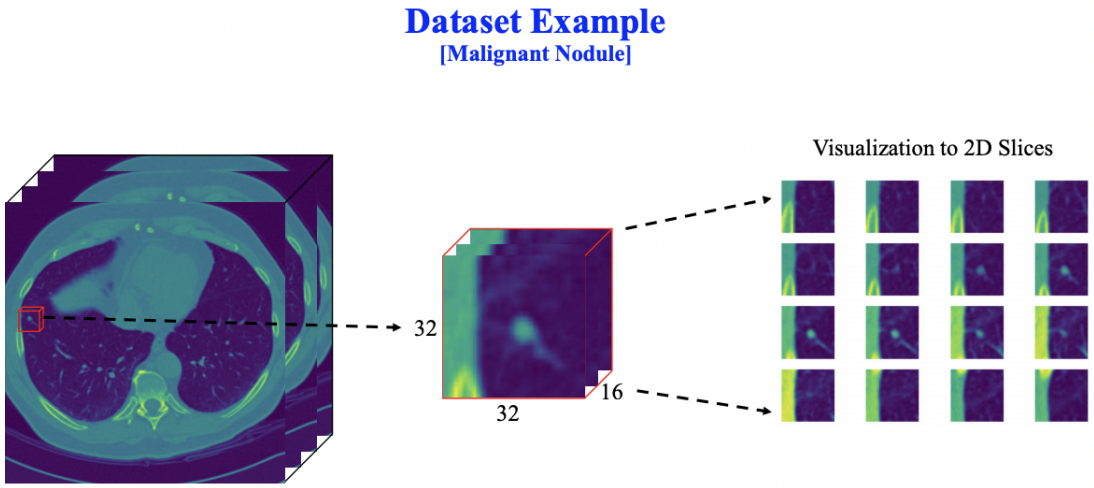
Task Description

- Objective: Given a lung nodule CT image, classify benign vs. malignant
- Dataset: CT image patches from LUNA16 Challenge
    - https://luna16.grand-challenge.org/
    - Input: 3,000 3D patches with a lung nodule in the center point
        - Size: 32cm × 32cm × 16cm, Nodule Diameter: 8.31 ± 4.76cm
        - 3,000 patches from 888 subjects • 1,500 benign + 1,500 malignant patches
        - Split into 800 train subjects, 88 test subjects
            - Train: 1,400 benign + 1,400 malignant patches, Test: 100 benign +100 malignant patches
            - Test data will not be available to students

Label (Boolean): Benign (False) vs. Malignant (True)



No preprocessing is performed, except for resampling voxels to $1cccc$ × $1cccc$ × $1cccc$ spacing. Perform your own preprocessing.

https://koreaoffice-my.sharepoint.com/:f:/g/personal/wltjr1007_korea_edu/EksffXy1lbxHqWK8yiI3sZ8B4uZeFP7nTXiZ_1yo9VnCKQ?e=GpJBsj

How to load the training data:

```
import numpy as np
dat = np.load("/content/trn_dat.npy") #Should contain (2800,32,32,16) float32 numpy array
lbl = np.load("/content/trn_lbl.npy") #Should contain (2800,) boolean numpy array
```

https://koreaoffice-my.sharepoint.com/personal/wltjr1007_korea_edu/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fwltjr1007%5Fkorea%5Fedu%2FDocuments%2FClass%2F2020%2D02%2F%EB%87%8C%EB%B0%8F%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D%EC%9E%85%EB%AC%B8%2Fassignment2%2Fstudent%20data&originalPath=aH

R0cHM6Ly9rb3JlYW9mZmljZS1teS5zaGFyZXBvaW50LmNvbS86ZjovZy9wZXJzb25hbC93bHRqcjEwMDdfa29yZWFfZWR1L0Vrc2ZmWHkxbGJ4SHFXSzh5aUkzc1o4QjR1WmVGUDduVFhpWl8xeW85Vm5DS1E_cnRpbWU9THFZQkNIaVEyRWc

Project Requirement

- Build the best model based on given train dataset “train_data.npy”
- Report scores for train, validation split
  - Randomly split the dataset into train (90%), validation (10%) split
- Analyze and discuss your models and results in markdown cells

Implementation Requirement

- Use the Google Colab (https://colab.research.google.com/)
- You may use any library of your choice (e.g. Scikit-learn, Tensorflow, PyTorch, …)
  - However, fully automated libraries are not allowed
- **(IMPORTANT) Submit two .ipynb files and model weight files**