooculogram, EOG) 및 인공산물(Artifact) 처리 등을 수행한 전처리 데이터도 제공한다. 각 자극 시청 후 피험자는 Valence, Arousal, Dominance 및 Liking 등의 정서 상태를 자기평가 방식으로 기록하였다. 이러한 데이터 구조는 EEG 기반 감정인식 모델의 성능평가에 널리 활용되고 있으며, 표 1은 DEAP Dataset[11]의 주요 구성요소를 나타낸다. 본 연구에서는 DEAP Dataset의 공식 전처리 데이터가 아닌 512Hz의 원본 BDF(BioSemi Data Format) 데이터를 사용하였으며, 구체적인 전처리 및 특징 추출 과정은 실험방법에서 설명한다.

Table 1. The DEAP Dataset structure

Item	Value
Dataset	DEAP Dataset
Number of Subjects	32
Number of EEG Channels	32
Peripheral Physiological Signals	8
Stimuli	40 Music Video Clips
Total Trials	1,280 (32 Subjects × 40 Trials)
Sampling Rate	512 Hz
Preprocessing	EOG Artifact Removal, Noise Removal, Downsampling
Emotion Labels	Valence, Arousal, Dominance, Liking
Rating Scale	9-point Self-assessment Scale(1-9)

2. A Hybrid Model for EEG-Based Emotion Recognition

최근 EEG 기반 감정 인식에서는 단일 모델의 한계를 극복하기 위해 서로 다른 알고리즘을 결합한 하이브리드(Hybrid) 접근 방식[14]이 활발히 연구되고 있다. 하이브리드 모델(Hybrid model)은 특징 추출(Feature extraction), 특징 선택(Feature selection), 시계열 분석(Time-series analysis) 등의 장점을 결합하여 성능 향상을 목표로 한다. Zhong[15]은 합성곱신경망(Convolutional Neural Network, CNN)과 순환신경망(Recurrent Neural Network, RNN) 구조를 결합한 심층 하이브리드 신경망(Deep Hybrid Network)을 제안하였다. 해당 연구는 CNN을 이용하여 EEG의 공간적 특징을 추출하고, 순환 구조를 통해 시간적 특성을 학습함으로써 단일 모델 대비 향상된 감정 분류 성능을 보고하였다. 이는 EEG 데이터가 공간 및 시간 특성을 동시에 포함하고 있음을 보여준다. Nguyen[14]은 DWT(Discrete Wavelet Transform)와 하이브리드 지능형 기법(Hybrid Intelligent Method)을 결합한 EEG 감정인식 모델을 제안하였다. 이 연구는 EEG 신호를 다중 주파수 대역으로 분해한 후 특징 추출과 분류를 결합하는 구조를 사용하여 감정인식 성능 향상을 확인하였다.

III. Experimental Procedure

1. Experimental Workflow

본 연구에서는 DEAP Dataset[11]에서 제공하는 원본 BDF EEG 데이터(raw data)를 이용하여 EEGLAB[12] 기반의 전처리를 수행하였다. 원본 EEG 신호의 샘플링 주파수는 512Hz이며, 본 연구에서는 32개의 EEG 채널을 확인한 후 모든 채널을 256Hz로 리샘플링(resampling)을 수행하였다. 이후의 전처리 및 특징 추출 과정에서는 256Hz로 통일된 EEG 신호를 사용하였으며, 추가적인 downsampling은 수행하지 않았다. 리샘플링된 EEG 신호의 주파수 대역 분리를 위해 EEGLAB[12]의 Basic FIR filter(pop_eegfiltnew)를 적용하였다. 그림 1은 전체적인 실험과정을 묘사하였다.

각 주파수 대역의 추출은 통과 대역 하한 주파수(Lower edge of the frequency pass band)와 통과 대역 상한 주파수(Higher edge of the frequency pass band)를 이용하여 분석대상 주파수 대역에 따른 경계값을 설정하는 방식으로 수행하였다. FIR filter의 차수(Filter order)는 별도의 고정값을 수동으로 지정하지 않고, pop_eeg filtnew()의 기본 자동 설정(Default automatic filter order)을 적용하였다. 최종적으로 쎄타(Theta), 슬로알파(Slow Alpha), 알파(Alpha), 베타(Beta) 및 감마(Gamma)의 5개 주파수 대역을 분석에 사용하였으며, 각 주파수 대역의 구체적인 범위는 표 2에 제시하였다.

www.kci.go.kr

Table 2. EEG Signal Bands Used in This Study

EEG Signal	Frequency (Hz)	Emotion/Cognition Processing
Theta	4-7Hz	Cognitive Load, Emotional Processing, Memory
Slow_Alpha	*8-10Hz*	*Relaxation, Attention Shifting*
Alpha	*8-12Hz*	*Arousal Regulation, Emotional Stability*
Beta	13-30Hz	Attention, Arousal, Cognitive Activity
Gamma	31-45Hz	Higher-order Cognitive Processing, Perceptual Integration

리샘플링된 EEG 신호의 주파수 대역 분리를 위해 EEGLAB의 Basic FIR filter(pop_eegfiltnew)를 적용하였다. 각 주파수 대역의 추출은 EEGLAB[12]에서 제공하는 통과대역 하한 주파수와 통과 대역 상한 주파수를 이용하여 수행하였으며, 분석대상 주파수 대역에 따라 해당 경계값을 설정하여 필터링하였다. FIR filter의 차수(Filter order)는 별도의 고정값을 수동으로 지정하지 않고, pop_eegfiltnew()에서 제공하는 기본 자동 설정(Default automatic filter order)을 적용하였다.

EEG 신호에 포함된 전원선 잡음(Line noise)을 처리하기 위해 EEGLAB[12]의 CleanLine 기능을 적용하였다. 실제 적용 설정에서는 제거 대상인 상용 전원 주파수(Line noise frequency)를 50Hz 및 120Hz로 설정하고, 유의한 정현파 성분 검출(Sinusoidal component detection)을 위한 p-value를 0.01, 주파수 대역폭(Bandwidth)를 2Hz로 설정하였다.

슬라이딩 윈도우 길이(Sliding window length)와 이동 간격(Step size)은 각각 4s 와 2s로 설정하였다. CleanLine과 ICA를 이용한 잡음 및 인공산물 처리 후, 전처리된 EEG 신호의 주파수 특성을 확인하고 각 trial의 분석 구간을 구성하였다. 분류를 위한 데이터 구성 과정에서는 각 trial의 EEG 신호를 약 3초 길이의 분석 구간(Epoch)으로 설정하였으며, Epoch 생성을 위한 별도의 슬라이딩 윈도우 또는 중첩(Overlap)은 적용하지 않았다.

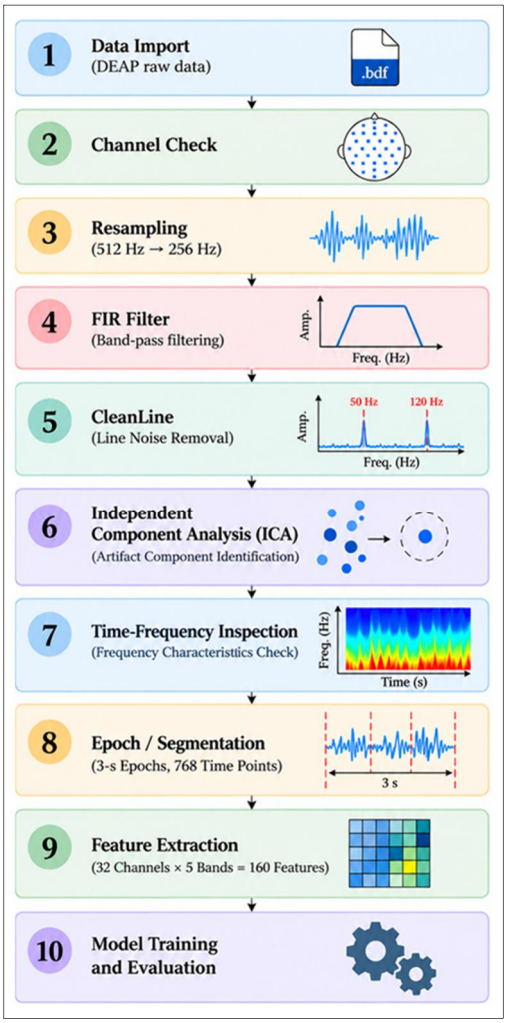


Fig. 1. Data processing and experimental procedure

하나의 분석 구간은 약 -1,000~1,996.0938ms로 설정하였으며, 256Hz의 샘플링 주파수를 기준으로 하나의 3초 분석 구간은 768개의 시간축 샘플링 포인트(Time-point samples)로 구성하였다. 여기서 768은 DEAP Dataset의 전체 1,280 trials 중 768개의 trial을 선택하거나 나머지 trial을 제외하였다는 의미가 아니다.

DEAP Dataset[11]의 기본적인 데이터 구조는 32 subjects × 40 trials = 1,280 trials이며, 768은 하나의 3초 EEG 분석 구간에서 256Hz의 샘플링 주파수에 의해 생성되는 시간축 샘플링 포인트의 수(256 samples/s × 3 s = 768 time points)를 의미한다. 따라서 전체 데이터 세트의 1,280 trials와 각 분석 구간을 구성하는 768개의 샘플은 서로 다른 데이터 차원을 나타낸다. Theta, Slow Alpha, Alpha, Beta 및 Gamma의 5개 주파수 대역별 특징값의 각 주파수 대역은 32개 EEG 채널에 대한 특징값으로 구성되므로, 하나의 분석 단위에 대한 입력 특징은 32 EEG channels × 5 frequency bands = 160 feature

www.kci.go.kr

values로 구성하였다. 추출된 160개의 특징값은 CSV(Comma-Separated Values) 형식으로 변환하여 이후 감정 분류 모델의 입력 데이터로 활용하였다[16].

또한 전처리된 EEG 신호의 주파수 상태를 확인하기 위해 EEGLAB[12]의 시간-주파수 관련 분석 기능을 이용하였다. 해당 과정은 EEG 신호의 주파수 특성을 확인하기 위한 목적으로 수행하였으며, 별도의 시간-주파수 변환 결과를 생성하여 분류 모델의 입력 특징으로 사용하지 않았다. 따라서 본 연구의 특징 데이터는 전처리된 EEG 신호에서 각 채널의 주파수 대역별 특징값을 구성하는 방식으로 생성하였다. 추출된 EEG 특징 데이터를 이용하여 감정 분류 실험을 수행하였다. 먼저 제안하는 SVM+PUK-LSTM 하이브리드 모델의 기본적인 분류 성능을 평가하기 위하여 기존 실험과 동일한 10-fold 교차검증(10-fold cross-valida tion)을 수행하였다. 전체 특징 데이터를 10개의 fold로 분할하고, 각 반복에서 9개의 fold를 학습 데이터로, 나머지 1개의 fold를 평가 데이터로 사용하여 모든 fold가 한 번씩 평가에 사용되도록 하였다. 이를 통해 제안 모델의 기존 평가 조건에서의 분류 성능을 산출하였다. 그러나 EEG 데이터에서는 동일 피험자로부터 생성된 신호가 서로 유사한 특성을 가질 수 있으므로, 일반적인 10-fold cross-validation에서 피험자 단위의 분리가 보장되지 않는 경우 동일 피험자의 데이터가 학습 fold와 평가 fold에 동시에 포함될 가능성이 있다. 따라서 본 연구에서는 10-fold cross-validation 결과를 기존 평가 조건에 따른 성능으로 제시하고, 이를 새로운 피험자에 대한 일반화 성능을 나타내는 피험자 간(Subject-independent) 평가와 구분하였다.

이러한 피험자 누수(Subject leakage) 가능성을 통제하고 새로운 피험자에 대한 모델의 일반화 성능을 평가하기 위하여 추가적으로 단일 피험자 제외(Leave-One-Subject-Out, LOSO) cross-validation [16]을 수행하였다. LOSO(Leave-One-Subject-Out) 평가에서는 단일 LSTM, SVM+RBF-LSTM 및 제안하는 SVM+PUK-LSTM 모델을 동일한 피험자 분할 조건에서 평가하였다.

DEAP Dataset[11]의 전체 32명 피험자에 대하여 각 fold에서 한 명의 피험자를 평가 대상으로 제외하고, 나머지 31명의 피험자 데이터를 학습 데이터로 구성하였다. 제외된 한 명의 피험자로부터 생성된 데이터는 해당 fold의 평가 데이터로만 사용하였으며, 이를 전체 피험자에 대해 반복하여 총 32-fold LOSO 평가를 수행하였다. 이 과정에서 동일 피험자로부터 생성된 모든 분석 데이터가 학습 데이터와 평가 데이터에 동시에 포함되지 않도록 피험자 단위로 분리하였다. 따라서 평가 대상 피험자의 데이터는 해당 LOSO fold의 평가 데이터에만 포함되고, 학습 데이터에는 포함되지 않도록 구성하였다. 또한 모델의 초기화 및 학습 과정에서 발생할 수 있는 무작위성에 따른 성능 변동을 확인하기 위하여 단일 LSTM 모델에 대해 랜덤 시드(Random seed)를 1, 42, 3407로 각각 설정하여 반복 수행한 LOSO 평가(repeated LOSO evaluation)를 추가적으로 수행하였다. 각 반복에서는 동일한 LOSO 피험자 분할 원칙을 유지하였으며, 이를 통해 서로 다른 초기화 조건에서의 모델 성능 안정성을 확인하였다.

2. Emotion Classification Based on the Arousal-Valence Plane

Arousal-Valence 평면은 감정 상태를 각성 수준(Arousal)과 정서가(Valence)의 두 차원으로 표현하는 정서 원형모형에 기초한다[17]. Arousal은 감정의 활성화 또는 각성 정도를 나타내며, Valence는 감정이 불쾌한 상태에서 유쾌한 상태까지 어느 위치에 해당하는지를 나타낸다. 이에 따라 공포는 높은 Arousal과 낮은 Valence, 지루함은 낮은 Arousal과 낮은 Valence, 흥분이나 행복은 높은 Arousal과 높은 Valence, 편안함은 낮은 Arousal과 높은 Valence 영역에 해당한다[17].

본 연구에서는 DEAP 데이터 세트에서 제공하는 각 trial의 자기평가 점수 중 Valence와 Arousal을 이용하여 감정 상태를 정의하였다[11]. 피험자 s의 n번째 시행(trial)에 부여된 원래의 Valence와 Arousal 점수를 각각 $V_{s,n}$과 $A_{s,n}$이라고 할 때, 두 점수는 모두 1점에서 9점 범위로 구성된다. 본 연구에서는 중앙 기준값인 5점을 임계값으로 적용하여 Valence와 Arousal을 각각 낮음(Low)와 높음(High) 상태로 이진화하였다. 수식 (1)의 $\overline{V}_{s,n}$과 $\overline{A}_{s,n}$은 원점수를 0 또는 1로 변환한 이진 상태 변수이다. 수식 (2)는 이진 상태 변수를 지시함수 I(·)로 동등하게 표현하며, 수식 (3)은 두 이진 상태의 조합으로 Arousal-Valence 기반 4-class 정답 레이블을 생성하는 규칙을 나타낸다.

$$\overline{V}_{s,n} = \begin{cases} 0, & V_{s,n} < 5 \\ 1, & V_{s,n} \geq 5 \end{cases}, \quad \overline{A}_{s,n} = \begin{cases} 0, & A_{s,n} < 5 \\ 1, & A_{s,n} \geq 5 \end{cases} \tag{1}$$

$$\overline{A}_{s,n} = I(A_{s,n} \geq 5), \ \overline{V}_{s,n} = I(V_{s,n} \geq 5) \tag{2}$$

$$y_{s,n}^{(AV)} = 2\overline{A}_{s,n} + \overline{V}_{s,n}, \ y_{s,n}^{(AV)} \in 0,1,2,3 \tag{3}$$

www.kci.go.kr

Arousal과 Valence의 이진 상태 조합에 따른 클래스 조건과 감정 상태를 Table 3에 제시하였다.

Table 3. Class-wise Emotional State

Class	Arousal	Valence	Emotional state
Class 0	$A_{s,n} < 5$	$V_{s,n} < 5$	LALV
Class 1	$A_{s,n} < 5$	$V_{s,n} \geq 5$	LAHV
Class 2	$A_{s,n} \geq 5$	$V_{s,n} < 5$	HALV
Class 3	$A_{s,n} \geq 5$	$V_{s,n} \geq 5$	HAHV

LALV는 Low Arousal-Low Valence를 의미하며, LAHV는 Low Arousal-High Valence를 의미한다. HALV는 High Arousal-Low Valence를 의미하고, HAHV는 High Arousal-High Valence를 의미한다.

따라서 본 연구의 4-class 분류는 Arousal과 Valence의 높고·낮음(high/low)의 조합으로 생성된 네 가지 상호 배타적 감정 상태를 구분하는 문제로 정의된다. 따라서 본 연구에서 분류 모델의 출력변수는 Arousal, Valence, Dominance 및 Liking의 네 정서 차원을 각각 예측하는 다중 출력(multi-output) 구조가 아니라, Arousal과 Valence의 이진 상태 조합으로 생성된 하나의 4-class 정답 레이블이다. 예를 들어 Arousal 점수가 7이고 Valence 점수가 3인 trial은 High Arousal-Low Valence에 해당하므로 HALV(Class 2)로 정의되며, Arousal 점수가 4이고 Valence 점수가 8인 trial은 Low Arousal-High Valence에 해당하므로 LAHV(Class 1)로 정의된다.

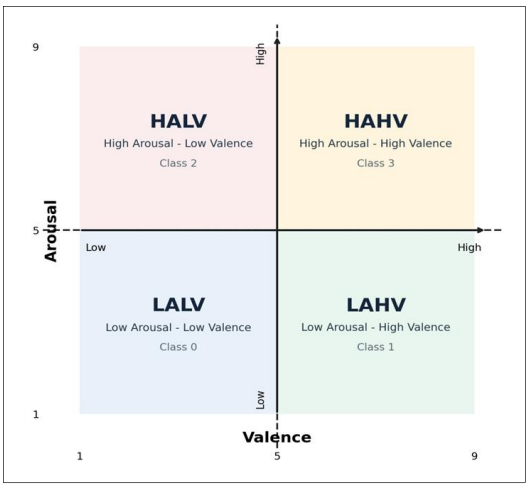


Fig. 2. Arousal-Valence plane

3. SVM Kernel Function and LSTM Network

SVM(Support Vector Machine)은 주어진 특징 공간에서 클래스 간의 마진(Margin)을 최대화하는 최적 결정 경계를 학습하는 지도학습 기반 분류 알고리즘이다[16]. 본 연구에서는 EEG 특징 공간에 존재하는 비선형 관계를 모델링하기 위하여 PUK(Pearson VII Universal Kernel) 커널을 적용한 SVM을 사용하였다. PUK 커널은 Pearson VII 함수에 기반한 범용 커널로서 커널의 폭과 형태를 조절하는 파라미터를 통해 다양한 비선형 데이터 분포를 표현할 수 있다[16,18].

PUK 커널은 Pearson VII 함수에 기반하여 커널의 형태와 폭을 조정할 수 있는 비선형 커널 함수이다. 두 특징 벡터 x_i, x_j에 대한 PUK 커널은 식 (4)과 같이 정의한다.

$$K_{PUK}(x_i, x_j) = \left| 1 + \left(\frac{2 \parallel x_i - x_j \parallel \sqrt{2^{1/w} - 1}}{\sigma} \right)^2 \right|^{-w} \quad (4)$$

여기서 ω는 PUK 함수의 형태를 조절하는 Omega parameter이고, σ는 커널의 폭을 조절하는 Sigma parameter이다. PUK 함수의 설정에서는 Omega(ω)=1.0, Sigma(σ)=1.0로 설정하였다. 따라서 본 연구에서 적용한 PUK 커널은 식 (5)처럼 표현할 수 있다.

$$K_{PUK}(x_i, x_j) = \frac{1}{1 + 4 \parallel x_i - x_j \parallel^2} \quad (5)$$

SVM-PUK은 위 커널을 이용하여 160차원 EEG 특징 벡터 간의 비선형 유사도를 계산하고, 이를 기반으로 클래스 간의 결정경계를 학습하였다[16]. 따라서 본 연구에서 SVM은 별도의 저차원 특징 벡터를 생성하는 특징 추출기로 사용되지 않으며, SVM의 출력값을 LSTM의 입력으로 전달하지 않는다. SVM-PUK과 LSTM은 동일한 EEG 특징 벡터를 각각 독립적으로 입력받아 분류를 수행하며, 두 분류기의 결과는 이후 하이브리드 분류 단계에서 결합된다.

LSTM(Long Short-Term Memory)은 memory cell과 input, forget 및 output gate를 이용하여 입력 특징 간의 의존 관계를 학습하는 순환 신경망이다[4]. 본 연구에서는 TensorFlow/Keras 2.13.0 환경에서 하나의 LSTM 계층을 구성하였으며, 은닉 유닛(Hidden units)은 10으로 설정하였다. 일반적인 LSTM에서는 순환 게이트 활성화 함수(Recurrent gate activation)로 시그모이드(Sigmoid) 계열의 함수가 널리 사용되지만, 본 연구의 실제 구현에서는 Keras LSTM 계층의 활성화 함수를 명시적으로 소프트맥스(Softmax)로 설정하여 실험을 수행하였

www.kci.go.kr

다. 또한 LSTM의 활성화 함수(activation function)에는 tanh를 적용하였다. 따라서 본 연구의 LSTM에서는 망각 게이트(Forget gate), 입력 게이트(Input gate) 및 출력 게이트(Output gate)의 게이트 활성화 함수(Recurrent activation)에 Softmax가 적용되며, 셀 후보값(Cell candidate) 및 셀 상태(Cell state)의 비선형 변환에는 tanh 활성화 함수가 사용된다.

LSTM 계층에서 생성된 출력은 4개의 출력 노드로 구성된 완전 연결층(Fully connected layer)에 전달되며, 최종 출력층에는 LSTM 내부의 게이트 활성화 함수(Gate activation)와 별도로 소프트맥스 활성화 함수(Softmax activation)를 적용하였다. 이를 통해 Arousal-Valence 조합으로 정의된 LALV, LAHV, HALV 및 HAHV의 4개 클래스에 대한 출력값을 산출하고 최종 감정 클래스를 결정하였다. 본 연구에서는 LSTM 내부 게이트의 순환 활성화 함수로 사용되는 Softmax와 최종 4-class 분류를 위한 출력층 Output Softmax를 서로 다른 역할로 적용하였다. 모델의 학습 조건은 은닉 유닛(Hidden units) 수 10, 에포크(Epochs) 30회, 배치 크기(Batch size) 64, Adam 최적화 기법, 학습률(Learning rate) 0.001 및 드롭아웃 비율(Dropout) 0.5로 설정하였다. 또한 입력 특징은 학습 데이터의 통계량을 기준으로 표준화한 후 모델 학습에 사용하였으며, 학습 데이터를 기준으로 표준화한 후 모델 학습에 사용하였다. 그림 3은 이러한 실제 구현 설정을 반영한 LSTM 계층과 4-class 출력 구조를 나타낸다.

4. SVM-LSTM Hybrid Classification Model

본 연구에서는 PUK 커널 기반 SVM과 LSTM의 서로 다른 분류 특성을 결합하기 위하여 SVM-LSTM 하이브리드 분류 모델을 구성하였다. 제안 모델은 SVM에서 생성된 특징 또는 분류 결과를 LSTM의 입력으로 전달하는 직렬 구조가 아니라, 전처리 및 특징 추출 과정을 통해 생성된 동일한 EEG 특징 벡터를 SVM-PUK과 LSTM에 각각 독립적으로 입력한 후 두 분류기의 결과를 결합하는 병렬 하이브리드 구조로 설계하였다.

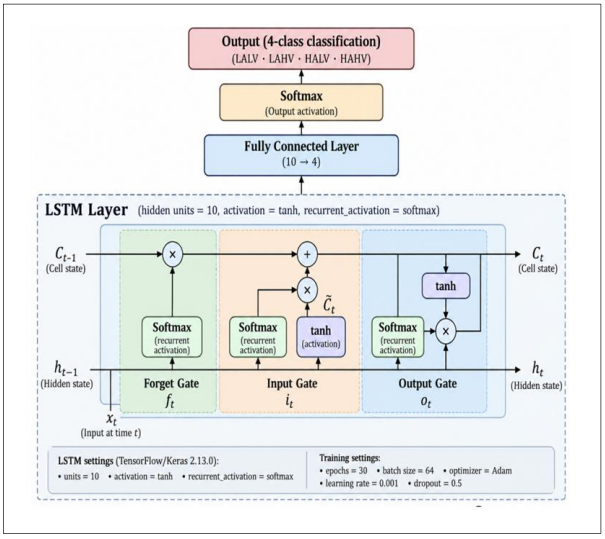


Fig. 3. LSTM Network for 4-class EEG Emotion Classification

전처리된 EEG 신호로부터 32개 채널과 Theta, Slow Alpha, Alpha, Beta 및 Gamma의 5개 주파수 대역에 대한 PSD 기반 특징을 추출하여 총 160차원의 특징 벡터 $X \in R^{160}$를 구성하였다. 생성된 특징 벡터는 SVM-PUK과 LSTM 분류기에 각각 입력된다. SVM-PUK은 PUK 커널을 이용하여 160차원 특징 공간의 비선형 관계와 클래스 간 결정경계를 학습하며, LSTM은 동일한 EEG 특징 벡터를 입력으로 사용하여 독립적인 분류 결과를 산출한다.

SVM-PUK과 LSTM에서 산출된 분류 결과는 Vote 메타 분류기를 이용하여 결합하고, 결합된 결과를 기반으로 최종 감정 클래스를 결정한다. 따라서 본 연구에서는 SVM-PUK과 단일 LSTM을 독립적인 기본 분류기(base classifier)로 구성하고, 두 모델의 분류 결과를 최종 의사결정 단계에서 통합하는 방식을 SVM-LSTM 하이브리드 분류 모델로 정의하였다. LSTM은 최종 실험에서 하나의 LSTM 계층과 하나의 출력층(Output layer)으로 구성하였으며, 은닉 유닛은 10, 학습 반복 횟수는 30 epochs로 설정하였다. LSTM 내부에서는 3절에서 설명한 바와 같이 Activation=tanh와 Recurrent_activation='Softmax'를 적용하였으며, 최종 4-class 분류 출력층에는 별도의 소프트맥스 활성화 함수를 적용하여 LALV, LAHV, HALV 및 HAHV에 대한 출력값을 산출하였다. 이때 LSTM 내부 게이트에 적용되는 게이트 활성화 함수(Recurrent Softmax)와 최종 분류 출력층의 Softmax는 서로 다른 역할을 수행한다. 세부적인 LSTM 학습 설정과 SVM의 PUK 커널 파라미터는 3절에서 제시한 설정을 동일하게 적용하였다.

www.kci.go.kr

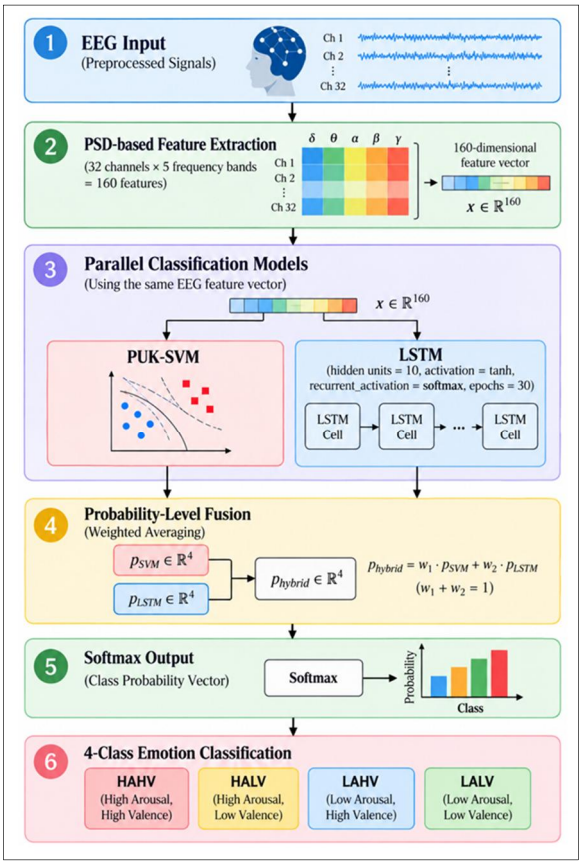


Fig. 4. Architecture of the Proposed SVM-LSTM Hybrid Model

최종 분류에서는 DEAP 데이터의 Arousal과 Valence 자기평가 점수에 5점을 임계값으로 적용하여 정의한 HAHV(High Arousal-High Valence), HALV(High Arousal-Low Valence), LAHV(Low Arousal-High Valence), LALV(Low Arousal-Low Valence)의 4개 감정 클래스를 대상으로 분류를 수행하였다. 그림 4는 전력 스펙트럼 밀도(Power Spectral Density, PSD) 기반 160차원 EEG 특징 생성, SVM-PUK과 LSTM의 병렬 분류 및 결과 결합을 통한 최종 4-class 감정 분류에 이르는 제안 SVM-LSTM 하이브리드 모델의 전체 구조를 나타낸다.

5. Performance Evaluation Metrics

분류 모델의 성능을 평가하기 위해 기존 실험에서는 피험자 의존(subject-dependent) 조건의 10-fold cross-validation을 수행하였으며, 추가적인 피험자 독립(subject-independent) 평가에서는 32-fold Leave-One-Subject-Out(LOSO) cross-validation을 수행하였다. 두 평가 방법의 성능 분석에는 Accuracy(정확도), Precision(정밀도), TP-rate(Recall), F1-score 및 표준편차(Standard Deviation, SD)를 평가 지표로 활용하였다. Accuracy는 전체 표본 중 올바르게 분류된 표본의 비율을 나타내는 지표로 식 (6)과 같이 산출하였다.

$$Accuracy(ACC) = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (6)$$

여기서 TP(True Positive)는 Positive Class를 Positive로 올바르게 분류한 경우, TN(True Negative)은 Negative Class를 Negative로 올바르게 분류한 경우를 의미한다. 반면 FP(False Positive)는 Negative Class를 Positive로 잘못 분류한 경우이며, FN(False Negative)은 Positive Class를 Negative로 잘못 분류한 경우를 의미한다. Precision은 모델이 Positive Class로 예측한 표본 중 실제 Positive Class에 해당하는 표본의 비율로 식 (7)과 같이 산출하였다.

$$\mathrm{Precision} = \frac{TP}{TP + FP} \quad (7)$$

TP-rate는 실제 Positive Class에 해당하는 표본 중 모델이 Positive Class로 올바르게 분류한 비율을 의미하며 Recall과 동일한 개념으로 식 (8)과 같이 산출하였다.

$$TP-rate(Recall) = \frac{TP}{TP + FN} \quad (8)$$

F1-score는 Precision과 Recall의 조화평균으로 두 평가 지표를 종합적으로 반영하며, 식 (9)와 같이 산출하였다[20].

$$F1-Score = \frac{2 \times (\mathrm{Precision} \times Recall)}{(\mathrm{Precision} + Recall)} \quad (9)$$

이진 분류에서는 Positive Class와 Negative Class를 기준으로 TP, TN, FP, FN을 산출하였다. 다중 클래스 분류에서는 각 클래스를 차례로 양성 클래스(Positive Class)로 설정하고 나머지 클래스를 음성 클래스(Negative Class)로 간주하는 일대다 방식(One-vs-Rest)을 적용하였다. 다중 클래스의 Precision, TP-rate 및 F1-score는 각 클래스의 표본 수를 가중치로 반영한 Weighted Average를 이용하여 산출하였으며, Accuracy는 전체 표본에 대한 정 분류율(Correctly Classified Instance s)을 기준으로 산출하였다.

또한 실험결과의 변동성을 분석하기 위해 Stand ard Standard Deviation(SD)을 사용하였으며, 식 (10)와 같이 산출하였다.

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(C_i - C')^2} \quad (10)$$

- C_i : i번째 실험에서 산출된 성능평가값
- C‘ : 전체 실험 성능 평가값의 평균
- n : SD분석에 사용된 실험결과의 개수

www.kci.go.kr

C_i는 (i)번째 실험에서 산출된 성능평가 값이며, C'는 전체 실험 성능평가 값의 평균, n은 SD 산출에 사용된 실험 결과의 개수를 의미한다.

기존의 피험자별 10-fold cross-validation 결과에 추가하여, 본 연구에서는 학습 데이터와 평가 데이터 간의 피험자 중복으로 인한 데이터 누수(Subject leakage)를 방지하고 학습 과정에서 관찰되지 않은 새로운 피험자에 대한 모델의 일반화 성능을 평가하기 위해 피험자 독립 평가(Subject-independent evaluation)를 수행하였다.

기존의 피험자별 10-fold cross-validation은 동일 피험자의 데이터 내에서 학습 및 평가가 수행되는 피험자 의존(Subject-dependent) 평가로 구분하였으며, 새로운 피험자에 대한 일반화 성능평가는 LOSO cross-validation을 활용하여 수행하였다.

6. Experimental Setup and Training Configuration

본 연구의 실험은 Windows 11 운영체제에서 Intel Core i5-13600K 프로세서와 NVIDIA GeForce RTX 5060 Ti 그래픽 카드를 사용하여 수행하였다. EEG 데이터의 전처리에는 MATLAB R2024a와 EEGLAB v2026.0.0[12]을 사용하였으며, 모델 구현 및 데이터 처리에는 Python 3.11을 기반으로 TensorFlow/Keras 2.13.0과 scikit-learn 1.2.2를 활용하였다. 실험에 사용된 주요 하드웨어 및 소프트웨어 환경과 모델 학습 설정은 표 4에 제시하였다.

SVM 기반 분류에서는 PUK 커널(Pearson VII function-based Universal Kernel) 및 RBF 커널(Radial Basis Function)을 적용하였으며, SVM-PUK의 주요 설정은 C=1.0, Omega=1.0, Sigma=1.0으로 설정하였다. LSTM 기반 모델은 은닉 유닛은 10으로 설정하였고, 30회 Epochs 동안 학습하였으며, 배치 크기는 64로 설정하였다. 최적화 알고리즘으로 Adam 최적화 기법을 사용하였으며, 학습률은 0.001, 드롭아웃 비율은 0.5로 설정하였다.

제안한 하이브리드 모델의 기존 성능평가는 피험자 의존(Subject-dependent) 조건의 10-fold cross-valida tion을 적용하여 수행하였다. 또한 동일 피험자의 데이터가 학습 데이터와 평가 데이터에 동시에 포함됨으로써 발생할 수 있는 피험자 데이터 누수를 방지하고, 학습 과정에서 관찰되지 않은 새로운 피험자에 대한 일반화 성능을 평가하기 위하여 피험자 독립(Subject-independent) Leave-One-Subject-Out(LOSO) cross-validation을 추가로 수행하였다.

Table 4. Experimental Hardware and Software Environment

Category		Specification
OS		Windows 11
Processor		Intel Core i5-13600K
Graphics Card		NVIDIA GeForce RTX 5060 Ti
MATLAB		R2024a
EEGLAB		v2026.0.0
Python		3.11
TensorFlow/Keras		2.13.0
scikit-learn		1.2.2
SVM Kernel		PUK / RBF
PUK Parameters	C	1.0
	Omega	1.0
	Sigma	1.0
LSTM_ configuration	Hidden Units	10
	Activation	tanh
	Recurrent(Gate) _Activation	Softmax
	Output Activation	Softmax
Epochs		30
Batch Size		64
Optimizer		Adam
Learning Rate		0.001
Dropout		0.5
Random Seed	Overall	1
	LSTM	1, 42, 3407
Proposed Model Evaluation		10-fold cross-validation
Subject-Independent Evaluation		32-fold LOSO
LOSO Training Subjects		31
LOSO Test Subject		1

LOSO 평가에서는 DEAP Dataset[11]의 전체 32명 피험자를 대상으로 각각의 fold 별로 한 명의 피험자를 평가 데이터로 완전히 분리하고, 나머지 31명의 피험자 데이터를 학습 데이터로 사용하여 총 32-fold의 평가를 수행하였다. 이에 따라 동일 피험자의 데이터가 학습 데이터와 평가 데이터에 동시에 포함되지 않는 피험자 비 중복(subject-disjoint) 조건을 유지하였다. 10-fold cross-validation 결과와 LOSO 결과는 서로 다른 평가 프로토콜(protocol)에 따른 결과이므로 각각 피험자 의존 성능과 피험자 독립 일반화 성능으로 구분하여 제시하였다.

본 연구에서는 실험 과정의 재현성을 확보하기 위하여 데이터 처리 및 모델 구현 과정 중 결과 재현에 직접 적으로 영향을 미치는 핵심 구현 절차와 설정을 코드 및 의사 코드 형태로 제시하였다. 구체적으로 3초 단위 EEG epoch 분할 및 전력 스펙트럼 밀도(Power Spectral Density, PSD) 특징 추출 과정, Arousal-Valence의 임계값을 이용한 4개 클래스(4-class) 레이블 생성 과정, 피험자 간 데이터 중복을 방지하기 위한 LOSO 기반 학습·평

www.kci.go.kr

가 데이터 분할 과정, 그리고 SVM-PUK의 주요 커널 설정(C=1.0, Omega=1.0, Sigma=1.0)을 포함하였다. 이를 통해 특징 생성, 클래스 정의, 피험자 독립 데이터 분할 및 분류기 설정에 이르는 주요 실험 절차를 명확히 하였으며, 관련 핵심 구현 코드와 의사 코드는 표 5에 제시하였다.

Table 5. Key Implementation Code and Pseudocode

```
EEG Preprocessing and PSD Feature Extraction
FS = 256
EPOCH_SIZE = 3 * FS
epochs = segment_eeg(eeg_signal)
feature_vectors = []

for epoch in epochs:
    features = extract_psd_features(epoch)
    feature_vectors.append(features)
X = np.asarray(feature_vectors)
Arousal-Valence Label Generation
if arousal >= 5 and valence >= 5:
    label = "HAHV"
elif arousal >= 5 and valence < 5:
    label = "HALV"
elif arousal < 5 and valence >= 5:
    label = "LAHV"
else:
    label = "LALV"
LOSO Subject Split
for test_subject in subjects:
    train_subjects = [
        s for s in subjects if s != test_subject]

    train_data = data[data["subject"].isin(train_subjects)]
    test_data  = data[data["subject"] == test_subject]

    X_train = train_data[features]
    y_train = train_data["label"]

    X_test = test_data[features]
    y_test = test_data["label"]
PUK-SVM Configuration
puk_svm = SMO()

puk_kernel = PUKKernel(
    omega=1.0,
    sigma=1.0 )

puk_svm.set_kernel(puk_kernel)
puk_svm.set_C(1.0)
```

IV. Experiments and Results

1. Four-Class Emotion Classification Performance under 10-Fold Cross-Validation

본 연구에서는 제안한 SVM+PUK-LSTM 하이브리드 모델의 감정 분류 성능을 평가하기 위하여 HAHV(High Arousal-High Valence), HALV(High Arousal-Low Valence), LAHV(Low Arousal-High Valence), LALV(Low Arousal-Low Valence)의 네 가지 감정 클래스를 대상으로 실험을 수행하였다. 각 데이터에 대해 10-fold 교차검증(10-fold-cross-validation)을 적용하였으며, 단순 정확도뿐만 아니라 정밀도(Precision), 민감도(Recall), F1 스코어(F1-score)와 매크로 평균(Macro average), 마이크로 평균(Micro average) 및 가중치 평균(Weighted average), 균형 정확도(Balanced Accuracy)를 함께 산출하여 모델의 분류 성능을 다각적으로 분석하였다.

Table 6. Overall performance of the proposed model across 32 datasets

Evaluation Metric	Mean(%)	SD(%)	Mean ± SD(%)
Accuracy	98.42	3.82	98.42 ± 3.82
Macro Precision	97.84	6.03	97.84 ± 6.03
Macro Recall	97.39	6.32	97.39 ± 6.32
Macro F1-score	97.55	6.31	97.55 ± 6.31
Weighted Precision	98.44	3.54	98.44 ± 3.54
Weighted Recall	98.42	3.82	98.42 ± 3.82
Weighted F1-score	98.38	3.89	98.38 ± 3.89
Micro Precision	98.42	3.82	98.42 ± 3.82
Micro Recall	98.42	3.82	98.42 ± 3.82
Micro F1-score	98.42	3.82	98.42 ± 3.82
Balanced Accuracy	97.39	6.32	97.39 ± 6.32

표 6은 32개 데이터에 대한 SVM+PUK-LSTM 모델의 전체 성능을 평균과 표준편차(Mean±SD)로 나타낸 것이다. 제안한 모델은 평균 Accuracy 98.42±3.82%를 나타냈으며, Macro Precision 97.84±6.03%, Macro Recall 97.39±6.32%, Macro F1-score 97.55±6.31%를 나타냈다. Weighted Precision, Recall 및 F1-score는 각각 98.44±3.54%, 98.42±3.82%, 98.38±3.89%로 나타났다. 또한 Micro Precision, Recall 및 F1-score는 모두 98.42±3.82%였으며, Balanced Accuracy는 97.39±6.32%로 나타났다. 이러한 결과는 전체 정확도뿐만 아니라 클래스별 성능을 동일하게 반영하는 Macro 및 Balanced Accuracy에서도 높은 분류 성능이 유지되었음을 보여준다.

www.kci.go.kr

Table 7. Mean class-wise performance

Class	N	Precision(%)	Recall(%)	F1-score(%)
HAHV	32	95.31±18.16	96.26±17.64	95.74±17.85
HALV	31	99.31±1.82	97.51±8.26	98.18±5.45
LAHV	32	98.52±4.64	97.69±5.50	98.09±5.02
LALV	32	98.22±4.46	98.05±3.75	98.13±4.04

표 7은 네 감정 클래스별 Precision, Recall 및 F1-score의 평균 성능을 나타낸다. HAHV의 Precision, Recall 및 F1-score는 각각 95.31±18.16%, 96.26±17.64%, 95.74±17.85%였으며, HALV는 각각 99.31±1.82%, 97.51±8.26%, 98.18±5.45%로 나타났다. LAHV는 98.52±4.64%, 97.69±5.50%, 98.09±5.02%, LALV는 98.22±4.46%, 98.05±3.75%, 98.13±4.04%를 나타냈다. 특히 HALV, LAHV 및 LALV의 F1-score가 모두 약 98% 수준으로 나타나 세 클래스에서 안정적인 분류 성능을 확인하였다. 반면 HAHV는 평균 성능 자체는 높았으나 다른 클래스에 비하여 표준편차가 크게 나타나, 데이터별 HAHV 분류 성능의 변동성이 상대적으로 컸음을 확인할 수 있다. 또한 HALV는 한 데이터에서 해당 클래스의 실제 표본이 존재하지 않아 유효한 31개 데이터(N=31)를 기준으로 클래스별 성능을 산출하였다.

Table 8. Aggregated confusion matrix across the 32 datasets

Actual / Predicted	HAHV	HALV	LAHV	LALV	Total
HAHV	8,709	22	37	14	8,782
HALV	70	5,589	7	24	5,690
LAHV	48	7	5,013	65	5,133
LALV	46	8	41	4,876	4,971
Total	8,873	5,626	5,098	4,979	24,576

표 8은 32개 데이터에서 산출된 confusion matrix를 HAHV, HALV, LAHV, LALV의 동일한 클래스 순서로 정렬하여 통합한 결과이다. 총 24,576개 표본 가운데 HAHV는 8,782개 중 8,709개, HALV는 5,690개 중 5,589개, LAHV는 5,133개 중 5,013개, LALV는 4,971개 중 4,876개가 각각 올바르게 분류되었다. 통합 혼동행렬(confusion matrix)의 대각 성분에 대부분의 표본이 집중되어 있으므로 네 개 감정 클래스 전반에서 높은 분류 성능이 나타났음을 확인할 수 있다. 그러나 클래스별 표본 수가 HAHV 8,782개에서 LALV 4,971개까지 차이를 보이므로, 본 연구에서는 전체 Accuracy뿐만 아니라 클래스별 Precision, Recall, F1-score 및 Macro average를 함께 제시하여 클래스 분포의 영향을 고려하였다.

Table 9. Macro average, Micro average, and Weighted-average performance

Average method	Precision(%)	Recall(%)	F1-score(%)
Macro average	97.84±6.03	97.39±6.32	97.55±6.31
Micro average	98.42±3.82	98.42±3.82	98.42±3.82
Weighted average	98.44±3.54	98.42±3.82	98.38±3.89

마지막으로 표 9는 Macro average, Micro average 및 Weighted average에 따른 성능을 비교한 결과이다. Macro average의 Precision, Recall 및 F1-score는 각각 97.84±6.03%, 97.39±6.32%, 97.55±6.31% 였으며, Micro average는 세 지표 모두 98.42±3.82%로 나타났다. Weighted average에서는 Precision 98.44±3.54%, Recall 98.42±3.82%, F1-score 98.38±3.89%를 나타냈다. Macro average와 Weighted average가 모두 높은 수준을 나타낸 것은 제안한 모델의 성능이 특정 다수 클래스의 높은 분류율에만 의존하지 않고 네 감정 클래스 전반에서 비교적 균형 있게 유지되었음을 보여준다.

또한 10-fold-cross-validation 결과에서 제안한 SVM+PUK-LSTM 모델은 전체 Accuracy뿐만 아니라 Macro F1-score, Weighted F1-score 및 Balanced Accuracy에서도 높은 성능을 나타냈으며, HAHV, HALV, LAHV 및 LALV의 네 감정 클래스에 대한 분류 성능 역시 전반적으로 높은 수준을 유지하였다. 다만 10-fold 교차검증 결과는 동일한 평가 프로토콜을 적용한 실험 내에서 모델의 분류 성능을 평가하기 위한 결과이며, 피험자 간 일반화 성능을 직접적으로 의미하지 않는다. 이에 본 연구에서는 학습 데이터와 평가 데이터 간의 피험자 중복을 배제한 피험자 독립 평가 성능을 검증하기 위하여 추가적으로 LOSO(Leave-One-S ubject-Out) 평가를 수행하고, 그 결과를 별도로 분석하였다.

2. Subject-Independent Performance Using LOSO Evaluation

본 연구에서는 기존 10-fold cross-validation 결과에 추가하여 학습 데이터와 평가 데이터 간의 피험자 중복에 따른 데이터 누수를 방지하고 새로운 피험자에 대한 모델의 일반화 성능을 평가하기 위해 LOSO(Leave-One-Subject-Out) 기반의 피험자 독립(Subject-independent) 평가를 수행하였다. DEAP Dataset[11]의 전체 32명 피험자를 대상으로 각각의 fold에서 한 명의 피험자를 평가 데이

www.kci.go.kr

터로 완전히 분리하고 나머지 31명의 피험자를 학습 데이터로 사용하여 총 32-fold LOSO 평가를 수행하였다.

본 연구에서 제안한 SVM+PUK-LSTM 하이브리드 모델을 주 평가 대상으로 설정하고, 제안모델의 성능을 검증하기 위한 비교모델로 단일 LSTM과 SVM+RBF-LSTM을 사용하였다. 모델 간 공정한 비교를 위해 세 모델 모두 랜덤 시드(Random seed) 값을 1로 설정하였고, 동일한 LOSO 피험자 분할 조건을 적용하였다. 각 모델의 성능은 피험자별 평가 결과를 기준으로 Accuracy, Macro 및 Weighted Precision·Recall·F1-score, Micro Precision·Recall· F1-score와 Balanced Accuracy를 산출하고, 32명 피험자 결과의 평균과 표준편차(Mean±SD)로 나타내었다. 또한 동일한 32명의 피험자별 LOSO Accuracy를 이용한 대응표본 통계 비교(paired statistical comparison)을 수행하여 제안모델과 비교모델 간 성능 차이의 통계적 유의성을 검증하였다.

Table 10. Subject-independent LOSO performance comparison

Metric	LSTM	SVM+RBF_LSTM	SVM+PUK_LSTM (Proposed)
Accuracy	82.77 ± 11.32	80.08 ± 7.91	90.11 ± 7.91
Macro Precision	80.26 ± 17.53	76.87 ± 17.08	89.29 ± 11.67
Macro Recall	78.21 ± 14.85	73.18 ± 14.67	89.07 ± 9.56
Macro F1-score	77.26 ± 17.01	73.02 ± 15.76	88.38 ± 11.04
Weighted Precision	84.89 ± 11.72	81.56 ± 11.01	91.03 ± 10.31
Weighted Recall	82.77 ± 11.32	80.08 ± 7.91	90.11 ± 7.91
Weighted F1-score	82.13 ± 12.61	79.04 ± 9.74	89.87 ± 9.19
Micro Precision	82.77 ± 11.32	80.08 ± 7.91	90.11 ± 7.91
Micro Recall	82.77 ± 11.32	80.08 ± 7.91	90.11 ± 7.91
Micro F1-score	82.77 ± 11.32	80.08 ± 7.91	90.11 ± 7.91
Balanced Accuracy	78.21 ± 14.85	73.18 ± 14.67	89.07 ± 9.56

표 10은 동일한 랜덤 시드 값이 1인 조건에서 단일 LSTM, SVM+RBF-LSTM 및 제안한 SVM+PUK-LSTM의 피험자 독립 LOSO 성능을 비교한 결과이다. 제안한 SVM+PUK-LSTM은 평균 Accuracy 90.11±7.91%로 가장 높은 성능을 나타냈으며, 단일 LSTM은 82.77±11.32%, SVM+RBF-LSTM은 80.08±7.91%로 나타났다. 따라서 제안 모델은 동일한 LOSO 평가 조건에서 LSTM 보다 7.34%p, SVM+RBF-LSTM 보다 10.03%p 높은 평균 정확도를 나타냈다. Accuracy 이외의 평가 지표에서도 제안 모델의 높은 성능이 확인되었다. SVM+PUK-LSTM의 Macro Precision, Macro Recall 및 Macro F1-score는 각각 89.29±11.67%, 89.07±9.56%, 88.38±11.04%였으며, 단일 LSTM의 80.26±17.53%, 78.21±14.85%, 77.26± 17.01%와 SVM+RBF-LSTM의 76.87±17.08%, 73.18± 14.67%, 73.02±15.76%보다 높게 나타났다. Weighted F1-score에서도 제안 모델은 89.87±9.19%로 단일 LSTM의 82.13±12.61%와 SVM+RBF-LSTM의 79.04±9.74% 보다 높은 성능을 나타냈다. 또한 클래스별 Recall을 동일하게 반영하는 균형 정확도(Balanced Accuracy)는 제안 모델에서 89.07±9.56%로 도출된 반면 LSTM과 SVM+RBF-LSTM에서는 각각 78.21±14.85%와 73.18±14.67%로 나타났다. 이는 제안 모델의 성능 향상이 전체 Accuracy에만 국한되지 않고 클래스 간 균형을 고려한 평가에서도 유지되었음을 보여준다.

Table 11. Class-wise LOSO performance

Model	Class	Precision (%)	Recall (%)	F1-score (%)
LSTM	HAHV	85.42 ± 16.69	87.57 ± 15.20	84.40 ± 12.18
	HALV	69.75 ± 40.91	64.52 ± 39.60	66.05 ± 38.93
	LAHV	87.06 ± 16.94	82.94 ± 18.90	82.14 ± 14.28
	LALV	80.91 ± 22.86	77.99 ± 24.59	77.91 ± 21.50
SVM+RBF_LSTM	HAHV	83.60 ± 18.81	80.93 ± 21.79	80.88 ± 17.79
	HALV	72.34 ± 32.35	64.16 ± 34.78	64.72 ± 31.39
	LAHV	76.00 ± 22.97	71.47 ± 25.15	71.03 ± 19.74
	LALV	71.52 ± 28.96	70.10 ± 31.21	69.65 ± 28.39
SVM+PUK_LSTM	HAHV	91.57 ± 19.01	88.19 ± 19.48	89.25 ± 18.14
	HALV	89.38 ± 21.81	88.02 ± 21.52	88.13 ± 20.73
	LAHV	87.54 ± 19.29	85.47 ± 23.16	85.49 ± 19.78
	LALV	87.72 ± 14.79	95.25 ± 7.59	90.45 ± 10.65

표 11은 동일한 랜덤 시드 값이 1인 조건에서 세 모델의 HAHV, HALV, LAHV 및 LALV 클래스별 Precision, Recall 및 F1-score를 비교한 결과이다. 제안한 SVM+PUK-LSTM 하이브리드 모델은 HAHV에서 Precision 91.57±19.01%, Recall 88.19±19.48%,

www.kci.go.kr

F1-score 89.25±18.14%를 나타냈으며, HALV에서는 각각 89.38±21.81%, 88.02±21.52%, 88.13± 20.73%를 나타냈다. LAHV의 Precision, Recall 및 F1-score는 87.54±19.29%, 85.47±23.16%, 85.49± 19.78%를 보였다. LALV 에서는 각각 87.72±14.79%, 95.25±7.59%, 90.45±10.65%로 나타났다. 특히 HALV 클래스에서 단일 LSTM의 F1-score는 66.05±38.93%, SVM+RBF-LSTM은 64.72±31.39%였던 반면, 제안 모델은 88.13±20.73%를 나타냈다. LALV 에서도 제안 모델의 F1-score는 90.45±10.65%로 LSTM의 77.91±21.50%와 SVM+RBF-LSTM의 69.65±28.39% 보다 높게 나타났다. 또한 네 개 클래스 중에서 제안 모델의 LALV Recall의 결과가 95.25±7.59%로 가장 높았다. 이러한 결과는 피험자 독립 평가 환경에서 제안 모델이 특정 감정 클래스에만 높은 성능을 보이는 것이 아니라 HAHV, HALV, LAHV 및 LALV 전반에서 비교적 높은 분류 성능을 유지했음을 보여준다. 특히 상대적으로 분류 변동성이 크게 나타난 HALV 에서도 제안 모델의 Precision, Recall 및 F1-score가 비교 모델보다 높은 수준을 유지하였다.

Table 12. Additional analysis of LSTM performance across random seeds under repeated LOSO

Seed	Accuracy (%)	Macro F1 (%)	Weighted F1 (%)	Balanced Accuracy (%)	Accuracy 95% CI
1	82.77 ± 11.32	77.26 ± 17.01	82.13 ± 12.61	78.21 ± 14.85	78.69-86.85
42	84.90 ± 9.00	78.42 ± 15.62	83.77 ± 10.94	79.35 ± 13.69	81.65-88.14
3407	87.26 ± 8.15	81.85 ± 13.97	86.38 ± 9.44	82.25 ± 12.60	84.32-90.19

앞서 제시한 세 모델은 동일하게 랜덤시드 값 1을 사용하였고, LOSO 비교를 본 연구의 주 성능평가로 사용하였다. 이와 별도로 딥러닝 기반 비교모델인 단일 LSTM의 무작위 초기화(random initialization)에 따른 성능 변동을 확인하기 위한 보조 분석(additional analysis)으로 랜덤시드(Random seed) 값을 1, 42 및 3407로 변경하여 repeated LOSO 평가를 수행하였다. 각 시드(Seed)에서도 동일한 피험자 단위의 학습 및 평가 분할을 유지하였으며, Table 12의 결과는 제안 모델과의 LSTM 기준모델(Baseline)의 시드 변화에 따른 변동성을 확인하기 위한 보조 결과로 제시하였다. 표 12에 나타난 바와 같이 LSTM의 평균 Accuracy는 seed=1에서 82.77± 11.32%, seed=42에서 84.90±9.00%, seed=3407에서 87.26± 8.15%로 나타났다. Accuracy의 95% 신뢰 구간(Confidence interval)은 각각 78.69-86.85%, 81.65-88.14%, 84.32-90.19%로 나타났다. Macro F1-score는 각각 77.26±17.01%, 78.42±15.62%, 81.85±13.97%였으며, Weighted F1-score는 82.13±12.61%, 83.77±10.94%, 86.38±9.44%로 나타났다. Balanced Accuracy 역시 시드에 따라 78.21±14.85%, 79.35±13.69%, 82.25± 12.60%의 범위를 나타냈다. 따라서 랜덤 시드의 변화에 따라 LSTM의 분류 성능에 일정 수준의 변동이 존재하는 것을 확인하였다. 특히 세 시드의 Accuracy는 약 82.77-87.26% 범위로 나타나 단일 실행 결과만으로 모델 성능을 해석하는 것보다 반복 평가를 함께 고려할 필요가 있음을 보여준다. 세 개의 랜덤 시드에서 LSTM의 평균 Accuracy는 82.77-87.26% 범위로 나타나 초기화 조건에 따른 일정 수준의 성능 변동이 존재함을 확인하였다. 따라서 단일 실행 결과만으로 모델 성능을 해석하기보다 무작위 초기화에 따른 변동성을 함께 고려할 필요가 있음을 확인하였다.

Table 13. Aggregated confusion matrix of SVM+PUK_LSTM

Actual / Predicted	HAHV	HALV	LAHV	LALV	Total
HAHV	4,260	174	114	101	4,649
HALV	100	2,569	109	112	2,890
LAHV	186	9	2,319	154	2,668
LALV	31	32	78	2,128	2,269
Total	4,577	2,784	2,620	2,495	12,476

표 13에서 제안한 SVM+PUK-LSTM의 32개 LOSO 평가에서 얻어진 혼동행렬(Confusion matrix)를 HAHV, HALV, LAHV 및 LALV의 동일한 클래스 순서로 정렬하여 통합한 결과이다. 전체 12,476개 평가 표본 가운데 HAHV는 4,649개 중 4,260개, HALV는 2,890개 중 2,569개, LAHV는 2,668개 중 2,319개, LALV는 2,269개 중 2,128개가 각각 올바르게 분류되었다. 통합 Confusion matrix의 대각 성분에 대한 정분류 표본이 집중되어 있어 네 감정 클래스 전반에서 높은 분류 성능이 나타났음을 확인할 수 있다. 전체 정분류 표본은 11,276개로, 통합 Confusion matrix를 기준으로 산출한 전체 정확도는 약 90.38%로 나타났다. 특히 LALV는 2,269개 중 2,128개가 올바르게 분류되어 높은 분류율을 나타냈으며, 이는 앞서 제시한 클래스별 Recall 결과와도 일관된 경향을 보였다.

www.kci.go.kr

Table 14. Statistical comparison of LOSO accuracy between the proposed and comparison models

Comparison	Mean Difference (%p)	Test	p-value	Holm-adjusted p	Effect Size rrbr_{rb}	95% CI of Mean Difference (%p)
SVM+PUK_LSTM vs LSTM	7.34	Wilcoxon	0.00123	0.00245	0.687	3.25–11.44
SVM+PUK_LSTM vs SVM+RBF_LSTM	10.03	Wilcoxon	0.000089	0.000266	0.806	5.88–14.17
LSTM vs SVM+RBF_LSTM	2.68	Wilcoxon	0.12396	0.12396	0.317	−2.03–7.39

마지막으로, 제안한 SVM+PUK-LSTM의 성능 향상이 비교모델과 통계적으로 유의한 차이를 나타내는지 검증하기 위하여, 동일한 32명 피험자의 LOSO Accuracy를 대응시켜 대응표본 통계 비교(Paired statistical comparison)을 수행하였다. 이를 위해 제안모델과 단일 LSTM, 제안모델과 SVM+RBF-LSTM의 비교를 포함하고, 추가적으로 두 비교모델 간 차이도 분석하였다. 모델 간 동일 피험자의 정확도 차이(Paired accuracy difference)의 정규성을 검토한 결과 일부 비교에서 정규성 가정이 충족되지 않아 비모수 검정인 Wilcoxon 부호순위 검정(Wilcoxon signed-rank test)을 적용하였다. 또한 세 모델 간 다중 짝별 비교(Pairwise comparison)에 따른 오류 증가를 통제하기 위해서 Holm-Bonferroni 보정(Correction)을 적용하였다.

표 14에서 제시한 바와 같이 제안한 SVM+PUK-LSTM은 LSTM보다 평균 Accuracy가 7.34%p 높았으며, Wilcoxon 부호순위 검정에서 통계적으로 유의한 차이를 나타냈다(p=0.00123). Holm-Bonferroni 보정 후에도 유의성이 유지되었으며(p_{adj}=0.00245), 순위 이연 상관계수(Rank-biserial correlation)은 0.687로 나타났다. 평균 대응 정확도 차이(Paired accuracy difference)의 95%, 신뢰 구간(Confidence interval)은 3.25-11.44%p로 나타났다. SVM+PUK-LSTM과 SVM+RBF-LSTM 간 평균 Accuracy 차이는 10.03%p 였으며, 이 차이 역시 통계적으로 유의하였다(p=0.000089, p_{adj}=0.000266). 순위 이연 상관계수는 0.806으로 나타났으며, 평균 차이의 95% Confidence interval은 5.88-14.17%p 이다. 반면 LSTM과 SVM+RBF-LSTM의 평균 Accuracy 차이는 2.68%p였으나, Holm-Bonferroni 보정 후에 통계적으로 유의하지 않았다(p_{adj}=0.12396). 해당 비교의 효과 크기(Effect size)는 0.317, 평균 차이의 95% Confidence interval은 2.03-7.39%p로 0을 포함하였다. 통계적 검정 결과, 동일한 피험자 독립(Subject-independent) LOSO 평가 조건에서 제안한 SVM+PUK-LSTM은 단일 LSTM 및 SVM+RBF-LSTM과 비교하여 통계적으로 유의하게 높은 Accuracy를 나타냈다. 특히 PUK 커널을 RBF 커널로 대체한 하이브리드 모델과의 비교에서도 통계적으로 유의한 성능 차이가 확인되어, PUK 커널 적용에 따른 성능 차이를 확인할 수 있었다. 반면 단일 LSTM과 SVM+RBF-LSTM 간 차이는 통계적으로 유의하지 않아, 단순한 평균 성능 차이만으로 두 비교모델 간 우열을 판단하기는 어려운 것으로 나타났다.

제안한 SVM+PUK-LSTM은 10-fold 교차검증에서 높은 분류 성능을 나타냈을 뿐만 아니라, 학습 및 평가 데이터 간 피험자 중복을 배제한 LOSO 평가에서도 평균 Accuracy 결과는 90.11±7.91%를 나타냈다. LOSO에서는 10-fold 교차검증에 비해 성능이 감소하였으며, 이는 새로운 피험자에 대한 일반화 평가가 샘플 단위(Sample-level) 교차검증보다 더 엄격한 평가 조건임을 보여준다. 그럼에도 제안모델은 동일한 시드(Seed)와 LOSO 조건에서 단일 LSTM 및 SVM+RBF-LSTM보다 높은 Accuracy, Macro F1-score, Weighted F1-score 및 Balanced Accuracy를 나타냈으며, 대응표본 통계 비교에서도 두 비교모델에 대한 성능 차이가 통계적으로 유의하였다. 또한 단일 LSTM의 랜덤 시드(Random seed)를 변경한 반복 LOSO(repeated LOSO) 결과에서 일정 수준의 성능 변동이 확인됨에 따라, 본 연구에서는 단일 평가 결과뿐만 아니라 피험자 독립 평가, 반복 실험 및 통계적 검정을 함께 고려하여 모델의 일반화 성능과 결과의 신뢰성을 검증하였다.

3. Comparison of Experimental Protocols and Performance with Previous and Recent Studies

기존 선행연구의 실험 조건을 분석한 결과, 각 연구는 감정 클래스 정의, 피험자 분할 방식, 입력 특징 및 평가 프로토콜에서 차이를 보였다. 특히 높은 분류 정확도를 보고한 일부 연구는 통합 샘플(Pooled sample) 기반의 10-fold-cross-validation을 사용한 반면, 새로운 피험자

www.kci.go.kr

에 대한 일반화 성능을 평가한 연구에서는 LOSO 기반 평가가 적용되었다. 따라서 서로 다른 평가 프로토콜에서 보고된 Accuracy를 동일한 조건의 직접적인 성능 비교로 해석하는 데에는 제한이 있다.

최근에는 피험자 간 EEG 분포 차이를 고려하기 위하여 그래프 신경망(Graph Neural Network, GNN), 어텐션 메커니즘(Attention mechanism), 트랜스포머(Transformer) 및 도메인 적응(Domain adaptation)을 활용한 피험자 독립 (Subject-independent) 감정인식 연구가 활발하게 수행되고 있다. 이에 본 연구에서는 기존 선행연구 중에서 DEAP Dataset[11]을 사용한 연구를 추가하여 피험자 간(Cross-subject) 평가 또는 피험자 독립(Subject-independent) 평가를 수행한 결과를 비교하였다. 최근 연구에서는 EEG 채널 간 공간적 관계를 모델링하기 위한 그래픽 합성곱(Graph convolution) 및 그래프 어텐션(Graph attention) 구조와 함께 채널/주파수 어텐션(Channel/frequency attention), 트랜스포머(Transformer) 기반 셀프 어텐션(Self-attention) 및 도메인 적응(Domain adaptation) 기법이 적용되고 있다. DGAT(Dynamic Graph Attention Network)[18]는 피험자 독립(Subject-independent) 평가에서 Valence 84.27±13.74%, Arousal 83.84±10.20%를 보고하였으며, FBAGAN(Frequency Band Attention Graph Convolutional Adversarial Network)[20]은 피험자 간(Cross-subject) 평가에서 Arousal 69.64%, Valence 65.18%를 나타냈다.

또한 MS-AGDA[21]는 다중 소스 도메인(Multi-source domain adaptation)을 적용하여 DEAP cross-subject 평가에서 65.59±5.50%를 보고하였다. 표 15의 결과는 기존 EEG 기반 감정인식 연구와 본 연구의 실험 조건 및 분류 성능을 비교한 결과이다. 기존 연구들은 감정 클래스 정의, 피험자 분할 방식, 교차검증 방법 및 입력 특징에서 서로 다른 실험 조건을 적용하고 있으며, 이에 따라 보고된 Accuracy 역시 서로 다른 평가 환경에서 산출되었다. 특히 일부 연구는 통합 표본(Pooled sample) 기반의 10-fold-cross-validation을 적용한 반면, Jirayu[22]와 Yao[23] 등은 피험자 간 일반화 성능을 고려한 평가를 수행하였다. 본 연구에서는 DEAP Dataset[11]의 32명 피험자를 대상으로 HAHV, HALV, LAHV 및 LALV의 4-class 분류를 정의하고 32-fold LOSO를 적용하여 피험자 독립 성능을 평가하였다. 따라서 표 15의 성능 값은 단순한 Accuracy의 직접 비교보다는 각 연구의 클래스 정의(Class definition)와 평가 프로토콜(Evaluation protocol)의 차이를 함께 고려하여 해석할 필요가 있다.

Table 15. Comparison of experimental protocols and performance with previous EEG-based emotion recognition studies

Method	Class definition	Subject protocol	Subjects	Validation	Input features	Result type	Accuracy (%)
DLN(SAE)+ PCA+CSA, 2014 [22]	V: Negative/Neutral/ Positive; A: Passive/Neutral/ Active	Independent	32	LOSO (31 train + 1 test)	PSD + hemispheric differences, PCA	5 repetitions; subject average	V 53.42 A 52.03
Merged LSTM, 2019 [24]	V/A/D/L binary	Not clearly reported as strict subject-independent	32	Reported evaluation	DWT + statistical features	Reported accuracy	V 84.89 A 83.85 D 84.37 L 80.72
CNN-HMMs, 2020 [15]	V/A/D High-Low binary	Pooled-segment evaluation	32	10-fold CV	MSE from θ/α/β/γ → spatial representation	10-fold Mean±SD	V 83.09±0.84 A 79.77±0.61 D 81.83±0.53
Multi-scale CNN, 2021 [25]	V/A binary + AV 4-class	Pooled-sample evaluation	32	10-fold CV	Multi-scale spatial/frequency EEG representation	10-fold average	V 98.36 A 98.27 4-class 98.38
Feature-Fused CNN (DBCN), 2022 [23]	V/A binary	Dependent & Independent	32	Subject-specific / subject-independent	Spatial-spectral EEG representation	Mean±SD	Independent: V 83.98±13.20 A 79.45±10.60
SVM+PUK_ LSTM, Proposed	AV 4-class: HAHV/HALV/LAHV/LALV	Subject-independent	32	32-fold LOSO	PSD, 32 ch × 5 bands = 160	32-subject Mean±SD	**90.11±7.91**

www.kci.go.kr

Table 16. Comparison with recent subject-independent EEG emotion recognition methods on DEAP

Study	Method / Category	Class definition	Subject protocol	Subjects	Validation	Input / Key feature	Accuracy (%)
DBCN, 2022 [23]	Feature-fused CNN	V/A binary	Independent	32	Subject-independent	Spatial-spectral representation	V 83.98±13.20 A 79.45±10.60
TSGCN, 2024 [21]	Temporal-spectral GCN + adversarial learning	V/A binary	Independent	32	LOOCV/LOSO	Temporal-spectral + dynamic graph features	V / A Mean±SD reported
DAMGCN, 2024 [26]	GCN + Dual Attention + Transformer self-attention	V/A/D labels	Independent	32	LOSO	DE + graph topology + channel/frequency attention	Subject-independent DEAP results
FBAGAN, 2024 [20]	GCN + Frequency-band Attention + Adversarial DA	V/A binary	Cross-subject	32	Cross-subject	Frequency-band + graph/channel relation	**A 69.64 V 65.18**
MSGDAN, 2024 [27]	Graph Neural Network + Multi-source Domain Adaptation	Cross-subject emotion	Cross-subject	DEAP	Cross-subject	Dynamic graph + functional connectivity + domain adaptation	DEAP cross-subject reported
DGAT, 2025 [19]	Dynamic Graph Attention Network	V/A binary	Independent	32	Subject-independent	DE + dynamic graph + multi-head attention	**V 84.27±13.74 A 83.84±10.20**
MS-AGDA, 2025 [21]	Multi-source unsupervised Domain Adaptation	Valence High/Low	Cross-subject	15 of 32	LOOCV	DE, 160-d input + adaptive gating/domain alignment	65.59±5.50
Proposed	SVM+PUK_LSTM	AV 4-class: HAHV/HALV/ LAHV/LALV	Independent	32	32-fold LOSO	PSD, 32 ch × 5 bands = 160	**90.11±7.91**

표 16에서는 최근 DEAP Dataset를 활용하여 피험자 독립(Subject-independent) 또는 피험자 간(Cross-subject) 평가를 수행한 EEG 감정인식 연구와 본 연구의 결과를 비교하였다. 최근 연구에서는 피험자 간 EEG 신호의 차이를 고려하기 위해 그래프 신경망(Graph Neural Network, GNN), 어텐션 메커니즘(Attention mechanism), 트랜스포머(Transformer) 및 도메인 적응(Domain adaptation)과 같은 방법이 활발하게 적용되고 있다. DGAT(Dynamic Graph Attention Network)[18]는 동적 그래프(Dynamic graph)와 멀티헤드 어텐션(Multi-head attention)을 적용하여 Valence 84.27±13.74%, Arousal 83.84±10.20%를 보고하였으며, FBAGAN[20]은 그래프 어텐션(Graph convolution), 주파수 대역 어텐션(Frequency-band attention) 및 적대적 도메인 적응(Adversarial domain adaptation)을 결합하여 Arousal 69.64%, Valence 65.18%를 나타냈다.

또한 MS-AGDA[21]는 다중 소스 도메인 적응(Multi-source domain adaptation)을 이용한 피험자 간(Cross-subject) 평가에서 65.59±5.50%를 보고하였다.

본 연구의 SVM+PUK-LSTM은 32명의 피험자를 대상으로한 4-class LOSO 평가에서 90.11±7.91%의 Accuracy를 나타냈다. 다만 최신 연구의 상당수가 Valence 또는 Arousal의 이진 분류를 수행하고 있으며, 본 연구의 Arousal-Valence 4-class 분류와 과제 정의가 다르므로, 표 16의 결과는 동일 조건의 직접적인 성능 우열보다는 최근 피험자 독립 (Subject-independent) 감정인식 방법의 평가 경향을 비교하기 위한 자료로 해석하였다.

V. Conclusions

본 연구에서는 EEG 기반 감정인식의 분류 성능과 새로운 피험자에 대한 일반화 성능을 향상시키기 위해 SVM+PUK-LSTM 하이브리드 모델을 제안하였다. DEAP Dataset의 EEG 신호를 256Hz로 리샘플링 한 후 Theta, Slow Alpha, Alpha, Beta 및 Gamma의 5개 주파수 대역으로 분리하고, CleanLine과 ICA를 이용하여 잡음 및 인공산물 성분을 처리하였다. 이후 3초 단위의 비중첩 에포크(Non-overlapping epoch)로 분할하고, 32개 EEG 채널과 5개 주파수 대역의 PSD를 이용하여 160차원 특징벡터를 구성하였다. 감정 상태는 Arousal과 Valence를 기준으로 HAHV, HALV, LAHV 및 LALV의 4-class로

www.kci.go.kr

정의하였다.

제안모델의 10-fold 교차검증 결과 Accuracy는 98.42±3.82%, Macro F1-score는 97.55±6.31%로 나타났다. 또한 피험자 간 데이터 중복을 배제한 32-fold subject-independent LOSO 평가에서는 Accuracy 90.11±7.91%, Macro F1-score 88.38±11.04%, Weighted F1-score 89.87±9.19%, Balanced Accuracy 89.07±9.56%를 나타냈다. 시드(Seed)값을 동일하게 1로 설정한 LOSO 조건에서 단일 LSTM과 SVM+RBF-LSTM의 Accuracy는 각각 82.77±11.32%와 80.08±7.91%로 나타났다. Wilcoxon 부호순위 검정과 Holm-Bonferroni 보정을 적용한 결과, 제안모델은 단일 LSTM(p_{adj}=0.00245) 및 SVM+RBF-LSTM(p_{adj}=0.000266)과 통계적으로 유의한 차이를 나타냈다.

동일한 시드 값(1) 및 LOSO 조건에서 단일 LSTM과 SVM+RBF-LSTM을 비교모델로 적용하였으며, 대응표본 통계 비교(Paired statistical comparison)을 통해 제안한 SVM+PUK-LSTM과 비교모델 간의 성능 차이의 통계적 유의성을 검증하였다.

이러한 결과를 통해 제안모델은 높은 10-fold-cross-validation 성능뿐만 아니라 새로운 피험자를 대상으로 한 피험자 독립(Subject-independent) 환경에서도 높은 4-class 감정 분류 성능을 유지하였으며, 동일 조건의 비교모델과의 통계적 검증을 통해 PUK 커널을 적용한 하이브리드 구성의 성능 기여를 확인하였다. 다만 선행연구들은 클래스 정의(Class definition)과 피험자 분할(Subject split) 및 평가 프로토콜(Evaluation protocol)이 서로 다르므로 Accuracy의 직접적인 수치 비교에는 제한이 있으며, 본 연구에서는 이러한 실험 조건의 차이를 함께 고려하여 결과를 분석하였다.

최근 EEG 감정인식 연구에서는 GNN(Graph Neural Network)을 이용한 채널 간 공간 관계 학습, Attention 및 Transformer 기반 특징 학습, 도메인 적응(Domain adaptation)을 활용한 피험자 간(Cross-subject) 분포 차이 완화가 주요 연구 방향으로 활용되고 있다. 향후 연구에서는 본 연구의 한계와 최근 EEG 감정인식 연구 동향을 고려하여 다음과 같은 방향으로 연구를 확장하고자 한다. 첫째, 현재 사용한 채널별 주파수 대역 특징에서 나아가 EEG 전극 간 공간적 관계와 피험자 간 신호 분포 차이를 반영할 수 있도록 그래프 기반 표현(Graph-based representation), Attention 및 도메인 적응 기법을 SVM+PUK-LSTM 구조와 결합하고자 한다. 둘째, 제안모델의 성능을 보다 엄격하게 검증하기 위해 CNN(Convolutional Neural Network), BiLSTM (Bidirectional Long Short-Term Memory) 및 Transformer 기반 모델을 동일한 4-class 정의와 동일한 Subject-independent LOSO protocol에서 비교하여 모델 구조에 따른 일반화 성능 차이를 분석할 예정이다. 셋째, DEAP 뿐만 아니라 DREAMER, SEED-IV 및 AMIGOS 등 서로 다른 EEG 데이터셋에 동일한 전처리 및 피험자 독립 평가 절차를 적용하여 데이터셋 변화에 따른 제안모델의 재현성과 범용성을 검증하고자 한다.

VI. Generative AI Ethics and Disclosure

본 연구에서는 그림 1, 2, 3 및 4의 도식 초안을 작성하기 위한 보조적 도구로 생성형 AI(OpenAI)를 활용하였다. 생성된 도식은 실제 연구 방법 및 실험 절차와의 일치성을 확보하기 위하여 저자가 직접 검토, 수정 및 검증하였다. 생성형 AI는 실험에 사용된 EEG 데이터 및 실험 결과의 생성이나 조작에는 사용되지 않았으며, 최종적으로 논문에 제시된 그림과 연구 내용의 정확성 및 연구윤리에 대한 책임은 저자에게 있다.

ACKNOWLEDGEMENT

This work was supported by project for Collabo R&D between Industry, University, and Research Institute funded by Korea Ministry of SMEsand Startups in 2025. (RS-2025-02316177)

REFERENCES

[1] W. Ma, Y. Zheng, et al., "A Comprehensive Review of Deep Learning in EEG-Based Emotion Recognition: Classifications, Trends, and Practical Implications," PeerJ Computer Science, Vol. 10, Article e2065, May 2024. DOI: 10.7717/peerj-cs.2065.

[2] M. Jafari, A. Shoeibi, et al., "Emotion Recognition in EEG Signals Using Deep Learning Methods: A Review," Computers in Biology and Medicine, Vol. 165, Article 107450, October 2023. DOI: 10.1016/j.compbiomed.2023.107450.

[3] S. Alhagry, A. A. Fahmy, and A. El-Khoribi, "Emotion Recognition Based on EEG Using LSTM Recurrent Neural Network," International Journal of Advanced Computer Science and

www.kci.go.kr

Applications (IJACSA), Vol. 8, No. 10, pp. 355-358, 2017. DOI: 10.14569/IJACSA.2017.081046.

[4] I. Aliyu, and C. G. Lim, "EEG Dimensional Reduction with Stack AutoEncoder for Emotional Recognition Using LSTM/RNN," Journal of the Korea Institute of Electronic Communication Sciences, Vol. 15, No. 4, pp. 717-724, 2020.

[5] S. G. Kim, T. S. Kim, and W. H. Lee, "EEG-Based Emotion Recognition Using a 3D Convolutional Neural Network with a Bottleneck Structure," KIISE Transactions on Computing Practices, Vol. 29, No. 5, pp. 251-256, 2023.

[6] C. E. Woo, H. S. Choi, and B. H. Kim, "Valence-Arousal Regression Model Using EEG with Multi-Output Prediction," Journal of Biomedical Engineering Research, Vol. 45, No. 5, pp. 279-285, 2024.

[7] J. A. Russell, "A Circumplex Model of Affect," Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980. DOI: 10.1037/h0077714.

[8] J. N. Njoku, A. C. Caliwag, et al., "Deep Learning Based Data Fusion Methods for Multimodal Emotion Recognition," Journal of KICS, Vol. 47, No. 1, pp. 79-87, 2022.

[9] H. W. Kang, and B. H. Kim, "ConTL: Improving the Performance of EEG-Based Emotion Recognition via the Incorporation of CNN, Transformer and LSTM," KIISE Transactions, Vol. 51, No. 5, 2024.

[10] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition," IEEE Transactions on Affective Computing, Vol. 13, No. 2, pp. 568-578, 2022.

[11] DEAP Dataset, "DEAP: A Database for Emotion Analysis Using Physiological Signals," https://www.eecs.qmul.ac.uk/mmv/datasets/deap/, [Accessed: Jun. 27, 2026].

[12] EEGLAB, "EEGLAB Toolbox," *https://sccn.ucsd.edu/eeglab/*, [Accessed: Jun. 27, 2026].

[13] S. Koelstra, et al., "DEAP: A Database for Emotion Analysis Using Physiological Signals," IEEE Transactions on Affective Computing, Vol. 3, No. 1, pp. 18-31, 2012. DOI: 10.1109/T-AFFC.2011.15.

[14] D. Nguyen, M. T. Nguyen, and K. Yamada, "Electroencephalogram-Based Emotion Recognition Using Hybrid Intelligent Method and Discrete Wavelet Transform," Applied Sciences, Vol. 15, No. 5, Article 2328, 2025. DOI: 10.3390/app15052328.

[15] Q. Zhong, Y. Zhu, D. Cai, L. Xiao, and H. Zhang, "Electroencephalogram Access for Emotion Recognition Based on a Deep Hybrid Network," Frontiers in Human Neuroscience, Vol. 14, Article 589001, 2020. DOI: 10.3389/fnhum.2020.589001.

[16] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed., Morgan Kaufmann, 2017.

[17] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing Real-Time Emotional Responses: A Personalized Assessment Based on Heartbeat Dynamics," Scientific Reports, Vol. 4, Article 4998, May 2014. DOI: 10.1038/srep04998.

[18] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, NY, USA, 1995.

[19] S. Ding, K. Wang, W. Jiang, C. Xu, H. Bo, L. Ma, and H. Li, "DGAT: A Dynamic Graph Attention Neural Network Framework for EEG Emotion Recognition," Frontiers in Psychiatry, Vol. 16, Article 1633860, 2025. DOI: 10.3389/fpsyt.2025.1633860.

[20] W. Chen, Y. Liao, R. Dai, Y. Dong, and L. Huang, "EEG-Based Emotion Recognitio n Using Graph Convolutional Neural Network With Dual Attention Mechanism," Frontiers in Computational Neuroscience, Vol. 18, Article 1416494, 2024. DOI: 10.3389/fncom.2024.1416494.

[21] J. Liu, Y. Ma, M. Houston, S. Gao, and Y. Zhang, "MS-AGDA: A Multi-Source Adaptive Gating Unsupervised Domain Adaptation for Cross-Subject and Cross-Session EEG-Based Emotion Recognition," Signal, Image and Video Processing, Vol. 19, Article 1297, 2025. DOI: 10.1007/s11760-025-04819-9.

[22] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component-Based Covariate Shift Adaptation," The Scientific World Journal, Vol. 2014, Article 627892, 2014. DOI: 10.1155/2014/627892.

[23] Q. Yao, H. Gu, S. Wang, and X. Li, "A Feature-Fused Convolutional Neural Network for Emotion Recognition from Multichannel EEG Signals," IEEE Sensors Journal, Vol. 22, No. 12, pp. 11954-11964, 2022. DOI: 10.1109/JSEN.2022.3172133.

[24] A. Garg, A. Kapoor, A. K. Bedi, and R. K. Sunkaria, "Merged LSTM Model for Emotion Classification Using EEG Signals," Proceedings of the 2019 International Conference on Data Science and Engineering (ICDSE), pp. 139-143, Kochi, India, 2019. DOI: 10.1109/ICDSE47409.2019.8971484.

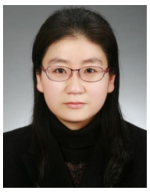
[25] T.-D.-T. Phan, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "EEG-Based Emotion Recognition by Convolutional Neural Network with Multi-Scale Kernels," Sensors, Vol. 21, No. 15, Article 5092, 2021. DOI: 10.3390/s21155092.

[26] S. Chen, Y. Wang, X. Lin, X. Sun, W. Li, and W. Ma, "Cross-Subject Emotion Recognition in Brain-Computer Interface Based on Frequency Band Attention Graph Convolutional Adversarial Neural Networks," Journal of Neuroscience Methods, Vol. 411, Article 110276, 2024. DOI: 10.1016/j.jneumeth.2024.110276.

[27] J. Wang, X. Ning, W. Xu, Y. Li, Z. Jia, and Y. Lin, "Multi-Source Selective Graph Domain Adaptation Network for Cross-Subject EEG Emotion Recognition," Neural Networks, Vol. 180, Article 106742, 2024. DOI: 10.1016/j.neunet.2024.106742.

www.kci.go.kr

Authors



Hyun-Ju Lee received the Doctor of Engineering degree in Computer Engineering from Sejong University, Seoul, South Korea, in 2019. Dr. Lee served as a Part-Time Lecturer at University College, Yonsei University, Seoul, South Korea, in 2021, the Department of Artificial Intelligence Applications, Seoul National University of Science and Technology, Seoul, South Korea, in 2025, and the Department of Biosystems Engineering, College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, in 2026. Her research interests include artificial intelligence, biosignal processing, EEG signal processing, emotion recognition, machine learning, and deep learning.

Dong-Il Shin received the Ph.D. degree in Computer Science from the University of North Texas, Denton, TX, USA, in 1997. Dr. Shin joined the faculty of the Department of Computer Engineering at Sejong University, Seoul, South Korea, in 1998. He is currently a Professor in the Department of Computer Engineering at Sejong University. His research interests include information security, machine learning, data mining, and biometric data processing.

www.kci.go.kr