동적 가중치 기반 멀티모달 감정 인식과 공감형 AI를 활용한 실시간 멘탈케어 모바일 애플리케이션

*신다온, 장예연, 박소연, 이다혜

국립공주대학교

*202201343@smail.kongju.ac.kr, wsx0329@naver.com, soyean414@gmail.com, dahye2480@naver.com,

Real-Time Mental-Care Mobile Application Using Dynamic-Weight Multimodal Emotion Recognition and Empathic AI

*Daon Shin, Yeyeon Jang, Soyeon Park, Dahye Lee

Kongju National University

요 약

현대 사회에서 정신 건강 관리의 필요성이 커지고 있으나, 대면 상담은 비용·접근성·즉시성의 한계가 존재한다. 본 연구는 사용자의 언어(텍스트) 및 비언어(표정 이미지) 신호를 동시에 활용해 감정을 정밀하게 추정하고, 그 결과를 기반으로 공감형 상담 대화와 위험 감정 연속 감지(경고)를 제공하는 실시간 멘탈케어 모바일 애플리케이션을 제안한다. 제안 시스템은 (1) 한국어 대화 약 5만 건으로 학습한 KoBERT 기반 텍스트 감정 분류 및 키워드 추출, (2) FER2013 기반 이전 모델 대비 실제 한국인 웹캠 환경 적합성을 높이기 위해 AI Hub 한국인 표정 데이터로 교체한 CNN 기반 이미지 감정 분류, (3) 두 모달의 신뢰도에 따라 가중치를 자동 조절하는 동적 가중치 융합 알고리즘, (4) 동일 감정이 5회 이상 연속될 때 경고 플래그를 반환하는 위험 감지 로직, (5) Fine-tuned Gemini 기반 공감형 상담 기능을 포함한다. 실험 결과, 이미지 모델은 이전 대비 실사용 환경에서 감정 인식의 안정성이 향상되었고(예: 72%→86%), Label Smoothing 적용 시 테스트 정확도 약 87%를 기록하였다. 또한 Grad-CAM 분석을 통해 고해상도 입력에서 감정 판단 근거가 얼굴 핵심 영역(눈썹·입꼬리 등)으로 수렴함을 확인하였다.

Ⅰ. 서 론

우울감·고립감이 심화되는 반면 전문 심리 상담은 비용과 접근성의 장벽으로 즉각적·상시적 도움을 받기 어려운 사용자가 존재한다. 이에 따라 시공간 제약 없이 도움을 제공하는 디지털 멘탈케어 솔루션이 주목받고 있으나, 기존 대화형 AI는 텍스트 중심으로 동작해 표정과 같은 비언어 신호를 반영하지 못하고, 위험 징후를 놓칠 수 있다는 점이 한계로 지적된다.

본 연구는 텍스트 감정 분류를 넘어 표정 기반 감정 인식을 결합한 멀티모달 감정 추정과, 그 결과를 활용한 공감형 상담 및 반복 감정 경고를 하나의 모바일 서비스로 구현하는 것을 목표로 한다.

Ⅱ. 시스템 구조

제안 시스템은 모바일 앱(FlutterFlow), 백엔드(FastAPI), 데이터 저장소(Firebase Authentication/Firestore), 감정 분석 및 상담 생성 AI 모듈로 구성된다. 앱은 사용자 인증을 기반으로 게시판(일상 기록)과 채팅 기반 상담 기능을 제공한다. 게시판에서는 사용자가 작성한 글을 분석 API로 전달하여 감정 결과와 핵심 키워드를 표시하고, 긴 글의 경우 문장 단위 감정 흐름을 제공한다. 채팅에서는 멀티모달 감정 결과를 반영해 공감형 응답을 생성하며, 사용자 정보·게시글/댓글·상담 로그는 Firestore에 저장되어 앱에서 즉시 조회·반영되도록 설계하였다.

Ⅲ. 멀티모달 감정 인식 모델

텍스트 감정 분석은 약 5만 건의 한국어 대화 데이터로 학습한 KoBERT 기반 분류기를 사용해 문장 단위 감정을 예측하고, 형태소 분석(Okt)을 통해 핵심 키워드를 추출한다. 게시판과 같이 긴 텍스트는 문장 분리(kss) 후 문장별 감정을 순차적으로 산출하여 사용자가 '한 번의 라벨'이 아니라 '감정 변화의 흐름'을 확인할 수 있도록 구성하였다.

이미지(표정) 감정 인식은 초기 FER2013 기반 48×48 흑백 환경에서 Sadness-Neutral 오분류가 잔존했고, 단순 업샘플링(48→96)만으로는 정확도가 개선되지 않는 한계를 확인하였다(기존과 동일한 72%). 따라서 2학기에는 한국인 고해상도 표정 데이터로 전환하고(약 50만 장, RGB, 5개 감정 클래스), 얼굴 검출 또한 실시간 서비스 안정성을 위해 OpenCV Haar Cascade에서 MediaPipe 기반으로 교체하였다. 이를 통해 "얼굴 검출→영역 추출→리사이즈/정규화"의 전처리 흐름을 일관화하고, 실제 웹캠 입력에서 감정 인식의 안정성을 강화하였다.

Ⅳ. 동적 가중치 기반 감정 융합 및 위험 감지

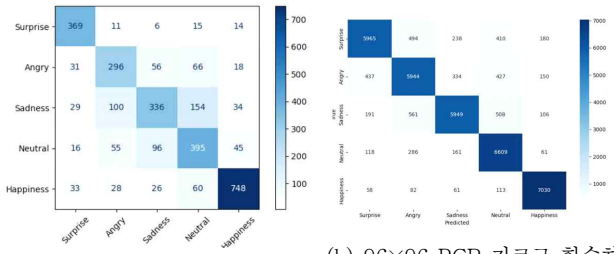
멀티모달 융합은 텍스트와 표정 예측을 동시에 고려하되, 실시간 환경에서의 출력 안정성을 최우선으로 설계하였다. 표정 정보는 카메라가 켜져 있는 동안 연속적으로 입력되지만, 텍스트는 발화 시점에만 발생하는 비연속 신호이므로 기본적으로 표정에 더 높은 비중을 두고(기본 7:3), 모달별 예측 신뢰도에 따라 가중치를 자동 조절한다. 예를 들어 이미지 예측 신뢰도가 낮으면 텍스트 비중을 높여 0.40:0.60으로 보정하고, 텍스트 신뢰도가 높거나 두 모달의 최빈 감정이 일치하면 0.50:0.50으로 정렬하여 결과를 안정화한다. 반대로 텍스트 신뢰도가 매우 낮아 정보성이 부족하면 이미지 1.0으로 처리하여 웹캠 기반 감정 흐름이 자연스럽게 유지되도록 설계하였다.

또한 대화 특성상 텍스트 감정이 한 번 출력되면 침묵 구간에서도 지속되어 표정 변화를 가릴 수 있으므로, 텍스트 감정은 발화 후 일정 시간(보

고서에서는 3초)만 유지하고 이후 중립으로 복귀하도록 하여 시간적 불균형을 완화하였다. 위험 감지는 단일 문장 '점수화'보다 반복 패턴에 주목하여, 동일 감정이 5회 이상 연속될 경우 경고 플래그를 반환하고 앱에서 팝업 안내를 제공하도록 구현하였다.

V. 실험 및 결과

이미지 모델의 성능 개선은 "데이터 도메인 적합성"과 "입력 정보량 확장"을 중심으로 확인되었다. 이후에는 48×48 흑백에서 96×96 RGB로 전환하여 픽셀 수가 약 12배 증가했고(2,304→27,648), 1채널 밝기 정보에서 RGB 3채널 정보를 활용하도록 변경하였다. 동시에 한국인 표정 데이터를 사용해 실제 사용자 환경에 더 잘 맞도록 조정한 결과, 보고서에서는 감정 정확도가 72%에서 86%로 향상되었다고 정리한다. 이는 '표정을 똑같이 지어도 매번 다르게 인식'되는 불안정성이 완화되며 실시간 표정 인식이 더 안정적으로 동작하게 되었음을 시사한다.



(a) 48×48 흑백 자료로 학습한 모델 혼동 행렬

(b) 96×96 RGB 자료로 학습한 모델 혼동 행렬

Fig. 1. 제안한 모델의 감정 예측 혼동행렬

학습 전략 비교에서는 여러 고도화 시도 중 Label Smoothing이 가장 단순하면서도 일반화 성능이 우수해 최종 선택되었고, Early Stop 없이 안정적 학습이 진행되며 검증 정확도 94.36%, 테스트 정확도 약 87%를 기록하였다. 또한 감정별 Recall 비교에서 Sadness(51→81), Neutral(65→93), Angry(63→80) 등 기존 취약 클래스가 크게 개선되어, 특히 이전에 문제였던 Sadness-Neutral 구분이 실질적으로 강화되었음을 확인할 수 있다.

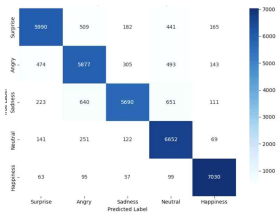


Fig. 2. Label Smoothing 혼동행렬

모델이 실제로 어떤 근거로 감정을 판단하는지에 대한 해석 가능성 분석으로 Grad-CAM 결과도 제시되었다. 48×48 모델은 얼굴 전반의 밝기 변화나 배경에 반응하는 등 판단 근거가 불안정한 반면, 96×96 모델은 슬픔에서 눈썹·내려간 입꼬리, 행복에서 광대·올라간 입꼬리·눈가, 분노에서 미간·눈썹 등 감정과 직접 관련된 핵심 부위에 집중하여 더 안정적으로 판단하는 양상을 보였다. 이는 해상도 증가와 데이터 교체가 단순 정확도 향상뿐 아니라, 사람의 표정 해석과 유사한 특징 참조로 이어졌다는 점에서 의미가 있다.

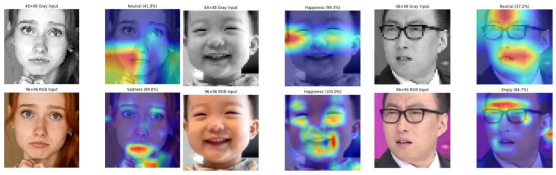


Fig. 3. 감정별 Grad-CAM 비교

앱 실행 동작은 실제 사용자 흐름에서 확인하였다. 사용자는 스플래시 이후 로그인/회원가입을 거쳐 메인 화면에 진입하며, 로그인 상태에 따라 초기 페이지가 분기되도록 구성하였다. 게시판에 글을 작성하면 감정 결과가 작성 시간 옆에 표시되고, 별도 팝업에서 핵심 키워드(Okt 기반)와 문장 흐름 기반 감정 추이 그래프가 제공된다(문장 분리는 kss 기반). 채팅에서는 입력 문장 감정과 웹캠 표정 감정이 동시에 계산되어 동적 가중치 융합 결과가 상담 응답에 반영되며, 동일 감정이 연속 누적될 경우 경고 문구 팝업이 출력된다. 또한 서버 전송/처리 상황은 프롬프트 로그로 모니터링하여 앱-API 연동이 정상 동작함을 점검하였다.

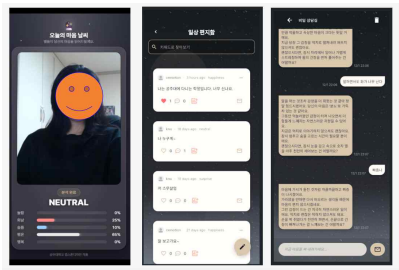


Fig. 4. 앱 실행 동작 화면

VI. 결 론

본 연구는 텍스트와 표정 정보를 결합한 멀티모달 감정 인식, 공감형 AI 상담, 연속 감정 기반 경고를 하나의 모바일 애플리케이션으로 통합한 실시간 멘탈케어 시스템을 제안하고 구현하였다. 특히 표정 인식은 데이터 도메인 적합성(한국인 표정)과 전처리 안정성(얼굴 검출 파이프라인)을 개선함으로써 실사용 환경에서 감정 출력의 일관성을 높였고, 학습 전략 측면에서는 Label Smoothing을 통해 취약 클래스의 Recall을 개선하였다. 또한 신뢰도 기반 동적 융합과 텍스트 감정 유지 시간 제한을 적용하여 실시간 대화에서의 흔들림을 완화하였다. 향후에는 사용자 평가 기반의 정량/정성 검증을 확대하고, 개인화 기준선 및 안전 대응 정책을 고도화하여 서비스 적용성을 강화할 예정이다.

ACKNOWLEDGMENT

본 연구는 2026년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음(2024-0-00073)

참 고 문 헌

[1] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv preprint arXiv:1307.0414, 2013.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT 2019 (Volume 1: Long and Short Papers), pp. 4171-4186, 2019. doi:10.18653/v1/N19-1423.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of CVPR 2016, pp. 2818-2826, 2016. doi:10.1109/CVPR.2016.308.