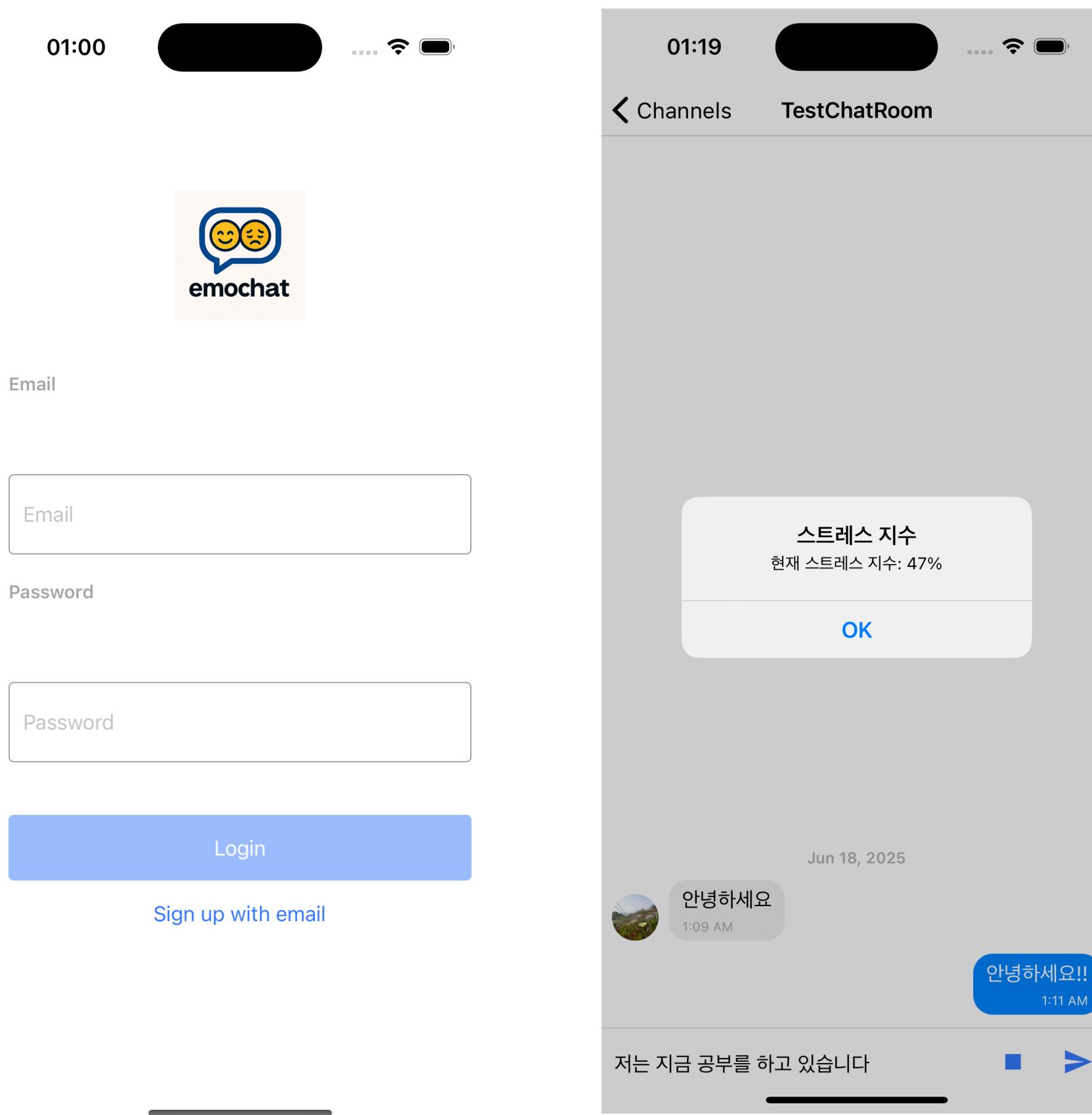




AI 기반 음성 감정 분석 및 개인화 추천 시스템 개발

성과발표회 발표자료

2025.06.18(수)



EmoChat

배경

스트레스는 번아웃 예방·삶의 질 향상을 위해 필수 관리 항목
기존 측정 방식(심전도·피부전도)은 침습적이고 일상 적용 어려움
음성 톤·속도·강도에 감정 정보 존재하나 채팅 앱은 미활용

솔루션

학습 단계: StressID(심전도·피부전도·호흡·얼굴·음성) 데이터로 고정밀 모델 훈련
추론 단계: 마이크·카메라만으로 스트레스 예측하는 지식 종류 기법 적용

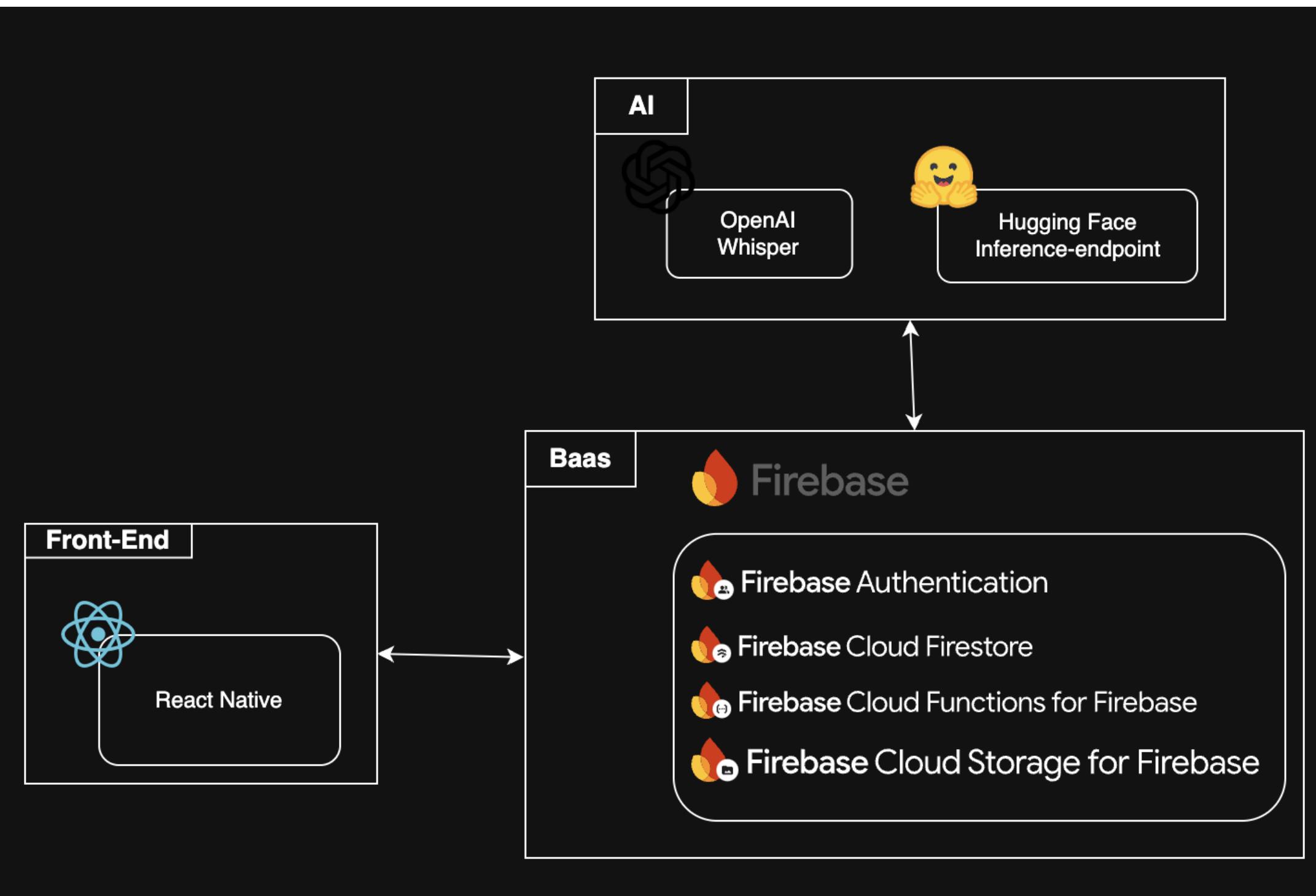
구현

Expo·ReactNative 채팅 앱에 통합
Whisper STT → 텍스트 변환 → Hugging Face Endpoint로 실시간 스트레스 계산
스트레스 높음 시 팝업 경고, 정상 시 말풍선 노란색 + 수치 표시

기대 효과

대화 중 심리 상태 인지·공감

전체 아키텍처



Front-end (React Native / Expo + GiftedChat)

사용자 음성 녹음 및 전송

STT 결과 및 스트레스 상태 실시간 표시

BaaS (Firebase)

Authentication: 소셜/이메일 로그인, 사용자 관리

Cloud Firestore: 채팅 메시지 실시간 저장·구독

Cloud Storage: 음성 파일 업로드·보관

Cloud Functions: 서비스 로직(알림, 데이터 후처리 등)

AI Services

OpenAI Whisper: 음성 → 텍스트(STT) 변환

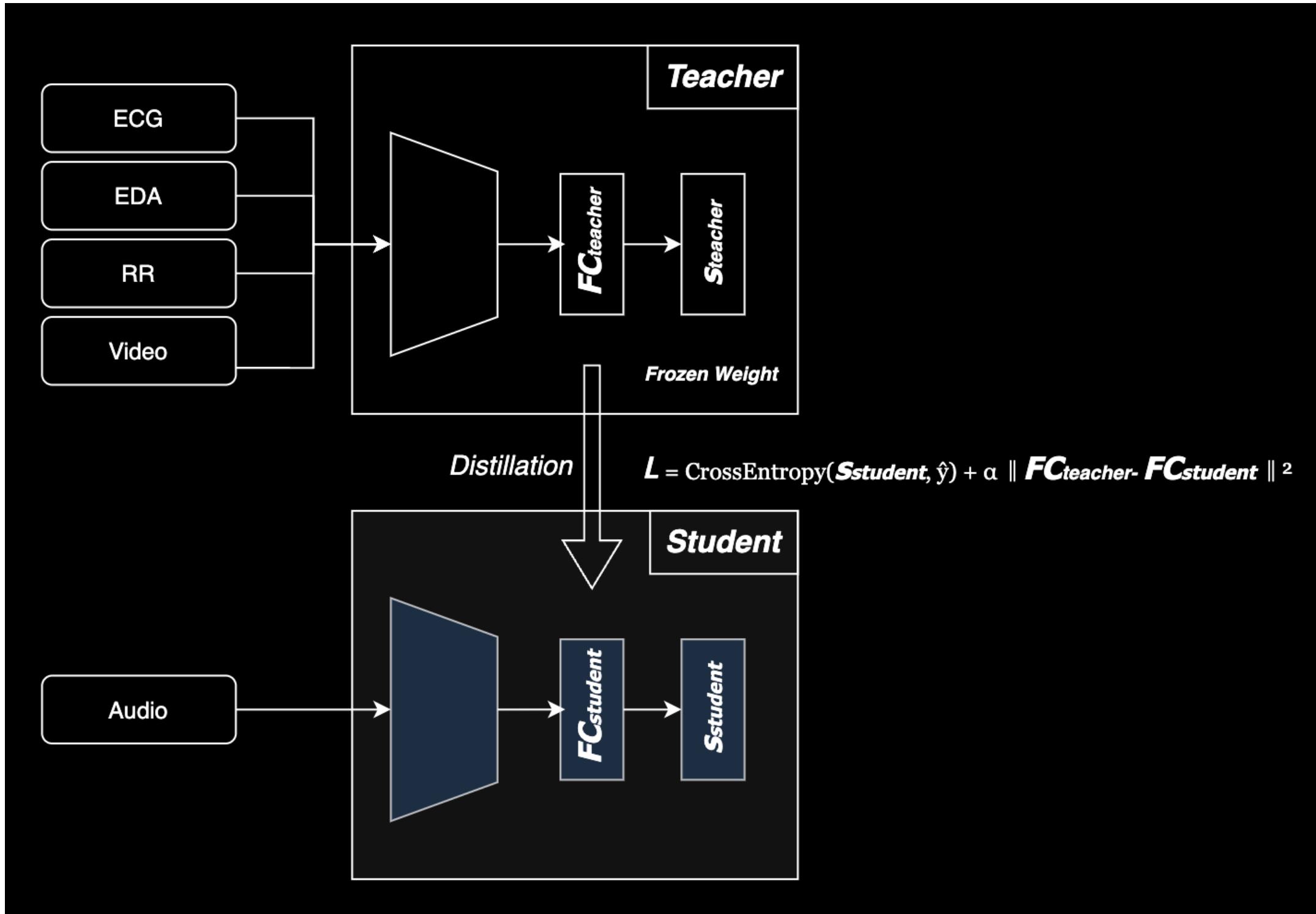
Hugging Face Inference Endpoint: 음성 분석 → 스트레스 지수 예측

데이터 흐름

앱에서 녹음 → BLOB 형태로 OpenAI Whisper 전송 → 텍스트 수신

동일한 BLOB을 Hugging Face Endpoint로 전송 → 스트레스 스코어 수신

텍스트와 스코어를 입력창 및 말풍선에 반영, 실시간 채팅은 Firestore로 동기화



모델 학습

voice-based-stress-recognition

침습적 생체신호와 비디오 데이터를 이용해 학습된 고정밀 Teacher 모델을 오디오만 입력으로 받는 경량 Student 모델로 지식 증류하는 과정

- * Teacher 모델

- * 입력: ECG, EDA, RR 간격, Video

- * 모달별 인코더 → 특징 벡터 결합 → 분류 헤드 → 스트레스 예측(logits)

- * 중간 특징 벡터는 학습용으로만 사용되고 가중치는 고정(frozen)

- * Student 모델

- * 입력: 스마트폰 마이크로 녹음한 오디오 → Word2Vec 임베딩

- * W2V 인코더 → 특징 벡터 → 분류 헤드 → 스트레스 예측(logits)

- * 학습 시 사용 손실

1. 분류 손실 : 학생 모델의 예측과 실제 레이블 간의 교차 엔트로피

2. 증류 손실 : Teacher와 Student 특징 벡터 간의 MSE

이 방식으로 학습 단계에는 침습적, 비디오 모달리티를

추론 단계에는 오디오만 사용해 실시간 비침습 스트레스 예측을 구현

[forwarder1121/voice-based-stress-recognition](#) like 0

Audio Classification PyTorch aether-raid/StressID English knowledge-distillation stress-detection audio custom_code License: mit

Model card Files and versions Community 1 Settings Deploy

Voice-Based Stress Recognition (StudentNet)

Model Card for forwarder1121/voice-based-stress-recognition

Model Details

- Model name: Voice-Based Stress Recognition (StudentNet)
- Repository: <https://huggingface.co/forwarder1121/voice-based-stress-recognition>
- License: MIT
- Library version: PyTorch ≥1.7

Model architecture:

A lightweight MLP-based StudentNet distilled from a multimodal TeacherNet trained on the StressID dataset.

Downloads last month 1,117



Inference Providers NEW

Audio Classification

This model isn't deployed by any Inference Provider. Ask for provider support

Dataset used to train forwarder1121/voice-based-stress-recog...

aether-raid/StressID

Viewer • Updated May 14 • 378 • 9

모델 학습 결과

voice-based-stress-recognition

실험 결과

TeacherNet (멀티모달)

Accuracy ≈ 0.82

Macro-F1 ≈ 0.80

UAR ≈ 0.79

StudentNet ($\alpha = 1e^{-6}$, 음성 임베딩 기반 지식 종류 모델)

Accuracy ≈ 0.76

Macro-F1 ≈ 0.74

UAR ≈ 0.73

모델 배포 및 활용 현황

Hugging Face 허브에 “voice-based-stress-recognition” 으로 공개
현재 다운로드 1.1K 회를 돌파하며 활발히 사용 중



Thank you