Classification prediction of non-small cell lung cancer
Baobao Geng
Marquette University

Introduction
Nowadays, lung cancer is a leading cause of death in both men and women worldwide.[1] Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC) are the most prevalent subtypes of lung cancer, and their distinction requires visual inspection by an experienced pathologist. [2] Nowadays, more and more researchers use machine learning to solve this question, such as convolutional neural network. In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imagery.[3]

Dataset used
We obtain whole-slide images of LUAD and LUSC from the Genomic Data Commons (GDC) portal which were generated by The Cancer Genome Atlas (TCGA), and they have made them publicly available. [4] The GDC is a research program of the National Cancer Institute (NCI). The mission of the GDC is to provide the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine. [5]

Figure 1a shows an example of slide images and Figure 1b is the basic information. For later analysis, we classify the dataset into three groups: normal tissue, LUAD and LUSC. The number of corresponding cases is respectively 528, 585 and 504. And the slides with a low information will be removed.
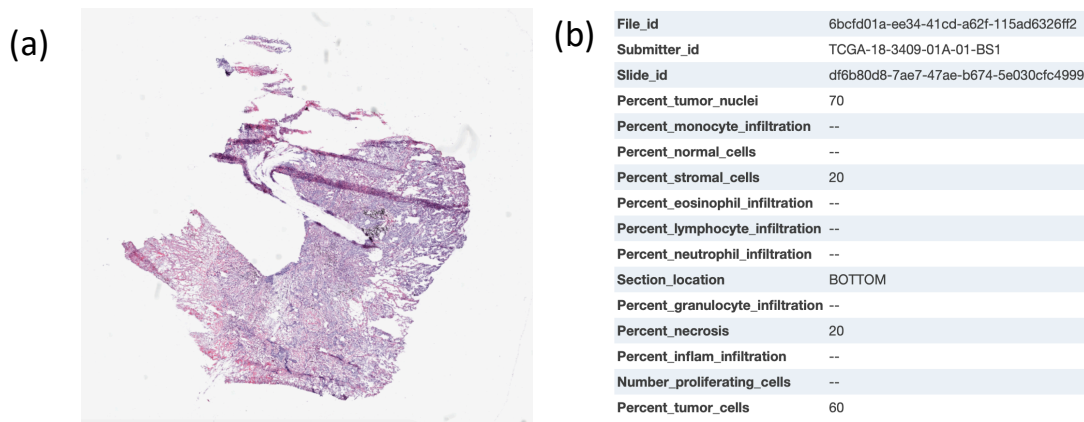

(a)

(b)

File_id	6bcfd01a-ee34-41cd-a62f-115ad6326ff2
Submitter_id	TCGA-18-3409-01A-01-BS1
Slide_id	df6b80d8-7ae7-47ae-b674-5e030cfc4999
Percent_tumor_nuclei	70
Percent_monocyte_infiltration	--
Percent_normal_cells	--
Percent_stromal_cells	20
Percent_eosinophil_infiltration	--
Percent_lymphocyte_infiltration	--
Percent_neutrophil_infiltration	--
Section_location	BOTTOM
Percent_granulocyte_infiltration	--
Percent_necrosis	20
Percent_inflam_infiltration	--
Number_proliferating_cells	--
Percent_tumor_cells	60

Figure 1. Dataset details. (a) Eosin stained histopathology images. (b) The details of the left picture.

Goal & Question
Use convolutional neural network or other deep-learning models to assist pathologists in the detection of lung cancer subtype. And, how to optimize the model to make the prediction more accurate?

References

[1] Nicolas C , Santiago O P , Theodore S , et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning[J]. Nature Medicine, 2018.

[2] Romaszko A M , Anna Doboszyńska. Multiple primary lung cancer: A literature review[J]. Advances in Clinical and Experimental Medicine, 2018, 27(5).

[3] Valueva M V , Nagornov N N , Lyakhov P A , et al. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation[J]. Mathematics and Computers in Simulation (MATCOM), 2020, 177.

[4] Wang Z , Jensen M A , Zenklusen J C . A Practical Guide to The Cancer Genome Atlas (TCGA).[M]// Statistical Genomics. Springer New York, 2016.

[5] Grossman R L , Heath A P , Ferretti V , et al. Toward a Shared Vision for Cancer Genomic Data[J]. New England Journal of Medicine, 2016, 375(12):1109.