

A Study on an Early Diagnosis System for Children Language Developmental Disorders Using KoBERT

Won-Cheol Park*

*Ph. D, Dept. of Computer Engineering, Kongju National University, Cheonan, Korea

[Abstract]

This study proposes an AI-based early diagnosis system for children's language developmental disorders using the Korean BERT (KoBERT) model. A dataset of 10,000 sentences was constructed, including typical expressions, autism spectrum disorder, stuttering, and communication disorders. The system was trained and evaluated to classify each type of disorder. Experimental results showed that the KoBERT-based system achieved the highest performance with an F1-score of 0.88, and the confusion matrix analysis confirmed accurate classification even between similar linguistic patterns. This study demonstrates that deep learning-based natural language processing (NLP) techniques can effectively support early diagnosis and intervention of language developmental disorders, contributing to clinical and educational applications.

▶ **Key words:** KoBERT, Language Developmental Disorders, Natural Language Processing (NLP), Deep Learning, Early Diagnosis

[요 약]

본 연구는 한국어 특화 딥러닝 언어 모델인 KoBERT를 활용하여 아동의 언어 발화 데이터를 분석하고, 언어 발달 장애를 조기 진단하는 AI 기반 시스템을 개발하였다. 총 10,000문의 데이터 셋을 구축하고, 정상 표현과 함께 자폐 스펙트럼 장애, 말더듬, 의사소통 장애 유형을 포함하여 모델을 학습하고 평가하였다. 실험 결과, KoBERT 기반 시스템이 F1-Score 0.88로 가장 우수한 성능을 기록하였으며, 혼동 행렬 분석에서도 유사 표현 간 정확한 분류 성능을 보였다. 본 연구는 딥러닝 기반 자연어 처리 기술이 아동 언어 발달 장애의 조기 진단과 중재에 효과적으로 활용될 수 있음을 확인하였으며, 향후 임상 및 교육 현장 적용 가능성을 제시한다.

▶ **주제어:** KoBERT, 언어 발달 장애, 자연어 처리, 딥러닝, 조기진단

I. Introduction

아동의 언어 발달은 인지 발달과 사회적 상호작용의 중요한 지표로, 아동의 전반적인 성장과 발달에 밀접한 영향을 미친다[1][2]. 그러나 언어 발달 장애는 초기 단계에서 명확하게 드러나지 않는 경우가 많아 조기 발견이 어렵고, 이로 인해 아동의 학습 능력과 사회성 발달에 심각한 영향을 초래할 수 있다[3-5]. 특히 자폐 스펙트럼 장애, 말더듬, 의사소통 장애와 같은 언어 발달 장애는 아동의 언어 표현과 발화 패턴에서 특징적으로 나타나지만, 전문가의 관찰과 면담만으로 정확히 진단하기에는 한계가 존재한다[6][7]. 기존 언어 발달 장애 진단 방식은 주로 전문가의 주관적 평가와 관찰에 의존하고 있으며, 이로 인해 객관성과 일관성을 확보하는 데 어려움이 따른다[8][9]. 특히 아동이 낯선 환경이나 긴장감으로 인해 평소와 다른 언어적 행동을 보이는 경우가 많아, 현장 진단의 신뢰성도 낮아질 수 있다[10]. 최근 국내외에서는 언어 발달 장애 아동의 조기 발견과 맞춤형 지원의 중요성이 더욱 강조되고 있다[11]. 언어 발달 지연이나 장애가 조기에 발견될 경우, 적절한 중재와 치료를 통해 아동의 언어 능력과 사회적 발달을 효과적으로 개선할 수 있다는 연구 결과들이 다수 발표되고 있다. 그러나 현실적으로 전문가의 직접 관찰과 면담만으로는 시간과 비용이 많이 소요되며, 아동의 상태에 따라 정확한 진단이 어려운 경우가 많다. 이러한 한계를 보완하기 위해 인공지능(AI)과 자연어 처리(NLP) 기술을 접목한 자동 진단 시스템의 필요성이 대두되고 있으며, 최근 한국어 처리할 수 있는 대규모 언어 모델의 발전은 이러한 시스템 개발 가능성을 높이고 있다. 특히 KoBERT와 같은 한국어 특화 딥러닝 모델은 문장 내 의미와 맥락을 효과적으로 분석할 수 있어, 아동의 언어적 특성을 정교하게 파악하는 데 유용하다[12]. 따라서 본 연구는 딥러닝 기반 자연어 처리 모델인 KoBERT를 활용하여 아동의 자연스러운 언어 발화 데이터를 분석하고, 언어 발달 장애를 조기에 진단할 수 있는 AI 기반 시스템을 개발하고자 한다. 특히 자폐 스펙트럼 장애(Autism Spectrum Disorder), 말더듬(Stuttering), 의사소통 장애(Communication Disorder), 정상 표현(Normal)을 포함한 다양한 언어적 특성을 분석하여 정확한 분류 모델을 구축하고, 전문가의 진단을 보조할 수 있는 체계를 마련하는 것을 목표로 한다. 이러한 시스템은 향후 실시간 분석과 모니터링 기능으로 확장 가능하며, 보다 효율적이고 객관적인 언어 발달 평가 도구로 활용될 수 있을 것이다. 이를 통해 언어 발달 장애의 조기 발견과 정확한 진단이 가능해

져, 아동의 효과적인 치료와 맞춤형 교육 계획 수립에 실질적으로 기여할 것으로 기대한다.

II. Preliminaries

1. Related works

1.1 Early Diagnosis of Language Disorders

아동의 언어 발달 장애 조기 진단과 딥러닝 기반 자연어 처리(NLP) 기술의 발전은 상호 보완적으로 발전해 왔다. 특히, 한국어 언어 모델인 KoBERT의 등장으로 한국어 텍스트 처리 분야에서 다양한 연구가 수행되고 있다. 언어 발달 장애는 아동의 전반적인 발달과 학습에 큰 영향을 미치며, 조기 발견과 개입이 중요하다. Kim(2022)은 언어 발달 장애 아동이 단순언어장애, 읽기장애, 자폐스펙트럼장애, 지적장애, 청각장애 등 다양한 사례군으로 나타난다고 보고하였다[13]. 또한, Thordardottir와 Ellis Weismer (2002)는 언어장애 아동이 일반 또래 아동에 비해 유의하게 적은 낱말을 회상한다고 지적하였다. 이러한 연구들은 언어장애 아동의 조기 진단과 개입의 필요성을 강조하고 있다[14].

1.2 Deep Learning-Based NLP Trends

딥러닝의 발전으로 자연어 처리 분야에서도 획기적인 성과가 나타나고 있다. 특히, BERT(Bidirectional Encoder Representations from Transformers)는 사전 학습된 언어 모델로서 다양한 자연어 처리 작업에서 우수한 성능을 보였다. 한국어 자연어 처리 분야에서도 BERT 기반 모델이 개발되어 활용되고 있다. KoBERT는 SKT Brain에서 개발한 한국어 특화 BERT 모델로, 뉴스, 위키백과 등의 한국어 데이터를 기반으로 학습되었다. 이 모델은 감성 분석, 문장 분류, 명사 추출 등 다양한 한국어 처리 작업에 활용되고 있다[15][16]. 최근 BERT 기반 모델들은 한국어 자연어 처리 분야에서 우수한 성능을 보이며 다양한 텍스트 분류 작업에서 활용되고 있다. 특히, 한국어 데이터에 특화된 KoBERT 모델은 감성 분석, 뉴스 분류, 욕설 탐지 등 다양한 과제에서 기존 머신러닝 기반 모델 대비 우수한 결과를 나타내고 있다[17]. 한 연구에서는 구글 지도 리뷰 데이터를 활용해 SVM, 1D-CNN, 2D-CNN, CNN-LSTM, KoBERT, KoGPT2 모델을 활용하여 리뷰 유용성 예측 모델을 설계하고 성능을 비교하였다. 실험 결과, CNN-LSTM 모델이 72.74%의 성능을 기록했고, KoBERT와 KoGPT2는 각각 73.22%, 75.74%로 딥러닝

모델보다 우수한 성능을 보였다. 최종적으로 대형 언어 모델과 딥러닝 모델을 통합한 모형이 76.37%의 정확도로 가장 높은 성능을 기록하며, 복잡한 텍스트 의미를 효과적으로 반영하고 예측 안정성을 높일 수 있음을 확인했다[18]. 또한, BERT 기반 모델의 성능은 데이터의 양과 질에 따라 안정적으로 증가하는 경향을 보이며, 약 10,000개 수준의 데이터셋에서도 0.85~0.90 수준의 성능이 보고되고 있다 [19]. 특히 문장 구조가 비교적 단순한 분류(task)일수록 KoBERT의 문맥 이해 능력이 정확도 향상에 크게 기여하는 것으로 알려져 있다. 최근 한국어 기반 대규모 언어 모델의 발전과 함께, KoBERT와 같은 딥러닝 모델을 다양한 한국어 자연어 처리 과제에 적용하려는 시도가 활발히 이루어지고 있다. 특히, KoBERT 모델은 한국어 문장 내 의미와 문맥을 효과적으로 분석할 수 있어, 복합적인 언어적 특성을 지닌 데이터에서도 높은 분석 성능을 보이는 것으로 평가된다. 이러한 모델의 발전은 아동의 언어 발달 상태를 자동으로 분석하고 조기 진단하는 시스템 개발에도 적용 가능성을 높이고 있다. 아동의 발화 데이터는 반복적 표현, 문법적 오류, 비정형적인 문장 구조 등 복합적인 언어적 특징을 지니고 있어 전통적인 기계학습 모델로는 분석에 한계가 존재한다. 반면, KoBERT와 같은 한국어 특화 딥러닝 모델은 이러한 복합적인 언어적 특성을 정교하게 학습하고 분석할 수 있다. 따라서, 아동 언어 발달 장애 조기 진단 시스템 구축에 있어 가장 효과적인 기술적 대안으로 평가된다. 기존 연구들은 주로 감성 분석, 뉴스 분류, 욕설 탐지 등 일반 자연어 처리 과제에 초점을 맞추어왔다. 이러한 기존 연구들과 비교할 때, 본 연구는 KoBERT 모델을 아동의 언어 발화 데이터에 적용하여 언어 발달 장애를 조기에 진단하는 새로운 접근을 시도하였다. 이는 기존 연구들과 차별화되는 시도로, KoBERT의 문맥 이해 능력을 임상적 조기 선별 시스템에 활용했다는 점에서 연구적 독창성을 지닌다.

III. The Proposed System

1. System Architecture

다음의 Fig. 1은 제안 시스템의 전체적인 구성도이며 크게 4단계로 구성되어 있다.

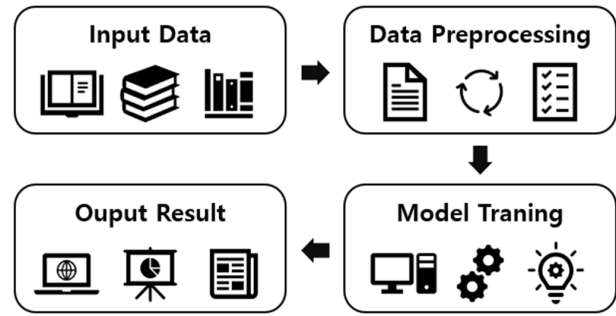


Fig. 1. Architecture of the proposed system

첫 번째 단계는 아동 대화 입력 단계로, 아동의 자연스러운 대화 데이터를 수집하고 입력하는 과정이다. 이 단계에서는 아동이 발화한 문장을 CSV파일로 입력하며, 말더듬 표현이나 반복적 문장 등이 포함될 수 있다.

두 번째 단계는 데이터 전처리 및 임베딩 과정으로, 입력된 아동 대화 데이터를 분석에 적합한 형태로 가공한다. 이를 통해 제안 시스템이 학습하고 예측할 수 있는 입력 형태로 데이터를 가공한다.

세 번째 단계는 제안 시스템의 학습 및 추론 단계로, 전처리된 데이터를 기반으로 KoBERT 모델이 아동의 언어 발화를 학습하고 각 문장을 장애 유형에 따라 분류한다. 학습 과정은 Hugging Face의 Trainer를 활용하여 구성하였으며, 세부 학습 전략과 하이퍼파라미터 설정은 이후 3.3절에서 자세히 설명한다. 학습이 완료된 모델은 테스트 데이터를 입력 받아 각 문장을 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현 중 하나로 분류한다. 예측 결과는 Confusion Matrix 및 시각화 그래프 형태로 출력되며, 이는 전문가가 직관적으로 판단을 내릴 수 있도록 돕는다.

마지막으로 장애 유형 분류 결과 출력 단계에서는 제안 시스템의 학습 및 추론 결과를 전문가가 직관적으로 확인하고 활용할 수 있도록 제공한다. 시스템은 입력된 아동의 대화 데이터를 분석한 후, 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 중 하나로 분류 결과를 출력한다. 분류 결과는 시각적으로 제공함으로써 전문가가 아동의 언어 발달 상태와 장애 여부를 빠르고 정확하게 파악하고, 이후의 상담이나 평가 과정에서 실질적으로 활용할 수 있도록 지원한다.

본 연구의 제안 시스템은 Python 3.12 환경에서 구현하였으며, 주요 라이브러리로는 Hugging Face의 transformers, datasets, 그리고 PyTorch (torch)를 사용하였다. 모델 학습과 추론은 Hugging Face의 Trainer 모듈을 기반으로 수행되었으며, 실험 과정의 재현성과 효율성을 확보하기 위해 로그와 결과 파일은 별도로 저장되도록 설계하였다. 제안 시스템의 학습과 추론 과정은

NVIDIA T4 GPU가 제공되는 Google Colab 환경에서 진행되었으며, GPU 자원을 활용하여 모델의 학습 속도와 성능을 최적화하였다. 데이터 전처리 및 결과 시각화에는 pandas, matplotlib, seaborn 라이브러리를 활용하여 시스템 전반의 처리 과정과 결과 분석을 지원하였다. 전체 구현 및 실험 코드는 Windows 기반 환경에서 작성 및 관리되었으며, 클라우드 환경과 연계하여 모델 학습을 수행하였다. 본 연구에서는 초기 실험 단계에서 SKT에서 제공하는 'skt/kobert-base-v1' 모델 사용을 검토하였으나, 해당 모델은 SentencePiece 기반 구조로 Hugging Face의 Trainer 및 AutoTokenizer와의 호환성 문제가 발생하였다. 이에 따라 실험의 효율성과 재현성 확보를 위해 Hugging Face에서 제공하는 'monologg/kobert' (SKT Brain KoBERT) 사전 학습 모델을 최종 선택하여 학습 및 평가를 진행하였다.

2. Data Construction and Preprocessing

본 연구에서는 언어 발달 장애 조기 진단을 위한 맞춤형 데이터셋을 직접 구축하였다. 데이터셋 구축 과정은 다음과 같은 절차로 진행되었다. 먼저 각 장애 유형별 실제 상담데이터 114건을 수집하고, 이를 토대로 각 장애 유형별 언어적 특성을 반영하여 초기 샘플 문장을 수작업으로 생성한 후, 데이터 다양성과 양을 확보하기 위해 문장 구조 변형, 반복 발화 삽입, 문법 오류 추가 등의 데이터 증강 기법을 적용하여 문장을 확장하였다. 이를 통해 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현의 특성을 균형 있게 반영한 총 10,000문장의 데이터셋을 완성하였다. 구체적으로 데이터셋은 자폐 스펙트럼 장애 2,088문장, 말더듬 2,043문장, 의사소통 장애 2,000문장, 정상 표현 1,869문장으로 구성되었으며, 각 클래스가 가능한 균형 있게 분포되도록 설계하였다. 각 문장은 해당 장애 유형의 언어적 특징을 반영하였으며, 실제 상담 현장에서 관찰될 수 있는 표현과 패턴을 적용하여 제작하였다. 말더듬 유형은 동일한 단어나 음절이 반복되는 표현을 중심으로 생성하였고, 자폐 스펙트럼 장애 유형은 대화의 흐름이 단절되거나 특정 문장이 반복되는 특징을 반영하였다. 의사소통 장애 유형은 문장의 구성이나 문법적 오류가 발생하거나 의사 전달이 불명확한 문장으로 구성하였다. 정상 유형은 일상 대화에서 나타나는 자연스러운 문장들로 구성하여 비교 기준으로 활용하였다. 구축된 데이터셋은 id, text, label, type의 4개 항목으로 구성된 CSV 파일 형태로 저장하였다. 이후 제안 시스템의 학습을 위해 데이터 전처리 과정을 수행하였다. 텍스트 데이터에서 불필요한 특수문자

와 공백을 제거하고, Hugging Face에서 제공하는 KoBERT 전용 토큰라이저를 활용하여 토큰화 및 정수 인코딩 과정을 거쳤다. 전처리된 데이터는 모델 학습에 최적화된 Tensor 형태로 변환되어 입력되며, 이 과정은 제안 시스템의 성능 향상에 중요한 역할을 한다.

3. Model Training Inference Process

본 연구의 제안 시스템은 KoBERT 모델을 기반으로 학습과 추론을 수행한다. 학습 과정에서는 전처리된 데이터를 입력하여 제안 시스템이 아동의 언어 발화 데이터를 이해하고, 장애 유형을 분류할 수 있도록 훈련하였다. 학습은 Hugging Face의 Trainer와 TrainingArguments를 활용하여 진행되었으며, 총 10 Epoch 동안 수행되었다. 학습률 (learning rate)은 $2e-5$ 로 설정하여 모델이 안정적으로 학습할 수 있도록 하였고, GPU 환경에서 효율적인 학습을 위해 배치 크기는 8로 지정하였다. 또한 매 Epoch마다 모델의 성능을 평가하는 evaluation_strategy = 'epoch'을 적용하여 학습 진행 상황을 지속적으로 확인하였다. 학습 과정 중 10 step마다 로그를 기록해 손실값 변화를 모니터링하고, 학습 결과와 로그는 각각 './kobert_results'와 './logs' 디렉터리에 저장되도록 구성하였다. 학습이 완료된 후, 테스트 데이터를 입력하여 장애 유형을 예측하고 최종 분류 결과를 도출하였다. 제안 시스템은 입력된 문장을 분석하여 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 중 하나로 분류하며, 최종 출력 결과는 전문가가 직관적으로 확인하고 활용할 수 있도록 제공된다.

IV. Experimental Results

1. Experimental Design

본 연구에서는 앞서 3.2절에서 구축한 총 10,000문장의 아동 언어 발화 데이터셋을 기반으로 실험을 진행하였다. 데이터셋은 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현 네 가지 클래스로 구성되었으며, 각 클래스는 가능한 균형 있게 분포되도록 설계하였다. 각 문장은 장애 유형별 언어적 특성을 반영하여 생성되었으며, 데이터 구축 과정에 대한 구체적인 설명은 3.2절에 제시하였다. 전체 데이터셋은 학습용(80%)과 테스트용(20%)으로 분리하여 활용하였으며, 학습 데이터는 모델 학습에, 테스트 데이터는 모델 성능 검증에 사용하였다. 성능 비교를 위해 제안한 KoBERT 기반 모델 외에 전통적인 기계학습 및 딥러닝 기반 모델을 함께 실험하였다. 비교 모델로는 TF-IDF 기반 Support Vector

Machine (TF-IDF + SVM), TF-IDF 기반 Extreme Gradient Boosting (TF-IDF + XGBoost), 그리고 BiLSTM(Long Short-Term Memory) 모델을 선정하였다. 비교 모델로 TF-IDF + SVM, TF-IDF + XGBoost, BiLSTM을 선정할 이유는 각각 전통적인 기계학습 기법, 앙상블 학습 기법, 딥러닝 기반 시퀀스 모델을 대표하기 때문이다. 이를 통해 다양한 접근법 간 성능 차이를 객관적으로 비교하고, 문맥 기반 모델(KoBERT)이 아동 언어 데이터 분류 정확도 향상에 기여하는지를 실험적으로 검증하고자 하였다. 이러한 비교를 통해 KoBERT 기반 제안 시스템의 성능을 객관적으로 평가하고, 기존 접근 방식과의 차별성을 확인하고자 하였다. 모든 모델은 동일한 데이터셋을 기반으로 학습 및 테스트를 수행하였다. 모델 성능 평가는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score를 기준으로 진행되었으며, 추가적으로 분류 결과의 직관적 해석을 위해 Confusion Matrix와 성능 비교 그래프를 활용하였다. 이러한 실험 설계를 통해 KoBERT 기반 제안 시스템의 분류 성능을 객관적으로 검증하고, 기존 접근 방식 대비 차별성과 우수성을 입증하고자 하였다.

2. Model Performance Results and Discussion

본 연구에서는 제안한 KoBERT 기반 모델과 비교 모델들의 성능을 평가하기 위해 동일한 테스트 데이터를 대상으로 실험을 수행하였다. 성능 평가는 정확도, 정밀도, 재현율, F1-Score를 기준으로 측정하였으며, 모델의 전반적인 분류 성능을 비교하였으며 모델별 최종 성능 비교 결과는 다음의 Table 1과 같다.

Table 1. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
1	0.78	0.77	0.76	0.765
2	0.82	0.81	0.80	0.805
3	0.80	0.79	0.78	0.785
4	0.88	0.87	0.89	0.88

Table 1에 제시된 모델은 1번 모델(TF-IDF + SVM), 2번 모델(TF-IDF + XGBoost), 3번 모델(BiLSTM), 4번 모델(KoBERT)이다. 실험 결과, 4번 모델인 KoBERT 기반 제안 시스템은 다른 비교 모델들에 비해 전반적으로 우수한 성능을 보였다. 특히 F1-Score에서 KoBERT 모델은 0.88로, 전통적인 TF-IDF 기반 기계학습 모델과 BiLSTM 딥러닝 모델을 모두 상회하는 결과를 기록하였다. 이는 KoBERT 모델이 한국어 문맥과 표현 구조를 보다 효과적

으로 이해하고 학습할 수 있는 구조적 강점을 지니고 있음을 보여준다. TF-IDF + SVM과 TF-IDF + XGBoost 모델은 상대적으로 낮은 성능을 기록하였으며, 이는 전통적인 벡터화 방식이 문장의 문맥적 의미를 충분히 반영하지 못하기 때문으로 판단된다. 특히, 아동의 언어 발화에서 나타나는 반복적 표현, 문법적 오류, 문장 구성의 불안정성 등 복잡한 언어적 특징을 효과적으로 처리하기에는 한계가 있는 것으로 분석된다. BiLSTM 모델 역시 KoBERT 보다는 낮은 F1-Score를 기록하였으나, 전통적인 TF-IDF 기반 모델보다는 상대적으로 높은 성능을 보였다. 이는 BiLSTM 모델이 문장의 순차적 구조를 고려하여 학습하기 때문으로 판단되지만, 한국어의 특수성과 긴 문맥적 정보 처리 측면에서는 KoBERT 모델이 더욱 효과적인 것으로 확인되었다. 또한, KoBERT 모델의 학습 과정에서 Epoch이 증가함에 따라 Training Loss와 Validation Loss가 지속적으로 감소하는 양상을 보였다. 특히 5 Epoch 이후부터 Validation Loss가 급격히 감소하였으며, 최종 10 Epoch에서는 다음의 Fig. 2, Table 2와 같은 학습 손실 결과를 기록하였다.

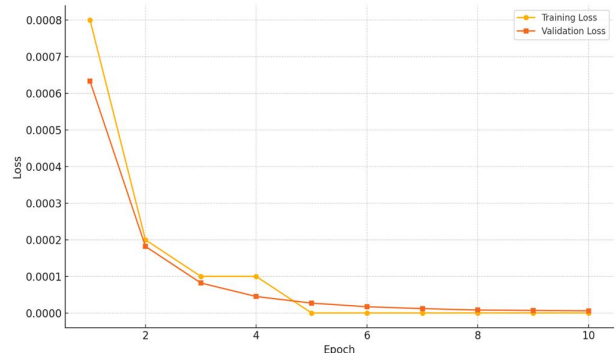


Fig. 2. Loss Reduction Trend over Epochs

KoBERT 모델은 학습 과정에서도 Loss 값이 빠르게 감소하고 안정적으로 수렴하는 모습을 보여주었으며, 이는 과적합 없이 효과적인 학습이 이루어졌음을 의미한다. 특히, 5 Epoch 이후부터 Validation Loss가 급격히 감소하고 10 Epoch에서 0.000006까지 수렴하는 결과는 KoBERT 모델의 학습 안정성과 성능의 신뢰성을 높이는 근거가 된다. 결과적으로, KoBERT 기반 제안 시스템은 아동의 언어 발화 데이터를 분석하여 장애 유형을 정확하게 분류할 수 있는 가능성을 보여주었으며, 기존 전통적인 분석 방법에 비해 실질적인 성능 향상과 함께 실무 적용 가능성을 높였다는 점에서 연구적 의미가 크다고 판단된다.

Table 2. Training and Validation Loss per Epoch

Epoch	Training Loss	Validation Loss
1	0.000800	0.000633
2	0.000200	0.000182
3	0.000100	0.000082
4	0.000100	0.000045
5	0.000000	0.000027
6	0.000000	0.000017
7	0.000000	0.000012
8	0.000000	0.000008
9	0.000000	0.000007
10	0.000000	0.000006

3. Confusion Matrix Analysis

각 모델의 분류 성능을 보다 구체적으로 분석하기 위해 혼동 행렬(Confusion Matrix)을 생성하였다. 다음의 Fig. 3은 KoBERT 모델의 혼동 행렬 결과이다.

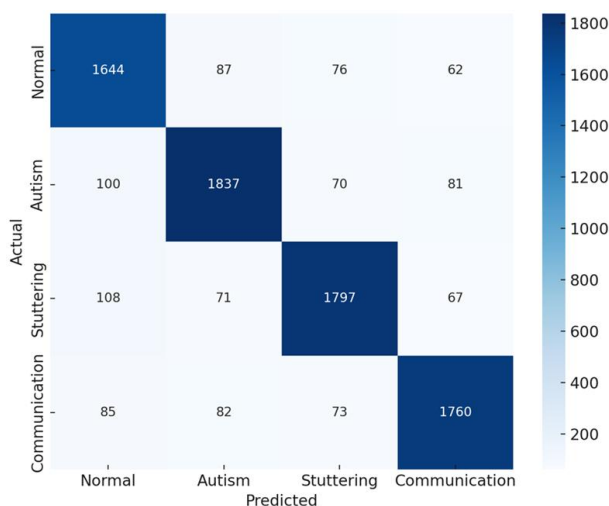


Fig. 3. Confusion Matrix of KoBERT

KoBERT 모델은 가장 안정적이고 우수한 분류 성능을 기록했으며, 전반적으로 오분류가 매우 적은 것으로 나타났다. 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현 모든 클래스에서 높은 True Positive(TP) 값을 기록하였다. 자폐 스펙트럼 장애의 경우 총 2,088개 중 1,837개가 정확히 분류되었고, 오분류된 251건은 다른 클래스(주로 의사소통 장애 등)로 소량 오분류되었다. 의사소통 장애는 1,760건이 정확히 분류되었고, 오분류는 240건이 발생하여 오차 범위가 매우 낮았다. 이러한 결과는 KoBERT 모델이 한국어 문맥과 표현 구조를 효과적으로 학습하여, 유사 표현 간 혼동 없이 장애 유형을 정확히 구분할 수 있음을 보여준다.

비교 모델인 TF-IDF + SVM, TF-IDF + XGBoost, BiLSTM은 KoBERT에 비해 전반적으로 오분류가 많은 양상을 보였다. 먼저 SVM 모델은 자폐 스펙트럼 장애 2,088개 중 1,597개를 정확히 분류하였고, 491건이 주로 의사소통 장애로 오분류되었다. 정상 표현은 1,869개 중 1,429개가 정확히 분류되었으며, 440건이 자폐 또는 말더듬으로 잘못 분류되는 등 전반적인 분류 성능이 낮았다. XGBoost 모델은 자폐 스펙트럼 장애 2,088개 중 1,680개를 정확히 분류하고, 408건을 오분류하는 등 SVM 대비 개선된 성능을 보였으나, 말더듬과 의사소통 장애 간 혼동이 여전히 존재하였다. BiLSTM 모델은 자폐 스펙트럼 장애에서 1,639개를 정확히 분류하고, 449건을 오분류하였다. 정상 표현에서도 1,467개를 정확히 분류하였지만 402건이 오분류되는 결과를 보였으며, KoBERT 모델 대비 낮은 성능을 기록하였다. 이러한 결과는 전통적인 벡터 기반 모델(SVM, XGBoost)이나 순차 학습 기반 모델(BiLSTM)이 아동 발화 데이터의 복합적인 언어 패턴과 문맥 정보를 KoBERT만큼 효과적으로 학습하지 못했음을 의미한다. 종합적으로, KoBERT 기반 제안 시스템은 모든 비교 모델 대비 가장 높은 분류 정확도와 안정적인 성능을 보였으며, 아동의 언어 발달 장애 조기 진단을 위한 실질적인 적용 가능성을 확인할 수 있었다.

V. Discussion

본 연구에서는 딥러닝 기반 자연어 처리 모델인 KoBERT를 활용하여 아동의 언어 발화 데이터를 분석하고, 언어 발달 장애를 조기에 진단할 수 있는 가능성을 실험적으로 검증하였다. 실험 결과, KoBERT 기반 제안 시스템은 정확도, 정밀도, 재현율, F1-Score 모든 지표에서 기존의 전통적 기계학습 모델(TF-IDF + SVM, TF-IDF + XGBoost) 및 BiLSTM 모델 대비 가장 우수한 성능을 보였다. 모델 성능 평가를 위한 시각화 과정에서는 scikit-learn과 matplotlib 라이브러리를 활용하여 Confusion Matrix를 생성하였으며, 이를 통해 KoBERT 모델의 분류 성능과 오분류 양상을 직관적으로 분석하였다. KoBERT 모델은 정상, 자폐 스펙트럼 장애, 말더듬, 의사소통 장애 모두에서 높은 True Positive(TP) 값을 기록하였다.

오류 분석 결과, 주요 오분류는 자폐 스펙트럼 장애와 의사소통 장애 간 경계가 모호한 경우에서 발생하였다. 예를 들어, "나는... 음... 그... 학교..."처럼 발화의 흐름이 끊

기거나 반복되는 경우, 의사소통 장애와 자폐 스펙트럼 장애 간 혼동이 나타났다. 그러나 오분류 비율은 전체 데이터셋 대비 매우 낮은 수준(약 5% 이하)으로 나타나, KoBERT 모델의 높은 문맥 이해력과 분류 성능을 뒷받침하였다.

처리 속도 측면에서는, NVIDIA T4 GPU 환경(Google Colab)에서 문장당 평균 0.02초 이내로 추론이 완료되었으며, 이는 실시간 또는 준실시간 시스템 구축에 적합한 수준임을 확인하였다. 학습 과정에서는 총 10 Epoch 동안 Training Loss와 Validation Loss가 안정적으로 감소하여 과적합 없이 효과적인 학습이 이루어졌음을 입증하였다.

임상 환경 활용 가능성 측면에서, 본 시스템은 전문가(언어치료사, 임상 심리사 등)가 아동 대화 데이터를 입력하여 초기 언어 발달 이상 징후를 조기에 탐지하고, 자동 분류 결과를 진단 보조 자료로 활용할 수 있도록 설계되었다. 특히 임상 환경에서 초기 선별 도구(Screener)로 활용될 수 있다. 시스템은 아동의 대화 데이터를 입력받아 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현 중 하나로 자동 분류 결과를 제공하고, 이를 통해 전문가가 추가적인 심층 진단 및 최종 평가를 수행하는 데 참고자료로 삼을 수 있도록 지원한다. 이처럼 시스템 결과는 전문가의 전문적 판단을 보조하는 형태로 사용되며, 빠른 초기 선별과 효율적인 개입 결정을 가능하게 하여 실질적인 임상 활용 가치를 높인다. 이를 통해 기존의 주관적 평가 방식 대비 객관성과 신뢰성을 크게 향상시킬 수 있으며, 치료 계획 수립 및 조기 개입을 위한 효율적인 지원 도구로 기능할 수 있다. 다만, 제안 시스템은 아동 발화의 표면적 언어 패턴(반복, 문법 오류, 의미 단절 등)을 중심으로 분석을 수행하기 때문에, 표정, 제스처 등 비언어적 표현이나 아동의 심층적 인지·정서 상태까지는 반영하지 못하는 한계가 있다. 또한, 유사한 언어적 특징을 공유하는 자폐 스펙트럼 장애와 의사소통 장애 간에서는 경계가 모호하여 일부 오분류가 발생할 수 있다. 따라서 본 시스템의 결과는 진단을 위한 참고 자료로 활용하되, 최종 진단은 반드시 전문가의 임상적 판단과 추가 평가를 통해 이루어져야 한다.

VI. Conclusions

본 연구에서는 딥러닝 기반 자연어 처리 모델인 KoBERT를 활용하여 아동의 언어 발화 데이터를 분석하고, 언어 발달 장애를 조기 진단할 수 있는 시스템을 개발

하였다. 특히 자폐 스펙트럼 장애, 말더듬, 의사소통 장애, 정상 표현을 대상으로 총 10,000문장의 데이터셋을 구축하고, KoBERT 기반 모델의 성능을 기존 기계학습 및 딥러닝 모델들과 비교·평가하였다. 실험 결과, KoBERT 기반 제안 시스템이 정확도, 정밀도, 재현율, F1-Score 모든 측면에서 가장 우수한 성능을 보였으며, 특히 F1-Score 0.88로 기존 모델들을 크게 상회하는 성능을 기록하였다. 혼동 행렬 분석에서도 KoBERT 모델은 유사한 언어적 특성을 가진 장애 간에서도 정확하게 분류하는 모습을 보여주었으며, 이는 한국어 문맥과 표현 구조를 효과적으로 학습했기 때문으로 판단된다. 반면, 전통적인 TF-IDF 기반 SVM 및 XGBoost 모델과 BiLSTM 모델은 복합적인 언어 표현에서 오분류가 많이 발생하여 KoBERT 대비 성능이 낮은 것으로 나타났다. 이러한 결과는 아동의 언어 발화에서 나타나는 반복적 표현, 문법적 오류, 문장 불완전성과 같은 특성을 처리하기 위해서는 문맥적 이해가 가능한 모델의 적용이 필수적임을 보여준다. 본 연구의 주요 기여점은 다음과 같다. 첫째, 한국어 기반 언어 발달 장애 진단을 위한 대규모 맞춤형 데이터셋을 구축하고, 모델 성능 평가에 활용했다는 점이다. 둘째, KoBERT 기반 모델의 우수성을 실험적으로 입증하고, 복합적인 언어 패턴 속에서도 높은 분류 정확도를 달성했음을 확인했다. 셋째, 향후 전문가의 임상 진단을 보조할 수 있는 실질적인 AI 시스템으로 확장 가능성을 제시했다. 향후 연구 방향으로서는 더 많은 양의 실제 상담 데이터를 활용하여 추가 학습 및 검증을 진행할 예정이다. 특히 오분류 사례를 기반으로 Fine-tuning을 진행하거나, 대화 문맥을 더 길게 반영할 수 있는 모델 구조 개선을 통해 진단 정확도를 더욱 높이는 방안을 고려할 수 있다. 또한, 실시간 분석 시스템으로 확장하여 아동의 언어 발달 상태를 지속적으로 모니터링하고, 이상 징후 발생 시 전문가에게 즉각적으로 알림을 제공하는 기능을 개발할 계획이다. 이러한 연구를 통해 아동 언어 발달 장애 조기 진단의 효율성과 정확성을 한층 향상시키고, 실제 교육 및 의료 현장에서 활용 가능한 AI 기반 솔루션으로 발전시켜 나갈 수 있을 것으로 기대한다.

REFERENCES

- [1] S.D. Park, and Y.J. Jang, "A Study on Effects of the Interaction of Children with Parents, Teacher and Peers on Their Language and Cognitive Development," JOEC, Vol. 22, No. 2, pp. 229-251, June 2016. DOI: 10.24159/joec.2016.22.2.229
- [2] S.R. Kwak, and J.S. Lee, "Effects of Sociality and Cognitive-Language

- Development in Early Childhood," *KJCEC*, Vol. 15, No. 3, pp. 265-291, September 2015. UCI : G704-SER000008863.2015.15.3.001
- [3] M.H. Lim, and H.J. Kwon, "Environmental Factors in Autism and Autistic Spectrum Disorder," *J Korean Acad Child Adolesc Psychiatry*, Vol. 22, No. 1, pp. 1-7, March 2011. UCI : G704-001951.2011.22.1.001
- [4] Y.M. Choi, and S.Y. Kim, "Attentional Bias toward Angry Faces in Typically Developing Children and Children with Autism Spectrum Disorder," *KOSES*, Vol. 27, No. 3, pp. 121-134, September 2024. DOI: 10.14695/KJSOS.2024.27.3.121
- [5] G.H. Seo, "Attentional Process and Memory Function in Children with Autism Spectrum Disorders," *J.E.B.D*, Vol. 22, No. 4, pp. 251-277, December 2006. UCI : G704-000501.2006.22.4.030
- [6] J.I. Kwon, J.S. Kim, A.M. Woo, H.J. Kim, M.E. Jung, H.S. Kim, and Y.J. Ko, "Diagnosis of Speech Disorder and Language Developmental Delay in Korean Preschoolers," *ARM*, Vol. 30, No. 4, pp. 309-314, December 2006. UCI : G704-000430.2006.30.4.016
- [7] H.E. Chung, S.Y. Kim, and D.S. Yim, "The Relationship between Linguistic Disfluencies and Cognitive Abilities in School-aged Children with and without Vocabulary Delay," *CSD*, Vol. 28, No. 2, pp. 240-254, June 2023.
- [8] S.H. Kim, and G.H. Hong, "Parents' Perceptions and Demands regarding Telepractice for Children with Language Development Disorders," *Jour. of KoCon.a*, Vol. 24, No. 11, pp. 842-853, November 2024. DOI: 10.5392/JKCA.2024.24.11.842
- [9] J.K. Kang, "Spelling Development Characteristics of Children With Speech Sound Disabilities According to Multilingual Evaluation Factors," *JSLHD*, Vol. 33, No. 4, pp. 1-10, December 2024.
- [10] H.M. Baek, and Y.G. Choi, "A Case Study on Enhanced Milieu Teaching for Acquisition of Communication Functions of Nonverbal Children with Autism Spectrum Disorder," *JOURNAL OF SPECIAL EDUCATION & REHABILITATION SCIENCE*, Vol. 61, No. 4, pp. 263-286, December 2022. DOI: 10.23944/Jrsers.2022.12.61.4.12
- [11] J.B. Kim, J.S. Seo, Y.H. Kim, S.Y. Chung, I.G. Lee, K.T. Hwang, and B.C. Lee, "The Efficacy of Speech and Language Therapy for Children with Speech and Language Delays according to the Etiologies," *Clin Exp Pediatr*, Vol. 48, No. 9, pp. 924-928, September 2005. UCI : G704-000560.2005.48.9.004
- [12] S. Lee, H. Jang, Y. Baik, S. Park, and H. Shin, "KR-BERT: A Small-Scale Korean-Specific Language Model," arXiv preprint arXiv:2008.03979, August 2020. DOI: 10.48550/arXiv.2008.03979
- [13] S.H. Kim, and H.S. Kim, "Overseas Literature Review of Nonspecific Language Impairment (NLI) Research," *JSLHD*, Vol. 31, No. 2, pp. 53-62, April 2022. DOI: 10.15724/jslhd.2022.31.2.053
- [14] S. Ellis Weismer, and E. Thordardottir, "Cognition and Language," in **Disorders of Language Development**, P.J. Accardo (Ed.), York Press, pp. 21-37, 2002.
- [15] K.H. Eom, and D.S. Kim, "Automated Classification Model for Online Public Opinions in a Political Arena: KoBERT based Sentiment Analysis," *KPSR*, Vol. 20, No. 3, pp. 167-191, September 2021. DOI: 10.30992/KPSR.2021.09.20.3.167
- [16] J.Y. Kim, D.K. Lee, and S.Y. Cheon, "Literary Emotion Classification Using KoBERT," *Journal of Korean Studies*, No. 87, pp. 5-31, December 2023. DOI: 10.17790/kors.2023.12.87.5
- [17] J.W. Hyeon, J.I. Lee, and H.K. Cho, "Sentiment Analysis of News on Corporation Using KoBERT," **Korean Accounting Review**, Vol. 47, No. 4, pp. 33-54, August 2022. DOI: 10.24056/KAR.2022.08.002
- [18] E.M. Kim, S.J. Nam, T.I. Kim, and T.H. Hong, "The Prediction of Review Helpfulness by Integrating Large Language Models and Deep Learning Based on KoBERT and KoGPT2," *Journal of Intelligence and Information Systems*, Vol. 30, No. 2, pp. 195-209, June 2024. DOI: 10.13088/jiis.2024.30.2.195
- [19] 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 (NAACL-HLT)**, Vol. 1, pp. 4171-4186, Minneapolis, Minnesota, June 2019. DOI: 10.18653/v1/N19-1423

Authors



Won-Cheol Park received the B. S., M. S. and Ph. D. degrees in Computer Engineering from Kongju National University in 2011, 2013, and 2018, respectively. He is currently teaching in the Department of Computer

Engineering, Kongju National University. He is interested in cloud computing, big data, image processing and machine learning.