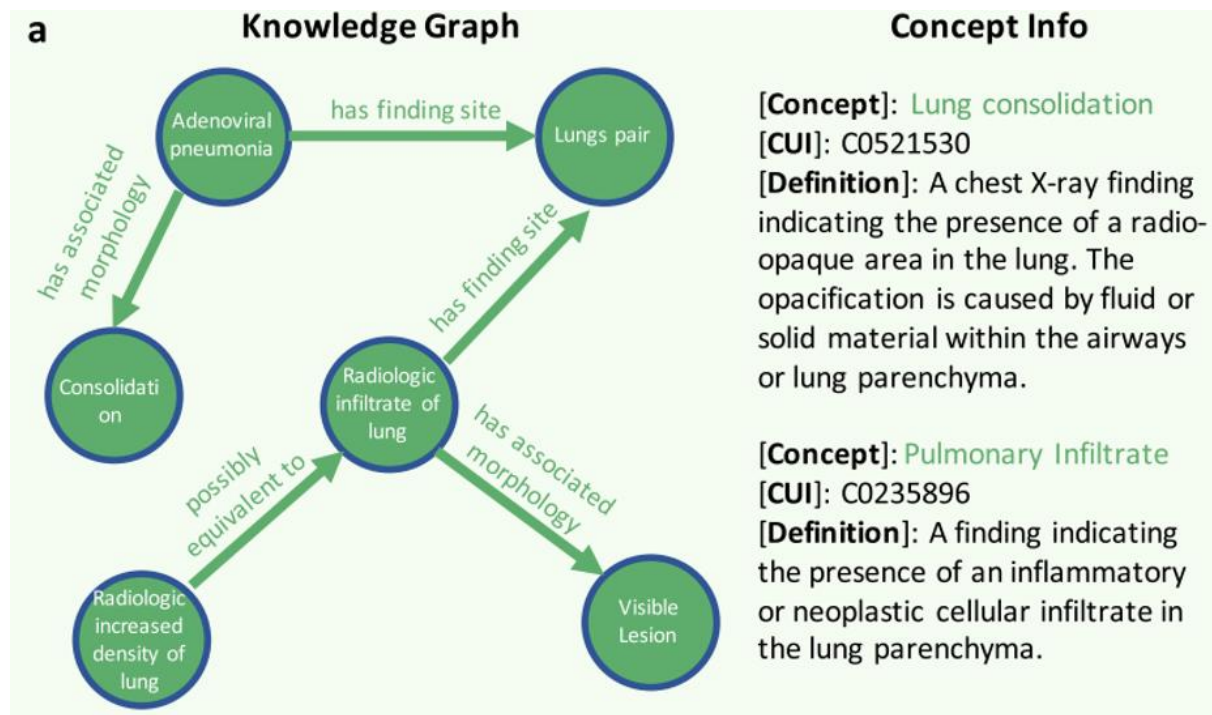


Introduction

- 연구 배경
 - 최근 대규모 데이터에서 사전학습된 비전-언어 모델(BERT, GPT, CLIP 등)은 자연어 처리와 컴퓨터 비전에서 성공적인 성과를 거두었지만, 의료 도메인에서는 세밀한 진단과 전문 지식의 필요성으로 인해 발전이 더딘 상황.
 - 기존의 비전-언어 모델은 의료 영상과 텍스트 데이터를 단순히 정렬시키는 방식으로 동작했으나, 이러한 접근법은 훈련 중 보지 못한 질병이나 방사선 소견에 대한 일반화가 어려움.
- 연구 목표
 - 본 논문은 흉부 X-선 영상과 방사선 보고서를 활용하여 의료 도메인 지식을 통합한 Knowledge-enhanced Auto Diagnosis (KAD) 모델을 제안.
 - KAD는 기존의 단순한 이미지-텍스트 정렬 방식을 넘어, 의료 지식 그래프(Unified Medical Language System, UMLS)와 임상 엔티티 추출 기술을 활용하여 의료 도메인 지식을 시각적 표현 학습에 적용.

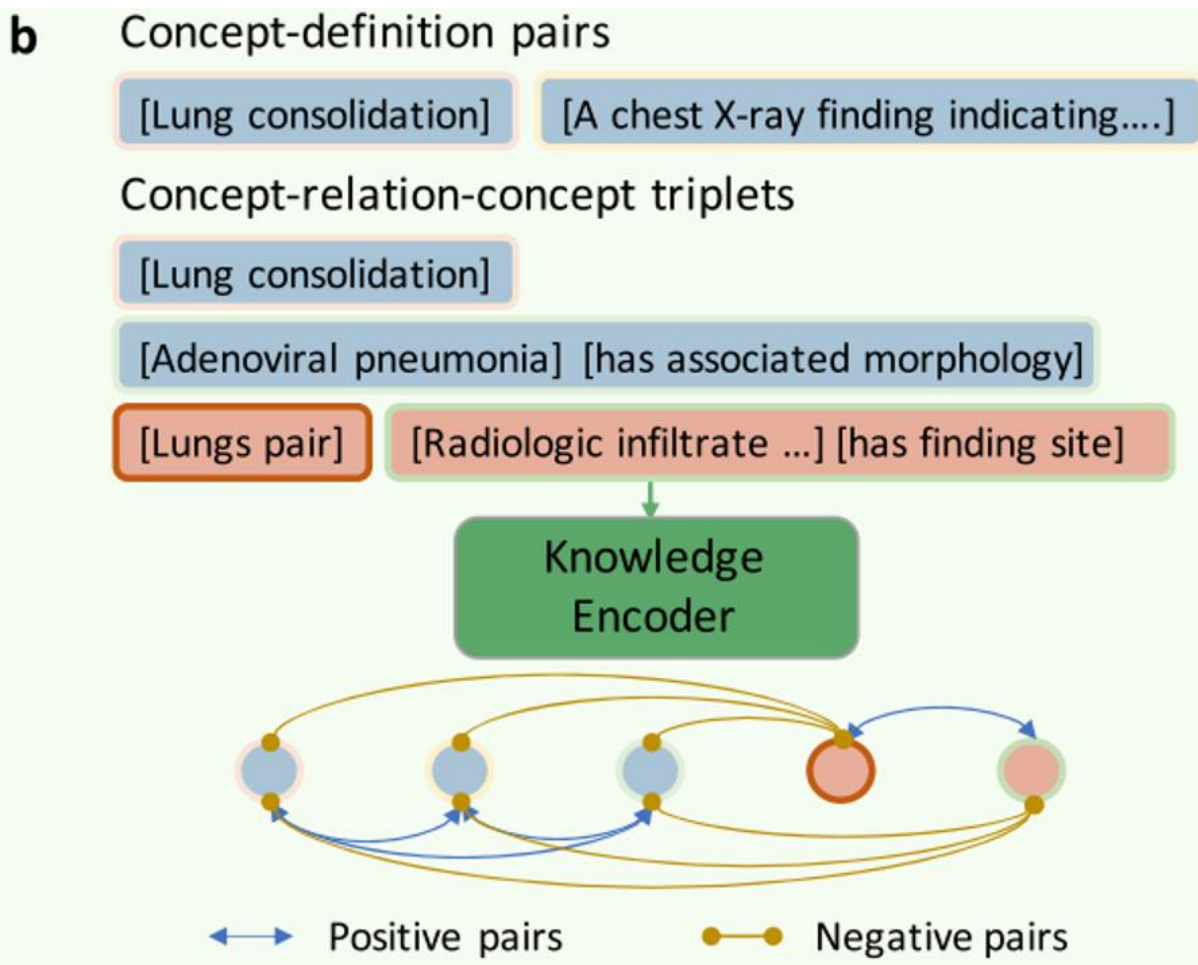
Methods



Knowledge Base

- a. Knowledge Base

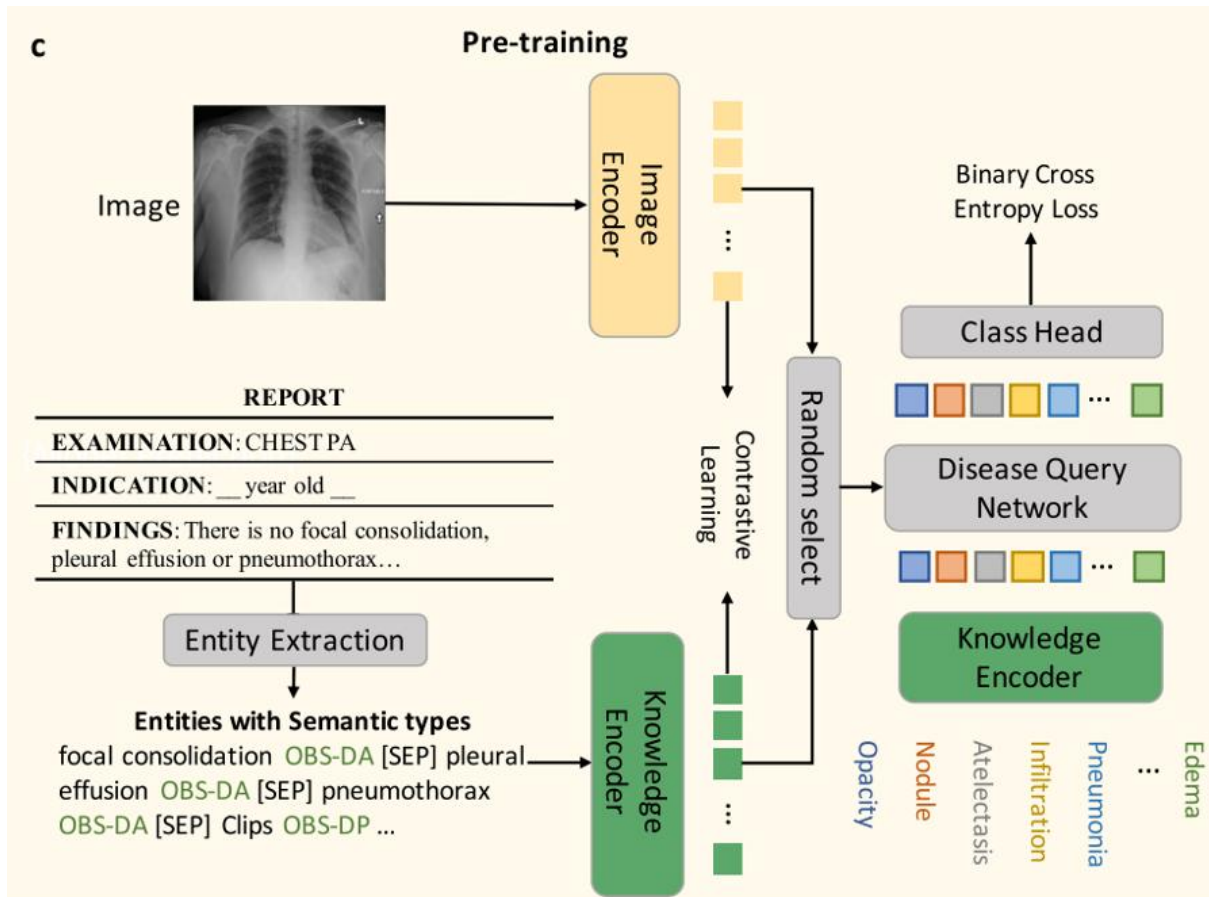
- UMLS을 기반으로 의료 개념(entities)과 관계(edge)를 학습.
- 의료 개념은 고유 식별자(CUI), 정의, 동의어 등을 포함하며, 관계는 "triplet" 형식(예: [개념1, 관계, 개념2])으로 표현.
- Knowledge Graph: 개념-관계-개념(triplets)으로 구성된 그래프. 예를 들어, "폐렴"과 "폐" 사이의 관계를 정의.
- Concept Info List: 각 개념에 대한 정의와 동의어를 포함한 정보 목록.
- 의료 도메인에서 사용되는 다양한 개념과 이들 간의 관계를 구조화하여 다단계 추론을 가능하게 함.



Knowledge Encoder

- b. Knowledge Encoder
 - 지식 그래프에서 추출된 텍스트 데이터를 사용하여 텍스트 표현 학습.
 - PubMedBERT를 기반

- Contrastive learning을 통해 동일한 CUI를 가진 텍스트 표현 간 유사성을 최대화하도록 학습.
- 의료 도메인 지식을 효과적으로 학습하여 이후 이미지-텍스트 학습에 활용.

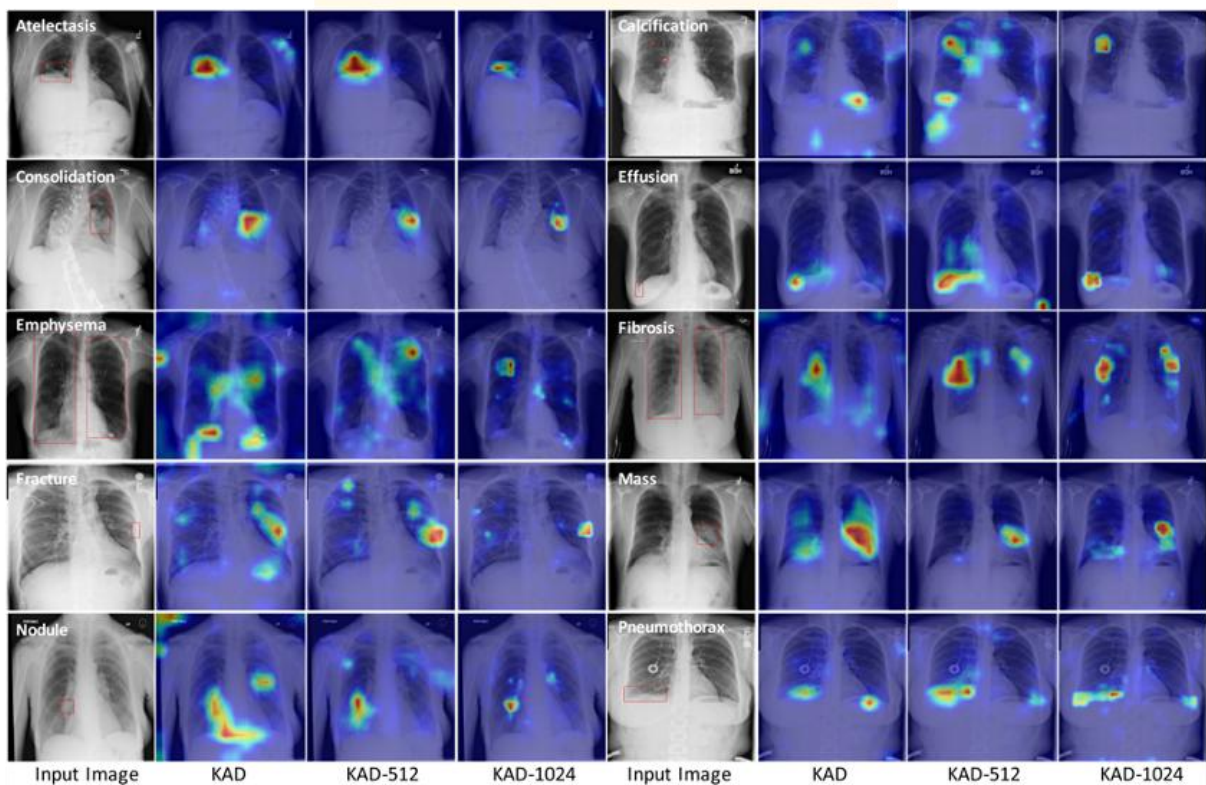
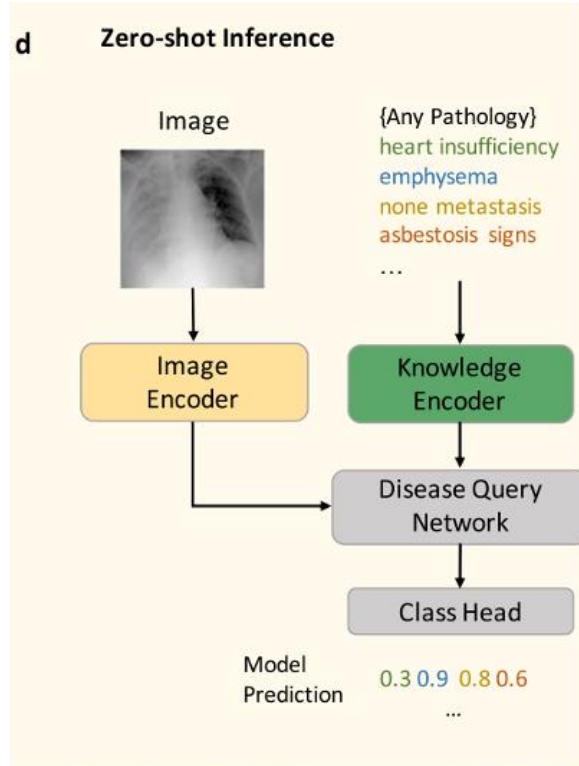


- c. Image-Text Contrastive Learning
 - Image Encoder:
 - X-선 이미지를 입력받아 특징을 추출.
 - ResNet-50 또는 Vision Transformer(ViT) 백본 사용.
 - Entity Extraction:
 - 방사선 보고서에서 임상 엔티티와 관계를 추출.
 - 세 가지 방법 사용:
 - Heuristic rules: 규칙 기반 접근으로 UMLS와 spaCy 패키지를 활용.

- RadGraph: 방사선 보고서에서 엔티티와 관계를 추출하는 사전 훈련된 모델.
- ChatGPT: 자연어 처리 모델을 이용해 보고서를 분석하고 엔티티를 추출.
- 사전 학습된 지식 인코더를 사용하여 X-선 이미지와 추출된 엔티티 간의 contrastive learning
- Disease Query Network, DQN
 - Transformer 기반 구조로 설계.
 - 질병 이름을 쿼리로 입력받아 이미지 특징과 상호작용하며 병리학적 소견의 존재 가능성을 예측.
 - Attention 맵을 생성하여 모델의 예측 근거를 시각적으로 제공.



- 학습 과정
 - 1단계: 지식 인코더 학습
 - UMLS 지식 그래프에서 개념 정의 및 관계를 사용하여 contrastive learning 수행.
 - 동일한 CUI를 가진 텍스트 표현 간 유사성을 최대화하여 의료 도메인 지식을 학습.
 - 2단계: 비전-언어 표현 학습
 - X-선 이미지와 방사선 보고서 쌍을 사용하여 cross-modal contrastive learning 수행.
 - DQN을 통해 질병 쿼리와 이미지 특징 간 상호작용 학습.
- 손실 함수:
 - Contrastive loss (L_{contrast}): 이미지와 텍스트 간 유사성을 높이기 위한 손실.
 - Binary cross-entropy loss (L_{dqn}): DQN의 질병 존재 여부 예측 성능을 최적화하기 위한 손실.
 - 총 손실: $L_{\text{KAD}} = L_{\text{contrast}} + L_{\text{dqn}}$.



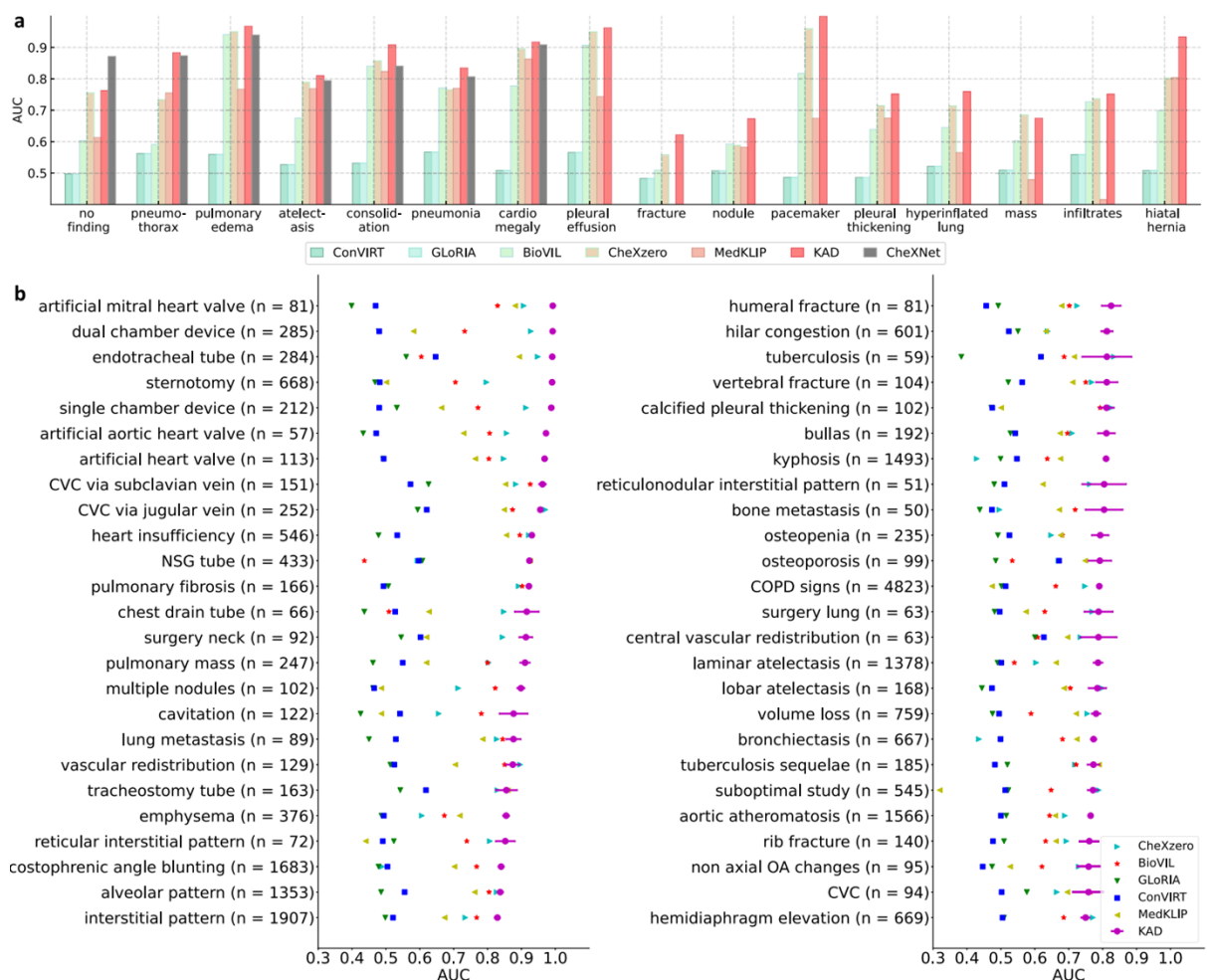
d. Inference Stage

- 질병 이름을 쿼리로 입력하면 DQN이 입력 이미지에서 해당 병리의 존재 가능성을 출력.
- Attention map을 통해 모델이 예측한 근거를 시각적으로 제공하여 설명 가능성을

높임.

- 원본 이미지 : 빨간 박스는 방사선 전문의가 표시한 병변 영역.
- 어텐션 맵 : 빨간색에서 파란색까지의 스펙트럼이 원본 이미지 위에 표시.
 - 빨간색 : 높은 주목도
 - 파란색 : 낮은 주목도

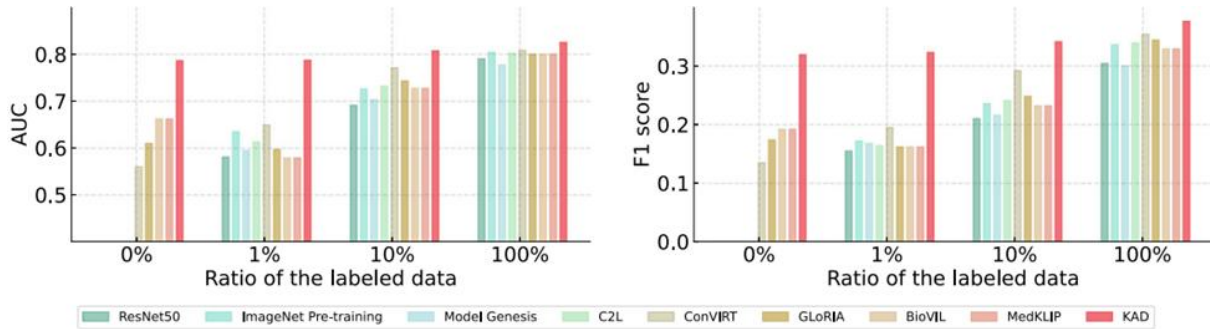
Results



KAD 모델의 PadChest 데이터셋에서의 zero-shot 비교

- a. Seen Classes (훈련 중 관찰된 클래스)
 - CheXNet(완전 지도 학습 모델)과 비교했을 때, KAD는 일부 병리에서 더 나은 성능을 보임.
- b. Unseen Classes (훈련 중 관찰되지 않은 클래스)

- 테스트셋에서 총 177개의 unseen 클래스 중:
 - 31개 클래스에서 $AUC \geq 0.900$.
 - 111개 클래스에서 $AUC \geq 0.700$.
- 특히, 드문 병리(long-tail distribution)에서도 강력한 일반화 성능을 보임.
 - long-tail 분포: 대부분의 클래스가 적은 샘플 수를 가짐.



KAD 모델의 ChestXray14 데이터셋에서의 fine-tuning 비교

- Fine-tuning에서는 라벨 데이터의 비율(1%, 10%, 100%)을 다양하게 조정하여 모델의 데이터 효율성을 분석.
- AUC(Area Under the Curve), F1 점수 등 다중 병리학적 소견에 대한 매크로 평균 (metric macro average)을 사용.
- 특히, 라벨 데이터의 1%만 사용했을 때도 기존 SOTA 모델보다 월등히 높은 AUC 와 F1 점수를 달성.

Ablation study

Methods	ChestXray14				CheXpert			
	AUC	MCC	F1	ACC	AUC	MCC	F1	ACC
Ablation on proposed modules								
KAD w/o Stage1	0.752	0.228	0.274	0.748	0.894	0.546	0.620	0.858
KAD w/o random select	0.751	0.242	0.290	0.780	0.878	0.571	0.671	0.812
KAD w/o DQN	0.672	0.144	0.109	0.747	0.822	0.419	0.508	0.806
Ablation on entity extraction module								
w/ UMLS	0.773	0.268	0.308	0.833	0.904	0.562	0.635	0.867
w/ ChatGPT	0.784	0.284	0.336	0.845	0.887	0.573	0.622	0.888
w/ RadGraph (KAD)	0.789	0.280	0.323	0.816	0.905	0.589	0.647	0.875

KAD 모델의 모듈별 분석, 이미지 인코더 및 해상도에 따른 분석

- 주요 실험 구성:
 - KAD w/o Stage1: Stage 1(지식 그래프 기반 텍스트 인코더 학습)을 생략하고 PubMedBERT를 그대로 사용.

- KAD w/o random select: Disease Query Network(DQN) 학습 시 이미지 특징만을 key와 value로 사용.
- KAD w/o DQN: DQN 없이 cross-modal contrastive learning만 수행.
- Entity Extraction Module:
 - w/ UMLS: RadGraph 대신 UMLS 기반 휴리스틱 규칙 사용.
 - w/ ChatGPT: ChatGPT를 사용하여 엔티티 추출.
 - w/ RadGraph: RadGraph를 사용하여 엔티티 추출(KAD 기본 설정).
- Stage1과 DQN은 KAD 모델의 성능에 가장 중요한 모듈로 확인됨.
- 엔티티 추출 도구(RadGraph, ChatGPT, UMLS)는 모두 경쟁력 있는 결과를 제공하지만, RadGraph가 가장 높은 성능을 보임.