

Konyang University

# KYMEKA Dataset Manual

Research director Prof. Key-Sun Choi

Team KYMEKA

2024. 03. 10

# Index

Chapter 1 Introduction·····	01
Section 1 Background·····	01
Section 2 Related works ·····	01
Section 3 Purpose·····	03
 Chapter 2 KYMEKA Dataset·····	01
Section 1 Type-1.0: Dataset collected at Konyang Hospital·····	01
Section 2 Type-2.1: MIMIC-CXR 유래 진단과정 주석 데이터세트·····	01
 Chapter 3 Manual Annotation·····	12
 Chapter 4 Downstream Task·····	14
Section 1 Generating causal section task·····	14
Section 2 Generative annotation task·····	17
 Reference·····	19

# Chapter 1 Introduction

## Section 1 Background

“이미지 캡셔닝(Image Captioning)”은 이미지를 입력받아, 그 특징에 따른 텍스트 묘사 생성, 분류, 인식, 등을 수행하는 인공지능 분야의 한 과제이다[1-3]. 특히 비전과 언어이해를 같이 수행하므로 더욱 중요하다고 하겠다. 이러한 이미지 캡셔닝을 의학 도메인의 CXR (Chest X-ray)에 적용한 분야가 “방사선학 보고서 생성(Radiology Report Generation)”이다. 다시 말하면, 방사선학 보고서 생성은 방사선 사진을 입력으로 받아 이미지 캡셔닝을 하며 자세한 Finding 과 Impression 을 획득하여 보고서를 출력하는 과제이다[4]. 그러나 이 타스크는 이미지 캡셔닝 뿐만 아니라 분야지식(domain knowledge)이 동원되어야 그 목적을 달성할 수 있다. 왜냐하면 병증, 증상, 원인이 포함된 소견과 추적진단은 이미지캡셔닝만으로는 불가능하기 때문이다[2].

## Section 2 Related works

최근 연구에서는 Radiology Report Generation Task 를 다루는 연구와 그 연구들에서 사용하는 데이터에 대해서 다룬다.

Source		Image captioning	KB based Impression	Causal Classification
Author	Title			
Jing et al. 2018	On the Automatic Generation of Medical Imaging Reports. [5]	O	X	X
Chen et al. 2020	Generating Radiology Reports via Memory-driven Transformer. [6]	O	X	X
Wang et al. 2022	Prior Knowledge Enhances Radiology Report Generation. [3]	O	O	X
Yang et al. 2023	Radiology report generation with a learned knowledge base and multi-modal alignment. [7]	O	O	X
Our Project		O	O	O

표 1. 관련 연구

방사선 보고서 생성 TASK의 연구들은 TASK의 확장을 위해 Knowledge Base 를 기반으로 보고서의 finding 과 impression 생성을 강화했다. 표 1 에서 볼 수 있듯이, 이 연구들은 흉부방사선 사진으로 image captioning 을 바탕으로 보고서를 생성한다. 또 질병의 추적 표현 (impression) 을 향상시키기 위해 여러 방법을 시도했다. 위의 연구들은 MIMIC-CXR 및 IU X-ray (Indiana University Dataset)를 바탕으로 수행되었다. Jing et al. 2018 의 연구에서는 방사선 보고서 생성을 위해 흉부방사선 사진을 입력하여 CNN (Convolutional Neural Network)으로 visual features, 사진에 따른 tags 를 semantic features 로 활용해서 보고서를 생성했다[5]. 생성에 활용된 방법은 Co-Attention 으로 features 를 함께 고려했다. Chen et al. 2020 의 연구는 흉부방사선 사진을 입력하여 multi-head attention 을 통해 방사선 보고서를 생성했다[6]. 이 연구는 정확한 생성을 위해 image captioning 의 localization 문제에 patch features 와 연결된 masked multi-head attention 을 적용했다. Wang et al. 2022 의 연구는 image captioning 에 해당하는 image encoder 와 disease 을 추적하여 impression 을 생성하기 위한 knowledge graph 을 함께 활용했다[3]. 이 연구는 흉부방사선 사진과 보고서 입력의 특징을 graph convolutional network 로서 보고서 생성을 위한 디코더로 사용한다. 분야 지식이 강조되는 방사선 보고서 생성에서 disease 을 추적하는 것으로 성능을 향상 기대했다. Yang et al. 2023 의 연구에서는 흉부방사선 사진을 입력하여 보고서를 생성하지만 생성의 방법에서 이미지에 대한 key embedding 으로 knowledge base 를 사용한다[7]. 이 방식은 이미지에 따라 knowledge base 를 자동 업데이트하여 보고서 생성에 필요한 KB 를 최적화하는 것에 기여가 있다.

우리의 연구는 진단의 원인을 포함하는 데이터를 기반으로 원인에 대한 분류 문제를 제공할 수 있다. 이는 관련 연구들에서 병증, 증상, 원인이 포함된 소견과 추적진단을 이미지캡셔닝만으로 불가능한 문제를 해결할 수 있는 데이터세트이다. 이 데이터를 통해 생성된 병변의 원인을 설명하는 보고서는 기존의 이미지캡셔닝, KB(Knowledge base) 기반의 Impression 생성 그리고 원인에 대한 분류를 통해 환자의 병증을 설명하는 것에 원인을 포함한 분야지식을 고려한다. 이렇게 고려된 분야지식은 원인을 고려하여 관련된 질병을 찾거나 원인을 언어로 표현하는 방식에서 기존의 연구와 차이가 있다.

Dataset name	Release year	Size	Resource				
			X-Ray	Report	Disease label	Lesion	Cause of disease
Our Dataset	2024	300	O	O	O	X	O
PadChest [8]	2020	206,222	O	O	O	O	X
MIMIC-CXR [9]	2019	371,920	O	O	O	X	X
CheXpert [10]	2019	224,316	O	X	O	O	X
National Institutes of Health Repository [11]	2017	112,120	O	X	O	X	X
PLCO [12]	2012	185,421	O	X	O	X	X

표 2. 관련 데이터세트

표 2는 관련 데이터세트를 비교한다. 표 2의 데이터세트는 모두 Radiograph를 기반으로 수집된 데이터세트이며, 표의 resource를 바탕으로 다양한 task에서 활용된다. 첫째로, X-ray 이미지와 병표지(disease label)로 구성된 데이터세트일 경우, X-ray 이미지분할을 통하여 병변을 인식하여 병표지를 분류하는 task를 수행할 수 있다. 둘째로, 첫째 데이터세트에 부가하여 병변위치(lesion location)이 추가되면 첫째 데이터세트보다 더 정확한 이미지캡셔닝 결과를 도출할 수 있다. 셋째로, 둘째 데이터세트에 소견보고서가 추가된다면 단순한 이미지캡셔닝에서 벗어나 방사선 보고서생성 task가 가능해진다. 넷째로, 진단의 원인을 포함하는 자원이 있다면 질병의 원인추론이 가능한 데이터세트가 된다. 이 문서의 데이터세트는 이 항목을 포함한다. 따라서 기존의 이미지캡셔닝 task에 인과추론을 포함하는 방사선학 보고서생성 task가 가능하게 된다.

## Section 3 Purpose

본 논문에서는 병진단의 원인을 포함하는 방사선 보고서생성을 수행할 수 있는 데이터를 소개하고 이를 활용하는 방법에 대해 기술한다. 이 데이터는 기존 데이터에서 부분적으로 포함하고 있는 병증의 해부학적 위치와 병증의 위치에 따른 척추 위치 이외에 병증의 원인을 모든 케이스에 대해 포함한다. 이 데이터는 최근 많은 연구자가 활용하는 MIMIC-CXR 데이터세트를 기반으로 영상 의학과 전공의로부터 수동 어노테이션을 통해 만들어졌다.

## Chapter 2 KYMEKA Dataset

### Section 1 Type-1.0: Dataset collected at Konyang Hospital

Type-1.0 는 건양대학교병원 영상의학과에서 획득한 데이터세트이다. 이 데이터는 후향적 연구로 수행되었으며, 2018.01.01.부터 2021.12.31.까지 건양대학교병원 호흡기과 내원 환자 대상 흉부 단순촬영 무작위 이미지와 의사의 주석 및 진단으로 구성된다. 의사의 주석은 흉부 방사선영상 판독문(Reading)에 포함된다. 의사의 진단은 Disease 에 포함되어 있다. 또한, 환자의 나이, 성별, 흡연력 그리고 과거력을 포함한다. 코호트 구성을 위해 최대한 다양한 샘플을 대상으로 수집되었다. 아래는 한 환자에 대한 데이터 구성과 설명이다.

“(300)\_reading\_EMR\_ver.1.0.xlsx”: 의사의 주석, 진단 및 환자 정보가 포함

- Case Title: 수집된 데이터의 고유 번호
- Sex: 환자의 성별
- Age: 환자의 나이
- Exam Name: 흉부방사선 촬영의 방식
- Exam Date: 촬영 일시
- HT: 고혈압 (0: No, 1: Yes)
- Diabet: 당뇨병 (0: No, 1: Yes)
- HDL: 고지혈증 (0: No, 1: Yes)
- Smoking: 흡연력 (-1: Previous, 0: No, 1: Yes)
- TB: 결핵력 (0 : 음성 / 1 : 양성 / 0+ : 양성 후 완치)
- ETC: 기타 과거력
- Conclusion: 의사의 판독문
- Disease: 병명

아래는 흉부방사선 사진을 제외한 Type-1.0 의 샘플로 사용자에게 제공되는 형식이다. 이 샘플은 하나의 샘플로 데이터 구성의 항목을 모두 가지고 있으며 환자의 과거력과 정보가 강조된 데이터세트이다.

```
[{Study_Num212489{Inspection Date{20180928}}{Inspection
Time{111715}}{Inspection Name{Chest PA(AP)}}{Sex{M}}{Birth{22080320}}{Conf
Date{20180928}}{Conf Time{144640}}{Reading Date{20180928}}{Reading
Time{144643}}{Smoking:0}{Diabet:0}{Diabet Period:null}{Diabet
Hospital:null}{HT:1}{HT Period:2244}{HT Hospital:1}{TB:1}{TB Period:2,268}{TB
Hospital:1}{HDL:1}{ETC:Alcoholism, dependence syndrome(2263), Sequelae of
respiratory andunspecified(2261), tuberculosis(2262 년), Conjunctivitis,
unspecified(2260), Unstable angina(2258), Acute myocardial infarction,
unspecified(2260), Angina pectoris(2263), TFCA(2274),stent a}{Reading:active
lesion in both lung fields.}{Disease:Pleural effusion}]
```

위 예시에 따르면, ID 는 212489 이며 검사한 항목은 Chest PA, 검사시간은 20180928, 111715 이다. 성별은 남성으로 생년월일과 시간 정보에 대한 부분은 데이터 비식별화를 위하여 임의의 수로 채워졌다. 이 케이스는 흡연력과 당뇨를 진단 받은 경험을 갖지 않으나 고혈압과 결핵에 대한 과거력이 존재한다. 또한 기타 과거력으로 알콜의존증 (2263 년 진단), 호흡기 휴유증 (2261 년 진단), 결막염, 협심증, 급성 심근경색을 진단 받았었다. 케이스에 대한 의사의 진단은 Pleural effusion 이고 판독에 따르면 양쪽 폐부에 모두 이상을 보였다.

## Section 2 Type-2.1: MIMIC-CXR 유래 진단과정 주석 데이터세트 (Disease diagnosis process annotated dataset augmented from MIMIC-CXR)

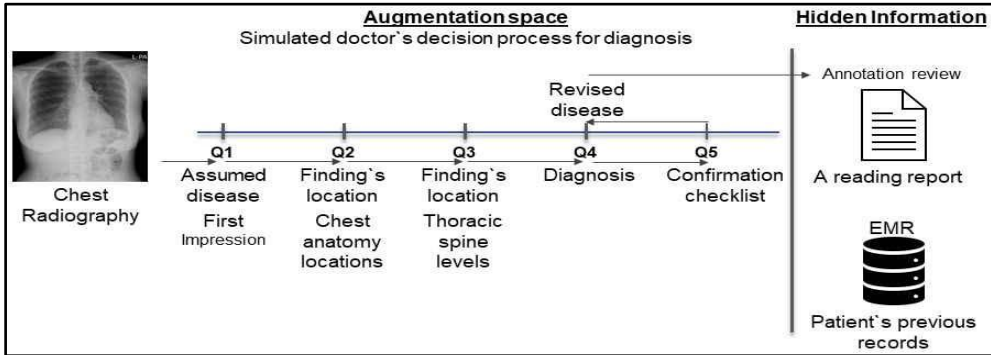


그림 1 dataset annotation process

Type-2.1 는 MIMIC-CXR 을 바탕으로 진단의 원인에 대한 데이터 증강을 수행된 데이터이다. 이 데이터는 진단 절차에 따른 진단의 원인을 물어보는 질문과 답변(Question & Answer)의 형태로 이루어져 있다. 이 데이터세트의 목표는 원 데이터에 질병진단의 원인을 증강하기 위한 것으로서 진단 절차 과정에서 그 진단의 원인이 도출되고 확인된다는 점에 주목을 한 것이다. 그림 1 과 같이, 질문은 왼쪽의 Q1 부터 Q5 까지의 순서대로 질의응답이 수행된다. 즉, 데이터세트는 5 개의 문답으로 구성돼 1 번부터 4 번까지는 의사가 흉부 방사선 사진이 주어지면 첫 인상과 해부학적 정보를 판단한 후 진단을 만드는 것이다. 5 번 문항은 검증을 위한 체크리스트로 진단 원인에 해당한다. 이 5 번 문항은 “의사 자격증 워크북” [13-20]에서 추출한 문항으로 질환의 원인에 대한 체크리스트이다. 마지막으로 이 원인 체크리스트를 살펴본 의사가 최종적으로 진단을 교정한다. 데이터주석과정에 참가한 의사에게는 판독 보고서와 환자의 이전 EMR 을 주석 검토에 활용할 수 있도록 제시 여부를 결정할 수 있다.



# Datasets Overview

- Organize causes and effects of data in QA format.
- Q1~4 are the effects of expert knowledge about the data.
- Q5 is the cause of the effect.
- Q4 option covers 94.398% of MIMIC-CXR.
- Connected to Korea's healthcare terminology standards.

## Products

300 data pairs consisting of DICOM, reports, QA.

## Original Source (MIMIC-CXR) Info

Image : DICOM files dir  
Report : Report files dir

## Question

Q1 : Patient's abnormal finding  
Q2 : Anatomy information from abnormal finding  
Q3 : Thoracic spine level from abnormal finding  
Q4 : Words that can be included in readings  
Q5 : Sub-question about words(Q4)

## Answer

A1 : Descriptive answer  
A2 : Multiple choice  
A3 : Multiple choice  
A4 : Multiple choice  
A5 : O/X selection for each question

그림 2 Dataset overview

그림 2는 데이터세트의 구성과 특성을 보인다. 데이터세트의 원본은 흉부 방사선영상, 보고서를 포함하며, MIMIC-CXR에서 등장하는 질병의 약 95%의 분포에 따라 선택한 원천 데이터를 대상으로 한 질병진단프로세스에 따른 문답(QA)으로 이루어져있다. 2023년 12월 현재 300쌍으로 구성되어 있다. 각 문답의 용어는 한국의료용어표준에 준거한다. 아래는 [Github](#)<sup>1</sup>에서 접근할 수 있는 이 데이터에 대한 설명이다.

“(Non-licensing)KYMEKA\_Datasets\_without\_log.csv”는 수동 주석으로 수집한 데이터 세트이다. 이 파일은 의사의 진단 원인을 문답 형태로 기록한 것으로서 그림 1,2에서 설명한 5개의 문답내이다. 이 파일은 아래 두 개의 파일과 연관되어 있다.

"Dataset1\_QA\_Case\_ver.1.0"는 데이터 세트에서 답변에 나올 수 있는 모든 가능한 객관식 보기에 대한 것이다. 질문과 보기는 Sheet로 나뉘어져 있다.

문답 내의 모든 용어는 보건복지부 용어표준집 “Korea's healthcare terminology standards\_V7.0.xlsx”을 따른다.

<sup>1</sup> Github: <https://github.com/KYMEKA/KYMEKA.github.io>

“(Non-licensing)KYMEKA\_Datasets\_without\_log.csv”:

- ID: Offset Annotation 의 ID
- Report\_dir: MIMIC-CXR 의 Report Directory
- DICOM\_dir: MIMIC-CXR 의 DICOM Directory
- A1: 질문 1 에 대한 답변
- A2: 질문 2 에 대한 답변
- A3: 질문 3 에 대한 답변
- A4: 질문 4 에 대한 답변, 질문 4 는 한 케이스에서 최대 3 개까지 선택이 가능
- A5: 질문 5(Column Name)과 답변 (Values)로서 A4 에서 설정한 질병 각각 하나마다 28 개 확인 체크리스트.

"QA\_Options\_ver.1.0.xlsx":

- 4 개의 Sheet 로 구성되며, 아래와 같다.
- “QA” sheet: 질의응답 형태의 설계를 확인할 수 있다. 이 Sheet 의 column 은 아래와 같다.
  - 분류: 진단의 원인과 결과를 표현한 분류
  - QA 번호: 질의응답의 넘버링
  - Worker Input: 작업자 (의사)에게 제공하는 자료
  - 질문형식: 주관식, 객관식, Checklist 중 하나로 질문의 성격
  - 질문: 작업자 (의사)에게 제공하는 질문
  - Normal Range: Checklist 에서 제공하는 질문의 가이드라인
  - Ref: Checklist 의 출처
  - Sample Answer: 질문의 Sample
  - 보기 위치: 객관식의 경우에 답변의 선택지 위치 sheet
- “QA2 보기” sheet: 질의응답 2 의 선택지
- “QA3 보기” sheet: 질의응답 3 의 선택지
- “QA4 보기” sheet: 질의응답 4 의 선택지

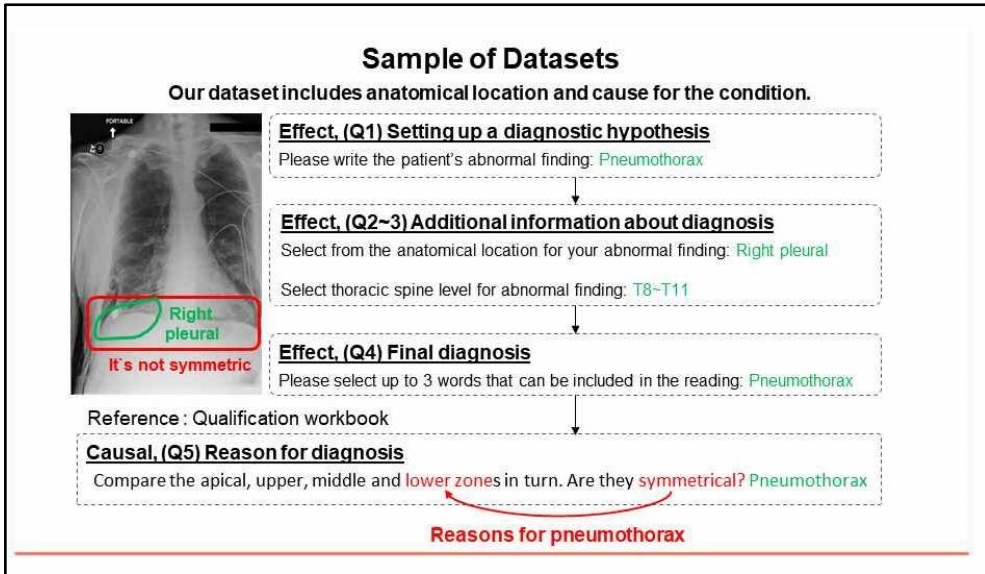


그림 3 Dataset sample

그림 3은 데이터세트의 샘플이다. 데이터세트에는 해부학적 위치와 이유가 포함된다. 첫 번째 질문은 전문가에게 흉부 방사선 사진을 보여주고 진단에 대한 가설을 작성하도록 하는 것이다. 두 번째와 세 번째 질문은 진단에 필요한 추가 정보를 만든다. 그림은 오른쪽 흉막 및 척추 레벨 t8~11을 기흉의 해부학적 위치로 보여준다. 네 번째 질문은 방사선 사진의 최종 진단이다. 이것은 진단에 이르는 가설을 보여주는 한 예이다. 다섯 번째 문제는 질병의 원인을 선정하는 checklist이다. 이 데이터세트는 질문 5를 진단의 이유로 생각한다. 이는 진단의 원인인 5번 문항을 다른 문항과 함께 고려하여 명시한다.

그림 3에서 볼 수 있듯이, 사진에서 pneumothorax 진단 이유는 여러 구역의 비대칭성 때문에 설정되었다. 답변 2에서 볼 수 있듯이, 오른쪽 흉막이 질병의 해부학적 위치가 된다. 그래서 답변 5의 여러 구역 중 Lower Zone이 비대칭성의 원인이며, 이 원인이 진단의 이유가 된다.

## Section 3 How to access datasets

### - Type-1.0 Licensing.

Type-1.0 데이터세트는 건양대병원에서 직접 획득한 데이터세트로 병원의 임상 환자를 대상으로 후향적으로 획득되어 과거력과 같은 민감한 정보를 포함한다. 데이터사용을 위해서는 건양대학교 병원의 IRB(Institutional Review Board) 규정에 따라야 한다. 이에 따라, 본 KYMEKA 관리팀은 건양대학교 병원의 IRB 심사에 따라 데이터 접근을 허가한다. 데이터에 접근하고자 하는 개인, 팀, 조직에서는 데이터 사용 허가를 심사 받아 그 결과를 KYMEKA 팀에 제출해야 한다. 또한, 본 데이터는 연구 용도로만 제약한다.

### - Type-2.0 Licensing.

Type-2.0 데이터세트는 MIMIC-CXR로서 physionet에서 배포하는 공개 데이터세트이다. 데이터세트를 사용하기 위해서는 관련 교육을 수행하여, 데이터 주관 기관의—사용 인증을 허가받아야 한다. 데이터세트 자체가 인체 유래물에 해당하고 개인정보가 상당히 포함되어 있어 심사 과정 및 교육에 대한 절차는 아래와 같다.

#### - Physionet 회원가입

회원가입은 아래의 URL 을 통해서 수행된다.

<https://physionet.org/content/mimic-cxr/2.0.0/files/#files-panel>

- 데이터에 요구되는 교육 완료  
교육은 CITI Course 로 구성되며, 기본적인 소개는 <https://physionet.org/about/citi-course/> 에서 확인할 수 있다. 교육은 전자문서를 바탕으로 수행되는 시험의 형식이며 교육 사이트는 다음과 같다. (<https://www.citiprogram.org/members/index.cfm?pageID=50>) 교육은 다양하게 존재하지만, 데이터 획득을 위해 수행해야 할 리스트가 존재하며 교육 리스트는 위 URL 에서 확인할 수 있다.
- Certification 다운로드 및 인증서 제출  
해당 교육을 모두 완료 후에 "My Records"에서 인증서를 내려받을 수 있다. 이후에 내려받은 인증서를 <https://physionet.org/settings/training/>에서 제출한다.
- 데이터 활용 동의서 작성  
<https://physionet.org/content/mimic-cxr/2.0.0/files/#files-panel> 위 링크의 가장 하단에 지금까지 소개한 내용을 포함한 마지막 과정이 있다. 데이터 사용 동의서를 클릭하시고 로그인 후에 서명을 수행하면 심사가 요청된다. 해당 과정은 2~8 주 정도의 심사 기간이 소요된다.
- KYMEKA 팀으로의 인증 제출  
지금까지의 과정을 모두 수행하고 MIT 에서의 심사까지 취득한 연구자는 KYMEKA 팀으로 해당 인증서 및 KYMEKA 팀의 데이터사용 허가 문서(구글 설문 링크)를 제출한다. 그 후 권한 허가를 위해 함께 검토된다.

## Chapter 3 Manual Annotation

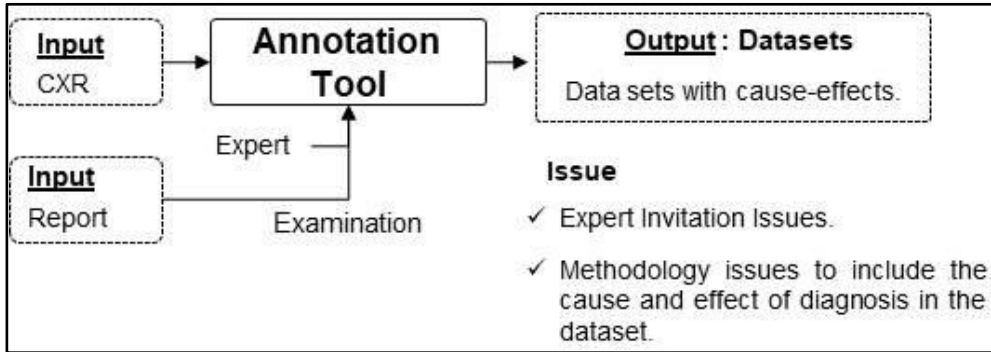


그림 4 Manual Annotation

여기서는 Type-2.0: MIMIC-CXR 진단절차 주석데이터세트(diagnosis process annotation dataset)를 위한 주석 작업을 소개한다. 우선 본 KYMEKA team에서 구성한 주석 플랫폼(annotation platform)을 소개한다. 그림 4는 수동 주석의 방법과 이슈를 소개한 그림으로 작업자는 전문가이고 그림 4와 같이 흉부방사선 영상을 주석 플랫폼에서 접근하여 의료전문가가 주석툴의 UI 플랫폼 화면에 따라 주석을 수행한다. 이때 이미 존재하는 소견서(report)는 클라우드소싱 정책에 따라 플랫폼에서 개방할 수도 있고, 결과물에 대한 검수단계에서 비교 데이터로 사용한다. 주석에 대한 결과물은 각 질문에 대한 답과 함께 작업자 행동이 기록되어, 질병진단의 원인과 결과를 포함하는 데이터가 된다. 완료된 작업물은 교차 검수를 하였다.

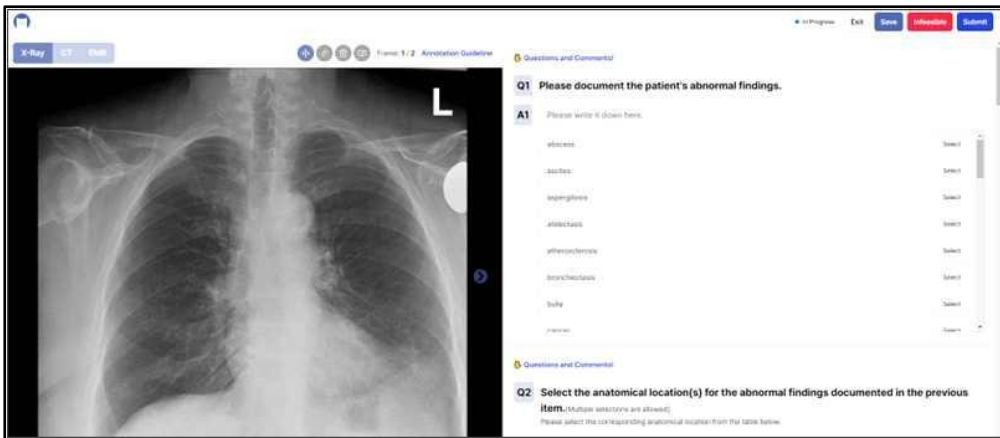


그림 5 Screen of Annotation Platform

그림 5에서 주식플랫폼의 질의응답 UI를 볼 수 있다. 이 플랫폼은 위치는 다음에서 시작한다. (<https://enterprise.labelr.io/>) 사용자 등록 등의 절차가 필요하며 아래에서 설명한다.

데이터에 대한 접근은 MIMIC-CXR licensing issue로 인하여 비공개되어있다. 접근을 하기 위해서는 KYMEKA github에 있는 구글 설문의 요구사항에 따라야 한다. (<https://github.com/KYMEKA/KYMEKA.github.io>) 비공개된 github 저장소에 초대되면, 주식 플랫폼에 대한 지침(guideline)에 접근할 수 있다. 지침에는 플랫폼과 작업에 대한 전반적인 절차가 소개되어 있다.

본 진단 원인 주식 프로젝트에서는 주식작업 전문가를 초청하고 있다. 또 진단 원인 주식방법을 같이 연구할 전문가를 초청하고 있다. 이 문제에 흥미가 있는 전문가는 KYMEKA github에 있는 구글 설문에 따라 작성을 부탁하며, 많은 전문가들의 흥미와 지원을 기다리고 있다. KYMEKA team에서는 이러한 문제를 해결하기 위해 해당 분야의 전문가를 항시 모집하고 있다.

## Chapter 4 Downstream Task

지금까지 구축한 데이터세트를 활용하는 과제를 소개한다. 첫 째 방사선학 보고서 생성(radiology report generation)에서 질병진단의 원인을 생성하는 과제이다. 두 번째는 데이터세트의 수동 주석의 어려움을 극복하고자 수행한 생성적 주석 과제이다. 모든 task 는 Type-2.0: MIMIC-CXR offset annotation dataset 를 기준으로 한다.

### Section 1 Generating causal section task

이 task의 목적은 입력으로 임의의 환자에 대한 방사선학 보고서가 주어졌을 때, 그 질병진단의 원인을 보고서로 생성하고자 하는 것이다. 이는 방사선학 보고서에 더 정확한 정보를 포함하는 작업이다. 질병진단의 원인을 포함하여 생성된 보고서로 의사에게 더 많은 정보를 제공하여 진단의 효과를 증대시킬 수 있다. 또, 환자에게 맞춤형 방사선학 보고서를 제공하여 의료 서비스의 질적 향상을 기대할 수 있다. 이 task는 MIMIC-CXR 의 보고서를 입력으로 활용하고 GPT 3.5 를 기반으로 보고서 생성(report generation)을 수행한다 [21]. 출력은 질병의 원인에 대한 보고서 섹션이다. 평가를 위해서 출력은 수동 주석으로 수집된 데이터의 보고서를 정답(golden standard)으로 하여 비교한다.



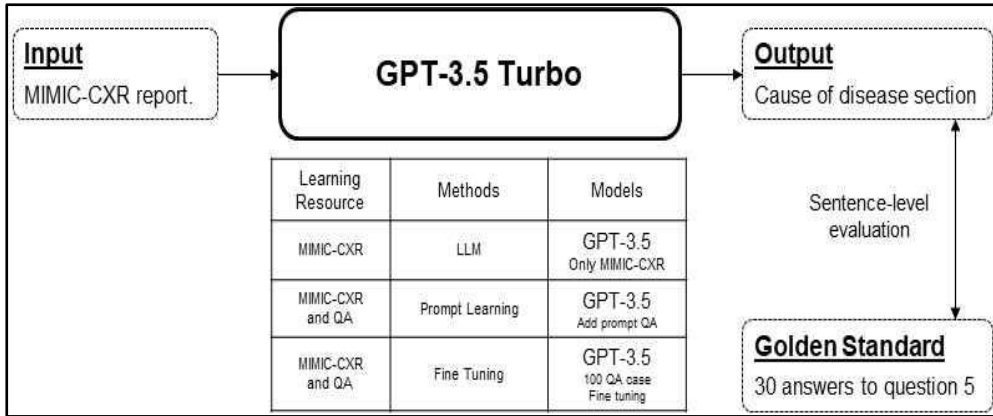


그림 6 Generating causal section task process

그림 6 은 이 task의 흐름도이다. 보고서 생성의 방법은 prompt learning 과 fine tuning 의 비교로 구성된다. 이 실험을 위하여 보고서, 질의응답 그리 고흡부방사선 사진과 연결된 데이터 쌍 300 건 중, 30 개를 평가용으로 사용했다.

Learning Resource	Methods	Models	Metric	Rouge-1	Rouge-2	Rouge-3
MIMIC-CXR	LLM	GPT-3.5 Only MIMIC-CXR	Recall	0.2066	0.0833	0.1614
			Precision	0.2308	0.0714	0.1795
			F1-score	0.2171	0.0714	0.1692
MIMIC-CXR and QA	Prompt Learning	GPT-3.5 Add prompt QA	Recall	0.3634	0.2111	0.3379
			Precision	0.3838	0.2804	0.3641
			F1-score	0.3692	0.2379	0.3470
MIMIC-CXR and QA	Fine Tuning	GPT-3.5 100 QA case Fine tuning	Recall	0.0905	0.0	0.0648
			Precision	0.1347	0.0	0.1044
			F1-score	0.1068	0.0	0.0791

표 3. Generating causal section task baseline result

(1) Only MIMIC-CXR GPT-3.5	(2) Fine tuning 100 case GPT-3.5
It's important to note that the specific cause for pneumothorax in this case is not provided in the report. The information only describes changes in the size of the right pneumothorax compared to a prior study.	The pneumothorax in this case may be attributed to a combination of factors, including trauma and anatomical location. The right pneumothorax observed at the T8~T11 thoracic spine level in the right pleural space indicates a localized issue in the upper to middle region of the right lung. The lack of symmetry in the apical, upper, middle, and lower zones suggests an asymmetric distribution of air in the pleural space, further confirming the presence of pneumothorax.

그림 7 causal section generation task result samples: (1) left side: report generation by only hospital (here, MIMIC-CXR) report, (2) right side: by using fine tuning 100 case and GPT-3.5 with KyMeka causal dataset.

표 3 과 그림 7 은 이 task의 결과와 그 샘플이다. 그림 7 의 두 가지 실험방법으로 생성한 보고서의 “원인 섹션”을 비교해보자. 좌측은 학습 없이 GPT 3.5 를 통해 생성한 보고서이며, 우측은 100 case 의 데이터로 fine tuning 한 보고서이다. 앞서 그림 3 에서 KyMeka diagnosis causal 데이터세트를 그림 7(2)에서 사용한 데이터 샘플이다. 그림 8(1) 과 같이 KyMeka 데이터를 학습하지 않았을 때 진단의 원인을 찾을 수 없다고 생성하고 있다. 반면, KyMeka 데이터로 fine tuning 한 경우에는 진단의 원인이 초록색으로 표기되어 생성된다. 이는 그림 2 에서 샘플의 해석과 유사하다. 이 실험에서 방사선학 보고서(Radiology report)의 원인 섹션 생성 결과를 문장 단위로 평가하기 위해서 GPT-3.5 Turbo 를 사용하여, 보고서의 cause of disease section 을 생성하여 KyMeka 데이터세트인 전문의 주석과 비교했다. 이 task는 병변을 안정적으로 설명하는 정확한 보고서 생성을 위해 원인과 결과 관계를 포함하는 데이터세트가 유효함을 보여 준다.

## Section 2 Generative annotation task

이 task의 목적은 데이터주석을 생성 AI를 활용한 자동 주석 생성이다. 이 task의 baseline은 LLM based on transformer를 사용했다. 생성적 주석 task의 전체적인 방향은 문장 생성 (A1), 해부학적 위치 (A2~3), 진단 (A4), 원인 분류 (A5)로 나누어진다. 그림 8는 원인 분류 (A5)의 task 수행 흐름도이다.

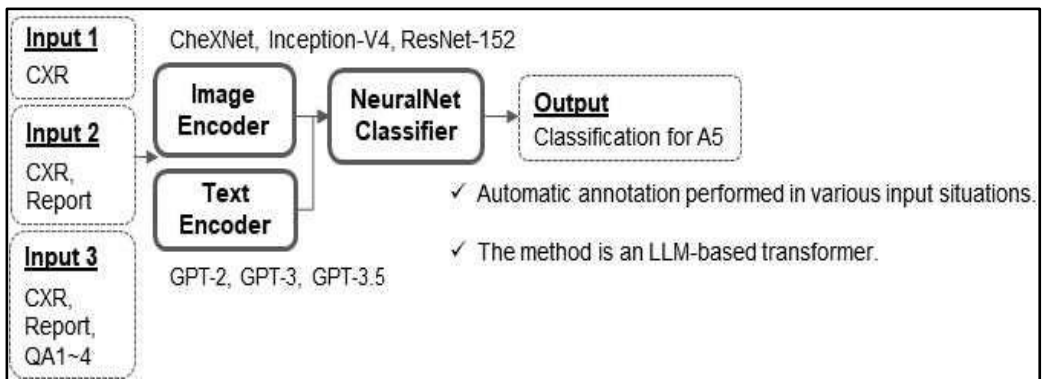


그림 8 Generative annotation downstream task process

그림 8에서 볼 수 있듯이, 이미지 인코더는 CheXNet, Inception-V4, ResNet-152를 사용했고 텍스트 인코더의 경우에는 GPT-2, 3, 3.5를 사용했다[22-26]. 이후 인코더의 가중치를 합하여 neural network 기반 분류기로 답변 5를 선택한다.

Input 1		Input 2 (with report)			Input 3 (with report and A1-4)		
IMG_E	Accuracy	Txt_E	IMG_E	Accuracy	Txt_E	IMG_E	Accuracy
ResNet	0.0698	GPT-2	CheXNet	0.4100	GPT-2	CheXNet	0.5341
Inception	0.0120	GPT-3	CheXNet	0.4438	GPT-3	CheXNet	0.6211
CheXNet	0.4311	GPT-3.5	CheXNet	0.5141	GPT-3.5	CheXNet	0.6254

표 4. 생성적 주석의 베이스라인 결과

표 4은 그림 8 절차의 결과이다. 표 4에서 볼 수 있듯이, 다양한 입력 상황에서 개별적인 실험이 수행되었다. 실험의 결과는 분류의 문제로 정확도 비교를 수행했다. 입력 1은 흉부 방사선 사진만을 입력으로 사용하여 수행되었으며, 텍스트 자원이 없으므로 이미지 인코더를 사용했다. 이 이미지 인코더의 성능은 CheXNet이 정확도 0.45로 낮은 성능을 보였지만 다른 인코더보다 약 40% 이상의 높은 정확도를 보였기 때문에, 입력 2와 3에서는 이미지 인코더를 CheXNet으로 고정하여 사용한다. 입력 2에서는 이미지와 리포트를 모두 사용한 생성적 주석을 수행했다. 추가된 리포트를 인코딩하기 위해서 텍스트 인코더가 GPT 버전에 따라 나뉘어 실험되었다. 그 결과 GPT 3.5가 0.51로 가장 높은 정확도를 보였으며, GPT 3과는 약 7%의 성능 차이를 보였다. 이미지만을 사용한 입력 1에 대해서는 약 8%의 향상된 성능을 보였다. 마지막 입력 3에서는 이미지와 리포트, 1~4의 질문에 대한 응답을 입력으로 사용했다. 응답의 형태는 주관식에 해당하는 문장과 객관식 선택형으로 구성된다. 이 실험에서는 선택형 분류를 One-Hot 인코딩을 사용하지 않고 텍스트 인코더에 리포트와 함께 입력했다. 그 결과 GPT 3.5에서 0.63으로 가장 높은 성능을 보였다. 이는 입력 2와 약 11%의 성능 차이를 보이며, 진단의 결과 (응답 1~4)가 진단의 원인 분류에 영향을 끼치는 것으로 판단된다.

## Reference

1. You, Q., Jin, H., Wang, Z., Fang, C., & Luo, J. (2016). Image captioning with semantic attention. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4651-4659).
2. Kanimozhiselvi, C. S. et al. "Image Captioning Using Deep Learning." 2022 International Conference on Computer Communication and Informatics (ICCCI) (2022): 1-7.
3. Wang S, Tang L, Lin M, Shih G, Ding Y, Peng Y. Prior Knowledge Enhances Radiology Report Generation. AMIA Jt Summits Transl Sci Proc. 2022 May 23;2022:486-495. PMID: 35854760; PMCID: PMC9285179.
4. Pahadia, M., Khurana, S., Geha, H., & Deahl, S. T. I. (2020). Radiology report writing skills: A linguistic and technical guide for early-career oral and maxillofacial radiologists. *Imaging Science in Dentistry*, 50(3), 269.
5. Jing, B., Xie, P., & Xing, E. (2017). On the automatic generation of medical imaging reports. arXiv preprint arXiv:1711.08195.
6. Chen, Z., Song, Y., Chang, T. H., & Wan, X. (2020). Generating radiology reports via memory-driven transformer. arXiv preprint arXiv:2010.16056.
7. Yang, S., Wu, X., Ge, S., Zheng, Z., Zhou, S. K., & Xiao, L. (2023). Radiology report generation with a learned knowledge base and multi-modal alignment. *Medical Image Analysis*, 86, 102798.
8. Bustos, A., Pertusa, A., Salinas, J. M., & De La Iglesia-Vaya, M. (2020). Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66, 101797.
9. Johnson, A. E., Pollard, T. J., Greenbaum, N. R., Lungren, M. P., Deng, C. Y., Peng, Y., ... & Horng, S. (2019). MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. arXiv preprint arXiv:1901.07042.

10. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597).
11. Cuticchia, A. J., Cooley, P. C., Hall, R. D., & Qin, Y. (2006). NIDDK data repository: a central collection of clinical trial data. *BMC medical informatics and decision making*, 6, 1-6.
12. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).
13. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
14. Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., ... & Huang, X. (2023). A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv preprint arXiv:2303.10420*.
15. Jones J, Knipe H, Yadegarfar M, et al. Chest x-ray (summary). Reference article, Radiopaedia.org (Accessed on 06 Sep 2023) <https://doi.org/10.53347/rID-42801>.
16. Breatnach, E., Abbott, G. C., & Fraser, R. G. (1984). Dimensions of the normal human trachea. *American Journal of Roentgenology*, 142(5), 903-906.
17. Haskin, P. H., & Goodman, L. R. (1982). Normal tracheal bifurcation angle: a reassessment. *American Journal of Roentgenology*, 139(5), 879-882.
18. Ludeña, T., Lozano-Samaniego, A., Maldonado, S., & Salas, F. (2022). El hilio pulmonar, dos signos radiológicos clásicos para descifrarlo. *Radiología*, 64(1), 60-68.

19. Ngo D, Slinger C, Chieng R, et al. Lung hilum. Reference article, Radiopaedia.org (Accessed on 06 Sep 2023) <https://doi.org/10.53347/rID-43948>.
20. Magera S, Sereke SG, Okello E, Ameda F, Erem G. Aortic Knob Diameter in Chest Radiographs of Healthy Adults in Uganda. Reports in Medical Imaging. 2022;15:21-29.
21. Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., ... & Huang, X. (2023). A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. arXiv preprint arXiv:2303.10420.
22. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.
23. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
24. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
25. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
26. Jones J, Knipe H, Yadegarfar M, et al. Chest x-ray (summary). Reference article, Radiopaedia.org (Accessed on 06 Sep 2023) <https://doi.org/10.53347/rID-42801>.