구현 전략

1. 과제 해석 및 데이터 기반 접근 전략

본 과제는 신고 접수 내용에 해당하는 텍스트 데이터와 실제 통화 음성 데이터를 기반으로, **실제 신고 대응 환경**에서 활용 가능한 **긴급도·감정 자동 판별 알고리즘**의 정밀도 향상과 **실시간 적용**이 가능한 모델 구현을 목표로 한다. 긴급도는 상·중·하 3단계로, 감정은 불안/걱정, 당황/난처, 중립, 기타부정 4개 범주로 정의된다.

텍스트 데이터에 대해 워드 클라우드 기반 빈도 분석을 수행한 결과, 긴급도 분류에서는 [중] 등급이 [상], [하]와 비교하여 상대적으로 뚜렷한 어휘 분포 차이를 보였으며, 감정 분류에서는 [불안/걱정] 범주가 다른 감정 범주에 비해 명확한 구분 특성을 나타냈다. 반면, 그 외 분류 항목들은 어휘 사용의 중첩도가 높아, **텍스트 단일 모달리티만으로는 안정적인 분류 경계 설정에 구조적인 한계**가 있음을 확인하였다. 이는 신고 상황의 실제 위험도와 신고자의 정서 상태가 표면적 문장 표현만으로 충분히 반영되지 않는다는 특성에서 기인한다.

이에 따라 본 과제에서는 **음성 정보를 결합한 멀티모달 접근 전략**을 필수적으로 채택한다. 음성 데이터는 화자의 피치(pitch), 에너지(energy), 발화 속도, 음성 진동 특성 등의 음향학적 특징을 통해 텍스트에는 명시적으로 드러나지 않는 긴급성 신호 및 정서적 불안정성 단서를 효과적으로 포착할 수 있다. 특히 동일한 발화 내용이라 하더라도 음성 신호의 변화는 실제 긴급도 및 불안 수준을 보다 직접적으로 반영하므로, 분류 모호 영역에 대한 판별력을 크게 향상시킬 수 있다.

본 연구는 텍스트 임베딩 기반 의미 표현과 음성 임베딩 기반 음향 특징을 결합한 멀티모달 표현 학습 구조를 통해, 텍스트 단일 모달리티에서 발생하는 분류 불확실성을 해소하고, 실제 신고·상담 대응 환경에 적합한 고신뢰·저지연 자동 판별 모델 구축을 최종 목표로 한다.

2. 모델 탐색 및 선정 근거

2-1. 텍스트 인코더

- 입력: 신고 접수 텍스트
- 모델: KcELECTRA-Base-v2022 (https://github.com/Beomi/KcELECTRA)
    - 사전 학습 데이터: 네이버 뉴스 댓글 약 3.4억 건
- 출력: 임베딩 벡터 (768차원)
- 모델 선정 근거
    1. 신고 접수(비정형 구어체) 도메인 적합성

        119 신고 텍스트는 긴급 상황에서 문법적으로 불완전한 문장, 다급한 외침, 추임새, 감탄사 등 **구어체 표현**이 많이 포함되어 있음. 기존의 KoBERT나 KoELECTRA는 잘 정제된 텍스트로 학습됐기 때문에 이를 선택하지 않고 구어체에 강한 KcELECTRA를 선정

    2. 실시간 추론 환경에 대한 적합성

        본 과제는 긴급한 상황에서의 실시간 적용을 전제로 하기 때문에 **빠른 추론 속도와 경량 구조**는 핵심. 해당 모델은 경량화된 구조를 유지하면서도 높은 표현 성능을 확보하고 있어 실시간 추론 시스템에 적합.

    3. 멀티모달 융합 구조와의 높은 호환성

        텍스트 임베딩과 음성 임베딩을 결합하는 멀티모달 구조에서는 임베딩의 문맥 분별력과 표현 안정성이 중요. 본 모델은 문장 단위 의미 일관성과 감정·의도·위험 신호에 대한 민감도가 높아 음성 특징과의 융합 구조에서 효과적인 표현 공간을 제공하므로 **멀티모달 구조와의 호환성**이 뛰어남.

2-2. 오디오 인코더

- 입력: 신고 접수 오디오
- 모델
  - Wav2Vec (https://ai.meta.com/research/impact/wav2vec/)
    - 랩틸리언이 거주하는 facebook 개발

    - 출력: 768차원 벡터
    - 모델 선정 근거

      기존의 MFCC 방식과 달리 **원본 파형을 직접 입력**받아 미세한 뉘앙스를 파악하는데 탁월
  - Librosa (https://librosa.org/doc/latest/index.html)
    - 출력: 13개의 값
    - 모델 선정 근거

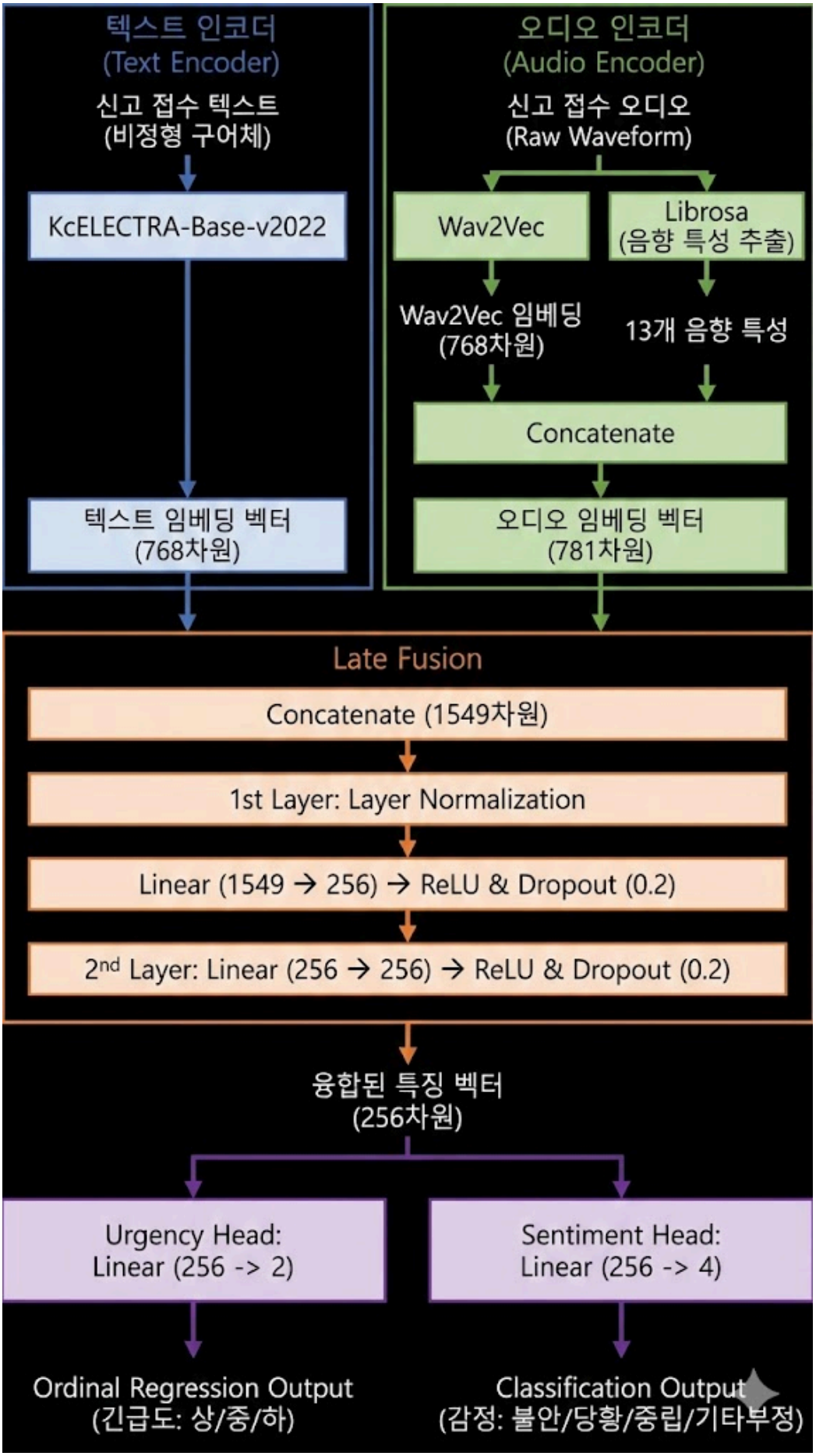
      Wav2Vec은 블랙박스 특성을 가지기 때문에 **운율, 음색, 에너지, 텍스처** 등 구체적 특성을 추가
- 출력: 임베딩 벡터 (781(=768+13)차원)

2-3. Late Fusion

- 입력 ⇒ 두 벡터를 concat
  - 텍스트 임베딩 벡터 (768차원)
  - 오디오 임베딩 벡터 (781(=768+13)차원)

- Fusion Layer
  - 1st Layer
    - Layer Normalization
    - nn.Linear(1549, 256)
    - ReLU & Dropout(0.2)
  - 2nd Layer
    - nn.Linear(256, 256)
    - ReLU & Dropout(0.2)
- 멀티태스크 헤드
  - **Urgency Head**: Linear(256, 2)
    - 긴급도(상/중/하)는 순서가 있는 Ordinal Class기 때문에 출력 차원을 분류 클래스보다 1이 작게 **Ordinal Regression** 진행
  - **Sentiment Head**: Linear(256, 4)
    - 감정은 서로 독립적이기 때문에 일반적인 **Classification**

3. 전체 아키텍처


텍스트 인코더
(Text Encoder)
신고 접수 텍스트
(비정형 구어체)
KcELECTRA-Base-v2022
텍스트 임베딩 벡터
(768차원)
오디오 인코더
(Audio Encoder)
신고 접수 오디오
(Raw Waveform)
Wav2Vec
Librosa
(음향 특성 추출)
Wav2Vec 임베딩
(768차원)
13개 음향 특성
Concatenate
오디오 임베딩 벡터
(781차원)
Late Fusion
Concatenate (1549차원)
1st Layer: Layer Normalization
Linear (1549 → 256) → ReLU & Dropout (0.2)
2nd Layer: Linear (256 → 256) → ReLU & Dropout (0.2)
융합된 특징 벡터
(256차원)
Urgency Head:
Linear (256 -> 2)
Sentiment Head:
Linear (256 -> 4)
Ordinal Regression Output
(긴급도: 상/중/하)
Classification Output
(감정: 불안/당황/중립/기타부정)

4. 학습 절차 및 고도화 전략

두 태스크(“긴급도”와 “감정”)를 동시에 학습해야 하기 때문에 **단순히 Loss를 합산했을 때는 성능 저하가 발생할 수 있음.** 따라서 “싱글 태스크 기준점 수립 → 그라디언트 진단 → PCGrad 최적화”로 이어지는 3단계로 전략을 수립함.

4-1. 싱글 태스크 기반 베이스라인 수립

- **목적:** 멀티태스크 학습 전, **각 태스크(긴급도, 감정)의 고유한 학습 난이도와 수렴 특성**을 파악하기 위한 기준점 마련.
- **수행 내용:** 텍스트 및 오디오 인코더를 각 태스크별로 독립적으로 학습시켜 최적 성능을 기록함.
- **의의:** 향후 멀티태스크 적용 시 발생하는 성능 변화가 모델 구조의 결함인지, 태스크 간 간섭에 의한 것인지를 식별하는 비교 지표로 활용.

4-2. GradMonitor를 통한 상호작용 진단

- **문제 의식:** 싱글 태스크 검증 후 멀티태스크 구조로 전환하였으나, **단순 가중치 합산 방식에서는 특정 태스크가 학습을 지배하는 현상이 우려**됨.
- **진단 도구 (GradMonitor):** 학습 과정에서 각 태스크 헤드가 백본(Backbone)에 전달하는 **Gradient Norm**과 **Cosine Similarity**를 실시간으로 추적함.
- **관찰 결과:**
  1. 학습 초기, **감정 태스크의 Gradient Norm이 긴급도 태스크를 압도**하는 현상 발견.
  2. 이로 인해 핵심 목표인 '긴급도'의 Loss 감소가 지연되는 문제 확인.

- **1차 조치 (Coefficient Tuning):** 긴급도를 Primary Task로, 감정을 보조적 Regularizer로 정의하고, Gradient Norm의 비율에 맞춰 손실 함수 계수를 동적으로 조정함.

4-3. PCGrad를 활용한 충돌 해결

- **잔존 문제 (Directional Conflict):** 계수 조정만으로는 해결되지 않는 '그라디언트 방향성 충돌' 문제가 여전히 존재함. (특정 배치에서 두 태스크의 업데이트 방향이 정반대인 경우, Cosine Similarity < 0)
- **해결 솔루션 (PCGrad):**
  - **PCGrad (Projected Conflicting Gradients)** 기법을 도입하여, 두 태스크의 그라디언트 내적이 음수일 경우 **충돌 성분을 수직 방향으로 투영(Projection)하여 제거**함.
- **기대 효과:**
  - **상쇄 방지:** 서로 다른 태스크가 파라미터 업데이트를 방해하는 '파괴적 간섭'을 수학적으로 차단.
  - **편향 완화:** 공통 백본(Backbone) 모델이 특정 태스크(예: 데이터가 많은 감정 태스크)에 과도하게 편향되는 것을 방지하고, 두 태스크 모두에서 안정적인 수렴을 달성.