

GPT-4를 활용한 의료 문답 모델 학습용 데이터 생성 및 검증 프레임워크

유진식^a, 김상윤^b, 류제원^c, 김미숙^{d*}

^a세종대학교 소프트웨어융합대학 컴퓨터공학과
E-mail: sssjinsig@gmail.com

^b한양대학교 공과대학 융합전자공학부
E-mail: ndkim11@naver.com

^c인제대학교 의과대학
E-mail: ryoojewon@gmail.com

^{d*}세종대학교 소프트웨어융합대학 데이터사이언스학과
서울특별시 광진구 능동로 209 세종대학교
E-mail: misuk.kim@sejong.ac.kr

요약

본 연구는 고성능 거대 언어 모델인 Generative Pre-trained Transformer(GPT-4)를 활용하여 의료 데이터를 생성하는 새로운 접근법을 제시한다. GPT-4의 강력한 생성 능력을 이용하여 높은 수준의 의료 문답 데이터를 생성하고, 이를 바탕으로 BERT기반 의료 언어 모델을 미세 조정하였다. 이 방법으로 훈련된 모델은 F1 점수 0.70을 달성하였음을 보여주며, 이는 실제 의료 문진 데이터 없이도 의료 문답 모델을 학습시킬 수 있는 가능성을 보여준다. 이 연구는 GPT-4를 이용한 다양한 도메인의 데이터 생성의 가능성을 잘 보여주며, 특히 본 논문의 의료데이터 생성방법을 이용한 Generative QA model로까지의 확장 가능할 것이다.

주요용어:

GPT-4, Medical Generated Data, BERT, Extractive QA

1. 서론

최근 AI분야의 눈에 띄는 발전으로 다양한 분야에서 데이터 확보가 더욱 중요하다. 컴퓨터 비전, 자연어 처리, 음성인식 등의 AI 분야에서 데이터를 확보하기 위해 데이터 증강 방식이 활용되고 있다. 특히 최근 다양한 Large Language Model(LLM)의 등장으로 자연어

처리 분야에서의 데이터 수요가 급격하게 늘어나고 있다. 자연어처리 분야에서의 데이터 증강 방식은 텍스트 데이터증강을 통하여 많은 자연어 처리(NLP) 작업에서 제한된 샘플 크기의 문제를 극복하기 위한 효과적인 전략이 되고 있다. 특히 대상 도메인의 데이터가 부족하고 품질이 떨어지는 few-shot 학습이 필요한 문제에서 데이터 증강 기법이 효과적으로 활용될 수 있다. 다시 말해, 데이터의 품질이 떨어지는 문제를 완화하기 위해서 자연스럽게 널리 사용되는 전략은 데이터 불변성을 더 잘 포착하고 샘플 크기를 늘리는 데이터 확대를 수행하는 것이다. 그러나 현재 텍스트 데이터 확대 방법은 생성된 데이터의 정확한 라벨을 보장할 수 없거나 다양성을 충분히 보장할 수 없다[1]. 또한, 데이터 증강의 방식은 현재 보유하고 있는 도메인의 데이터에 대해서만 증강을 통하여 추가적인 데이터를 생성하여 사용할 수 있다. 이는 도메인 한계와 기존 데이터에 의존해야 하는 한계점이 명확히 존재한다.

따라서, 본 논문에서는 기존의 데이터 기반의 증강 방식의 한계점들을 보완하는 기존 구축된 데이터와 관계없는 새로운 데이터 확보 방식을 제안하고자 한다. Generative Pre-trained Transformer(GPT-4)와 같은 LLM을 활용하여 새로운 데이터를 생성하고, 제한적인 도메인인 소화기 내과의 임상적 문제 상황에 대한 원인 감별 task에 활용할 수 있는 미세 조정(Fine-tuning)을 위한 데이터셋을 구축하였다.

* 교신저자

2. 관련 연구

2.1. BioBERT

Bidirectional Encoder Representations from Transformers for Biomedical Text Mining(BioBERT)[2]는 BERT 기반의 생의학 분야 텍스트 마이닝 모델이다. 기존의 BERT의 경우 특수한 분야의 단어에 대한 학습보다 일반적인 상황에 대처를 잘 하기 위해 중점이 보편적인 어휘 사용에 있다[3]. 따라서 BERT는 생의학학 분야의 특수한 용어가 포함된 생의학 문장을 이해하지 못한다. 따라서 BERT를 생의학 분야에 특화하여 문장을 학습시키면 복잡한 생의학 문장을 이해하게 할 수 있다. BioBERT는 이를 수행하기 위해 2가지의 단계를 거쳐서 데이터 학습이 이루어진다. 생의학적 Pre-training 단계에서는 4.5B의 PubMed Abstract 단어와 13.5B의 PMC full-text 단어로 구성된 학습셋을 이용하였다. 또한 생의학적 Named Entity Recognition(NER), Relation Extraction(RE), Question Answering(QA) 3가지의 텍스트 마이닝을 수행할 수 있도록 미세조정 되었다. BioBERT의 학습 데이터가 학술 문장이며 학습을 위해 모델을 통해 생성된 문장은 없다는 특징이 있다.

2.2. ChatDoctor

ChatGPT와 같은 LLM은 다양한 분야에서 인간 답변과 유사한 답변을 하는 등 놀라운 성능을 보이고 있지만, 여전히 의료 분야에서는 낮은 정확도와 의학적으로 적절한 조언 혹은 진단을 내릴 능력은 부족하다는 문제점이 있다. 이 문제를 해결하기 위해 ChatDoctor[4]는 의사와 환자의 문진을 모방하는 모델을 제안하였다. ChatDoctor는 의사와 환자의 문진을 모방하기 위해 환자와 의사의 문답 데이터를 기반으로 QA 모델을 학습시켰다. QA 모델을 학습시키기 위해 데이터셋은 HealthCare Magic이라는 온라인 의학 컨설팅 사이트에서 10만개, icliniq에서 만개의 실제 의사와 환자 대화내용을 데이터셋으로 사용하였다. 추가적으로 ChatDoctor는 각 질병에 대해 증상, 원인, 검사 방법, 일반적인 처방 약물 등의 정보를 하나의 Disease DB 지식 참고파일로 만들었다. 이는 보편적인 LLM을 학습시킬 때의 Wikipedia와 같은 역할을 의료 분야에 대해 수행한다. 학습 단계에서 ChatDoctor는 Disease DB의 지식을 기반으로 의사-환자 문답 데이터를 사용하여 의료 QA 특화 모델을 학습시켰다. 실제 문답 데이터뿐만 아니라 더 많은 학습 데이터를 얻기 위해 Large Language Model Meta AI(LLaMA)를 통해 문답 데이터를 생성하여 학습셋을 구성하였다.

3. 제안 방법

3.1. Extractive QA 모델 학습 Overview

GPT-4를 통하여 의료 데이터 생성을 하고 데이터의 성능평가를 하기 위하여 BERT 모델을 이용한 Extractive QA[5] 모델을 생성하였다. 데이터의 생성 후 모델 미세 조정을 수행하고, 성능 평가를 진행하였다. Extractive QA모델은 Stanford question answering dataset 인 SQUAD dataset[6]형식을 이용하였다. 데이터의 도메인은 의료분야 중 소화기내과 분야로 한정하였다. QA 모델 학습을 위한 Overview는 그림 1과 같다.

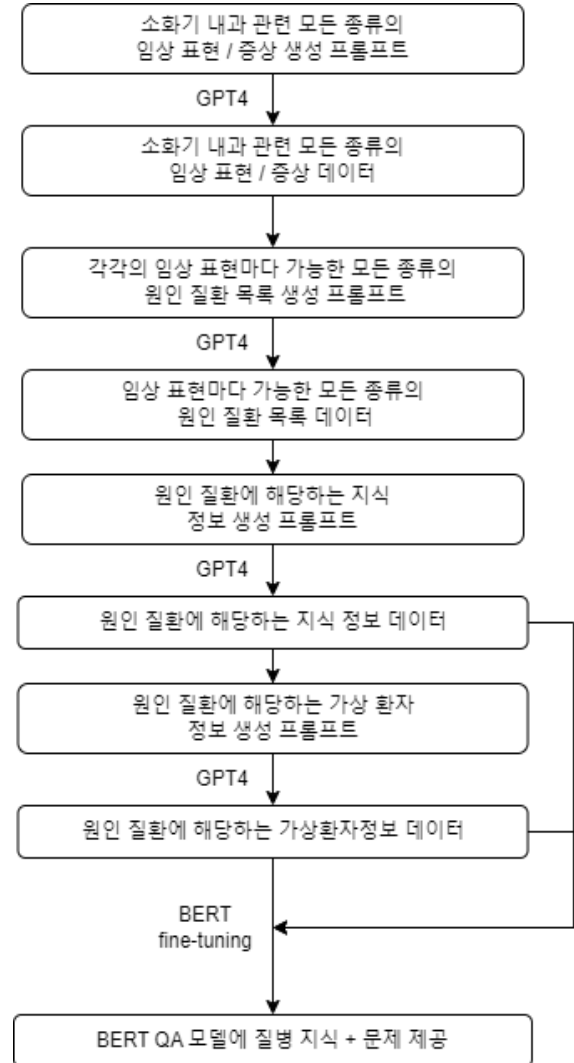


그림 1. QA 모델 학습 Overview

원인질환 데이터 생성: 소화기 내과 관련 모든 종류의 임상 표현/증상 데이터를 소화기내과 관련 모든 임상 표현/증상 데이터 생성 프롬프트를 GPT-4에 입력하여 생성하였다. 생성된 소화기내과 관련 모든 임상 표현/증상 데이터를 각각의 임상표현마다 가능한 모든 종류의 원인 질환 목록 생성 프롬프트에 첨부하여 각각의 임상 표현마다 가능한 모든 종류의 원인 질환 목록 데이터를 생성한다. 소화기 내과 관련 모든 종류의 임상 표현/증상 데이터를 통해 생성된 원인질환 목록 데이터의 개수는 총 45개 종류의 원인질환이 생성되었다.

지식 정보 데이터 생성: 생성된 원인질환 목록 데이

터를 원인 질환에 해당하는 필요한 지식 정보 생성 프롬프트에 첨부하여 원인 질환에 해당하는 필요한 지식 정보 데이터를 생성한다. 45개의 원인질환 데이터를 바탕으로 생성된 지식 정보 데이터는 원인질환 1개 당 1개의 지식 정보를 생성하였으므로 총 45개의 데이터를 생성하였다. 이에 관련된 자세한 내용은 3.2절에서 다루도록 한다.

가상환자정보 데이터 생성: 생성된 지식 정보 데이터를 바탕으로 원인 질환에 해당하는 가상환자정보 생성 프롬프트에 첨부하여 원인질환에 해당하는 가상환자정보 데이터를 생성한다. 45개의 지식정보를 바탕으로 생성된 가상환자정보 데이터는 지식정보 1개 당 18개의 가상환자정보를 생성하였으므로 총 810개의 데이터가 생성되었다. 이에 관련된 자세한 내용은 3.3절에서 다루도록 한다.

생성된 지식정보 데이터와 가상환자정보 데이터를 SQUAD dataset 형식으로 전처리(Preprocessing)하여 BERT모델에 학습시킨다. 최종적으로 GPT-4로 생성한 의료데이터로 BERT 모델의 미세조정을 수행하고 135개의 검증 데이터셋으로 recall, precision, F1-score 평가 지표로 성능평가를 진행하였다.

3.2. Generating Disease Knowledge

각 질병에 대해서 지식 정보 데이터 생성을 위해 GPT에게 요구사항을 상세하게 작성하였고 그림 2와 같다.

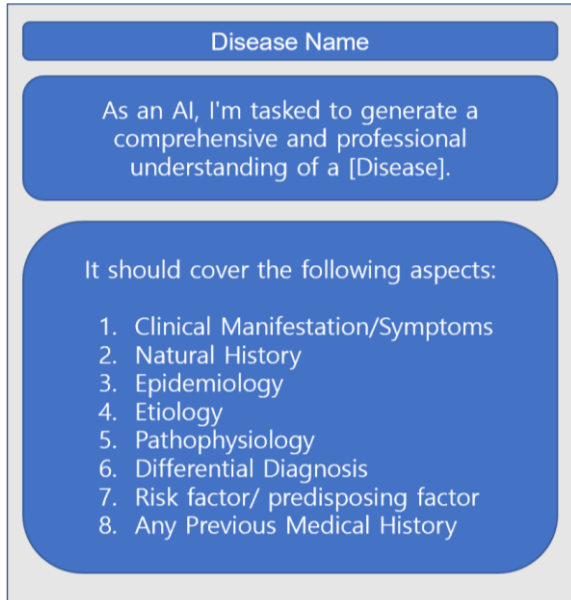


그림 2. 질병 별 지식 정보 생성 프롬프트

프롬프트는 지식을 생성하고자 하는 질병의 명칭을 기입하였다. 그 다음으로 서술 방향 및 어조와 문장구성 방식을 전문적인 의료 지식 생성이라는 목적에 부합하도록 프롬프트에 요구하였다. 마지막으로 세부적인 지식 내용을 지정하여 주었다. 질병 생성 프롬프트에서 질병 지식 생성에 포함시킨 세부 항목들에 대한 설명은

표 1과 같다.

표 1. 질병 지식 생성에 포함된 항목 및 설명

Aspect	Description
Clinical Manifestation /Symptoms	하나의 질환에 대해서 나타내는 증상에 대한 설명
Natural History	시간의 경과에 따른 질병 진행 서술 예) 감상선암, 폐암, 췌장암 각각의 경과가 다름
Epidemiology	나이, 성별, 인종에 대한 설명 예) 협심증의 여러 증상이 겹치는 20대 여자의 경우 협심증으로 확진을 할 수 있는가? 나이 정보가 없다면 협심증의 가능성이 매우 높지만 나이 정보가 20대라는 정보로 그 가능성이 급격하게 감소함
Etiology	의심되는 원인에 대한 질문 예) 장염에 걸리면 의사는 질병의 원인을 알고 있기 때문에 최근에 뭐 잘못 먹은 것은 없는지 질문을 하여 원인을 파악함
Pathophysiology	역학, 원인과 유사함. 과학적, 생리학적으로 이 질병이 어떤 원인으로 유발이 되는지 서술한 정보
Differential Diagnosis	같은 증상에 대해서 고려해야 할 질병 목록을 나타냄
Risk Factor / Predisposing Factor	다른 데이터를 통한 선행요인에 대한 정보 예)비만, 당뇨는 주로 저소득층에서 많이 생김. 이 팩터가 낮은 사람들에게서 더 빈번하게 질병이 생기는 통계가 있음
Previous Medical History	환자의 과거력 예) 고지혈증 과거력이 있다면 현재에도 고지혈증의 가능성이 높음

3.3. Generating Clinical Scenario Case

가상 환자 정보 데이터를 생성하기 위해 GPT에 요구사항을 상세하게 작성하였고 그림 3과 같다.

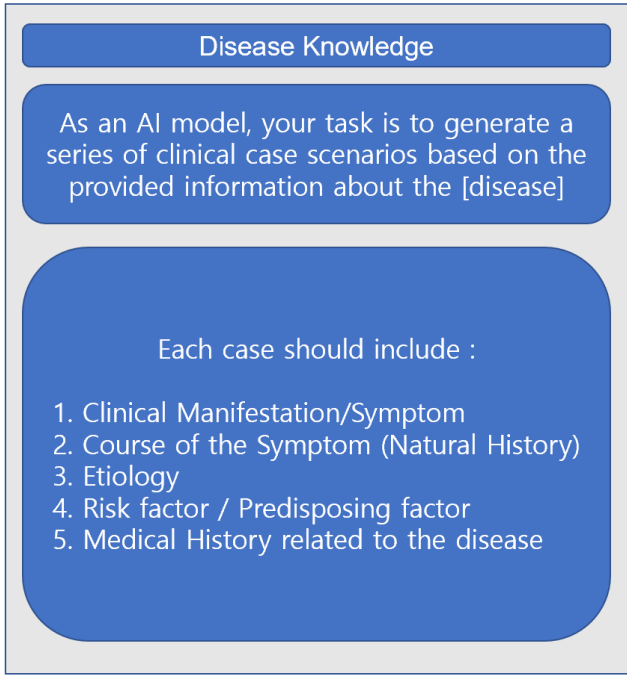


그림 3. 가상환자 생성 프롬프트

프롬프트에는 생성하고자 하는 질병에 대해 앞서 생성한 질병 지식 데이터를 입력한다. 그 다음으로 서술 방향 및 어조와 문장구성 방식을 임상 증상 시나리오 즉 가상환자 데이터 생성을 위한 목적에 부합하도록 프롬프트를 통해 요구하였다. 마지막으로 질병 지식 생성 프롬프트에 포함시킨 세부 항목들 중 일부분을 선택하여 수록하였다.

3.4 BERT: Medical fine-tuning

3.1절에 소개한 구축한 데이터들을 바탕으로 SQUAD 형식에 맞추어 810개의 데이터를 가진 데이터셋을 학습 및 검증에 활용하였다. 이를 5:1의 비율로 학습 데이터셋과 검증 데이터셋으로 분리하여 학습과 성능평가를 진행하였다. 학습 데이터 가공 과정은 표 2와 같다.

표 2. 문답 학습 데이터 가공 과정

Step 1	각 질병의 지식과 환자의 증상을 담은 질문을 하나의 Context 파일로 조합해주는 과정
Step 2	Step 1에서 지식 + 질문으로 조합된 텍스트안에 포함된 모든 개행 문자를 제거 후 한 줄의 Context 데이터로 변환
Step 3	Context에서 정답과 index를 찾아서 저장
Step 4	Context, 질문, 정답, Index를 SQUAD 형식의 csv 파일로 조합

기존의 사전 학습된 BERT를 이용하여 미세조정을 수행하였다.

3. 실험결과

3.1. 평가 지표

일반적으로 의학적 진단은 문장이 길고 진단의 근거도 다양하기에 정답과 오답이 명확한 이진 분류(Binary Classification) 문제는 아니지만 현재 모델에서 예측 또는 실험을 잘 했는지 못 했는지를 알기 위해서는 예측 결과를 그림 4와 같이 Confusion Matrix에 나타낸다.

		Positive	Negative	
Predicted Label	True Positive(TP) 252	False Positive(FP) 207	Positive	
	False Negative(FN) 12	True Negative(TN)	Negative	
		True Label		

그림 4. Confusion Matrix [7]

Confusion Matrix는 흔히 Error Matrix로도 알려져 있고 실험 데이터셋을 단순히 정답과 오답이 아닌 오답과 미탐을 포함하여 표로 나타낸 것이다. 앞서 언급한 것과 같이 Confusion Matrix에 실험 결과를 나타내기 위해서는 TP, FP, FN를 일반적인 이진 분류 문제와는 다르게 다음과 같이 정의하였다.[8][9]

- **True positive (TP):** 정답 단어 중에 예측에 포함되어 있는 단어의 수
- **False positive (FP):** 예측 단어들 중에 정답 단어가 아닌 단어의 수
- **False negative (FN):** 정답 단어 중 예측 단어에 없는 단어의 개수

예를 들어 “Liver Cirrhosis”가 정답인 경우 모델에서 “Cirrhosis may present”라는 정답이 나왔다. 이때 예측 문장 중 Cirrhosis 만이 정답 단어에 포함된다. 따라서 TP는 1이다. 그러나 “may present”는 정답이 아닌 예측 단어들로 FP는 2가 된다. 또한 Liver는 예측 단어에 포함되지 않았으므로 FN는 1개이다.

3.2. 정량적 실험결과

45개의 질병에 대하여 각각 3개의 검증데이터를 구성하였고 총 135개의 데이터로 성능평가를 한 결과를 표 3에 나타내었다.

표 3. 성능평가 결과표

	Precision	Recall	F1 score
값	0.55	0.95	0.7

True Positive 총 252개 False Negative 12개 False Positive 총 207개이다. Precision 과 Recall은 각각

0.55 와 0.95이고, F1 score는 0.70이다. 이를 통해 본 연구에서 생성한 데이터가 자연어처리 모델 학습에 이용될 가능성을 확인하게 되었다. 더 나아가 데이터 생성이라는 방법을 통하여 모델 학습용 데이터셋을 구축하는 것에 대한 가능성 또한 발견하게 되었다.

3.3. 정성적 실험 결과

제안한 모델을 통해 각 질의에 해당하는 진단 결과를 정성적으로 확인해 보았고, 표 4와 같다.

표 4. BERT 모델 출력의 정성적 분석

	Ground Truth	Prediction
Best Case	Splenic Infarction	Splenic Infarction
Too Much Extra Token	GERD	GERD can range from Mild to Severe and
Missing Answer Token	Liver cirrhosis	Cirrhosis may present
Worst Case	Gastroenteritis (viral, bacterial)	viral, bacterial) 1. Clinical Manifestation

“Best case”에서는 Splenic Abscess (비장농양)과 Splenic Infarction (비장경색) 중 Splenic Infarction을 정확히 진단해냈다. 그러나 이를 제외한 “Too Much Extra Token” 혹은 “Missing Answer Token” 그리고 “Worst Case” 모두 정답 레이블보다 더 많은 토큰을 Context에서 정답 레이블로 예측했다. 정답에 해당하는 질병 명사와 추가적으로 동사까지 정답으로 예측하는 경우가 대부분이었다. 따라서 정량적 평가에 비해 정성적 평가 결과 Extractive QA 모델의 한계점이 더 명확히 나타난다[10]. 따라서 Extractive 모델에서 Abstractive 모델로 발전시키는 방향으로 추후 연구 방향 설정이 필요하다.

4. 결론

본 연구에서는 최근 화제가 되고 있는 거대 언어 모델 LLM을 이용하여 수준 높고 질 좋은 의료 문답 데이터를 생성하고 오직 생성된 의료 문답 데이터를 이용하여 의료용으로 사전학습 언어모델을 미세 조정하였다. GPT-4와 같은 우수한 성능의 언어모델로 생성된 데이터를 활용하여 학습된 Extractive QA 모델은 0.70의 F1 score를 얻을 수 있었다. 이는 생성된 학습 데이터의 수를 늘리고 모델을 더욱 정교하게 미세조정한다면 실

제 문진 데이터 없이 성능이 발전된 의료 문답 모델을 학습시킬 수 있는 가능성을 확인하였다. 본 논문에서는 Extractive QA 모델을 활용하여 생성된 데이터로 학습한 모델의 성능을 분석하였지만 추후 연구 진행될 Abstractive QA 모델을 활용하여 더 많은 데이터를 학습할 수 있다면 False Positive 값이 감소하여 더 높은 F1 score를 달성할 것을 기대해볼 수 있을 것이다.

Acknowledgments

This work was supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2020040126).

References

[1]Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Wei Liu, Ninghao Liu, Sheng Li, Dajiang Zhu, Hongmin Cai, Lichao Sun, Quanzheng Li, Dinggang Shen, Tianming Liu, Xiang Li .(2023), “AugGPT: Leveraging ChatGPT for Text Data Augmentation”, Forthcoming

[2]Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J.(2020) BioBERT: a pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics*

[3]Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019) “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), pages 4171–4186

[4]Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, You, Zhang(2023) “ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge”, Forthcoming

[5] Luo, Man & Hashimoto, Kazuma & Yavuz, Semih & Liu, Zhiwei & Baral, Chitta & Zhou, Yingbo. (2022). Choose Your QA Model Wisely: A Systematic Study of Generative and Extractive Readers for Question Answering.*Proceedings of the 1st Workshop on Semiparametric Methods in NLP: Decoupling Logic from Knowledge*, pages 7–22

[6]Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. (2016) “SQuAD: 100,000+ Questions for Machine Comprehension of Text”. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Association for Computational Linguistics

- [7]Deepak Kumar Sharma, Mayukh Chatterjee, Gurmehak Kaur, Suchitra Vavilala(2022).“3-Deep learning applications for disease diagnosis”, Deep Learning for Medical Applications with Unique Data, Academic Press, pages 31–51
- [8]Van Rijsbergen, C. J. (1979),*Information Retrieval (2nd ed.)*. Butterworth-Heinemann.
- [9]Y. Sasaki (2007). The truth of the F-measure, Technical Paper, School of Computer Science, University of Manchester
- [10]Miyoungh Ko,Jinhyuk Lee,Hyunjae Kim,Gangwoo Kim,Jaewoo Kang(2020)"Look at the First Sentence: Position Bias in Question Answering", EMNLP

Appendix

소화기 내과에서 겪는 모든 종류의 임상표현/증상을 생성한 결과는 다음과 같다.

Oral Cavity and Esophagus:

1. Dysphagia (difficulty swallowing)
2. Odynophagia (painful swallowing)
3. Heartburn
4. Regurgitation (the sensation of fluid or food coming back up)
5. Chest pain related to eating
6. Water brash (sudden excess of saliva)
7. Halitosis (bad breath)

Stomach and Duodenum:

1. Epigastric pain (pain in the upper middle part of the abdomen)
2. Nausea
3. Vomiting
4. Hematemesis (vomiting blood)
5. Early satiety (feeling full quickly)
6. Anorexia (loss of appetite)
7. Weight loss
8. Bloating or feeling gassy

Small Intestine and Colon:

1. Diarrhea
2. Constipation
3. Hematochezia (passing fresh blood through the rectum)
4. Melena (passing dark, tarry stools due to upper GI bleed)
5. Fecal incontinence (inability to control bowel movements)
6. Abdominal pain or cramps

7. Flatulence (passing gas)
8. Steatorrhea (excess fat in feces, often causing them to be foul-smelling and greasy)
9. Change in bowel habits
10. Tenesmus (constant urge to pass stools)

Liver, Gallbladder, and Pancreas:

1. Jaundice (yellowing of the skin or eyes)
2. Ascites (buildup of fluid in the abdomen)
3. Pruritus (itching)
4. Dark urine
5. Clay-colored or pale stools
6. Hepatomegaly (enlarged liver)
7. Spider angiomas (spider-like blood vessels visible under the skin)
8. Palmar erythema (reddening of the palms)
9. Right upper quadrant pain
10. Caput medusae (dilated veins around the belly button)
11. Asterixis (flapping tremor of the hand when the wrist is extended)
12. Dupuytren's contracture (thickening and tightening of the tissue under the skin of the palm)
13. Gynecomastia and testicular atrophy in men (due to liver disease)

General GI Symptoms:

1. Fatigue
2. Fever
3. Anemia (low red blood cell count)
4. Nutrient deficiencies

각각의 임상표현에 대한 가능한 모든 원인들

1. Gastrointestinal Causes

- Peptic ulcers (gastric or duodenal)
- Gastritis
- Gastroenteritis
- Gastroesophageal reflux disease (GERD)
- Gastric cancer
- Pancreatitis
- Gallbladder disease (such as gallstones or cholecystitis)
- Hiatal hernia
- Dyspepsia (indigestion)
- Pyloric stenosis
- Celiac disease
- Inflammatory bowel disease (such as Crohn's disease or ulcerative colitis)
- Small intestinal bacterial overgrowth (SIBO)

2. Cardiovascular Causes

- Myocardial infarction (heart attack)
- Angina pectoris
- Aortic dissection

3. Respiratory Causes

- Pneumonia (especially right lower lobe)
- Pleurisy

4. Musculoskeletal Causes

- Costochondritis
- Muscle strain

5. Metabolic Causes

- Diabetic ketoacidosis
- Hypercalcemia

6. Psychogenic Causes

- Stress
- Anxiety
- Somatoform disorders

7. Others

- Pregnancy (especially ectopic pregnancy)
- Hepatitis or liver disease
- Spleen disease
- Kidney disease, including kidney stones
- Certain infections like helicobacter pylori
- Certain medications that can cause stomach upset

3. Epidemiology:

Achalasia is relatively rare, with an estimated incidence of approximately 1 per 100,000 individuals per year. It affects men and women equally and can occur at any age, but it's more commonly diagnosed in middle-aged or older adults.

4. Etiology:

The exact cause of achalasia is unknown. It's thought to occur when the nerves in the esophagus become damaged, leading to a failure of the esophageal muscles and the lower esophageal sphincter (LES) to function properly. Some theories suggest a viral cause, and there is also a rare familial form suggesting a genetic component.

5. Pathophysiology:

In achalasia, the loss of inhibitory ganglion cells in the myenteric plexus of the esophagus leads to a failure of relaxation of the lower esophageal sphincter, and a lack of peristalsis in the esophageal body. This results in the classic symptoms of dysphagia and regurgitation.

6. Differential Diagnosis:

Conditions to consider in the differential diagnosis include esophageal cancer, gastroesophageal reflux disease (GERD), esophageal stricture, and other motility disorders such as diffuse esophageal spasm and nutcracker esophagus.

7. Risk factor/predisposing factor

This condition is caused by damage to the nerves of the esophagus, though the exact reason for this damage is unknown. It may be related to an autoimmune condition or infection.

8. Any Previous Medical History that is related to the disease:

There is no known previous medical history linked to the development of achalasia. It can coexist with autoimmune conditions, but it is not clear if there is a direct correlation.

그림 2 질병 지식 생성 프롬프트의 결과는 다음과 같다.

Achalasia

1. Clinical Manifestation/Symptoms:

Achalasia presents with difficulty swallowing (dysphagia), regurgitation of undigested food, chest pain, weight loss, and occasionally, coughing at night. The dysphagia typically occurs with both solids and liquids, and tends to progressively worsen over time.

2. Natural History:

Achalasia is a chronic, progressive disorder. The course of the disease varies, but most patients will experience progressive difficulty in swallowing and worsening regurgitation over time. Treatment can help manage symptoms, but there's no cure for achalasia.

그림 3 가상환자 생성 프롬프트의 결과는 다음과 같다.

A 62-year-old woman presents to the clinic with a 6-month history of difficulty swallowing. She describes it as a feeling of food sticking in her chest, and it occurs with both solids and liquids. She also reports instances of undigested food coming back up several hours after eating, and she's noticed a decrease in her weight over

the past few months. She has a history of thyroid disease, but no history of gastrointestinal disorders.

What is the most likely diagnosis of the patient?

The most likely diagnosis for this patient is Achalasia. Her symptoms of progressive dysphagia for both solids and liquids, regurgitation of undigested food, and weight loss are characteristic of this condition. While her history of autoimmune thyroid disease is not directly linked to achalasia, there is a potential for other autoimmune conditions to coexist.