

파이썬 기반 CNN 딥러닝을 활용한 정상치아/충치 이미지 분류와 치과 추천 프로젝트

빅데이터 분석 전문가 양성과정 31기
김윤규 한경민 정명찬

목 차

- 기획 목표
- 데이터 수집, 정제
- 분석 모델링
- 시각화
- 결론, 보완 사항

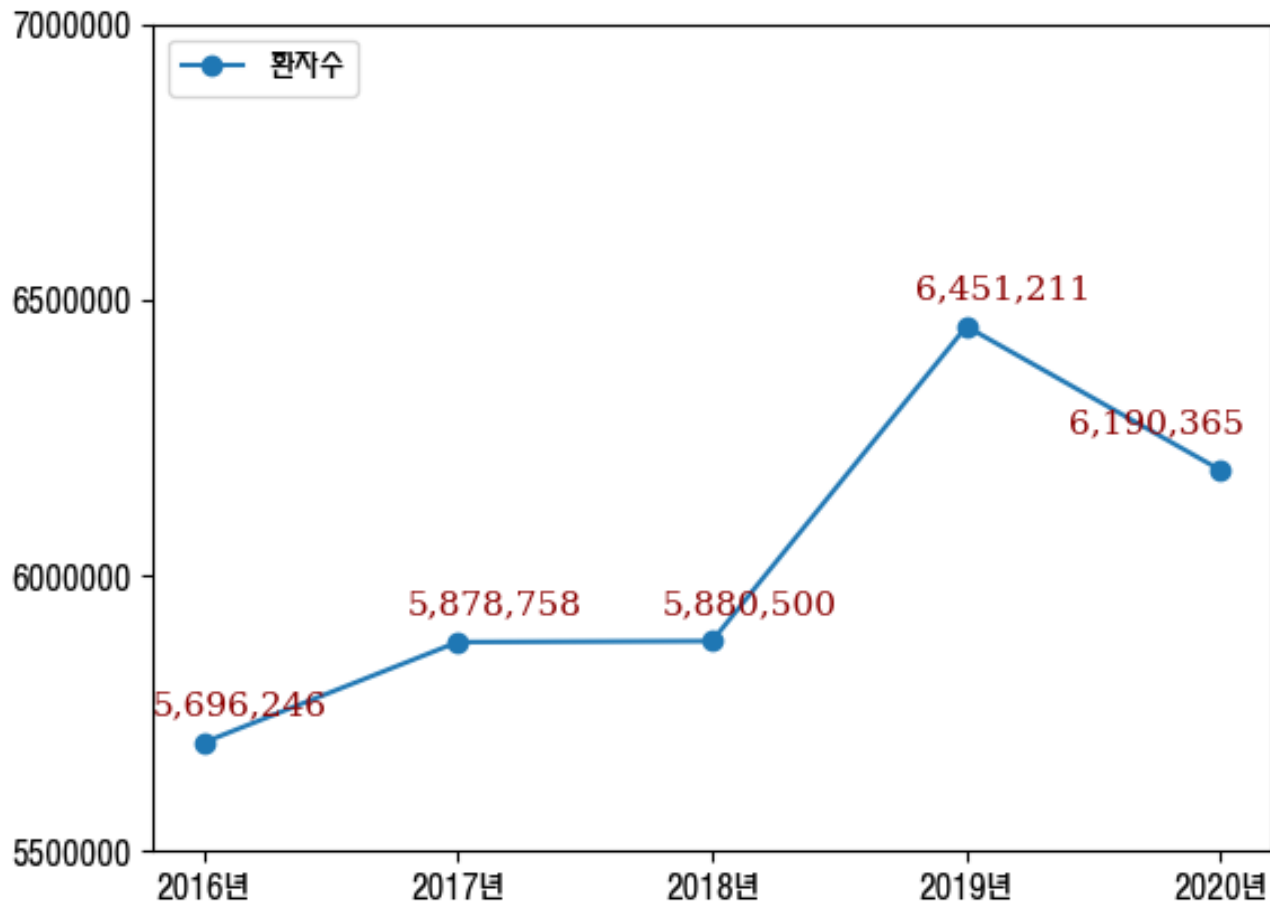
A dimly lit dental clinic. In the foreground, a black dental chair is tilted back. Above it, a dental light fixture hangs. To the right, a large piece of dental equipment with multiple tubes and a control panel is visible. The background shows white cabinets and a sink area. The overall atmosphere is professional and clinical.

기획 목표

현황 파악 통계

```
import matplotlib.pyplot as plt
```

2016 ~ 2020 총치 내원 환자 수

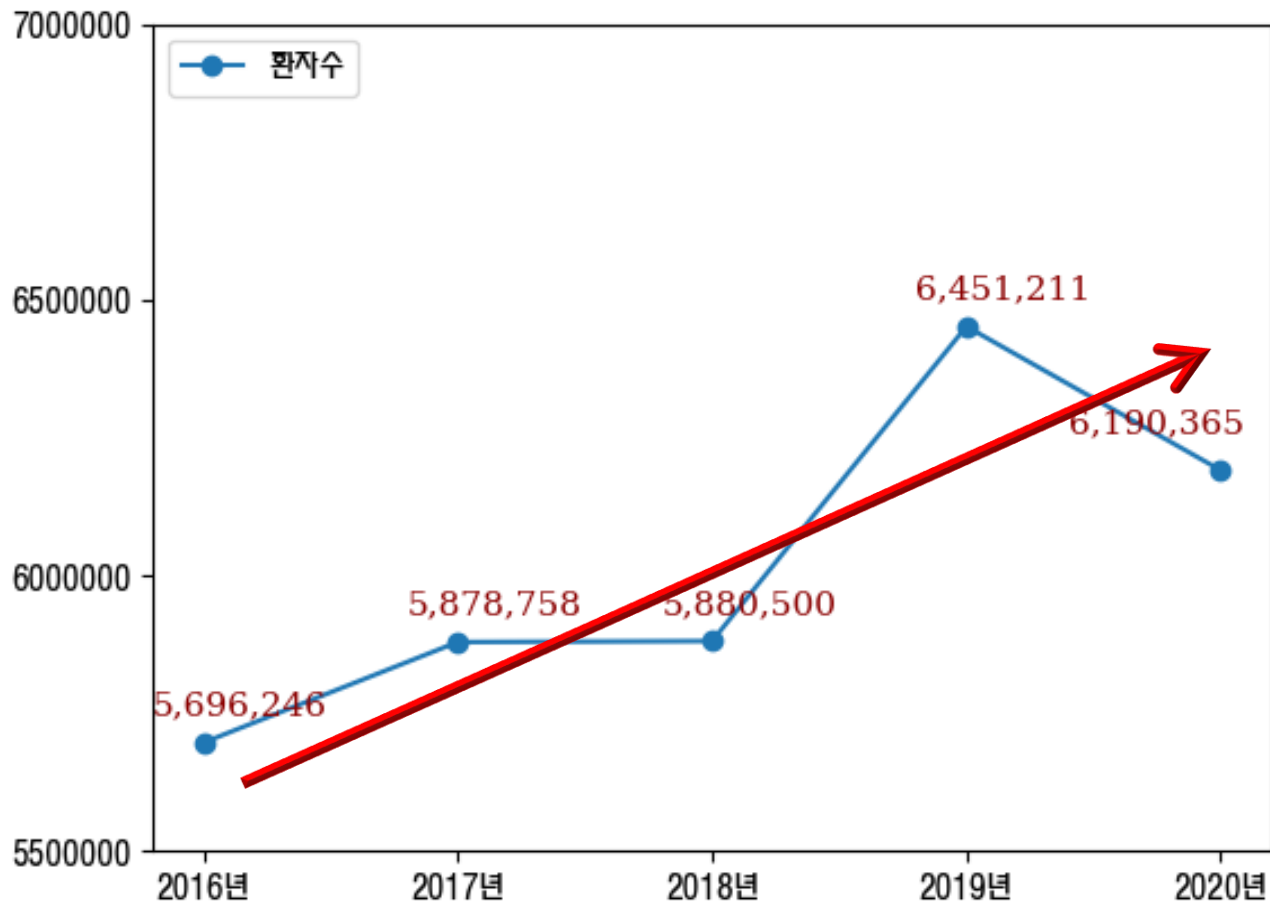


출처 : 보건의료빅데이터개방시스템

현황 파악 통계

```
import matplotlib.pyplot as plt
```

2016 ~ 2020 총치 내원 환자 수

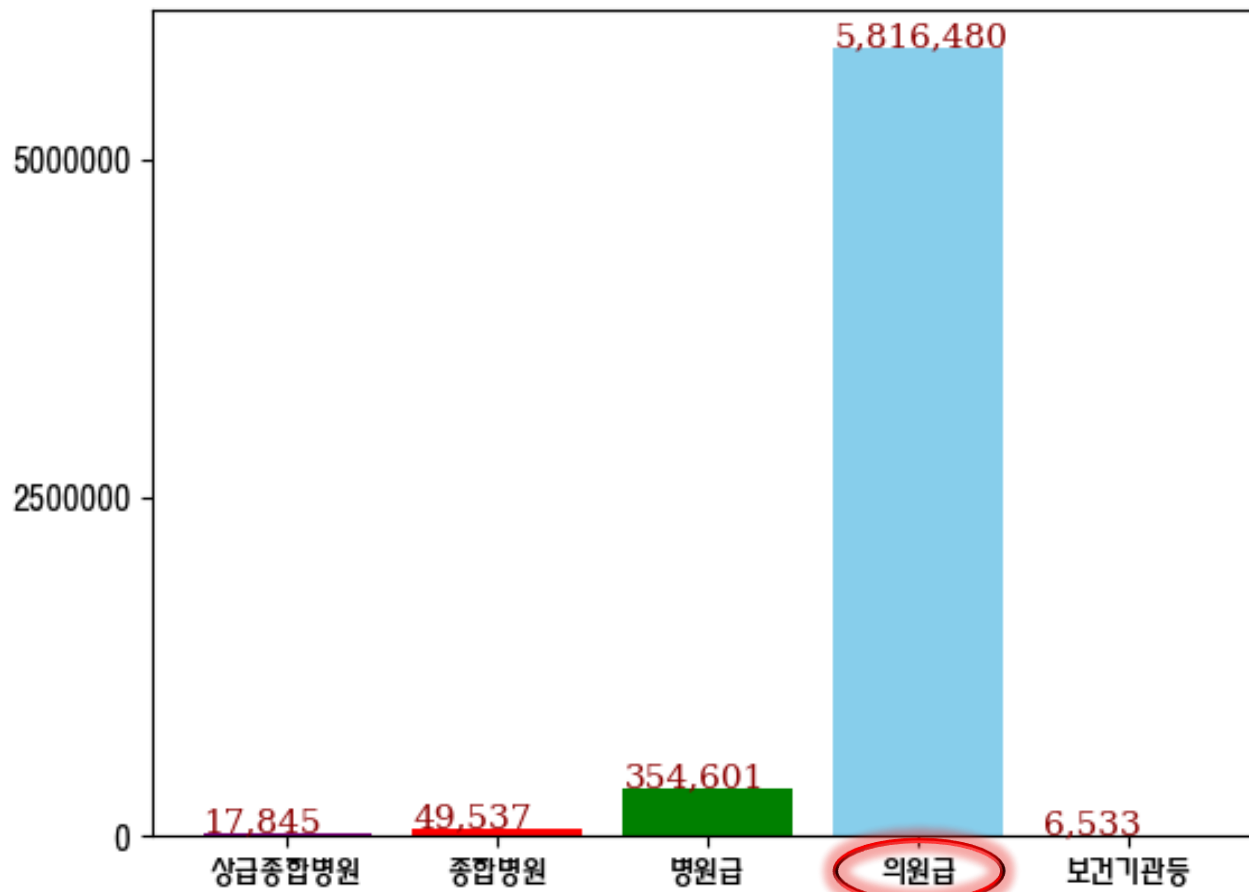


출처 : 보건의료빅데이터개방시스템

현황 파악 통계

```
import matplotlib.pyplot as plt
```

2020년 요양기관별 내원 환자 수



출처 : 보건의료빅데이터개방시스템

니즈 파악 – 네이버 지식인 질문

Q 충치인가요? 아프진 않은데 찍어보니깐 너무 까매서 불안하네요....ㅜㅜ 충치인가요? 혹시 심각한가요? 비용은 얼마정도 드나요?

Q 어금니 사이에 충치인지 궁금합니다 아프지도 않고 시리지도 않은데 전에 어금니에 충치가 나여 봤더니 어금니 사이에도 검은색이 있더라구요 이게 초기 어...

Q 충치치료 질문드립니다. 충치가 많으면...다른 이빨에도 충치가 옮기나요?

Q 사랑니 발치 후 어금니 옆 충치 ... 옆면에서 충치가 보인다고 합니다. 충치 진행은 어금니 표면부터 신경까지의 거리를 1이라 했을... 신경치료를 하면 치아에 ...

Q 앞니 충치 치과치료 앞니에 이정도 크기의 충치가 생겼는데 치료하는 비용이 어느정도 들까요? 충치가 많이 진행된 상태인가요??

Q 충치일까요ㅜㅜ

질문 내용 워드클라우드 시각화

from bs4 import BeautifulSoup | from wordcloud import WordCloud



질문 내용 워드클라우드 시각화

음식을 씹는 치아
충치가 가장 많이 생기는 치아

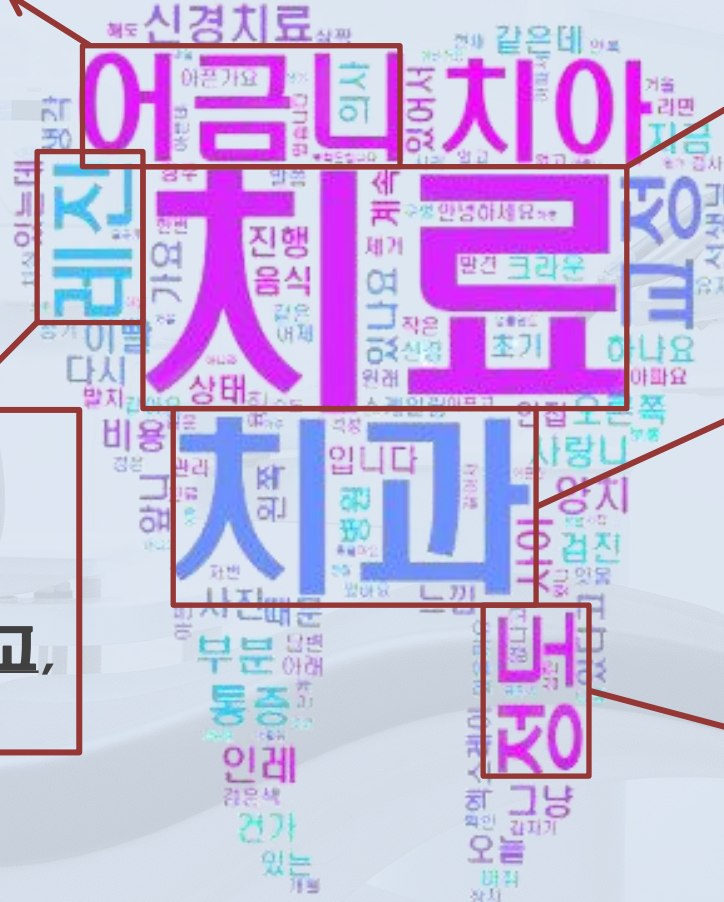
치료 필요성
충치 치료의 종류

치과 방문 필요성
우수 치과 추천 니즈

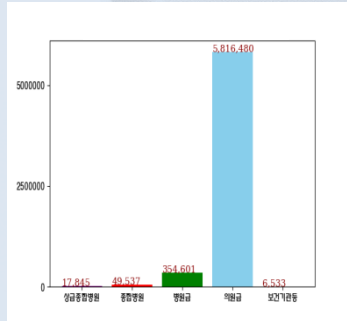
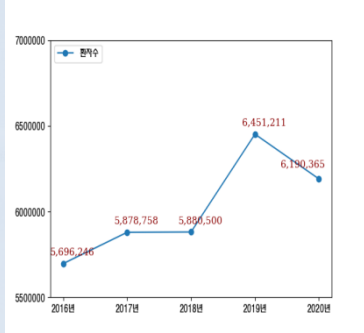
가장 많이 쓰이는
치과용 수복재료

아말감보다 부작용 적고,
금보다 비용 저렴

치아 우식 정도 판단



분석 방향 설정



CNN 이미지 분류

치아 사진을 바탕으로
충치인지 판단하는 시스템 구현

Folium 지도시각화

충치일 경우,
거주지 근처 우수 치과 추천





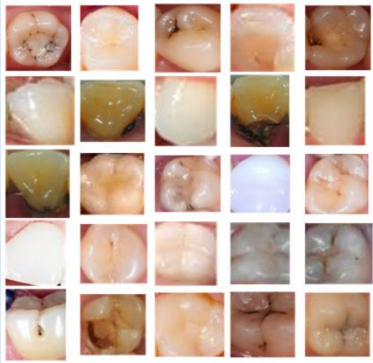
데이터 수집, 정제

이미지 학습, 검증에 이용한 데이터



데이터 준비 과정

```
from bs4 import BeautifulSoup |  
from selenium import webdriver |  
import urllib.request as req
```



네이버, 구글 이미지 스크래핑

정상치아, 충치 이미지 수집

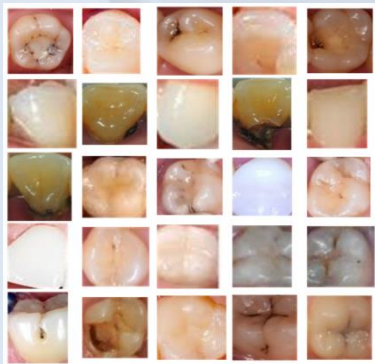
데이터 정제 작업

주제와 맞지 않는 데이터 삭제

이미지 당 한 개의 이를 나타내도록
이미지 편집

데이터 준비 과정

```
from bs4 import BeautifulSoup |  
from selenium import webdriver |  
import urllib.request as req
```



네이버, 구글 이미지 스크래핑

정상치아, 충치 이미지 수집

데이터 정제 작업

주제와 맞지 않는 데이터 삭제

이미지 당 한 개의 이를 나타내도록
이미지 편집

정제 후 이미지 데이터

총 **722**개

데이터 준비 과정

```
from bs4 import BeautifulSoup |  
from selenium import webdriver |  
import urllib.request as req
```

정제 후 이미지 데이터

총 **722**개

학습데이터

577개

검증데이터

145개

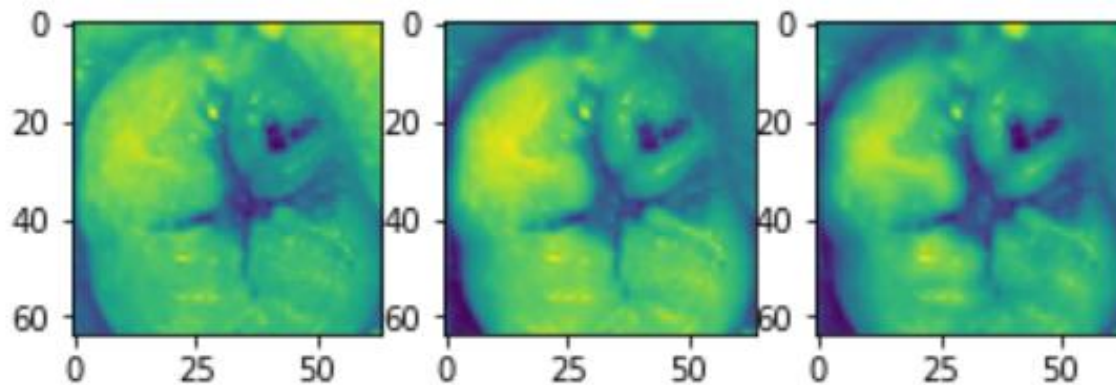
A dark, grayscale image of a dental clinic. In the foreground, a dental chair is visible, tilted back. Above it, a dental light fixture hangs. To the right, there is a control panel with various buttons and a monitor. The background shows a cabinet with drawers. The overall scene is dimly lit, with the text '분석 모델링' (Analysis Modeling) overlaid in the center.

분석 모델링

이미지 RGB 변환

```
from PIL import Image  
from tensorflow.keras.preprocessing.image import ImageDataGenerator
```

```
plt.subplot(1,3,1)  
plt.imshow(x[50][:,:,0])  
plt.subplot(1,3,2)  
plt.imshow(x[50][:,:,1])  
plt.subplot(1,3,3)  
plt.imshow(x[50][:,:,2]);
```



CNN 모델 형성 input_size : (64, 64)

from tensorflow.keras.layers import Conv2D, MaxPool2D, Input, Flatten, Dense, Dropout |
from keras.preprocessing.image import img_to_array, load_img, array_to_img

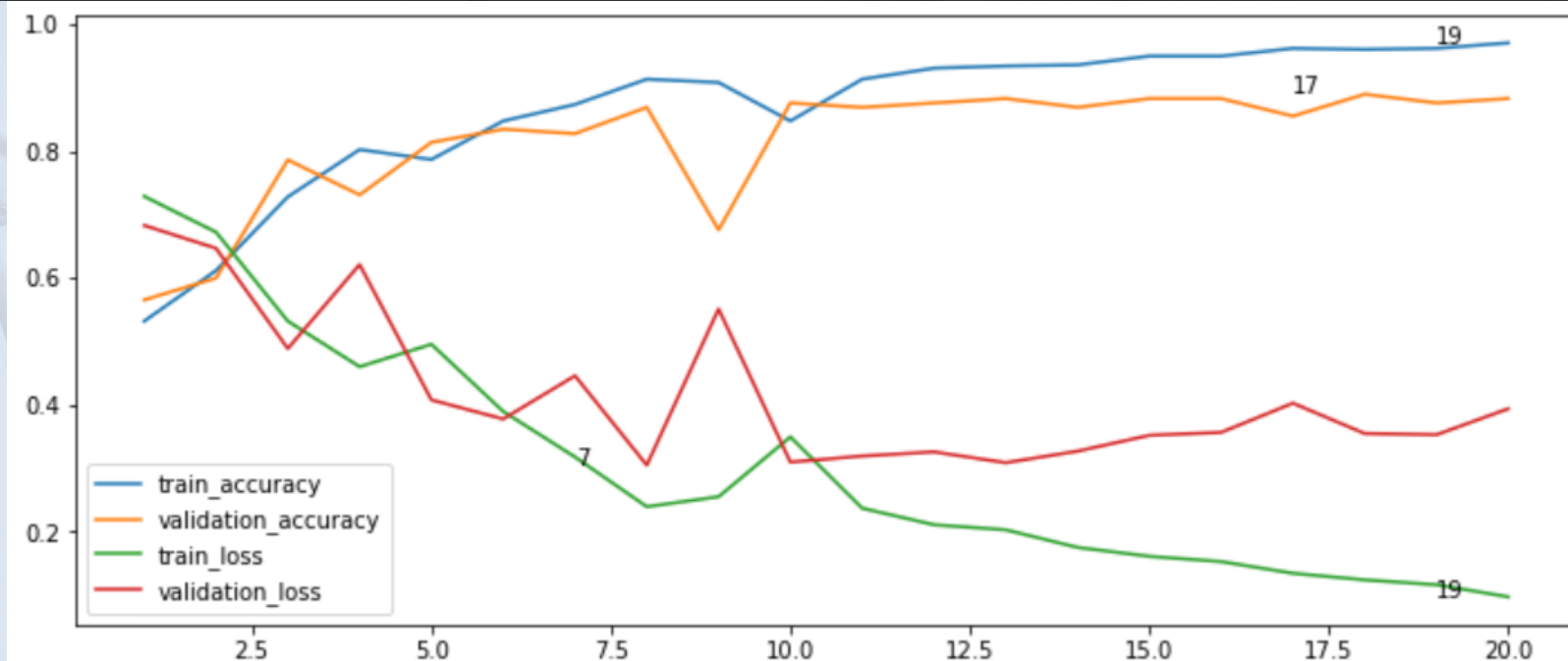
입력층

```
model_cnn64 = Sequential()  
model_cnn64.add(Conv2D(filters=32, kernel_size=(3,3), strides=(1,1), padding='same', input_shape=(64,64,3), activation='relu'))  
model_cnn64.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn64.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn64.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn64.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn64.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn64.add(Flatten())  
model_cnn64.add(Dense(512, activation='relu'))  
model_cnn64.add(Dropout(0.5))  
  
model_cnn64.add(Dense(2, activation='softmax'))  
model_cnn64.summary()
```

은닉층

출력층

Accuracy, Loss 도출 input_size : (64, 64)



이미지
Input_size
(64,64)

Learning_rate :
0.001
Epochs :
20
Batch_size :
32

Train Accuracy 최대값:
0.970537245273590
Test Accuracy 최대값:
0.889655172824859
Train loss 최소값:
0.097569189965724
Test loss 최소값:
0.305091440677642

CNN 모델 형성 input_size : (128, 128)

```
from tensorflow.keras.layers import Conv2D, MaxPool2D, Input, Flatten, Dense, Dropout |  
from keras.preprocessing.image import img_to_array, load_img, array_to_img
```

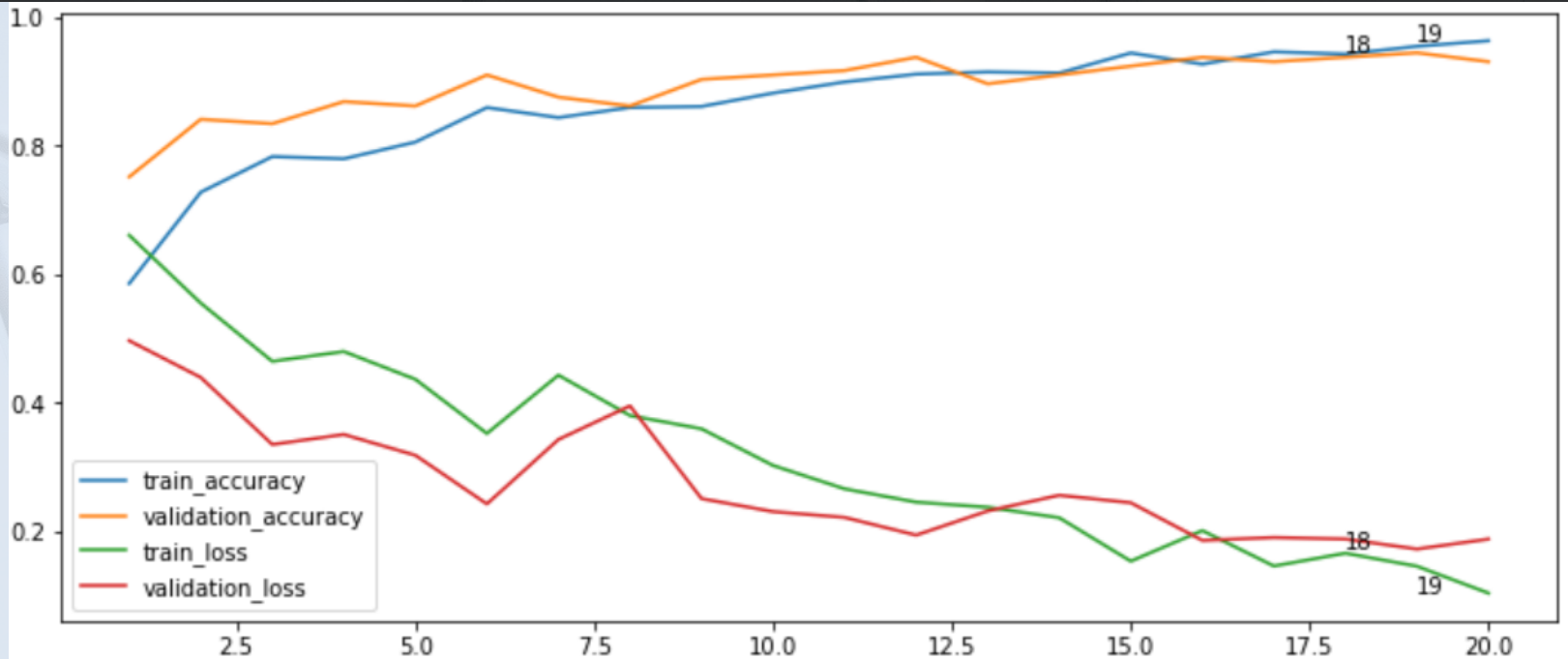
입력층

```
model_cnn128 = Sequential()  
model_cnn128.add(Conv2D(filters=32, kernel_size=(3,3), strides=(1,1), padding='same', input_shape=(128,128,3), activation='relu'))  
model_cnn128.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn128.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn128.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn128.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn128.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn128.add(Flatten())  
model_cnn128.add(Dense(512, activation='relu'))  
model_cnn128.add(Dropout(0.5))  
  
model_cnn128.add(Dense(2, activation='softmax'))  
model_cnn128.summary()
```

은닉층

출력층

Accuracy, Loss 도출 input_size : (128, 128)



이미지
Input_size
(128,128)

Learning_rate :
0.001
Epochs :
20
Batch_size :
32

Train Accuracy 최대값:
0.963604867458343
Test Accuracy 최대값:
0.944827556610107
Train loss 최소값:
0.1039920821785926
Test loss 최소값:
0.1726340055465698

CNN 모델 형성 input_size : (224, 224)

from tensorflow.keras.layers import Conv2D, MaxPool2D, Input, Flatten, Dense, Dropout |
from keras.preprocessing.image import img_to_array, load_img, array_to_img

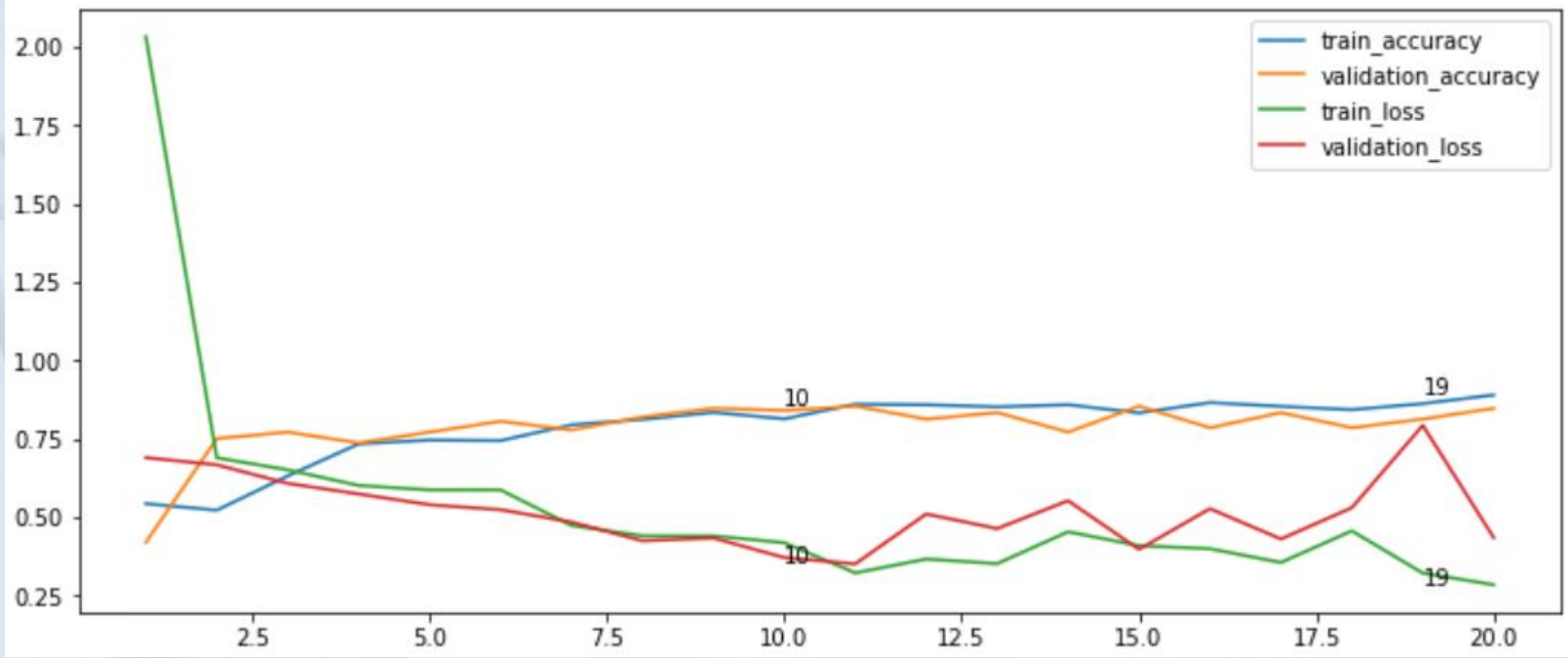
입력층

```
model_cnn224 = Sequential()  
model_cnn224.add(Conv2D(filters=32, kernel_size=(3,3), strides=(1,1), padding='same', input_shape=(224,224,3), activation='relu'))  
model_cnn224.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn224.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn224.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn224.add(Conv2D(filters=64, kernel_size=(3,3), strides=(1,1), padding='same', activation='relu'))  
model_cnn224.add(MaxPool2D(pool_size=(2,2)))  
  
model_cnn224.add(Flatten())  
model_cnn224.add(Dense(512, activation='relu'))  
model_cnn224.add(Dropout(0.5))  
  
model_cnn224.add(Dense(2, activation='softmax'))  
model_cnn224.summary()
```

은닉층

출력층

Accuracy, Loss 도출 input_size : (224, 224)

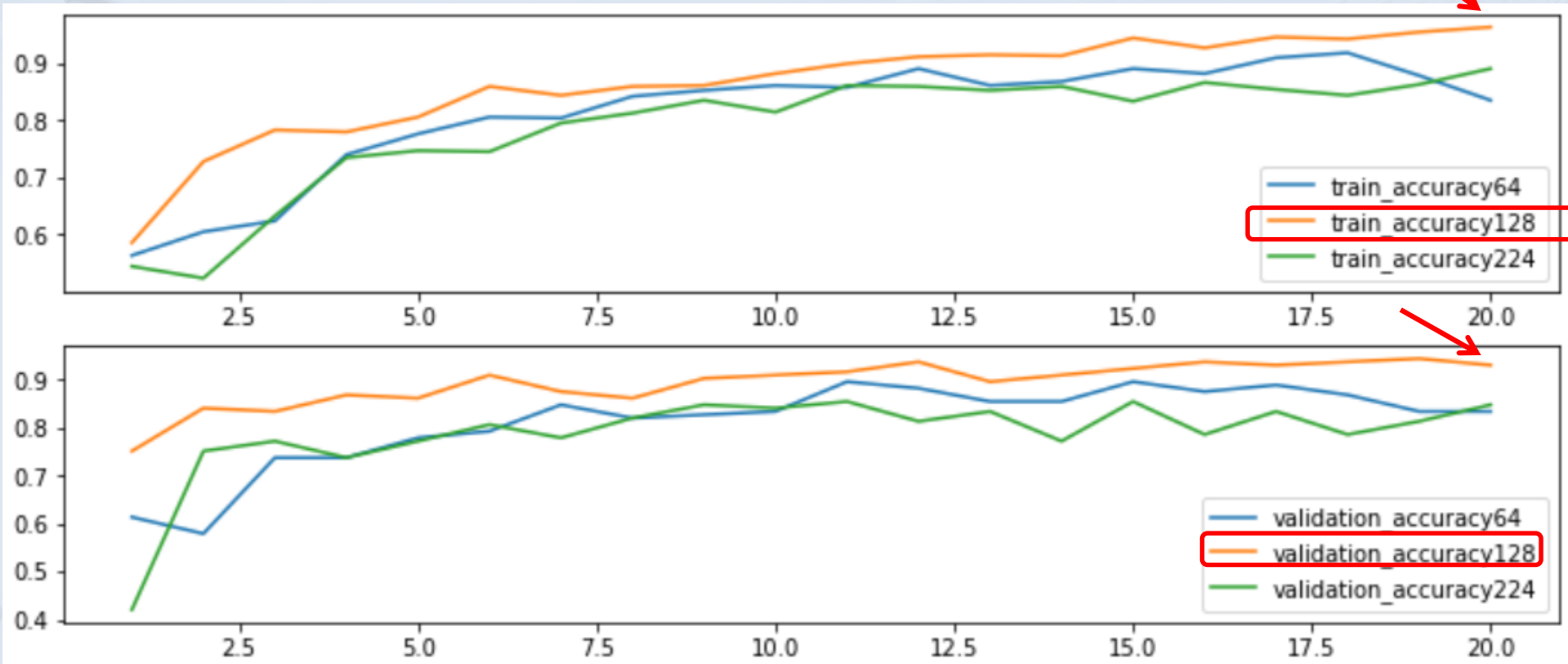


이미지
Input_size
(224,224)

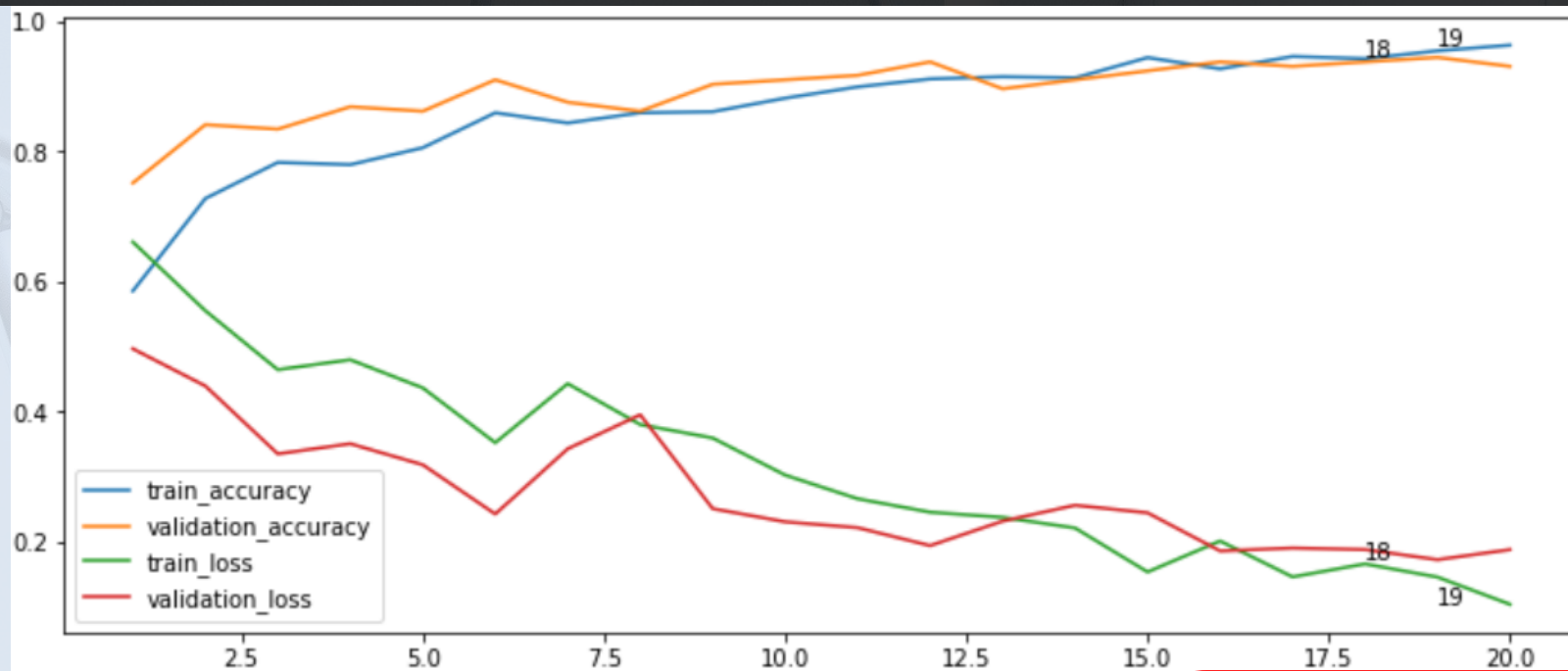
Learning_rate :
0.001
Epochs :
20
Batch_size :
32

Train Accuracy 최대값:
0.890814542770385
Test Accuracy 최대값:
0.8551723957061768
Train loss 최소값:
0.2855668663978576
Test loss 최소값:
0.3522111177444458

Accuracy 비교



Input_size 128 * 128 모델 선택



BEST

이미지
Input_size
(128,128)

Learning_rate :
0.001
Epochs :
20
Batch_size :
32

Train Accuracy 최대값:
0.963604867458343
Test Accuracy 최대값:
0.944827556610107
Train loss 최소값:
0.1039920821785926
Test loss 최소값:
0.1726340055465698

시각화

A dark, grayscale photograph of a dental clinic. In the foreground, a black dental chair is tilted back. Above it, a dental operating light hangs from the ceiling. To the right, a dental unit with various controls and hoses is visible. The background shows white cabinets and a sink area. The Korean word '시각화' (Visualization) is written in white in the center of the image.

리뷰 우수 치과 정보 수집

모두닥

의사회원 | 회사소개 | 인재채용

리뷰쓰기 가격정보 ▾ 건강정보

가 본 사람들의 솔직한 병원 후기

인증된 리뷰 393,315개 회원 수 1,348,027명 의사 수 91,988명

출처 : <https://www.modoodoc.com/>

리뷰 우수 치과 정보 수집

모두닥 의사회원 | 회사소개

리



서울 > 서초구

강남더블유치과의원

★★★★★ 9.6 리뷰 12

총치치료 (3) 사랑니발치 (2) 치과검진 (2) 임플란트 (2) 치아교정 (2)

치과 스케일링 (1) 사각턱수술 (1)

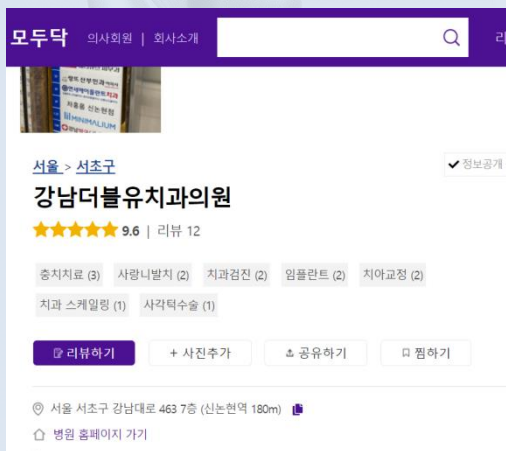
리뷰하기 + 사진추가 공유하기 찜하기

서울 서초구 강남대로 463 7층 (신논현역 180m)

병원 홈페이지 가기

리뷰 우수 치과 정보 수집

```
from bs4 import BeautifulSoup |  
import urllib.request as req
```



치과 선정 기준

- 서울시 소재 한정
- 리뷰 평점 8.0 이상
- 의원급

좌표(위도, 경도) 정보

GeocodingTool64 프로그램 이용
주소 바탕으로 위도, 경도 정보 획득

리뷰 우수 치과 정보 수집

```
from bs4 import BeautifulSoup |  
import urllib.request as req
```



치과 데이터
총 **483**개

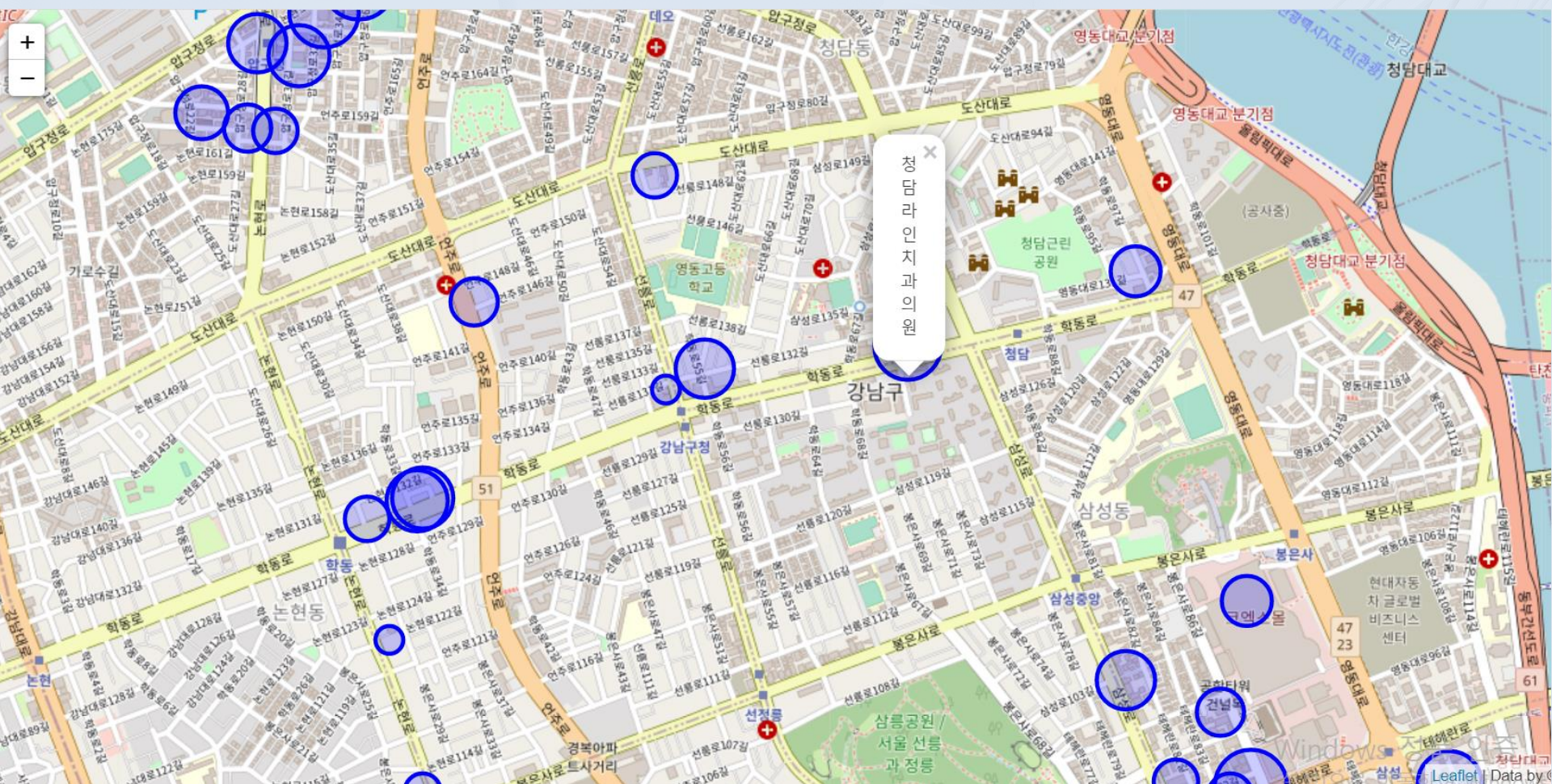
치과 선정 기준

- 서울시 소재 한정
- 리뷰 평점 8.0 이상
- 의원급

좌표(위도, 경도) 정보

GeocodingTool64 프로그램 이용
주소 바탕으로 위도, 경도 정보 획득

Import folium





결론, 보완 사항

결론, 보완 사항

- 충치 판단과 치과 추천 **니즈 충족 가능성** 높음
- 높은 정확도(약 **95%**)의 이미지 분류 모델 구현
- 사진을 찍은 후 하나의 이만 나타나도록 **편집해야 하는**
번거로움 존재
- 지방의 치과도 반영할 수 있도록 **데이터 확보** 필요
- **웹, 앱**에서 실질적으로 활용할 수 있도록 **프로그래밍** 필요



감사합니다