

SMART-MDR: Stacked Multimodal Architecture with Robust Text-Tabular Data for Medical Department Recommendation

Yang-Hoon Ham*, Seong-Min Lee**, Min-soo Kim**, Chanhee Kwak***

*Student, Dept. of AI Convergence Engineering, Kangnam University, Yongin, Korea

**Student, Dept. of Industry Data-science, Kangnam University, Yongin, Korea

***Associate Professor, Dept. of AI Convergence Engineering, Kangnam University, Yongin, Korea

[Abstract]

This study proposes a multimodal system that integrates subjective symptom text with structured data to recommend the most appropriate medical department. 2.89 million medical records provided by the National Health Insurance Service, was used to determine key diagnosis codes and medical department data. During data preprocessing, irrelevant departments were excluded, and data augmentation was performed using a large language model (LLM) to enhance data quality. The proposed model was designed with a dual-track architecture combining KM-BERT for text data processing and XGBoost for structured data analysis, with a CatBoost-based Stack-Ensemble model for final prediction. Experimental results showed that the proposed model achieved an accuracy of 0.748 and an F1-score of 0.732 compared to existing single models.

▶ **Key words:** Multimodal, Natural Language Processing, Late-Fusion, Medical System, Large Language Model

[요 약]

본 연구는 환자의 주관적 증상 텍스트와 정형 데이터를 융합하여 최적의 진료과를 추천하는 멀티모달 추천 시스템을 제안한다. 국민건강보험공단에서 제공하는 289만여 건의 진료내역 데이터를 기반으로 주요 상병코드 및 진료과 데이터를 수집하고, 이를 학습 데이터로 사용하였다. 데이터 전처리 과정에서 불필요한 진료과를 제거하고, 대규모 언어 모델(LLM)을 활용해 데이터 증강을 수행하여 데이터의 품질을 향상시켰다. 제안된 모델은 KM-BERT와 XGBoost 기반의 이중 트랙 구조로 설계되었으며, 최종 예측에는 CatBoost 기반 Stack-Ensemble 모델을 적용하였다. 실험 결과, 제안된 모델은 기존 단일 모델에 비해 정확도와 F1-Score가 각각 0.748, 0.732를 보였다.

▶ **주제어:** 멀티모달, 자연어처리, 후기 융합, 의료 시스템, 거대언어모델

- First Author: Yang-Hoon Ham, Corresponding Author: Chanhee Kwak
- *Yang-Hoon Ham (eric9898@kangnam.ac.kr), Dept. of AI Convergence Engineering, Kangnam University
- **Seong-Min Lee (sbqlsbql1234@naver.com), Dept. of Industry Data-science, Kangnam University
- **Min-soo Kim (msbg6349@gmail.com), Dept. of Industry Data-science, Kangnam University
- ***Chanhee Kwak (chk@kangnam.ac.kr), Dept. of AI Convergence Engineering, Kangnam University
- Received: 2025. 10. 31, Revised: 2025. 12. 06, Accepted: 2025. 12. 10.

I. Introduction

외래진료의 과다, 특정 의료기관의 과잉 이용, 임상 의사 수 부족 등 우리나라의 의료서비스 접근성과 질 저하의 주요 원인으로 꾸준히 지적되고 있다[1]. 이에 따라 환자들은 경제적 및 시간적 부담을 안게 되며, 중소병원과 의원은 경영난에 직면하고 있다. 특히 의료기관 간 기능 구분이 명확하지 않은 상황에서 의료자원의 비효율적 활용이 문제가 되고 있으며, 대형병원으로의 환자 쏠림 현상이 심화됨에 따라 의원과 병원급 의료기관 간의 직접적인 경쟁이 발생하고 있다[2]. 이러한 현상이 지속될 시, 일차의료기관의 붕괴로 이어져 국민의 의료 접근성과 건강보험재정에 부정적인 영향을 미칠 수 있다.

환자들이 경미한 증상에도 상급병원을 이용하는 사례가 빈번한 이유는 자신의 증상에 대한 정보 부족과 적합한 진료과 선택의 어려움에서 기인한다. 이는 외래진료 과다와 같은 문제를 더욱 악화시키며, 환자의 요구, 잘못된 진료 지침, 의료 시스템의 구조적 문제 등이 복합적으로 작용한 결과로 볼 수 있다[3]. 예컨대, 단순한 감기 증상에도 과도한 주사나 링거액 투여를 요구하는 경향은 의료자원의 불필요한 낭비를 초래한다.

본 연구는 이러한 문제를 해결하기 위해 환자가 느끼는 주관적 증상을 텍스트로 입력받아 상병코드를 예측하고, 그에 적합한 진료과를 추천하는 모델을 제안한다. 이를 위해 트랜스포머 인코더 기반의 NLP 모델인 BERT와 Tree 기반 앙상블 모델인 XGBoost를 결합하여, 비정형 데이터와 정형 데이터를 효과적으로 활용할 수 있는 fusion 모델 구조를 설계하였다. 이를 통해 두 종류의 서로 다른 데이터를 학습한 모델은 환자의 주관적인 증상과 성별, 나이와 같은 추가적인 정보를 통한 개인 맞춤 의료서비스를 제공하여 기존 연구와는 차별화된 접근을 시도하였다[4]. 본 연구가 제안하는 모델은 의료 지식이 부족한 환자들에게 적절한 진료과 선택을 도와 의료서비스 접근성을 향상시키고, 제한된 의료 자원 활용의 최적화를 기대할 수 있다.

본 논문의 나머지 구성은 2장에서 의료 추천 시스템에 관한 발전과 그 한계점을 짚어보면서 제안하는 연구의 필요성에 대해 강조한다. 3장에서는 사용한 데이터셋과 제안하는 모델 구조의 유효성을 검증하며 4장에서 모델 실험 결과를 제시하고 이에 대한 해석을 진행한다.

1. Related works

정형 데이터를 활용한 의료 추천 시스템은 의료 데이터의 효율적 분석과 개인화된 의료 서비스 제공을 목표로 여러 연구가 수행됐다. Markov Decision Process와 Q-Learning을 포함한 강화학습 기법을 사용하여 의료 빅 데이터를 분석한 연구에서는, 환자의 상태 변화에 따른 최적의 의료행위를 제안하는 모델을 개발하였다[6]. 이를 통해 당뇨병 환자와 같은 만성질환 환자들에게 맞춤형 치료를 제공하는 데 유용한 결과를 도출하였다.

한편, 정형 데이터를 텍스트로 변환하고 이를 대규모 언어 모델을 통해 분석하여 데이터 증강을 수행하는 연구는 의료 데이터의 양적, 질적 한계 극복에 기여하였다[7]. 이와 같은 선행연구들은 정형 데이터 분석을 통해 의료 추천 시스템의 정확도와 신뢰성을 강화하고, 환자 맞춤형 진단 및 치료의 가능성 확장에 중요한 역할을 한다.

기존의 추천 시스템은 주로 정형 데이터(structured data)를 중심으로 설계되어, 의료 분야에서 중요한 정보를 포함하는 비정형 데이터(unstructured data)를 충분히 활용하지 못하는 한계를 보인다. 정형 데이터는 환자의 연령, 성별, 혈압, 혈액 검사 수치 등 명확히 구조화된 정보로 구성되며, 이러한 데이터는 수치화와 분석이 용이하다. 의료 현장에서 생성되는 비정형 데이터는 환자의 증상 기술, 병력, 의사의 소견 등 텍스트 형태로 제공되는 경우가 많으며, 환자의 상태를 더 깊이 이해하는 데 중요한 정보를 포함한다. 하지만 기존 시스템이 비정형 데이터를 효과적으로 처리하지 못할 경우, 환자의 주관적 상태나 의사의 진단 맥락이 분석 과정에서 배제될 가능성이 있다[8].

결과적으로, 정형 데이터에만 의존한 시스템은 임상적 맥락을 충분히 반영하지 못해 의료 결정의 정확도가 저하될 수 있다. 그러나 비정형 데이터는 방대한 양의 의료 정보를 포함하고 있음에도 불구하고, 이를 구조적으로 분석하고 정량적으로 활용하기 어려운 한계가 존재한다. 이에 따라, 비정형 데이터를 직접 활용하는 대신, 대규모 언어 모델(LLM)을 활용하여 의미를 추출하고 정형 데이터와 연계하는 방식이 효과적인 해결책이 될 수 있다. LLM은 의료 기록, 의사 소견, 논문 등에서 중요한 정보를 추출하고 구조화하여 정형 데이터 기반의 분석과 결합할 수 있도록 도와주며, 이를 통해 보다 정교하고 맥락을 고려한 의료 의사결정을 지원할 수 있다[9].

II. Preliminaries

III. The Proposed Scheme

본 연구는 한국어 진료과 추천 시스템을 구축하기 위해 텍스트 데이터(환자의 증상 설명)와 정형 데이터(환자 정보 및 진료 이력)를 통합적으로 처리하는 멀티모달 앙상블 모델을 제안한다. 텍스트 분석 단계에서는 Transformer encoder 기반의 모델인 BERT를 활용하였다.

BERT는 자연어 처리에서 문맥의 양방향성을 활용하여 언어 이해 능력을 향상시키는 사전 학습된 모델로, Transformer 구조를 기반으로 한다[10]. 기존 연구에서도 BERT를 의료 분야에 적용하여 전자의무기록(Electronic Health Records, EHR) 분석, 의료 텍스트 분류, 임상 진단 예측 등의 작업에서 높은 성능을 보인 바 있다[11].

본 연구는 BERT 모델 중에서 의료 데이터로 사전학습된 KM-BERT를 활용하여 의료 도메인에 특화된 텍스트 임베딩을 생성하고, 트리 기반 앙상블 모델인 XGBoost를 활용하여 정형 데이터를 처리하는 병렬 트랙 구조를 설계하였다. KM-BERT는 고려대학교 의료 빅데이터 연구소에서 사전 학습된 Transformer encoder 구조 기반의 언어 모델로, 약 1억 1천 6백만 단어로 이루어진 한국어 의료 코퍼스를 바탕으로 학습되었다[12]. 해당 코퍼스는 의료 교과서, 건강 정보 뉴스, 의료 연구 논문 등 다양한 한국어 의료 문서로 구성되어 있어, 의료 문맥에 대한 높은 이해도를 제공한다.

제안된 모델은 텍스트 데이터와 정형 데이터의 융합 기법을 통해 두 데이터 간 상호보완적 정보를 학습하여 추천 성능을 향상시키고자 한다. 특히, KM-BERT는 한국어 증상 설명에서 도출되는 복잡한 의미를 효과적으로 추출할 수 있으며, XGBoost는 환자의 수치적 데이터를 효율적으로 분석할 수 있으므로 이 둘을 동시에 활용하여 각각의 장점을 극대화하고자 하였다. 이러한 접근 방식은 선행연구에서 제안된 텍스트와 테이블 데이터를 융합하여 예측 성능을 개선한 기법을 진료과 추천 시스템 영역에 적용한 것으로, 기존 연구들이 보여준 우수한 성능을 의료 추천 시나리오에서도 재현하고자 한다[1].

1. Building structured and unstructured datasets

1.1 Data Collection

본 연구는 국민건강보험공단에서 제공하는 진료내역정보 데이터를 기반으로 환자의 증상 설명 및 진료내역 정보를 도출하였다. 해당 데이터셋은 2023년 기준 총 2,891,197건의 환자 진료 데이터를 포함하고 있으며,

Table 1. The description of medical history Data

Data count	2,891,197
Variables	User serial number
	Gender
	Age range
	Department
	Main illness code
	Number of nursing days
	Number of hospital days
Total prescription days	

Table 1과 같이 총 8개의 컬럼 중 연구에 필요한 7개의 컬럼만 추출하여 사용하였으며, 이들 컬럼은 성별, 연령대, 진료과, 주상병 코드, 간호 일수, 입원 일수, 총 처방 일수이다.

각 상병코드를 텍스트 데이터로 변환하기 위해 건강보험심사평가원에서 제공하는 상병마스터 데이터를 활용하여 코드와 한글 병명을 매핑하였다. 예를 들어, 'J00' 코드는 '급성 비인두염(감기)'로, 'K29' 코드는 '위염'으로 변환하여 텍스트 데이터로 활용하였다. 또한, 위키피디아 및 네이버 의료 지식백과에서 상병코드별 증상을 나타내는 텍스트 데이터를 추가로 수집하여 진료과와 관련된 다양한 증상 표현을 확보하였다.

1.2 Data Augmentation

추가적으로, 본 연구에서는 LLM(Large Language Model)을 활용하여 데이터를 증강하였다. 사용한 모델은 GPT-4o 모델이며 기존 데이터를 증강하고 부족한 증상 텍스트를 보완하는 역할을 수행했다. 이를 위해 OpenAI의 API를 활용하였으며, 기존에 수집한 진료과별 환자의 증상 텍스트 데이터를 기반으로 추가적인 증상 설명 텍스트를 생성하였다.

예를 들어, LLM을 활용해 환자의 증상 데이터를 재구성하거나 새로운 문맥에서의 표현을 생성함으로써 텍스트 데이터의 질적, 양적 향상을 달성하였다. 이러한 접근 방식은 증상 설명의 다양한 표현 방식과 맥락 차이를 반영하여, 모델의 일반화 성능을 개선하는 데 기여할 수 있다 [13]. 또한, 의료 데이터와 같은 복잡한 구조를 다룰 때 LLM이 텍스트의 의미적 맥락을 풍부하게 확장하고, 기존 데이터에서 부족했던 정보의 틈을 메우는 데 효과적이다 [14]. 이를 바탕으로 본 연구는 GPT-4o 모델을 활용하여 환자의 증상 데이터와 상병코드 간의 관계를 더욱 다양하게 표현하고, 기존 데이터셋의 한계를 보완하는 과정을 수행하였다.

1.3 Data preprocessing

본 연구에서는 효율적인 진료과 추천 시스템의 학습 데이터를 구축하기 위해 진료과목을 선별하는 과정을 수행하였다. 이를 위해 2020년부터 2023년까지 상급종합병원 및 종합병원에서 공통적으로 운영되는 주요 진료과목 목록을 초기 후보군으로 설정하고, 국민건강보험공단 데이터를 활용하여 진료과목별 빈도 분석을 실시하였다.

Table 2의 빈도 분석 결과 진료내역 데이터의 수가 상대적으로 부족하여 신뢰도 확보가 어려운 하위 세 개 진료과(진단검사의학과, 결핵과, 기타)를 학습 데이터에서 제외하였다. 또한, 영상의학과와 같은 다른 진료과로부터 의뢰를 받아 진료를 수행하는 특성이 있어 텍스트 기반 추천 방식에 적합하지 않다고 판단하였다. 또한, 진단검사의학과(Laboratory), 결핵과(Tuberculosis) 등은 연간 샘플 수가 10건 미만으로, 이는 전체 데이터의 0.01%에도 미치지 못한다. 이러한 극소수 데이터에 오버 샘플링을 적용할 경우 과도한 과적합과 현실과 동떨어진 인공적인 패턴(Artifacts)이 생성될 위험이 있어 학습에서 제외하였다.

위와 같은 정보를 기반으로, 최종 학습 데이터로는 국민건강보험공단의 진료내역정보 데이터에서 주요 진료과목과 환자의 증상 텍스트와 상병코드 간의 매핑이 명확하며, 제안하는 모델의 학습 및 평가에 적합한 데이터 구성이 가능한 진료과목을 기준으로 선정하였다. 또한 진료과목 선정 과정에서는 WHO의 국제질병분류(ICD) 체계와 국내 상급종합병원의 진료과 운영 기준을 바탕으로 검토를 진행하였으며, 기존 연구에서 제시된 진료과 추천 모델의 기준과 비교하여 선정의 타당성을 검증하였다. 최종 데이터셋은 Table 3과 같다.

2. Multimodal data processing

모달리티(modality)는 인간이 정보를 받아들이는 방식이나, 어떤 형태로 나타나는 현상을 말한다[15]. 인간은 시각, 청각, 촉각, 후각, 미각 등 다양한 감각을 통해 세상을 경험하며, 이러한 멀티모달 정보를 효과적으로 처리하는 것은 AI의 자연스러운 이해와 판단을 가능하게 한다 [17,18]. 그러나 이러한 각기 다른 특성을 가진 데이터를 통합하는 과정에서 많은 기술적 도전과제가 발생한다.

2.1 Modality Fusion

Modality Fusion은 멀티모달 데이터 처리에서 서로 다른 모달리티로부터 정보를 결합하여 예측 또는 의사결정을 수행하는 과정을 의미한다. Modality Fusion은 데이터의 특성, 결합 방식, 학습 모델의 구조 등에 따라 다양한 방법으로 구현될 수 있다. 본 연구에서 접목한 fusion 기

Table 2. Frequency of medical department visits in the dataset

Department	2020	2021	2022	2023
Internal Medicine	17,757	17,374	21,320	22,320
Otolaryngology	14,077	12,465	19,519	20,331
Ophthalmology	12,505	13,080	13,214	13,721
Orthopedics	10,739	11,406	12,371	12,677
Pediatrics	5,571	5,270	8,924	7,545
Dermatology	7,206	7,208	7,160	7,400
Obs and Gynae	3,930	4,022	4,138	4,064
Urology	3,796	3,765	3,744	3,817
Family Medicine	2,565	2,393	3,365	3,278
Anesthesiology and Pain	2,746	2,913	3,162	3,219
Surgery	2,703	2,792	3,104	3,092
Neurosurgery	1,876	2,013	2,148	2,143
Psychiatry	1,481	1,694	1,913	2,063
Rehabilitation Medicine	1,230	1,308	1,444	1,472
Neurology	736	827	935	968
Radiology	689	725	771	755
Plastic Surgery	128	155	190	216
Thoracic Surgery	102	112	129	126
Others	4	4	6	5
Laboratory	4	4	8	3
Tuberculosis	5	2	2	2

Table 3. Final dataset

Final dataset		
Data count	48,000	
Dependent Variable	Department	
Predictor Variable	Table data	Gender
		Age range
		Main illness code
		Number of nursing days
		Number of hospital days
		Total prescription days
	Text data	Patient symptom text

술은 데이터의 특성과 학습 모델의 구조 측면이다.

먼저, 데이터셋 구축 시 fusion 적용 방식은 다음과 같다. 멀티모달은 이질적이며, 상호 보완적인 특징을 가지고 있다[15]. 이를 결합하면 데이터의 불완전성을 보완하고 단일 모달리티보다 더 높은 예측 성능과 신뢰성을 얻을 수 있다. 예를 들어, 오디오-비주얼 음성 인식에서는 음성과 영상 데이터를 융합해 노이즈 환경에서도 정확한 인식이 가능해진다. 또한, 특정 모달리티의 데이터가 누락되더라도 다른 모달리티로 이를 보완할 수 있어 강건한 모델 구축이 가능하다. 더불어, 모달리티 간 복합적 관계를 학습함으로써 복잡한 패턴 이해와 고차원 예측이 가능해 다양한 응용 분야에 활용된다. 이러한 이유로 Fusion은 멀티모달 머신러닝에서 성능 향상과 정밀도 제고를 위한 필수적인 기술로 간주된다.

따라서, 본 연구에서 구축한 진료과 추천 데이터셋은 환자가 말하는 증상 설명 텍스트와 더불어 해당 증상을 뜻하는 증상 코드, 증상에 대한 입내원 일수, 약 처방 일수 등 테이블 데이터를 진료과에 매핑하였다. 이를 통해 진료과에 대한 다양한 정보를 결합하여 데이터의 정합성을 높일 수 있게 된다.

2.2 Late Fusion

모달리티 결합 방법은 크게 Early Fusion(초기 융합), Late Fusion(후기 융합)으로 구분된다[15]. Early Fusion은 여러 모달리티에서 추출된 feature를 초기 단계에서 결합하여 하나의 입력 데이터로 처리한다. 이는 데이터의 전처리 단계에서 모든 모달리티 데이터를 하나의 입력 데이터로 통합하지만, 이질적인 특성을 가진 멀티모달 데이터를 조화롭게 통합하기 어렵다는 한계가 있다.

Late Fusion은 이러한 문제를 효과적으로 해결할 수 있는 접근 방식이다. Late Fusion은 각각의 모달리티로부터 독립적인 예측 결과를 생성한 후, 이를 결합하여 최종 결론을 도출하는 방식이다. 이 접근법은 각 모달리티가 독립적으로 처리될 수 있는 유연성과 확장성을 제공한다[16]. 또한, 단일 모달리티의 데이터가 누락되더라도 나머지 모달리티로부터 결과를 도출할 수 있어 강건성이 뛰어난 장점이 있다.

본 연구의 데이터는 텍스트 데이터와 테이블 데이터로 구성되어 있어 데이터의 형식적 차이와 정규화 방법의 불균형이 발생할 수 있다. 이러한 상황에서 Early Fusion을 적용할 경우 전처리 과정의 복잡성과 정보 손실의 위험이 증가할 수 있다. 따라서, 본 연구에서는 Late Fusion 방식을 채택하여 텍스트와 테이블 데이터 각각에 대해 독립적인 모델을 학습한 후, 최종적으로 예측 결과를 결합하여 최적의 진료과를 추천하고자 한다.

2.3 Stack-Ensemble

본 연구에는 Late Fusion 구조 기반 멀티모달 네트워크 외에도 Stack-Ensemble 방식을 추가하였다. Stack-Ensemble은 Late Fusion 멀티모달 네트워크와 다양한 테이블 기반 모델(XGBoost 등)을 Base 모델로 활용하며, 이들로부터 5-Fold 교차 검증을 통해 생성된 Out-of-Fold Predictions을 Stacker(meta model)의 입력으로 사용하는 방식이다[17]. 이 방식은 Stacker 입력 이전의 Base 모델들이 포착한 다양한 패턴을 비선형적으로 결합하여 최종 예측 성능을 극대화할 수 있는 전략이며 Stack-Ensemble은 Late Fusion이 포착하지 못한 비선형적 상호작용과 복잡한 패턴을 추가로 학습하며 이를 통해 두 방식은 서로 다른 수준에서 데이터를 결합하여 상호

보완적인 관계를 형성할 수 있게 된다. 결과적으로, 텍스트와 테이블 데이터의 상호작용을 효과적으로 학습함으로써 진료과 추천 시스템의 정확도를 한층 높일 수 있다.

3. Model design

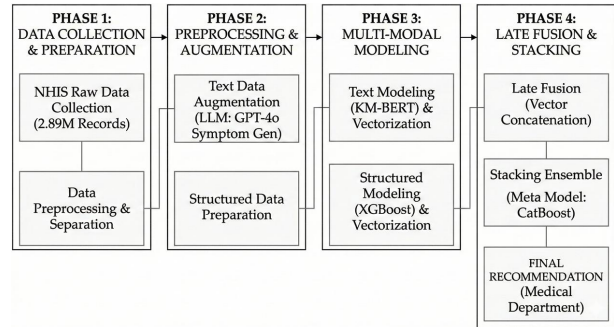


Fig. 1. The overall workflow for department prediction

본 연구에서 제안하는 멀티모달 진료과 추천 프레임워크의 전체 절차는 Figure 1과 같다. 이는 데이터의 특성에 따라 정형 정보와 비정형 텍스트를 분리하여 처리하는 이원화된 구조를 갖는다. 특히 데이터 준비 단계에서는 GPT-4o 기반의 텍스트 증강을 도입하여 단순 질병 코드를 풍부한 임상 증상 텍스트로 변환함으로써 정보의 희소성을 보완하였다. 각 데이터는 KM-BERT(텍스트)와 XGBoost(정형)를 통해 독립적으로 학습되며, 최종적으로 특징 벡터를 결합하는 Late Fusion 방식과 CatBoost 메타 모델을 적용한 스택킹 앙상블을 통해 예측 성능을 최적화하였다.

3.1 Model architecture overview

Figure 2는 본 연구에서 제안하는 텍스트 데이터와 테이블 데이터를 각각 독립적으로 처리하는 이중 트랙 구조 모델이다. 텍스트 데이터 처리 모델은 KM-BERT 모델을 활용하여 사용자가 입력한 증상 텍스트를 분석하고, XGBoost는 환자의 나이, 성별, 진료 기록 등 테이블 데이터를 처리한다. XGBoost는 테이블 데이터의 복잡한 패턴과 변수 간 상호작용을 학습하며, KM-BERT가 포착하지 못할 수 있는 정보를 보완하는 역할을 한다.

두 모델에서 생성된 결과는 벡터 형태로 변환되어 단순 벡터 결합 방식(concatenation)으로 통합된다. 결합된

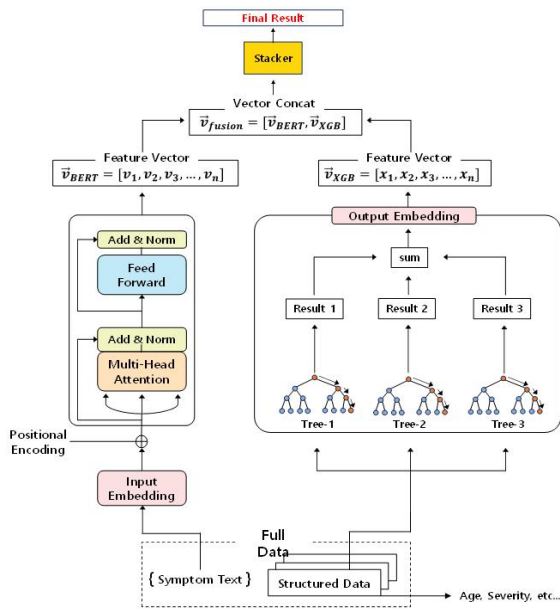


Fig. 2. Model architecture

벡터는 Stack 모델의 입력으로 사용된다. 여기서 Stack 모델은 개별 모델이 학습한 특성을 종합하여 최종적인 예측을 수행하는 역할을 한다. 앞서 2.3절에서 소개한 바와 같이, Stack-Ensemble 방식은 Base 모델(XGBoost, KM-BERT)의 예측 결과를 통합하여 비선형적인 관계를 학습하고 최적의 결합 방식을 찾아낸다. 이를 통해 모델의 성능을 극대화할 수 있으며, 단일 모델보다 더욱 정교한 진료과 추천이 가능해진다.

3.2 Experimental setting

KM-BERT 모델의 학습 시 주요 하이퍼파라미터는 다음과 같이 최적화하였다. 코사인 어닐링 스케줄러를 사용하여 초기 학습률을 $2e-5$ 로 설정하고, 훈련 과정에서 점진적으로 감소시켜 안정적으로 수렴하도록 하였다. 배치 크기는 16으로 설정하였으며, epoch는 15로 설정하였다. 추가로, 조기 종료를 3으로 적용하여 손실이 3번 이상 감소하지 않는다면 학습이 종료되게 설정하였다.

XGBoost 모델의 하이퍼파라미터는 Optuna를 통해 다양한 하이퍼파라미터 조합을 탐색하여 최적의 파라미터 조합을 선정하였다. 최적화 과정에서는 Weighted F1-score 스코어를 사용하여 모델의 성능을 평가하고 개선했다.

IV. Performance Evaluation

1. Performance evaluation and cross-validation

본 연구에서는 Accuracy, Precision, Recall, F1-score와 같은 분류 모델의 성능 평가 지표를 사용하여 모델을 평가하였다. 특히, 다중 클래스 분류 문제의 특성을 고려하여 각 클래스에 대한 분류 성능을 상세히 분석하였다. 데이터셋은 Train-test split 비율을 7:3으로 설정하여 학습과 테스트에 사용하였고 모델의 일반화 성능을 평가하고 과적합을 방지하기 위해 5-fold 교차 검증을 수행하였다. 이를 통해 모델이 다양한 데이터 샘플에 대해 안정적인 성능을 유지하도록 하였다.

2. Performance comparison

본 연구에서는 텍스트와 테이블 데이터를 처리하여 진료 과목 코드를 예측하는 멀티모달 분류 모델을 개발하고, 개발한 모델과 단일 모델의 성능을 비교 평가하였다. KM-BERT와 XGBoost를 각각 텍스트 및 정형 데이터 처리에 활용하였으며, 두 모델의 출력값을 최종적으로 스택 모델에 결합하여 예측을 수행하였다. 스택 모델로는 CatBoost, LightGBM, Random Forest를 비교하였다. 각각의 모델을 Stack 모델로 사용하여 최종 예측 성능을 비교한 결과, CatBoost 0.748, LightGBM 0.670, Random Forest 0.637 순으로 CatBoost가 가장 높은 성능을 보여 최종 스택 모델로 선정되었다. 다만, 다중 클래스 분류 문제의 특성을 고려할 때, 특정 진료과에 따라 성능 차이가 발생하였다. 예를 들어, 비교적 데이터가 균형적으로 분포된 진료과에서는 높은 성능을 보이는 반면, 일부 클래스(정신건강의학과, 신경외과)의 경우 데이터 샘플이 상대적으로 적거나 클래스 간 증상 유사성이 높아 혼동이 발생할 수 있다. 따라서, 향후 연구에서는 클래스 별 성능 차이를 줄이기 위해 데이터 증강, 클래스 가중치 조정, 추가적인 모델 최적화 등을 고려할 필요가 있다.

Table 4는 단일 모델과 Stack-Ensemble 모델과의 성능 비교를 기록한 표이다. 성능 비교 결과를 종합적으로 보면, KM-BERT와 XGBoost 개별 모델의 예측 성능은 단독으로 사용했을 때 한계가 있었다. 반면, CatBoost 기반의 스택 모델은 모든 주요 성능 지표에서 가장 높은 결과를 기록하였다. 특히, 본 연구에서 수행한 오분류 분석(Error Analysis) 결과, 제안된 앙상블 모델은 단일 모델들이 놓친 모호한 케이스들을 효과적으로 교정하는 양상을 보였다. 예를 들어, 텍스트 모델은 환자의 주호소에 포함된 특정 키워드에 과몰입하여 내과로 오분류하는 경향이 있었으나, 스택 모델은 정형 데이터의 연령 및 진료 이

Table 4. Model performance comparison

	XGBoost	KM-BERT	Stack Ensemble Model
Precision	0.31	0.43	0.74
Recall	0.23	0.36	0.71
F1-score	0.28	0.39	0.732
Accuracy	0.34	0.53	0.748

력 정보를 결합하여 이를 소아청소년과로 올바르게 보정하였다. 이는 기존 연구들에서 보고된 바와 같이, 이질적인 모달리티(Tabular·Text)가 상호보완적인 임상 정보를 제공하여 단일 모델 대비 성능이 향상된다는 일반적인 경향성과도 일치한다[18]. 즉, 텍스트의 의미론적 정보와 정형 데이터의 인구통계학적 특성이 결합됨으로써 진료과 분류의 결정 경계가 더욱 명확해진 결과로 해석된다.

결과적으로, CatBoost 기반의 스택 모델은 범주형 변수가 다수 포함된 데이터와 멀티모달 데이터 구조를 처리하는 데 있어 우수한 성능을 보였으며, 진료과목코드와 같은 복잡한 타겟 변수를 예측하는 데 가장 적합한 접근 방식임을 확인하였다. 본 연구의 결과는 멀티모달 데이터 처리 및 의료 데이터 분석 분야에서 스택 모델의 가능성을 제시하며, 향후 연구와 실제 적용에서 중요한 참고 자료로 활용될 수 있다.

V. Conclusions

본 연구는 의료 서비스의 접근성 향상과 제한된 의료 자원의 최적화를 목표로 환자의 주관적 증상 텍스트와 정형 데이터를 결합한 진료과 추천 모델을 제안하였다. 이를 위해 KM-BERT와 XGBoost 기반 Late Fusion 멀티모달 모델을 설계하고, CatBoost와 Stack-Ensemble 전략을 도입해 예측 성능을 극대화하였다.

데이터셋 구축 과정에서 진료과별 상병코드와 증상 데이터를 수집하고, LLM(대규모 언어 모델)을 활용한 데이터 증강을 적용하여 실제 환자의 민감한 임상 텍스트를 직접 수집하지 않고도 다양한 텍스트 데이터를 확보하였다. 다만, 의료 데이터 특성상 실제 의료기관의 전자 의무 기록(EMR) 데이터를 추가로 수집하고 모델을 검증할 필요가 있다. 정형 데이터 및 의사의 진찰 기록을 포함한 EMR 데이터를 활용한다면 모델 입력의 품질을 향상시키고 임상 적용 가능성을 강화할 수 있을 것이다.

또한, 본 연구에서 제안하는 Stack-ensemble 모델의

실험 결과, 정확도와 F1-Score가 지속적으로 향상되었으며, 개별 모델을 단독으로 사용할 때보다 더 높은 예측 성능을 보였다. 이는 본 연구의 융합 전략이 데이터의 구조적 복잡성과 다양한 타겟 변수를 효과적으로 처리했음을 시사한다. 그러나, 일부 클래스의 경우 데이터 샘플이 상대적으로 적어 클래스 간 증상 유사성이 높아 혼동이 발생할 가능성이 있다.

따라서 향후 연구에서는 클래스 별 성능 차이를 줄이기 위한 기법을 고려할 필요가 있다. 또한, 기존에 수집하지 못하였던 환자의 정형 데이터나 의사의 진찰 데이터 등 EMR 데이터를 추가로 확보한다면 본 연구에서 제안한 모델의 입력 품질을 크게 향상시키며 활용 가능성 또한 강화할 수 있을 것이다.

REFERENCES

- [1] H. J. Kim, Y. H. Kim, H. S. Kim, J. S. Woo, and S. J. Oh, "The Impact of Increased Outpatient Cost-Sharing on Utilization of Outpatient Services at General and Tertiary Hospitals," *The Journal of Health Administration*, Vol. 23, No. 1, pp. 19-34, 2013. DOI: 10.1186/s12913-017-2076-8
- [2] K. H. Kim, "A Study on Improving Outpatient Concentration in Tertiary Hospitals," *Healthcare Policy Forum*, Vol. 8, No. 4, pp. 106-113, 2010.
- [3] Y. S. Cheong, and S. K. Park, "What Factors Promote Overtreatment in Korea?: Causative Considerations and Solutions to Overtreatment," *The Journal of the Korean Society for Medical Ethics*, Vol. 19, No. 3, pp. 375-389, 2016.
- [4] S. H. Lee, and J. Y. Kang, "Classification Modeling for Predicting Medical Subjects using Patients' Subjective Symptom Text," *The Journal of Korean Big Data Society*, Vol. 6, No. 1, pp. 51-62, 2021. DOI: 10.36498/kbigdt.2021.6.1.51
- [5] S. M. Kim, "A study on changes in outpatient waiting time before and after EMR introduction using queueing analysis," M.S. thesis, Kosin Univ., 2017.
- [6] M. M. Afsar, "Personalized Recommendation Using Reinforcement Learning," Ph.D. Thesis, University of Calgary, Alberta, Canada, 2022.
- [7] R. Sampson, "Data augmentation for text generation from structured data," MS. thesis, Simon Fraser Univ., 2023.
- [8] Y. Choi, and D. Lim, "Integrating Clinical Records and Medical Imaging for Disease Prevention and Treatment," *The Journal of Korea Society of Computer and Information*, Vol. 25, No. 2, pp. 45-58, March 2015.
- [9] S. J. Choi, "Machine learning methods for multi-modal data

analysis," Research Report TRKO201700011357, Pohang Univ. of Science and Technology, 2017.

- [10] 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, pp. 4171-4186, June 2019. DOI: 10.18653/v1/N19-1423
- [11] J. H. Lee, W. J. Yoon, S. D. Kim, D. H. Kim, S. K. Kim, C. H. So, and J. W. Kang, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, Vol. 36, No. 4, pp. 1234-1240, Sep. 2019. DOI: 10.48550/arXiv.1901.08746
- [12] Y. J. Kim, J. H. Kim, J. M. Lee, M. J. Jang, Y. J. Yum, S. T. Kim, U. S. Shin, Y. M. Kim, H. J. Joo, and S. H. Song, "A Pre-trained BERT for Korean Medical Natural Language Processing," *Scientific Reports*, Vol. 12, No. 13847, 2022. DOI: 10.1038/s41598-022-17806-8
- [13] Q. Liu, X. Wu, X. Zhao, Y. Zhu, Z. Zhang, F. Tian, and Y. Zheng, "Large language model distilling medication recommendation model," *arXiv preprint*, arXiv:2402.02803, 2024. DOI: 10.48550/arXiv.2402.02803
- [14] Y. T. Lee, "Enhancing Medication Recommendation with LLM Text Representation," *arXiv preprint*, arXiv:2407.10453, 2024. DOI: 10.48550/arXiv.2407.10453
- [15] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 2, pp. 423-443, 2019. DOI: 10.1109/TPAMI.2018.2798607
- [16] K. Gadzicki, R. Khamsehshari, and C. Zetsche, "Early vs Late Fusion in Multimodal Convolutional Neural Networks," Proceedings of the 2020 IEEE 23rd International Conference on Information Fusion (FUSION), pp. 1-6, Rustenburg, South Africa, 2020. DOI: 10.23919/FUSION45008.2020.9190246
- [17] X. Shi, J. Mueller, N. Erickson, M. Li, and A. J. Smola, "Benchmarking multimodal automl for tabular data with text fields," *arXiv preprint*, arXiv:2111.02705, 2021. DOI: 10.48550/arXiv.2111.02705
- [18] A. Kline, H. Wang, Y. Li, S. Dennis, M. Hutch, Z. Xu, F. Wang, F. Cheng, and Y. Luo, "Multimodal machine learning in precision health: A scoping review," *npj Digital Medicine*, Vol. 5, Nov. 2022. DOI: 10.1038/s41746-022-00712-8

Authors



Yang-Hoon Ham is currently a student for the B.S. degree in Data Science at Kangnam University, Korea. He is interested in MLLM, LLM, VLM.



Seong-Min Lee is currently a student for the B.S. degree in Data Science at Kangnam University, Korea. He is interested in machine learning, artificial intelligence, and data analysis.



Min-soo Kim is currently a student for the B.S. degree in Data Science at Kangnam University, Korea. He is interested in natural language processing and generative AI.



Chanhee Kwak received the B.S. in computer science, M.S., and Ph.D. degrees in management engineering in KAIST, Korea. He joined the faculty of the department of AI convergence engineering at Kangnam University, Yong-in in 2020. He is currently an associate professor in the department of AI convergence engineering at Kangnam university, Yongin, Korea. His research interests include data analytics systems and machine learning techniques.