



# Code States Project 1

# 딥러닝을 이용한 피부암 진단 원격 의료 서비스

13기 – 14팀

오지혜

박윤아

김강호

안동선

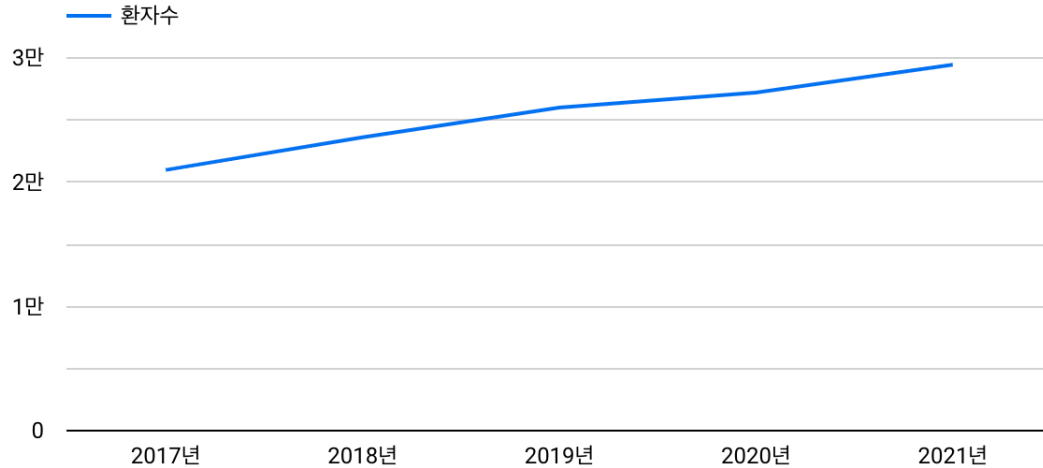
이준호

# 목차

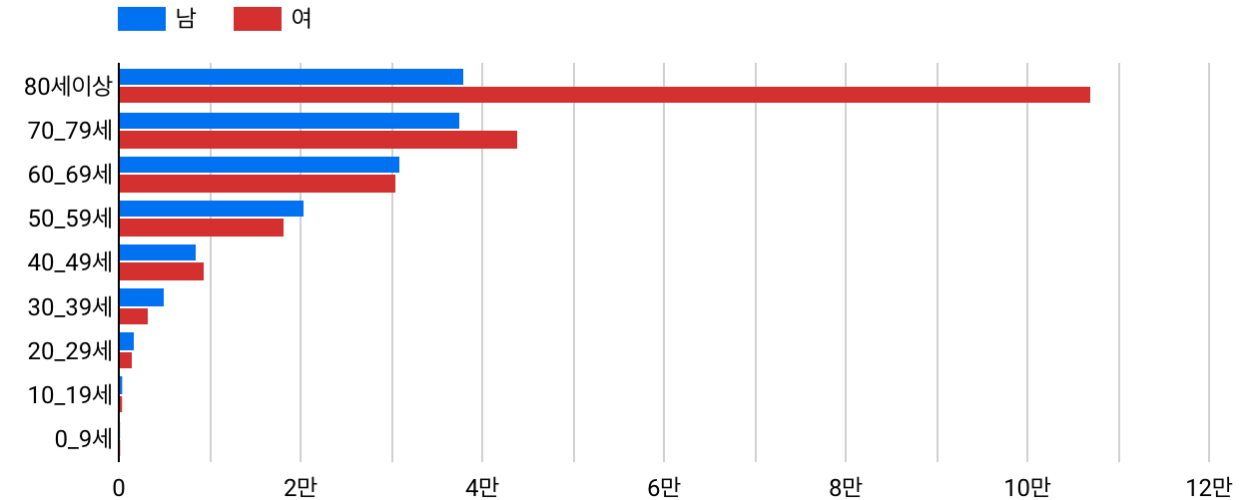
01. 프로젝트 개요
02. 프로젝트 팀 구성 및 역할
03. 프로젝트 수행 절차 및 방법
04. 프로젝트 수행 결과
05. 자체 평가 의견

# 01 프로젝트 개요 - 프로젝트 선정 배경

## 국내 피부암 환자 현황



<2017-2021 피부암 환자 수>

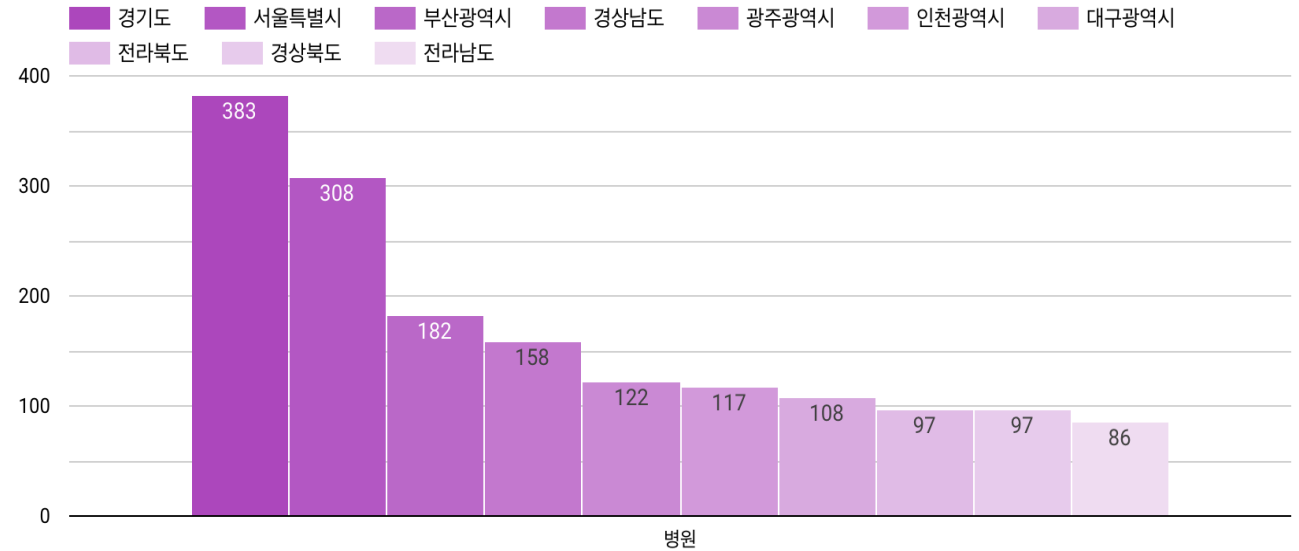
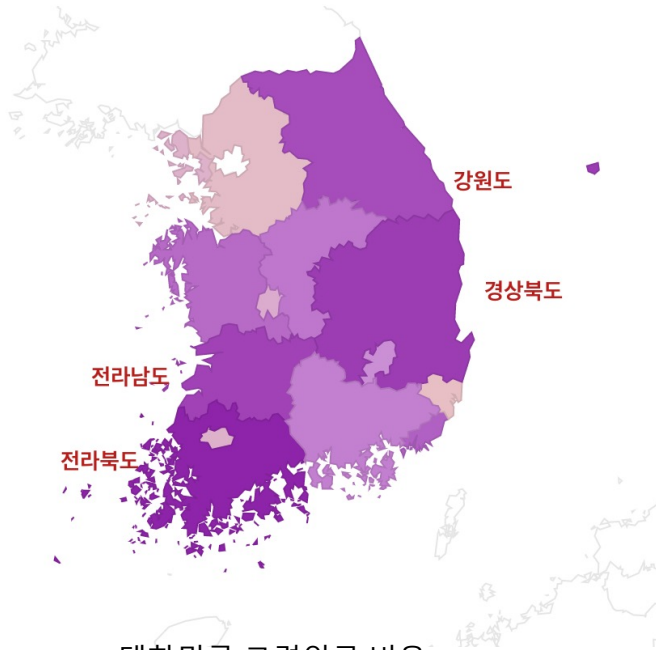


<2022년 피부암 연령대별 피부암 환자 내원일 수>

국내 피부암 환자는 매년 증가 추세를 보이고 있고,  
그중 50세 이상의 고령 환자 비율이 높으며 여자 환자 비율이 높다

# 01 프로젝트 개요 - 프로젝트 선정 배경

## 국내 의료 서비스 현황



<2022년 기준 전국 병원 분포>

### 고령인구 비율이 높은 순위

1. 전라남도
2. 경상북도
3. 전라북도

### 전국 병원 분포 낮은 순위

1. 전라남도
2. 경상북도
3. 전라북도



지방의료감소로 인해 고령인구가 많은 지방이 상대적으로 의료 혜택을 보기 어려운 상황

# 01 프로젝트 개요 - 프로젝트 서비스 기획

## 비대면 진료 앱 후기 분석

	word	weight ▾	frequency
1.	비대면	<div><div></div></div>	<div><div></div></div>
2.	기록을	<div><div></div></div>	<div><div></div></div>
3.	진료기록	<div><div></div></div>	<div><div></div></div>
4.	비대면진료	<div><div></div></div>	<div><div></div></div>
5.	건강관리	<div><div></div></div>	<div><div></div></div>
6.	가입도	<div><div></div></div>	<div><div></div></div>
7.	건강기록을	<div><div></div></div>	<div><div></div></div>
8.	병원이	<div><div></div></div>	<div><div></div></div>

<3점 이상 후기 가중치, 빈도>

	word	weight ▾	frequency
1.	비대면	<div><div></div></div>	<div><div></div></div>
2.	처방전	<div><div></div></div>	<div><div></div></div>
3.	병원이	<div><div></div></div>	<div><div></div></div>
4.	코로나	<div><div></div></div>	<div><div></div></div>
5.	고객센터	<div><div></div></div>	<div><div></div></div>
6.	내원	<div><div></div></div>	<div><div></div></div>
7.	카톡	<div><div></div></div>	<div><div></div></div>
8.	배송이	<div><div></div></div>	<div><div></div></div>

<3점 이하 후기 가중치, 빈도>

## 긍정 평가와 사용자 반응

- 비대면 진료가 편리함
- 수도권외의 경우 연결된 병원이 많아서 좋음
- 지방의 경우에도 수도권 병원에서 비대면 진료를 볼수 있어 편리
- 거동이 불편하신 분들이 원격 진료를 볼 수 있어 편리

## 부정평가와 사용자 반응

- 지방의 경우 연결된 병원이 없어서 불편함
- 비대면 진료를 신청했는데 병원에서는 내원 하라고 함
- 거주지역외에 타지역 병원과 연결되어 내원 이 불편함

# 01 프로젝트 개요- 웹 크롤링수행

## ① 원격 의료앱 크롤링



### 사용 기술 및 한계점

- Python Selenium, Chromedriver
- 동적 데이터를 자동화하여 가져올 수 있는 프레임 워크와 Chrome 브라우저를 제어하기 위한 드라이버를 사용하여 크롤링 진행

# 크롤 드라이버 세팅

```
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')
chrome_options.add_argument('--disable-blink-features=AutomationControlled')
driver = webdriver.Chrome(chrome_driver, options = chrome_options)
```

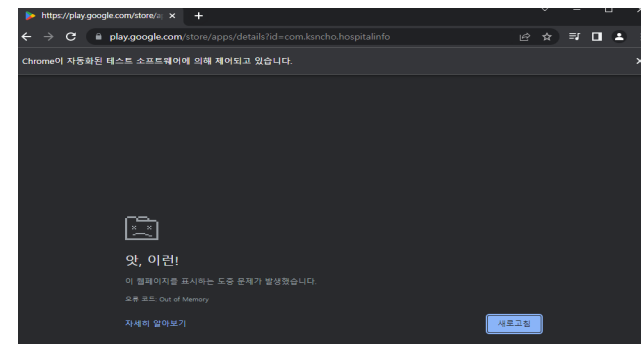
- 데이터를 가져오기 위한 클래스 생성, 배열, 데이터 추출 작업 진행

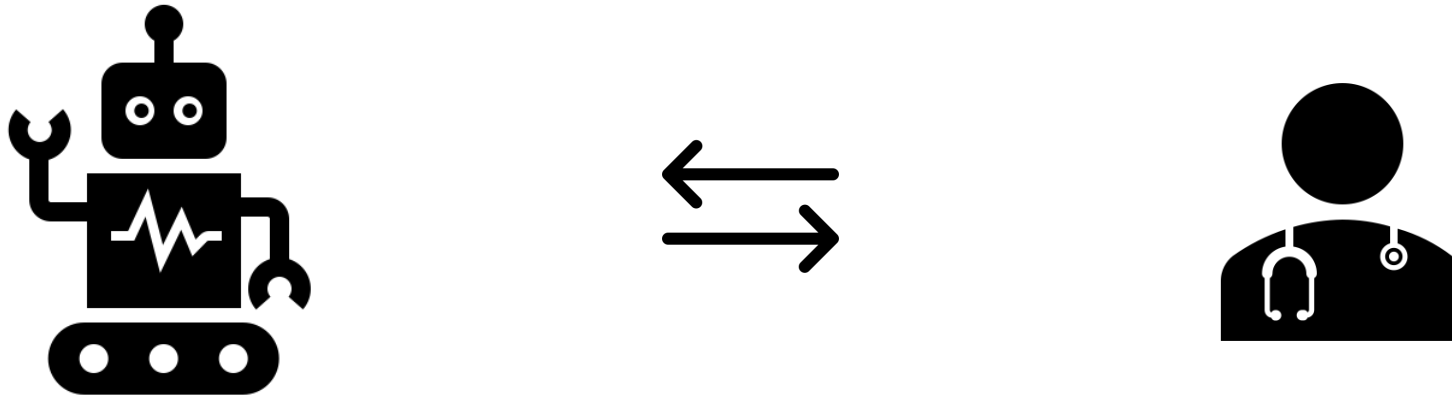
```
# 리뷰 데이터 클래스 접근
review_source = soup_source.find_all(class_ = 'RHolpe')
# 리뷰 데이터 저장용 배열
dataset = []
# 데이터 넘버링을 위한 변수
review_num = 0
# 리뷰 1개씩 접근해 정보 추출
for review in review_source:
    review_num += 1
    # 리뷰 등록일 데이터 추출
    date_full = review.find_all(class_ = 'bp9Aid')[0].text
    date_year = date_full[0:4] # 연도 데이터 추출
    # 해당 단어가 등장한 인덱스 추출
    year_index = date_full.find('년')
    month_index = date_full.find('월')
    day_index = date_full.find('일')
    # 월(Month) 데이터 추출
    date_month = str(int(date_full[year_index+1:month_index]))
```

- Selenium을 활용한 작업이 눈에 보이는 콘텐츠는 모두 크롤링이 가능하지만 브라우저를 활용하는 기술이기에 컴퓨터 자원을 많이 소비

```
import selenium
import requests
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from bs4 import BeautifulSoup
import time
from time import sleep
import random
import pandas as pd
```

- Request 라이브러리를 통해 속도 측면을 보완했지만 데이터 크기가 많이 큰 경우 컴퓨터 메모리 다운 현상 발생



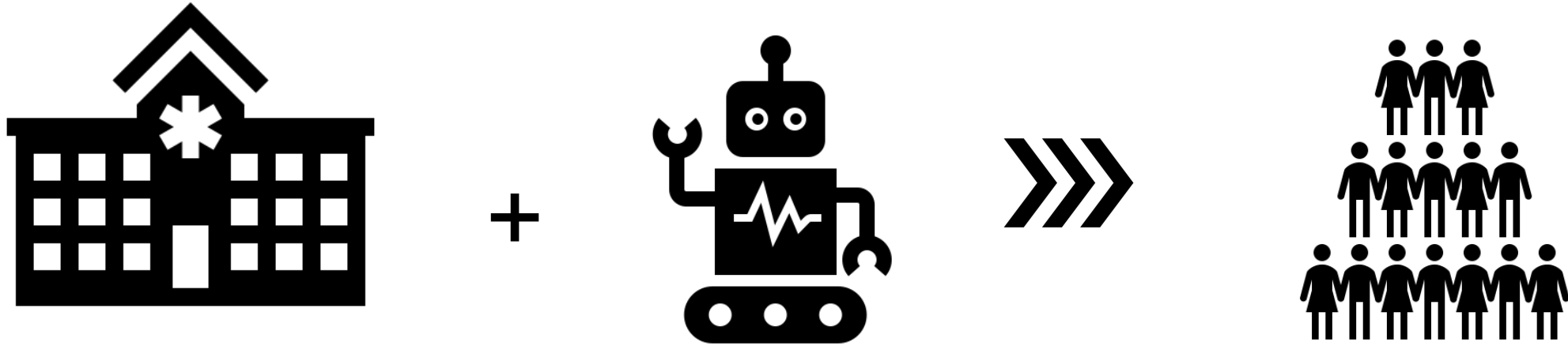


### 딥러닝을 이용한 의료 서비스 사례: 암진단 모델 왓슨

- 클라우드 기반 플랫폼으로 방대한 데이터를 분석
- 암환자 치료 확률을 높이기 위해 활용
- 의사들에게 치료법 권고 및 조언 가능

→ 실제 길병원, 부산대 병원, 전남대 병원에서 이용되었음





## 프로젝트 주제: 딥러닝을 이용한 피부암 진단 원격의료 서비스

- 100 % 원격 진료 서비스로 수도권 병원의 서비스를 의료 취약 계층 분들이 누릴 수 있게 하자
- 피부암과 같은 자가진단이 어려운 질병을 딥러닝 머신 구현을 통해 간단하고 정확히 진단 받도록 하자

## 02

## 프로젝트 팀 구성 및 역할

훈련생	역할	담당 업무
오지혜	DA	<ul style="list-style-type: none"> <li>▶ 데이터 수집 및 분석</li> <li>▶ 서비스 시스템 기획</li> <li>▶ 웹페이지 시나리오 작성</li> <li>▶ 발표자료 및 발표진행</li> </ul>
박윤아	DA	<ul style="list-style-type: none"> <li>▶ 모델 분석 주제 제시</li> <li>▶ 데이터 수집 및 분석</li> <li>▶ 서비스 시스템 기획</li> </ul>
김강호	DS DE	<ul style="list-style-type: none"> <li>▶ 프로젝트 기획서, ds파트 발표자료 취합</li> <li>▶ 사전학습 모델을 활용한 딥러닝 모델 제작 및 최적화</li> <li>▶ 웹 앱 구성 및 취합(백엔드 기능 일부 구현)</li> </ul>
안동선	DS DE	<ul style="list-style-type: none"> <li>▶ 프로젝트 기획서 초안 작성</li> <li>▶ 데이터셋 수집 및 라벨링</li> <li>▶ 다양한 모델을 fine tuning 후 비교</li> <li>▶ 앱스토어 후기 크롤링</li> <li>▶ 웹앱 구현</li> </ul>
이준호	DE	<ul style="list-style-type: none"> <li>▶ 원격의료 앱 데이터 크롤링</li> <li>▶ 웹페이지 구현(프론트 엔드)</li> <li>▶ 웹 페이지 시연</li> </ul>

## 03 프로젝트 수행 절차 및 방법

### 사전기획

8/30 - 9/1

- 프로젝트 주제 선정 및 기획
- 기획안 초안 작성

### 데이터 분석

9/2 - 9/8

- 서비스 기획 데이터 분석
- 대시보드 작성

### 서비스 기획

9/7- 9/8

- 전체 서비스 구조 기획
- 웹 페이지 시나리오 작성

### 데이터 수집 및 선정

8/31 - 9/2

- 모델링 데이터 수집
- 모델링 데이터 선정
- 서비스 기획 데이터 수집
- 활용 방안 논의

### 모델링

9/2 - 9/7

- 전이학습 모델 선정
- Fine tuning 후 성능 비교
- 배치 정규화 & Dropout 적용
- 분류기 구조 변경
- 모델 작동 시간 측정

### 웹 서비스 구축

9/8 - 9/13

- 디자인 및 기능 기획
- 부트스트랩 적용
- 웹 구현

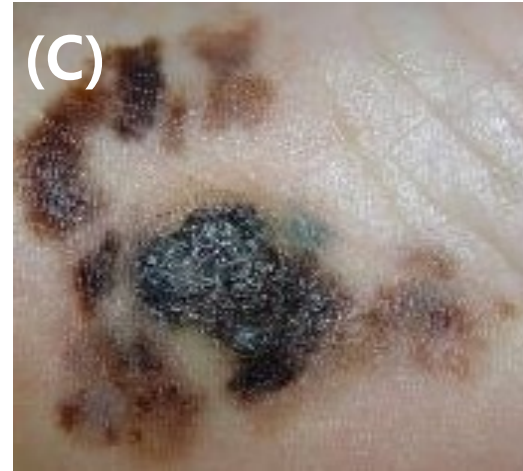
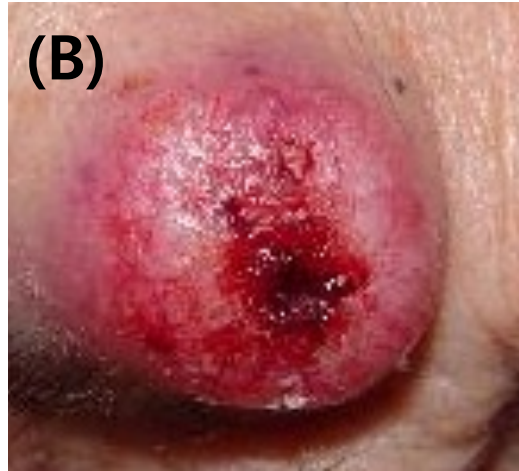
# 04

## 프로젝트 수행 결과

결과 제시 ① 탐색적 분석 및 전처리



학습 데이터 소개 (Train/dev set)



(A). 기저세포암

(B). 편평세포암

(C). 흑색종 악성

(D). 흑색종 양성

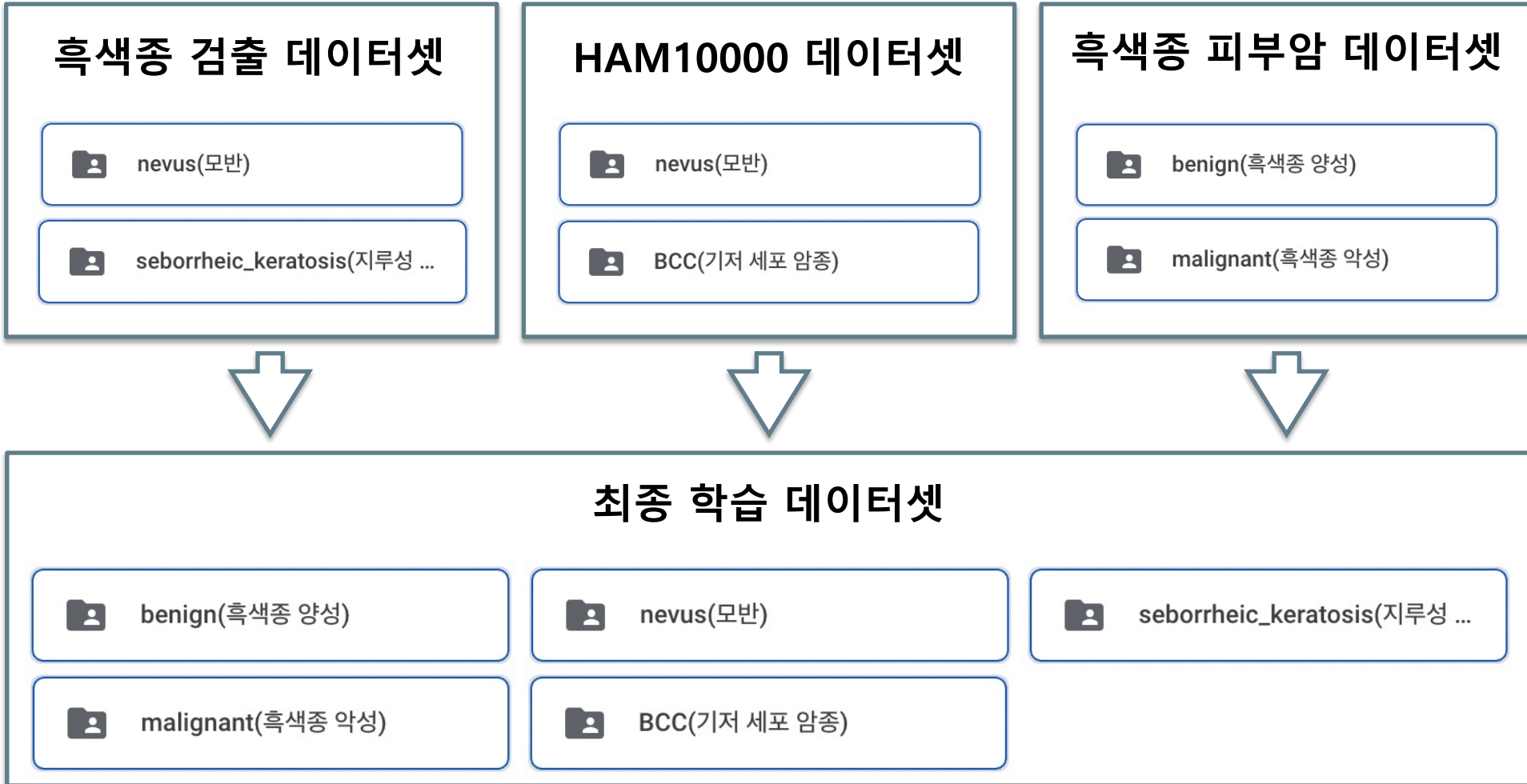
(E). 모반

## 04 프로젝트 수행 결과

결과 제시 ① 탐색적 분석 및 전처리



### 학습 데이터 소개 (Train/dev set)



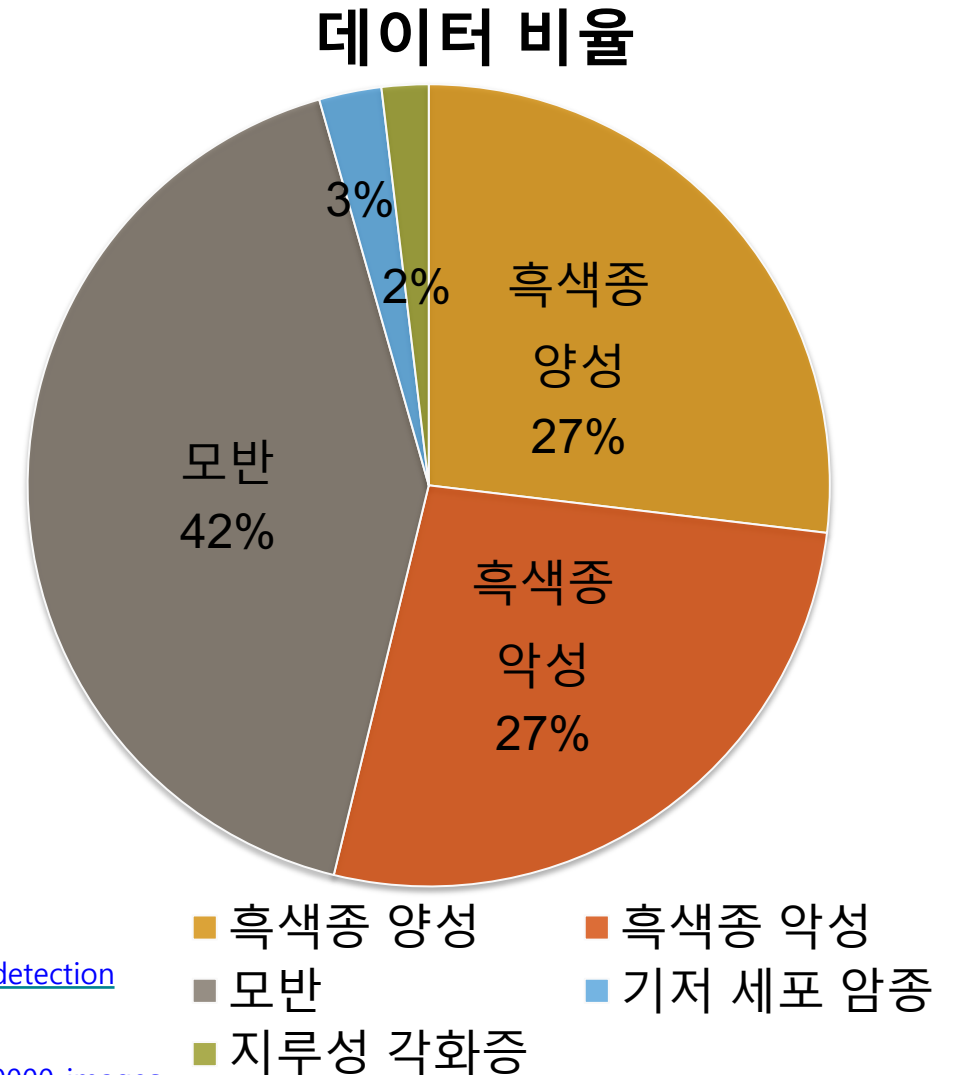
## 04 프로젝트 수행 결과

결과 제시 ① 탐색적 분석 및 전처리



학습 데이터 소개 (Train/dev set)

상태	데이터 개수(개)
흑색종 양성	5515
흑색종 악성	5119
모반	8570
기저 세포 암종	516
지루성 각화증	387



데이터 출처

- 흑색종 검출 데이터셋: <https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection>

- HAM10000 데이터셋: <https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification?select=masks>

- 흑색종 피부암 데이터셋: <https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images>

## 04 프로젝트 수행 결과

### 결과 제시 ② 모델 비교



#### 사전학습 모델 성능 비교

- 동일한 학습데이터를 적용해서 사전학습 모델을 Fine tuning 했다.
- 모델크기와 모델의 분류성능지표(Accuracy, precision, recall)를 기준으로 모델을 선정했다.

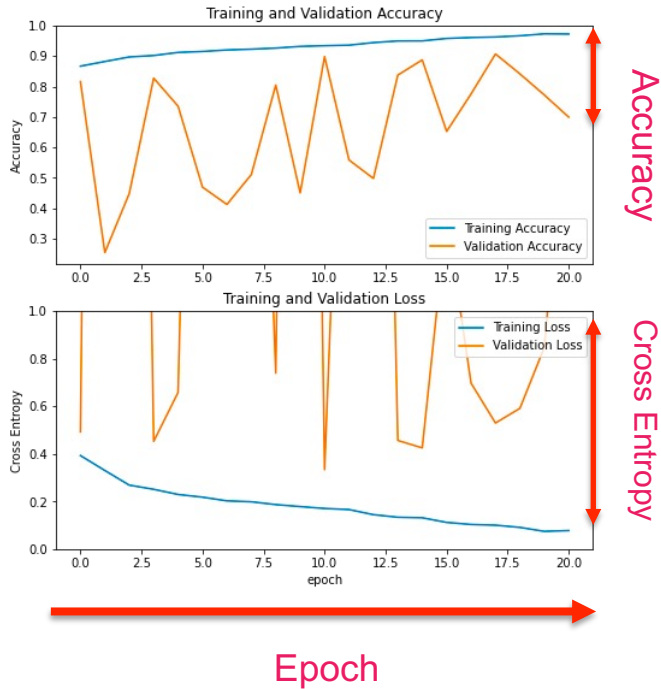
모델	모델 크기 (MB)	loss	accuracy	precision	recall
ResNet50	316.7	0.3332	0.8983	0.9067	0.8868
InceptionResNetV2	681.8	0.1715	0.9337	0.9357	0.9287
EfficientNetV2M	222.4	0.1527	0.9491	-	-
EfficientNetV2S	268.9	0.1452	0.9521	0.9521	0.9521
EfficientNetV2B3	183.1	0.1406	0.9451	0.9455	0.9431
EfficientNetV2B1	107.9	0.1387	0.9501	0.9511	0.9501
EfficientNetB1	82.1	0.227	0.9462	0.9462	0.9541

# 04 프로젝트 수행 결과

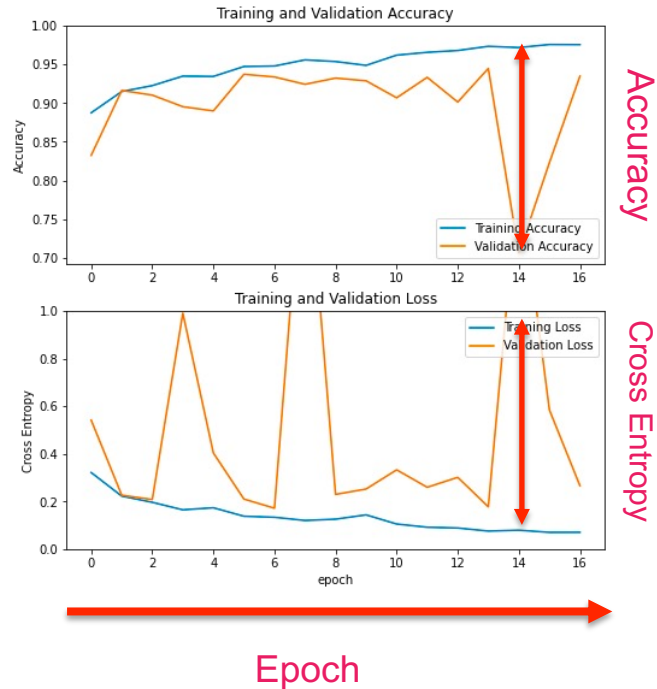
## 결과 제시 ③ 모델 평가 및 개선



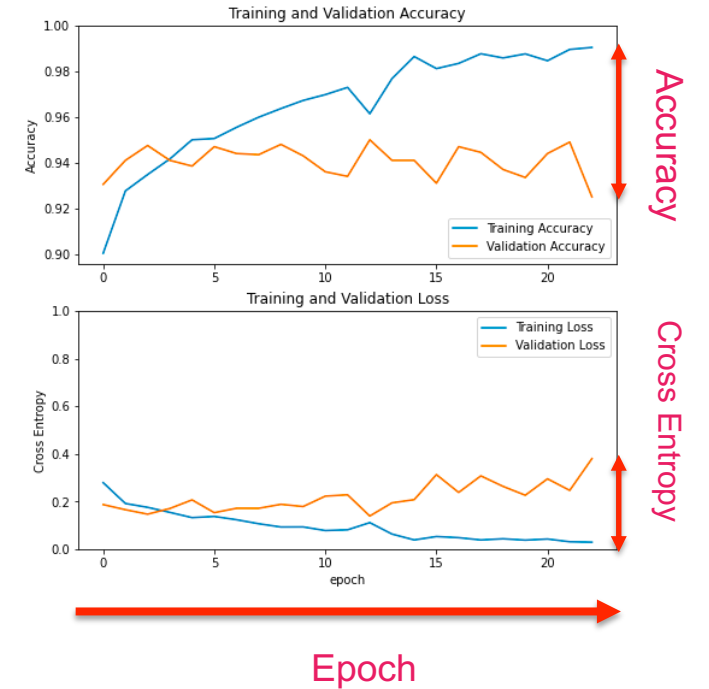
### ResNet50



### InceptionResNetV2



### EfficientNetV2B1



에폭이 증가할수록 학습이 불안정하게 진행된다. → 배치 정규화

학습이 진행 될수록 과적합이 심해진다. → Dropout & Early Stopping



# 04 프로젝트 수행 결과

결과 제시 ③ 모델 평가 및 개선



## 모델 선정 이유

	안동선	김강호
선정모델	EfficientNetV2B1	EfficientNetB1
선정이유 및 과정	<ul style="list-style-type: none"><li>- 여러가지 사전학습 모델을 사용해 보고 학습데이터 셋과 잘 맞는 모델을 선정</li><li>- 모델경량화 + 빠른 학습을 위해 만들어진 모델</li><li>- 다른 모델들 보다 파라미터의 개수가 작으면서 더 뛰어난 예측성능을 보임</li></ul>	<ul style="list-style-type: none"><li>- 모델의 파라미터 개수가 작아 경량화와 성능 모두 만족가능</li><li>- 헬스케어/의료 분야에서의 이미지 분류 성능이 상위권[1], [2]</li><li>- EfficientNetB1관련 참고자료가 많았음</li></ul>

참고논문 :

[1] Khan, E.; Rehman, M.Z.U.; Ahmed, F.; Alfouzan, F.A.; Alzahrani, N.M.; Ahmad, J. Chest X-ray Classification for the Detection of COVID-19 Using Deep Learning Techniques. *Sensors* 2022, 22, 1211. <https://doi.org/10.3390/s22031211>

[2] Ravi, V., Acharya, V. & Alazab, M. A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. *Cluster Comput* (2022). <https://doi.org/10.1007/s10586-022-03664-6>

## 04 프로젝트 수행 결과

결과 제시 ③ 모델 평가 및 개선



최종 모델 구성

	안동선	김강호
데이터 전처리	<ul style="list-style-type: none"><li>- Train/test set(9:1) 분리</li><li>- Batch size : 32</li><li>- 이미지 사이즈 : 256 * 256 으로 통일</li></ul>	<ul style="list-style-type: none"><li>- Train/test set(9:1) 분리</li><li>- Batch size : 4</li><li>- 이미지 사이즈 : 256 * 256 * 3</li><li>- 불균형 데이터의 랜덤 오버샘플링</li><li>- 데이터 증강(좌우반전)</li></ul>
학습모델 구성	<ul style="list-style-type: none"><li>- Base_model : EfficientNetV2B1 사용</li><li>- BatchNormalization</li><li>- Dropout(0.3) 적용</li><li>- EarlyStopping 적용</li><li>- 출력층 : Softmax 함수(다중분류) 적용</li></ul>	<ul style="list-style-type: none"><li>- Base_model : EfficientNetB1 선정</li><li>- BatchNormalization</li><li>- Regularizer<ul style="list-style-type: none"><li>&gt; 가중치 감소법 적용</li></ul></li><li>- Dropout(0.45) 적용</li><li>- 출력층 : Softmax 함수(다중분류) 적용</li></ul>

# 04 프로젝트 수행 결과

결과 제시 ④ 최종 모델 분석



## 분류 결과(Confusion Matrix)

모델 예측분류 수			A	B	C	D	E
실제분류	A	안동선	480	26	0	0	0
		김강호	264	8	0	0	0
	B	안동선	32	526	0	0	0
		김강호	21	253	0	0	0
	C	안동선	0	0	834	7	11
		김강호	0	0	397	7	15
	D	안동선	0	0	4	41	0
		김강호	0	0	2	16	0
	E	안동선	0	0	14	0	30
		김강호	0	0	1	0	20

A : 흑색종 양성

B : 흑색종 악성

C : 모반

D : 기저세포 암종

E : 지루성 각화증

@ 최종사용 모델

- 안동선 : EfficientNetV2B1

>> 이미지 2005개 예측

- 김강호 : EfficientNetB1

>> 이미지 1004개 예측

## 04 프로젝트 수행 결과

결과 제시 ④ 최종 모델 분석



### 최종 모델 성능

		Precision	Recall	F1-score	Support
A	안동선	0.9375	0.9486	0.9430	506
	김강호	0.9263	0.9706	0.9479	272
B	안동선	0.9529	0.9427	0.9477	558
	김강호	0.9693	0.9234	0.9458	274
C	안동선	0.9789	0.9789	0.9789	852
	김강호	0.9925	0.9475	0.9695	419
D	안동선	0.8542	0.9111	0.8817	45
	김강호	0.6957	0.8889	0.7805	18
E	안동선	0.7317	0.6818	0.7059	44
	김강호	0.5714	0.9524	0.7143	21
안동선		Accuracy : 0.9531			2005
김강호		Accuracy : 0.9462			1004

A : 흑색종 양성  
B : 흑색종 악성  
C : 모반  
D : 기저세포 암종  
E : 지루성 각화증

@ 최종사용 모델  
- 안동선 : EfficientNetV2B1  
- 김강호 : EfficientNetB1

## 05 자체 평가 의견

오지혜

- 데이터 수집 과정에서 어떤 데이터가 유용한지 감을 잡게 되었다
- 데이터 분석을 통해 프로젝트 아이디어 구상과 서비스 기획까지 해볼 수 있어 직무적으로 좋은 경험이었다.
- 팀원들과 소통을 통해 전체적인 프로젝트 파이프 라인을 배울 수 있었다
- 시간 부족과 컴퓨터 환경 부족으로 서비스 기획에 맞춰 서비스 구현을 하지 못한 것이 아쉽다

박윤아

- 전체적인 프로젝트를 기획해보면서 프로젝트 프로세스 이해도를 높일 수 있어서 좋았습니다.
- 부족한 커뮤니케이션 스킬을 높일 수 있었고 협업 하는 방법을 배울 수 있었던 프로젝트였습니다.
- 기획에 집중해서 이전에 진행한 프로젝트처럼 모델의 성능, 기능을 다루지 못해서 아쉬웠습니다.

김강호

- 내가 제작한 딥러닝 모델의 피부암 분류 성능이 좋게 나왔다.
- >> 불균형 데이터였음에도 불구하고 정확도는 물론 precision, recall, f1-score 모두 좋게 나왔으며, 특히 모반과 구별하기 어려운 흑색종의 분류성능도 잘 나왔다.
- 팀 프로젝트를 통해, 팀원들과의 소통을 어떻게 해야 할지 약간은 감을 잡은 것 같다.
- 웹앱 구현을 잘 하지 못했으며, 이번 프로젝트에서 제작한 딥러닝 모델을 컴퓨터 하드웨어 관련 문제 때문에 웹앱에 적용하지 못했다.
- 팀원들간의 소통이 잘 되지 않아 서로의 의견이 건도는 현상이 약간 있었던 것 같다.

## 05 자체 평가 의견

### 안동선

데이터셋을 검색하고 폴더 별로 라벨링을 해서 학습에 필요한 데이터셋을 직접 만들어 보는 경험을 해서 좋았습니다.

여러가지 사전학습 모델을 경험하고 비교해 볼 수 있어서 좋았습니다.

팀원들과 소통하면서 프로젝트를 진행하는 경험을 할 수 있어서 좋았습니다.

TensorFlow가 로컬 환경에서 구동되지 않아서 제작한 딥러닝 모델을 웹앱에 적용하지 못한 것이 아쉽습니다.

웹앱에서의 딥러닝 모델의 이미지 분류시간을 예상하기 위하여 코랩에서 시뮬레이션을 돌렸으나 조건을 통일하는 데 실패해서 분류시간을 일반화 하지 못했습니다.

### 이준호

- 주기적인 온라인 미팅을 통해 수행한 팀원과 일정 논의 및 프로젝트 진행은 흥미로운 경험이었습니다.
- Section 수업의 과제나 프로젝트 진행 시 활용 데이터는 공공 데이터나 kaggle에서 가져왔는데 크롤링을 통한 데이터 수집은 Section3를 제대로 복습하는 시간이었습니다.
- CP1에서는 웹 구현에 집중해서 개인 프로젝트에서 진행했던 데이터 전처리, 딥러닝 모델 구축 및 학습 업무를 수행하지 못해서 아쉬웠습니다.

**DA**

전체적인 프로젝트 파이프 라인을 배울 수 있었고 프로젝트 아이디어 구상과 서비스 기획까지 해볼 수 있어 직무적으로 좋은 경험이었다  
기획에 집중해서 이전에 진행한 프로젝트처럼 모델의 성능, 기능을 다루지 못해서 아쉬웠다

**DS**

딥러닝 모델의 분류 성능이 잘 나왔다.  
하드웨어 관련 문제로 딥러닝 모델을 웹에 적용하지 못해서 아쉽다.  
상호간의 커뮤니케이션 역량을 발전 할 수 있는 좋은 계기가 되었다.

**DE**

백엔드 측면에서 완성도가 부족해서 아쉬웠다.  
HTML, css, Java Script를 더 공부해서 백엔드 기능도 구현하고 싶다.

**감사합니다 !**