JESK J Ergon Soc Korea 2026; 45(1): 1-12
https://dx.doi.org/10.5143/JESK.2026.45.1.1
https://jesk.or.kr eISSN:2093-8462

A Study on a Nonverbal Signal-Based Analysis for Assessing the User Satisfaction

Seong-Taek Kim, Sang-Ho Kim

Kumoh National Institute of Technology, Department of Industrial Engineering, Gumi, 39177

비언어적 신호 분석 기반의 사용자 만족도 평가 방안 연구

김성택, 김상호

금오공과대학교 산업공학과

Corresponding Author
Sang-Ho Kim
Kumoh National Institute of Technology, Department of Industrial Engineering, Gumi, 39177
Email : kimsh@kumoh.ac.kr

Received : October 31, 2025
Revised : January 06, 2026
Accepted : March 03, 2026

Copyright@2026 by The Human Factors & Ergonomics Society of Korea. All right reserved.

Objective: This study explores the feasibility and responsiveness of a multimodal approach to estimate user satisfaction by analysing physiological and non-verbal behavioural signals collected in natural, unguided interaction conditions.

Background: Traditional satisfaction assessments have primarily relied on self-report questionnaires; however, this method fails to capture subtle emotional shifts or real-time satisfaction changes occurring during interactions with AI systems. Furthermore, it suffers from limitations in fully reflecting users' actual emotional states due to the influence of recall bias and self-censorship.

Method: This study utilized a multimodal dataset comprising EEG, ECG, PPG, GSR, and facial expression data collected during spontaneous voice-based human–AI interactions. After preprocessing and feature extraction, Support Vector Machine (SVM) models were independently trained under two classification settings. First, user satisfaction was classified into four discrete states defined by the joint combination of stability and favorability. Second, separate binary classification models were trained to predict stability and favorability individually, using the same dataset and cross-validation protocol but optimized for each classification task.

Results: In four-class satisfaction classification, EEG-based models achieved the highest average accuracy (68.8%), followed by facial expression-based models (67.4%) and physiological signal-based models (65.1%). In the independently trained binary classification analyses, substantially higher performance was observed across all modalities. EEG-based models achieved F1-scores of 86.6% for stability and 85.5% for favorability, while facial models achieved F1-scores of 86.2% and 84.4%, respectively. Physiological signals showed comparable performance for stability (85.7%) but lower F1-scores for favorability (81.7%) under natural interaction conditions.

Conclusion: Our results demonstrate that user satisfaction can be estimated continuously in real-time from non-invasive signals, and offer real promise for the development of adaptive and emotionally responsive interaction systems.

Application: The proposed approach contributes to the advancement of real-time multimodal satisfaction estimation models and provides useful insights for the design of emotionally adaptive human-AI interaction systems.

Keywords: User satisfaction, Nonverbal signal, Evaluation methodology, Human AI interaction

1. Introduction

최근 인간공학 및 사용자 경험(UX) 분야에서는 제품이나 서비스의 사용 과정에서 발생하는 감정 반응을 기반으로 사용자 만족도(User Satisfaction)를 정량적으로 평가하려는 시도가 활발히 이루어지고 있다(Picard, 1997). 그러나 기존의 만족도 평가는 주로 설문조사나 인터뷰와 같은 정성적 방법에 의존해 왔으며, 이로 인해 실시간 감성 변화나 미묘한 만족도 차이를 포착하기 어렵다는 한계가 존재한다(Shin et al., 2021). 또한 이런 사용자의 언어 기반 응답은 응답자의 자기 검열(Self-censorship)이나 기억 왜곡(Recall bias)이 작용할 수 있다(Bolger et al., 2003; Robinson and Tannenberg, 2019).

이러한 한계를 극복하고자 표정, 시선, 제스처, 몸짓 등의 비언어적 행동 단서(Nonverbal behavioral cues)나 EEG, ECG, PPG, GSR 등과 같은 생리 신호(Physiological signals)를 분석함으로써 실제 상호작용 중 사용자의 감정 상태를 정량적으로 평가하려 시도하고 있다.

생리 신호와 비언어적 행동 단서를 단일 모달리티가 아닌 멀티모달로 이용해 접근하는 연구도 활발히 진행되고 있다. Fu et al. (2022)은 EEG, 호흡(RES), PPG, GSR 신호를 동시에 활용한 멀티모달 융합과 전이 학습 기반 도메인을 통해 감정 인식 정확도를 향상시켰다. Saffaryazdi et al. (2022)은 얼굴 미세표정을 생리 신호와 결합해 감정 추정의 신뢰도를 높이자 제안하였다. Apicella et al. (2021)은 시각 자극(OASIS dataset)을 활용한 유도형 실험에서 EEG 기반 감정 인식 정확도를 검증하였는데, 8채널 건식 전극 시스템을 이용해 개인 내 96.1%, 교차 피험자 간 80.2%의 높은 정확도로 정서적 Valence(긍정/부정)를 분류하였다. 또한 Zhuang et al. (2018)은 영화 영상을 통해 여섯 가지 정서를 유도한 후, 참가자가 해당 장면을 회상하도록 설계하여 자기 유도 감정을 분석하였다. 그 결과, EEG 기반으로 긍정/부정 감정 이진 분류에서 87.36%, 여섯 범주 감정 분류에서는 54.52%의 평균 정확도를 달성하였다.

기존 연구들은 주로 영상, 음성, 이미지 등의 인위적인 자극에 대한 반응을 유도한 다음, 이러한 반응에서 발생한 신호들을 분석하는 데 집중해 왔지만(Calvo and D'Mello, 2010), 이러한 통제된 환경은 실제 서비스 사용 상황에서 발생하는 자연스러운 감정의 변화를 충분히 반영하지 못한다. 이러한 한계를 극복하기 위해 비유도(Non-induced) 상황에서 생리 및 행동 신호를 활용한 연구가 제안되고 있으며, 실제 상호작용 맥락에서 감정의 자연적 변화를 탐색하려는 시도가 이루어지고 있다(McDuff et al., 2019). Zhao et al. (2023)은 사용자-AI 상호작용의 복잡한 맥락을 반영하기 위해 데이터 수집의 생태학적 타당성(Ecological validity)을 강화하는 방향이 중요하다고 지적하고 멀티모달 감정 인식 연구들은 통제된 환경(Lab-controlled setting)을 넘어 실세계 상호작용 환경으로 확장되어야 한다고 강조하고 있다.

따라서 본 연구는 비유도 환경에서 수집된 다양한 비언어적 모달리티 신호를 기반으로 사용자 만족도를 예측하고, 실제 상호작용 맥락에서의 실시간 만족도 추정 가능성을 검증함으로써, 향후 감성 적응형 설계(Emotionally Adaptive Design)의 기반을 마련하고자 한다.

2. Method

2.1 Dataset

본 연구는 Choi et al. (2024)이 구축한 "EEG Dataset for the Recognition of Different Emotions Induced in Voice-User Interaction"을 사용하였다. 이 데이터셋은 음성 기반 인공지능과의 실제 대화 맥락에서 EEG, ECG, PPG, GSR, Facial 신호를 동시 측정한 것으로, 인위적인 감정 유도 없이 자연스러운 상호작용 중에 발생한 감정 반응을 반영한다. 참가자들은 미리 정해진 감정을 느끼도록 요구 받지 않았으며, 일상적인 대화처럼 음성 명령을 하고 AI의 응답을 들으며 실험이 진행되었다. 이렇게 해서 얻은 데이터는 현실적인 사용자 경험에서 자연스럽게 생기는 감정을 반영한다. 실험은 조용한 실내 실험실 환경에서 진행되었다. EEG, ECG, PPG, GSR은 Brain Products 사의 ActiChamp EEG amplifer, Facial은 Logitech사의 HD C920 웹캠으로 기록되었다.

데이터셋은 총 44명의 참가자(남성 26인, 여성 18인, 평균나이 24.64±2.13세)가 공개에 동의한 데이터를 포함하며 각 참가자는 총 80회의 VUI 상호작용(trial)을 수행하였다. 이는 네 가지 VUI 응답 유형(아동/성인 음성 × 간단/자세한 정보 제공)에 대해 각 20회씩 수행된 실험으로 구성되어 있다. 각 상호작용에 대한 데이터는 EEG(63채널, 1,000Hz), ECG, PPG, GSR (1,000Hz), Facial (30fps)이 동일한 타임스탬프에서 동시에 기록되었다. 이때 신호에 잡음이 많은 참가자와 트라이얼, 센서 측정 누락이 발생한 데이터는 제거되어 최종

적으로 39인 2,883개의 상호작용에 대한 데이터가 제공되었다. 참가자별로 평균 7.39±6.08개의 트라이얼이 제거되었으며, 이러한 전처리로 인해 5개 모달리티의 싱크는 모든 트라이얼에 대해 누락 없이 제공되었다.

참가자는 제시된 문장을 읽어 음성 명령을 수행한 뒤, VUI 시스템이 제공하는 음성 응답을 듣고 3초간의 대기 시간을 가진 후 9개의 형용사 쌍으로 이루어진 7점 척도 설문을 매 트라이얼마다 self-report 방식으로 응답하였다(예: Unstable-Stable, Hard to Understand-Easy to Understand, Uncertain-Certain 등). 이 설문 점수들은 Kansei 공학 기반 요인분석(Factor analysis)을 통해 고유값(Eigenvalue)이 1 이상인 것을 기반으로 두 개의 감성 요인으로 축소되었다. 연구진은 요인부하량(Factor loading)을 기반으로 두 요인에 참가자의 반응이 얼마나 안정적인가(안정성, Stability), 얼마나 마음에 들었는가(선호도, Favorability)로 명명하였다. 안정성은 시스템 응답에 대한 예측 가능성과 신뢰성, 선호도는 사용자가 응답에 대해 느끼는 호감도 및 상호작용의 긍정적 인상을 의미한다. Figure 1은 도출된 안정성과 선호도에 따라 4가지 클래스를 할당한 것이다. 사분면은 해당 데이터셋에서 도출한 안정성과 선호도를 음수, 양수로 구분하여 나눈 것이다. Class 1은 양의 안정-양의 선호 상태를 나타내고 Class 3은 음의 안정-음의 선호 상태를 나타낸다. 각 클래스의 트라이얼 데이터 수는 Table 1과 같다.

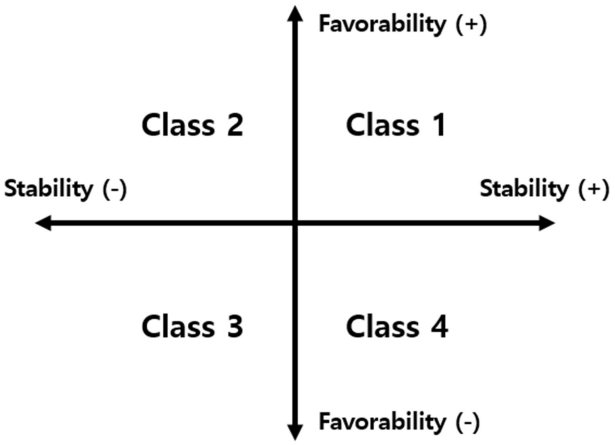

Figure 1. Classification of user emotional states across stability-favorability quadrants

Table 1. Number of trials in each class

Class	Trial (exp)
Class 1	807
Class 2	849
Class 3	651
Class 4	576

2.2 Preprocessing

EEG, ECG, PPG, GSR, Facial 신호는 센서 특성에 따라 다른 전처리 과정을 거쳤다. 모든 신호는 실험 단위(exp)별로 정렬되고 분리되었으며, 노이즈 제거, 정규화, 리샘플링 과정을 통해 분석 가능한 형태로 변환되었다. 일부 결측치는 인접 샘플을 이용해 보정하였다.

EEG: EEG 데이터는 데이터셋에서 제공된 전처리된 상태를 그대로 이용하였다. 59~61Hz 노치 필터와 1~80Hz 대역통과 필터를 적용한 뒤, 1,000Hz에서 200Hz로 다운샘플링 되었다. Independent Component Analysis (ICA)를 통해 눈깜빡임 및 근전도 잡음을 제거하고, 노

이즈가 심한 채널(FT9, T7, TP9, FT10, T8, TP10)을 제외하였다. ±75μV를 초과하는 구간은 제거되었으며, 분석은 음성 인터페이스의 응답(answer) 구간으로 한정되었다. 이렇게 전처리된 데이터에서 주파수 영역 특징(Frequency-domain features)에 따라 Delta, Theta, Alpha, Beta, Gamma 5가지 특징을 추출하였다(Klimesch, 1999).

ECG: 0.5~40Hz 대역통과(band-pass) 필터를 적용하고 기준선을 제거(detrending)한 후, 200Hz로 다운샘플링 하였다. 이후 Pan-Tompkins 알고리즘을 이용하여 QRS complex를 검출하고, 이를 기반으로 RR 간격(RR interval)과 PR 간격(PR interval)을 추출하였다(Pan and Tompkins, 1985). 이러한 간격 변동은 심장의 전기적 안정성과 자율신경계 활동을 반영하는 지표로 활용되었다(Shaffer and Ginsberg, 2017).

PPG: 0.5~5Hz 대역통과 필터를 적용하여 노이즈를 제거하고 기준선 변동을 보정하였다. 200Hz로 리샘플링한 후, 파형의 수축기(Systolic) 및 이완기(Diastolic) 피크를 검출하고, 피크 간 간격과 진폭 변화를 계산하여 심박수(HR), 심박변이도(HRV), 및 말초혈류 안정성 지표를 산출하였다(Charlton et al., 2018).

GSR: 0.01~0.5Hz 대역통과(band-pass) 필터를 적용하고 기준선을 제거한 뒤, 200Hz로 다운샘플링하였다. 전처리된 신호는 지속 성분(Skin Conductance Level, SCL)과 순간 반응 성분(Skin Conductance Response, SCR)으로 분리되었으며, SCR 구간에서 피크의 개수(Number of peaks), 피크의 크기(Peak amplitude), 피크 간격(Peak interval)을 계산하여 교감신경 활성도 변화를 정량화하였다(Benedek and Kaernbach, 2010).

Facial: FaceReader 소프트웨어를 이용하여 43개의 얼굴 근육 움직임(Action Unit)과 7가지 기본 감정 확률(Neutral, Happy, Sad, Angry, Surprised, Disgusted, Scared)을 추출하였다(Den Uyl and Van Kuilenburg, 2005).

Table 2는 각 센서에서 추출한 특징을 정리한 것이다.

Table 2. Summary of extracted features from multimodal signals

Signal type	Main features	Number of features
EEG	Frequency-domain features (Delta, Theta, Alpha, Beta, Gamma)	5 × 59
ECG	QRS complex, RR interval, PR interval	5
PPG	HR, HRV, Systolic peak, Diastolic peak	4
GSR	Number of peaks, Peak amplitude, Peak interval	3
Facial	43 action units, 7 emotion expressions	50

2.3 Analytical framework

본 연구는 전처리된 생리 및 행동 신호를 이용하여 안정성과 선호도를 예측하기 위한 분석 프레임워크를 구성하였다. 각 신호에서 추출된 특징(feature)은 표준화(Z-score) 후 분류 모델의 입력으로 사용되었으며, 모델 학습에는 Support Vector Machine (SVM)을 사용하였다.

SVM은 본 연구의 데이터 구조에 적합한 비선형 분류 알고리즘으로, 샘플 수가 제한적이면서도 다양한 형태의 특징(feature vector)을 다룰 수 있다는 장점을 가진다(Hsu et al., 2010). 본 연구의 데이터는 시계열 신호를 요약한 형태의 통계적 특성으로 구성되어 있어, LSTM이나 CNN 등의 대규모 데이터셋을 요구하는 딥러닝 모델보다 SVM이 과적합 위험이 낮고(Vapnik, 1995), 상대적으로 적은 데이터에서도 안정적인 성능을 확보할 수 있다. 또한 SVM은 입력 차원이 높더라도 RBF 커널을 통해 복잡한 결정 경계를 유연하게 학습할 수 있기 때문에, 다중 생리 신호의 이질적 특징을 통합적으로 다루는 데 유리하다(Chang and Lin, 2011). 모델은 안정성과 선호도의

수준(음양)에 따라 정의된 네 가지 클레스를 분류하는 사진 분류(quaternary classification) 문제와 안정성과 선호도 각각의 이진 분류 문제 두 가지로 설정되었다. 데이터 불균형 문제를 해결하기 위해 각 클래스에서 동일하게 500개의 데이터를 추출하여 총 2,000개의 데이터셋을 구성하였다. 이후 5-fold 교차검증을 수행하였으며, 사진 분류에서는 데이터 수를 동일하게 맞추기 위해 각 fold에서 클래스별 400개의 학습 데이터와 100개의 검증 데이터를 사용하였다. 이진 분류 분석은 4진 분류와 동일한 데이터 분할을 사용하였으며, 기존의 4개 클래스 레이블을 안정성과 선호도 기준으로 음양의 두 범주로 재구성하여 수행하였다. 데이터 분할 및 교차검증 절차는 4진 분류 실험과 동일하게 유지하였다. 최종 결과는 다섯 개 fold에서 얻은 성능 지표의 평균값으로 산출하였다.

본 연구는 데이터셋에서 제공된 5개 신호를 특성에 따라 세 가지 그룹으로 분류하였다. 이는 센서 부착의 난이도, 측정 시 드는 비용, 해석의 난이도, 센서 사용시 사용자가 느낄 수 있는 부담, 데이터 획득 방식 등의 현실적 요인을 고려하여 이루어졌다.

EEG 그룹: 두부(Head)에 부착되는 고비용, 고정밀 센서이다. 인지, 감정 변화, 반응과 관련된 두뇌의 피질 반응을 반영한다.

Physiological 그룹(ECG, PPG, GSR): 피부 접촉을 기반으로 하는 저비용 생리 신호 그룹이다. 자율신경계 활성과 반응을 포괄적으로 반영한다. 이 그룹은 세 가지 그룹 중에서 유일하게 멀티모달 방식의 그룹이다.

Facial 그룹: 비접촉(Non-contact) 방식으로 촬영된 얼굴 영상이다. 표정 및 미세한 얼굴 근육 움직임을 포착하고 추출하여 겉으로 드러나는 사용자의 반응을 반영한다.

다양한 연구에서 EEG는 가장 정밀도가 높은 감정 분석 수단으로써 사용된다. 하지만 실제 사용자의 일상적인 사용 환경에서는 EEG와 같은 고비용, 고부담의 센서를 사용하기 어렵다. 또한 이동 제한, 착용시의 불편함 등의 사용자에게 가해지는 부담으로 인해 반응이 손상될 수 있다. Physiological 그룹과 Facial 신호는 비교적 부담이 적다는 공통점이 있지만 별도로 비교하기 위해 그룹을 분리했다. 이는 Facial 신호를 통해 비접촉형 사용자 만족도 측정 시스템의 가능성을 평가하고자 위해서였다. 또한 Facial 신호는 비접촉, 저비용, 저난도라는 장점을 가져 이상적인 방식으로 보이나 사용자의 얼굴이 카메라의 정면을 향해야 한다는 단점이 존재하기 때문이기도 한다.

Table 3은 각 그룹에 포함된 센서, 측정 방식, 난이도와 코스트를 정리한 것이다.

Table 3. Classification of sensor groups and their practical characteristics

Group	Included sensors	Measurement type	Difficulty & Cost
EEG	EEG	Contact (Head-mounted)	High
Physiological signals	ECG, PPG, GSR	Contact (Skin-based)	Medium
Facial	Facial video	Non-contact (Camera-based)	Low

3. Results

3.1 EEG group performance

EEG 그룹은 평균 정확도 68.8%로 세 그룹 중 가장 높은 성능을 보였다. 혼동행렬 분석 결과, 네 개 클래스(Class 1-4) 전반에서 균형 잡힌 분류 경향을 보였으며, 특히 Class 1(양의 안정-양의 선호)에서 가장 높은 예측 일치를 보였다. EEG 신호는 사용자의 인지적 안정성과 긍정적 감정 상태를 효과적으로 구분하며, 전반적으로 가장 안정적이고 일관된 분류 성능을 보였다. 또한 EEG 기반 모델은 클래스 간 구분 경향이 뚜렷한 경향을 보인다. 이는 Figure 2에서 확인할 수 있다.

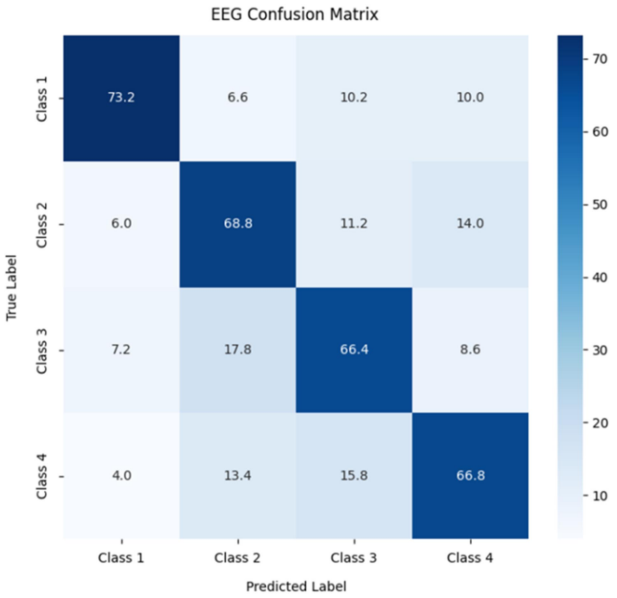

Figure 2. 4-class classification EEG confusion matrix heatmap

EEG 기반 이진 분류 결과는 Figure 3에 나타나 있다. 안정성 분류에서는 Balanced Accuracy 86.5%와 F1-score 86.6%를 기록하여 고·저 안정성 상태를 안정적으로 구분할 수 있음을 확인하였다. 선호도 분류 역시 Balanced Accuracy 85.8%, F1-score 85.5%로 비교적 높은 성능을 보였다. 이는 EEG 신호가 만족도의 개별 차원을 효과적으로 반영함을 시사한다.

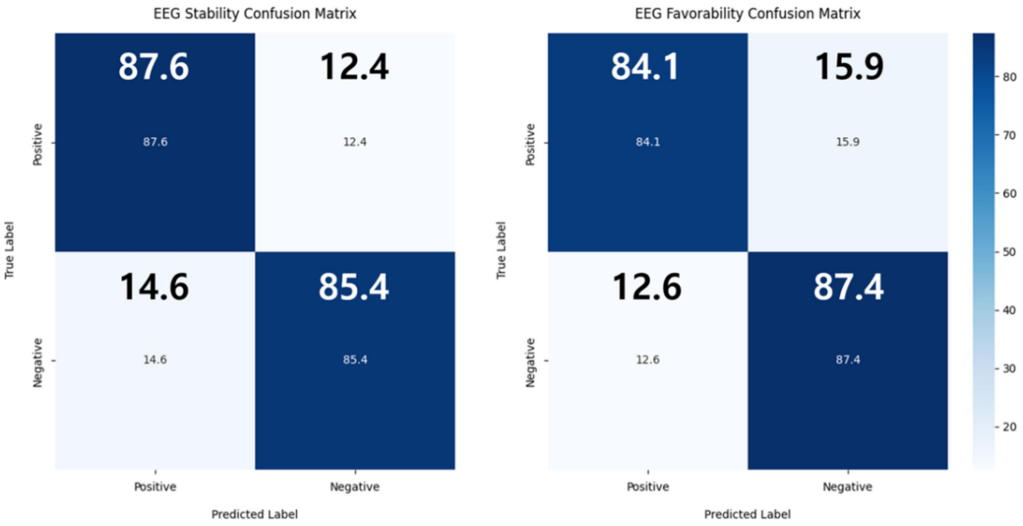

Figure 3. 2-class classification EEG confusion matrix heatmap

3.2 Physiological group performance (ECG, PPG, GSR)

Physiological 그룹은 평균 정확도 65.05%로 세 그룹 중에선 가장 낮은 정확도를 보였다. Figure 4에서 확인할 수 있듯, Class 2(음의

안정-양의 선호), Class 3(음의 안정-음의 선호)에서 상대적으로 높은 정확도를 보인다. 이는 음의 안정 상태에서 나타나는 심박수, 혈류량, 피부전도 반응 등의 생리적 지표가 클래스 분류 정확도에 영향을 미친 것으로 보인다. 또한 Class 1(양의 안정-양의 선호)과 Class 4(양의 안정-음의 선호), Class 2(음의 안정-양의 선호)와 Class 3(음의 안정-음의 선호)을 서로 오분류하는 비율이 높은 것을 확인할 수 있다. 이는 안정성 구분의 성능은 좋지만 선호도 구분에 있어서는 한계를 드러내는 것으로 해석할 수 있다.

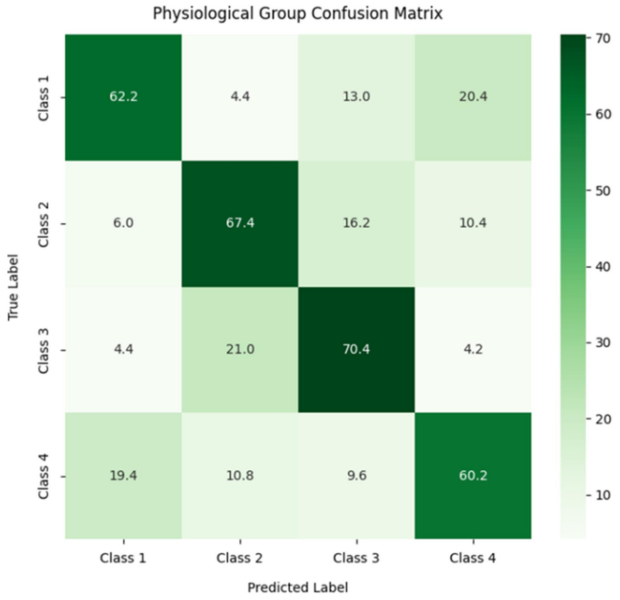

Figure 4. 4-class physiological signals confusion matrix heatmap

Physiological 그룹의 이진 분류 결과는 Figure 5에 나타나 있다. 안정성 분류에서는 Balanced Accuracy 85.6%, F1-score 85.7%로 비교적 안정적인 성능을 보였으며, 선호도 분류에서는 Balanced Accuracy와 F1-score가 모두 81.7%로 나타났다. 전반적으로 생리 신호는

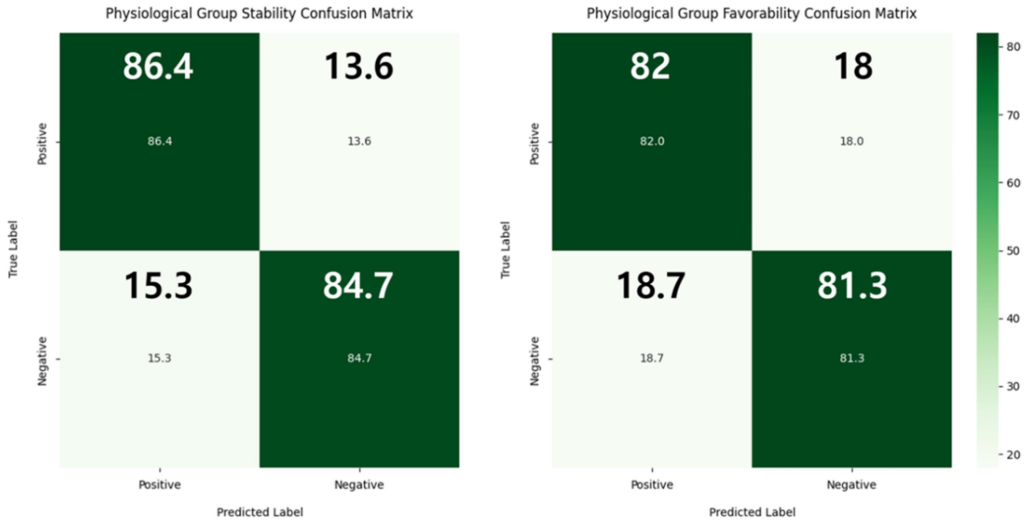

Figure 5. 2-class physiological signals confusion matrix heatmap

안정성과 선호도의 전반적인 경향을 포착할 수 있었으나, EEG 기반 결과에 비해 성능이 다소 낮게 나타났다. 특히 선호도의 경우 분류 정확도가 비교적 낮게 나타났는데, 이는 사진 분류에서 나타난 문제와 같은 결과로 해석된다.

추가적으로 해당 그룹에 포함된 모달리티를 각각 사진 분류 문제로 분석하였을 때는 약 32~39% 정도의 정확도를 보였다. 이는 Table 4에 나타나 있다. 이 경우 분류 성능이 현재 시도 중인 사진 분류 문제에서의 우연 정확도 수준인 25%에 근접하는 정확도를 보였다. 이진 분류 문제의 경우에도 F1 Score 56~66.8%, Balanced Accuracy 60~70.5% 수준인 모습을 보이며 비슷한 경향을 나타낸다.

Table 4. Classification accuracy for each modality of physiological group

Modality	4-class accuracy	2-class stability	2-class favorability
ECG	40.4%	F1 score: 66.8% Balanced accuracy: 70.5%	F1 score: 63.2% Balanced accuracy: 65.2%
PPG	34.0%	F1 score: 56.0% Balanced accuracy: 60.0%	F1 score: 60.4% Balanced accuracy: 63.4%
GSR	36.8%	F1 score: 58.2% Balanced accuracy: 61.8%	F1 score: 64.8% Balanced accuracy: 61.2%

3.3 Facial group performance

Facial 그룹은 평균 정확도 67.5%를 기록하였으며 Figure 6에 나타나 있다. 미소나 입꼬리의 상승 등은 양의 선호 상태에서 주기적으로 관찰되었고, 눈썹 하강이나 시선 회피는 음의 선호 상태일 때 반복적으로 관찰되었다. Class 3(음의 안정-음의 선호)와 Class 4(양의 안정-음의 선호)의 분류 성능이 비교적 높았으며 EEG와 마찬가지로 전반적으로 균형 잡힌 분류 성능을 보였다.

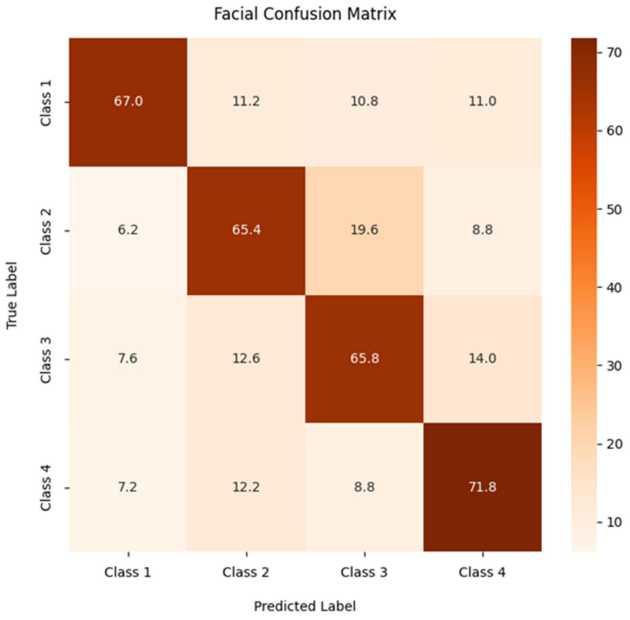

Figure 6. 4-class facial confusion matrix heatmap

Facial 그룹 이진 분류 결과는 Figure 7에서 볼 수 있다. 안정성 분류에서는 Balanced Accuracy 86.1%, F1-score 86.2%로 높은 성능을 보였으며, 선호도 분류에서는 Balanced Accuracy 84.5%, F1-score 84.4%를 기록하였다. 이진 분류 문제에서 Facial 그룹은 Physiological 그룹보다 다소 높은 분별력을 보였으며 EEG 그룹과 비슷한 수준의 성능을 보였다.

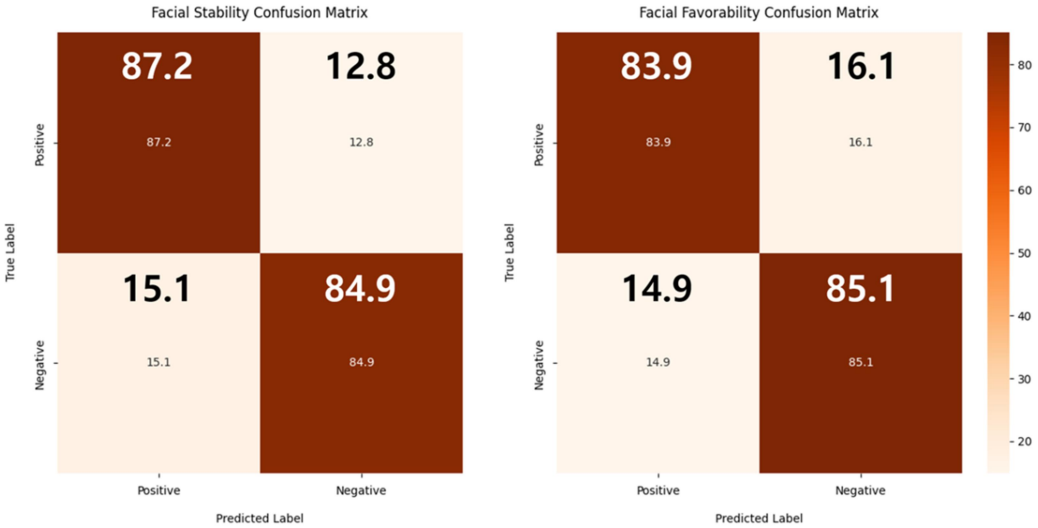

Figure 7. 2-class facial confusion matrix heatmap

4. Discussion

EEG는 사용자의 반응에서 나타나는 두뇌의 전반적인 반응, Physiological 그룹은 생리적 각성과 반응 강도, Facial은 비언어적 행동 단서인 외형적 감정 표현을 나타낸다. 세 그룹 간의 분류 성능을 비교해 보면 EEG 그룹은 세 모델에서 가장 높은 정확도와 일관된 분류 성능을 보인다. Physiological 그룹은 안정도 분류에서 비교적 높은 민감도를 보였지만 선호도 구분에선 한계를 보였다. Facial 그룹은 EEG 그룹보다는 정확도가 낮았으나 마찬가지로 일관된 분류 성능을 보인다. 전반적으로 세 그룹 모두 안정성과 선호도의 상호작용을 포착하였으며, 감정 상태의 복합성에 따라 신호 유형별 민감도가 다르게 나타났다. 이러한 결과는 사용자 만족도 평가 시 단일 신호보다 복수 신호 조합이 중요하다는 것을 보여준다. Physiological 그룹은 멀티모달한 방식을 사용했음에도 가장 낮은 분류 성능을 보였다. 하지만 이는 그룹 내의 단일 모달리티 기반으로 분류하였을 시 보다는 높은 수치인 것으로 확인되었다. 모든 그룹에서 안정성과 선호도에 따른 이진 분류 시에는 모두 80% 이상의 정확도를 보였으며, 이는 데이터셋에서 사용자의 만족도 상태를 나타내는 사분면의 수준 보다는 낮더라도 만족도의 한 축 기준으로는 만족도 추정이 가능하다는 것을 나타낸다.

본 연구는 EEG를 포함한 다양한 생리 및 행동 신호 중에서도 Physiological 신호(ECG, PPG, GSR)와 Facial이 실생활에서의 사용자 만족도 추정에서 실질적인 기여를 할 수 있음을 보여준다. 두 신호 모두 간단한 측정 절차를 통해 획득 가능하며, 특히 ECG와 GSR같은 경우에는 현재 성장하고 있는 웨어러블 디바이스에도 탑재되고 있는 만큼 안정적인 실제 환경에서의 활용 가능성을 가진다. Physiological 신호는 교감신경계의 활성 변화를 반영해 긴장, 피로, 각성 상태 등을 감지할 수 있고(Benedek and Kaernbach, 2010), Facial 신호는 표정과 미세한 행동 단서를 통해 감정의 방향을 추정할 수 있다(Ekman and Friesen, 1978). 또한 비언어적 행동은 언어보다 많은 정보를 포함하고 있어 사용자의 감정 상태를 효과적으로 탐지할 수 있다(Mehrabian, 1971). 즉, 두 그룹은 EEG에 비해 정밀도는 조금 낮지만, 사용자에게 부담이 적고 측정 효율이 높아 실제 환경에서 사용될 수 있을 것으로 보인다. Physiological 그룹은 선호도의 구분에서는 약점을 보이지만 안정도의 구분에선 다른 그룹과 근접한 성능을 보이는데, 이는 다른 그룹과의 통합 분석을 통해 분류 성능과 신뢰도를 높일 수 있을 것으로 기대된다. 특히 Facial 그룹은 사진 분류와 이진 분류 문제 모두에서 EEG 그룹보다는 정확도가 낮았으나 근접하는 성능을 내는 것을 바탕으로 측정 절차가 비교적 간단한 두 신호 그룹의 조합이 각 센서의 단점을 보완하고

상호작용성을 포착하는 멀티모달의 의의를 통해 더 높은 수준의 정확도를 이룰 것이라고도 기대할 수 있다.

본 연구는 최근 이루어지고 있는 감정 인식 모델이나 감성 분석 모델의 비해 다소 낮은 정확도를 보이고 있다. 이는 본 연구가 유도된 환경이 아닌 비유도 상황에서의 자연스러운 상호작용 데이터를 사용했기 때문이다. Kim et al. (2025)는 비유도된 상황에서의 데이터와 유도된 상황의 데이터를 같은 모델을 사용하였을 때의 결과를 비교하는 연구를 수행한 바 있다. 해당 연구에서 EEG를 포함한 멀티모달 데이터로 Valence와 Arousal 수치를 각각 이진 분류하였을 때 F1 Score가 크게는 12.2%의 차이를 보였다. 유도된 감정 데이터를 사용했을 때는 자극과 감정 상태가 명확하여 분류 정확도가 높게 나타났지만, 자연스러운 감정 데이터에서는 감정 강도 변화가 미미하고 다양한 맥락이 혼재해 있어 전반적인 분류 정확도가 낮은 모습을 보였다.

감정 인식 모델의 가치는 단순히 높은 분류 정확도를 보이는 것에 한정되지 않는다. 본 연구진은 감정 인식 모델의 목적과 가치가 궁극적으로는 실제 사용 맥락에서 발생하는 자연적 감정 반응을 얼마나 안정적으로 포착하고 인테페이스나 상호작용에 반영하는가에 있다고 본다. 본 연구는 EEG와 같이 측정 난이도가 높고 비용이 많이 드는 정밀한 장비 없이도 사용자의 안정성(Stability)과 선호도(Favorability)를 탐지할 수 있음을 검증함으로써, 현실적이고 접근 가능한 감정 추정 모델의 가능성을 제시한다. 이는 인위적 자극 없이 발생하는 사용자의 저강도 반응을 기반으로 자연스러운 사용자 만족도를 분석할 수 있는 새로운 접근 방식이라 할 수 있다.

이러한 결과는 향후 실시간 상호작용 시스템 설계에 시사점을 제공한다. AI 비서, 챗봇, 가상 휴먼 등과 같이 최근 확장되고 있는 AI 기반 인터페이스에서 사용자의 반응을 실시간으로 감지하고 발화, 톤, 표현, 강도 등의 응답 방식을 조정한다면, 사용자에게 더욱 몰입감 있는 경험을 제공할 수 있을 것으로 기대된다. 본 연구의 비유도 기반 접근은 위와 같은 감성 적응형 인터페이스(Emotionally adaptive interface) 개발의 기초가 된다. 본 연구는 유도되지 않은 자연스러운 상호작용 중 나타나는 사용자의 반응에 대한 생리적, 비언어적 행동 단서 데이터를 활용하여 사용자 만족도를 추정한다는 점에서 의미가 있다. 이는 향후 실생활에서 적용 가능한 사용자 상태 추정 및 반응 감지 모델의 개발에 있어서 중요한 기반이 될 것이다.

5. Conclusion

본 연구는 인공지능 기반 VUI 환경에서 생리 신호와 비언어적 행동 단서를 통해 사용자 만족도를 실시간으로 추정할 수 있는 가능성을 탐색하였다. 특히 본 연구는 인위적 감정 유도 없이 자연스러운 상호작용 중 발생한 데이터를 사용하였다. 연구 결과, EEG가 가장 높은 분류 정확도를 보였으나, ECG, PPG, GSR 등의 생리 신호와 Facial 신호 또한 만족도 변화의 주요 단서로 작용함을 확인하였다. 이러한 신호들은 측정이 비교적 용이하고 착용시에도 사용자에게 부담이 적어, 실제 환경에서의 실시간 사용자 만족도 추정 모델에서 사용될 잠재력을 지니고 있음을 알 수 있다.

본 연구는 기존 유도형 감정 실험의 한계를 넘어, 자연스러운 상호작용 맥락에서 사용자 만족도를 정량화할 수 있는 평가 체계를 제안하였다. 또한 EEG에 비해 사용자 부담과 측정 난이도가 낮은 생리 및 행동 신호만으로도 사용자 만족도 상태를 유의미하게 예측할 수 있음을 실험적으로 입증하였다. 이러한 접근은 향후 AI 기반 인터랙션 시스템의 개인화(personalization) 및 감성 적응형 설계(Emotionally Adaptive Design)를 위한 이론적 토대를 제공할 것이다.

다만 본 연구는 음성 인터페이스 환경에 한정된 데이터셋을 사용하였고, 감정 반응의 폭이 상대적으로 좁다는 한계를 가진다. 향후 연구에서는 사용자의 선호, 상황, 맥락 등 다양한 변수를 고려한 상호작용 환경에서의 검증이 필요하다. 현 단계에서는 Physiological 그룹만이 멀티모달한 방식으로 검증되었으나 향후 연구에서는 더욱 다양한 모달리티의 통합을 통해 멀티모달한 방식으로 정확도와 신뢰성을 높일 것이며, 개인별 반응 차이를 보정할 수 있는 개인화 모델링(Personalized modeling) 기법을 적용함으로써 예측의 일반화 성능을 향상시킬 것이다. 이러한 시도를 통해 사용자 부담이 적으며 측정 난이도가 낮은 신호 기반의 실시간 사용자 만족도 예측 기술을 구현함으로써 향후 지능형 상호작용 시스템의 핵심 요소로 발전할 수 있는 마련하는데 기여할 수 있을 것으로 기대된다.

Acknowledgement

본 연구는 국립금오공과대학교 교수연구년제에 의하여 연구된 실적물임.

References

Apicella, A., Arpaia, P., Mastrati, G. and Moccaldi, N., EEG-based detection of emotional valence towards a reproducible measurement of emotions, *Scientific Reports*, 11(1), 21615, 2021. doi:10.1038/s41598-021-00812-7

Benedek, M. and Kaernbach, C., Decomposition of skin conductance data by means of nonnegative deconvolution, *Psychophysiology*, 47(4), 647-658, 2010. doi:10.1111/j.1469-8986.2009.00972.x

Bolger, N., Davis, A. and Rafaeli, E., Diary methods: Capturing life as it is lived, *Annual Review of Psychology*, 54(1), 579-616, 2003. doi:10.1146/annurev.psych.54.101601.145030

Calvo, R.A. and D'Mello, S., Affect detection: An interdisciplinary review of models, methods, and their applications, *IEEE Transactions on Affective Computing*, 1(1), 18-37, 2010.

Chang, C.C. and Lin, C.J., LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1-27, 2011. doi:10.1145/1961189.1961199

Charlton, P.H., Celka, P., Farukh, B., Chowienczyk, P. and Alastruey, J., Assessing mental stress from the photoplethysmogram: a numerical study, *Physiological Measurement*, 39(5), 2018. doi:10.1088/1361-6579/aabe6a

Choi, G.Y., Shin, J.G., Lee, J.Y. and Kim, S.H., EEG dataset for the recognition of different emotions induced in voice-user interaction, *Scientific Data*, 11(1084), 2024. doi:10.1038/s41597-024-03887-9

Den Uyl, M.J. and Van Kuilenburg, H., The FaceReader: Online facial expression recognition, *Proceedings of Measuring Behavior 2005*, 589-590, 2005.

Ekman, P. and Friesen, W.V., Facial Action Coding System*, Consulting Psychologists Press*, 1978.

Fu, Z., Zhang, B., He, X., Li, Y., Wang, H. and Huang, J., Emotion recognition based on multi-modal physiological signals and transfer learning, *Frontiers in Neuroscience*, 16, 1000716, 2022. doi:10.3389/fnins.2022.1000716

Hsu, C.W., Chang, C.C. and Lin, C.J., A practical guide to support vector classification, Technical Report, Department of Computer Science, *National Taiwan University*, 2010.

Kim, S.T., Heo, I.S. and Kim, S.H., Validating the applicability of multimodal emotion classification models in natural interaction scenarios, *Proceedings of the Spring Conference of the Ergonomics Society of Korea*, 2025.

Klimesch, W., EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis, *Brain Research Reviews*, 29(2-3), 169-195, 1999. doi:10.1016/S0165-0173(98)00056-3

McDuff, D., Rowan, K., Choudhury, P., Wolk, J., Pham, T. and Czerwinski, M., A multimodal emotion sensing platform for building emotion-aware applications, *arXiv preprint*, arXiv:1903.12133, 2019.

Mehrabian, A., Silent messages: Implicit communication of emotions and attitudes, Wadsworth, Belmont, CA, 1971.

Pan, J. and Tompkins, W.J., A real-time QRS detection algorithm, *IEEE Transactions on Biomedical Engineering*, BME-32(3), 230-236, 1985. doi:10.1109/TBME.1985.325532

Picard, R.W., Affective Computing, *MIT Press*, Cambridge, MA, 1997.

Robinson, D. and Tannenberg, M., Self-censorship of regime support in authoritarian states: Evidence from list experiments in China, *Research & Politics*, 6(3), 2053168019856449, 2019. doi:10.1177/2053168019856449

Saffaryazdi, N., Lim, H., Rassovsky, Y. and Lim, S., Using facial micro-expressions in combination with EEG and peripheral signals for emotion recognition, *Frontiers in Psychology*, 13, 864047, 2022. doi:10.3389/fpsyg.2022.864047

Shaffer, F. and Ginsberg, J.P., An overview of heart rate variability metrics and norms, *Frontiers in Public Health*, 5, 258, 2017. doi:10.3389/fpubh.2017.00258

Shin, J.G., Choi, G.Y., Hwang, H.J. and Kim, S.H., Evaluation of emotional satisfaction using questionnaires in voice-based human–AI interaction, *Applied Sciences*, 11(4), 1920, 2021. doi:10.3390/app11041920

Vapnik, V.N., The Nature of Statistical Learning Theory, *Springer-Verlag*, New York, 1995.

Zhao, Z., Zhang, Y. and Hu, B., Emotion recognition with multi-modal peripheral physiological signals: A review, *Frontiers in Computer Science*, 5, 1264713, 2023. doi:10.3389/fcomp.2023.1264713

Zhuang, N., Zeng, Y., Yang, K., Zhang, C., Tong, L. and Yan, B., Investigating patterns for self-induced emotion recognition from EEG signals, *Sensors*, 18(3), 841, 2018. doi:10.3390/s18030841

Author listings

Seong-Taek Kim: rubunu1115@gmail.com
Highest degree: Bachelor of Engineering, Department of Industrial Engineering, Kumoh National Institute of Technology
Position title: Master's Student, Department of Industrial Engineering, Kumoh National Institute of Technology
Areas of interest: Human-AI Interaction, Usability Evaluation, User Satisfaction

Sang-Ho Kim: kimsh@kumoh.ac.kr
Highest degree: Ph.D., Department of Industrial Engineering, POSTECH
Position title: Professor, School of Industrial, Mathematics & Big Data Science Engineering, Kumoh National Institute of Technology
Areas of interest: Human-robot Interaction, Human Factors and Ergonomics, Occupational Safety and Health