



# SemEval-2025 Task 11: Bridging the Gap in Text-Based Emotion Detection

---

Shamsuddeen Hassan Muhammad et al.

Proceedings of SemEval-2025 (ACL), April 2025

**Seo Hyuneun**

Computer Engineering, Seokyeong University

2025.11.16





- \*\*32개 언어\*\* (7개 언어군) × \*\*100,000+ 인스턴스\*\*
- \*\*저자원 언어 포함:\*\* 하우사어(Hausa), 오로모어(Oromo), 요루바어(Yorùbá), 스와힐리어(Swahili) 등
- \*\*기준 한계 해결:\*\* 영어 중심/번역 의존에서 벗어나, 현지 커뮤니티와 협력하여 원본 데이터 수집
- \*\*출처 다양화:\*\* 소셜미디어 (Twitter, Reddit, YouTube), 연설, 뉴스, 문학 텍스트

7개 언어군(Language Families)의 전 세계적 분포





## 데이터셋 구성 - 레이블 설계

- \*\*6개 감정 클래스 (Multi-label):\*\* 하나의 텍스트에 여러 감정 동시 존재 가능
- \*\*감정 강도 (Intensity) - 11개 언어:\*\* 0 (없음) ~ 3 (고강도) 의 4단계 척도
- \*\*주석 신뢰도 (Quality):\*\* SHCMP 60~90% 달성 (고품질 검증 완료)

### 6 Emotion Classes (Multi-label)

Joy (기쁨)

Sadness (슬픔)

Anger (분노)

Fear (두려움)

Surprise (놀라움)

Disgust (혐오)

### 4-point Intensity Scale (0-3)

0 (None)

1 (Low)

2 (Medium)

3 (High)





## 실험 설계: 3가지 Track

---

### Track A: 다중레이블 감정 탐지

---

6개 감정의 \*\*존재(1) / 부재(0)\*\*를 예측

- \*\*평가:\*\* 평균 매크로 F1-score
- \*\*참가:\*\* 114팀 (최다)

### Track B: 감정 강도 예측

---

각 감정의 \*\*강도(0~3)\*\*를 예측

- \*\*평가:\*\* 피어슨 상관계수
- \*\*참가:\*\* 32팀

### Track C: 교차언어 감정 탐지

---

Target 언어의 훈련 데이터 \*\*없이\*\* 예측  
(Zero/Few-shot)

- \*\*평가:\*\* 평균 매크로 F1-score
- \*\*참가:\*\* 51팀





### Track A & B: Team Pai (1위)

- \*\*방법:\*\* LLM 양상을  
(ChatGPT-4o, DeepSeek-V3, Gemma-9b, Qwen-2.5, Mistral)
- \*\*기법:\*\* LoRA 미세조정 + 반복적 프롬프트 최적화
- \*\*후처리:\*\* XGBoost / 가중투표(Weighted Voting)  
양상을

### Track C: Team Deepwave (1위)

- \*\*방법:\*\* Gemma-2 미세조정
- \*\*기법:\*\* CoT(Chain-of-Thought) 프롬프팅
- \*\*전략:\*\* 작업 분해 (Task Decomposition)  
(1. 감성 키워드 식별 → 2. 극성 인식)





### 저자원 언어의 극적 향상 (Low-resource)

- \*\*오로모어 (Oromo):\*\* Baseline 0.126 → 0.616 (약 5배↑)
- \*\*요루바어 (Yorùbá):\*\* Baseline 0.092 → 0.461 (약 5배↑)

### 고자원 언어도 지속 개선 (High-resource)

- \*\*영어 (English):\*\* 최고 F1 0.823 (+16% vs Baseline)
- \*\*러시아어 (Russian):\*\* 최고 F1 0.901 (+7% vs Baseline)

(예: Oromo, Yoruba의 Baseline 대비 Top Team F1-score 비교)

### 여전한 언어별 격차 (The Gap)

고자원 언어 (F1 0.80~0.90) vs 저자원 언어 (F1 0.30~0.50)





## 압도적 트렌드: LLM + Prompting

- \*\*LLM 미세조정\*\* (Gemma-2, LLaMA-3, Qwen 등)
- \*\*프롬프팅:\*\* Few-shot, Zero-shot, Chain-of-Thought (CoT)

## 기타 방법론

- \*\*전통 모델:\*\* BERT 계열 (XLM-RoBERTa, mBERT)도 여전히 유효
- \*\*데이터:\*\* 대부분 팀이 외부 데이터 \*\*미사용\*\* (Few-shot만으로 충분)
- \*\*결과:\*\* Track C (교차언어)에서 저자원 언어 격차 더 심화

(예: LLM-based 70%, BERT-based 20%, Others 10%)





### 감정 강도(Intensity) 레이블의 가치

Track B의 0-3 강도 척도는 "기억 인출 시 감정 강도 변화" 추적에 직접 활용 가능 (e.g., 고강도 슬픔 vs 저강도 슬픔 텍스트)

### 교차언어 전이(Cross-lingual) 학습

한국어 감정 데이터 부족 시, 영어/중국어 등 고자원 언어 데이터로 학습된 모델을 전이(transfer)하여 활용 가능 (Track C 참고)

### LLM 기반 접근법의 효과 (CoT)

CoT 프롬프팅은 복잡한 감정 추론 과정을 명시화하여 모델의 해석 가능성을 높임 (Text → Image Caption → Image Emotion 파이프라인 응용)

### 확장 필요성 (Multimodal)

텍스트(Text)만으로는 한계. 텍스트, 이미지, 음성을 통합하는 멀티모달 접근 및 감정의 시간적 변화(Temporal Dynamics) 추적 필요





### Bottom Line

- SemEval-2025 Task 11은 \*\*저자원 언어\*\* 텍스트 감정 탐지의 새로운 표준 벤치마크를 제시
- \*\*LLM 기반 접근법\*\* (양상블, CoT)이 저자원 환경에서도 강력함을 입증
- \*\*감정 강도(Track B)\*\* + \*\*교차언어 전이(Track C)\*\* 실험은 우리 연구의 직접적인 참고 자료

### 우리 연구 적용 방향 (Next Steps)

1. 텍스트 감정(LLM) → 이미지 캡션 → 이미지 감정 매핑 파이프라인 구축
2. 기억 인출 시 감정 강도(0-3) 변화 추적 실험 설계
3. 한국어 감정 데이터 부족 문제 → 교차언어 전이(Cross-lingual Transfer)로 해결 모색

Next Seminar: 이미지 모달리티 논문 리뷰

