深度學習於醫學影像分析—**Deep learning in medical image analysis**
Spring 2021

Homework 1, due on 2021/4/12

1. In this homework, you are asked to analyze a dataset containing chest x-ray images with disease labels, provided by the E-Da Hospital, Taiwan. The goal is to predict subjects' disease statuses by using their chest x-ray images.

    The majority of the subjects in the dataset are disease-free while the rests might contain multiple sites of abnormalities（主動脈硬化(鈣化)、動脈彎曲、肺野異常、肺紋增加、脊椎病變、肺間肋膜增厚、心臟肥大）in one chest x-ray image. This is a typical **multi-label classification** problem.

2. Use the dataset "**hwk01_data.rar**" to build up your prediction models and test models' accuracy on the dataset "**hwk01_test.rar**". These two datasets can be downloaded from e3 (https://e3.nycu.edu.tw/) under "Homework 1".

    The file "**data.xlsx**" in "**hwk01_data.rar**" contains the information regarding subjects' disease labels and images of training data. Please read the file "**README.pdf**" first for the meaning of each data item in "**data.xlsx**".

    The file "**test.xlsx**" in "**hwk01_test.rar**" contains the information regarding testing images and their possible disease categories.

3. To build the prediction model for multi-label classification,

    a. Use convolutional neural networks (CNNs) to extract feature representations and identify possible diseases in these images. Utilize transfer learning to adopt the pre-trained model whose weights are trained from **ImageNet** to improve prediction accuracy.

    b. Perform the analysis under two CNN architectures: VGG16 and ResNet50. Output the prediction results for the test dataset based on each of the two architectures.

4. You may refer to the following links:

https://keras-cn.readthedocs.io/en/latest/other/application/

https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/

for the examples of using Python to perform the analyses mentioned in 3.

5. Load your created prediction models with the images in the dataset "**hwk01_test.rar**" and generate their predicted probabilities for all 7 disease categories (主動脈硬化(鈣化)、動脈彎曲、肺野異常、肺紋增加、脊椎病變、肺間肋膜增厚、心臟肥大).

    Use these predicted probabilities to fill in the file "**test.xlsx**" in "**hwk01_test.rar**". Upload two files "**VGG16.csv**" and "**ResNet50.csv**" for prediction results based on VGG16 and ResNet50 architectures, respectively.