

2023.05.15

권정을, 권혁민, 이지원, 윤수아



# Multimodal Multi-class Classification Using radiology reports and Chest radiographs: Unimodal과 Multimodal model의 성능 비교 연구

명지대학교 산업경영공학과

FOM 1기 권혁민



권정을: [rnjswjddmf@gmail.com](mailto:rnjswjddmf@gmail.com)  
권혁민: [prun000111@gmail.com](mailto:prun000111@gmail.com)  
이지원: [intelli2000e@gmail.com](mailto:intelli2000e@gmail.com)  
윤수아: [dbstndk456@mju.ac.kr](mailto:dbstndk456@mju.ac.kr)



# INDEX

1. 팀 소개
2. 프로젝트 주제 소개
3. 데이터 소개
4. 모델 소개
5. 실험방법
6. 논문 소개

# 01 | 팀 소개

권정을, 권혁민, 이지원, 윤수아

- Team Name:  
**흉부마법사들 (MOC : magician of chest)**



노예 권정을  
@hanmadeoutlier



19학번 권혁민  
운영위원장이 되고 싶은 남자  
(뽀아 달라는 뜻)



19학번 이지원  
@\_zzzonezzzz



20학번 윤수아  
@sua\_Ong



# 02 | 프로젝트 주제 소개

## 주제 선정 이유



노예 권정을

- 중환자실(ICU)에 입원하는 환자들을 위한 임상 의사결정 지원 시스템(CDSS) 관련 연구에 관심이 있음
- 멀티모달 머신러닝에 관심이 있음



19학번 권혁민

- Image Classification과 추천 알고리즘을 활용한 App 개발에 관심이 있음
- 자연어처리와 컴퓨터비전을 결합하는 Visual Question Answering 시스템에 관심이 있음



19학번 이지원

- Object Detection 기반의 제조현장 안전관리에 관심이 있음
- Image를 통해 산업 재해 위험을 평가하는 것에 관심이 있음



20학번 윤수아

- 의료, 헬스케어 등 사람을 위한 인공지능과 반려동물 의료 산업에 관심이 있음
- Image 데이터 처리 경험을 쌓고 싶음

Research of interest is  
**“Computer Vision”**

시각적 세계를 해석하고 이해하도록 컴퓨터를 학습시키는 인공지능 분야  
이미지, 영상 → 딥러닝

데이터 소싱, 모델링 난이도를 고려하여 모델이 이미 존재하며, 데이터 수집이 용이한 “medical domain”을 최종 선정

# 02 | 프로젝트 주제 소개

## 배경

- **CDSS(임상 의사 결정 지원 시스템)**

- ✓ 의료 전문가들이 환자를 진단하고 치료하는 과정에서 의사 결정에 도움을 주는 컴퓨터 프로그램
- ✓ 단순 진단과는 다른 개념으로, 데이터 과학자가 의사를 도울 수 있는 하나의 인터페이스

- **CXR-CV(Chest X-ray computer vision)**

- ✓ 폐렴, 암, 뇌 COVID-19
- ✓ 흉부 X-ray에서 볼 수 있는 특정한 패턴으로 진단이 가능
- ✓ 데이터 과학으로 높은 성능을 낼 수 있는 분야 중 하나



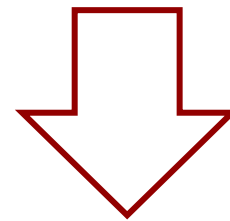
〈인공지능 개발 시스템의 작동예〉

# 02 | 프로젝트 주제 소개

## 기존의 CDSS의 문제점

### • 기존 CDSS(영상 의사 결정 지원 시스템)의 문제점

- ✓ 대부분 단일 데이터 양식(예: 환자의 X-RAY image)에 초점을 맞춤
- ✓ Complementation information(예: 환자의 임상 정보)를 고려하지 않음



- 여러 단서들(환자의 과거력, 촬영 요청 이유 등 )을 포함해서 훈련한다면 더욱 좋은 성과를 보일 것임
- 시각 및 언어 모델을 설계하면 **고효율의 질병분류**가 가능할 것임

# 02 | 프로젝트 주제 소개

## 폐 질환 진단 process

### • 폐 질환 진단 process

1. 중환자실에 환자가 입원한다.
2. 의료진이 환자를 관리한다.
3. 폐 질환이 의심될 경우 indication(촬영 요청 이유)을 작성하여 방사선과에 보낸다.
4. 방사선의사는 환자의 chest를 찍고 보고서(radiology reports, semi-structured)를 작성한다.
5. Foundation, impression section에는 관찰에 대한 내용을 기술한다.
6. 의료진이 보고서를 보고 진단을 내린다.

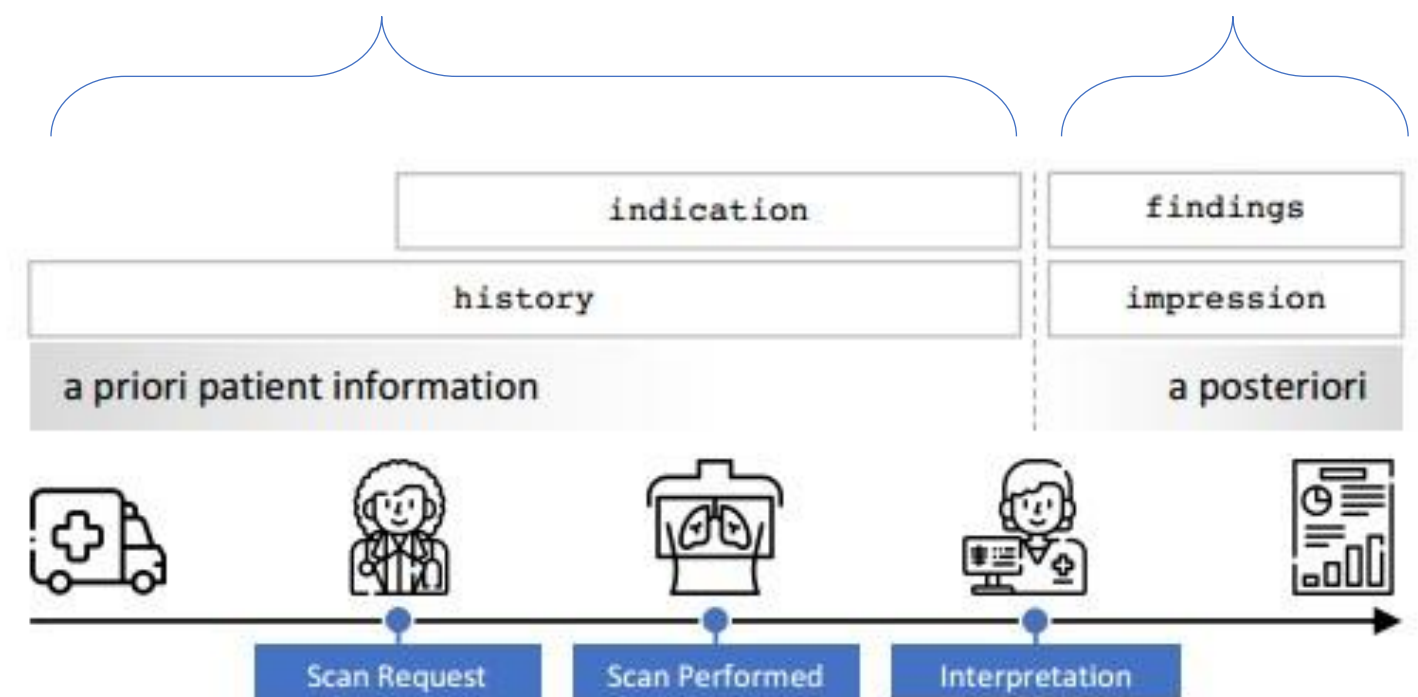
#### radiology reports

FINAL REPORT  
EXAMINATION: CHEST (PA AND LAT)  
INDICATION: \_\_\_M with hypoxia // ?pna, aspiration. ] Clinician의 indication  
COMPARISON: None  
FINDINGS:  
PA and lateral views of the chest provided. The lungs are adequately aerated.  
There is a focal consolidation at the left lung base adjacent to the lateral hemidiaphragm. There is mild vascular engorgement. There is bilateral apical pleural thickening.  
The cardiomeastinal silhouette is remarkable for aortic arch calcifications. The heart is top normal in size.  
IMPRESSION:  
Focal consolidation at the left lung base, possibly representing aspiration or pneumonia.  
Central vascular engorgement.

Radiologist의 진찰

의사 개입 X, 진단 전, 낮은 수집 비용  
→ 응용에 현실적

의사 개입 O, 진단 후, 높은 수집 비용  
→ 응용에 비현실적





# 02 | 프로젝트 주제 소개

## Multimodal

### • Multimodal(다중양식)

- ✓ 여러가지 형태의 정보를 통해 소통하는 환경
- ✓ Text, image, video, speech, etc..
- ✓ 인간의 "5가지 감각 기관"과 같이 다양한 정보를 처리하고 연결하는 모델을 만들어 세계를 이해하는 방법

### • Multimodal Learning

- ✓ 다양한 데이터 형태를 결합하여 더욱 정확하고 포괄적인 결과를 도출하는 머신 러닝 방법

Image and Trace:



Caption:

In the front portion of the picture we can see a dried grass area with dried twigs. There is a woman standing wearing light blue jeans and ash colour long sleeve length shirt. This woman is holding a black jacket in her hand. On the other hand she is holding a balloon which is peach in colour. On the top of the picture we see a clear blue sky with clouds. The hair colour of the woman is brownish.





# 02 | 프로젝트 주제 소개

## 주제

### • 소아/청소년의 흉부 X-ray를 활용한 폐 질환 분류 모델 개발 및 modal(데이터 유형)별 분류 성능 비교

- ✓ 가설: *Multimodal*(다중 양식) 데이터를 활용하면 분류 성능이 향상될 것이다.
- ✓ Unimodal(image or text) 기반 모델 성능과 Multimodal(image+text) 기반 모델 성능을 비교하고
- ✓ Text 보다는 Image, Image 보다는 Text+Image가 더욱 풍부한 정보를 표현할 수 있음을 증명
- ✓ 따라서 아래 **Result table**을 만드는 것이 본 프로젝트의 목적

Unimodal machine learning

Multimodal machine learning

Method	Modality		Accuracy	Precision		Recall		$F_1$		AUROC	
	Image	Text		Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro
CheXpert Labeler	✗	✓	80.6	9.3	13.4	18.8	27.0	8.5	17.9	51.13	53.3
BERT	✗	✓	85.1 $\pm$ 0.2	21.6 $\pm$ 1.0	32.9 $\pm$ 0.9	47.2 $\pm$ 5.8	54.7 $\pm$ 0.9	26.1 $\pm$ 1.1	41.1 $\pm$ 0.9	71.1 $\pm$ 1.2	81.7 $\pm$ 0.8
ResNet-50	✓	✗	86.0 $\pm$ 0.2	26.0 $\pm$ 1.1	43.7 $\pm$ 2.1	34.0 $\pm$ 1.8	57.4 $\pm$ 1.5	27.4 $\pm$ 0.4	49.5 $\pm$ 0.8	73.8 $\pm$ 0.5	84.4 $\pm$ 0.6
Attentive	✓	✓	86.8 $\pm$ 0.1	26.8 $\pm$ 0.5	44.2 $\pm$ 0.7	34.7 $\pm$ 0.2	61.3 $\pm$ 0.6	29.1 $\pm$ 0.4	51.4 $\pm$ 0.3	76.6 $\pm$ 0.3	86.0 $\pm$ 0.2
MMBT	✓	✓	87.7 $\pm$ 0.2	30.8 $\pm$ 0.3	47.8 $\pm$ 0.7	55.4 $\pm$ 1.8	64.7 $\pm$ 0.7	35.0 $\pm$ 0.6	55.0 $\pm$ 0.6	80.6 $\pm$ 0.1	87.8 $\pm$ 0.1

〈Result table: 학습에 사용한 modality에 따른 분류 성능〉

# 03 | 데이터 소개

## 데이터 구조

- 데이터 출처 : 고려대학교 안암 병원 외 8개 병원, AIHub 플랫폼에서 무료 제공

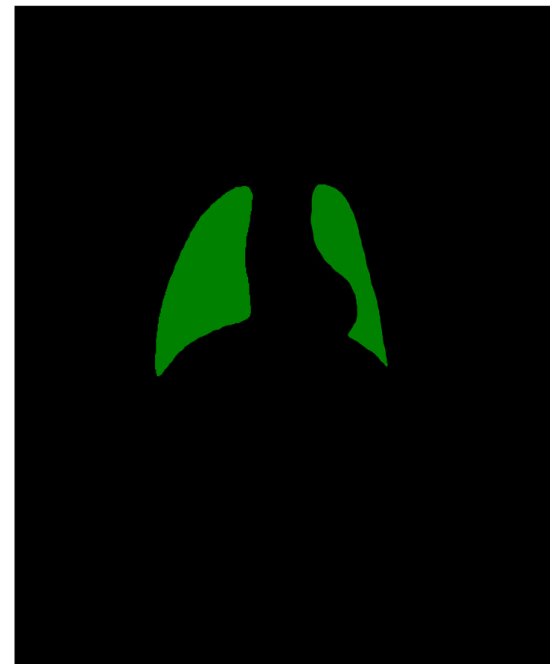


- 데이터 유형 : Image, Tabular(데이터프레임 형식의 부가 정보)

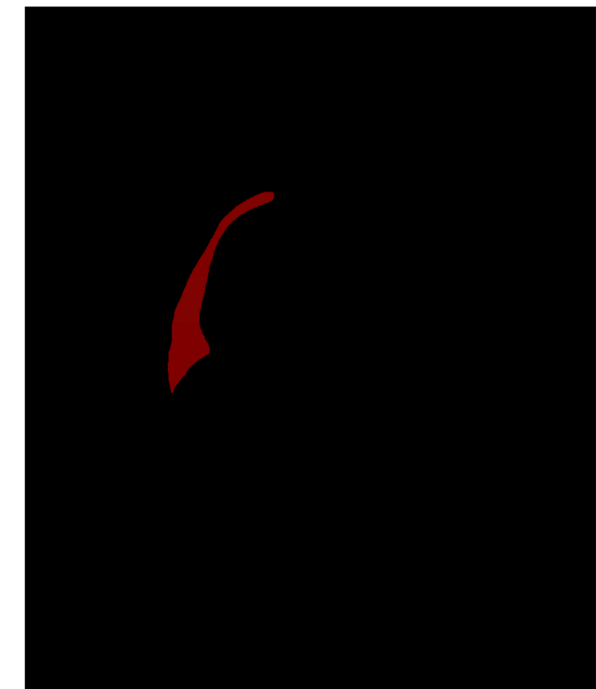
- ✓ 원본 데이터 : 비식별화 처리된 소아 흉부 X-ray 영상(Dicom형식)
- ✓ 라벨링 데이터 : 폐영역과 병변 영역이 Annotation 처리된 영상(PNG형식), 환자 나이, 몸무게, 질환종류 등의 임상정보 포함 메타데이터(JSON형식)



원본데이터(dicom)



폐영역 세그멘테이션(png)



병변영역 세그멘테이션(png)

# 03 | 데이터 소개

## 데이터 통계

### • 데이터 구축 규모

		확보건수
정상		6,767
폐렴(바이러스성, 세균성)		8,346
기타 폐질환		6,104
	공기누출(기흉, 종격동기흉)	1,161
	과다팽창	468
	무기폐	845
	신생아 호흡곤란증후군	2,351
	흉막삼출	1,279
총합계		21,217

### • 나이대 별, 질환 별 구축 규모 및 분포

연령대	질환종류		환자명수		영상건수	
0-2개월	Normal	정상	4,976	1,056		1,180
	RDS	신생아호흡곤란증후군		2,069		2,351
	Pnemonia (viral+bacteria)	폐렴 (바이러스성, 세균성)		1,293		1,584
	Air leak	공기누출 (기흉, 종격동기흉)		558		772
3개월-5세	Normal	정상	5,618	2,255		2,514
	Pnemonia (viral+bacteria)	폐렴 (바이러스성, 세균성)		2,285		2,706
	Hyperinflation	과다팽창		446		468
	Pleural effusion	흉막삼출		292		364
	Atelectasis	무기폐		340		403
6-10세	Normal	정상	3,726	1,207		1,358
	Pnemonia (viral+bacteria)	폐렴 (바이러스성, 세균성)		1,825		2,338
	Pleural effusion	흉막삼출		364		480
	Atelectasis	무기폐		330		442
11-15세	Normal	정상	3,416	1,433		1,715
	Pnemonia (viral+bacteria)	폐렴 (바이러스성, 세균성)		1,360		1,718
	Pleural effusion	흉막삼출		301		435
	Air leak	공기누출 (기흉, 종격동기흉)		322		389
총합계			17,736		21,217	




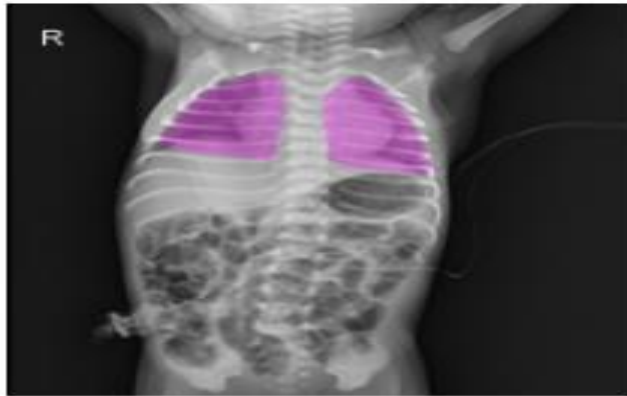




# 03 | 데이터 소개

## 데이터셋 – IMAGE(chest radiograph)

### • 데이터셋의 종류 및 형식

- ✓ 인공지능학습을 위한 소아흉부질환의 데이터셋은 X-ray 이미지와 폐/병변 부위를 Annotation한 labeled video(Polygon Segmentation) 및 임상정보 메타데이터를 포함한 labeled dataset으로 구성됨.

### • 질환 별 라벨링 예시

		
폐렴	RDS	공기누출
		
무기폐	과다팽창	흉막삼출

# 03 | 데이터 소개

## 데이터셋 – TEXT(radiology reports)

### • Patient's Demographics Information(in Tabular format)

- ✓ 데이터셋에는 환자의 인구통계학적 정보가 기록되어 있음(age\_group, age, sex, height, weight 등)
- ✓ 본 테이블에 수집되어 있는 “**report**” 정보를 활용할 수 있음
- ✓ “**report**”는 clinician(의료진) 또는 radiologist(방사선의)의 소견이 작성되어 있는 경우가 있으나, 수집되는 과정에서 필수 제약이 보장되어 있지 않기 때문에 모든 환자가 report를 가지고 있는지는 알 수 없음 → 탐색적 분석(EDA) 진행 후 판단할 필요가 있음

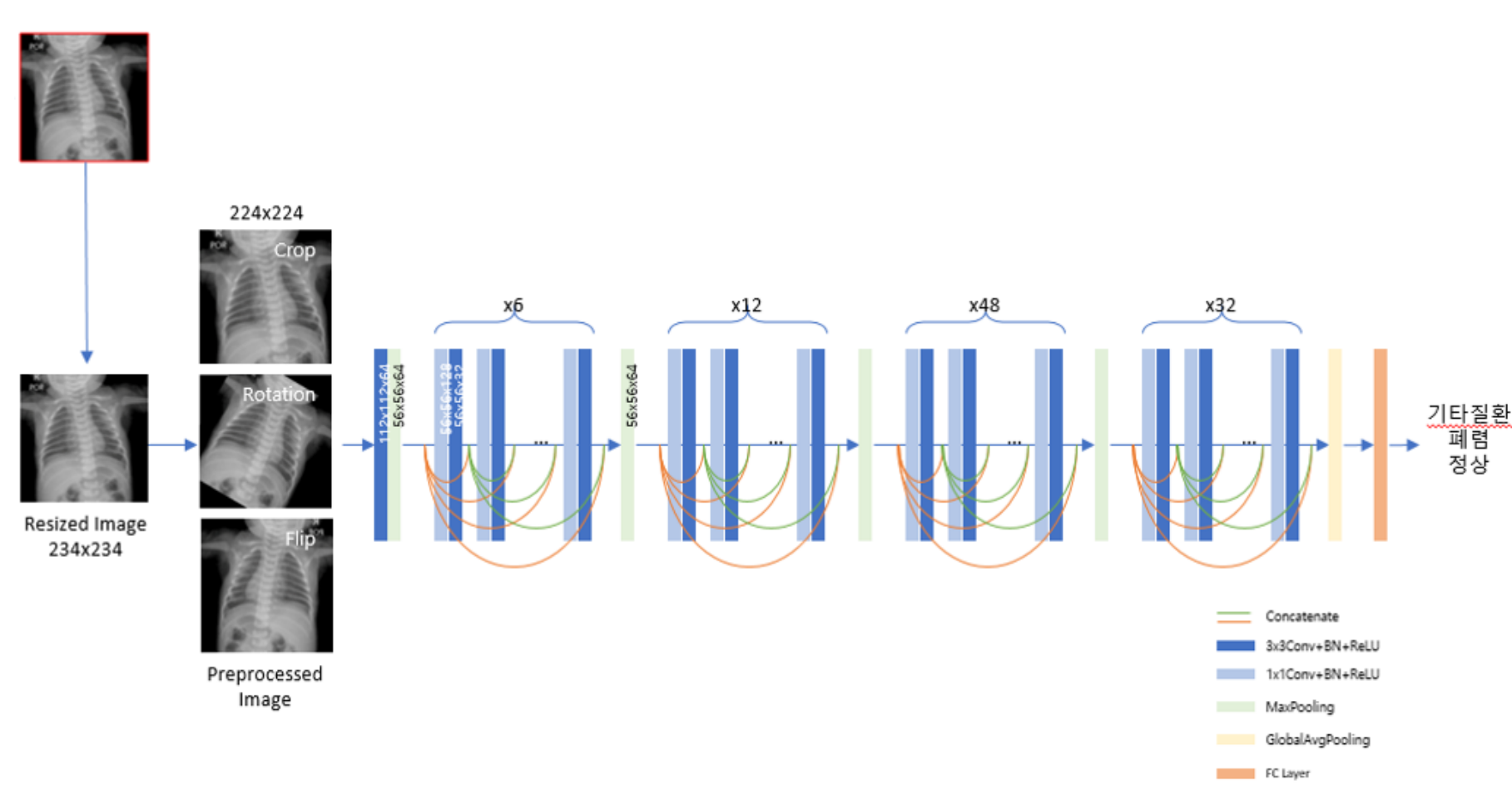
patient(환자정보)					
7	age_group	연령대	Y	string	[표] 참조
8	diagnosis	진단명	Y	string	[표] 참조
9	pneumonia_type	폐렴유형		string	1:바이러스성, 2:세균성,
					3:알수없음, 4:바이러스성+세균성
10	report	진단내역		string	
11	sex	성별	Y	string	M.남, F.여
12	age	나이		string	1세미만: 0~11개월
					1세이상: 1세~15세
13	height	키		number	cm
14	weight	몸무게		number	kg

# 04 | 모델 소개

## DenseNet

### • 활용 모델: DenseNet121

- ✓ Image(chest radiograph) modal 학습에 사용할 모델
- ✓ 121개의 레이어로 이루어진 DenseNet 모델 중 하나로, ImageNet 데이터셋으로 학습된 모델



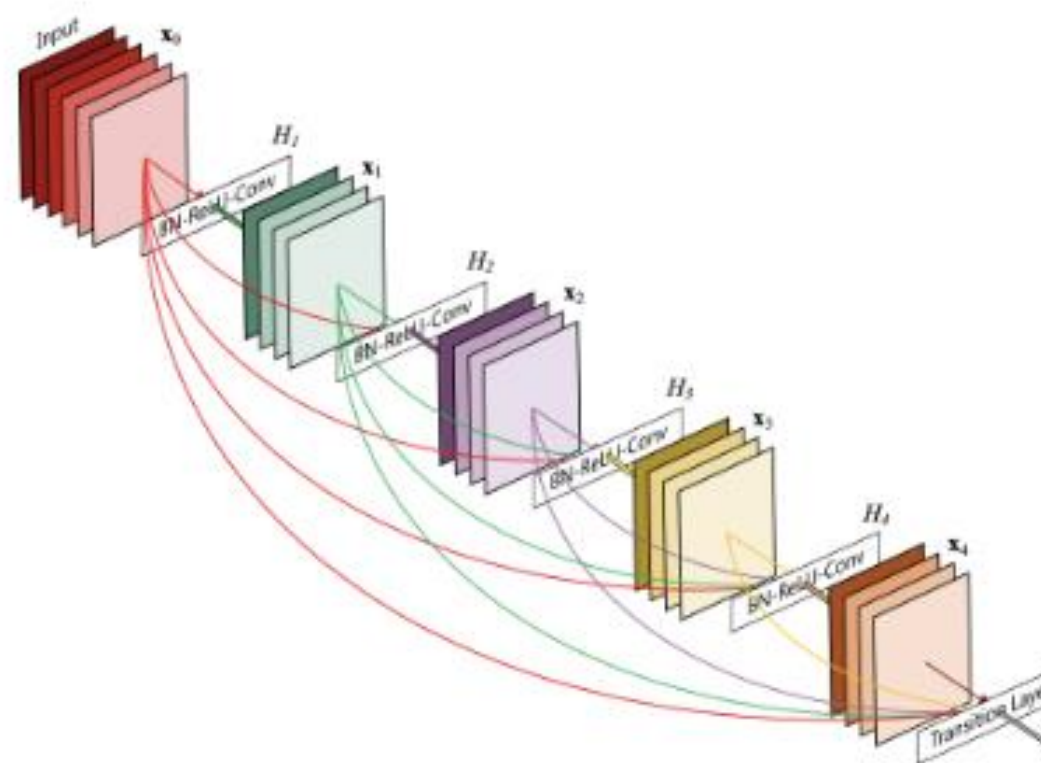


# 04 | 모델 소개

## DenseNet

### • Densely Connected Convolutional Networks

- ✓ 네트워크 각 레이어의 출력이 이전 레이어의 출력과 결합되는 방식으로 이루어지는 구조
- ✓ feature map들을 concatenate(연결)하는 방식 사용
- ✓ feature map들이 전체 네트워크에 연결(처음 레이어의 feature map을 마지막 레이어의 feature map까지 연결)
- ✓ 정보가 소실되는 것을 방지
- ✓ 적은 수의 파라미터로 깊은 모델을 만들 수 있음



# 04 | 모델 소개

## Data Augmentation

### • Data Augmentation

- ✓ 머신러닝 모델의 성능을 향상시키기 위해 데이터를 다양하게 변형시키는 기법
- ✓ 데이터의 다양성을 학습(과적합을 방지, 일반화 성능 향상)
- ✓ “report”가 있는 환자를 필터링 하면, 샘플 수가 감소할 것으로 예상되기 때문에 데이터 증강기법은 필수 과정으로 보임

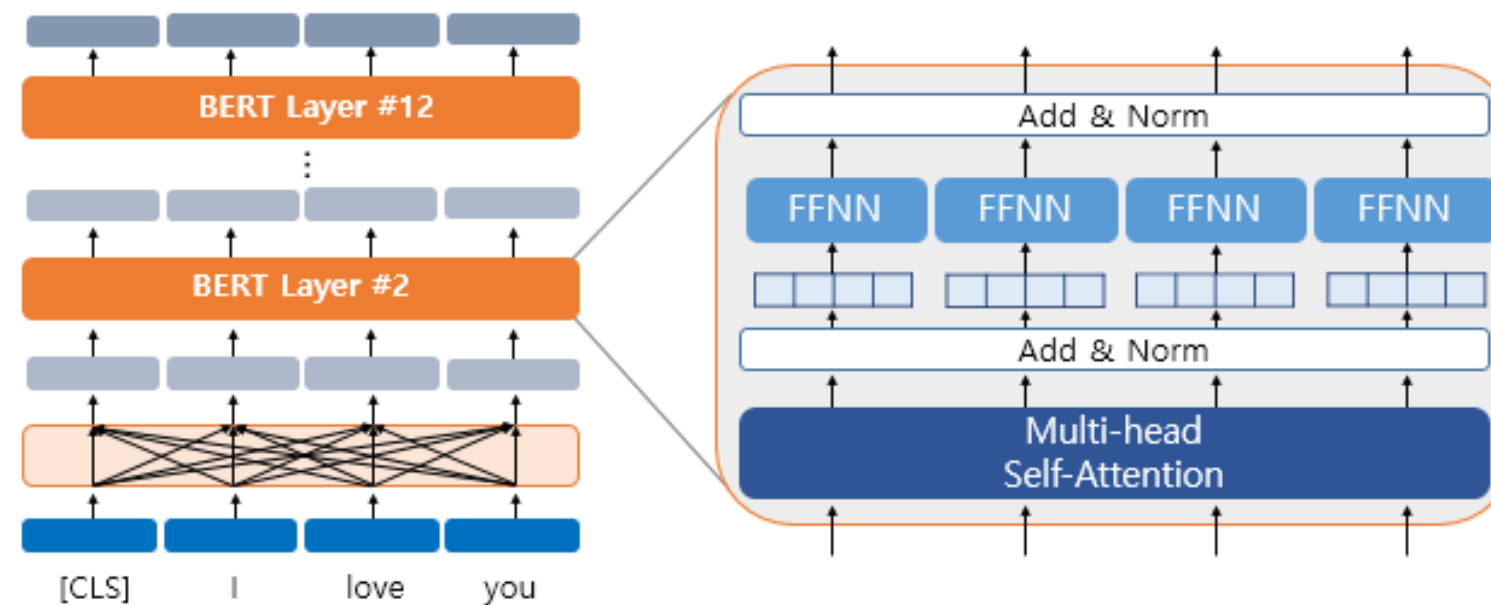


# 04 | 모델 소개

## BERT

### • BERT(Bidirectional Encoder Representations from Transformers)

- ✓ Text(Radiologist report) modal 학습에 사용할 모델
- ✓ 자연어 시퀀스를 입력했을 때, 문맥을 고려하여 수치화 된 벡터로 representation 하는 모델
- ✓ Transformer의 인코더 구조를 활용하여 학습되었기 때문에 attention mechanism의 장점을 그대로 반영





# 05 | 실험 방법

## Experiment

### 1. 관측치 필터링

- ✓ Radiology reports가 존재하는 환자를 대상으로 샘플을 구성
- ✓ 샘플로 구성된 환자들의 나이, 성별 비율을 확인하여 데이터가 불균형하지 않은지 확인

### 2. Radiology Reports 전처리 및 BERT representation

- ✓ Text data인 방사선의의 보고서를 전처리(특수 문자, 오타 등)
- ✓ 사전학습 된 BERT를 가져와 벡터로 표현

### 3. CXR(chest X-ray) representation

- ✓ 사전학습 된 Dense-Net을 가져와 finetuning
- ✓ Model 내 CNN feature map을 reshape하여 벡터로 표현

### 4. 성능 비교 실험

- ✓ Text 만 사용한 경우: 2의 결과를 BERT의 입력으로 넣고 BERT의 마지막 layer에 classifier를 추가하여 분류 성능 평가
- ✓ Image 만 사용한 경우: image를 단순히 Dense-Net의 입력으로 넣고 분류 성능 평가
- ✓ Text + Image 사용한 경우: 2, 3의 결과를 BERT의 입력으로 넣고, classifier를 추가하여 분류 성능 평가

# 06 | 논문 소개

## • 논문 개요

- ✓ Paper: “*Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers*”
- ✓ Author: Grzegorz Jacenków, Alison Q. O'Neil, Sotirios A. Tsaftaris
- ✓ 분야: **Computer Vision** and Pattern Recognition, Machine Learning
- ✓ Comments: Accepted at the **IEEE International Symposium on Biomedical Imaging (ISBI)** 2022 as an oral presentation.
- ✓ Submitted on 12 Feb 2022

## • 선정 이유

- ✓ 본 팀이 하고자 하는 바를 상세히 기술하고 있음(사용하고자 하는 데이터의 상황도 같음, report + image)
- ✓ 환자 진단의 사전 지식을 활용하여 더욱 정확한 이미지 진단이 가능함을 증명했음
- ✓ 이미지를 보고 질병을 분류하는 공모전 같은 대회들은 진단이 완료된 시점에서 수집된 데이터이기 때문에 실용성이 없음
- ✓ 하지만 본 논문은 **진단이 이루어지기 전의 정보만을 활용**하여 질병 예측을 하기 때문에 의사의 임상 업무를 도울 수 있는 실용적 예시를 보여줌
- ✓ Unimodal, Multimodal 실험을 상세히 진행하였음
- ✓ 본 프로젝트의 좋은 가이드라인이라고 생각했음

arXiv:2202.06076v1 [cs.CV] 12 Feb 2022

### INDICATION AS PRIOR KNOWLEDGE FOR MULTIMODAL DISEASE CLASSIFICATION IN CHEST RADIOGRAPHS WITH TRANSFORMERS

Grzegorz Jacenków<sup>1</sup> Alison Q. O'Neil<sup>1,2</sup> Sotirios A. Tsaftaris<sup>1,3</sup>

<sup>1</sup> The University of Edinburgh <sup>2</sup> Canon Medical Research Europe <sup>3</sup> The Alan Turing Institute

#### ABSTRACT

When a clinician refers a patient for an imaging exam, they include the reason (e.g. relevant patient history, suspected disease) in the scan request; this appears as the indication field in the radiology report. The interpretation and reporting of the image are substantially influenced by this request text, steering the radiologist to focus on particular aspects of the image. We use the indication field to drive better image classification, by taking a transformer network which is unimodally pre-trained on text (BERT) and fine-tuning it for multimodal classification of a dual image-text input. We evaluate the method on the MIMIC-CXR dataset, and present ablation studies to investigate the effect of the indication field on the classification performance. The experimental results show our approach achieves 87.8 average micro AUROC, outperforming the state-of-the-art methods for unimodal (84.4) and multimodal (86.0) classification. Our code is available at <https://github.com/jacenkow/mmbt>.

**Index Terms**— Multimodal Learning, Chest X-Ray Classification, Transformers, BERT

#### 1. INTRODUCTION

Chest radiography remains the most common imaging examination for the diagnosis and treatment of a variety of lung conditions such as pneumonia, cancer, and even COVID-19. Automation of X-ray interpretation could considerably improve healthcare systems, lowering costs and addressing the pressing challenge of expert shortage [1]. Yet, current techniques for clinical decision support mostly focus on a single modality (e.g. patient's X-ray) and do not take into account complementary information which might be already available in a hospital's database (e.g. patient's clinical history) [2], [3]. We are particularly interested in providing the indication field, i.e., the motivation for the patient's screening examination. This field may include the patient's history, a request to evaluate a particular condition, and other clues which can steer the radiologist's attention to particular imaging features. The indication field is often the only information provided by the referring physician [4], and can influence the interpretation of the imaging exam [5]. In this paper, we want to design a vision-and-language model that is able to use such text-based



**Fig. 1:** We consider the problem of classifying chest X-ray images given the patient information in a free-text form. We only use knowledge about the patient collected before the imaging examination and do not require radiologist intervention as opposed to most prior studies.

side information to aid and complement disease classification.

Current state-of-the-art methods for vision-and-language tasks (such as VisualBERT [6]) are mostly based on transformer architectures, which require extensive pre-training. The process typically involves using a dataset with annotated bounding boxes around the objects of interests, such as Conceptual Captions [7], to initialise the weights, which are later fine-tuned to the final task. Unfortunately, the biomedical community lacks domain-specific yet general multimodal datasets which could be used for pre-training large transformer networks. To address this problem, one could leverage existing unimodal models, and fine-tune the models to a multimodal task as proposed in multimodal BERT (MMBT) [8], which we evaluate on a biomedical task. As BERT does not provide the means to process imaging input, MMBT embeds image features from a ResNet-152 [9] classifier.

We evaluate the ability of a unimodally pre-trained BERT model to process biomedical imaging and non-imaging modalities during the fine-tuning step. Specifically, we use chest radiographs and the indication field from associated radiology reports to perform multi-label classification. The network can be pre-trained on unimodal datasets which are more common than multimodal, but it is still capable of learning multimodal interactions during the fine-tuning step.

**Contributions:** (1) We present a strong baseline for multimodal classification of chest radiographs; (2) We evaluate the model with the prior work achieving the new state-of-the-art

# 07 | Reference

- [1] Multimodal Machine Learning(<https://datascience0321.tistory.com/31>)
- [2] 소아 흉부 이미지 데이터(<https://aihub.or.kr/aihubdata/data/view.do?cumMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=652>)
- [3] <https://github.com/Jeong-Eul/Data-Mining-Study/tree/main/Paper/Indication>
- [4] BERT(<https://wikidocs.net/115055>)





Q & A

---

감사합니다