Apply EfficientNetV2 deep learning model to detect cases of lung diseases on chest X-ray images

Nguyễn Đức Anh

Advisor: Lương Trung Kiên

Abstract: The emergence of the COVID-19 pandemic has not only adversely affected the global economy, but also put great pressure on the medical systems of countries, especially developing countries. That also opens up opportunities to apply scientific and technological advances to support people in medical matters. Typically, applying image recognition technology to diagnose diseases. Chest X-ray radiographic (CXR) imagery enables earlier and easier lung disease diagnosis. Therefore, in this project, we propose a deep learning method using a transfer learning technique to classify lung diseases on CXR images to improve the efficiency and accuracy of computer-aided diagnostic systems' (CADs') diagnostic performance. Our proposed method is a one-step, end-to-end learning, which means that raw CXR images are directly inputted into a deep learning model (EfficientNet v2-S) to extract their meaningful features in identifying disease categories. Datasets are compiled based on multiple sources published on Kaggle about four classes of normal, pneumonia, covid-19, and tuberculosis. With a dataset of more than 22,000 lung X-ray images of various diseases, it ensures that the model is carefully trained to provide high recognition and prediction performance.

Keywords: X-ray, EfficientNet, Pneumonia; Puberculosis, Covid-19

1. Introduction

Lung disease is one of the top causes of death worldwide and includes covid-19, pneumonia, and tuberculosis diseases. For example, pneumonia is a disease that causes air sacs in the lungs. Its symptoms include fever, cough, shortness of breath, loss of appetite, stabbing chest pain, low energy, vomiting, and confusion, especially in older people. Research data shows that the most common and dangerous complication in patients infected with covid-19 is pneumonia. In the early days of infection, the virus attacks lung cells, mainly the protective layer of cilia, then peels off this protective layer. When the protective layer is lost, the respiratory tract will be contaminated with dirt fluids and viruses overflow. This attack method causes many Covid-19 patients to have inflammation of both lungs. When it reaches stage 3, severe lung damage will continue to spread, leading to acute respiratory failure (rapid breathing, shortness of breath, cyanosis). Tuberculosis is a disease provoked by an airborne bacterial infection in the lung. Symptoms of tuberculosis include a cough for three consecutive weeks, a loss of appetite that yields unintentional weight loss, chills, fever, and night sweats.

Therefore, early assessment and diagnosis can significantly reduce the life-threatening nature of lung diseases and improve the quality of life of suffering patients. In modern medical image modalities, imaging tests are extremely powerful tools that can help doctors diagnose a range of conditions. The most commonly used image modalities are the chest Diagnostics. Diagnosis is through X-ray imaging (CXR) and computed tomography (CT). They are diagnostic tools that allow doctors to see structures inside the body without cutting. Lung disease detection is currently performed through an examination of CXR images by a professional radiologist, due to its convenient and non-invasive assessment for overall findings of the chest situation in brief. It is also suitable for follow-up examination since disease changes can be observed more easily and earlier. However, there is a common human error that may be caused by the misreading of a CXR image due to the complex anatomical structure of the chest. Therefore, computer-aided diagnostic systems (CADs) are used to help radiologists overcome clinical decisions with more precise diagnosis and to minimize misreading. To improve the efficiency and accuracy of the diagnostic performance of CADs, the CXR image has been widely exploited by various methods.

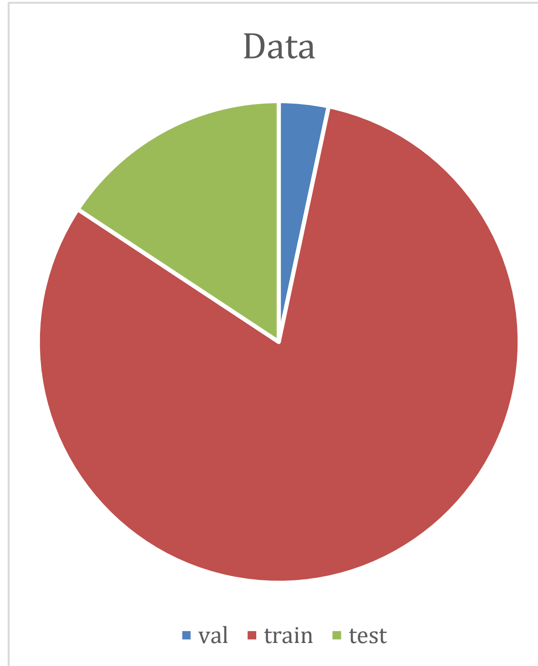
The deep learning (DL) method has outperformed traditional machine learning in medical image analysis on classification, segmentation, and detection tasks which apply to various data modalities ranging from a 1D signal to a 3D image. Compared to the natural RGB image, the CXR image is a grayscale image with one-channel information. The information of the CXR image is scanned by an X-ray, also called a radiograph. The X-ray sends radiation through the body. Areas with high levels of calcium (i.e., bones) block the radiation, causing them to appear white on the image. Soft tissues (i.e., lungs, heart, liver, muscle, and more) allow the radiation to pass through and appear in a range from gray to black on the image. The CXR image is commonly stored as a one-channel grayscale file in png or tiff format, which causes a unique challenge for the DL method. Some research applied the DL model to the one-channel grayscale CXR image by learning from scratch (with random, initialized weights), which yielded limited results. To overcome this limitation, an intuitive method is to collect more data. However, both the collection and labeling of medical data require experts and consume a lot of time and money. In this project, we propose the DL method using a transfer learning technique to classify lung diseases in CXR images to improve the efficiency and accuracy of a CADs' diagnostic performance. We exploited the convolutional neural network (CNN) model in general (EfficientNet v2-S model in specific) to classify lung diseases using transfer learning techniques with empirical hyperparameters to achieve a significant performance. Our proposed method was a one-step, end-to-end learning, which means that raw CXR images were directly inputted into a deep learning model (EfficientNet v2-S) to extract their meaningful features in identifying disease categories.

2. Data Preparation

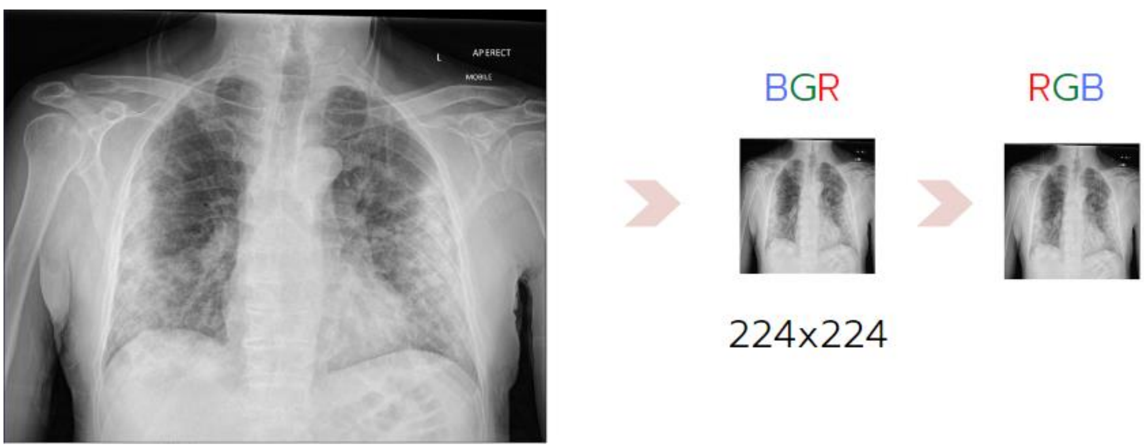
The dataset is compiled from sources on Kaggle including 23,182 files for 3 folders test (771), train (18768), validation (3643). Contains x-ray images of normal people or of patients with Covid-19 or Pneumonia or Tuberculosis.

The image will be in PNG format of the patient's lung X-ray with dimensions of 510x510 and has been labelled for tuberculosis, covid-19, pneumonia and normal.

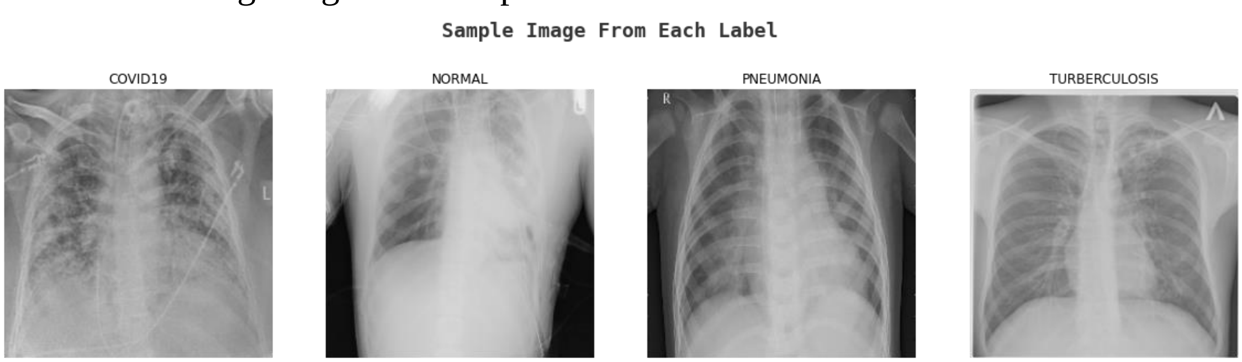
Because this is public data, there is no additional cost to the project. However, that also brings certain risks such as uneven image quality and unreliable reliability. Because to enrich the data, I have merged two data sets, so there may appear duplicate images that negatively affect the model.



After taking the image from the input with cv2.imread, we get an image in original size and formatted in BGR color code. To fit the model, we convert it back to RGB color code and resize it to 224x224



And the resulting image is as sample for each label



3. Methods

Every CNN network has three basic input parameters: width, depth, and resolution. A larger model has a better performance but the parameters will also increase. Therefore, the authors Mingxing Tan and Quoc V. Le found that when systematically scaling the width, depth and resolution parameters, the accuracy of the model can be increased, but the parameters do not increase too much. It's EfficientNet.

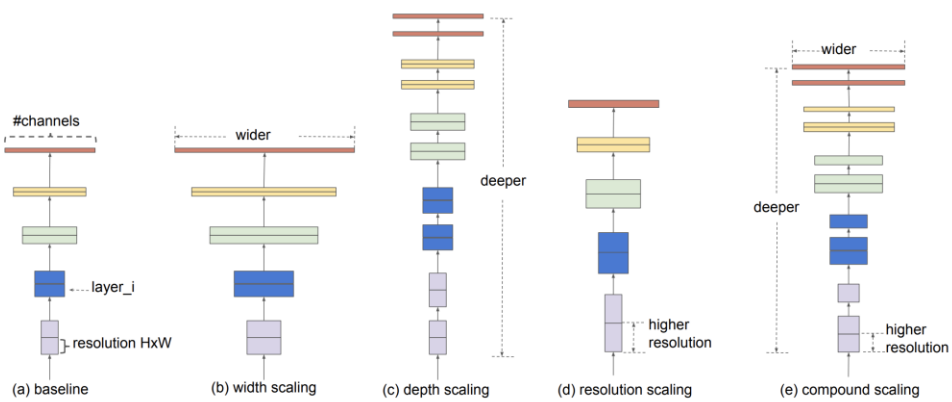


Figure 2. **Model Scaling.** (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

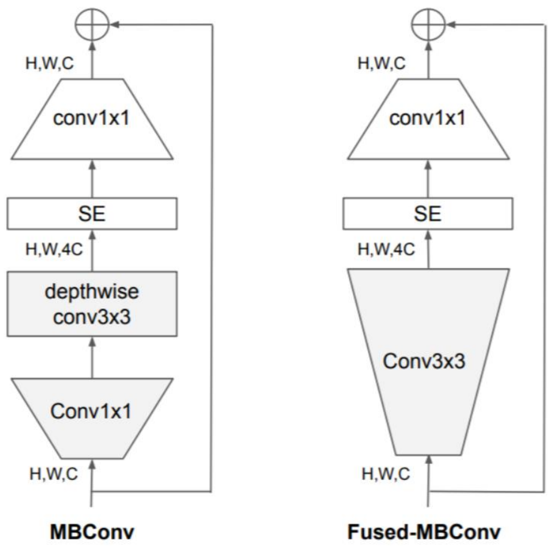
Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Table-1: EfficientNetV1-B0 baseline architecture

After the resounding success of EfficientNet architecture, Mingxin Tan and Quoc V. Le have launched a new cluster network with faster training speed and better parameter efficiency - EfficientNetV2.

We already know that the EfficientNetV1 architecture utilized MBConv layers with depthwise convolutions. Depthwise convolutions have fewer parameters and FLOPs than regular convolutions, but they often cannot fully utilize the modern accelerators. This means that a reduction in FLOPs doesn't necessarily lead to improvement in training speed.

Recently, Fused-MBConv was proposed which replaces the depthwise 3x3 convolution and expansion 1x1 convolution in MBConv with a regular 3x3 convolution as shown in figure above.



To compare these two building blocks and performance improvement, the authors of the EfficientNetV2 architecture gradually replaced the original MBConv in EfficientNetV1 with Fused-MBConv and the results are shown in table-2 below.

	Params (M)	FLOPs (B)	Top-1 Acc.	TPU imgs/sec/core	V100 imgs/sec/gpu
No fused	19.3	4.5	82.8%	262	155
Fused stage1-3	20.0	7.5	83.1%	362	216
Fused stage1-5	43.4	21.3	83.1%	327	223
Fused stage1-7	132.0	34.4	81.7%	254	206

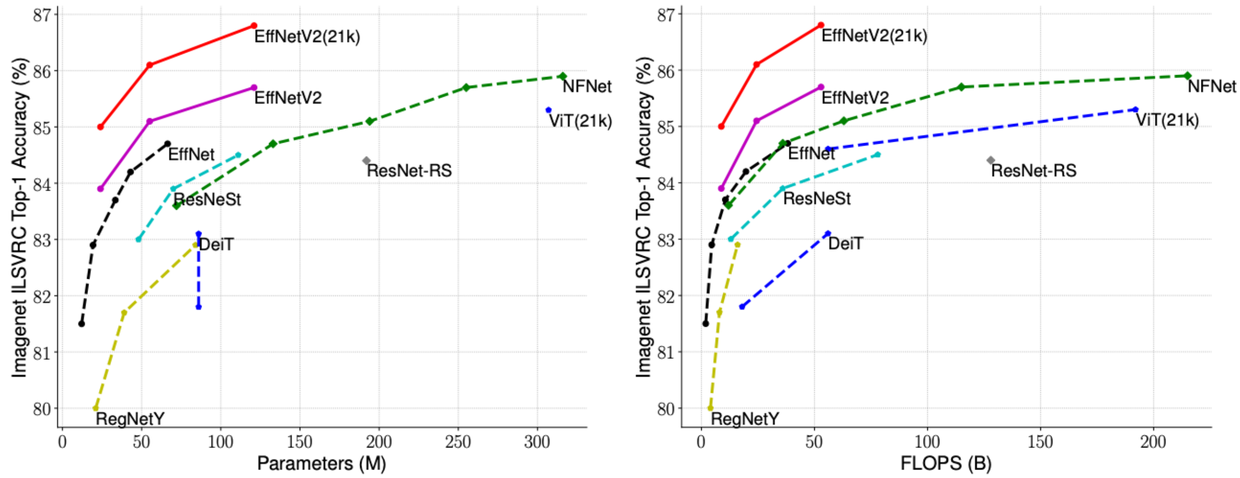
Table-2: Replacing MBConv with Fused-MBConv. No fused denotes all stages use MBConv, Fused stage1-3 denotes replacing MBConv with Fused-MBConv in stage {2, 3, 4} as in table-1.

As can be seen in table-2 above, it was observed that when Fused-MBConv is applied in the early layers (stages 1-3), this can lead to an improvment in the training speed with a small overhead on parameters and FLOPs.

Table-3: EfficientNetV2-S architecture

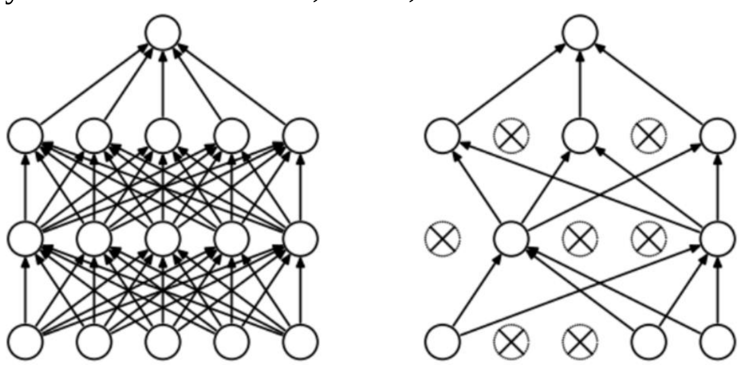
Stage	Operator	Stride	#Channels	#Layers
0	Conv3x3	2	24	1
1	Fused-MBConv1, k3x3	1	24	2
2	Fused-MBConv4, k3x3	2	48	4
3	Fused-MBConv4, k3x3	2	64	4
4	MBConv4, k3x3, SE0.25	2	128	6
5	MBConv6, k3x3, SE0.25	1	160	9
6	MBConv6, k3x3, SE0.25	2	272	15
7	Conv1x1 & Pooling & FC	-	1792	1

Compared to EfficientNetV1, the EfficientNetV2 architecture trains upto 11x faster while being 6.8x smaller. And much better than existing models.



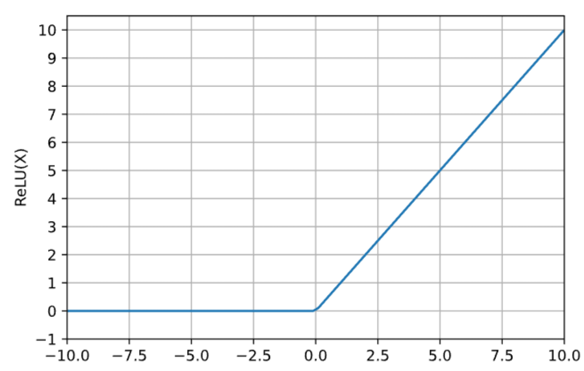
(a) Parameters (b) FLOPs

After obtaining Efficientnet V2s model with pre-train='imagenet'. I tuning a bit more to better fit this project. First, add a Dropout layer with ratio = 0.2 to combat over-fitting and Dropout forces the neural network to find more robust features, with the characteristic that they must be more useful, better, and tastier when associated with many other neurons.

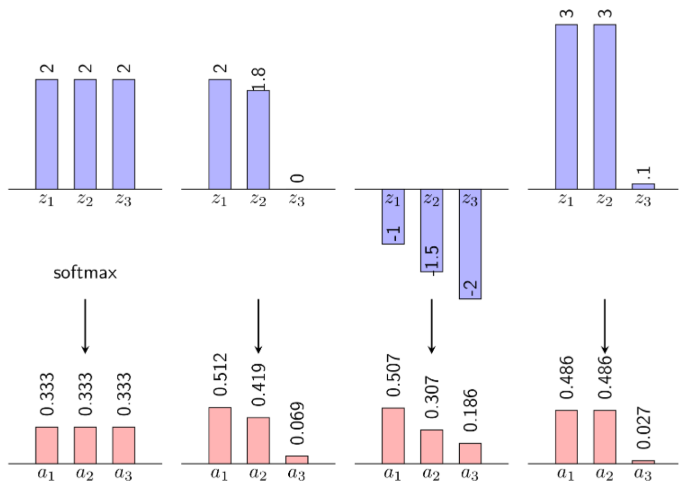


(a) Standard Neural Net (b) After applying dropout.

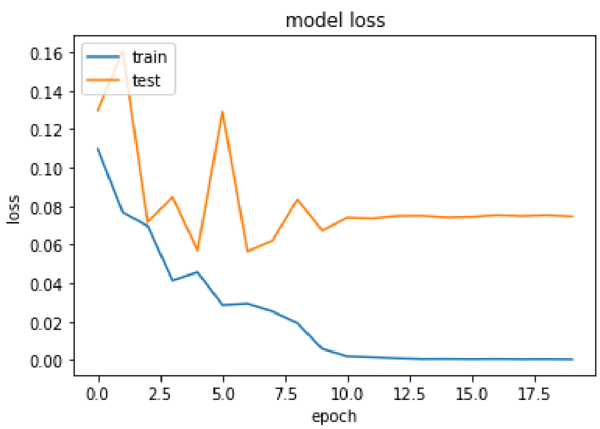
Then add a dense layer with activation='relu' with output = 1024. ReLU is proven to make training Deep Networks much faster (according to Krizhevsky et al.). This acceleration is attributed because the ReLU is calculated almost instantaneously and its gradient is also calculated extremely fast with a gradient of 1 if the input is greater than 0, zero if the input is less than zero.

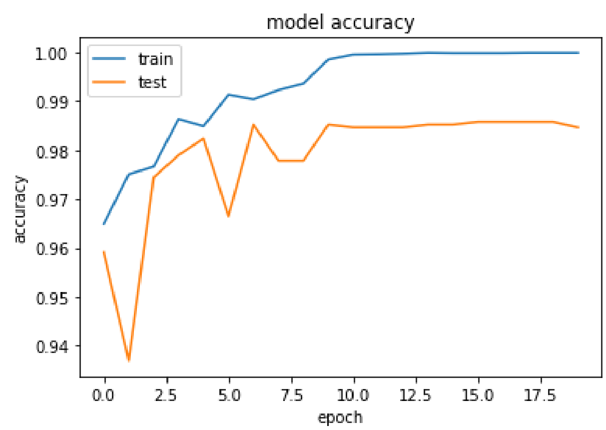


Finally add a class trait with enable='softmax'. It converts the results to a profit distribution and compresses the data in the range 0-1, so it's good for classification.



After training on the mini-data set with 4632 images and got the result accuracy = 0.972 when evaluating on the test set(10% dataset). Then I found that the model was suitable for the problem, so I decided to apply it to the full data set of 19554 images. Then the result is accuracy = 0.9867 when evaluating on the test set (10% dataset). And when plotting the training history of the model, it is found that the model is in good fit form.





4. Result & Application

AI can have many real-life applications, including in healthcare and industry. Bringing AI into healthcare can help reduce pressure on the medical industry as well as improve the quality of medical care in places with poor human resources. In the not too distant future, there may be automated medical stations operated 100% by AI without human intervention and supervision. And can bring them to poor developing countries so that more people have access to modern health care, improving people's quality of life.

The application of AI in industry can help reduce production costs while increasing product quality, thereby promoting the development of the industry.