

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



**BÁO CÁO MÔN HỌC**  
**NHẬN DẠNG THỊ GIÁC VÀ ỨNG DỤNG**  
**PHÁT HIỆN COVID-19 TỪ HÌNH ẢNH TIA X**

Nhóm thực hiện: **Phan Thanh Duy Tân – CH1901029**

**Đặng Hoàng Duyên – CH1901019**

**Châu Duy Cảnh – CH1902015**

**GIẢNG VIÊN GIẢNG DẠY: PGS.TS. LÊ ĐÌNH DUY**

Hồ Chí Minh, ngày 20 tháng 01 năm 2021

## MỤC LỤC

<b>I. TỔNG QUAN VỀ ĐỀ TÀI</b>	4
<b>II. CƠ SỞ LÝ THUYẾT</b>	5
1. Tổng quan về bệnh COVID-19	5
1.1. COVID-19 là gì?	5
1.2. Chẩn đoán COVID qua ảnh X-quang	7
1.3. Điều trị COVID-19	10
<b>III. MÔ HÌNH VÀ THUẬT TOÁN</b>	11
1. VGG-16 là gì?	11
2. Cách thức VGG-16 hoạt động	12
<b>IV. BỘ DỮ LIỆU</b>	20
1. Bộ dữ liệu	20
2. Phân chia bộ dữ liệu:	21
<b>V. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ</b>	21
1. Đánh giá kết quả đạt được:	21
2. Đề xuất hướng nghiên cứu	22
<b>VI. TÀI LIỆU THAM KHẢO</b>	23

## LỜI MỞ ĐẦU

Ngày nay, trí tuệ nhân tạo (AI) đã trở thành một công cụ quan trọng trong mọi lĩnh vực nói chung và các ứng dụng y tế nói riêng. AI được đón nhận trên toàn cầu và được dùng để thiết kế các ứng dụng y tế nhằm hỗ trợ nhân viên y tế trong việc chẩn đoán và điều trị bệnh nhân một cách có tác dụng và hiệu quả. Virus Corona 2019 (COVID-19) hay còn gọi là Hội chứng hô hấp cấp tính nặng do Corona-2 (SARS CoV-2) là bệnh lý với các triệu chứng có thể bao gồm: sốt, ho, khó thở, đau cơ, mệt mỏi với thời gian ủ bệnh được ước tính trong khoảng từ 2 đến 14 ngày. Trường hợp nghiêm trọng có thể phát triển thành viêm phổi, hội chứng suy hô hấp cấp tính, nhiễm trùng máu và sốc nhiễm trùng có thể dẫn đến tử vong Trong vòng một 1 năm, COVID-19 đã tàn phá các thị trấn và thành phố trên khắp thế giới. Tỷ lệ mắc bệnh và tỷ lệ tử vong của COVID-19 hiện nay vẫn đang tăng lên. Trong bài tiểu luận này, nhóm sẽ thảo luận về sự cần thiết của Phát hiện COVID-19 từ hình ảnh tia X để giúp các bác sĩ đưa ra các chẩn đoán và phác đồ điều trị phù hợp và hiệu quả.

## **I. TỔNG QUAN VỀ ĐỀ TÀI**

### **1. Tầm quan trọng của chủ đề:**

Bệnh viêm đường hô hấp cấp (COVID-19) được xếp vào danh sách bệnh truyền nhiễm nhóm A, là bệnh truyền nhiễm nguy hiểm, tuy nhiên bệnh khó phát hiện sớm do dễ nhầm lẫn với các bệnh ho thông thường. Người bệnh thường đi khám và phát hiện bệnh khi bệnh đã tiến triển nặng và gây nhiều tổn thương cho phổi. Tử vong xảy ra nhiều hơn ở người cao tuổi, người suy giảm miễn dịch và mắc các bệnh mãn tính kèm theo.

### **2. Bài toán:**

Từ ảnh chụp X-quang khi thăm khám, hệ thống sẽ tiến hành đánh giá nguy cơ mắc bệnh viêm đường hô hấp cấp (COVID-19).

### **3. Lý do chọn đề tài (bài toán):**

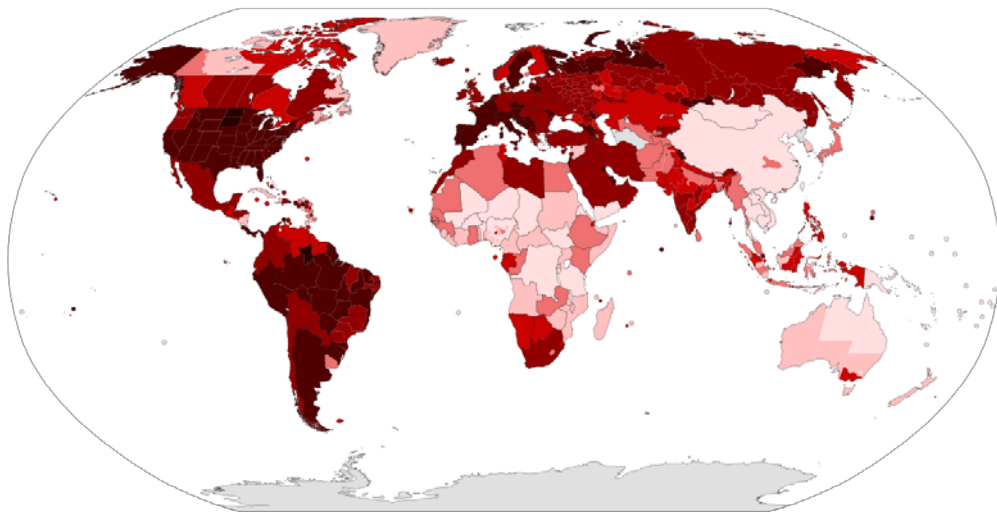
Mục tiêu phát hiện bệnh sớm, chính xác và điều trị kịp thời giúp giảm triệu chứng cũng như mức độ trầm trọng của bệnh. Giúp bác sĩ chẩn đoán nhanh và chính xác bệnh, theo dõi bệnh và cải thiện chất lượng sống của bệnh nhân.

## II. CƠ SỞ LÝ THUYẾT

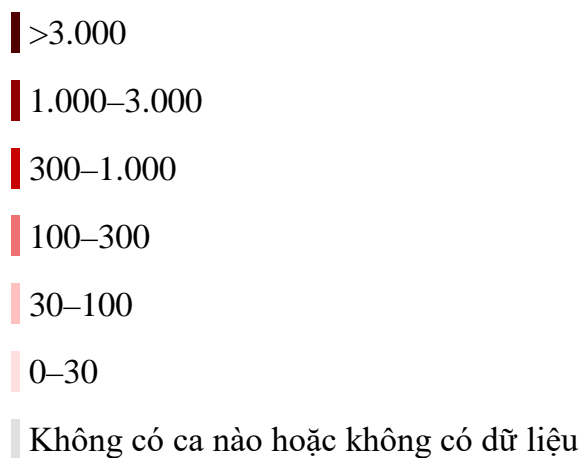
### 1. Tổng quan về bệnh COVID-19

#### 1.1. COVID-19 là gì?

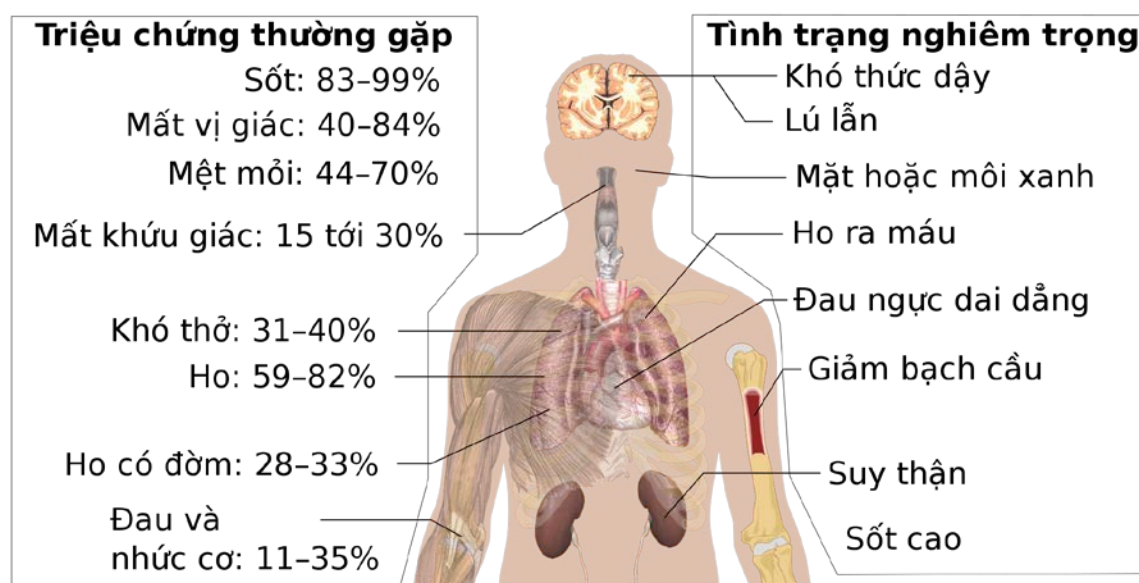
Virus gây sự bùng phát của dịch bệnh COVID-19 là SARS-CoV-2 (trước đây được gọi là virus Corona) được Tổ chức Y tế Thế Giới (WHO) đặt tên chính thức và thực hiện vào ngày 11/02/2020. Ca bệnh xác định đầu tiên được ghi nhận tại thành phố Vũ Hán, tỉnh Hồ Bắc, Trung Quốc ngày 03/12/2019. Ngày 11/03/2020, WHO nhận định dịch COVID-19 là đại dịch toàn cầu.



Bản đồ số ca nhiễm được xác nhận trên 100.000 người tính đến ngày 16 tháng 12 năm 2020



Các triệu chứng của bệnh nhân mắc bệnh từ nhẹ đến nặng bao gồm: sốt, ho và khó thở. Các triệu chứng này có thể xuất hiện từ 02 đến 14 ngày sau khi tiếp xúc nguồn bệnh. Tới khi khởi phát, virus gây sốt và có thể tổn thương đường hô hấp. Trường hợp nặng, gây viêm phổi và có thể nhiều cơ quan khác trong cơ thể khiến bệnh nhân tử vong, nhất là các trường hợp có bệnh nền. Các triệu chứng được báo cáo gồm sốt trong 90% trường hợp mắc bệnh, mệt mỏi và ho khan trong 80% trường hợp, 20% bị khó thở và suy hô hấp chiếm 15%. X-quang ngực đã tiết lộ các dấu hiệu ở cả hai phổi.



Chính phủ các quốc gia trên thế giới đã tiến hành phản ứng đáp trả nhằm bảo vệ sức khỏe người dân cũng như các nhóm cộng đồng trên toàn cầu, bao gồm: hạn chế đi lại, phong tỏa kiểm dịch, ban bố tình trạng khẩn cấp, sử dụng lệnh giới nghiêm, tiến hành cách ly xã hội, hủy bỏ các sự kiện đông người, đóng cửa trường học và những cơ sở dịch vụ, kinh doanh ít quan trọng, khuyến khích người dân tự nâng cao ý thức phòng bệnh, đeo khẩu trang, hạn chế ra ngoài khi không cần thiết, đồng thời chuyển đổi mô hình hoạt động kinh doanh, học tập, làm việc từ truyền thống sang trực tuyến. Để chủ động phòng, chống bệnh viêm đường hô hấp cấp do COVID-19, Bộ Y tế Việt Nam đã khuyến cáo người dân và cộng đồng chủ động thực hiện tốt một số biện pháp thông thường như: không đi du lịch đến các vùng có dịch bệnh; tránh mua bán, tiếp xúc với các loại động vật nuôi hoặc hoang dã, sử dụng các thực phẩm đã được nấu chín, giữ ấm cơ thể, tăng cường sức khỏe bằng ăn uống, nghỉ ngơi, sinh hoạt hợp lý, luyện tập thể thao ...

Một số tổ chức trên thế giới đang nghiên cứu vắc-xin cho bệnh này. Viện nghiên cứu Dịch tễ học và Vi sinh vật Gamaleya có trụ sở tại Moscow đã chế tạo thành công vắc-xin ngừa Covid-19 mang tên Sputnik V, được Bộ Quốc phòng Nga đăng kí vào 11/08/2020.

## 1.2. Chẩn đoán COVID qua ảnh X-quang

Các loại tổn thương phổi do COVID-19 có thể thấy trên hình ảnh X-quang là:

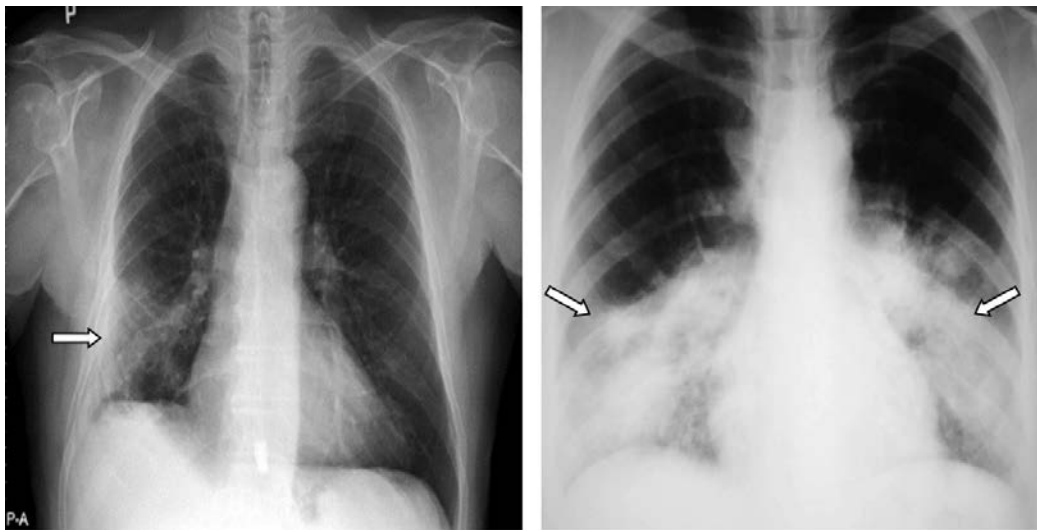
- Các dấu hiệu điển hình
  - Nốt mờ, đám mờ, kính mờ hình tròn đa ổ.
  - Đông đặc nhu mô phổi vùng ngoại vi đa ổ (nếu không phân bố ở ngoại vi thì coi là chưa xác định được)
- Các dấu hiệu không điển hình, có thể do COVID-19 hoặc các nguyên nhân khác
  - Đông đặc khu trú thùy phổi.
  - Tràn dịch màng phổi.
  - Mờ tổ chức kẽ quanh rốn phổi.
  - Dày thành phế quản, dày vách liên tiểu thùy.
  - Xẹp phổi.
  - Bệnh lý hạch lympho.

Hình ảnh kính mờ (ground glass opacity - GGO)	là tổn thương đông đặc không hoàn toàn, có tỷ trọng cao hơn nhu mô phổi xung quanh vẫn có thể thấy đường bờ các mạch máu hoặc phế quản bên trong tổn thương đó.
Hình ảnh nốt mờ	là những hình mờ có đường kính dưới 3cm, dạng hình tròn, có thể đơn độc, có thể rải rác trong nhu mô phổi. Các nốt phổi thường có ranh giới rõ, được bao quanh bởi nhu mô phổi và không liên tục với rốn phổi hay trung thất.
Hình ảnh dày thành phế quản	là những tổn thương thể hiện thành của phế quản dày lên, do có sự tích tụ dịch hay chất nhày xung quanh thành phế quản, trong mô kẽ.

Hình ảnh dày vách liên thùy

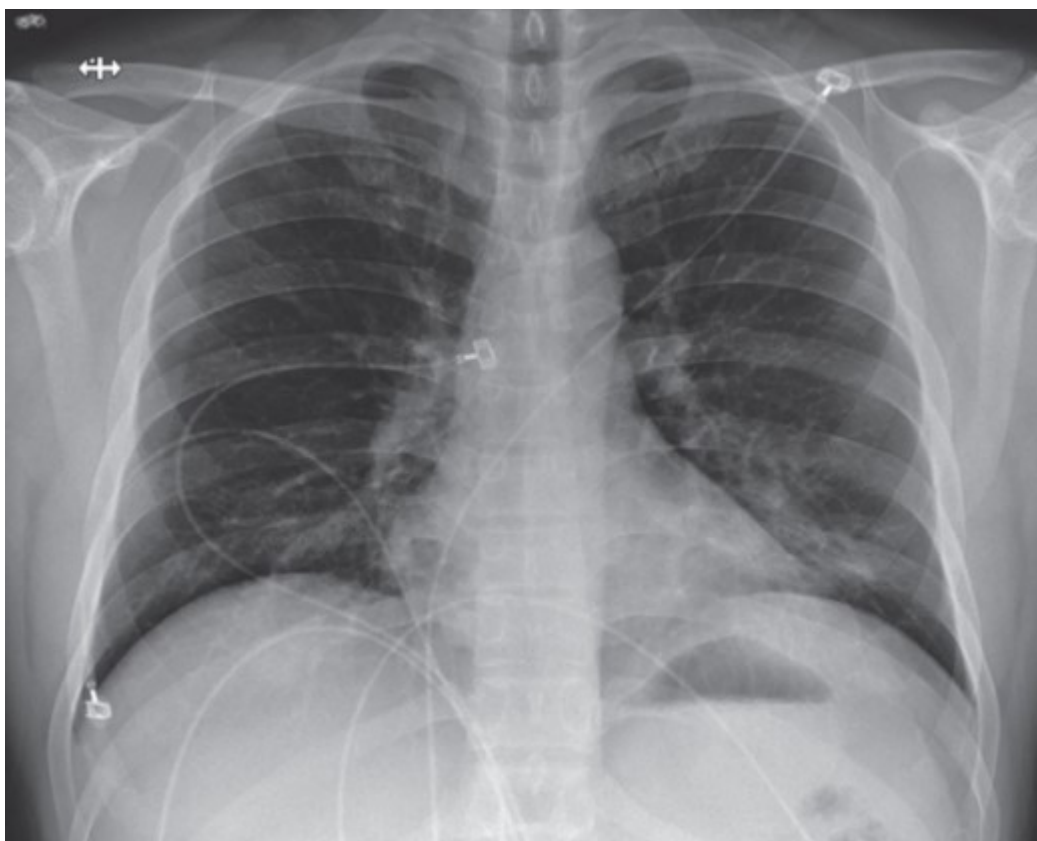
liên quan đến dịch rãnh liên thùy, thâm nhiễm tế bào hoặc xơ hóa. Trong viêm phổi do virus, dày các vách liên thùy gặp trong tổn thương lan tỏa trong ARDS.

Hình ảnh tổn thương phổi do virus bao gồm các dấu hiệu không đặc hiệu. Ở giai đoạn sớm hình ảnh X quang có thể bình thường. Các dấu hiệu hay gặp là tổn thương dạng kính mờ, đông đặc ở ngoại vi, thùy dưới của phổi. Ít khi gặp dấu hiệu tạo hang hay tràn dịch, tràn khí màng phổi. Các tổn thương trên phim tiến triển nhanh. Nếu không điều trị kịp thời có thể nhanh chóng dẫn đến suy hô hấp cấp tính. Nếu tiến triển tốt các tổn thương nhanh chóng mất đi, có thể để lại các tổn thương xơ hai phổi.

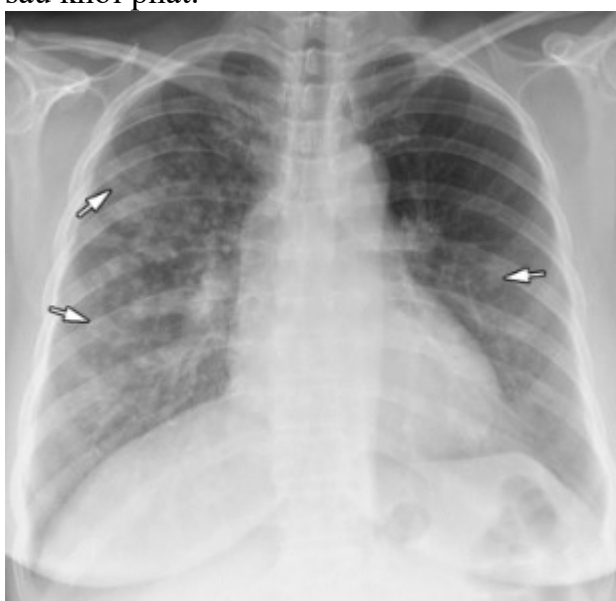


Hình ảnh tổn thương kính mờ phổi phải (trái) và đông đặc thùy dưới hai phổi (phải) trên X-quang phổi thẳng.

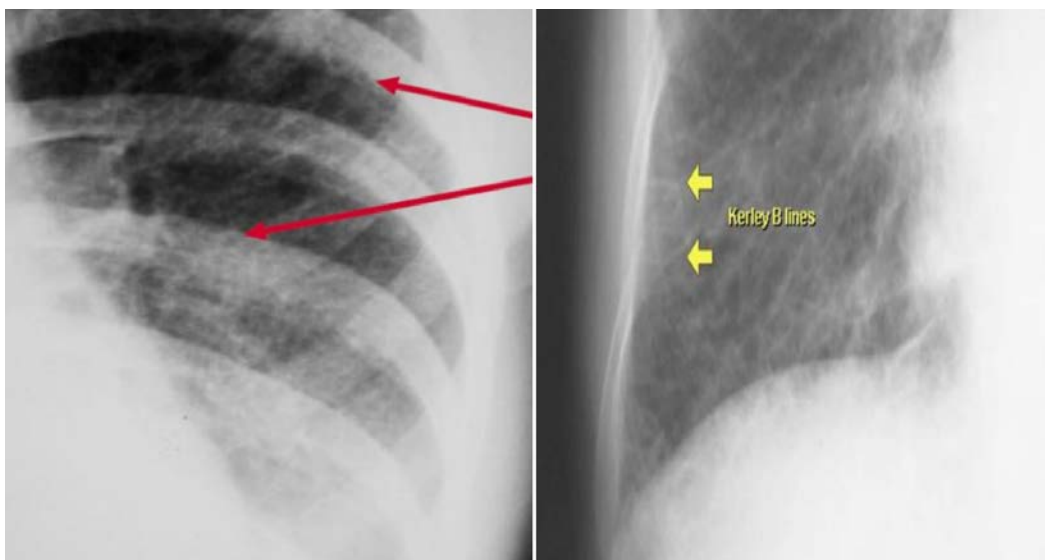




Hình ảnh lưới và đông đặc nền phổi hai bên trên X-quang phổi thẳng chụp ngày thứ 9 sau khởi phát.



Hình ảnh nốt mờ trên X-quang phổi thẳng.



Hình ảnh dày thành phế quản (trái) và đường Kerley B do phù mô kẽ (phải) trên X-quang phổi thẳng.

### 1.3. Điều trị COVID-19

Tập trung chính là điều trị suy hô hấp. Bệnh nhân cần được theo dõi sát sao, đặc biệt ngày thứ 7-10 của bệnh, sử dụng các dấu hiệu lâm sàng, các thang điểm cảnh báo sớm, theo dõi tiến triển hằng ngày X-quang phổi của bệnh nhân để phát hiện và xử trí kịp thời các biến chứng/tiến triển nặng của bệnh.

Với bệnh nhân suy hô hấp nặng, bộ Y tế Việt Nam khuyến cáo đặt ống nội khí quản sớm, và thở máy xâm nhập, và chỉ cân nhắc các biện pháp hỗ trợ hô hấp không xâm nhập cho từng trường hợp cụ thể, chứ không áp dụng thường quy và cần theo dõi sát bệnh nhân.

Khi các dấu hiệu lâm sàng được cải thiện và hai mẫu liên tiếp bệnh phẩm đường hô hấp (cả dịch tỵ hầu và dịch họng), lấy cách nhau  $\geq 24$  giờ âm tính với COVID-19 thì bệnh nhân có thể xuất viện. Sau khi ra viện, bệnh nhân tiếp tục được cách ly tại nhà thêm 14 ngày. Bệnh nhân cần được ở phòng riêng thoáng khí, đeo khẩu trang, ăn riêng, vệ sinh tay, hạn chế tiếp xúc với các thành viên khác trong gia đình và không được ra ngoài. Bệnh nhân cần được đo thân nhiệt 2 lần/ngày và khám lại ngay nếu sốt hoặc có dấu hiệu bất thường khác.

Về các thuốc kháng virus đặc hiệu (như lopinavir/ritonavir, chloroquine, hydroxychloroquine, remdesivir...): Do chưa có đủ các bằng chứng về hiệu quả và an toàn của những thuốc này trong điều trị COVID-19 nên bộ Y tế chưa khuyến cáo áp

dụng thường quy trong điều trị (ngoài phạm vi sử dụng trong các nghiên cứu thử nghiệm lâm sàng ở Việt Nam).

### III. MÔ HÌNH VÀ THUẬT TOÁN

#### 1. VGG-16 là gì?

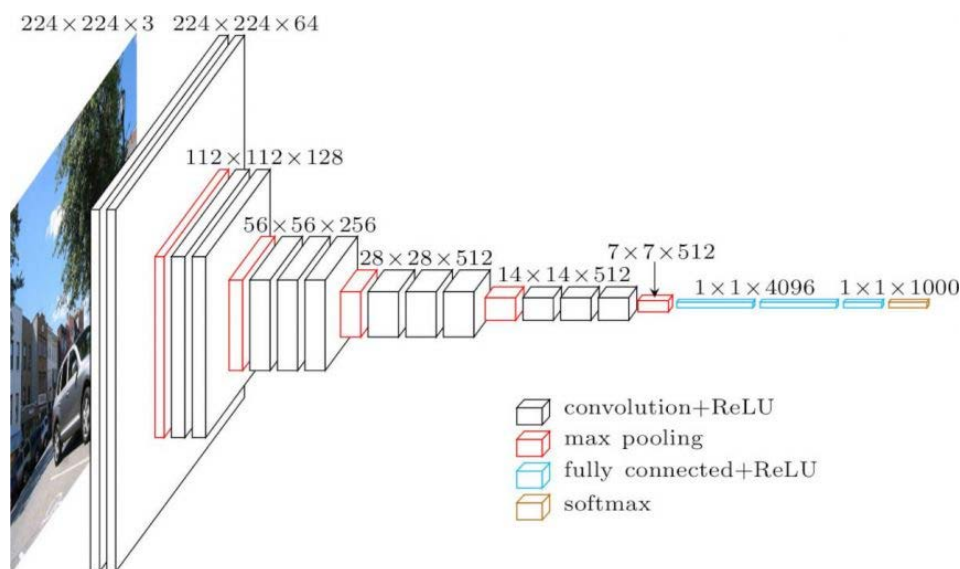
Kiến trúc mạng VGG đã được Simonyan và Zisserman giới thiệu trong bài báo năm 2014 “Very Deep Convolutional Networks for Large Scale Image Recognition”. VGG đặc trưng bởi sự đơn giản của nó, chỉ sử dụng các convolutional layers  $3 \times 3$  xếp chồng lên nhau theo chiều sâu ngày càng tăng. Giảm kích thước phải xử lý bằng max pooling. Với hai lớp fully-connected layers, mỗi lớp có 4.096 nút. Tiếp sau đó là một lớp softmax classifier

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input ( $224 \times 224$ RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

16 và 19 đại diện cho số lớp trọng lượng trong mạng (cột D và E trong hình). Mạng VGG-16 sâu hơn so với AlexNet và số lượng tham số của nó lên tới 138 triệu tham số. Đây là một trong những mạng có số lượng tham số lớn.

Vào năm 2014, thì 16 và 19 layer networks đã được xem là kiến trúc có độ sâu tốt. Tuy nhiên, hiện nay một số kiến trúc mới đã đạt được nhiều hơn thế như kiến trúc ResNet train với độ sâu 50-200 cho ImageNet với độ sâu hơn 1,000 cho CIFAR-10.

## 2. Cách thức VGG-16 hoạt động



Input đầu vào cho conv1 layer là ảnh với kích thước  $224 \times 224$  RGB. Hình ảnh sẽ đi qua các lớp convolutional layer, ), trong đó các bộ lọc được sử dụng với receptive field rất nhỏ:  $3 \times 3$ . Sải của convolution được cố định là 1 pixel; phần đệm spatial của convolution layers giúp spatial resolution được bảo toàn sau khi chạy qua lớp convolution. Spatial pooling sẽ gồm 5 layers max-pooling. Max-pooling được thực hiện với cửa sổ trượt là  $2 \times 2$ , với sải là 2.

Ba lớp Fully-Connected (FC) layers theo sau một chồng các lớp convolution layers: hai lớp đầu tiên có 4096 channels mỗi lớp, lớp thứ ba có 1000 channels và thực hiện

phân loại 1000-way ILSVRC. Lớp layer cuối là soft-max layer. Config cho các lớp Fully-Connected (FC) layers là giống nhau ở mọi network.

Nhược điểm của VGG-16 là thời gian train lâu và kiến trúc của hệ thống khá nặng.

Code

#import lib cần thiết

```
import numpy as np # linear algebra
import random

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os

import tensorflow as tf

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv2D, Flatten, Dropout, MaxPooling2D
from tensorflow.keras.preprocessing.image import ImageDataGenerator, array_to_img,
img_to_array, load_img

from tensorflow.keras import optimizers

from tensorflow.keras import backend as K

from tensorflow.keras.models import load_model

from tensorflow.keras.applications.vgg16 import VGG16
from tensorflow.keras.applications.vgg16 import preprocess_input
from tensorflow.keras.callbacks import ModelCheckpoint

import PIL

import matplotlib.pyplot as plt

import json

from IPython.display import Image as disp_image
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
import seaborn as sns

import matplotlib.pyplot as plt

from PIL import Image

import cv2

import shutil

from glob import glob

# Helper libraries

import matplotlib.pyplot as plt

import math
```

```
%matplotlib inline#
```

## #Thư mục dữ

```
data_root='/input/covidct/'
```

```
path_positive_cases = os.path.join('/input/covidct/NCOVID/')
```

```
path_negative_cases = os.path.join('/input/covidct/NonCOVID/')
```

```
# file hình ảnh định dạng png hoặc jpg
```

```
positive_images_ls = glob(os.path.join(path_positive_cases,"*.png"))
```

```
negative_images_ls = glob(os.path.join(path_negative_cases,"*.png"))
```

```
negative_images_ls.extend(glob(os.path.join(path_negative_cases,"*.jpg")))
```

## #2 lớp NCOVID và NonCOVID

```
covid = {'class': NCOVID, 'path': path_positive_cases, 'images': positive_images_ls}
```

```
nonCovid = {'class': NonCOVID, 'path': path_negative_cases, 'images': negative_images_ls}
```

## # Tạo thư mục train - test

```
subdirs = ['train/', 'test/']
```

```
for subdir in subdirs:
```

```
    labldirs = ['NCOVID', 'NonCOVID']
```

```
    for labldir in labldirs:
```

```
        newdir = subdir + labldir
```

```
        os.makedirs(newdir, exist_ok=True)
```

## #Phân chia tập huấn luyện và tập kiểm tra

```
random.seed(237)
```

## #Tỷ lệ bộ kiểm tra

```
test_ratio = 0.2
```

```
for cases in [covid, nonCovid]:
```

```
    total_cases = len(cases['images'])#Tất cả hình ảnh
```

```
    num_to_select = int(test_ratio * total_cases)
```

```
    print(cases['class'], num_to_select)
```

```
    list_of_random_files=random.sample(cases['images'], num_to_select)
```

```
    for files in list_of_random_files:
```

```
        shutil.copy2(files, 'test/' + cases['class'])
```

## #Tạo tập huấn luyện và tập thử nghiệm

```
for cases in [covid, nonCovid]:
```

```
    image_test_files = os.listdir('test/' + cases['class'])
```

```
for images in cases['images']:

    if images.split('/')[1] not in (image_test_files):

        shutil.copy2(images, 'train/' + cases['class'])
```

**#Sử dụng điều chỉnh hình ảnh VGG để nâng cao hình ảnh**

```
img_height = 224
img_width = 224
channels = 3
batch_size = 32
epochs = 200
train_datagen = ImageDataGenerator(rotation_range=40,
                                    width_shift_range=0.2,
                                    height_shift_range=0.2,
                                    rescale=1./255,
                                    shear_range=0.2,
                                    zoom_range=0.2,
                                    #horizontal_flip=True,
                                    fill_mode='nearest',
                                    validation_split=0.2)
```

```
test_datagen = ImageDataGenerator(rescale=1./255)
training_set = train_datagen.flow_from_directory(
    './train',
    target_size = (img_height, img_width),
    batch_size = batch_size,
    class_mode = 'binary',
    subset = 'training',
    shuffle=True)
validation_set = train_datagen.flow_from_directory(
    './train',
    target_size = (img_height, img_width),
    batch_size = batch_size,
    class_mode = 'binary',
```

```

        subset = 'validation',
        shuffle=False)
test_set = train_datagen.flow_from_directory(
    './test',
    target_size = (img_height, img_width),
    batch_size = 1,
    shuffle = False,
    class_mode = 'binary')
print(training_set.class_indices)
img_files = os.listdir('train/NonCOVID')
img_path = img_files[np.random.randint(0,len(img_files))]
img = cv2.imread('train/NonCOVID/{}'.format(img_path))

# VGG16 trọng số

model = VGG16
(weights='input/vgg16weight/vgg16_weights_tf_dim_ordering_tf_kernels_n
otop.h5', include_top=False, input_shape = (img_width, img_height,
channels))

for layer in model.layers[:-5]:
    layer.trainable = False

# Xóa lớp mạng được kết nối đầy đủ

# Thiết kế mạng

top_model = Sequential()

top_model.add(model)

top_model.add(Flatten())

top_model.add(Dense(256, activation='relu'))

top_model.add(Dropout(0.5))

top_model.add(Dense(1, activation='sigmoid'))

print(model.summary())

print(top_model.summary())

# Huấn luyện mạng

```



```

top_model.compile(loss='binary_crossentropy',

                  optimizer=optimizers.RMSprop(lr=1e-4, decay=1e-6),

                  metrics=['accuracy'])

history = top_model.fit_generator(

    training_set,

    steps_per_epoch=training_set.n // batch_size,

    epochs=epochs,

    validation_data=validation_set,

    validation_steps=validation_set.n // batch_size)

# Lưu mạng

top_model.save('covid_model.h5', save_format='h5')

# Các chỉ số phân tích đồ thị

acc = history.history['accuracy']

val_acc = history.history['val_accuracy']

loss = history.history['loss']

val_loss = history.history['val_loss']

epochs = range(len(acc))

plt.plot(epochs, acc, 'b', label='Training Accuracy')

plt.plot(epochs, val_acc, 'r', label='Validation Accuracy')

plt.title('Training and validation accuracy')

plt.legend()

plt.figure()

plt.plot(epochs, loss, 'b', label='Training Loss')

plt.plot(epochs, val_loss, 'r', label='Validation Loss')

plt.title('Training and validation loss')

plt.legend()

plt.show()

```



```

verbose=1)

print('Test loss: ', test_pred[0])

print('Test accuracy: ', test_pred[1])

test_pred = top_model.predict(test_set,

                               steps=test_set.n,

                               use_multiprocessing=False,

                               verbose=1)

test_pred_class = (test_pred >= 0.5)*1

# Ma trận

cm = confusion_matrix(test_set.classes, test_pred_class)

ax= plt.subplot()

sns.heatmap(cm, annot=True, ax = ax); #annot=True to annotate cells

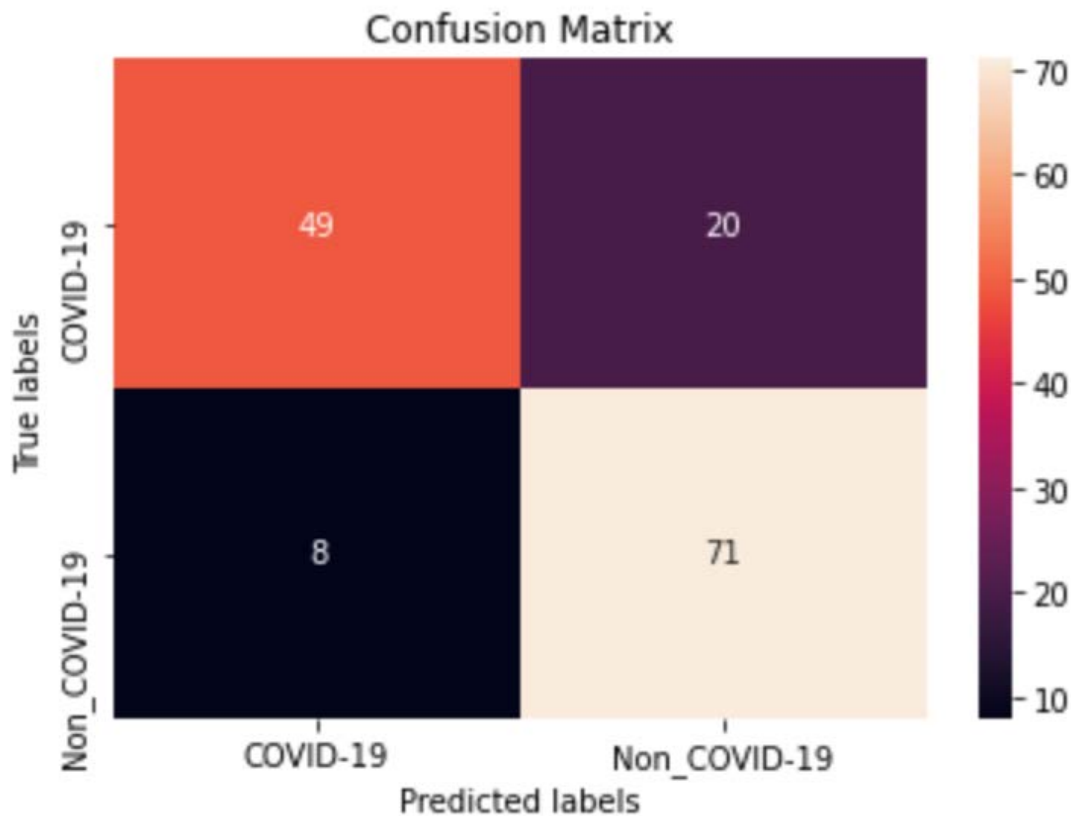
ax.set_xlabel('Predicted labels');ax.set_ylabel('True labels');

ax.set_title('Confusion Matrix');

ax.xaxis.set_ticklabels(['COVID-19', 'Non_COVID-19']);

ax.yaxis.set_ticklabels(['COVID-19', 'Non_COVID-19']);

```



## IV. BỘ DỮ LIỆU

### 1. Bộ dữ liệu

Bộ dữ liệu gồm :

- COVID19-1.zip - COVID19-31.zip: 31 tệp zip chứa ảnh chụp CT viêm phổi do COVID-19.
- CP-1.zip - CP-32.zip: 32 tệp zip chứa ảnh chụp CT của bệnh viêm phổi thông thường.
- Normal-1.zip - Normal-27.zip: 27 tệp zip chứa scan CT của các normal controls.
- metadata.csv : tệp csv chứa metadata, bao gồm id bệnh nhân, scan id, tuổi, giới tính, bệnh hiểm nghèo, chức năng gan, chức năng phổi và thời gian tiến triển.
- unzip\_filenames.csv: tệp csv làm tài liệu tham khảo để kiểm tra và đảm bảo rằng tất cả các tệp đã được tải xuống thành công.

- lesions\_slices.csv: tệp csv kiểm tra và theo dõi các lát cắt có tổn thương (lesions\_slices) trong tập dữ liệu.

## 2. Phân chia bộ dữ liệu:

Dữ liệu dùng để huấn luyện (Train): 27201(NonCovid) - 12520 (NCovid)

Dữ liệu dùng để thẩm định (Validation): 9107(NonCovid) - 4529 (NCovid)

Dữ liệu dùng để kiểm thử (Test): 9450(NonCovid) - 4346(NCovid)

## V. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ

### 1. Đánh giá kết quả đạt được:

**Confusion matrix:**

	Positive	Negative
Positive	TP (True Positive )	FN (False Negative )
Negative	FP (False Positive)	TN (True Negative )

Trong đó: Positive: dương tính

Negative: âm tính

TP: Kết quả trả về đúng (dương tính - dương tính)

TN: Kết quả trả về đúng (âm tính - âm tính)

FP: Kết quả trả về sai (âm tính - dương tính)

FN: Kết quả trả về sai (dương tính - âm tính)  
("Input" - "Output")

**Kết quả kiểm thử:** Trên 67153 mẫu dữ liệu (ảnh)

	Positive	Negative
Positive	93.23%	3.80%
Negative	4.15%	93.44%

**Sensitivity** (độ nhạy):  $TP / (TP+FN) = 90.5\%$

**Specificity** (độ đặc hiệu):  $TN / (TN+FP) = 91.12\%$

**Precision** (độ chính xác):  $TP / (TP+FP) = 90.78\%$

**Recall** (độ phủ):  $|TP| / (|TP|+|FN|) = 50\%$

**F1 score** (tính trung bình có trọng số của Precision và Recall):

$$2*TP / (2*TP + FP + FN) = 48\%$$

**Accuracy** (sự chính xác):

$$(TP + TN) / (TP + TN + FN + FP) = 94.1 \%$$

**Đánh giá chung về kết quả đạt được:**

Nhìn chung kết quả là chấp nhận được, tuy nhiên, về bộ dữ liệu hiện tại lấy từ internet có thể đã qua chỉnh sửa, nhóm sẽ tiến hành thử nghiệm trên dữ liệu thật của bệnh viện và đánh giá lại kết quả của mô hình.

**2. Đề xuất hướng nghiên cứu**

Do đề tài còn hạn chế về mặt thời gian nên chưa thể đi sâu vào phân tích, cải thiện tính chính xác và kết quả đạt được.

Cần phải thực hiện trên nhiều mô hình khác nhau để thấy được sự khác biệt và đánh giá được độ hiệu quả, tính chính xác từ đó phân tích và cải thiện mô hình cho kết quả tốt hơn.

## VI. TÀI LIỆU THAM KHẢO

<http://ncov-ai.big.ac.cn/download?lang=en>  
<https://www.pyimagesearch.com/2020/03/16/detecting-covid-19-in-x-ray-images-with-keras-tensorflow-and-deep-learning/>  
<https://viblo.asia/p/mot-vai-hieu-nham-khi-moi-hoc-machine-learning-4dbZNoDnlYM>  
<https://github.com/jeremykohn/rid-covid>  
<https://colab.research.google.com/drive/1DmWD4xilACIks7joVkiS9LoZwYgcHIgS>  
<https://github.com/sanjeevpalla/COVID-19-Detection> => here  
<https://medium.com/@sanjeev.palla/detecting-covid-19-in-x-ray-images-using-vgg16-cnn-model-4258c5ec8058>  
<https://github.com/zeeshannisar/COVID-19>  
<https://github.com/kaushikjadhav01/COVID-19-Detection-Flask-App-based-on-Chest-X-rays-and-CT-Scans>  
<https://colab.research.google.com/github/IliasPap/COVIDNet/blob/master/COVIDNet.ipynb>  
<https://github.com/ieee8023/covid-chestxray-dataset/tree/master/scripts>  
<https://github.com/muhammedtal0/COVID-19>  
<https://www.youtube.com/watch?v=nHQDDAAzIsI>  
<https://towardsdatascience.com/covid-19-detector-flask-app-based-on-chest-x-rays-and-ct-scans-using-deep-learning-a0db89e1ed2a>  
<https://github.com/ieee8023/covid-chestxray-dataset>  
<https://github.com/UCSD-AI4H/COVID-CT/tree/master/Data-split>  
Bài báo:  
<https://arxiv.org/pdf/2007.10785.pdf>  
<https://www.medrxiv.org/content/10.1101/2020.05.22.20110817v1.full.pdf>  
<https://core.ac.uk/download/pdf/323462348.pdf>  
[https://assets.researchsquare.com/files/rs-34534/v1\\_stamped.pdf](https://assets.researchsquare.com/files/rs-34534/v1_stamped.pdf)  
<https://www.benhvien108.vn/dac-diem-ton-thuong-phoi-do-covid-19-tren-hinh-anh-x-quang-va-cat-lop-vi-tinh.htm>  
<https://radiologyassistant.nl/chest/covid-19/covid19-imaging-findings>  
<https://www.uclahealth.org/radiology/workfiles/pdf/UCLA-covid19-chest-radiographic-findings>  
Mathias P., Wouter V.E, et al (2020), CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19 - Definition and Evaluation, Radiology, 296(2), 97-104.  
Tomás Franquet, MD, PhD. Imaging of Pulmonary Viral Pneumonia. Radiology: Volume 260: Number 1—July 2011.  
Lan T. Phan, Thuong V. Nguyen, Quang C. Luong, Thinh V. Nguyen, Hieu T. Nguyen, et al, Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam, The New England Journal of Medicine, January 28, 2020.  
<https://www.sciencealert.com/wuhan-coronavirus-can-be-infectious-before-people-show-symptoms-official-claims>  
<https://neurohive.io/en/popular-networks/vgg16/>