

우리는
하나조

흉부 X-ray를 이용한 소아 폐렴 예측



흉부 X-ray를 이용한 소아 폐렴 예측

목차

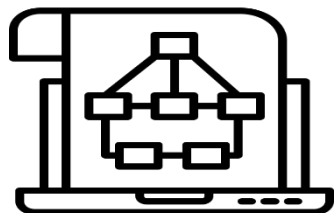
1. 조원 및 역할 소개
2. 프로젝트 주제 및 선정 배경
3. 프로젝트 수행 절차 및 과정
4. 프로젝트 결과
5. 자체 평가 의견

01

조원 및 역할 소개

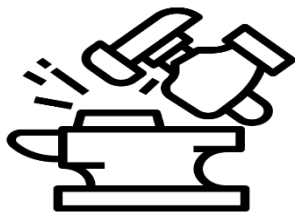
01 조원 및 역할 소개

조성진



모델 구현

남태우



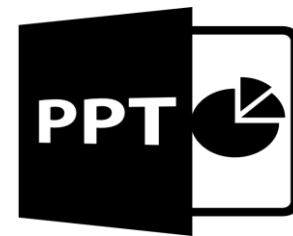
데이터 가공

이지안



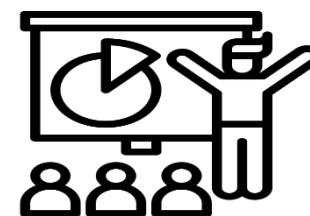
데이터 시각화

정은정



ppt 제작

이호재



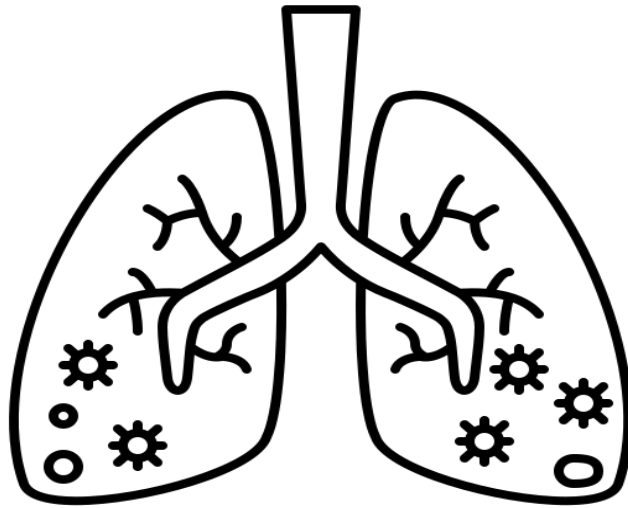
ppt 발표

02

프로젝트 주제 및 선정 배경

1. 페럼이란?
2. 페럼의 종류

2-1 폐렴이란?



폐 실질에 염증이 생기는 질환

감염성 폐렴

세균성 폐렴

바이러스성 폐렴

곰팡이성 폐렴

비감염성 폐렴

화학물질노출

방사선 감염

2-1 폐렴이란?

성인 폐렴

대개 하부 폐부에서 발생

CT, X-ray

VS

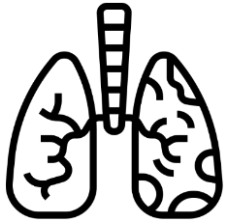
소아 폐렴

대개 상부 폐부에서 발생

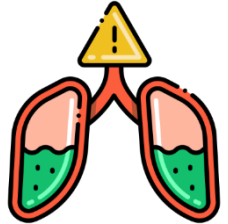
폐가 작아서 CT 검사 시 혼동이 올 수도 있음

장기들이 미성숙하여 성인들보다 방사선 노출에 민감. 사진을 잘못 판독하여 재촬영 하는 일이 없도록 유의 필요

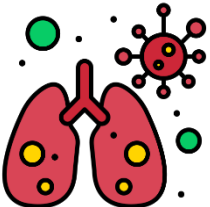
2-1 폐렴이란?



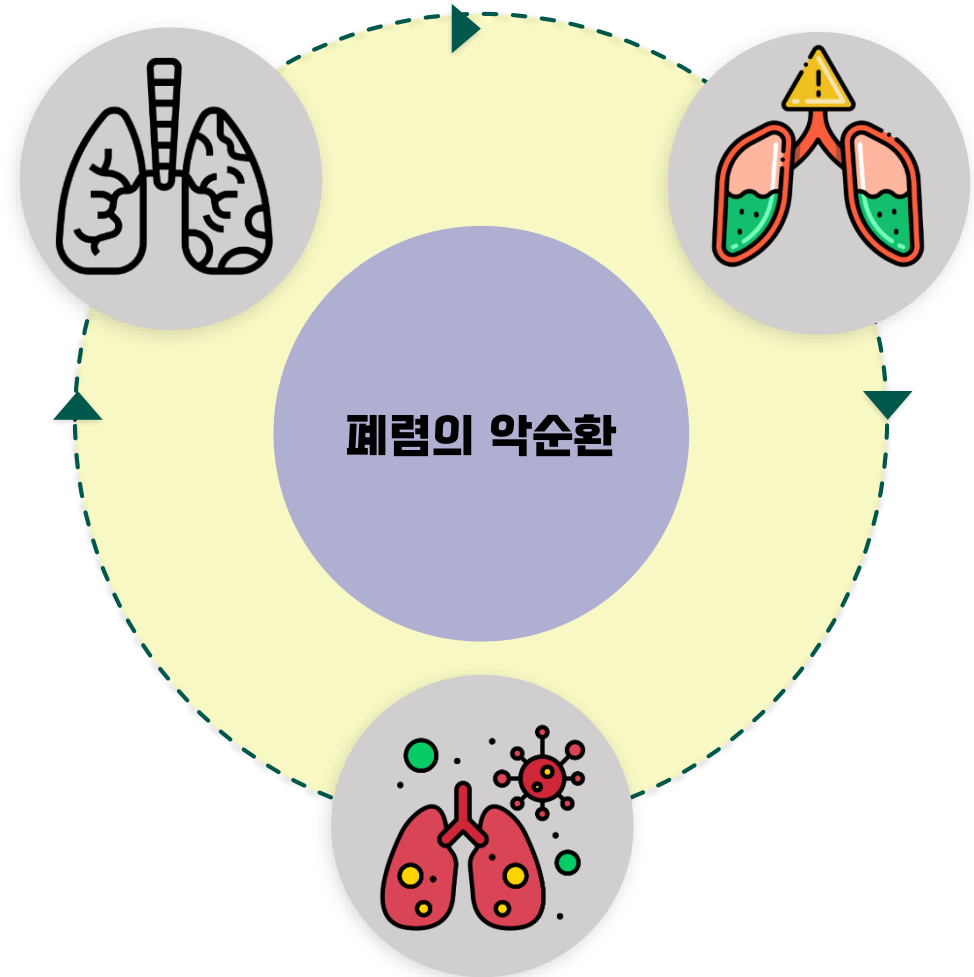
폐조직의 손상



병원체가 자라기 좋은
환경

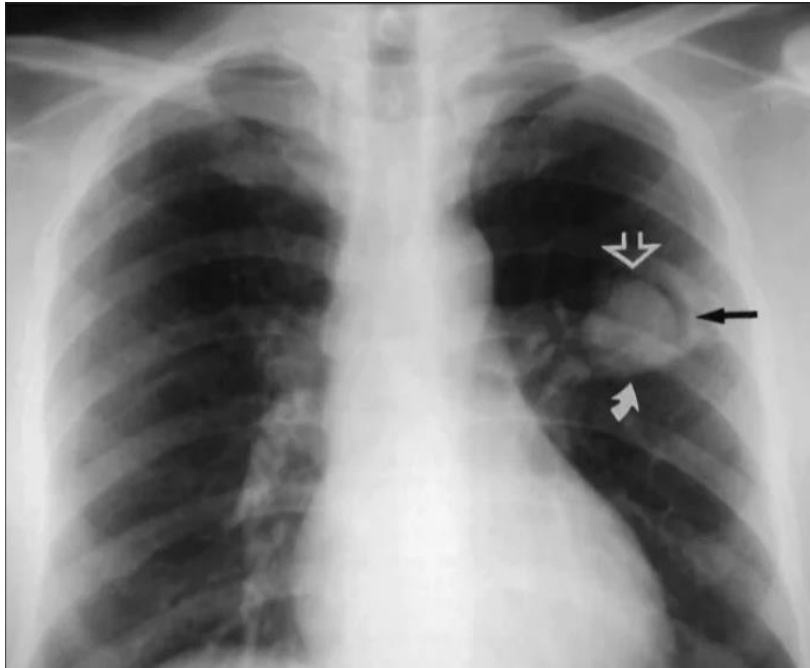


염증 생성



2-2 폐렴의 종류

■ 세균성 폐렴(Bacterial Pneumonia)



주로 면역결핍증 이나 만성질환을 가지고 있는 환자에게 관찰

주로 폐기능이 떨어진 경우에 합병되는 경우가 많음

영아기에 흔히 나타남, 전체 발병의 70%가 1세 미만

포도알균

연쇄
상구균

폐렴구균

2-2 폐렴의 종류

■ 바이러스성 폐렴(Viral Pneumonia)



신생아를 제외하고 소아와
청소년에서 가장 흔한 폐렴의 원인

호흡기세포융합바이러스(Respiratory
syncytial virus, RSV)의 경우 영아기 폐렴의
가장 흔한 원인
종종 무호흡을 동반, 천식의 발병 위험을 높임

코로나

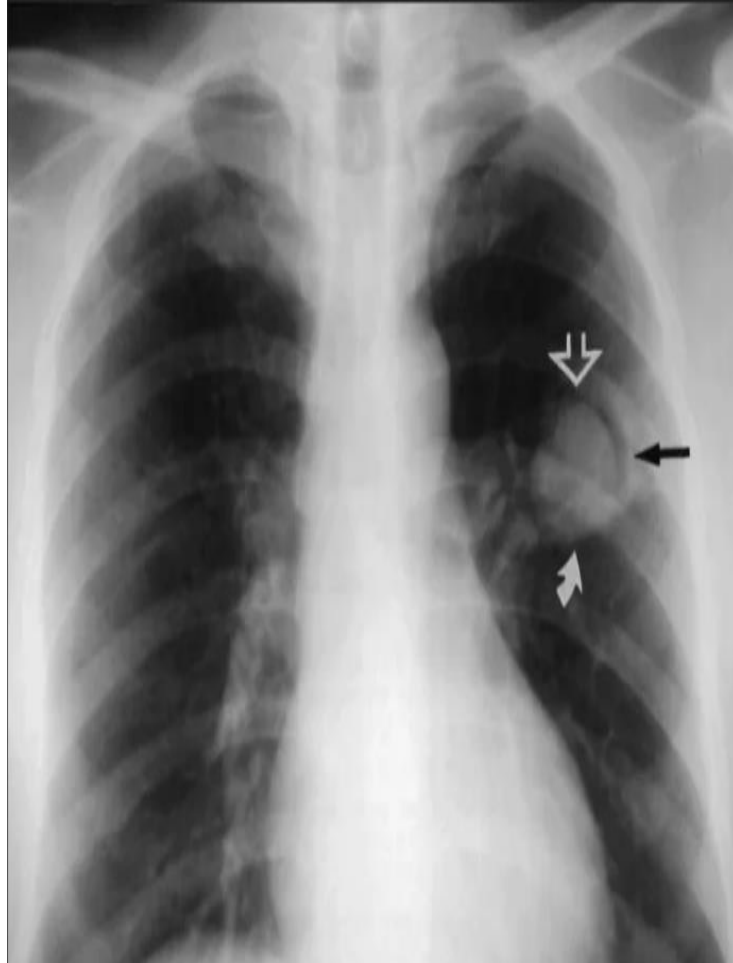
아데노
바이러스

호흡기
세포융합
바이러스

2-2 폐렴의 종류



정상적인 폐



세균성 폐렴



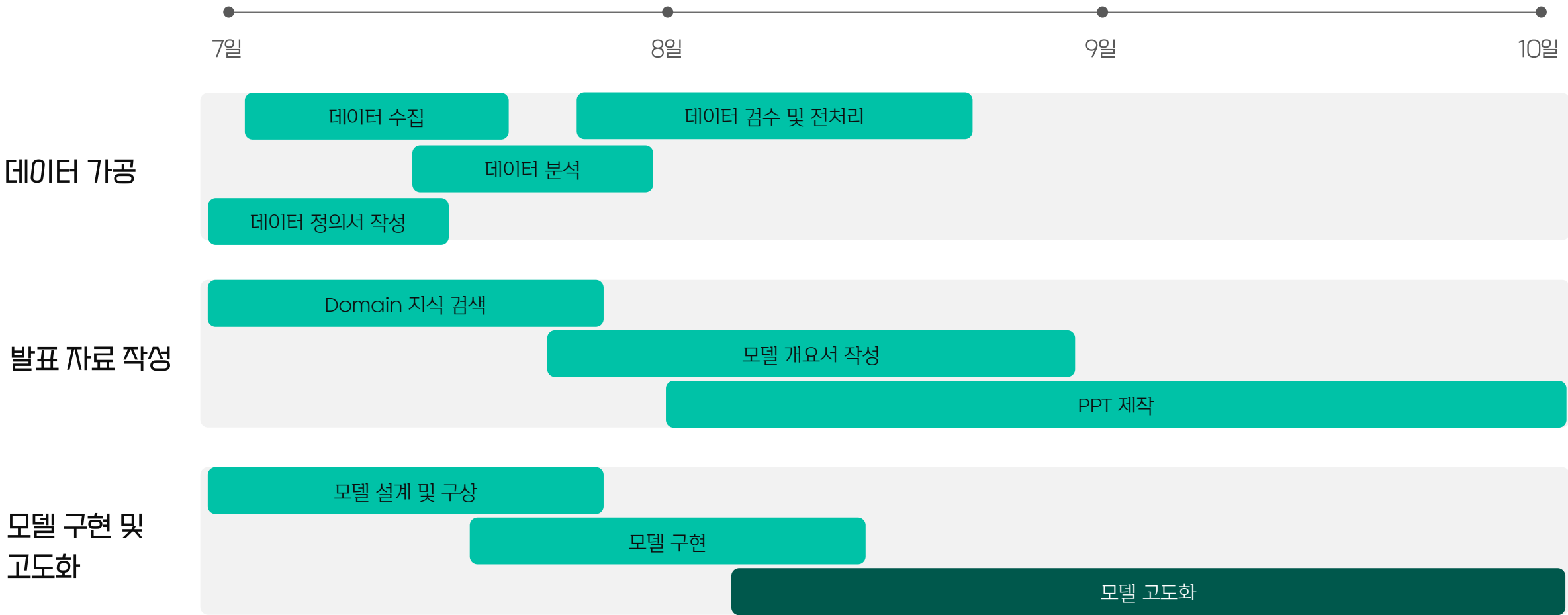
바이러스성 폐렴

03

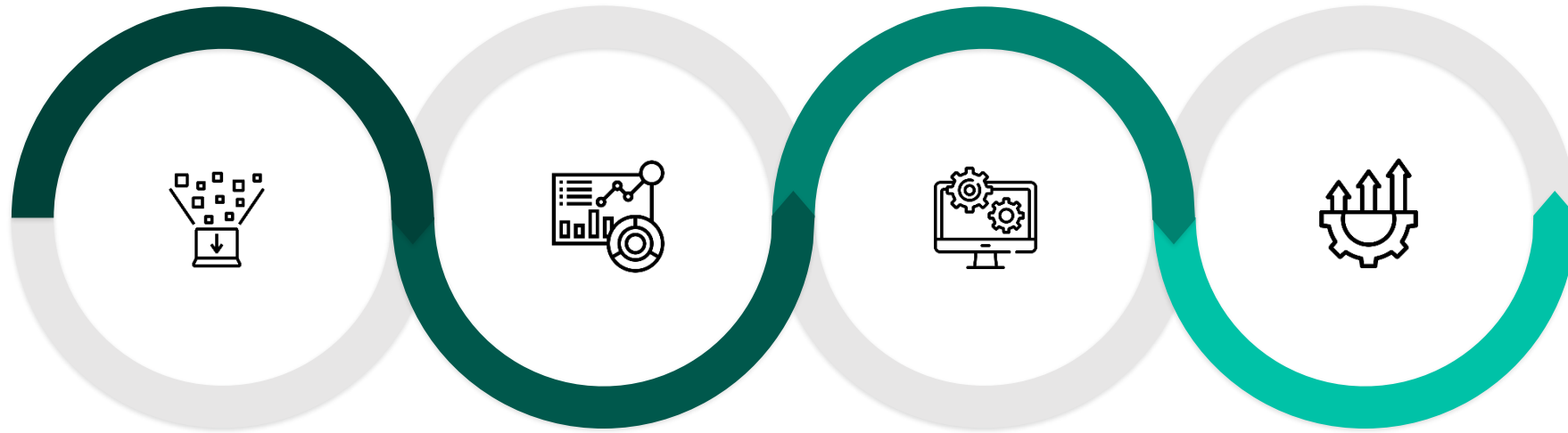
프로젝트 수행 절차 및 과정

1. 프로젝트 수행 일정
2. 프로젝트 과정

3-1 프로젝트 수행 일정



3-2 프로젝트 과정



01 데이터 수집

02 데이터 전처리

데이터 분석
데이터 전처리

03 모델 구현

사용 모델
성능 평가 지표
모델 설계

04 모델 고도화

모델 시행
오류 수정

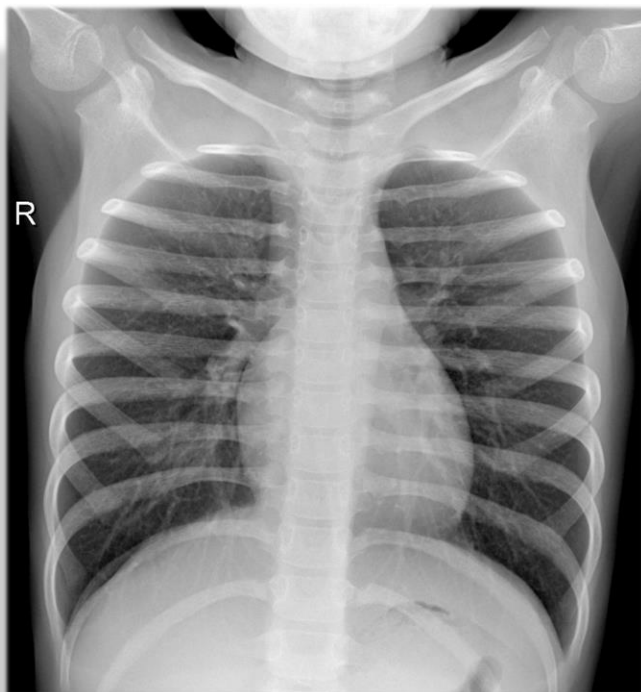
3-2 프로젝트 과정

■ 데이터 수집

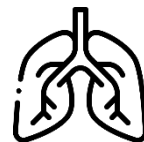


3-2 프로젝트 과정

입력 데이터



- 총 데이터 수 5840
- Train data 4672
- Test data 1268



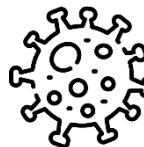
정상인

- Class 0
- Train data 1227 개 데이터



세균성 폐렴

- Class 1
- Train Data 2238 개 데이터

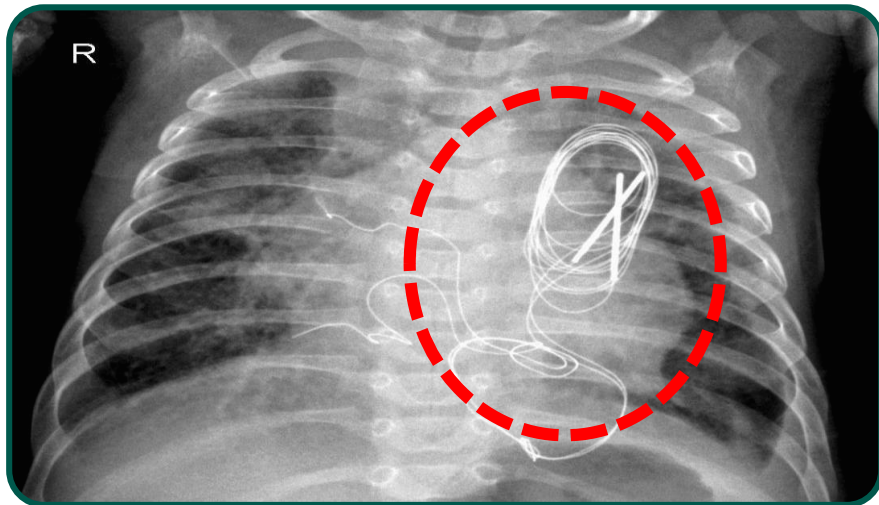


바이러스성 폐렴

- Class 2
- Train Data 1207 개 데이터

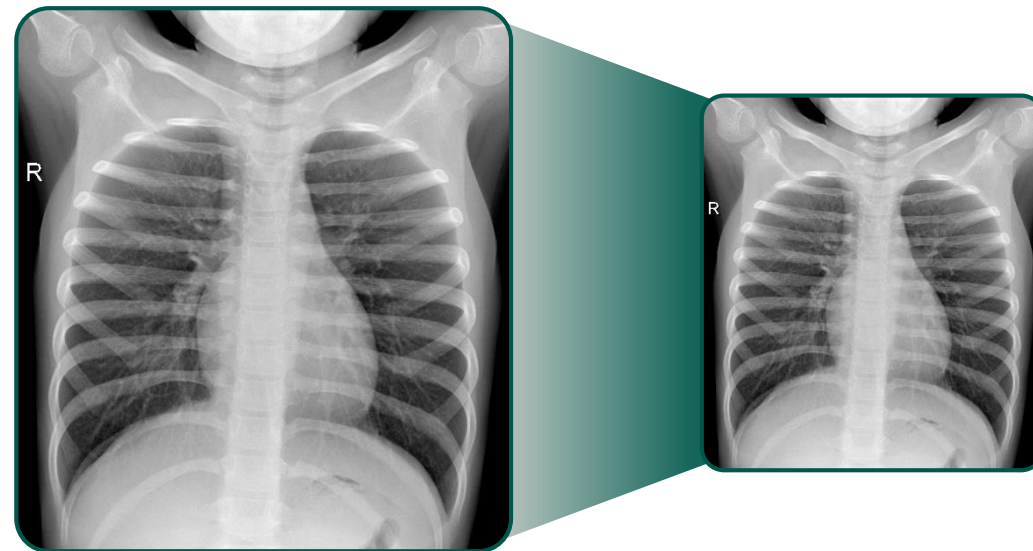
3-2 프로젝트 과정

데이터 전처리



이물질 제거

폐를 가리는 이물질 사진 제거
150개 제거



사이즈 조절

이미지 사이즈 224 * 224

3-2 프로젝트 과정

모델 설계

모델 채택

모델 성능 지표

01

VGG Net

02

RES Net

03

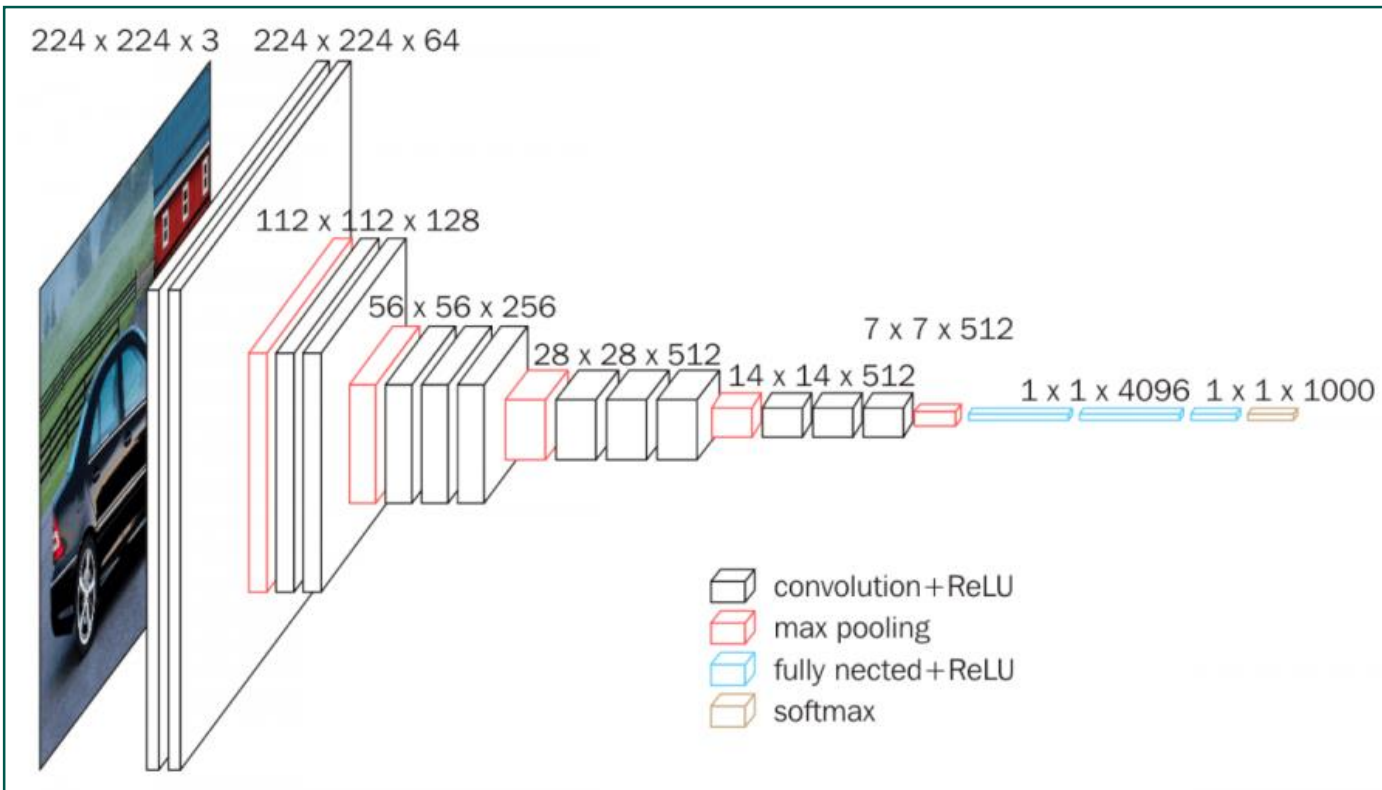
Grad CAM

04

F1 Score

3-2 프로젝트 과정

모델 설계



VGG Net 이란?

☑ 소개

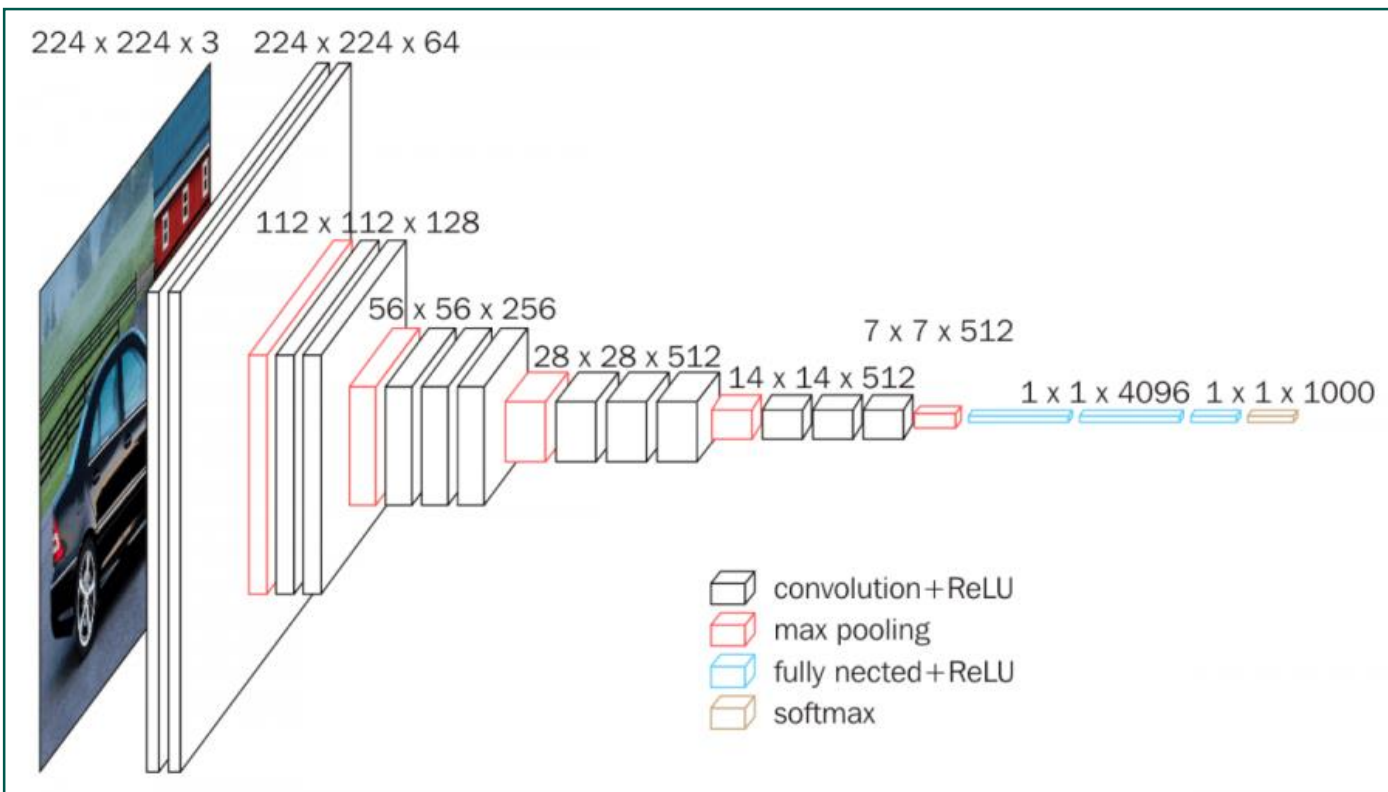
옥스퍼드 대학 연구팀 VGG에 의해 개발된 16개 또는 19개의 layer로 구성된 모델. 2014년 ImageNet 이미지 인식 대회에서 준우승.

☑ 구성

☑ 특징

3-2 프로젝트 과정

모델 설계



VGG Net 이란?

✓ 소개

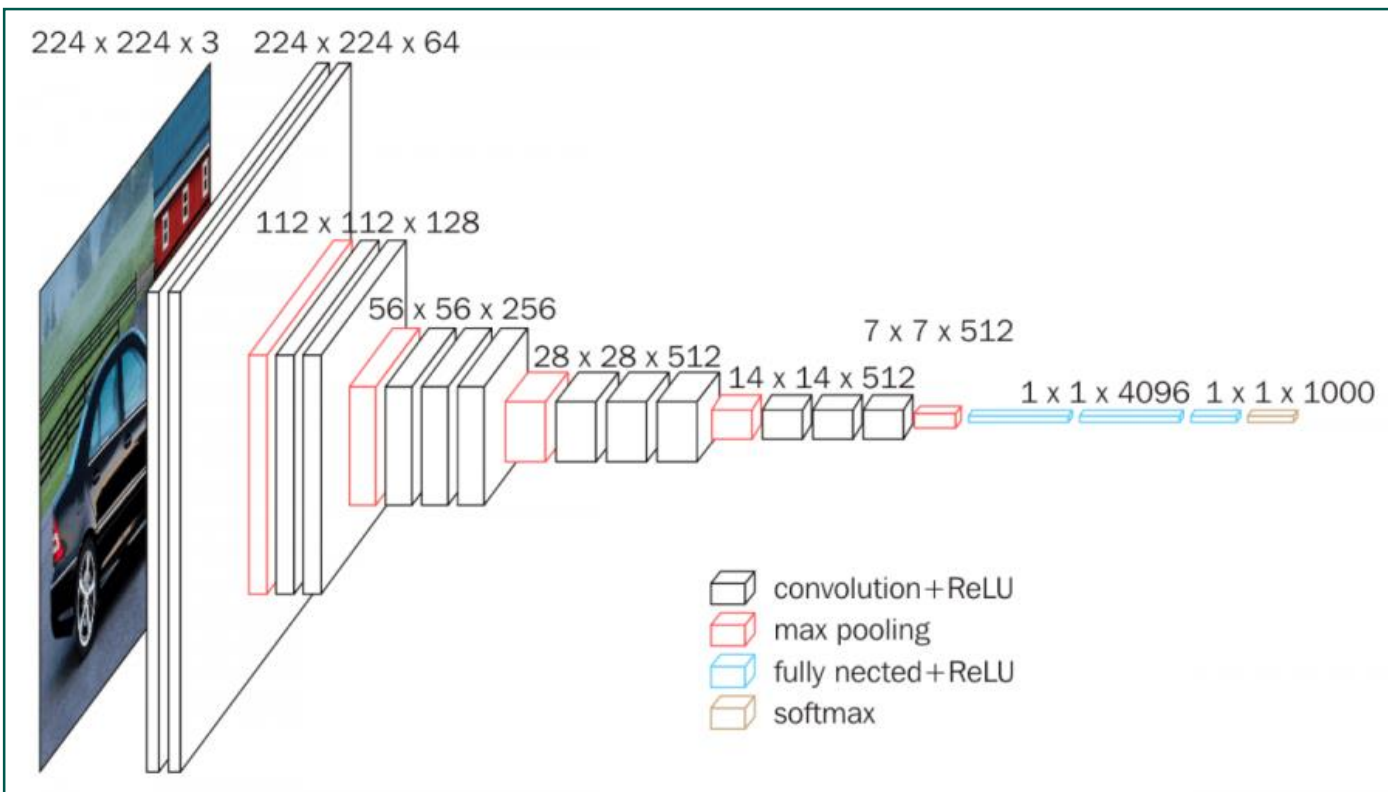
✓ 구성

13개의 Convolution Layer
3개의 Fully - Connected Layer
3 * 3 의 필터 / Stride 1 / Padding 1
2 * 2 의 Max Pooling

✓ 특징

3-2 프로젝트 과정

모델 설계



VGG Net 이란?

✓ 소개

✓ 구성

✓ 특징

여러 개의 layer를 쌓아 ReLU를 더 많이 적용

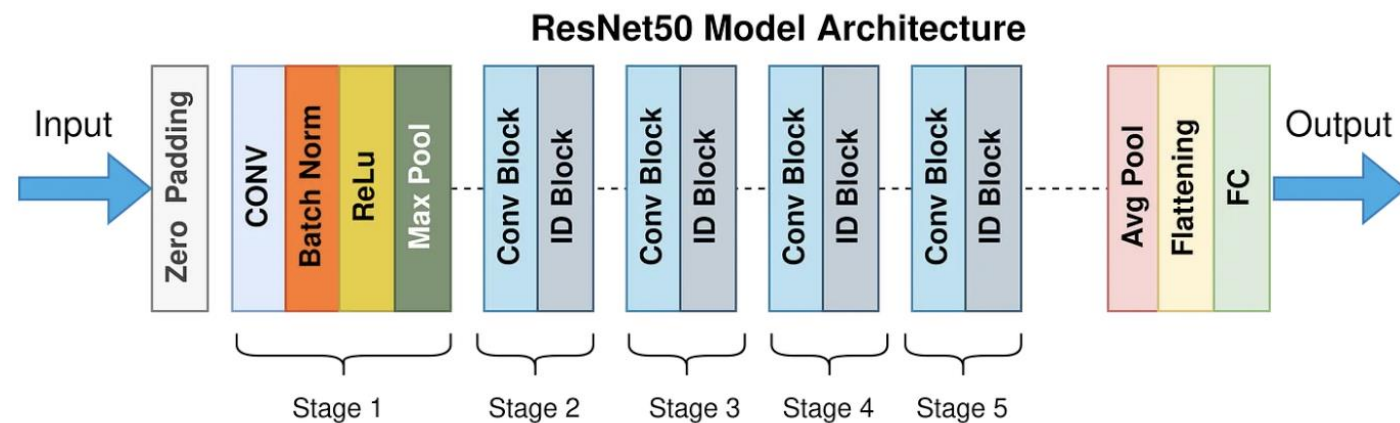
→ 비선형성 증가

최종단의 fully-connected layer 파라미터의 개수 증가

→ 학습 시간이 오래 걸린다는 단점 존재

3-2 프로젝트 과정

모델 설계



Resnet-50 Model architecture

RES Net 이란?

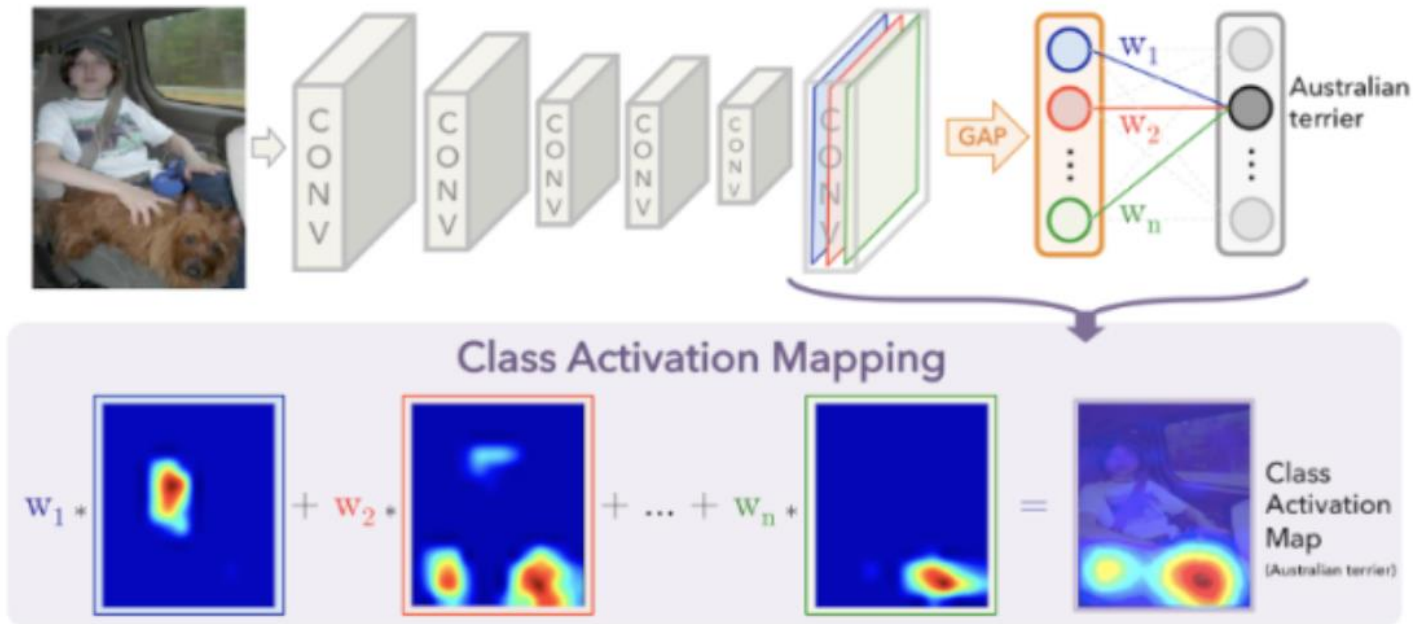
✓ 특징

Layer가 깊어질수록 신경망의 가중치를 업데이트 하는 과정에서 기울기가 작아져 에러가 커지는 문제

→ 해결하기 위해 만들어진 Residual Neural Network

3-2 프로젝트 과정

모델 설계



Grad CAM 이란?

특징

- 이미지 분석 모델에서 마지막 Convolution Layer의 결과값을 원본 이미지 사이즈로 변경
 - Global Average Pooling을 적용(이하 'x')
 - 뒤에 적용해주는 Dense Layer의 가중치 값들을 x값에 곱해주어 나온 값들을 더해준 것을 히트맵으로 표시
 - 히트맵과 이미지 겹쳐서 표시

3-2 프로젝트 과정

모델 설계

Precision

예측값 중에서 실제 값으로 예측한 비율

Recall

실제값 중에서 제대로 예측한 비율

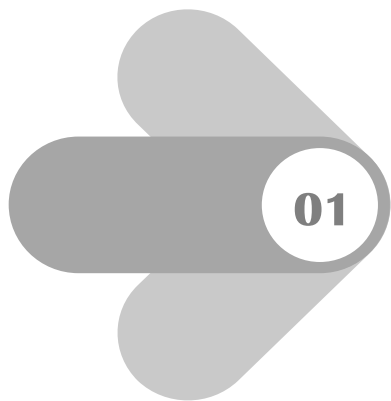
F1 Score

Precision과 Recall의 조화평균을 나타낸 값

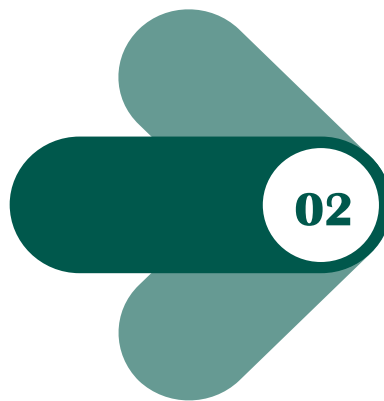
예측값 실제값	Class 0	Class 1	Class 2
0	T	F	F
1	F	T	F
2	F	F	T

3-2 프로젝트 과정

모델 고도화

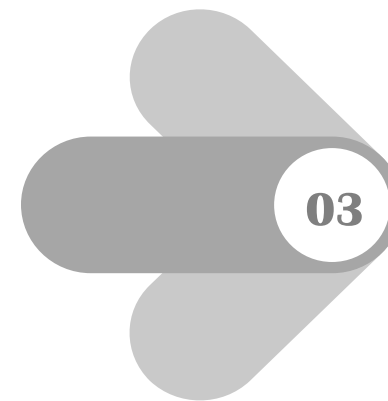


1차 시도



Dropout 추가

Dropout : 0.2

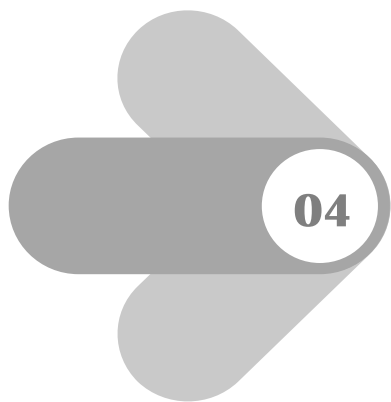


Output Layer 개수 변경

Third Dense : 1000 → 3

3-2 프로젝트 과정

모델 고도화



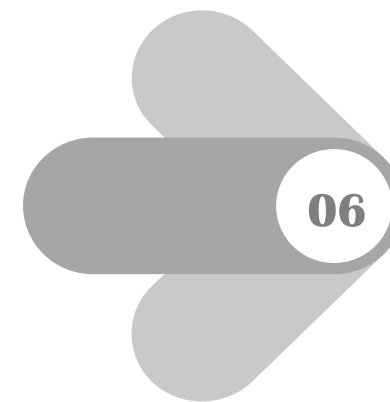
Hidden Layer 개수 변경

First Dense : 4096 -> 5000
Second Dense : 4096 -> 1000



Activation Function 변경

ReLU -> Leaky ReLU

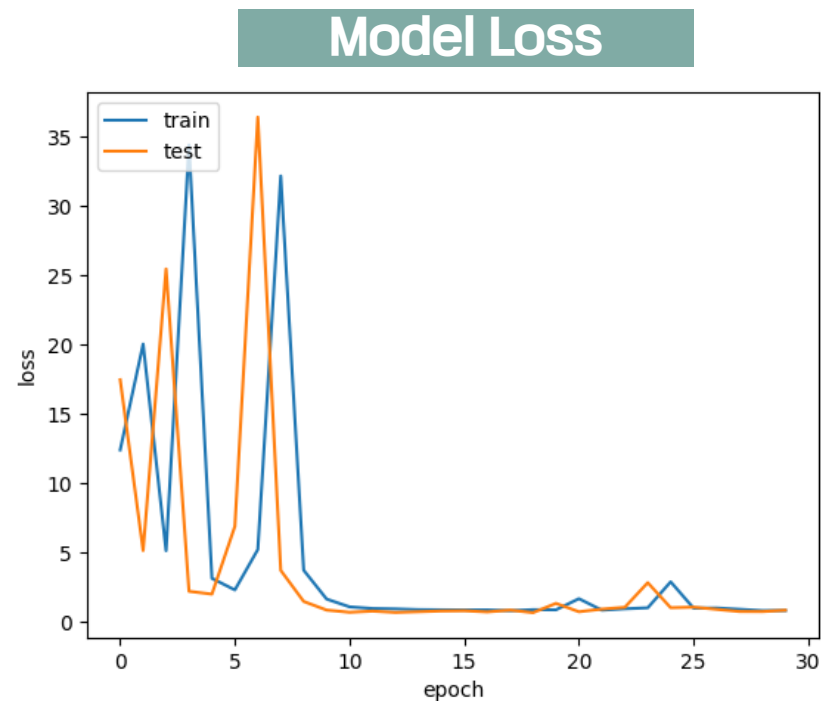
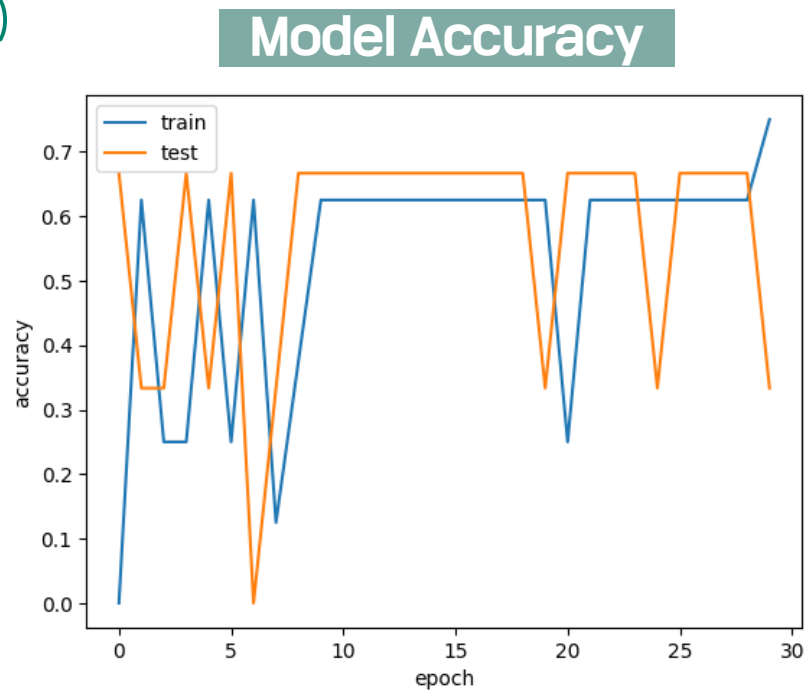


Hidden Layer 추가 및 개수 변경

First Dense : 5000 -> 4096
Second Dense : 1000 -> 4096
Third Dense : 3 -> 1000
Fourth Dense : 3

3-2 프로젝트 과정

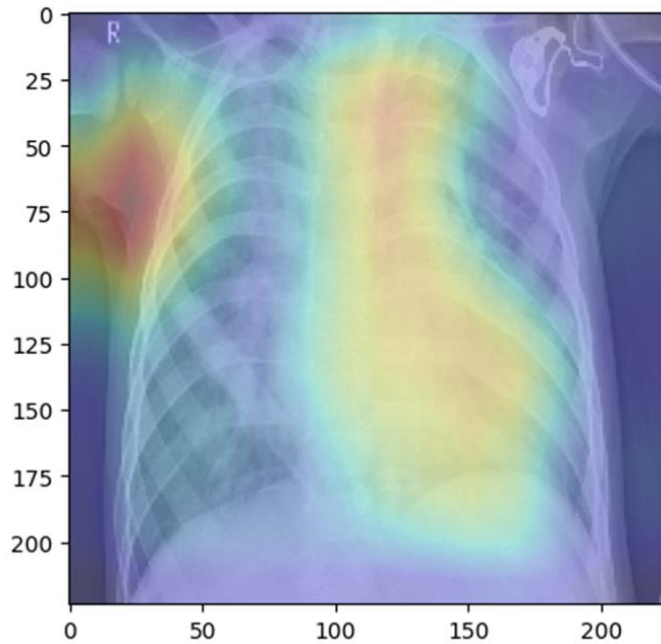
1차 시도 (VGG16)



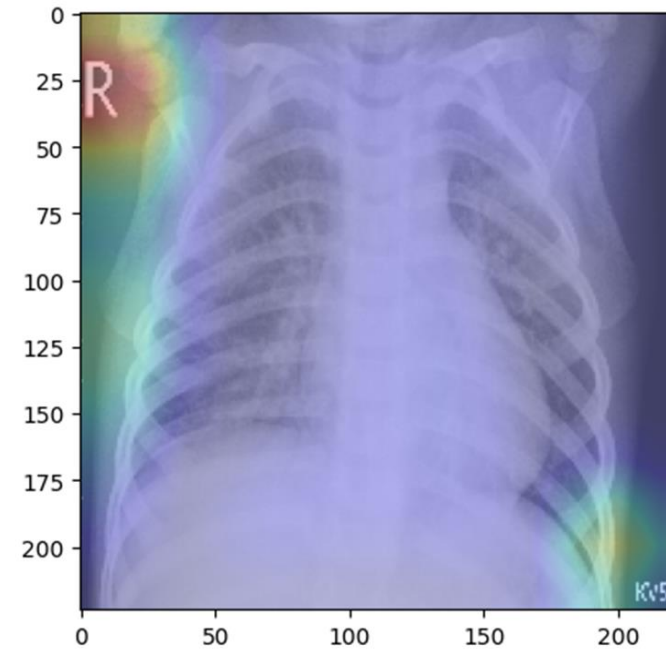
Train 정확도가 증가하나 Test 정확도가 감소함

3-2 프로젝트 과정

1차 시도 (VGG16)



Grad CAM



이미지 내에 R이라는 글자가 너무 크면 그 부분에 **가중치 할당**이 높게 이루어지는 문제

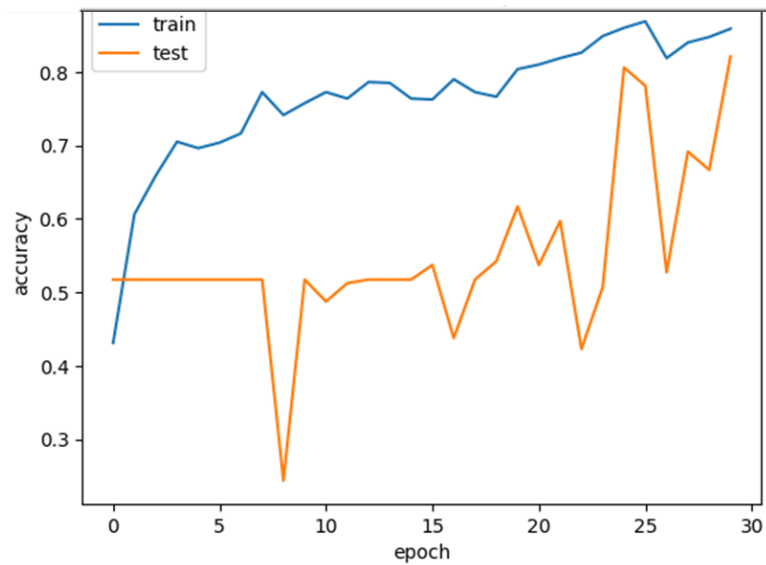
→ Dropout 추가, Image scale down

3-2 프로젝트 과정

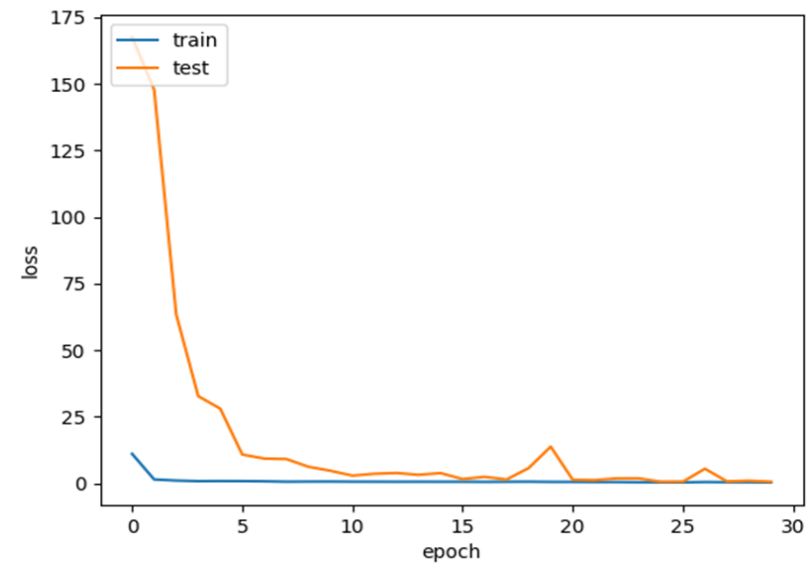
2차 시도

(VGG16 + Dropout)

Model Accuracy



Model Loss



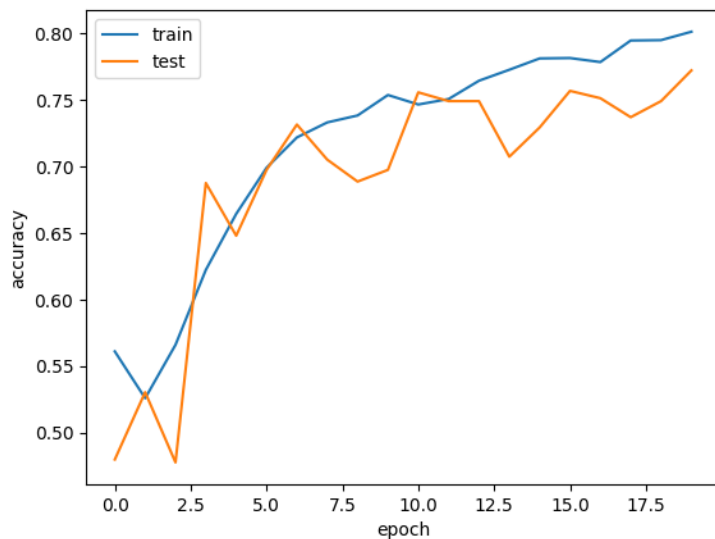
Test accuracy 증가

3-2 프로젝트 과정

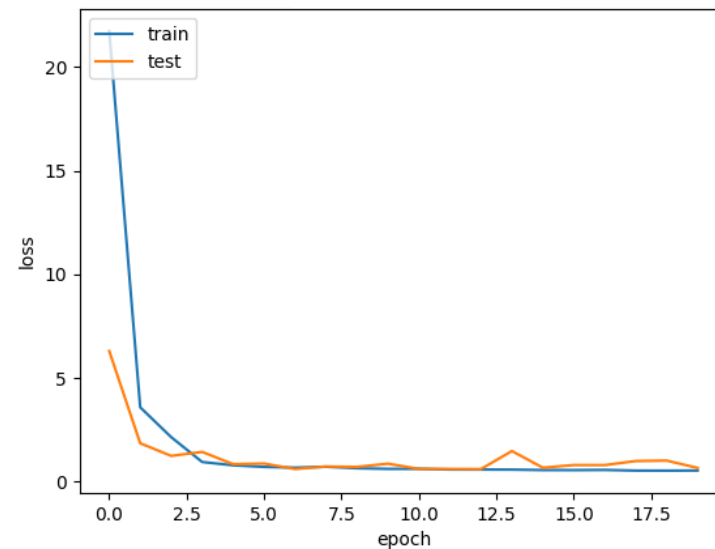
3차 시도

(VGG16 + Dropout + Third Layer : Dense 3)

Model Accuracy



Model Loss



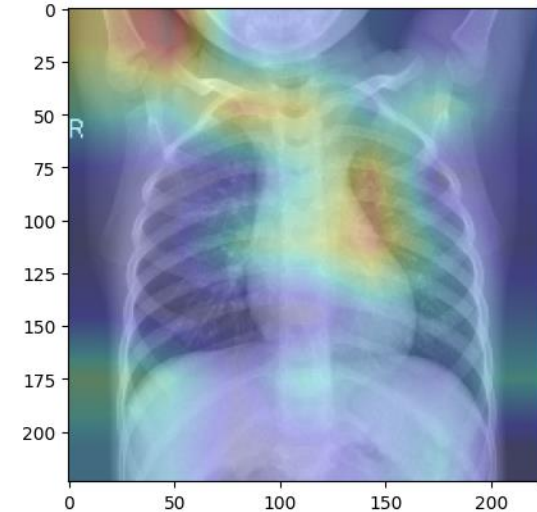
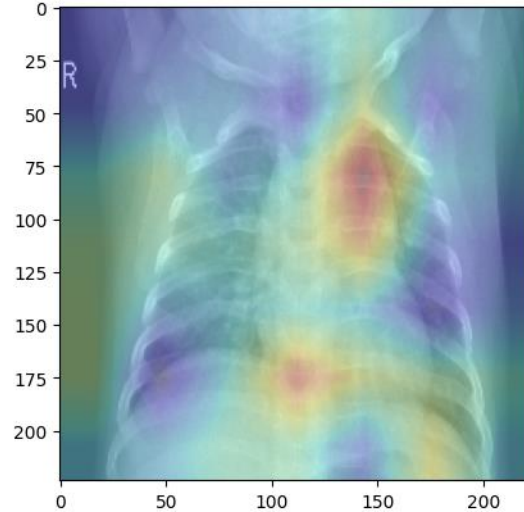
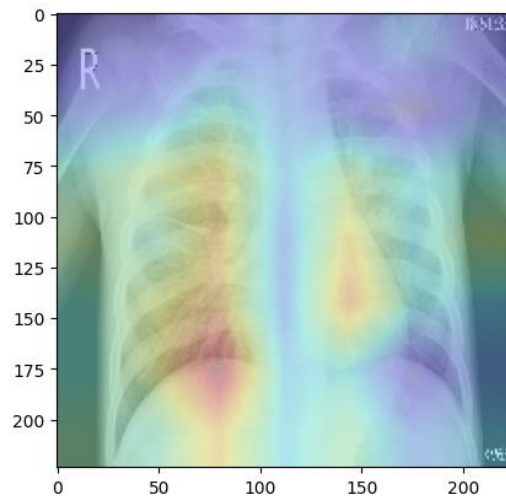
Accuracy 와 loss 수치 개선

3-2 프로젝트 과정

3차 시도

(VGG16 + Dropout + Third Layer : Dense 3)

Grad CAM

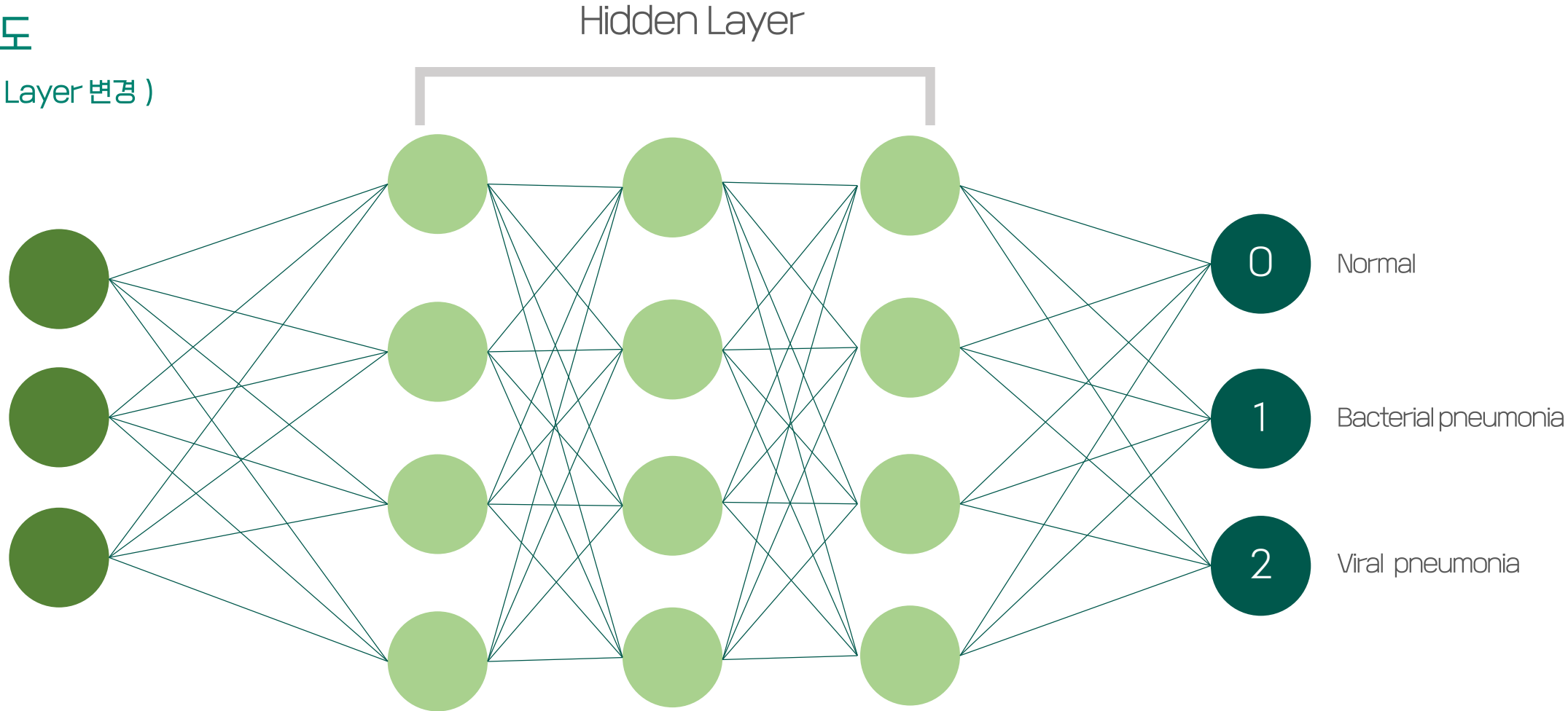


폐 말고 다른 부위 보고 판별하는 문제 발생

3-2 프로젝트 과정

4차 시도

(Hidden Layer 변경)

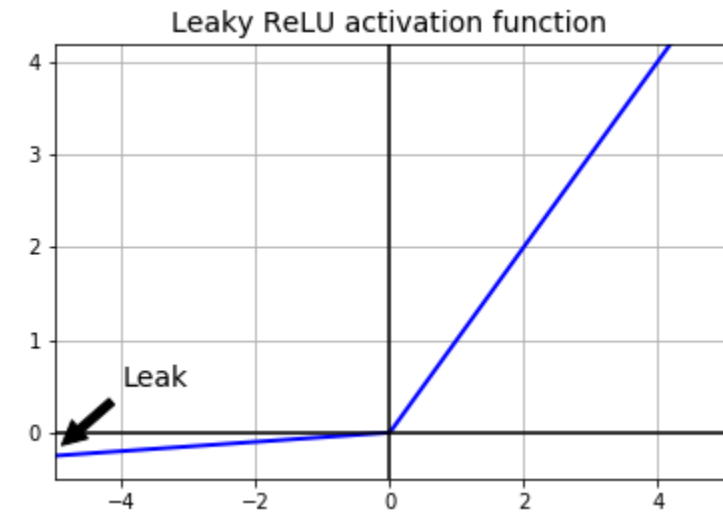
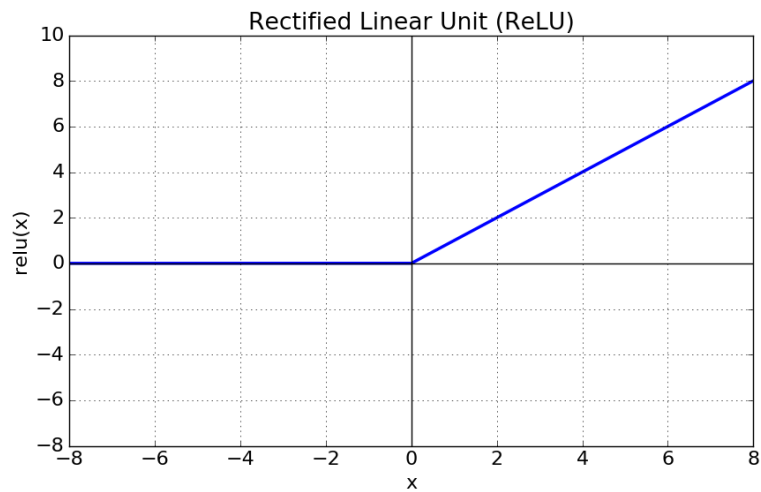


First Dense : 4096-> 5000, Second Dense : 4096 -> 1000, Third Dense : 1000->3

3-2 프로젝트 과정

5차 시도

(Activation Function 변경)

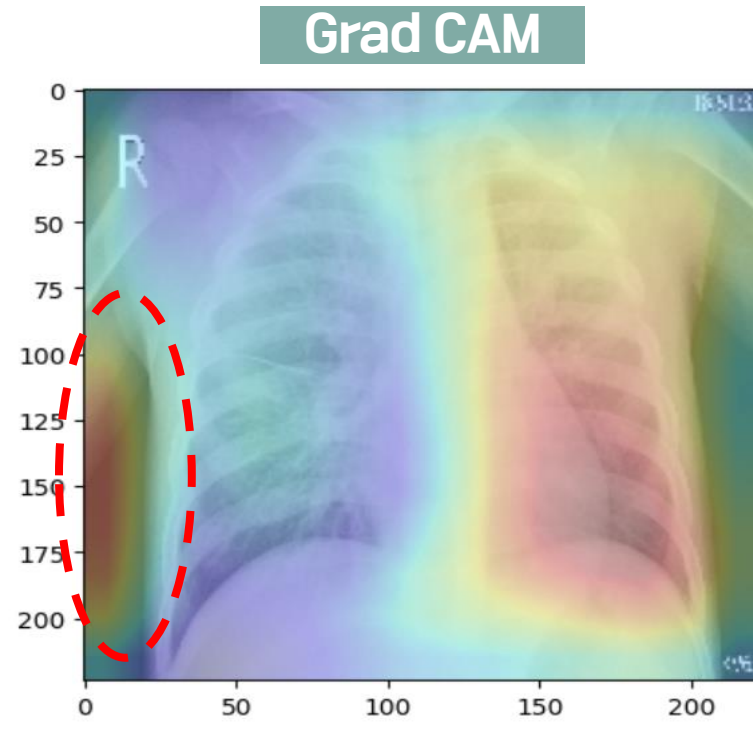


Activation Function **Leaky ReLU**로 변경

3-2 프로젝트 과정

5차 시도

(Activation Function 변경)

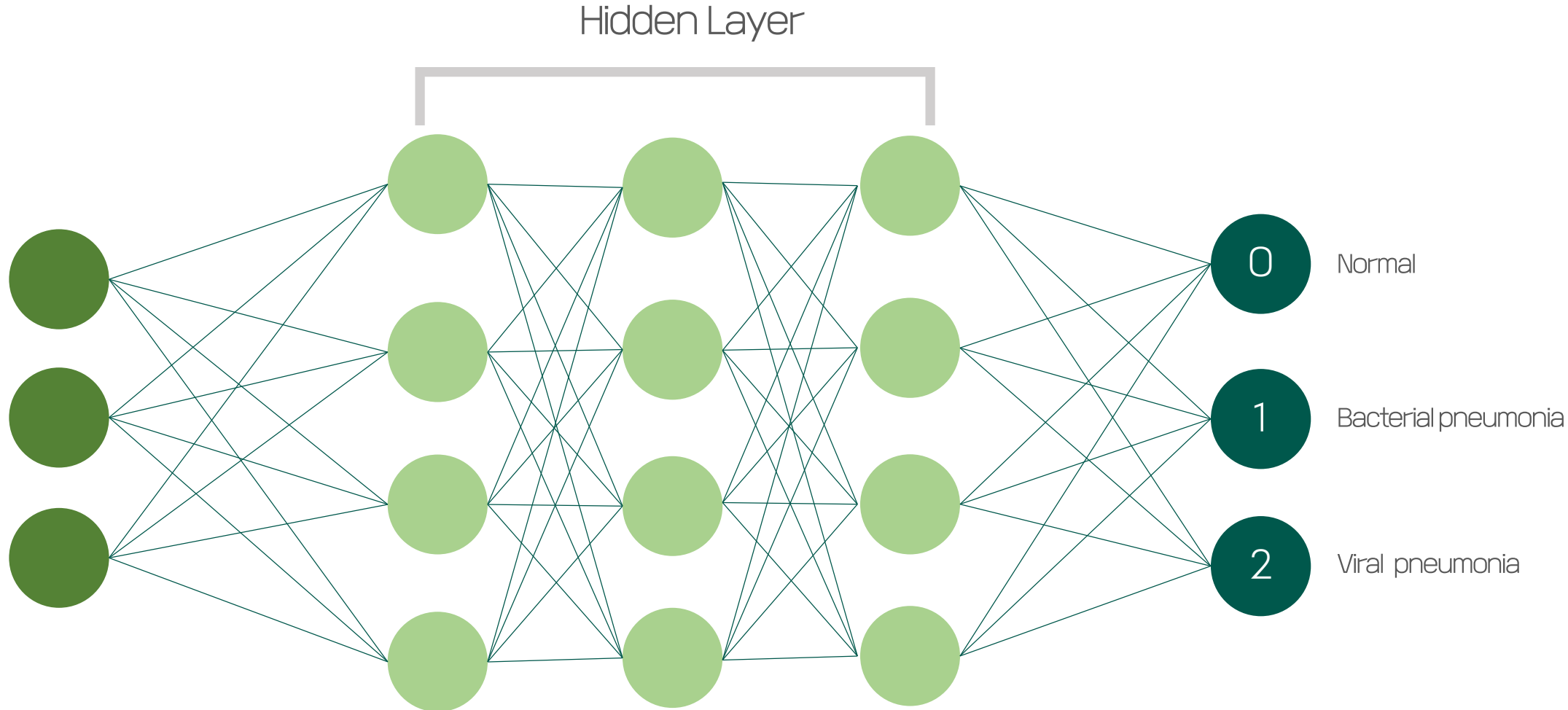


잘 찾은 사진 내에서도 폐 밖을 가리키는 문제

04

프로젝트 결과

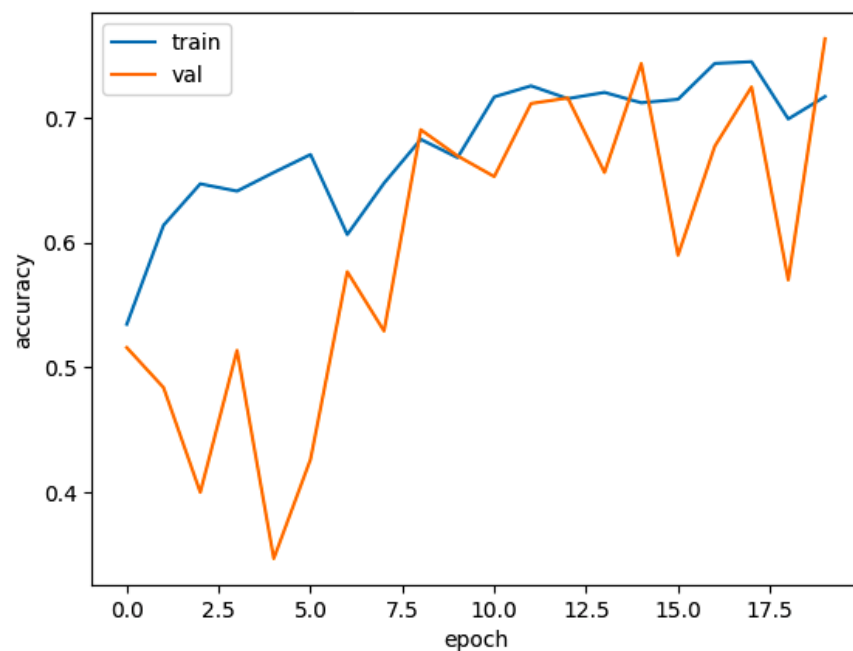
04 프로젝트 결과



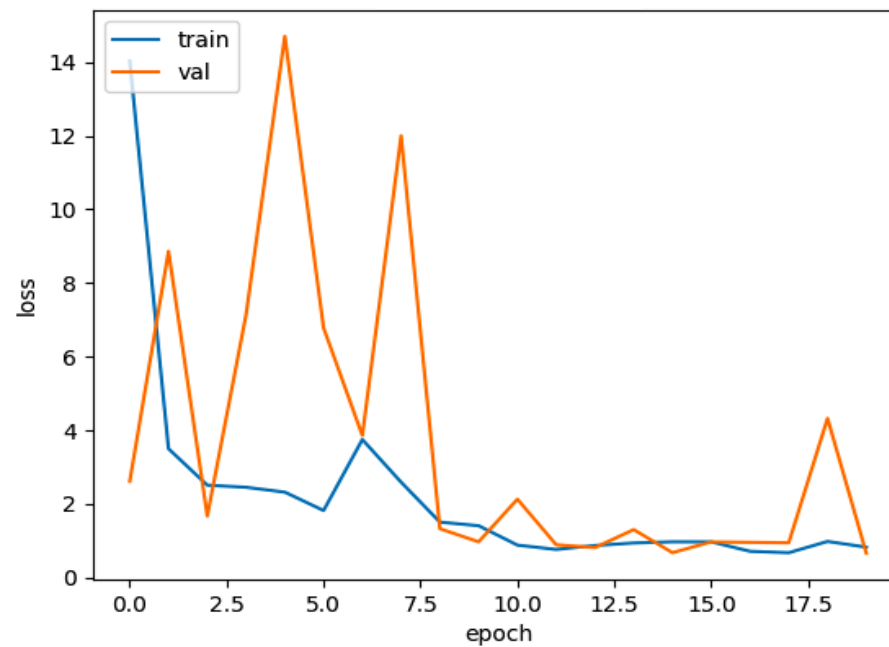
First Dense : 4096 -> 5000, Second Dense : 1000 -> 4096, Third Dense 3 -> 1000, Fourth Dense : 3

04 프로젝트 결과

Model Accuracy

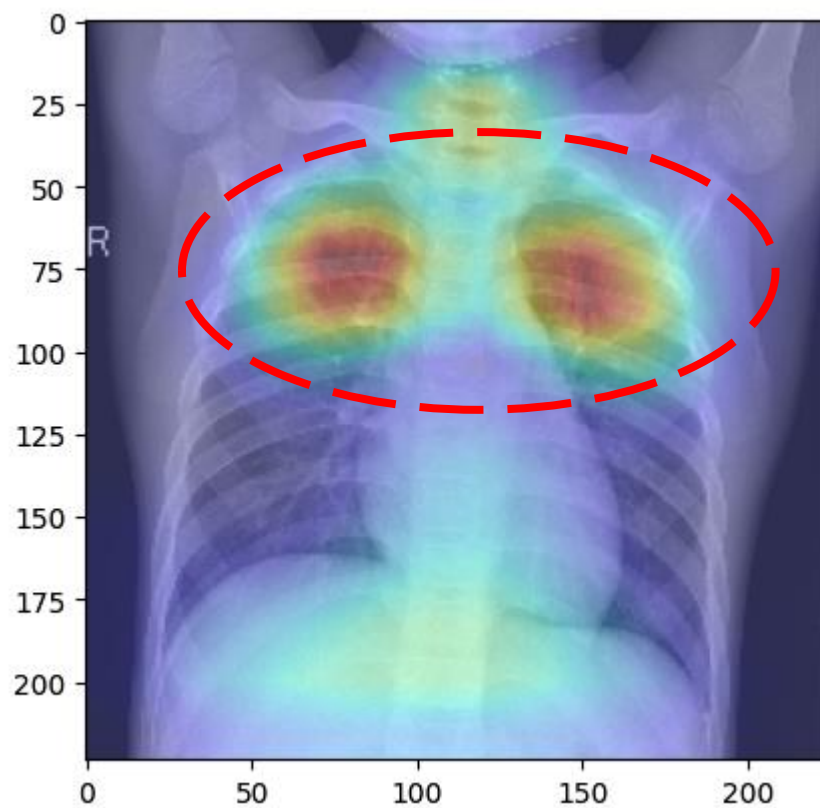


Model Loss



04 프로젝트 결과

Grad CAM



04 프로젝트 결과

F1 score

```
Precision : 0.7868517
```

```
Recall : 0.7632392
```

```
F1 Score : 0.7748656186349338
```

05

자체 평가 의견

05 자체 평가 의견

모델 학습에 용의하게 데이터 전처리

- 주어진 데이터 중 학습에 방해되는 데이터(이물질)제거

Grad-CAM 적용

- XAI(설명 가능한 인공지능)을 위해 시각화 방법으로 Grad-CAM 선정

모델 고도화를 위한 다양한 시도

- 정확성을 높이기 위해 6번 재시도

05 자체 평가 의견

기술적 한계

- Labeling 되지 않은 Test Data 활용 실패
- 이물질 사진 제거 시 명확하지 않은 기준



이물질 있는 사진도 모델에 적용할 수 있도록 개선방법을 모색해야함

기능적 한계

- 이미지 크기(1400 * 1024)로 인해 모델 학습 시 오랜 시간 소요
- 개발 환경 성능이 저조하여 모델 학습이 중단되는 현상 발생

Q & A



Thank you