

PNEUMONIA CLASSIFICATION MODEL

내향인들

CONTENTS

1

프로젝트 개요

3

성능 개선 & 튜닝

2

모델 선정

4

결론 및 회고

데이터 설명

train.csv

정상: 1341

폐렴: 3875

데이터 불균형

test.csv

정상: 234

폐렴: 390

val.csv

정상: 8

폐렴: 8

배경 및 목표

배경:

kaggle의 chest x-ray pneumonia 데이터를 활용

CNN 모델을 통해 폐렴을 분류

목표:

2종 오류를 줄이고, 데이터 증강을 통해 모델 예측력 향상

4가지 모델을 비교 후 최종 모델 선정 후 모델 개선 및 튜닝

PNEUMONIA



NORMAL



평가 지표

F1-Score

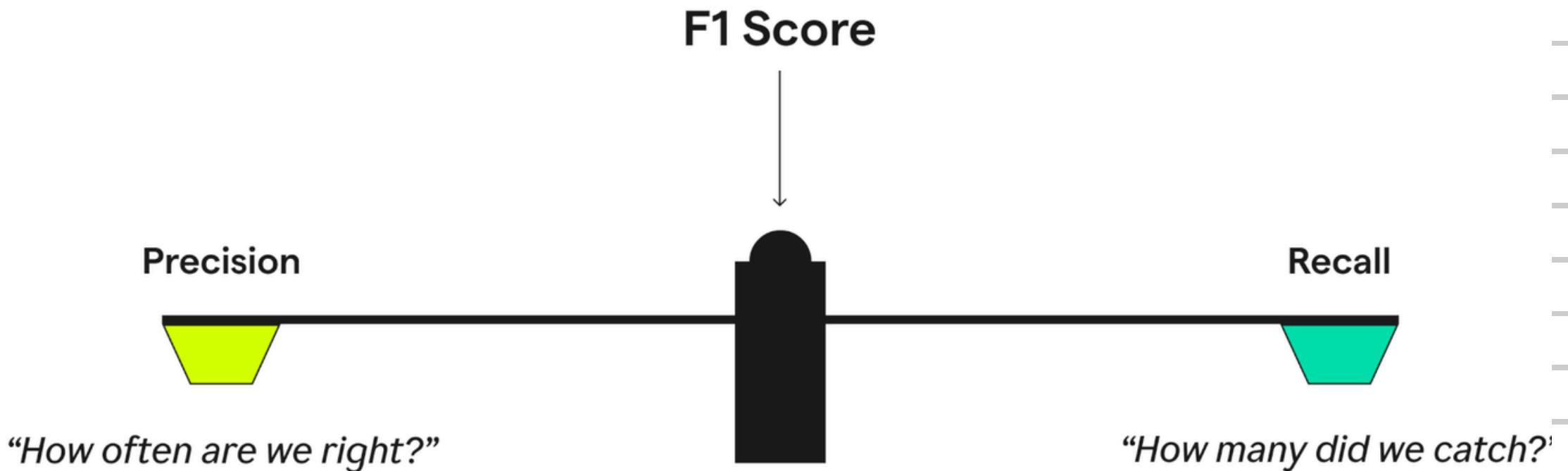
AUC-ROC Curve

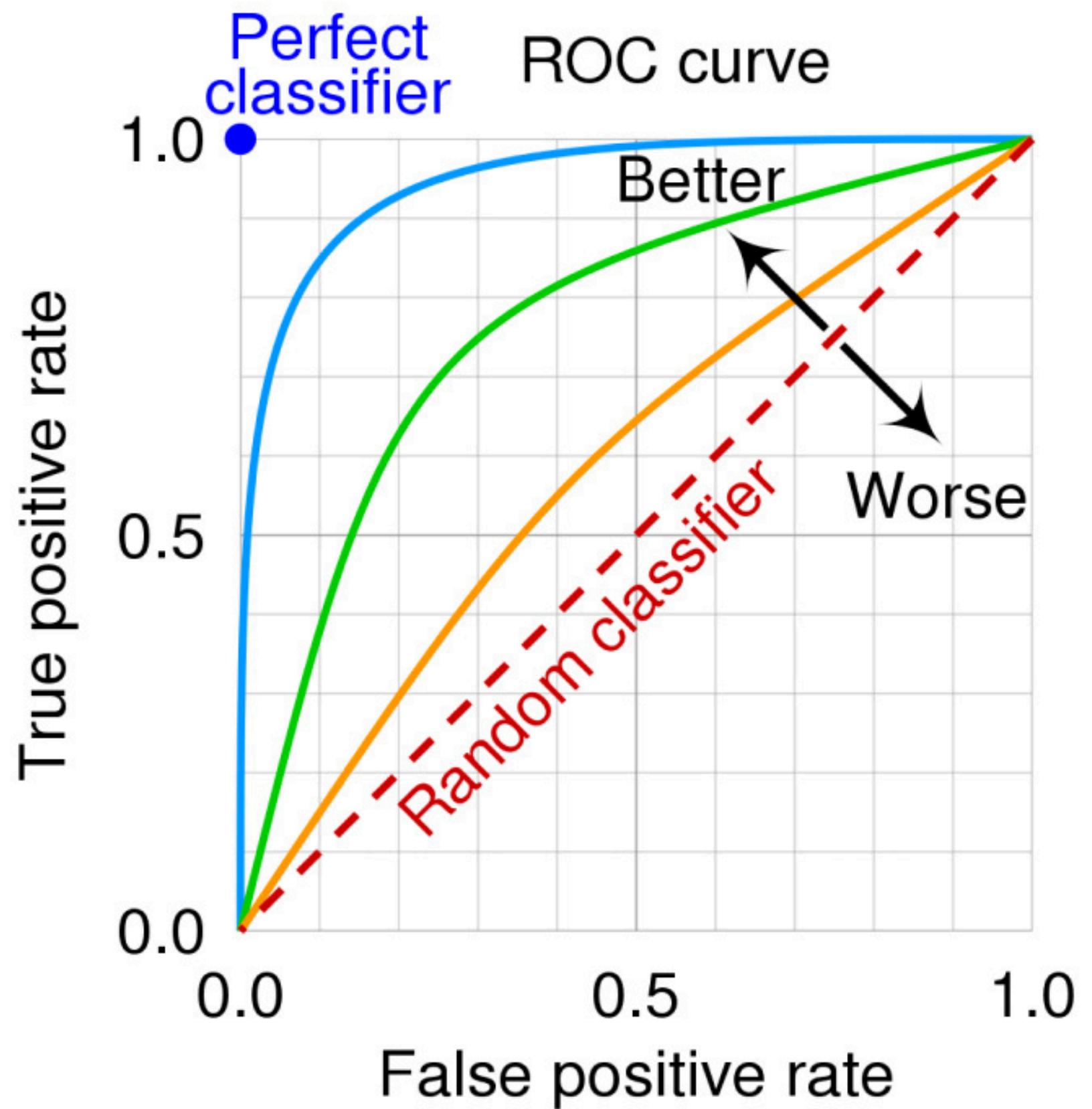
F2-Score

AUC-PR Curve

F1-Score

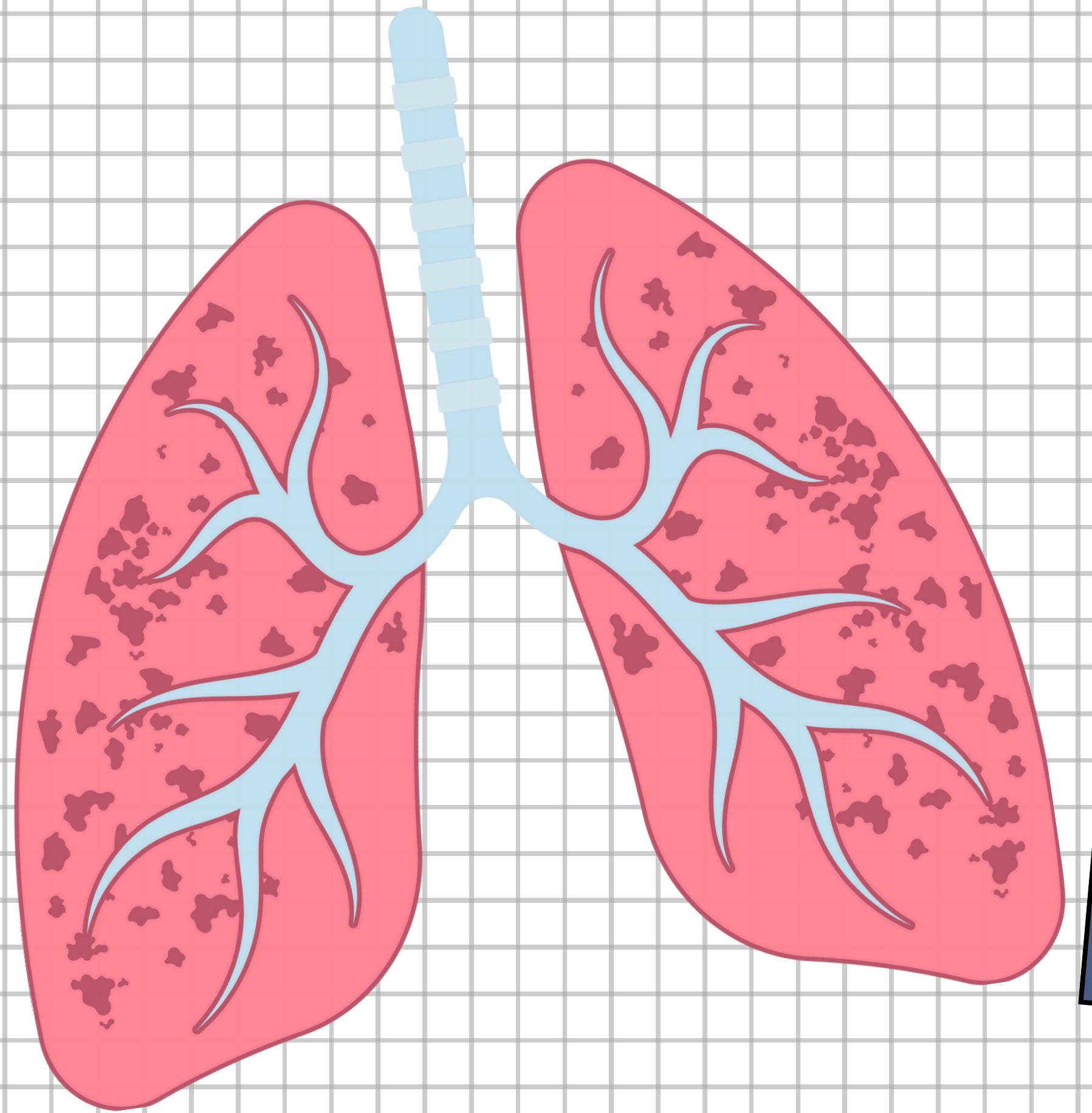
정밀도 & 재현율의 조화 평균
‘폐렴 환자를 놓치지 않으면서도, 정상 환자를
폐렴이라고 잘못 진단하는 경우가 적음을 판단’
“균형 잡힌 폐렴 진단 성능”





AUC-ROC Curve

가능한 모든 임계값에 대한 모델의 성능을 종합적으로 평가 => 단일 임계값에 의존하지 않으면서 모델의 전반적인 분류 성능을 종합적으로 판단
“전체적으로 폐렴 vs 정상 잘 가려내는 능력”



추가 지표 고려

데이터의 불균형

: 정상 데이터 1341개 / 폐렴 데이터
3875개

모델 학습 시 정상 클래스에 대한 예
측 성능 우려

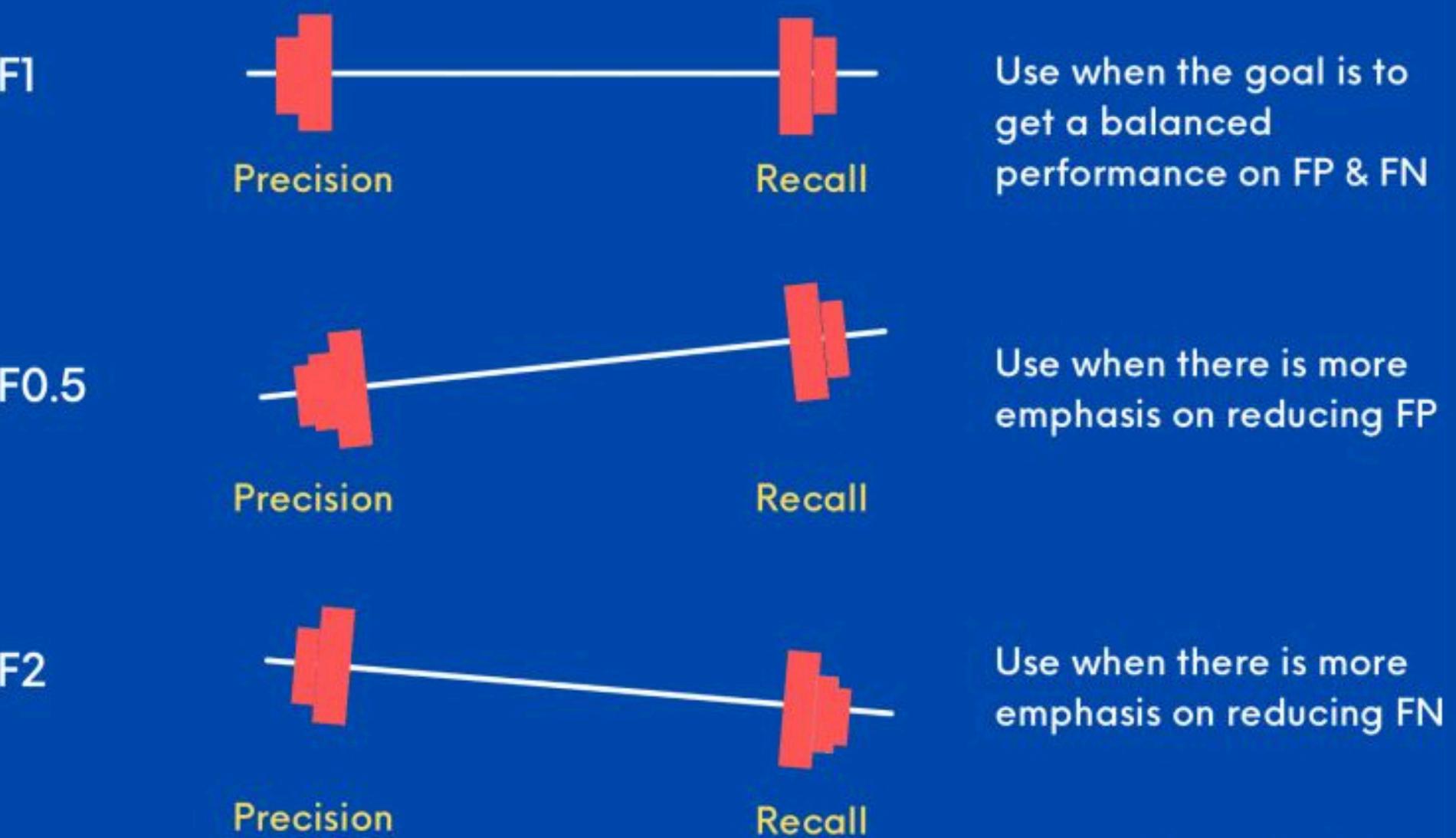
질병 탐지의 기능

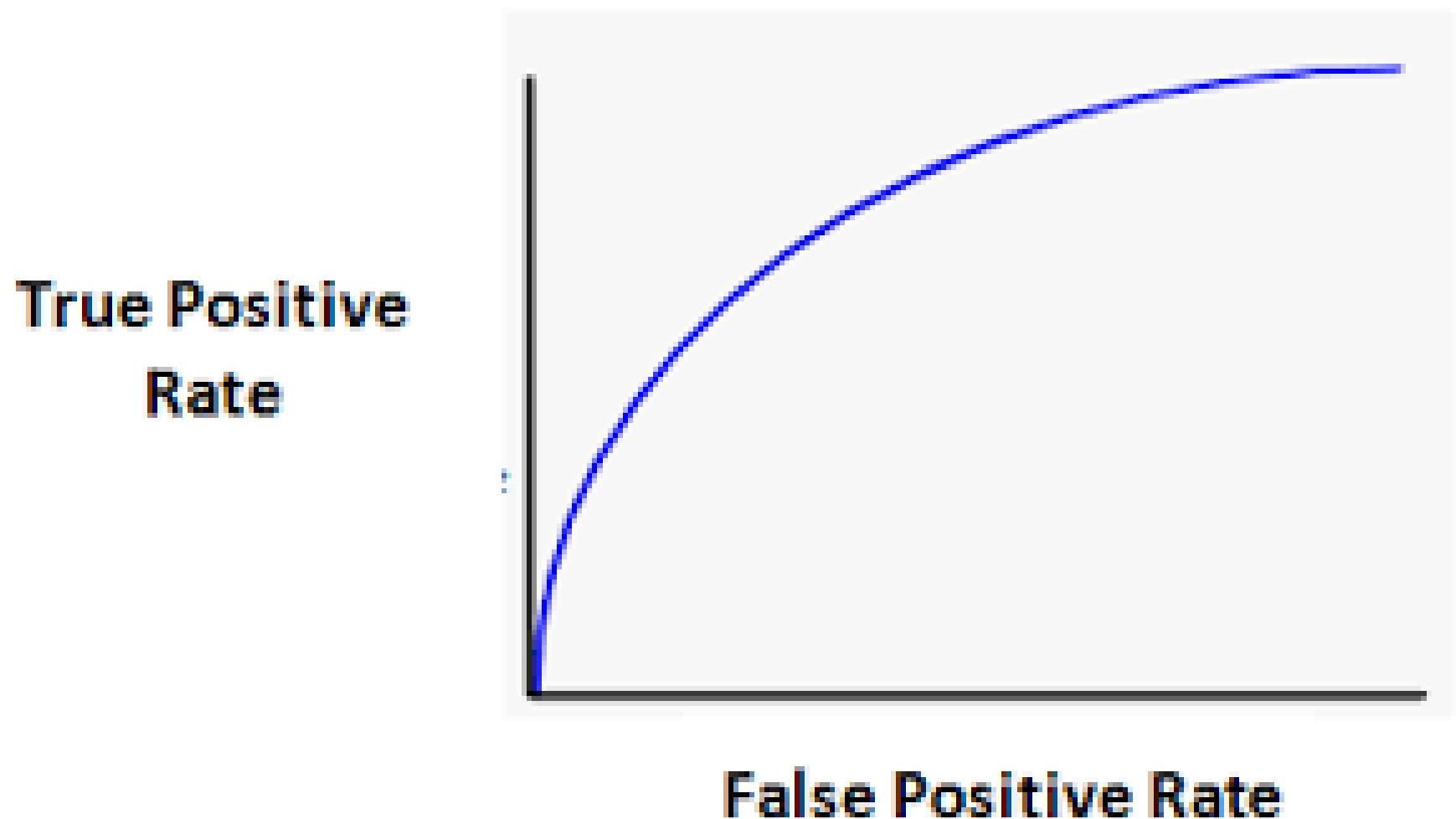
: 폐렴 환자를 놓치지 않는 것이 중요!
Recall의 값이 더 잘 반영되어야 함

F2-Score

F1-Score보다 Recall에 더 가중치를 둔 지표
높을수록 실제 폐렴 환자를 최대한 놓치지 않음
“폐렴 환자를 절대 놓치지 않는 능력”

F-Beta Score (F1 and Cousins)





PR-AUC Curve

ROC-AUC는 불균형 데이터셋에서 모델 성능을 과대평가 가능
임계값을 변화시킬 때의 정밀도와 재현율 사이의 관계,
클래스 불균형에 강하고, 실제 진단 상황과 더 밀접한 지표
“오진을 줄이면서도 환자를 놓치지 않는 능력을 평가”

데이터 전처리

데이터 증강

- 기존 이미지에 회전, 이동, 뒤집기 등 변형하여 학습 데이터의 다양성과 양을 증가
- 학습 중 실시간으로 적용하여, 데이터셋 크기를 늘리지 않음
- 데이터 다양성 증가로 모델 일반화 능력 향상 효과 기대

클래스 가중치 조정

- 가중치를 부여하여 소수 클래스 중요도 높여 학습 성능 향상
- 학습 중 손실 함수에 반영되어 소수 클래스를 더 잘 인식하도록 유도
- 불균형 데이터에서 민감도 개선 효과 기대

CNN MODELS

VGG16

- 단순한 구조 + 깊은 층으로 이미지 특징을 잘 잡는 모델
- 사진 속 작은 패턴(눈,코,입)을 잘 찾아냄

ResNet50

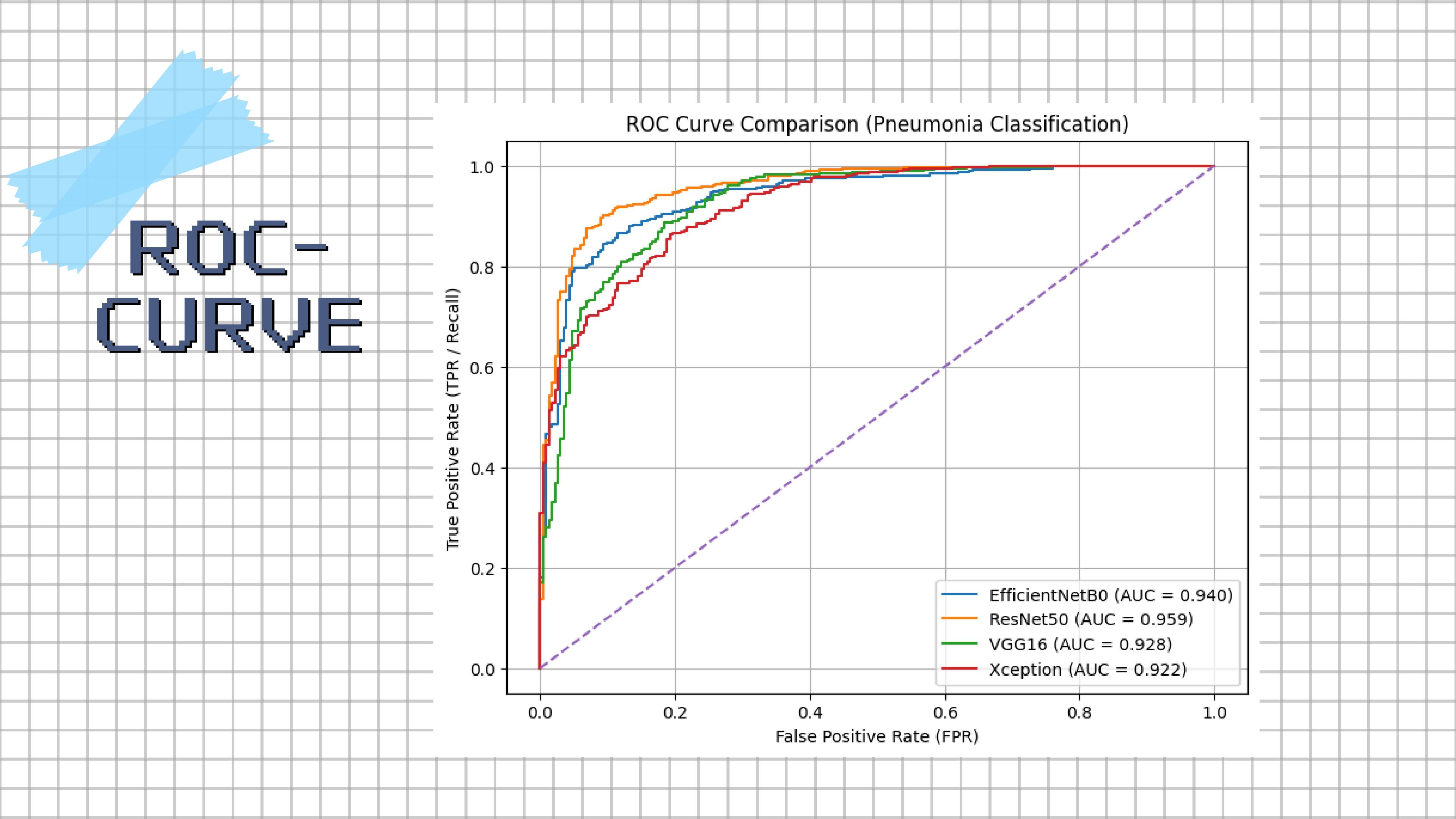
- Residual(잔차)연결로 층이 깊어져도 성능이 무너지지 않음
- 복잡한 패턴도 안정적으로 학습 가능 (의료 이미지 같은 복잡한 데이터에 강함)

Efficient NetB0

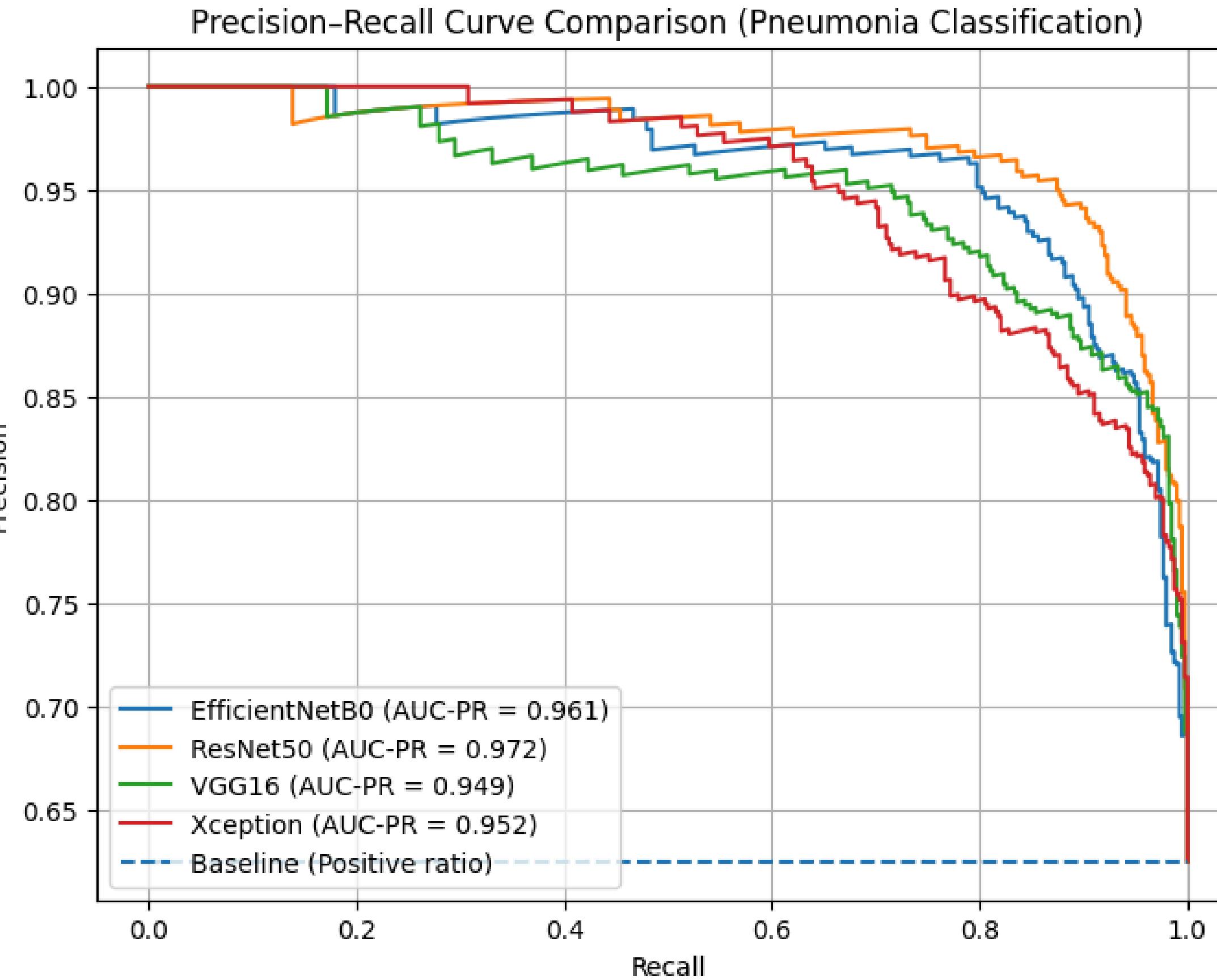
- 크기는 작지만 정확도는 높은 모델
- ‘복합 스케일링’ 방법을 통해 적은 파라미터로도 좋은 성능 달성

Xception

- 계산을 단순화해 빠르고 가벼운 모델
- 이미지 처리에서 속도를 중시할 때 유리함



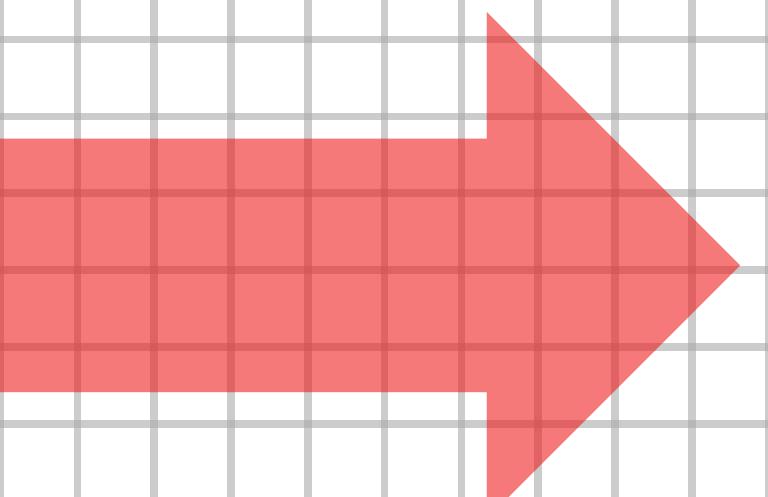
PR-
ROC
CURVE



모델 선정

EfficientNetB0 vs ResNet50 vs VGG16 vs Xception 성능 비교

지표	EfficientNetB0	ResNet50	VGG16	Xception
F1 Score	0.8942	0.9141	0.8973	0.6472
F2 Score	0.9291	0.9301	0.9230	0.5368
F0.5 Score	0.8619	0.8986	0.8730	0.8146
ROC AUC	0.94	0.96	0.93	0.92



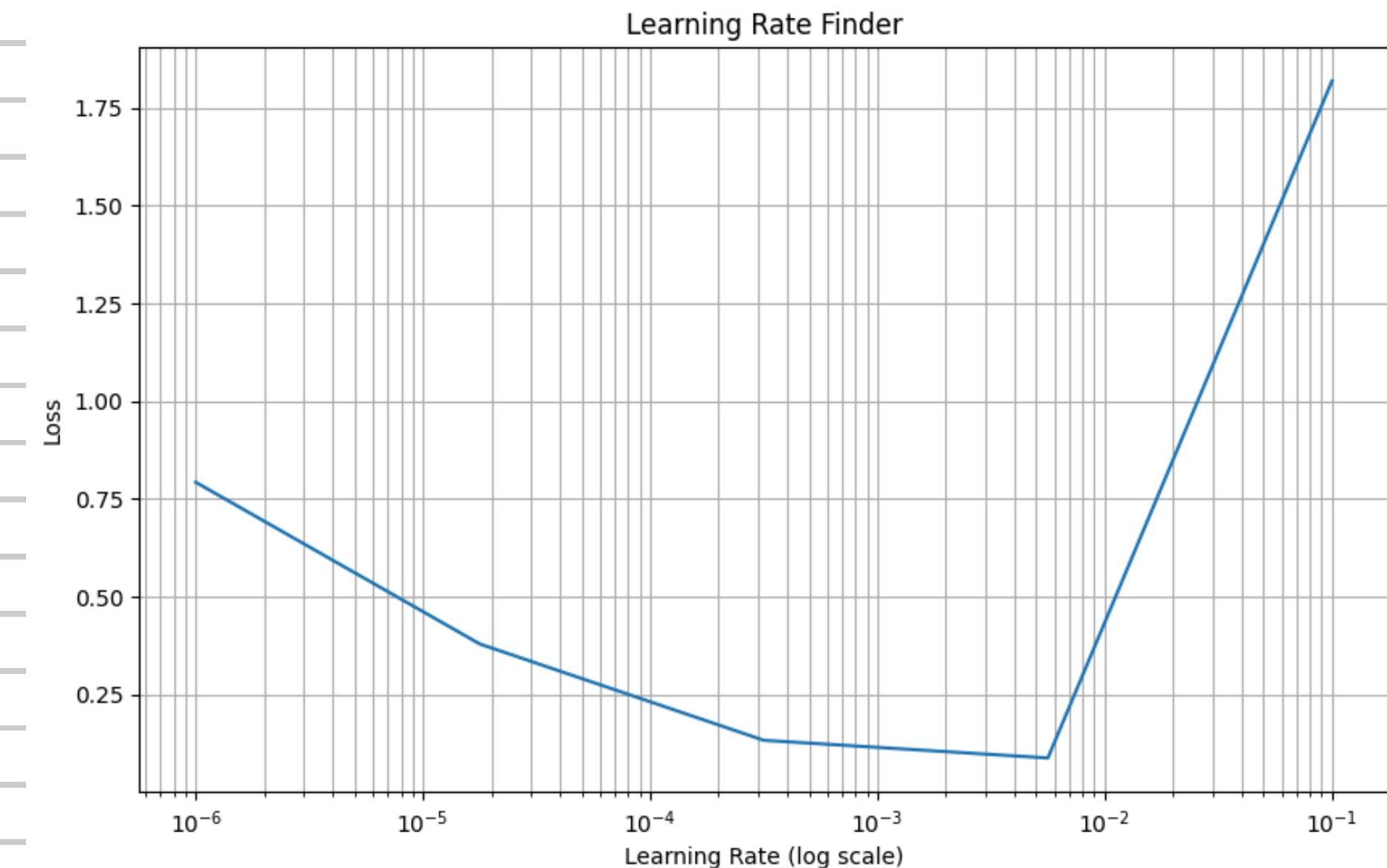
ROC Curve, P-R Curve,
성능지표 모두에서 가장 좋은 성능을 보
여준 ResNet50을 최종 모델로 선정

성능 개선 & 튜닝

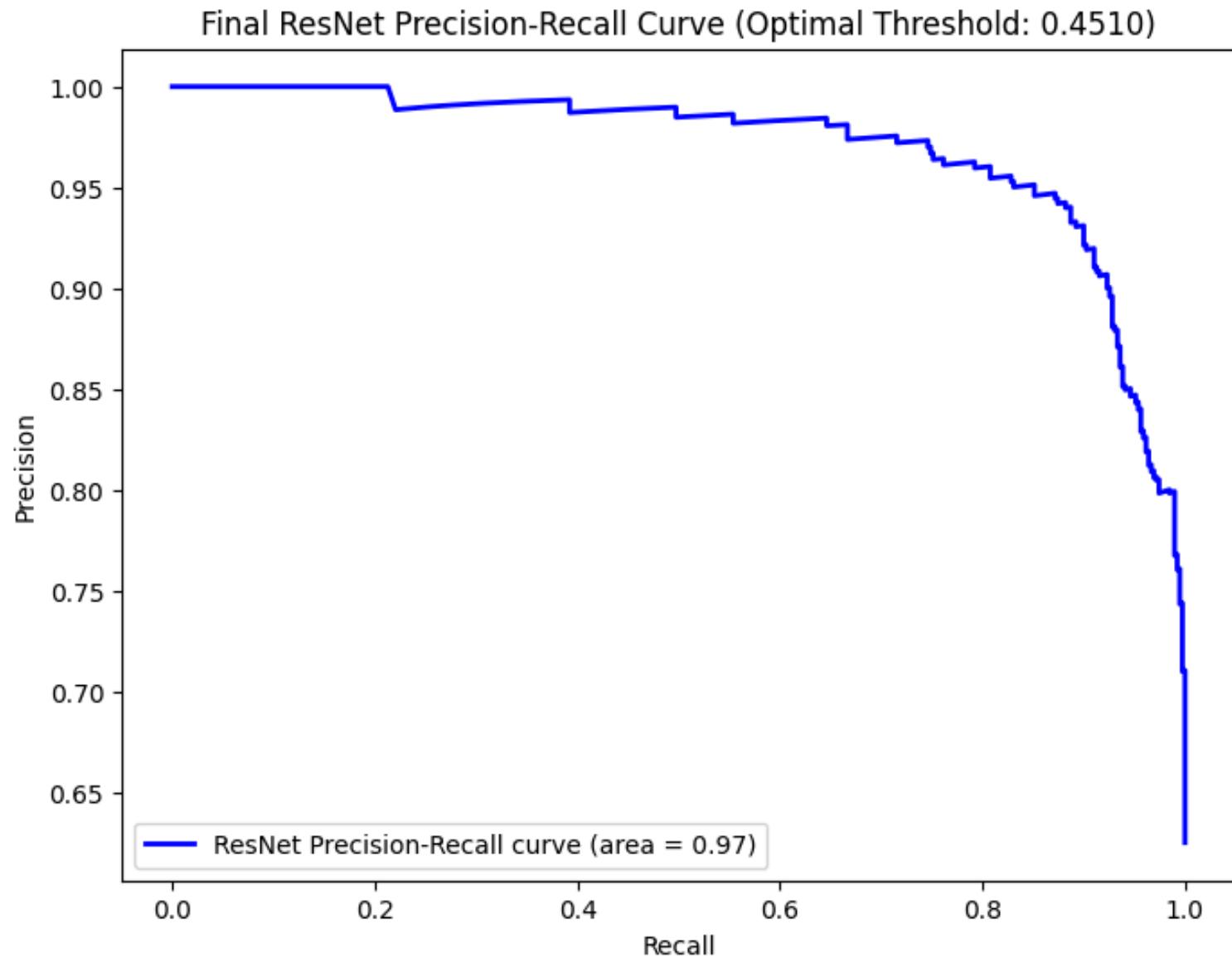
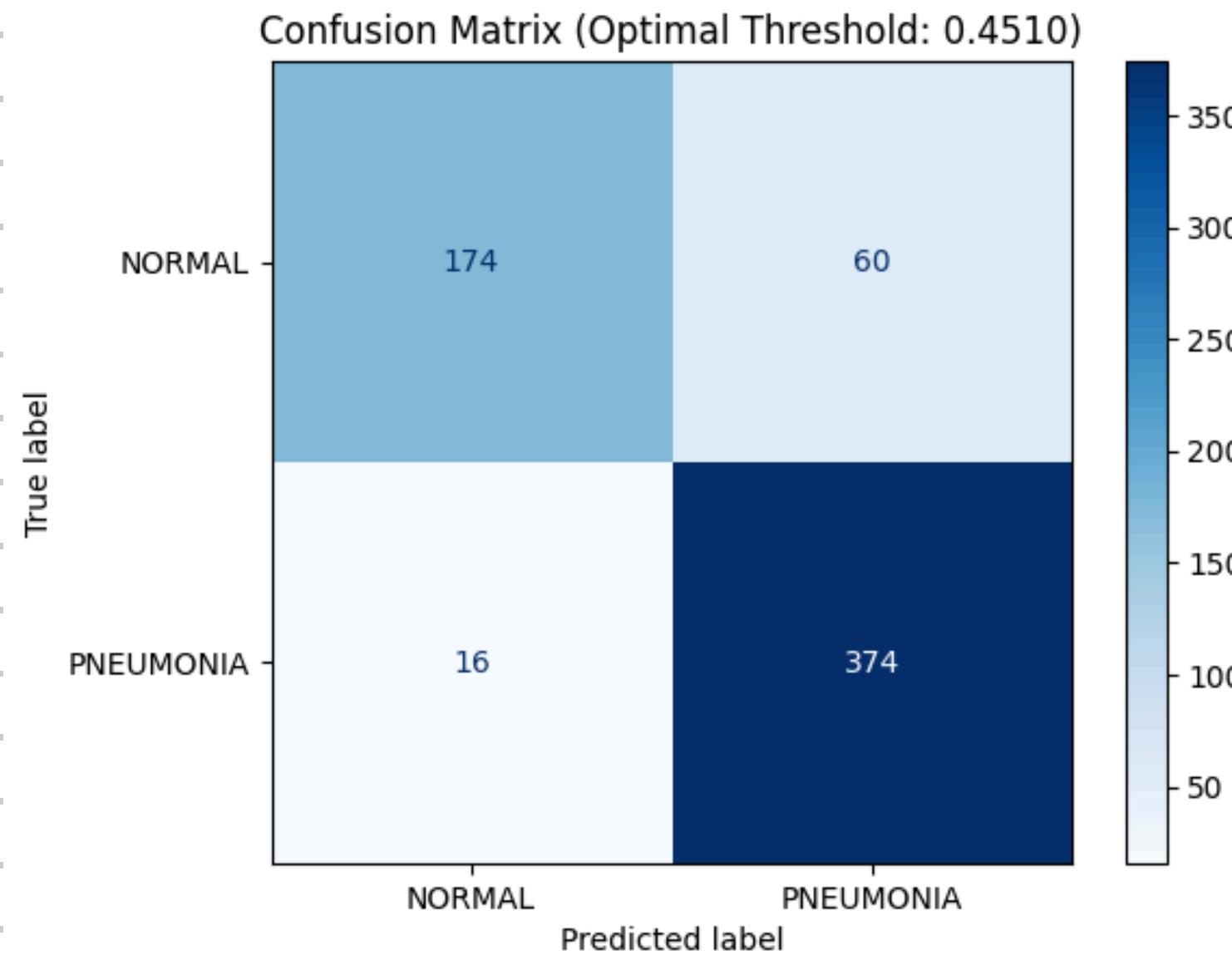
임계값 조정 : 0.3 ~ 0.5 사이의 값을 50개로
나누어 최적의 임계값을 찾음 => 0.451

학습률 조정 : $1e-6 \sim 1e-1$ 범위에서 Loss값이
가장 적은 최적의 학습률을 찾음 => $1e-3$

Threshold	Precision	Recall	F1 Score	F2 Score
0	0.300000	0.832230	0.966667	0.894425
1	0.304082	0.832230	0.966667	0.894425
2	0.308163	0.835920	0.966667	0.896552
3	0.312245	0.835920	0.966667	0.896552
4	0.316327	0.845291	0.966667	0.901914



결론



질병 예측 모델의 특성상 미탐지 오류(False Negative)를 최소화하는 것이 중요
다양한 딥러닝 모델을 비교해 ResNet모델을 선정, 데이터 증강과 클래스 가중치 전처리 기법의 시너지,
F2 스코어 등 다양한 지표들을 활용하여 예측 정확도를 유지하면서도 미탐지 오류를 효과적으로 줄일 수
있는 최적의 모델을 구현

회고

권태성

- 폐렴 이미지 분류 미니 프로젝트는 딥러닝 이론을 직접 적용하고 새로운 모델 아키텍처를 배우며 지식을 확장하는 소중한 경험이었습니다. 또한, 팀원들과 협업하여 스트레스 없이 팀워크 기술을 향상시키고, 실질적인 경험과 협업의 가치를 깨달을 수 있었던 의미 있는 시간이었습니다.

김현욱

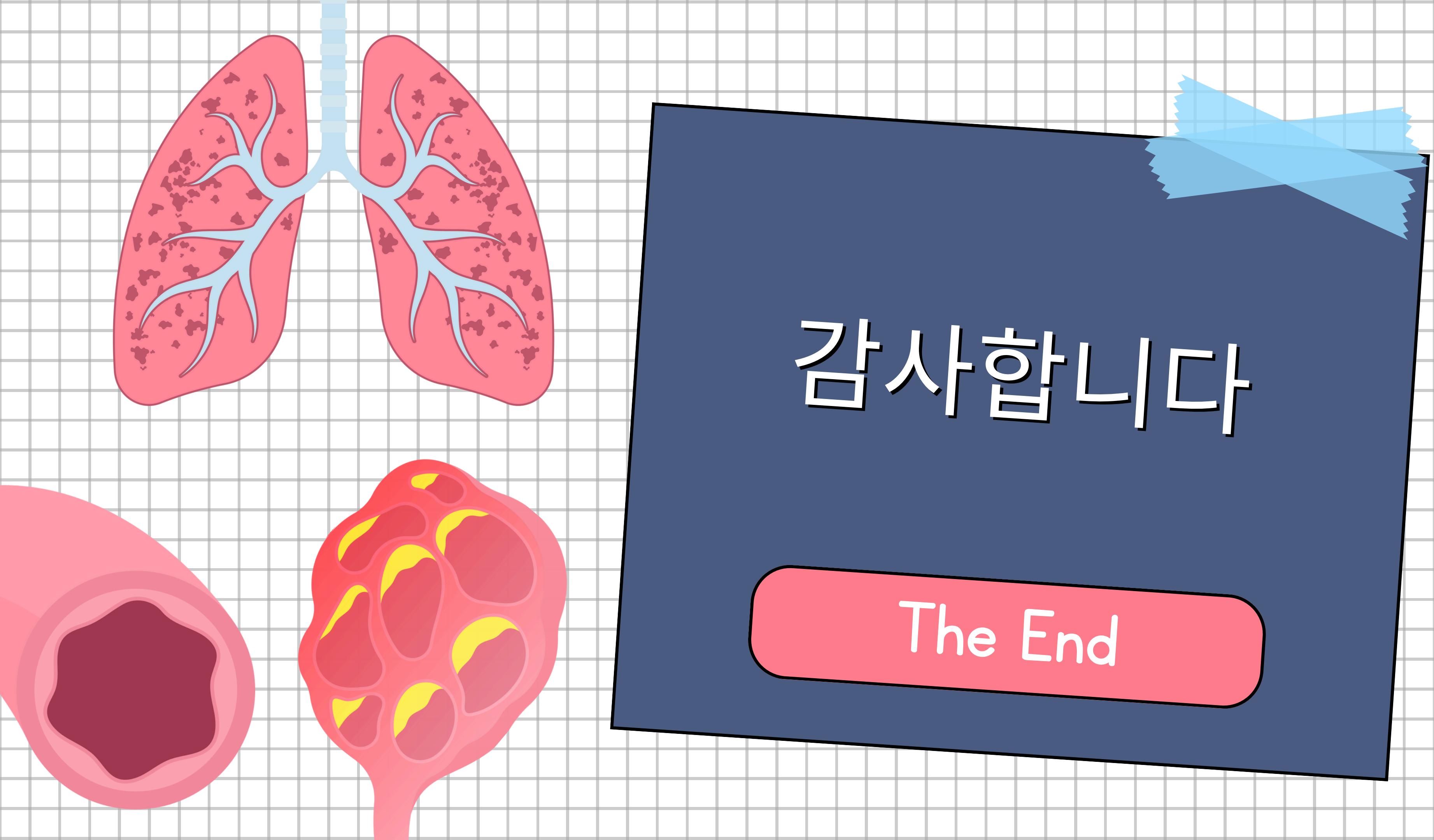
- 모델 선정 시에 가장 최신의 검증된 모델이 좋은 결과가 나올 것이란 제 예측이 틀렸고 이로 인해 데이터 특성과 용도에 따라 적합한 모델이 있음을 깨닫는 계기가 되었고 이론으로는 헷갈리던 혼동행렬과 지표들을 프로젝트 중 실제로 적용을 해보면서 이해에 큰 도움이 되었습니다. 의학 분야데이터를 더 잘 학습할 수 있는 모델 개발이 필요함을 느끼고 모델 개발에 흥미가 생기는 프로젝트였습니다.

나계훈

- 이번 프로젝트를 통해 의료 데이터에서 클래스 불균형과 데이터 부족 문제에 관하여 이해하게 됐습니다. 데이터 증강과 클래스 가중치 조정의 조합을 통해 모델 성능 향상을 할 수 있었습니다. 딥러닝 모델이더라도 전처리가 중요한 경우가 있다는 걸 이해하게 됐습니다. 앞으로도 다양한 전처리와 조합 전략을 적용할 계획입니다.

정안식

- 프로젝트를 진행하며 1종 오류(정상 환자를 폐렴으로 진단)와 2종 오류(폐렴 환자를 놓침) 사이의 균형을 맞추는 것이 어려웠습니다. 결국 F1 Score와 ROC AUC 같은 지표를 활용해 어떤 오류를 더 중요하게 다룰지 고민할 수 있었습니다.



감사합니다

The End