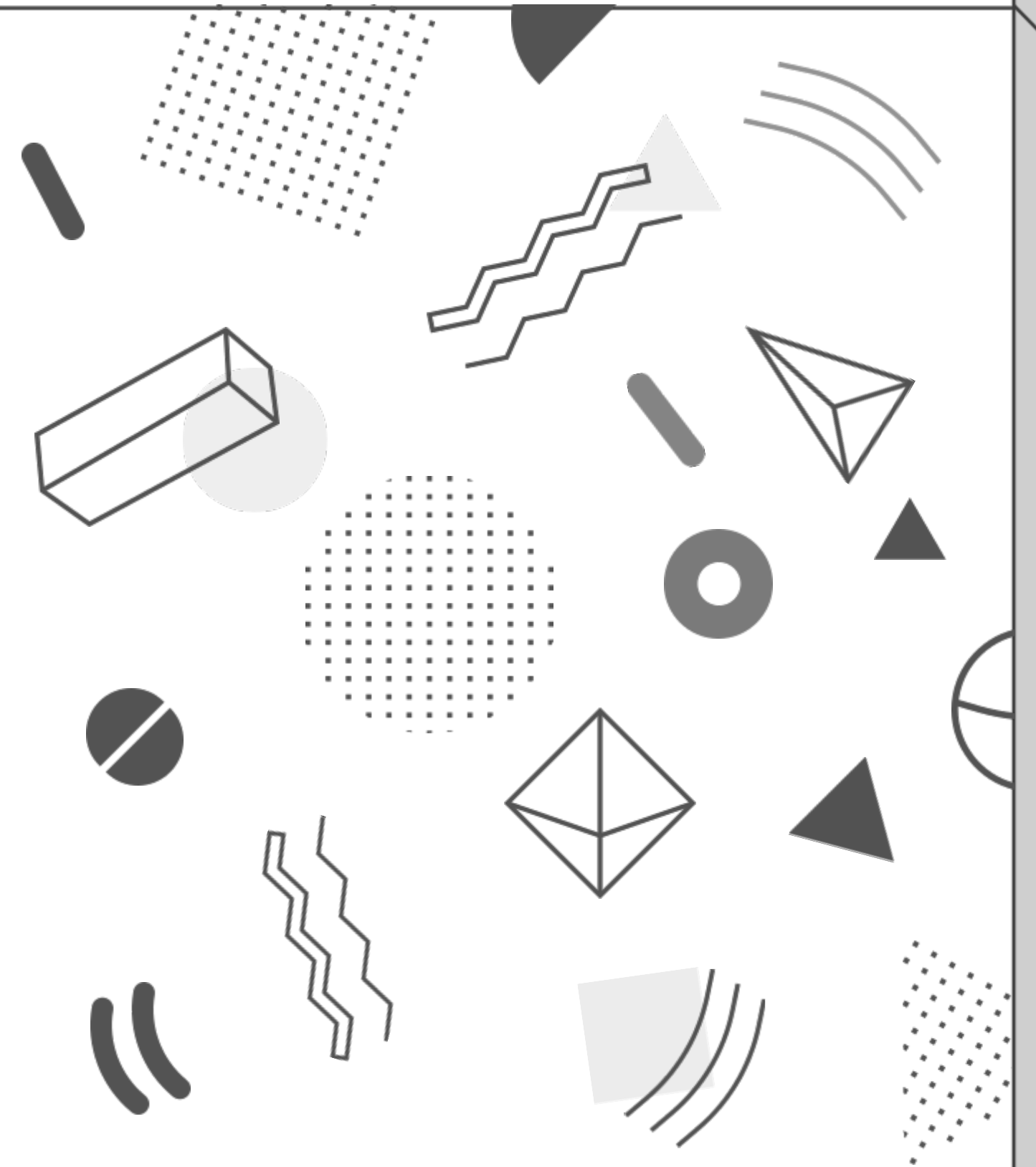




2022-2 딥러닝 프로젝트 결과보고서

# 흉부 X-ray 사진으로 호흡기질환 분류하기

2270322 박다연



# 목차

1. 문제 설명
2. 문제 해결을 위한 접근 방법
3. 사용한(수집한) 데이터 설명
4. 인공지능 모델의 입력과 출력에 대한 상세 설명
5. 딥러닝 모델 상세 설명
6. 인공지능 학습과정의 결과물
7. 학습된 인공지능을 문제 해결에 적용한 방법 및 결과물 설명
8. 제안서의 내용에 대비하여 변경된 부분에 대한 사유 및 설명



# 1. 문제 설명

폐렴은 세균이나 바이러스 등의 미생물로 인한 감염으로 발생하는 폐의 염증으로, 폐 증상과 전신적인 증상이 나타난다. 결핵은 결핵균에 의해 감염되는 질병이며, 역사상 가장 많은 생명을 앗아간 감염 질환이다. 코로나19는 신종 코로나 바이러스에 의해 감염되는 질병이며, 6억 명의 감염자의 6백만명의 사망자를 발생시켰다. 이 세 질환은 증상이 비슷하여 혼동하기 쉬우며, 결핵과 코로나 19는 전파력이 강하고 폐렴은 치사율이 강한 만큼 정확하고 빠른 진단이 필요하다.

결핵, 폐렴 등 감염성 폐질환을 진단하는 방법 중 가장 간단하고 경제적인 방법은 X-ray 촬영이다. X-ray의 검사시간은 1~5분이며, 검사비용은 만원 이하이다. 하지만 X-ray 촬영만으로는 무슨 질병에 걸렸는지 알지 못하고, 반드시 의사의 진단이 있어야 한다. 의사의 진단 없이 X-ray 촬영만으로 결핵, 폐렴, 코로나19를 정확하고 빠르게 구분할 필요가 있다.

## 2. 문제를 해결하기 위한 접근 방법

정상, 폐렴, 결핵, 코로나 환자의 흉부 X-ray 사진을 분류(Classification)하도록 인공지능 모델을 학습시킨다.  
이 인공지능 모델을 이용하여 환자가 어느 질병에 걸렸는지 흉부 X-ray 사진만으로 판별할 수 있도록 한다.


### 문제를 해결하는 데 인공지능 접근 방법이 좋은 이유

인간이 X-ray로 질병을 진단하려면 최소 6년은 학습해야 한다.  
하지만 인공지능 모델은 충분한 데이터만 있다면 학습하는데 며칠, 학습된 모델로 판별하는 것은 1초 안에 가능하다.  
인공지능을 사용하는 것이 훨씬 효율적이라고 할 수 있다.  
의사가 없어도 질병을 판별할 수 있기 때문에 의사가 몇 없는 열악한 환경에서도 유용하게 사용할 수 있을 것이다.

미래에 신종 코로나 바이러스만큼 강력한 신종 호흡기 질환이 발생할 수 있다.  
신종 호흡기 질환에 걸린 사람들의 X-ray 사진만 충분히 있다면 이 인공지능 모델에 새로운 질병을 학습시킬 수 있고,  
학습된 인공지능 모델을 이용하여 신속한 진단이 가능할 것이다.

### 3. 사용한 데이터 설명

#### 1) 폐렴 환자의 흉부 X-ray 데이터

 PAUL MOONEY · UPDATED 5 YEARS AGO

▲ 5358


New Notebook

Download (2 GB)

⋮

## Chest X-Ray Images (Pneumonia)

5,863 images, 2 categories




<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

Paul Mooney(kaggle) - Chest X-Ray Images (Pneumonia)

폐렴 환자의 흉부 X-ray 사진 4273장, 일반인의 흉부 X-ray 사진 1583장이 들어있다.

### 3. 사용한 데이터 설명

#### 2) 결핵 환자의 흉부 X-ray 데이터

 TAWSIFUR RAHMAN AND 2 COLLABORATORS · UPDATED A YEAR AGO

▲ 124

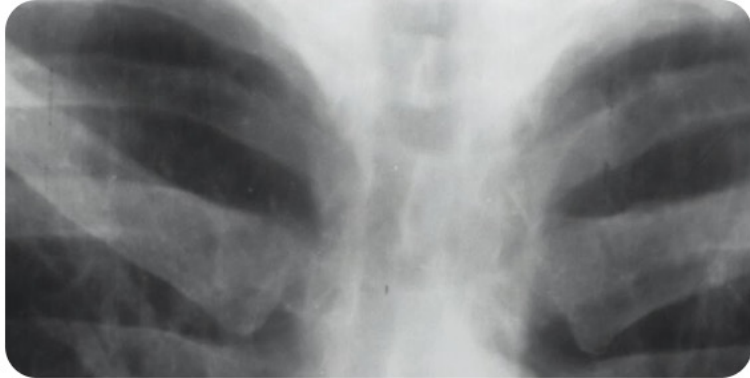
New Notebook

Download (696 MB)

⋮

## Tuberculosis (TB) Chest X-ray Database

The largest TB Chest X-ray Database



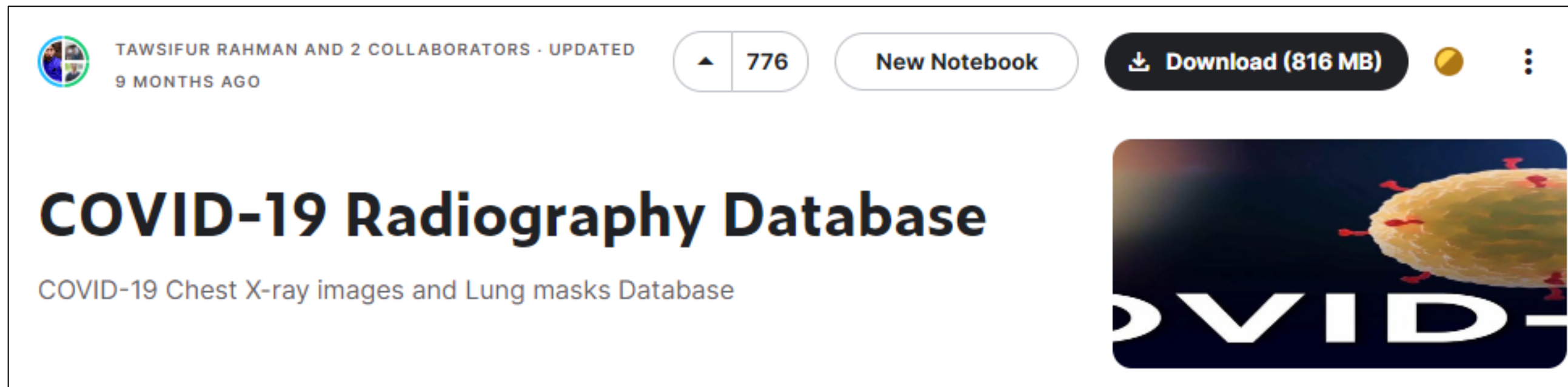
<https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>

Tawsifur Rahman 외 2인(kaggle) - Tuberculosis (TB) Chest X-ray Database

결핵 환자의 흉부 X-ray 사진 700장, 일반인의 흉부 X-ray 사진 3500장이 들어있다.

## 3. 사용한 데이터 설명

### 3) 코로나19 환자의 흉부 X-ray 데이터



<https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>


Tawsifur Rahman 외 2인(kaggle) - COVID-19 Radiography Database

코로나19 환자의 흉부 X-ray 사진 3616장, 일반인의 흉부 X-ray 사진 10192장이 들어있다.



### 3. 사용한 데이터 설명

#### 4) 일반인의 흉부 X-ray 데이터 수집

 PAUL MOONEY · UPDATED 5 YEARS AGO

▲ 5358


New Notebook


Download (2 GB)

⋮

### Chest X-Ray Images (Pneumonia)

5,863 images, 2 categories



 TAWSIFUR RAHMAN AND 2 COLLABORATORS · UPDATED A YEAR AGO

▲ 124

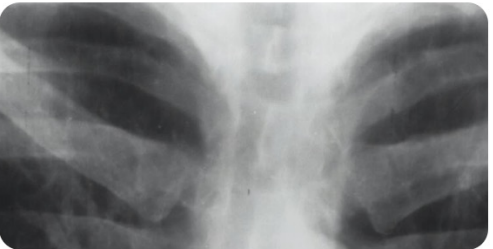
New Notebook


Download (696 MB)

⋮

### Tuberculosis (TB) Chest X-ray Database

The largest TB Chest X-ray Database



 TAWSIFUR RAHMAN AND 2 COLLABORATORS · UPDATED 9 MONTHS AGO

▲ 776


New Notebook

Download (816 MB)

⋮

### COVID-19 Radiography Database

COVID-19 Chest X-ray images and Lung masks Database



앞서 소개한 dataset에 들어있는 일반인의 흉부 X-ray 사진 16275장을 이용한다.



## 4. 인공지능 모델의 입력과 출력에 대한 상세 설명

### 1) 입력

이 프로젝트에서 입력으로 사용할 Feature는 X-ray 사진이다.

측면에서 찍은 사진, 지나치게 흐릿한 사진, 공백이 많은 사진은 직접 삭제해 주었다.

opencv의 cvtColor() 함수를 이용하여 이미지를 흑백으로 변환하고,  
resize() 함수를 이용하여 이미지의 크기를 256 \* 256 픽셀로 줄였다.



```
def img_convert(filename):  
    try:  
        img = cv2.imread(filename)  
        img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY) # 흑백 사진으로 변환  
        img = cv2.resize(img, (256, 256)) # 256 * 256 사이즈로 변환  
    except:  
        print(filename, ": 이미지 파일이 맞는지 확인해 주세요.", file=sys.stderr)  
    return img
```



## 4. 인공지능 모델의 입력과 출력에 대한 상세 설명

### 1) 입력

#### 데이터의 노멀라이즈

Feature로 사용하는 X-ray사진은 256 \* 256 픽셀, 흑백사진이다.  
이를 노멀라이즈 하기 위해 한 픽셀당 255로 나누었다.

```
# normalize
self.train_X = np.asarray(self.train_X) / 255.0
self.train_Y = np.asarray(self.train_Y)
self.test_X = np.asarray(self.test_X) / 255.0
self.test_Y = np.asarray(self.test_Y)
```

#### 데이터의 분할

인공지능의 학습과 테스트를 수행하기 위해서 데이터를 8:2로 나누었다.

폐렴 환자의 흉부 X-ray 사진 4265장 => 학습용 3412장, 테스트용 853장으로 나누었다.

결핵 환자의 흉부 X-ray 사진 696장 => 학습용 557장, 테스트용 139장으로 나누었다.

코로나19 환자의 흉부 X-ray 사진 3596장 => 학습용 2879장, 테스트용 717장으로 나누었다.

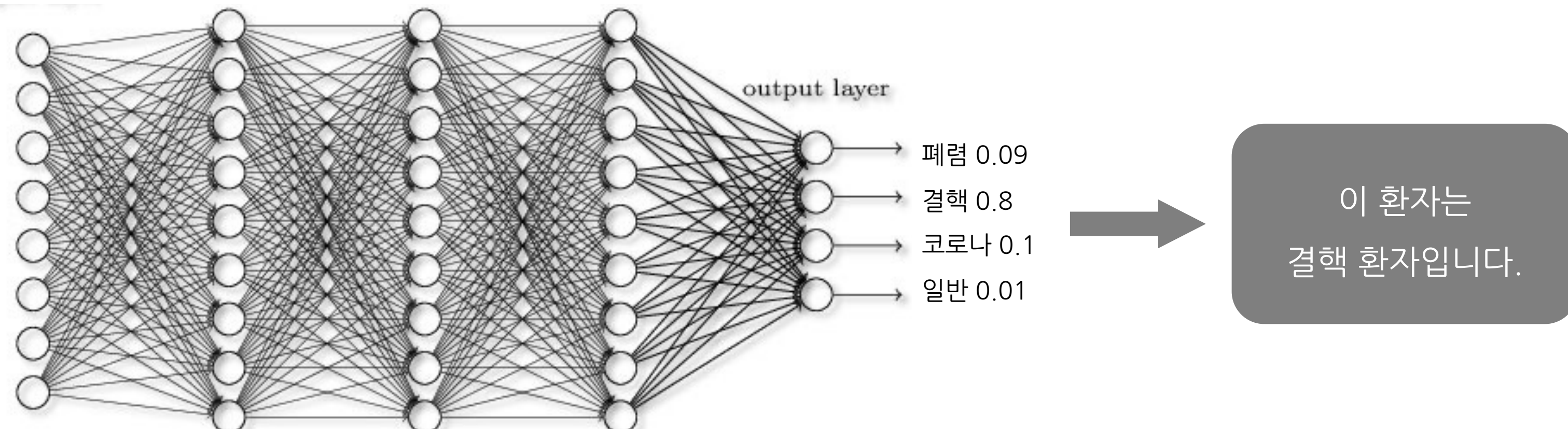
일반인의 흉부 X-ray 사진 15287장 => 학습용 12330장, 테스트용 2957장으로 나누었다.

## 4. 인공지능 모델의 입력과 출력에 대한 상세 설명

### 2) 출력

폐렴, 결핵, 코로나, 일반 총 4개의 카테고리 분류(Classification)해야 하기 때문에 마지막 층의 출력값은 4개이고 활성화함수는 SoftMax로 설정하였다.

만약 인공지능의 출력의 결과가 폐렴 0.09, 결핵 0.8, 코로나 0.1, 일반 0.01이라면 이 환자는 결핵 환자라고 결론을 내린다.



# 5. 딥러닝 모델 상세 설명

층수	종류	크기	활성화 함수
1층	CNN	16, (3, 3)	ReLu
-	MaxPool	(2, 2)	-
2층	CNN	32, (3, 3)	ReLu
-	MaxPool	(2, 2)	-
3층	CNN	32, (3, 3)	ReLu
-	MaxPool	(2, 2)	-
4층	CNN	64, (3, 3)	ReLu
-	MaxPool	(2, 2)	-
5층	CNN	64, (3, 3)	ReLu
-	MaxPool	(2, 2)	-
6층	CNN	64, (3, 3)	ReLu
7층	Flatten	-	-
8층	FNN	256	ReLu
-	Dropout	rate=0.7	-
9층	FNN	4	SoftMax

# 5. 딥러닝 모델 상세 설명

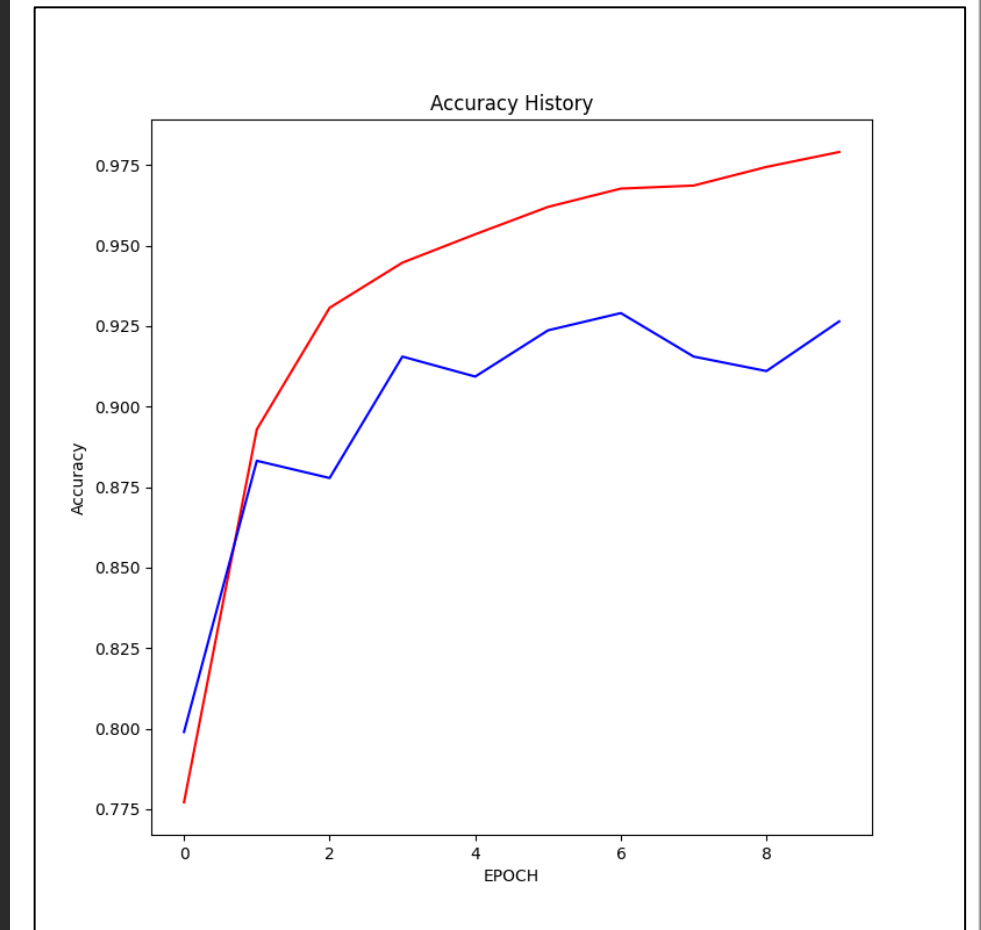
- 1층: 256 \* 256 흑백 사진에 (3, 3) 필터 16개를 적용한다. => (254, 254, 16)  
MaxPool(2, 2)을 적용한다. => (127, 127, 16)
- 2층: (127, 127, 16)에 (3, 3) 필터 32개를 적용한다. => (125, 125, 32)  
MaxPool(2, 2)을 적용한다. => (62, 62, 32)
- 3층: (62, 62, 32)에 (3, 3) 필터 32개를 적용한다. => (60, 60, 32)  
MaxPool(2, 2)을 적용한다. => (30, 30, 32)
- 4층: (30, 30, 32)에 (3, 3) 필터 64개를 적용한다. => (28, 28, 64)  
MaxPool(2, 2)을 적용한다. => (14, 14, 64)
- 5층: (14, 14, 64)에 (3, 3) 필터 64개를 적용한다. => (12, 12, 64)  
MaxPool(2, 2)을 적용한다. => (6, 6, 64)
- 6층: (6, 6, 64)에 (3, 3) 필터 64개를 적용한다. => (4, 4, 64)
- 7층: Flatten한다. => (1024)
- 8층: FNN을 통해 학습하며, 오버피팅을 방지하기 위해 DropOut을 0.7로 설정한다.
- 9층: 분류를 위해 SoftMax 활성화 함수를 사용하고, 출력은 4개가 되도록 한다.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 16)	160
max_pooling2d (MaxPooling2D)	(None, 127, 127, 16)	0
conv2d_1 (Conv2D)	(None, 125, 125, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 32)	0
conv2d_2 (Conv2D)	(None, 60, 60, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 32)	0
conv2d_3 (Conv2D)	(None, 28, 28, 64)	18496
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 64)	0
conv2d_4 (Conv2D)	(None, 12, 12, 64)	36928
max_pooling2d_4 (MaxPooling2D)	(None, 6, 6, 64)	0
conv2d_5 (Conv2D)	(None, 4, 4, 64)	36928
flatten (Flatten)	(None, 1024)	0
dense (Dense)	(None, 256)	262400
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1028
Total params: 369,828		
Trainable params: 369,828		
Non-trainable params: 0		

## 6. 인공지능 학습과정의 결과물

### 1) 결과물

```
***** TRAINING START *****
2022-12-14 15:31:38.244655: W tensorflow/core/platform/profile_utils/cpu_utils.cc:128] Failed to get CPU frequency: 0 Hz
Epoch 1/10
600/600 [=====] - 230s 380ms/step - loss: 0.5537 - accuracy: 0.7880 - val_loss: 0.4593 - val_accuracy: 0.8369
Epoch 2/10
600/600 [=====] - 258s 430ms/step - loss: 0.2570 - accuracy: 0.9090 - val_loss: 0.3549 - val_accuracy: 0.8909
Epoch 3/10
600/600 [=====] - 265s 442ms/step - loss: 0.1993 - accuracy: 0.9330 - val_loss: 0.3549 - val_accuracy: 0.8796
Epoch 4/10
600/600 [=====] - 269s 448ms/step - loss: 0.1515 - accuracy: 0.9470 - val_loss: 0.3644 - val_accuracy: 0.8913
Epoch 5/10
600/600 [=====] - 329s 549ms/step - loss: 0.1283 - accuracy: 0.9563 - val_loss: 0.2595 - val_accuracy: 0.9246
Epoch 6/10
600/600 [=====] - 321s 535ms/step - loss: 0.1100 - accuracy: 0.9594 - val_loss: 0.3132 - val_accuracy: 0.9123
Epoch 7/10
600/600 [=====] - 292s 486ms/step - loss: 0.0930 - accuracy: 0.9667 - val_loss: 0.2479 - val_accuracy: 0.9361
Epoch 8/10
600/600 [=====] - 288s 480ms/step - loss: 0.0827 - accuracy: 0.9711 - val_loss: 0.3062 - val_accuracy: 0.9376
Epoch 9/10
600/600 [=====] - 349s 583ms/step - loss: 0.0708 - accuracy: 0.9754 - val_loss: 0.3676 - val_accuracy: 0.9295
Epoch 10/10
600/600 [=====] - 332s 553ms/step - loss: 0.0633 - accuracy: 0.9776 - val_loss: 0.3196 - val_accuracy: 0.9363
```



최종 val\_accuracy는 0.9363이며, 오버피팅이 발생하였다.



## 6. 인공지능 학습과정의 결과물

### 2) 오버피팅

오버피팅을 줄이기 위해 CNN 레이어 수 변경, CNN 필터 수 변경, Dropout 비율 변경, FNN 노드 수 변경 등의 시도를 해 보았다. 하지만 오버피팅을 해소하지 못하였다.

오버피팅의 원인을 알아내기 위해 어떤 데이터에서 오류가 많이 발생하는지 분석해 보았다.

결핵 사진과 코로나 사진으로 질병을 입력했을 때 오류가 많이 발생하였다.

결핵 환자의 X-Ray 사진은 700장으로 양이 너무 적고,

코로나에 걸리면 결국 바이러스성 폐렴으로 발전하기 때문에 코로나와 폐렴의 구분이 잘 되지 않아

오버피팅이 발생한 것으로 추측된다.



## 7. 학습된 인공지능을 문제 해결에 적용한 방법 및 결과물 설명

인공지능 학습과 질병 예측을 하나의 프로그램에서 수행할 수 있도록 프로그램을 제작하였다.

### 프로그램 작동방식

- 프로그램을 실행하면 메뉴가 출력된다. (사진 1)  
사용자는 인공지능 모델을 학습시킬지, 만들어진 모델을 통해 질병을 알아낼지 선택할 수 있다.
- 만약 1번을 선택한다면 인공지능 모델 학습을 시작한다.
- 만약 2번을 선택한다면 사용자가 입력한 사진으로 질병을 예측한다.  
사용자는 photos/ 밑에 사진을 직접 가져다 놓는 방식으로 데이터를 입력한다.  
photos/ 밑에 바로 사진을 넣어도 되고 사진이 들어있는 디렉토리를 통째로 넣어도 된다.  
예측 결과는 표 형태로 출력된다. (사진 2)  
상세한 예측 결과는 엑셀 파일에 저장된다. (사진 3)

```
===== 메뉴 =====  
1. 인공지능 모델 학습시키기  
2. 흉부 X-ray 사진으로 질병 알아내기  
3. 종료하기  
=====  
입력 >>
```

(사진 1) 메뉴

## 7. 학습된 인공지능을 문제 해결에 적용한 방법 및 결과물 설명

파일명	확률	결과
../photos/tuberculosis/Tuberculosis-6.png	96.639%	normal
../photos/tuberculosis/Tuberculosis-7.png	99.626%	tuberculosis
../photos/tuberculosis/Tuberculosis-5.png	99.246%	tuberculosis
../photos/tuberculosis/Tuberculosis-4.png	93.524%	tuberculosis
../photos/tuberculosis/Tuberculosis-1.png	100.000%	tuberculosis
../photos/tuberculosis/Tuberculosis-3.png	100.000%	tuberculosis
../photos/tuberculosis/Tuberculosis-2.png	99.991%	tuberculosis
../photos/tuberculosis/Tuberculosis-10.png	98.150%	tuberculosis
../photos/tuberculosis/Tuberculosis-9.png	99.180%	tuberculosis
../photos/tuberculosis/Tuberculosis-8.png	99.877%	tuberculosis
../photos/pneumonia/person1_virus_13.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_7.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_6.jpeg	99.986%	pneumonia
../photos/pneumonia/person1_virus_12.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_bacteria_1.jpeg	100.000%	pneumonia
../photos/pneumonia/person2_bacteria_3.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_9.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_bacteria_2.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_11.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_8.jpeg	100.000%	pneumonia
../photos/normal/covid-normal9.png	100.000%	normal
../photos/normal/covid-normal8.png	99.999%	normal
../photos/normal/covid-normal10.png	100.000%	normal
../photos/normal/covid-normal3.png	99.978%	normal
../photos/normal/covid-normal2.png	100.000%	normal
../photos/normal/covid-normal1.png	100.000%	normal
../photos/normal/covid-normal5.png	100.000%	normal
../photos/normal/covid-normal4.png	100.000%	normal
../photos/normal/covid-normal6.png	100.000%	normal
../photos/normal/covid-normal7.png	100.000%	normal
../photos/covid/COVID-1.png	100.000%	covid
../photos/covid/COVID-2.png	100.000%	covid
../photos/normal/covid-normal4.png	100.000%	normal

(사진 2) 예측결과 출력

	A	B	C	D	E	F
1	파일명	코로나19	일반	폐렴	결핵	결과
2	../photos/tuberculosis/Tuberculosis-6.png	0.00400787	96.63892984	3.7136E-06	3.357057273	normal
3	../photos/tuberculosis/Tuberculosis-7.png	0.368361198	0.002878158	0.002530872	99.62622523	tuberculosis
4	../photos/tuberculosis/Tuberculosis-5.png	0.422926946	0.102878083	0.228279945	99.24591184	tuberculosis
5	../photos/tuberculosis/Tuberculosis-4.png	0.240489212	2.568848245	3.667042777	93.5236156	tuberculosis
6	../photos/tuberculosis/Tuberculosis-1.png	4.50098E-14	1.33731E-16	9.7121E-20	100	tuberculosis
7	../photos/tuberculosis/Tuberculosis-3.png	1.04908E-05	1.77656E-19	4.38136E-20	99.99998808	tuberculosis
8	../photos/tuberculosis/Tuberculosis-2.png	0.008764079	1.20497E-09	1.33761E-08	99.99123812	tuberculosis
9	../photos/tuberculosis/Tuberculosis-10.png	1.803652942	0.016658014	0.029741044	98.14994931	tuberculosis
10	../photos/tuberculosis/Tuberculosis-9.png	0.81450725	0.003470585	0.001747833	99.1802752	tuberculosis
11	../photos/tuberculosis/Tuberculosis-8.png	0.117083162	0.003582464	0.002561547	99.87677336	tuberculosis
12	../photos/pneumonia/person1_virus_13.jpeg	1.74574E-05	6.8982E-05	99.99991655	1.67672E-09	pneumonia
13	../photos/pneumonia/person1_virus_7.jpeg	7.10453E-11	4.84673E-07	100	1.06805E-12	pneumonia
14	../photos/pneumonia/person1_virus_6.jpeg	0.000784119	0.013375378	99.98569489	0.000137628	pneumonia
15	../photos/pneumonia/person1_virus_12.jpeg	3.35138E-09	9.40437E-07	100	3.08781E-12	pneumonia
16	../photos/pneumonia/person1_bacteria_1.jpeg	3.08089E-08	0.000385264	99.99961853	3.13141E-10	pneumonia
17	../photos/pneumonia/person2_bacteria_3.jpeg	4.45635E-07	4.9012E-06	100	4.12963E-09	pneumonia
18	../photos/pneumonia/person1_virus_9.jpeg	9.19979E-07	0.000200851	99.99979734	6.61424E-09	pneumonia
19	../photos/pneumonia/person1_bacteria_2.jpeg	4.29079E-09	8.06774E-05	99.99991655	6.21652E-11	pneumonia
20	../photos/pneumonia/person1_virus_11.jpeg	6.3132E-12	1.0022E-08	100	1.83089E-14	pneumonia
21	../photos/pneumonia/person1_virus_8.jpeg	7.75114E-12	1.35881E-08	100	1.4575E-15	pneumonia
22	../photos/normal/covid-normal9.png	2.80748E-07	100	3.80919E-13	1.86432E-08	normal
23	../photos/normal/covid-normal8.png	0.000567903	99.9994278	4.0367E-09	5.07982E-06	normal
24	../photos/normal/covid-normal10.png	1.30922E-09	100	1.7624E-16	1.38602E-11	normal
25	../photos/normal/covid-normal3.png	0.019372039	99.97801185	0.000105142	0.002511648	normal
26	../photos/normal/covid-normal2.png	8.42626E-07	100	5.31441E-13	1.96513E-08	normal
27	../photos/normal/covid-normal1.png	3.81756E-06	99.99998808	1.46293E-08	1.15317E-05	normal
28	../photos/normal/covid-normal5.png	2.18629E-08	100	9.34498E-16	6.25806E-13	normal
29	../photos/normal/covid-normal4.png	3.63898E-05	99.99996424	3.52024E-11	7.73731E-07	normal
30	../photos/normal/covid-normal6.png	0.005872645	99.99363422	2.26405E-07	0.000484067	normal
31	../photos/normal/covid-normal7.png	4.23636E-07	100	1.87432E-14	5.66075E-10	normal
32	../photos/covid/COVID-1.png	99.95562434	0.044029951	1.6573E-05	0.000326157	covid
33	../photos/covid/COVID-2.png	99.99837875	9.50203E-06	0.001233875	0.00037593	covid
34	../photos/covid/COVID-3.png	99.20380116	0.791977905	6.42333E-06	0.004216052	covid

(사진 3) 상세 예측결과가 저장된 엑셀 파일

# 7. 학습된 인공지능을 문제 해결에 적용한 방법 및 결과물 설명

## 실제 실행 결과

결핵 환자, 폐렴 환자, 일반인, 코로나 환자의 흉부 X-Ray 사진을 10장씩 모아 photos/ 디렉토리 안에 넣고 프로그램을 실행시켰다.

총 40장의 사진 중 37장의 질병을 맞추었다.

나머지 3장은 결핵을 일반인으로, 코로나를 일반인으로, 코로나를 폐렴으로 잘못 예측하였다.

정답: 결핵, 예측값: 일반인

정답: 코로나, 예측값: 일반인

정답: 코로나, 예측값: 폐렴

파일명	확률	결과
../photos/tuberculosis/Tuberculosis-6.png	96.639%	normal
../photos/tuberculosis/Tuberculosis-7.png	99.626%	tuberculosis
../photos/tuberculosis/Tuberculosis-5.png	99.246%	tuberculosis
../photos/tuberculosis/Tuberculosis-4.png	93.524%	tuberculosis
../photos/tuberculosis/Tuberculosis-1.png	100.000%	tuberculosis
../photos/tuberculosis/Tuberculosis-3.png	100.000%	tuberculosis
../photos/tuberculosis/Tuberculosis-2.png	99.991%	tuberculosis
../photos/tuberculosis/Tuberculosis-10.png	98.150%	tuberculosis
../photos/tuberculosis/Tuberculosis-9.png	99.180%	tuberculosis
../photos/tuberculosis/Tuberculosis-8.png	99.877%	tuberculosis
../photos/pneumonia/person1_virus_13.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_7.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_6.jpeg	99.986%	pneumonia
../photos/pneumonia/person1_virus_12.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_bacteria_1.jpeg	100.000%	pneumonia
../photos/pneumonia/person2_bacteria_3.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_9.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_bacteria_2.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_11.jpeg	100.000%	pneumonia
../photos/pneumonia/person1_virus_8.jpeg	100.000%	pneumonia
../photos/normal/covid-normal9.png	100.000%	normal
../photos/normal/covid-normal8.png	99.999%	normal
../photos/normal/covid-normal10.png	100.000%	normal
../photos/normal/covid-normal3.png	99.978%	normal
../photos/normal/covid-normal2.png	100.000%	normal
../photos/normal/covid-normal11.png	100.000%	normal
../photos/normal/covid-normal5.png	100.000%	normal
../photos/normal/covid-normal4.png	100.000%	normal
../photos/normal/covid-normal6.png	99.994%	normal
../photos/normal/covid-normal7.png	100.000%	normal
../photos/covid/COVID-1.png	99.956%	covid
../photos/covid/COVID-2.png	99.998%	covid
../photos/covid/COVID-3.png	99.204%	covid
../photos/covid/COVID-7.png	99.974%	normal
../photos/covid/COVID-6.png	99.999%	covid
../photos/covid/COVID-4.png	91.718%	pneumonia
../photos/covid/COVID-5.png	99.566%	covid
../photos/covid/COVID-8.png	100.000%	covid
../photos/covid/COVID-9.png	97.276%	covid
../photos/covid/COVID-10.png	95.908%	covid

## 8. 제안서의 내용에 대비하여 변경된 부분에 대한 사유 및 설명

### 1. 코로나19 dataset 변경

기존 코로나19 dataset은 사진 수가 468장으로 매우 적고, 데이터 품질도 좋지 않아

<https://github.com/ieee8023/covid-chestxray-dataset> 에서

<https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database> 으로 변경하였다.

### 2. 사진 사이즈 변경

기획서에는 사진을  $300 * 300$ 으로 변환한다고 작성하였다.

하지만, 코로나19 dataset 사진의 사이즈가 모두  $256 * 256$  이므로 다른 사진의 사이즈도  $256 * 256$ 으로 통일하였다.

### 3. 데이터 분할 비율 변경

기획서에는 실행용 파일 10장, 학습용 80%, 테스트용 20%로 데이터를 분할하였다.

실행용 사진과 테스트용 사진을 구분하는 것에 의미가 없다고 판단하여 실행용 사진을 따로 분리하지 않았다.

## 8. 제안서의 내용에 대비하여 변경된 부분에 대한 사유 및 설명

### 4. CNN 필터 수와 Dropout 비율 변경

성능을 개선하고 오버피팅을 줄이기 위해 다양한 조건으로 인공지능 모델을 학습시켜 본 결과, 가장 성능이 높은 모델 구조를 발견하였다.

이 모델 구조에 맞게 필터 수와 Dropout 비율을 변경하였다.

### 5. 엑셀 파일 저장 기능 추가

프로그램을 종료하더라도 실행 결과가 사라지지 않도록 실행 결과를 엑셀 파일로 제공하였다.

감사합니다

