

# Lung Cancer Segmentation

Deep Neural Networks Final Project

---

Salem AlAthari, Zehui Chen, Kat Desai, Manjit Singh

April 25, 2024

# Research Question

Can we use deep learning to identify lung cancer in CT scan images?

# The Dataset

- Fusion of Kazakh Research Institute of Oncology & Radiology and LIDC-IDRI dataset
- 972 CT images labeled by radiologists using the Lung-RADS System
- Each image is labeled with a mask overlaying the cancer

*Nam, Diana; Panina, Alexandra; Pak, Alexandr (2024),  
“Lung cancer segmentation dataset with Lung-RADS class”,  
Mendeley Data, V1, doi: 10.17632/5rr22hgzw.1*

## Files



lung\_cancer\_test.pkl

1.1 GB 



lung\_cancer\_train.pkl

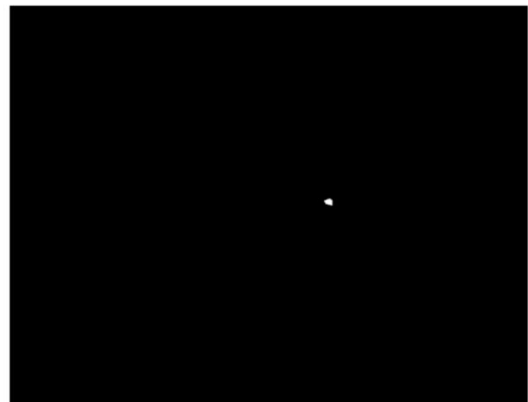
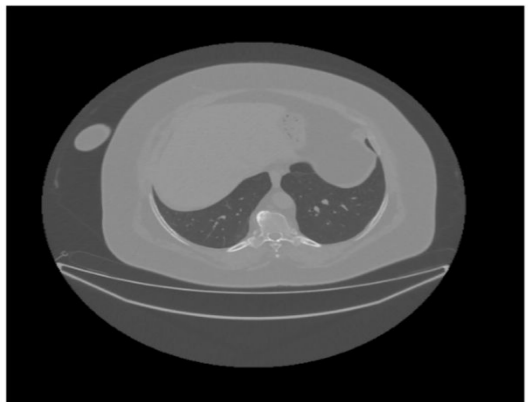
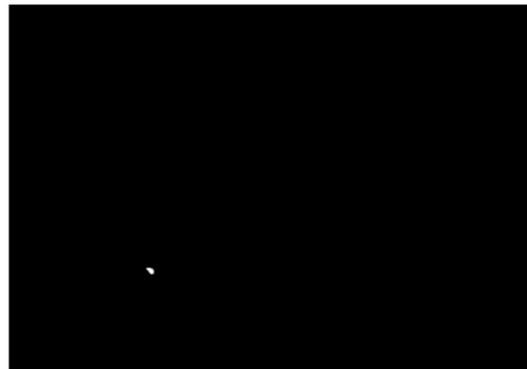
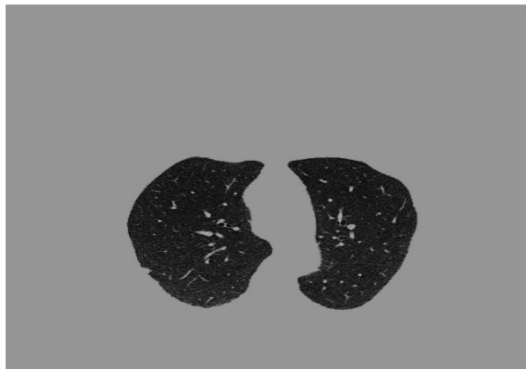
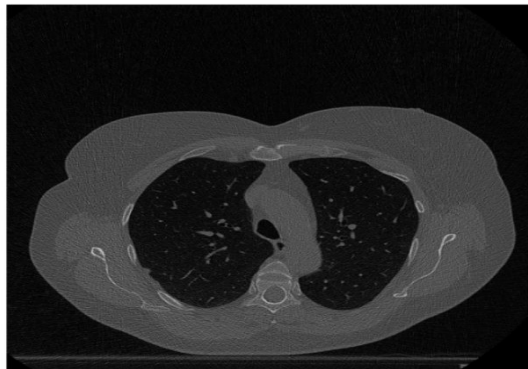
4.15 GB 

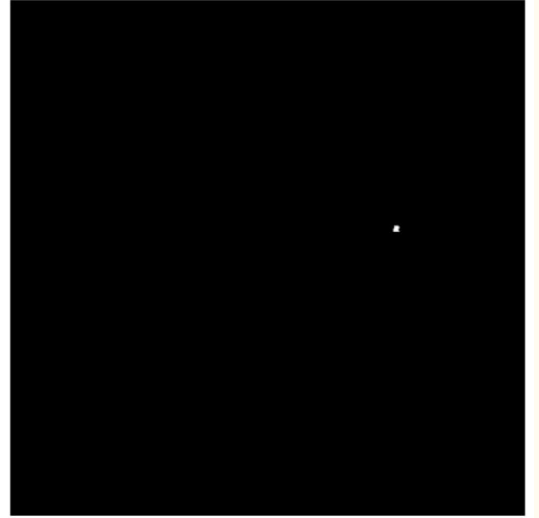
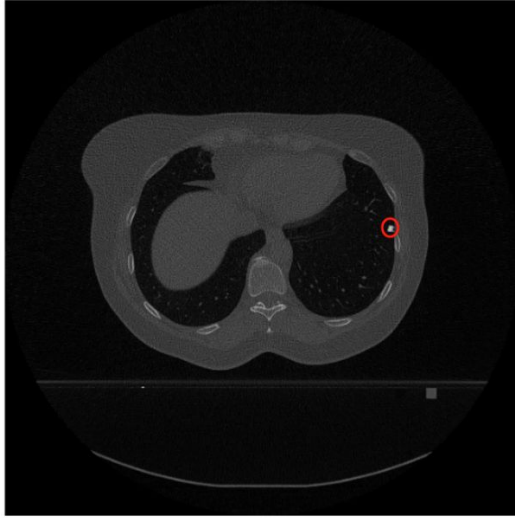
# Dataset

	label1	mask	hu_array	hu_array_old
0	LR2	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -1024.0, -1024.0, -1024.0,...
1	LR2	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -1024.0, -1024.0, -1024.0,...
2	LR2	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -1024.0, -1024.0, -1024.0,...
3	LR2	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -1024.0, -1024.0, -1024.0,...
4	LR2	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -1024.0, -1024.0, -1024.0,...
...	...	...	...	...
703	LR4B	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1015.0, -1024.0, -972.0, -975.0, -1013.0, -...
704	LR4B	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1022.0, -1024.0, -957.0, -987.0, -1022.0, -...
705	LR4B	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1024.0, -951.0, -998.0, -1019.0, -...
706	LR4B	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1009.0, -968.0, -1022.0, -1007.0, ...
707	LR4B	[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[[[-0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0, -0.0...	[[[-1024.0, -1005.0, -957.0, -1017.0, -1013.0, ...

708 rows  $\times$  4 columns

# Dataset

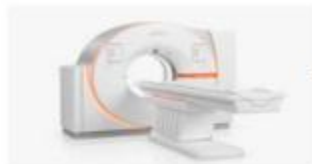




# Literature Review - Key Points

- Most research in the current field is done using pre-trained models and using transfer learning to get it to work with another dataset.
- Models used for segmentation stem from the UNET architecture (E.G. MobileNetV2).
- More modern research focuses on segmenting multiple CT scans for a 3D segmentation.

### Segmentation part



CT scan machine



Data collection



Segmentation  
network



Segmentation  
output

### Classification part



Segmented  
images



Classification  
network



Final prediction



benign



Malignant



# Literature Review - Citations

- Riaz, Z., Khan, B., Abdullah, S., Khan, S., & Islam, M. S. (2023). Lung Tumor Image Segmentation from Computer Tomography Images Using MobileNetV2 and Transfer Learning. Bioengineering (Basel, Switzerland), 10(8), 981.  
<https://doi.org/10.3390/bioengineering10080981>
- Said, Y., Alsheikhy, A. A., Shawly, T., & Lahza, H. (2023). Medical Images Segmentation for Lung Cancer Diagnosis Based on Deep Learning Architectures. Diagnostics (Basel, Switzerland), 13(3), 546.  
<https://doi.org/10.3390/diagnostics13030546>
- Primakov, S.P., Ibrahim, A., van Timmeren, J.E. et al. Automated detection and segmentation of non-small cell lung cancer computed tomography images. Nat Commun 13, 3423 (2022).  
<https://doi.org/10.1038/s41467-022-30841-3>

# Tentative Development Schedule

**Apr 11-18.....**Dataset sourcing and literature review

**Apr 18-25.....**Finetune a pretrained model with this dataset

**Apr 25-May 2.....**Develop CNN baseline model from scratch

**May 2-9.....**Performance enhancement

**May 9-16.....**Presentation & Paper