

Enhanced LoRA-SimMatch Model for Mental Health-Related Text Classification in Social Media Environments

Myeong-Guk Jeong*, Ji-Ho Kim*, Dana Yang**

*Student, Dept. of Computer Software, Korean Bible University, Seoul, Korea

**Professor, Dept. of Artificial Intelligence, Korean Bible University, Seoul, Korea

[Abstract]

Mental health text classification based on social media is limited by the scarcity of labeled data and the computational burden associated with full-parameter fine-tuning. To address these limitations, this paper proposes an Enhanced LoRA-SimMatch(ELSM) model that applies LoRA to a pre-trained language model and integrates it with SimMatch-based semi-supervised learning. Through this structure, the issue of pseudo-label noise in semi-supervised learning is effectively refined, enabling precise classification of mental health texts. In experimental results, while training only 1.05% of the total parameters, ELSM achieves relative improvements of 8.04% in Accuracy and 7.84% in Macro-F1, alongside a 36.3% reduction in ECE compared to existing baselines, demonstrating superior classification performance and prediction reliability. As future research, we plan to evaluate the domain extensibility of ELSM in low-resource labeled-data environments across various text processing tasks, including legal document classification, medical text analysis, and toxic text detection.

▶ **Key words:** Semi-Supervised Learning, LoRA, SimMatch, PLM, Mental Health Text Classification

[요 약]

소셜 미디어에서 정신건강 관련 텍스트 분류는 정답 라벨의 희소성과 전체 파라미터 미세조정
에 따른 학습 부담이 있다. 본 논문에서는 이러한 한계를 해결하기 위해 사전 학습 언어모델에
LoRA를 적용하고 SimMatch 기반 준지도학습과 결합한 ELSM(Enhanced LoRA-SimMatch) 모델을
제안한다. ELSM은 기존 SimMatch의 클래스 수준 확률과 인스턴스 수준 유사도 결합 구조에
mxai 인코더와 LoRA를 적용한다. 이러한 구조를 통해 준지도학습의 의사 라벨 노이즈 문제를
효과적으로 정제하고 정신건강 텍스트를 정밀하게 분류한다. 실험 결과 ELSM은 전체 파라미터의
1.05%만 학습하면서 기존 베이스라인 대비 Accuracy 8.04%, Macro-F1 7.84% 향상과 ECE 36.3%
감소를 달성하여 우수한 분류 성능과 예측 신뢰도를 입증했다. 향후 연구로는 법률, 의료, 악성
텍스트 탐지 등 소량 라벨 데이터 환경에서의 도메인 확장성을 검증할 예정이다.

▶ **주제어:** 준지도학습, 저랭크 적응, 유사도 기반 학습, 사전 학습 언어모델, 정신건강 텍스트 분류

- First Author: Myeong-Guk Jeong, Corresponding Author: Dana Yang
- *Myeong-Guk Jeong (jmgjmg102@bible.ac.kr), Dept. of Computer Software, Korean Bible University
- *Ji-Ho Kim (bear356se@bible.ac.kr), Dept. of Computer Software, Korean Bible University
- **Dana Yang (dana1112@bible.ac.kr), Dept. of Artificial Intelligence, Korean Bible University
- Received: 2026. 08. 10, Revised: 2026. 09. 17, Accepted: 2026. 09. 24.

I. Introduction

세계보건기구(World Health Organization, WHO)와 글로벌 질병부담연구(Global Burden of Disease Study, GBD)에 따르면 자살은 2021년 전 세계적으로 주요한 공중보건 문제이며 10~29세에서 주요 사망 원인 중 하나로 보고되었다[1,2]. 이는 정신건강 분류 체계의 필요성을 시사한다. 청년층은 심리적 위기 상황에서 디지털 환경을 통해 자신의 상황을 표현하고 도움을 구하는 경향이 강하며 검색 기록의 변화와 검색량은 통계적으로 자살률과 유의미한 연관성을 보인다[3,4,5]. 특히 트위터나 블로그와 같은 소셜 미디어 플랫폼에 축적되는 문맥 정보와 온라인 자기표현 데이터는 위기 징후를 조기에 탐지하는데 핵심적인 단서가 된다[6].

소셜 미디어 데이터 기반 정신건강 텍스트 분류는 심리적 위기를 탐지하기 위한 연구 분야로 최근 ACL(Association for Computational Linguistics) 등 전산언어학 학회에서 BERT(Bidirectional Encoder Representations from Transformers)와 같은 PLM(Pre-trained Language Model)을 활용한 기법이 주로 사용되고 있다[7,8]. 이러한 모델 연구에서는 정신건강 관련 표현에 나타나는 복잡한 문맥 정보와 감정적 어조의 미묘한 변화를 포착하는 어휘 및 문맥적 표현 능력이 중요하다[9]. 그러나 정신건강 분야의 특성상 환자의 프라이버시 문제와 전문의의 직접적인 진단이 필수이므로 대규모 소셜 데이터 전체에 공인된 정답 라벨을 부여하는 것은 어렵고 많은 비용을 요구하는 한계가 있다[10]. 이러한 데이터 희소성 문제를 해결하기 위해 최근 소량의 라벨 데이터와 대량의 비라벨 데이터를 함께 활용하여 모델의 성능을 높이는 준지도학습(Semi-Supervised Learning)이 연구되고 있다[11]. 그러나 기존의 준지도학습 연구들은 주로 특정 알고리즘을 개별적으로 적용하거나 단일 SNS 데이터에 국한하여 성능을 검증하였다. 또한 대규모 데이터를 반복 학습하는 과정에서 가중치 전체를 업데이트하기 때문에 많은 학습 시간을 소요하는 한계를 지니고 있다[12].

이에 본 연구에서는 준지도학습에 LoRA(Low-Rank Adaptation)를 적용한 ELSM(Enhanced LoRA-SimMatch) 모델을 제안한다. 특히 정신건강 분류 분야는 정답 라벨이 부족하므로 대량의 비라벨 데이터를 활용하는 준지도학습이 필요하다. 이를 검증하기 위해 Reddit과 Twitter의 서로 다른 데이터 구조를 동일한 입력 형태와 분류 클래스로 정형화한다. 이러한 환경을 바탕으로 정신건강 분류 영역에서 라벨 데이터 희소성을 완화하기 위해 의미와 인스턴

스 이중 유사도 결합으로 의사 라벨을 정제하는 ELSM 모델의 성능을 입증한다.

본 논문의 구성은 다음과 같다. 2장에서는 PLM 기반의 정신건강 분류와 지도학습 및 준지도학습 관련 연구를 소개한다. 3장에서는 환경 설계와 ELSM 메커니즘을 기술한다. 4장에서는 모델별 분류 성능을 분석한다. 마지막으로 5장에서는 연구의 성과를 정리하고 향후 발전 방향을 제시하며 결론을 맺는다.

II. Preliminaries

1. PLM-based Mental Health Text Classification

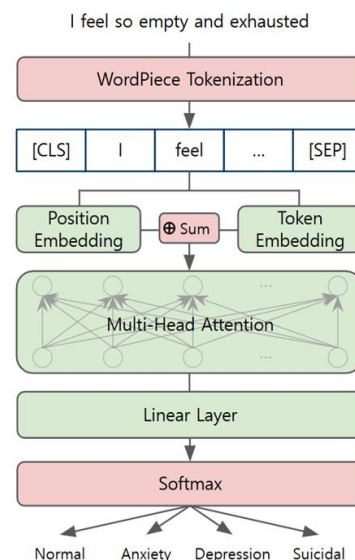


Fig. 1. Overview of the BERT Model Architecture

자연어 처리 분류 과제는 양방향 문맥 표현을 학습하는 BERT의 등장으로 PLM을 활용하는 구조가 표준이다[13]. Fig. 1을 보면 입력된 문장은 토큰화를 거쳐 문장의 시작과 끝을 의미하는 CLS(Classification)와 SEP(Separator)가 포함된 시퀀스로 변환되고 Token Embedding과 Position Embedding의 출력값을 합산한다. 합산된 벡터는 BERT 인코더 블록 내부의 Multi-Head Attention을 통해 양방향 문맥 정보를 학습한다. 최종 레이어는 전체 문맥 표현이 압축된 CLS 토큰의 출력 벡터를 추출하고 Linear Layer와 Softmax 함수를 연산한다. 이후 최종적으로 Normal(정상), Anxiety(불안), Depression(우울), Suicidal(자살) 4개 클래스의 확률을 산출하고 분류 결과를 확정한다.

이러한 연산 과정을 통해 모델은 문장 내 미묘한 감정 변화와 어휘적 맥락을 정교하게 식별한다[9]. 이를 통해 범

용 PLM은 심리적 위기 징후 탐지 영역으로 확산되어 정신 건강 분류 평가에서 높은 정확성을 입증하며 정신건강 분류 연구의 주축이 되었다[7]. 그러나 일반적인 분류용 PLM은 임베딩 이방성(Anisotropy)으로 인해 문장 간 유사도 변별력이 저하되는 한계가 있다[14].

이에 본 연구는 대조 학습(Contrastive Learning)을 통해 이방성을 완화하고 MTEB(Massive Text Embedding Benchmark) 벤치마크에서 임베딩 성능이 검증된 mixedbread-ai/mxbai-embed-large-v1을 백본으로 적용한다[15,16]. 해당 모델은 표면적인 어휘 중첩이 잦은 정신건강 텍스트 간의 미묘한 정서적 차이를 정밀하게 구별하는 데 적합하다.

2. Supervised learning

지도학습은 정답 라벨을 기반으로 입력 데이터와 목표 클래스 간의 매핑 관계를 최적화하는 방식이다[17]. 특히 자연어 처리 분야의 텍스트 분류 과제에서는 입력 문장을 벡터화한 뒤 손실 함수를 최소화하는 방향으로 신경망 가중치를 갱신한다. 또한 SNS 데이터를 활용한 정신건강 분류 영역에서도 지도학습은 사용자들의 자살 위기 징후나 위험 수준을 다중 클래스로 분류하는 표준적인 프레임워크로 활용되고 있다[10].

이러한 방식은 정제된 라벨 환경에서 모든 매개변수가 목표 과제에 맞추어 직접 수렴하므로 탐지 성능이 안정적이다[15]. 그러나 모델 규모가 확장되면서 가중치 매개변수 전체를 업데이트하는 구조는 독립적인 매개변수 집합을 유지해야 하므로 많은 학습시간이 소요된다[18]. 또한 소셜 미디어 글만으로는 작성자의 정신건강 상태를 직접 확인하기 어렵다[19]. 따라서 본 연구는 라벨 데이터의 확보가 어렵고 학습시간에 대한 제약이 존재하는 환경에서 지도학습의 한계를 극복할 대안으로 준지도학습을 적용하고 기존 지도학습 알고리즘과 비교를 통해 제안하는 모델의 우수성을 검증한다.

3. Semi-supervised learning

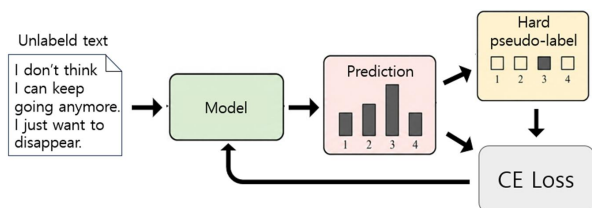


Fig. 2. Pseudo-Labeling Flow for SNS Text Classification

준지도학습은 일부 정답 라벨이 부여된 데이터와 다수의 비라벨 데이터를 함께 활용하여 분류 모델을 학습하는 방식으로 Fig. 2는 준지도학습의 대표적인 알고리즘인 의사 라벨링(pseudo-labeling)이다[20,21]. 이 방식은 모델이 비라벨 데이터에 대한 예측을 수행하고 그중 신뢰도가 높은 결과를 임시 정답인 의사 라벨로 채택한다. 의사 라벨은 예측 확률 분포 전체를 사용하는 소프트 의사 라벨과 최대 확률 클래스를 하나의 정답으로 변환하는 하드 의사 라벨로 구분된다. 이후 모델은 의사 라벨을 정답처럼 사용하여 CE Loss(Cross Entropy Loss)를 계산하고 학습을 진행한다. 이를 통해 정신건강 분류에서는 정답 라벨이 없는 다수의 텍스트에 의사 라벨을 부여하여 부족한 라벨 데이터를 보완할 수 있다. 그러나 모델이 스스로 생성한 예측 결과를 다시 학습에 사용하기 때문에 초기 오분류가 후속 학습에서 반복적으로 강화되는 확증 편향 문제가 발생한다[22].

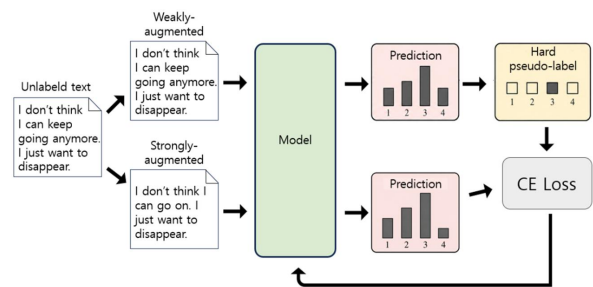


Fig. 3. FixMatch Flow for SNS Text Classification

이를 보완하기 위해 Fig. 3은 의사 라벨의 확증 편향 문제를 완화하는 FixMatch 알고리즘이다[23]. FixMatch는 먼저 약한 증강(Weak augmentation)이 적용된 비라벨 데이터의 예측 중 신뢰도가 높은 샘플만 의사 라벨로 채택한다. 이후 동일 데이터에 강한 증강(Strong augmentation)을 적용한 모델의 예측이 앞서 생성한 의사 라벨과 일치하도록 CE Loss를 계산하고 모델에 적용한다. 이는 동일한 입력에 대하여 어휘나 문장 구조의 변형이라도 텍스트의 의미를 보존하여 클래스 예측이 일관성 있게 유지된다. 이러한 방식은 정신건강 분류에서 문장의 핵심 의미를 유지하는 범위에서 표현을 변형할 수 있다. 또한 데이터 증강 전후의 문장이 동일한 정신건강 클래스로 분류되도록 학습하여 어휘와 문장 표현의 다양성에 대한 모델의 예측 안정성을 높일 수 있다. 그러나 텍스트 기반의 정신건강 분류에서는 단순한 데이터 증강만으로 의미적 뉘앙스의 차이를 구별하기 어려운 한계가 있다[19].

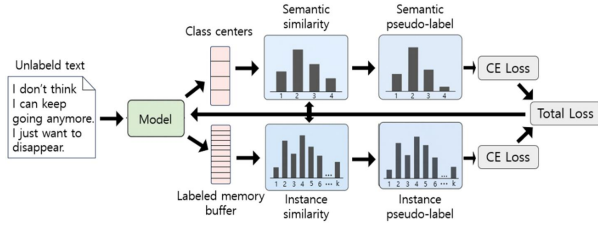


Fig. 4. SimMatch Flow for SNS Text Classification

이에 착안한 SimMatch는 Fig. 4와 같이 클래스 수준의 의미 유사도(Semantic similarity)와 개별 라벨 샘플 수준의 인스턴스 유사도(Instance similarity)를 동시에 고려하는 알고리즘이다[24]. Class centers를 통해 비라벨 샘플과 각 클래스 간의 의미 유사도를 계산하고 Labeled memory buffer를 통해 개별 라벨 샘플 간의 인스턴스 유사도를 계산한다. 이를 바탕으로 의미 의사 라벨(Semantic pseudo-label)과 인스턴스 의사 라벨(Instance pseudo-label)을 상호 보완하여 생성하며 동일한 비라벨 샘플에 서로 다른 증강을 적용할 때 일관된 클래스 예측을 생성하도록 학습한다. 결과적으로 정신건강 텍스트는 동일한 위험 클래스에서도 표현 방식이 다양하고 서로 다른 클래스에서도 유사한 정서적 표현이 나타날 수 있다[19].

이러한 환경에서 SimMatch와 같이 클래스 수준과 개별 문장 수준의 유사도를 함께 고려하는 방식은 정신건강 클래스를 세밀하게 구분한다[24]. 따라서 본 연구는 제한된 라벨 환경에서 정신건강 분류를 위해 대표적인 준지도학습 알고리즘인 Pseudo-Labeling, FixMatch, SimMatch를 동일한 조건에서 비교한다. 이를 바탕으로 비라벨 데이터를 효과적으로 활용하는 SimMatch 프레임워크와 LoRA를 적용하여 탐지 성능을 높인 ELSM 모델을 제안한다.

III. The Proposed Scheme

1. Proposed model

본 연구에서 제안하는 정신건강 텍스트 분류 모델 ELSM(Enhanced LoRA-SimMatch)은 소셜 미디어 데이터의 정답 라벨 희소성 문제를 완화하기 위해 고안되었다. ELSM은 컴퓨터 비전 분야에서 제안된 SimMatch의 의미·인스턴스 유사도 결합 구조를 소셜 미디어 정신건강 텍스트 분류 환경에 맞게 확장한 준지도학습 프레임워크이다. 원본 SimMatch는 CNN으로 추출한 시각적 특징을 기반으로 이미지의 클래스 예측과 개별 이미지 간 유사도 관

계를 학습한다. 반면 ELSM은 대조 학습으로 사전 학습된 mx-bai-embed-large-v1을 인코더로 도입하여 문장의 문맥을 반영한 텍스트 임베딩을 추출한다. 추출한 텍스트 임베딩은 정신건강 범주에 대한 클래스 수준 예측과 라벨 문장과 인스턴스 수준 유사도를 함께 계산한다. 또한 데이터 증강에서는 이미지 변환 대신 정서적 의미 보존을 고려한 약한 텍스트 증강과 역번역 기반 강한 증강을 하나의 파이프라인으로 통합하였다. 이를 통해 문장의 표현을 다양화하면서도 증강 전후에 일관된 정신건강 범주를 예측하도록 학습한다. 원본 SimMatch와 ELSM의 주요 구성 비교는 Table 1과 같다.

Table 1. Structural and Methodological Comparison between SimMatch and Proposed ELSM

Property	Original SimMatch	ELSM
Task	Image Classification	Mental Health Text Classification
Backbone	CNN	mx-bai-embed-large-v1
Embedding Space	Visual feature space	Text embedding space
Parameter training	Full-parameter training	LoRA-based fine-tuning
Augmentation	Weak/strong image augmentation	Weak text augmentation and back-translation
Similarity information	Semantic and instance	Semantic and instance

ELSM은 위와 같은 아키텍처를 바탕으로 Fig. 5와 같이 구성되어 4단계로 동작한다.

- Step 1: 수집된 소셜 미디어 정신건강 텍스트를 정형화하고 증강 데이터를 생성한다.
- Step 2: mx-bai-LoRA 모듈로 학습 효율성을 높이며 정서 문맥 특징 벡터를 추출한다.
- Step 3: 이중 출력 헤드를 통해 의미 유사도와 인스턴스 유사도를 연산하여 고품질 의사 라벨을 정제한다.
- Step 4: 복합 손실 함수를 최적화하여 최종 텍스트 범주를 분류한다.

이어지는 3.2절부터 3.5절까지는 제안한 아키텍처의 각 단계별 구성 요소와 필요성을 순차적으로 상세히 기술한다.

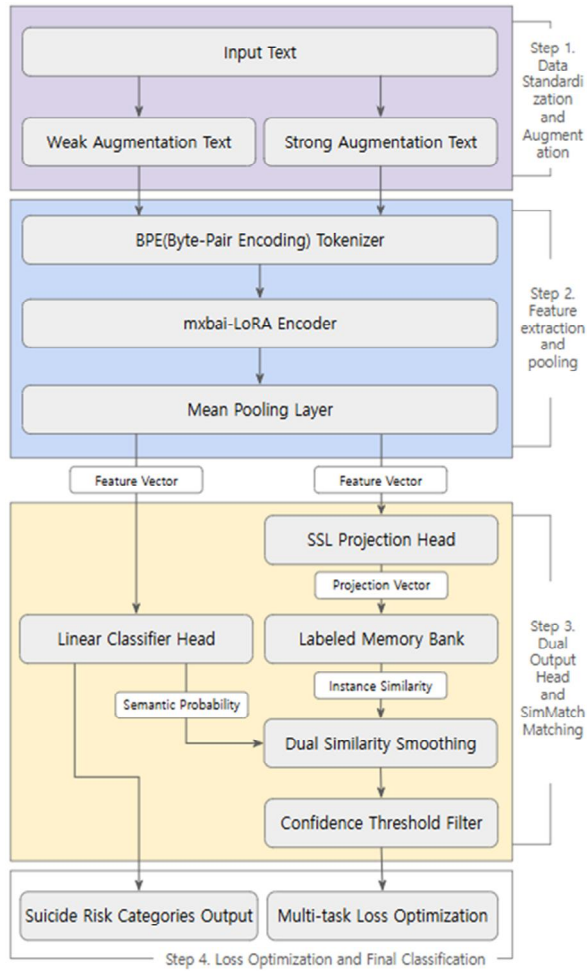


Fig. 5. The ELSM Architecture for Mental Health-Related Text Classification

2. Data Standardization and Augmentation (Step 1)

SNS는 데이터 작성 양식과 메타데이터 구조가 서로 상이하다. 예를 들어 Reddit 게시글은 제목과 본문이 분리된 마크다운 서식이며 Twitter 텍스트는 멘션, 해시태그, URL 링크가 포함된 단문 형태로 표현된다. 본 연구에서는 이처럼 SNS별로 노이즈와 불필요한 메타데이터를 정규표현식 기반 규칙으로 정제하여 순수 정서적 문맥 텍스트로 변환하고 단일 텍스트 구조로 통합한다.

이러한 환경적 구성을 기반으로 Fig. 5의 Step 1에서는 수집된 비정형 소셜 미디어 텍스트를 모델 학습에 적합하게 가공하고 준지도학습의 일관성 정규화를 위한 데이터 증강을 수행한다. 먼저 입력된 정신건강 Input Text x 를 Weak Augmentation Text x_w 와 Strong Augmentation Text x_s 두 가지 형태로 변환한다. x_w 는 x 의 정서적 뉘앙스를 그대로 유지하는 데이터이며 x_s 는 x 의 의미를 보존하면서 역번역을 통해 표현을 다변화한 데이터이다. 이를 통해 x 에 대해 두 가지 변형 텍스트를

생성하여 모델이 표현 변화에도 일관된 클래스를 예측하도록 유도한다.

3. Feature Extraction and Pooling (Step 2)

이전 단계에서 도출한 x_w 와 x_s 를 기반으로 이 단계에서는 텍스트의 핵심 정서 문맥을 담은 고차원 특징 벡터 v 를 효율적으로 추출한다. Fig. 5의 Step 2 모듈 구조와 같이 입력 텍스트 x_w 와 x_s 는 BPE Tokenizer를 거쳐 최대 512개 길이의 토큰 시퀀스로 변환하여 mxbai-LoRA 인코더 모듈에 입력한다. mxbai-LoRA 인코더 모듈은 사전 학습된 mxbai의 가중치 W_0 를 고정하고 Attention Layer에 LoRA 어댑터 행렬 A 와 B 를 추가한다. 입력 텍스트 $x \in \{x_w, x_s\}$ 에 대한 모듈의 토큰 특징 행렬 h 는 다음 식과 같이 연산한다.

$$h = W_0 x + \Delta W x = W_0 x + B A x$$

여기서 $\Delta W = B A$ 는 저랭크 행렬의 곱으로 표현하는 가중치 변화량이며 전체 가중치를 갱신하지 않고 LoRA 어댑터 매개변수만 선택적으로 학습한다. 이러한 방식은 사전 학습 모델의 표현력을 유지하고 제한된 라벨 데이터 환경에서 과도한 가중치 변화를 방지한다. Fig. 5에서 제시한 것처럼 mxbai-LoRA 인코더를 통해 추출된 h 는 Mean Pooling Layer에 입력하여 정신건강 텍스트 전체의 문맥 정보가 압축된 1024차원의 특징 벡터 v 로 변환한다. v 는 다음 단계인 이중 출력 헤드에서 의미 예측 확률 연산과 인스턴스 유사도 계산을 위한 공통 입력값으로 사용한다.

4. Dual Output Head and SimMatch Matching (Step 3)

Step 3은 5개의 모듈로 구성된다. 먼저 Fig. 5에서 Step 3 왼쪽 상단의 Linear Classifier Head 모듈은 텍스트의 정서적 맥락을 바탕으로 정신건강 클래스를 직접 판별한다. 오른쪽 상단의 SSL Projection Head 모듈은 유사도 연산에 최적화된 저차원 임베딩 공간으로 특징을 전달한다. 다음 단계인 Labeled Memory Bank 모듈은 라벨 데이터가 부족한 환경에서 과거 학습된 정답 샘플들을 잊지 않고 보관하여 비라벨 문장과 비교한다. 이어 Dual Similarity Smoothing 모듈은 의미 예측 확률과 인스턴스 유사도의 두 관점을 가중 결합하여 단일 예측 오차를 보완하고 정제된 의사 라벨을 합성한다. 마지막으로 Confidence Threshold Filter 모듈은 예측 확률이 임계값을 초과하는 신뢰도 높은 고품질 샘플만 선별하여 노이

즈 수렴을 방지하는 역할을 맡는다.

이러한 모듈 구조를 기반으로 Step 3에서는 이전 단계에서 추출한 v 를 두 개의 독립된 헤드로 전달하여 클래스 수준의 의미 유사도와 텍스트 수준의 인스턴스 유사도를 함께 연산한다. 이를 통해 단일 예측 오차를 보완하고 고품질 의사 라벨을 정제한다. 먼저 SSL Projection Head 모듈은 v 를 128차원으로 축소하고 L2 정규화 연산을 적용하여 프로젝션 벡터 v_p 를 출력한다. 이후 Labeled Memory Bank 모듈에서 보관 중인 정답 라벨 특징 z 와의 코사인 유사도를 연산하고 유사도 스케일링을 조절하는 매개변수 t 를 적용한 Softmax 연산을 통해 인스턴스 유사도 q 를 도출한다.

$$q = \text{Softmax}\left(\frac{\text{sim}(v_p, z)}{t}\right)$$

이와 동시에 Linear Classifier Head 모듈은 v 를 입력받아 4개의 정신건강 클래스에 대한 Semantic Logits 점수를 산출하고 Softmax 연산을 적용하여 의사 확률 분포 p_{sem} 으로 변환한다. 이를 통해 Dual Similarity Smoothing 모듈은 두 헤드에서 각각 산출한 p_{sem} 과 q 를 9:1 비율로 가중 결합하여 정제된 정신건강 의사 라벨 $\hat{p}$ 를 산출한다.

$$\hat{p} = \alpha \cdot p_{sem} + (1 - \alpha) \cdot q \quad (\alpha = 0.9)$$

마지막으로 Confidence Threshold Filter 모듈은 예측 확률이 임계값 0.95를 초과하는 신뢰도 높은 텍스트만 학습에 반영하도록 지시함수 기반 Mask를 적용한다.

$$Mask = \mathbb{I}(\max(\hat{p}) \geq 0.95)$$

이러한 과정으로 정제된 의사 라벨과 선별된 텍스트는 다음 단계에서 지도학습 손실과 준지도학습 일관성 손실을 통합하여 모델 전체를 최적화하는 데 활용한다.

5. Loss Optimization and Final Classification (Step 4)

손실 최적화 단계에서는 Fig. 5의 Step 4에서 제시하는 Multi-task Loss Optimization 구조에 따라 지도학습 손실과 준지도학습 일관성 손실을 적용한 복합 손실 L 을 계산하고 모델을 최적화한다. 전체 복합 손실 L 은 다음 식과 같이 연산한다.

$$L = L_s + \lambda_u L_{sem} + \lambda_{in} L_{in}$$

여기서 L_s 는 정답 라벨이 존재하는 정신건강 텍스트에 대한 지도학습 Cross-Entropy Loss이다. L_{sem} 은 x_s 의 예측 분포와 $\hat{p}$ 간의 의미적 일관성을 강제하는 손실이며 L_{in} 은 개별 텍스트 간 인스턴스 유사도 관계를 보존하는 일관성 손실이다. 비지도 손실 가중치 λ_u 와 인스턴스 손실 가중치 λ_{in} 은 각각의 일관성 손실 반영 비율을 조절한다. 이때 mxbai 인코더의 가중치는 유지하고 학습 과정에서는 Step 2에서 추가한 LoRA 어댑터 행렬만 선택적으로 갱신하여 모델의 학습 시간을 크게 절감한다. 최종적으로 L 기반의 모델 학습이 완료된 이후 Fig. 5의 Step 4에서 제시하는 Mental Health Categories Output 구조에 따라 훈련된 mxbai 인코더와 선형 분류 헤드를 통해 입력 텍스트의 정서 문맥을 분석한다. 이를 바탕으로 Normal, Anxiety, Depression, Suicidal 4개의 클래스 중 가장 높은 확률을 갖는 클래스를 판별한다.

IV. Evaluation

1. Experimental Setup

본 연구에서는 정신건강 관련 텍스트를 Normal, Anxiety, Depression, Suicidal 네 가지 클래스로 분류하여 모델의 성능을 평가한다. 학습 데이터로는 Kaggle에 공개된 Depression Detection using Sentiment Analysis 데이터셋을 사용하였다. 해당 데이터셋은 Reddit과 Twitter에서 수집된 공개 텍스트를 포함하고 있으며 데이터셋에 포함된 라벨은 선행 연구의 분류 체계를 참고하여 Normal, Anxiety, Depression, Suicidal 4개 클래스만 선별하였다[19]. 평가 데이터로는 독립적인 벤치마크 검증을 위해 Hugging Face에 공개된 ourafla의 Mental Health Text Classification Dataset을 사용하였다. 해당 데이터셋 또한 동일한 4개 클래스로 구성되어 있으며 두 공개 데이터셋 모두 사용자 개인정보 보호를 위해 계정 식별자가 제거된 상태로 제공되므로 데이터 분할은 게시물 단위로 수행되었다. 이때 발생할 수 있는 데이터 누출(data leakage)을 방지하기 위해 정규화 과정에서 중복 문장을 제거하였으며 학습 데이터와 평가 데이터 간의 중복 텍스트를 배제하여 데이터의 독립성을 확보하였다.

실험에 사용된 전체 데이터셋의 구성과 클래스별 표본 수는 Table 2와 같다. 학습 데이터는 라벨 텍스트 3,716

개와 비라벨 텍스트 34,119개로 구성하였으며 평가에는 학습과 분리된 3,192개의 라벨 텍스트를 테스트 데이터로 활용한다. 이를 기반으로 지도학습 모델은 라벨 데이터만 사용하고 준지도학습 모델은 라벨 데이터와 비라벨 데이터를 함께 활용한다. 실험에 적용된 세부 하이퍼파라미터 설정은 Table 3과 같으며 모든 비교 실험에는 동일한 데이터 분할과 학습 조건을 적용한다. 추가적으로 모델의 예측 정확도와 더불어 정신건강 분류와 같은 고위험 도메인에서의 예측 신뢰도를 검증하기 위해 Accuracy, Macro-F1과 함께 ECE(Expected Calibration Error)를 평가지표로 활용하였다. 특히 의사 라벨링 기반 준지도학습에서 발생하기 쉬운 Overconfidence 편향을 정량화하기 위해 예측 확률 구간을 15 bins로 균등 분할하여 모델의 확신도와 실제 정확도 간의 평균 오차를 측정하였다.

Table 2. Class distribution and dataset statistics

Class/Category	Train	Test	Total
Normal	929	798	1,727
Anxiety	929	798	1,727
Depression	929	798	1,727
Suicidal	929	798	1,727
Labeled	3,716	3,192	6,908
Unlabeled	34,119	-	34,119
Overall Total	37,835	3,192	41,027

Table 3. Hyperparameter Configuration

Property	Value
Maximum sequence length	512
Epochs	5
steps per epoch	2000
batch size	4
Learning rate	2e-4
Random seed	42
LoRA rank	8
LoRA scale	16
LoRA dropout	0.1

본 실험은 세 단계로 나누어 진행한다. 동일한 Full fine-tuning 조건에서 지도학습과 준지도학습의 결과를 비교하여 비라벨 데이터 활용에 따른 성능 변화를 검증한다. 이후 각 학습 방식에 LoRA를 적용하여 전체 파라미터를 학습하는 방식 대비 학습 효율성과 분류 성능을 비교 분석한다. 마지막으로 기본 지도학습 모델과 ELSM 모델의 클래스별 성능 및 특징 공간 표현을 비교하여 단일 지표만으로 파악하기 어려운 세밀한 정신건강 분류 성능을 다각도로 평가한다.

2. Comparison of Supervised and Semi-supervised Learning Performance

Table 4는 Full fine-tuning 조건에서 Supervised와 준지도학습 방법의 성능을 비교한 결과이다. 모든 준지도학습 알고리즘은 지도학습 대비 우수한 Accuracy와 Macro-F1을 기록하였으며 ECE도 낮게 나타났다. 특히 SimMatch는 Supervised 대비 Accuracy와 Macro-F1을 각각 5.89%, 6.14% 향상시켰으며 ECE를 29.45% 감소시켜 예측 신뢰도를 크게 개선하였다. 이러한 결과는 비라벨 데이터 활용이 분류 성능과 예측 신뢰도 향상에 기여함을 보여준다. 또한 인스턴스 간 유사도와 클래스 수준의 특징 관계를 함께 활용하는 방식이 클래스 간 표현 중첩이 큰 정신건강 텍스트 분류에 긍정적으로 작용한 것으로 나타난다.

Table 4. Full Fine-Tuning Comparison

Method	Accuracy	Macro-F1	ECE
Supervised	0.6855	0.6877	0.2937
Pseudo-labeling	0.7014	0.6990	0.2832
FixMatch	0.7043	0.7113	0.2557
SimMatch	0.7259	0.7299	0.2072

Fig. 6은 학습 단계에 따른 각 방법의 Macro-F1 변화를 나타내며 최종 성능 평가는 이 과정에서 가장 높은 성능을 기록한 최적 모델 가중치(best checkpoint)를 사용하여 산출했다. 모든 방법에서 성능이 일정하게 증가하지 않았으며 학습 과정에서 반복적으로 상승하고 하락했다. SimMatch는 대부분의 학습 단계에서 다른 방법보다 높은 Macro-F1을 유지했고 8,000 step 부근에서 다시 성능이 상승했다. 반면 FixMatch와 Pseudo-labeling은 6,000 step 이후 성능이 크게 감소한 뒤 일부 회복하는 경향을 보였다. Supervised는 변화 폭이 비교적 작았지만 전반적인 성능은 준지도학습 방법보다 낮았다. 이러한 결과는 SimMatch가 최종 성능뿐만 아니라 학습 과정에서도 상대적으로 높은 분류 성능을 보였음을 나타낸다.

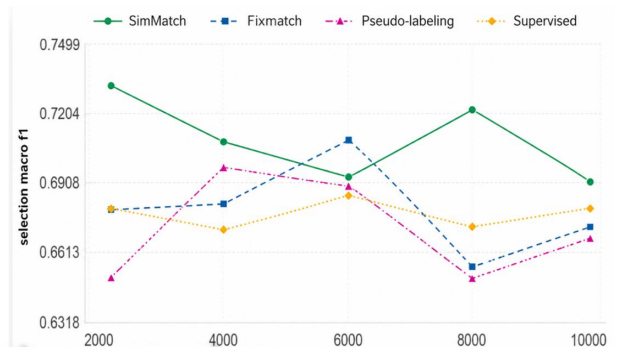


Fig. 6. Macro-F1 Comparison

3. Performance Comparison Based on LoRA Application

Table 5는 Full fine-tuning과 LoRA 조건에서 학습한 모델의 성능을 비교한 결과이다. ELSM은 지도학습 모델 대비 Accuracy 8.04%, Macro-F1 7.84% 높았으며 ECE도 36.3% 감소하여 제안 모델이 가장 우수한 예측 신뢰도를 보였다. 또한 SimMatch-Full과 비교할 때도 ELSM은 Accuracy와 Macro-F1을 각각 2.03%, 1.6% 향상했고 ECE를 9.7% 낮추었다. 이러한 결과는 SimMatch 기반 LoRA 모델의 분류 성능과 예측 신뢰도를 함께 향상하는데 효과적임을 보여준다.

Table 5. LoRA Comparison

Method	Accuracy	Macro-F1	ECE
Full Supervised	0.6855	0.6877	0.2937
Full SimMatch	0.7259	0.7299	0.2072
LoRA Supervised	0.7133	0.7112	0.2652
LoRA Pseudo-labeling	0.7246	0.7245	0.2522
LoRA FixMatch	0.7334	0.7298	0.2099
ELSM	0.7406	0.7416	0.1871

Table 6은 단일 시드 편향을 배제하고 모델의 안정성을 검증하기 위해 세 가지 무작위 시드를 적용한 반복 실험 결과이다. ELSM은 SimMatch-Full 대비 평균 Accuracy와 Macro-F1에서 각각 3.13%, 2.36% 향상했다. 특히 ELSM의 최저 Macro-F1이 기존 모델의 최고 성능을 상회하였으며 Macro-F1 표준편차가 또한 절반 수준으로 감소하여 특정 random seed에 의존하지 않고 안정적인 성능을 유지하였다. 또한 평균 ECE는 19.34% 감소하여 무작위 시드 변화 조건 전반에서 성능 우위와 예측 신뢰도 개선을 확인하였다.

Table 6. Performance stability comparison across random seeds

Method	Seed	Accuracy	Macro-F1	ECE
SimMatch-Full	41	0.7083	0.7125	0.2176
	42	0.7259	0.7299	0.2072
	43	0.7093	0.7169	0.2607
Mean $\pm$ Std	-	0.7145 $\pm$ 0.0099	0.7198 $\pm$ 0.0090	0.2285 $\pm$ 0.0284
ELSM	41	0.7350	0.7359	0.1805
	42	0.7406	0.7416	0.1871
	43	0.7350	0.7328	0.1853
Mean $\pm$ Std	-	0.7369 $\pm$ 0.0032	0.7368 $\pm$ 0.0045	0.1843 $\pm$ 0.0034

Table 7은 Full fine-tuning 대비 ELSM의 파라미터 수와 학습 시간 및 추론 시간을 비교한 결과이다. ELSM은 전체 파라미터의 1.05%만 학습하여 학습 시간을 기존 대비 29.7% 단축하였으며 표본당 추론 시간 또한 기존 대비 11.36% 단축하였다. 이를 통해 제안 모델이 학습 파라미터 수와 훈련 시간을 줄이면서 추론 단계의 연산 효율성을 안정적으로 확보하였다.

Table 7. LoRA Parameter Comparison

Parameter Type	Full fine-tuning	ELSM
Total Params	335,145,988	338,701,316
Trainable Params	335,145,988	3,559,428
Trainable Ratio	100%	1.05%
Train Time	62.19 min	43.70 min
Inference Time	9.86 ms	8.74 ms

4. Ablation Study on SimMatch Hyperparameters

Table 8은 confidence threshold를 0.95로 고정하고 semantic similarity와 instance similarity의 결합 가중치 α 에 따른 성능 변화를 비교한 결과이다. 단일 유사도만 적용한 조건들과 비교했을 때 ELSM은 Instance-only($\alpha=0.0$) 대비 Accuracy와 Macro-F1이 각각 1.86%, 1.51% 향상되었으며 ECE는 19.39% 감소하여 예측 신뢰도가 개선되었다. 또한 Semantic-only($\alpha=1.0$) 대비 Accuracy와 Macro-F1이 각각 2.92%, 2.33% 향상되었으며 ECE는 동등한 수준을 유지하였다. 이는 semantic similarity를 중심으로 활용하면서 instance-level 정보를 보조적으로 반영하는 구성이 단일 유사도 모델 대비 분류 성능과 예측 신뢰도 향상에 효과적임을 보여준다.

Table 8. Effect of Semantic-Instance Weight α

α	Accuracy	Macro-F1	ECE
0.0	0.7271	0.7306	0.2321
0.1	0.7259	0.7305	0.2222
0.3	0.7221	0.7232	0.2389
0.7	0.7265	0.7258	0.2310
0.9 (ELSM)	0.7406	0.7416	0.1871
1.0	0.7196	0.7247	0.1844

Table 9는 α 는 0.9로 고정하고 의사 라벨의 confidence threshold에 따른 성능 변화를 비교한 결과이다. threshold 0.95에서 Accuracy 0.7406과 Macro-F1 0.7416으로 가장 높은 분류 성능을 보였다. 이는 0.95가 본 실험에서 가장 적절한 설정이며 threshold를 0.99까지 높이면 ECE가

0.2301로 증가하여 지나치게 엄격한 의사 라벨 선택이 예측 신뢰도 측면에서는 불리하게 작용함을 확인하였다.

Table 9. Ablation Study on Confidence Threshold

Confidence Threshold	Accuracy	Macro-F1	ECE
0.00	0.7133	0.7155	0.1825
0.70	0.7253	0.7240	0.1961
0.80	0.7287	0.7292	0.1422
0.90	0.7190	0.7223	0.1707
0.95 (ELSM)	0.7406	0.7416	0.1871
0.99	0.7372	0.7392	0.2301

5. Class-wise and Feature Space Analysis

Table 10은 Supervised-Full과 ELSM의 클래스별 F1-score를 비교한 결과이다. 구체적으로 Normal, Anxiety, Depression 클래스의 F1-score가 각각 7.74%, 6.95%, 4.57% 향상되었다. 특히 본 연구의 핵심 타깃인 Suicidal 클래스는 F1-score가 기존 0.6271에서 0.7029로 12.09% 가장 큰 폭으로 상승하였다. 이는 ELSM이 특정 클래스에 한정되지 않고 전반적인 분류 성능을 개선했음을 보여준다.

Table 10. Class-wise F1-score

Category	Supervised-Full F1	ELSM F1
Anxiety	0.7588	0.8115
Depression	0.5860	0.6128
Normal	0.7790	0.8393
Suicidal	0.6271	0.7029

Fig. 7은 테스트 데이터에서 추출한 특징 벡터를 UMAP으로 시각화한 결과이다. Supervised-Full(a)에서는 클래스 간 특징 분포가 전반적으로 겹쳤으나 ELSM(b)에서는 Anxiety, Depression, Suicidal 클래스가 비교적 구별 가능한 영역을 형성했다. Normal 클래스의 일부 특징도 독립된 군집을 나타냈다. 이러한 특징 공간의 구조적 분리 양상은 앞서 확인한 클래스별 F1-score의 전반적인 상승 경향을 뒷받침한다.

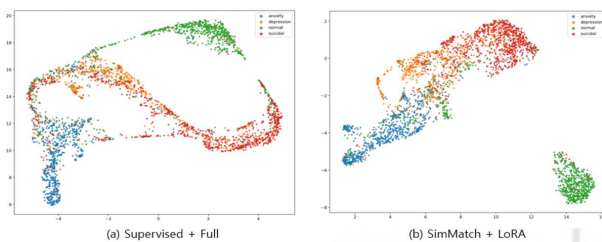


Fig. 7. UMAP Feature Visualization

추가적으로 UMAP 시각화의 주관적 해석을 보완하기 위해 Table 11과 같이 정량적 군집 분리도 평가를 수행하였다. ELSM은 Supervised-Full 대비 Silhouette이 63.58% 향상되었고 Davies-Bouldin은 27.55% 감소하였다. Calinski-Harabasz는 112.66%, Centroid Inter/Intra는 55.67% 향상되었다. 이를 통해 ELSM이 지도학습 기반 Full fine-tuning보다 클래스 내부의 응집도와 클래스 간 분리도가 전반적으로 향상된 표현 구조를 형성했음을 확인하였다.

Table 11. Quantitative clustering evaluation in feature space

Metric	Supervised-Full	ELSM
Silhouette	0.1535	0.2511
Davies-Bouldin	2.0370	1.4758
Calinski-Harabasz	720.94	1533.09
Centroid Inter/Intra	1.4741	2.2947

종합하면 ELSM은 Full fine-tuning 기반 지도학습보다 전체 성능과 클래스별 F1-score를 높였으며 ECE도 낮췄다. 또한 전체 파라미터의 1.05%만 학습하면서도 가장 높은 Macro-F1을 기록했다. UMAP 결과에서도 클래스 간 분리 구조가 상대적으로 뚜렷하게 나타났다. 이러한 결과는 제한된 라벨 데이터와 다수의 비라벨 데이터를 함께 활용하는 준지도학습과 일부 파라미터만 학습하는 LoRA를 적용한 방식이 정신건강 관련 텍스트 분류에 효과적임을 보여준다.

V. Conclusions

본 연구에서는 SNS 기반 정신건강 분류 분야의 정답 라벨 데이터 희소성의 한계를 완화하기 위해 ELSM을 제안한다. ELSM은 mxbai-LoRA 인코더 모듈로 사전 학습 모델의 문맥 표현력을 보존하면서 학습 효율성을 확보하고 이중 출력 헤드 구조로 정신건강 텍스트를 세밀하게 분류한다. 성능 평가 결과 ELSM은 정신건강 텍스트 분류에서 Accuracy 0.7406, Macro-F1 0.7416, ECE 0.1871을 달성하여 기존 지도학습 모델과 준지도학습 모델 대비 우수한 탐지 정확도와 높은 예측 신뢰도를 입증했다.

그러나 본 연구는 한계점과 윤리적 고려사항이 있다. 제안 모델의 예측 결과는 전문의의 실제 임상 진단을 대체할 수 없으며 사전 스크리닝 및 보조 도구 목적으로 제한해야 한다. 특히 고위험 상태의 사용자를 놓치는 False Negative

와 사용자에게 부정적인 영향을 유발하는 False Positive의 오분류 위험을 방지하기 위해 실제 서비스 적용 시에는 전문가 검증이 수반하는 인간 참여형(Human-in-the-loop) 체계가 필요하다. 또한 본 평가는 모델 간 비교의 공정성을 위해 클래스 균등 환경에서 수행되었으나 실제 SNS의 클래스 불균형으로 인해 발생할 수 있는 비라벨 데이터 편향 문제를 포괄적으로 검증하는 데 한계가 있다.

향후 연구로는 ELSM을 다양한 도메인으로 확장하는 연구를 추진할 예정이다. 특히 법률 문서 분류, 의료 기록 분석, 사이버불링 및 악성 텍스트 탐지 등 라벨링 비용이 높고 소량의 라벨 데이터만 존재하는 다양한 비정형 텍스트 처리 분야에 ELSM을 적용하여 본 프레임워크의 도메인 확장성과 범용 분류 우수성을 검증하고자 한다[25].

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MOE) (No. RS-2026-25601684).

REFERENCES

- [1] World Health Organization, "Suicide worldwide in 2021: Global health estimates," World Health Organization, 2025.
- [2] GBD 2021 Suicide Collaborators, "Global, regional, and national burden of suicide, 1990-2021: A systematic analysis for the Global Burden of Disease Study 2021," *The Lancet Public Health*, Vol. 10, No. 3, pp. e189-e202, March 2025. DOI: 10.1016/S2468-2667(25)00006-4
- [3] K. Wong, C. S. Chan, M. Chan, C. Wong, Q. Cheng, C. Xiong, and P. Yip, "Who seeks help online? Comparing online and offline help-seeking preferences amongst youths with suicidal ideation," *Journal of Affective Disorders*, Vol. 292, pp. 21-29, September 2021. DOI: 10.1016/j.jad.2021.05.056
- [4] P. A. Areán, A. Pratap, H. Hsin, T. K. Huppert, K. E. Hendricks, P. J. Heagerty, T. Cohen, C. Bagge, and K. A. Comtois, "Perceived utility and characterization of personal Google search histories to detect data patterns proximal to a suicide attempt in individuals who previously attempted suicide: Pilot cohort study," *Journal of Medical Internet Research*, Vol. 23, No. 5, Art. No. e27918, May 2021. DOI: 10.2196/27918
- [5] J. Y. Son, J. H. Han, S. C. Kim, W.-S. Choi, and H. J. Hong, "Korean adolescent suicide and search volume for 'self-injury' on internet search engines," *Frontiers in Psychiatry*, Vol. 14, Art. No. 1186754, June 2023. DOI: 10.3389/fpsy.2023.1186754
- [6] B. Kysnes, G. J. Hjetland, E. Haug, I. Holsen, and J. C. Skogen, "The association between sharing something difficult on social media and mental well-being among adolescents. Results from the 'LifeOnSoMe'-study," *Frontiers in Psychology*, Vol. 13, Art. No. 1026973, November 2022. DOI: 10.3389/fpsyg.2022.1026973
- [7] I. Levkovich, and M. Omar, "Evaluating of BERT-based and Large Language Mod for Suicide Detection, Prevention, and Risk Assessment: A Systematic Review," *Journal of Medical Systems*, Vol. 48, Art. No. 113, December 2024. DOI: 10.1007/s10916-024-02134-3
- [8] R. J. Zevallos, A. M. Schoene, and J. E. Ortega, "The first multilingual model for the detection of suicide texts," *Proceedings of the Second Workshop on Scaling Up Multilingual & Multi-Cultural Evaluation*, pp. 1-11, Abu Dhabi, United Arab Emirates, January 2025.
- [9] A. O. J. Ibitoye, O. O. Oladimeji, and O. F. W. Onifade, "Contextual emotional transformer-based model for comment analysis in mental health case prediction," *Vietnam Journal of Computer Science*, Vol. 12, No. 3, pp. 277-299, August 2025. DOI: 10.1142/S2196888824500192
- [10] J. Wu, K. He, O. P. Kampman, G. S. K. Chung, W. W. B. Goh, Q. Lin, J. Xu, Y. Du, X. Shang, E. Cambria, and M. Feng, "Language-Centered Mental Healthcare with Large Language Models: A Comprehensive Survey Across the Care Continuum," *SenticNet*, preprint, 2026. <https://sentic.net/language-centered-mental-healthcare.pdf>
- [11] N. Farruque, R. Goebel, S. Sivapalan, and O. R. Zaïane, "Depression symptoms modelling from social media text: an LLM driven semi-supervised learning approach," *Language Resources and Evaluation*, Vol. 58, No. 3, pp. 1013-1041, September 2024. DOI: 10.1007/s10579-024-09720-4
- [12] P. Zhang, Z. Mai, Q.-H. Nguyen, and W.-L. Chao, "Revisiting Semi-Supervised Learning in the Era of Foundation Models," *Advances in Neural Information Processing Systems*, Vol. 38, pp. 66049-66078, San Diego, CA, USA, and Mexico City, Mexico, December 2025. DOI: 10.52202/085713-1984
- [13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171-4186, Minneapolis, MN, USA, June 2019. DOI: 10.18653/v1/N19-1423
- [14] B. Li, H. Zhou, J. He, M. Wang, Y. Yang, and L. Li, "On the Sentence Embeddings from Pre-trained Language Models," *Proceedings of the 2020 Conference on Empirical Methods in*

- Natural Language Processing (EMNLP), pp. 9119-9130, Online, November 2020. DOI: 10.18653/v1/2020.emnlp-main.733
- [15] T. Gao, X. Yao, and D. Chen, "SimCSE: Simple Contrastive Learning of Sentence Embeddings," Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 6894-6910, Online and Punta Cana, Dominican Republic, November 2021. DOI: 10.18653/v1/2021.emnlp-main.552
- [16] L. Caspari, K. Ghosh Dastidar, S. Zerhoubi, J. Mitrovic, and M. Granitzer, "Beyond benchmarks: Evaluating embedding model similarity for retrieval augmented generation systems," arXiv preprint arXiv:2407.08275, 2024. DOI: 10.48550/arXiv.2407.08275
- [17] Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. S. Yu, and L. He, "A Survey on Text Classification: From Traditional to Deep Learning," ACM Transactions on Intelligent Systems and Technology, Vol. 13, No. 2, Art. No. 31, April 2022. DOI: 10.1145/3495162
- [18] Z. Han, C. Gao, J. Liu, J. Zhang, and S. Q. Zhang, "Parameter-efficient fine-tuning for large models: A comprehensive survey," arXiv preprint arXiv:2403.14608, 2024. DOI: 10.48550/arXiv.2403.14608
- [19] R. Sawhney, H. Joshi, S. Gandhi, and R. R. Shah, "Towards ordinal suicide ideation detection on social media," Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 22-30, Virtual Event, Israel, March 2021. DOI: 10.1145/3437963.3441805
- [20] X. Yang, Z. Song, I. King, and Z. Xu, "A Survey on Deep Semi-Supervised Learning," IEEE Transactions on Knowledge and Data Engineering, Vol. 35, No. 9, pp. 8934-8954, September 2023. DOI: 10.1109/TKDE.2022.3220219
- [21] D.-H. Lee, "Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks," ICML 2013 Workshop on Challenges in Representation Learning, Atlanta, GA, USA, June 2013.
- [22] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8, Glasgow, UK, July 2020. DOI: 10.1109/IJCNN48605.2020.9207304
- [23] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," Advances in Neural Information Processing Systems, Vol. 33, pp. 596-608, Online, December 2020.
- [24] M. Zheng, S. You, L. Huang, F. Wang, C. Qian, and C. Xu, "SimMatch: Semi-Supervised Learning with Similarity Matching," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14471-14481, New Orleans, LA, USA, June 2022. DOI: 10.1109/CVPR52688.2022.01407
- [25] F. van der Sluis, and E. L. van den Broek, "Model interpretability enhances domain generalization in the case of textual complexity modeling," Patterns, Vol. 6, No. 2, Art. No. 101177, February 2025. DOI: 10.1016/j.patter.2025.101177

Authors



Myeong-Guk Jeong is currently pursuing the B.S. degree at Korean Bible University, Seoul, South Korea. His research interests include artificial intelligence, natural language processing, semi-supervised learning and text analysis.



Ji-Ho Kim is currently pursuing the B.S. degree at Korean Bible University, Seoul, South Korea. His research interests include natural language processing, deep learning and data analysis.



Dana Yang received the B.S. degree in the Department of Computer Software from Korean Bible University in 2014, the M.S and Ph.D. degree in computer science and engineering from Ewha Womans University.

Seoul, South Korea, in 2023. She is currently an Assistant Professor with the Department of Artificial Intelligence, Korean Bible University. Her research interests include machine learning, blockchain, authentication, steganography, flying adhoc network (FANET) security, and D2D network security. She received the Outstanding Paper Award from the 20th International Conference on Advanced Communication Technology (ICACT), in 2018.