Image Processing based Lung Cancer Detection from a Low Dose CT Scan Image

Machine Perception
DS/NC/ESD 863/2016-T2-DNE863

	Group 33	
Team 1	Team 2	Team 2
Amit Kumar (MT2016009)	Shreekantha A N (MS2016010)	Vaishnav Ram Savarni K R (MS2016006)

ज्ञानमुत्तमम्

Overview

- Motivation
- Problem Statement
- Dataset Description
- Our Approach
- Segmentation
- Blob Analysis
- Feature Extraction
- Classifier design
- Future Scope
- Conclusion

Motivation

Number of Cancer cases by type in India (2012)

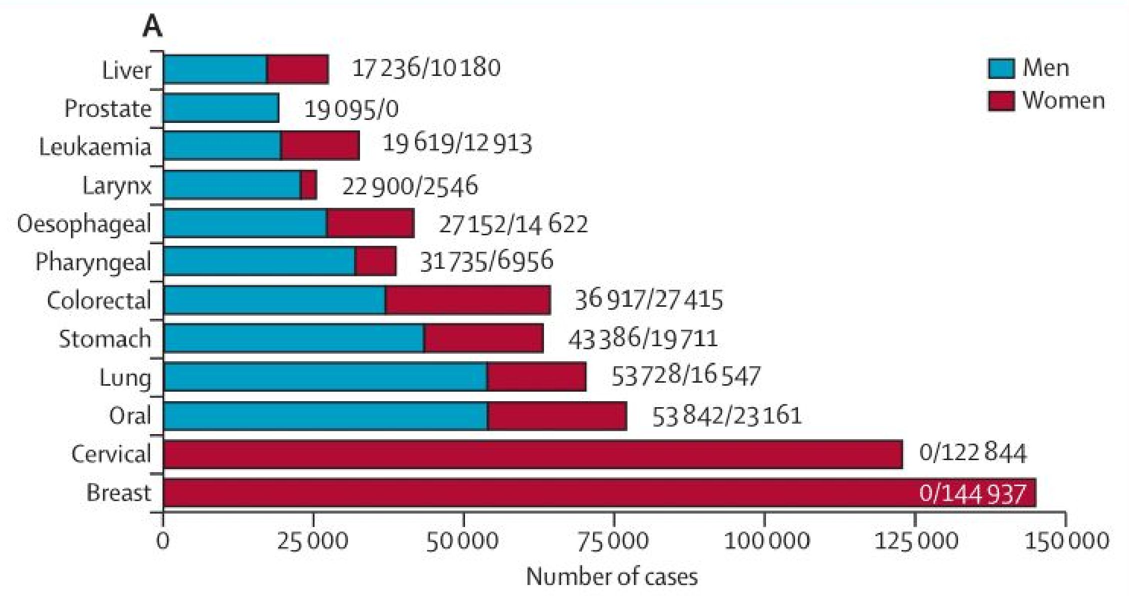


Image courtesy: Mallath MK, Taylor DG, Badwe RA, et al. The growing burden of cancer in India: epidemiology and social context. Lancet Oncol. 2014;15(6):e205-12.

Motivation

Number of deaths due to Cancer in India (2012)

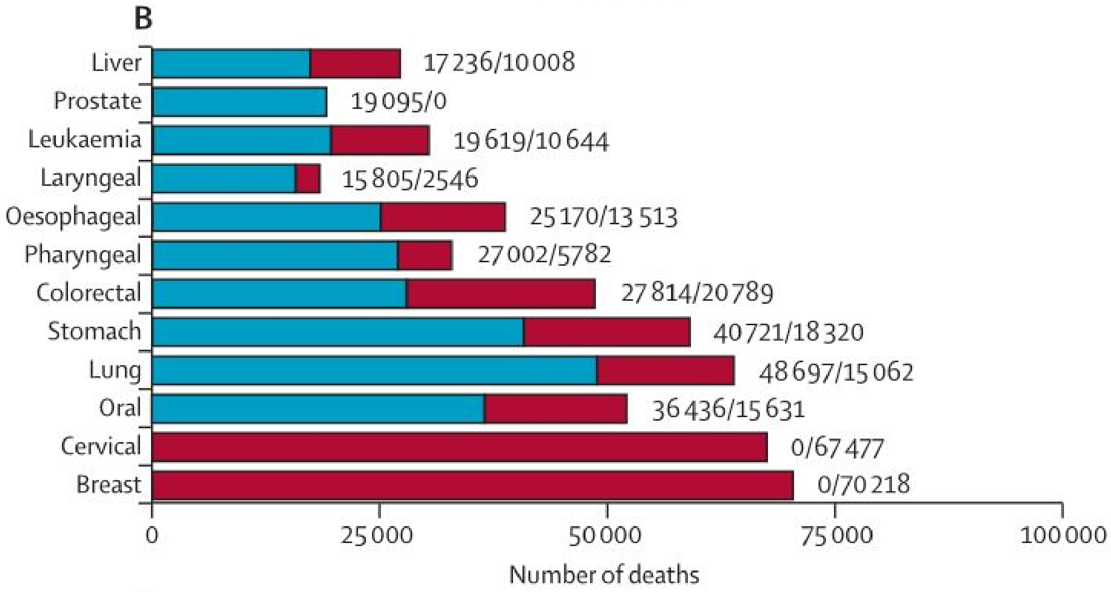


Image courtesy: Mallath MK, Taylor DG, Badwe RA, et al. The growing burden of cancer in India: epidemiology and social context. Lancet Oncol. 2014;15(6):e205-12.

Motivation

Most common cancers in Indian Men and Women

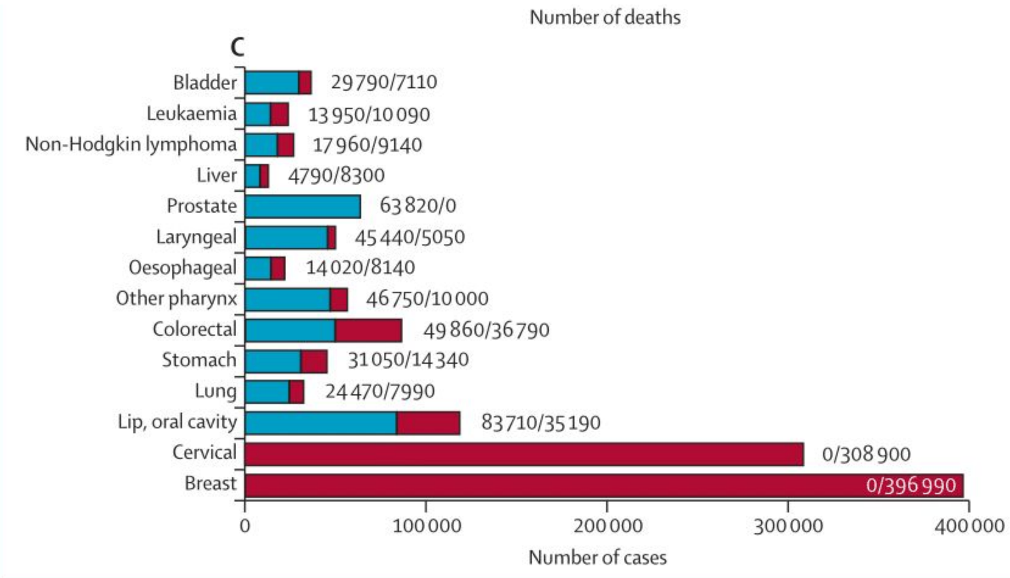


Image courtesy: Mallath MK, Taylor DG, Badwe RA, et al. The growing burden of cancer in India: epidemiology and social context. Lancet Oncol. 2014;15(6):e205-12.

Motivation

- One year ago, the office of the U.S. Vice President spearheaded a bold new initiative, the Cancer Moonshot, to make a decade's worth of progress in cancer prevention, diagnosis, and treatment in just 5 years.
- Data Science Bowl will be a critical milestone in support of the Cancer Moonshot by convening the data science and medical communities to develop lung cancer detection algorithms.
- Using a data set of thousands of high-resolution lung scans provided by the National Cancer Institute, participants will develop algorithms that accurately determine when lesions in the lungs are cancerous.
- Kaggle Dataset contains 2121 patient data with each patient containing approximately 200 images of CT scans.
- The dataset is quite extensive and is around 100GB in size.
- The problem and the dataset is quite promising to achieve meaningful result and contribute something to the community.

Problem Statement

- Low-dose computed tomography (CT) scans can reduce lung cancer deaths by 20 percent[1]
- Challenge of reducing false positive rate
- Reducing patient anxiety, costly and unnecessary diagnostic work
- Develop algorithms that accurately determine when lesions in the lungs are cancerous.
- Get patients earlier access to life-saving interventions, and give radiologists more time to spend with their patients.

[1] Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395-409.

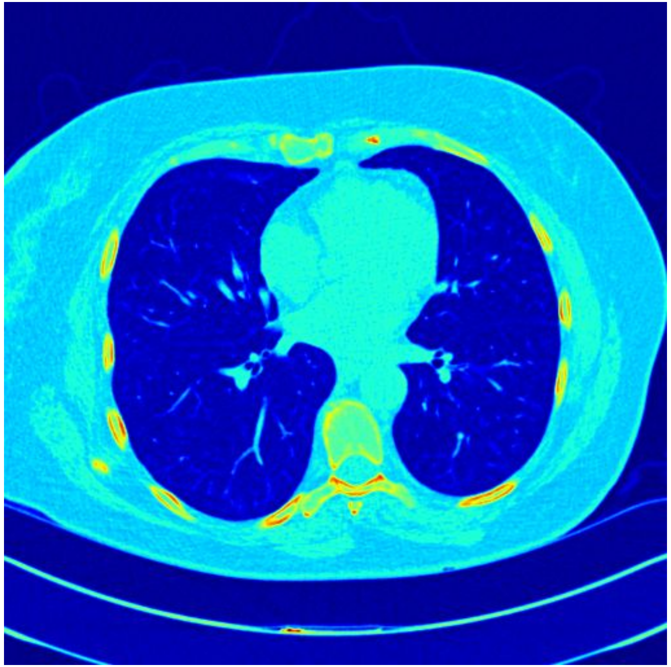
Dataset description

Kaggle Dataset:

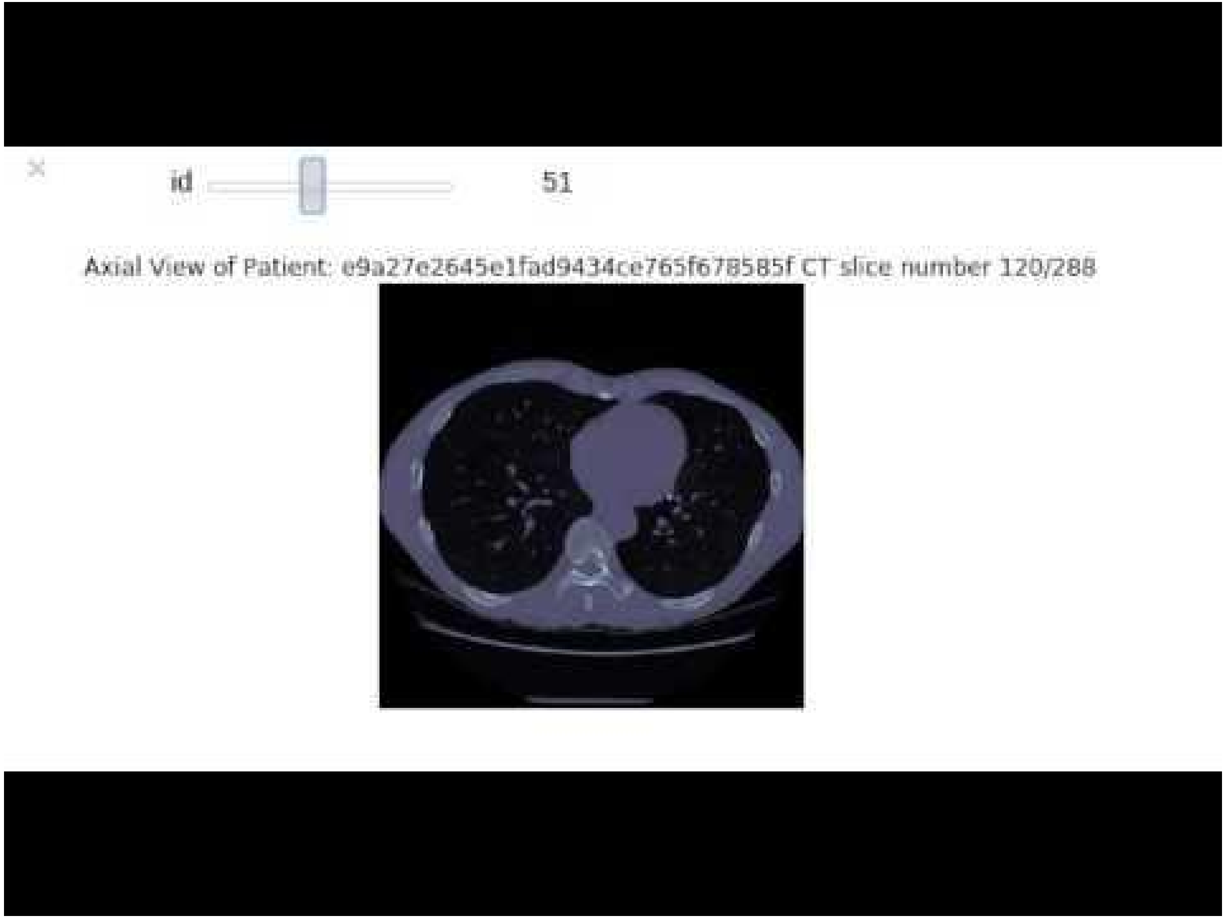
- Over a thousand low-dose CT images from high-risk patients in DICOM format
- Each image contains a series with multiple axial slices of the chest cavity
- Each image has a variable number of 2D slices, which can vary based on the machine taking the scan and patient.
- The DICOM files have a header that contains the necessary information about the patient id, as well as scan parameters such as the slice thickness.
- The images in this dataset come from many sources and will vary in quality

Dataset description

Kaggle Dataset:

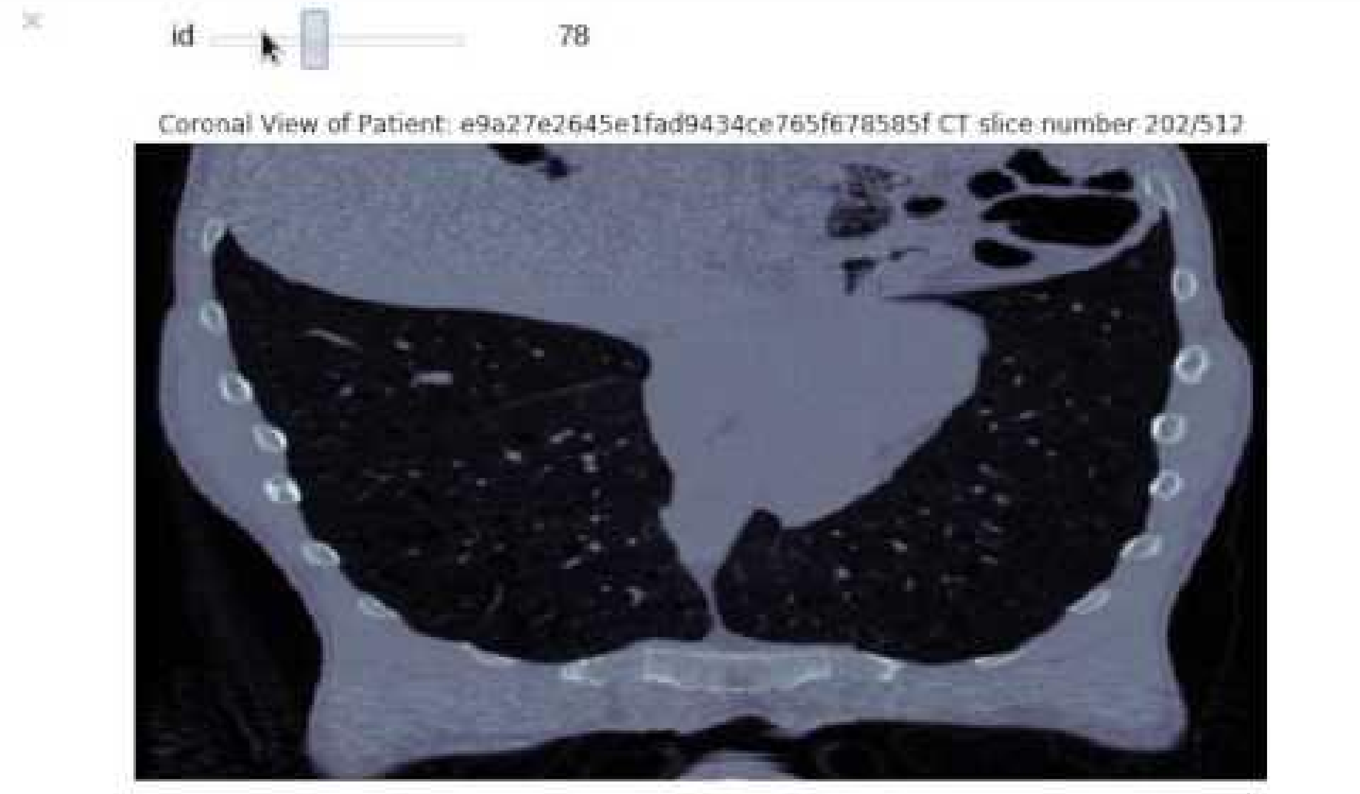


Axial view of CT scans

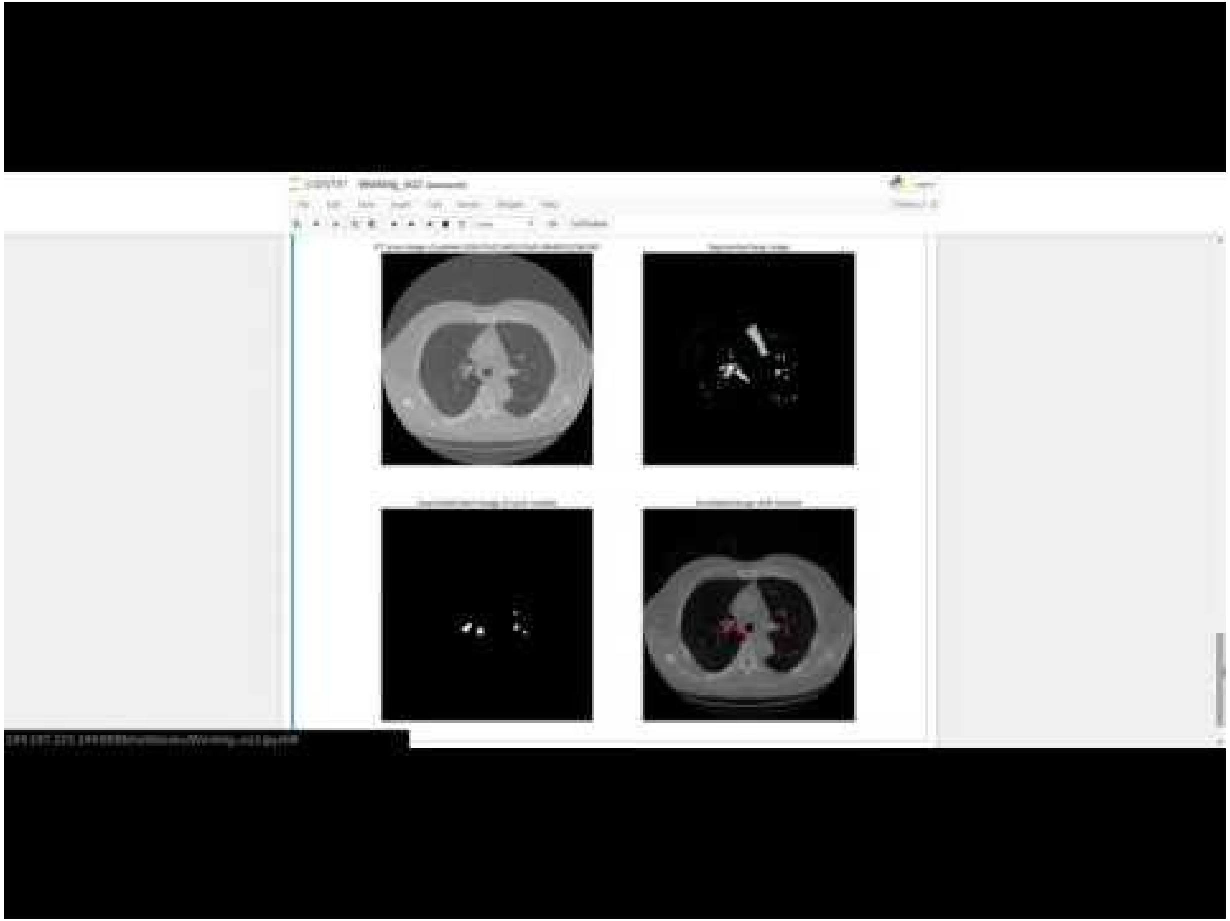


Sagittal View of CT scans

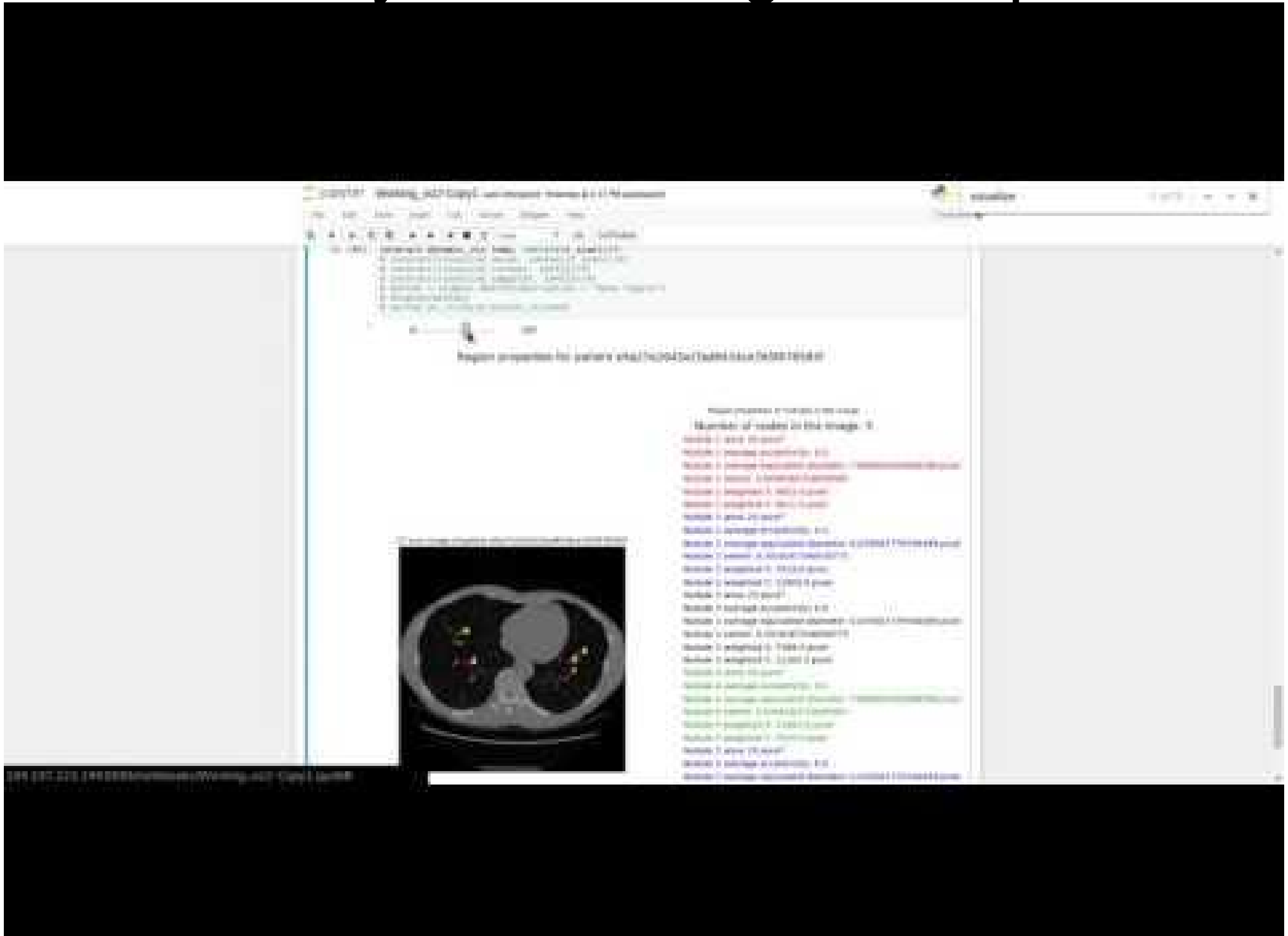
Coronal View of CT scans



Segmentation process



Blob Analysis and Region Properties



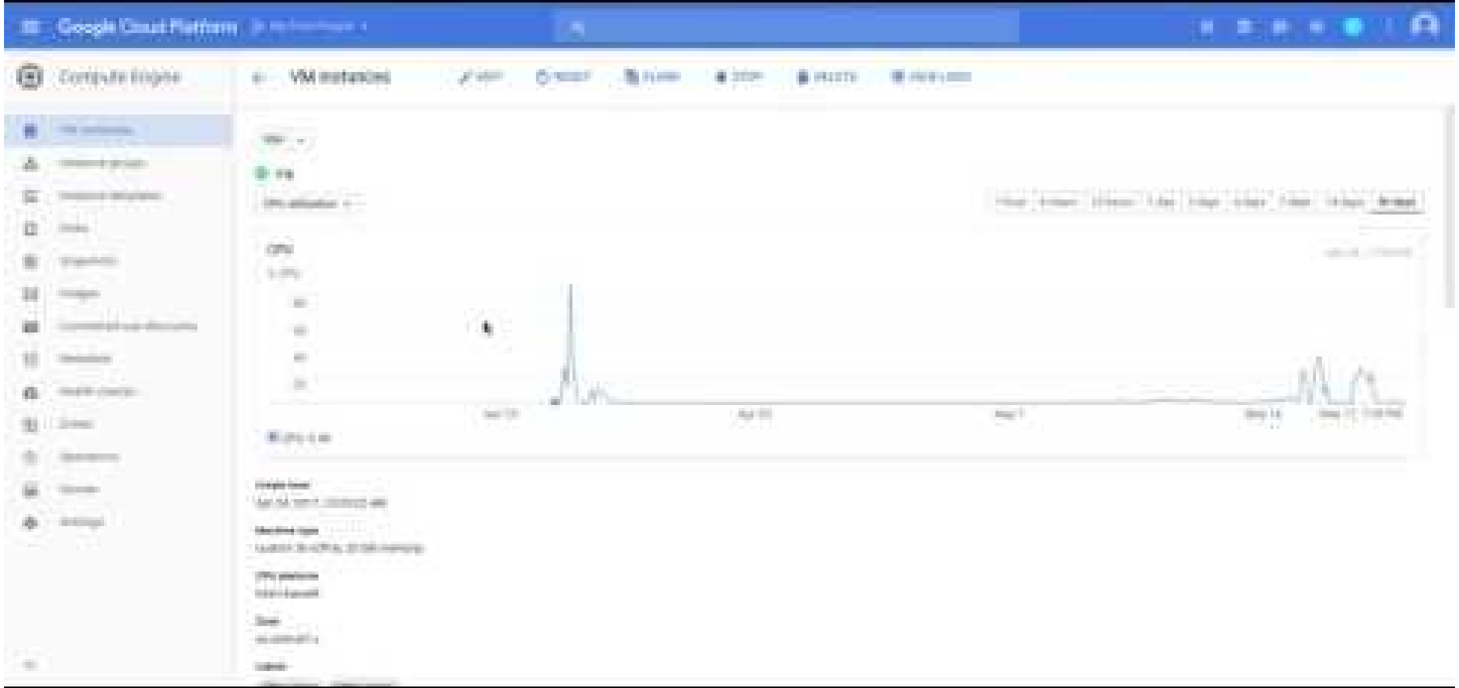
Feature extraction

```
[avg_area,max_area, avg_eccentricity,
avg_equivalent_diameter,
std_equivalent_diameter, total_extent,
weightedX, weightedY, num_nodes,
num_nodes_per_slice]
```

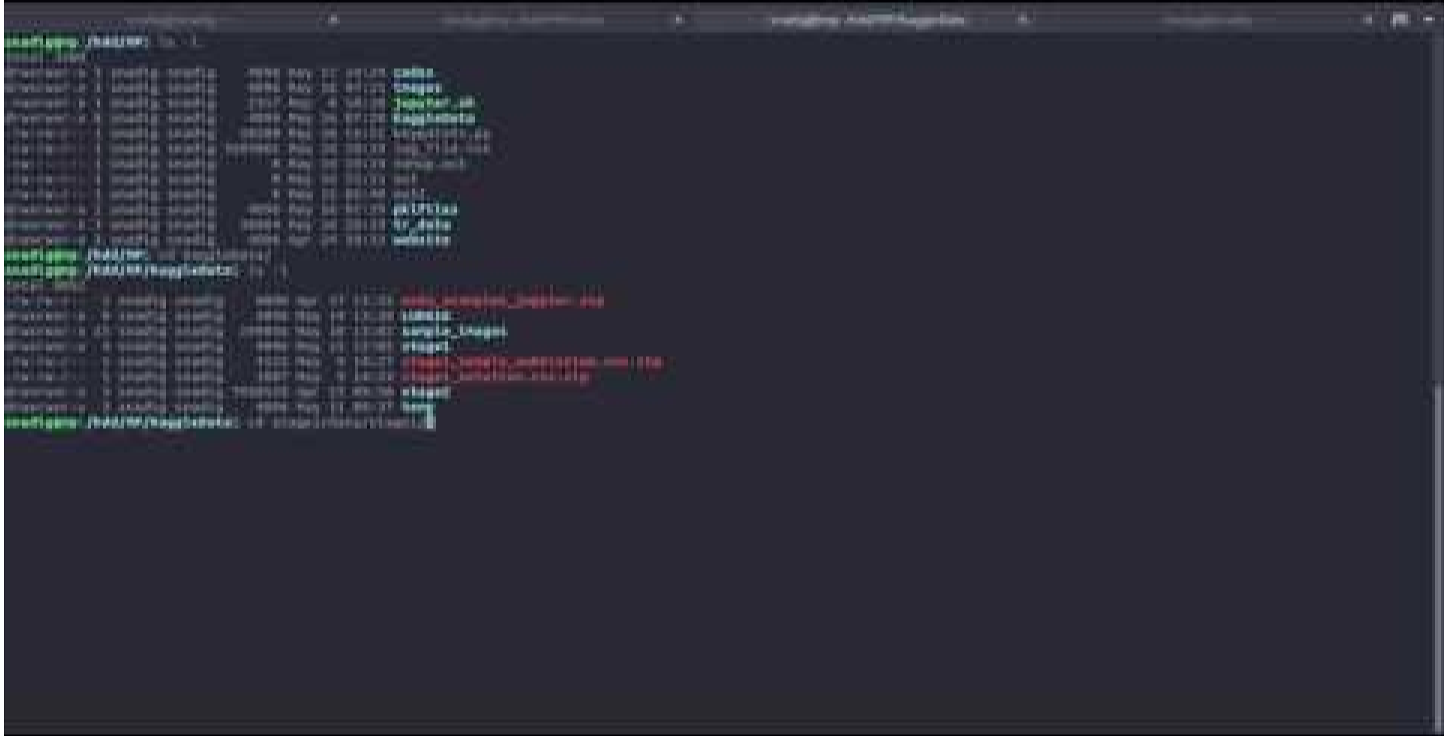
Feature extraction

```
    regions = regionprops(labeled_image, image_nob, cache=True)
    for rp in regions:
        total_area += rp.area
        areas.append(rp.area)
        avg_eccentricity += rp.eccentricity
        avg_equivalent_diameter += rp.equivalent_diameter
        eqi_diams.append(rp.equivalent_diameter)
        total_extent += rp.extent
        weightedX += rp.centroid[0]*rp.area
        weightedY += rp.centroid[1]*rp.area
        num_nodes += 1
weightedX = weightedX / total_area
weightedY = weightedY / total_area
avg_area = total_area / num_nodes
avg_eccentricity = avg_eccentricity / num_nodes
avg_equivalent_diameter = avg_equivalent_diameter / num_nodes
std_equivalent_diameter = np.std(eqi_diams)
max_area = max(areas)
num_nodes_per_slice = num_nodes * 1. / n_slices
return np.array([avg_area, max_area, avg_eccentricity, avg_equivalent_diameter, std_equivalent_diameter, \
                 total_extent, weightedX, weightedY, num_nodes, num_nodes_per_slice])
```

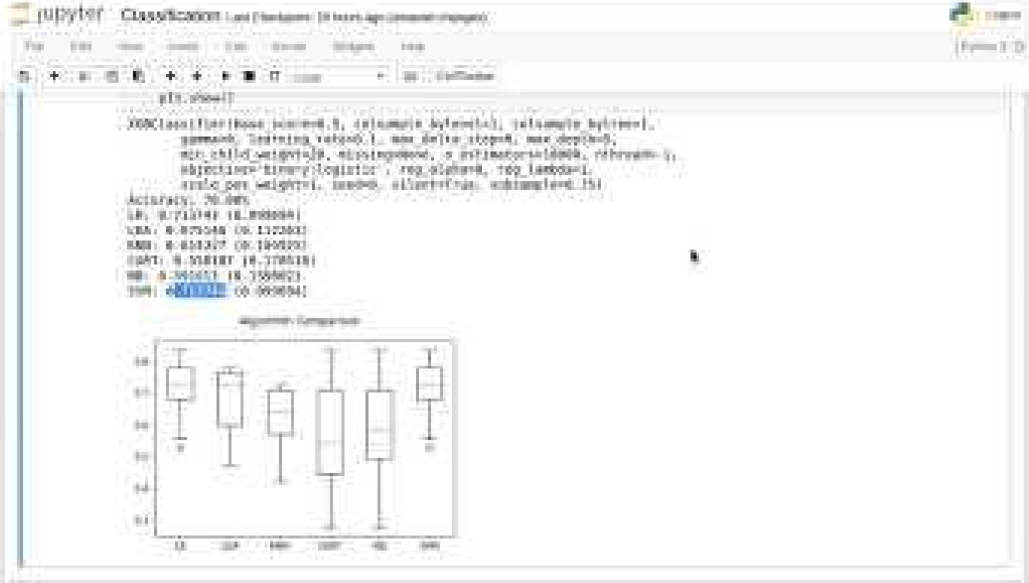
Google Cloud Dashboard



GCP ssh access



Classifier training and results



Future Scope

- Improvement in segmentation process using sophisticated machine learning techniques like CNNs (3D segment analysis using CNNs)
- Better segmentation using U-net
- Classification using 3D Convolutional Neural Networks
- Challenge of reducing false positive rate to a greater extent
- Training with more datasets like LUNA16

Conclusion

- An attempt to detect Lung Cancer lesions using Low Dose CT scan images
- Challenging problem both computationally and intellectually
- Contribution to the community
- Learning cloud tools for large scale Machine Learning tasks
- Testing with different models
- Exposure to Medical Imaging and challenges
- Satisfactory results using classical machine learning models
- Work on improving the accuracy using more sophisticated models in the future

Thank you

Image Processing based Lung Cancer Detection from a Low Dose CT Scan Image

	Group 33	
Team 1	Team 2	Team 2
Amit Kumar (MT2016009)	Shreekantha A N (MS2016010)	Vaishnav Ram Savarni K R (MS2016006)

Machine Perception
DS/NC/ESD 863/2016-T2-DNE863

ज्ञानमुत्तमम्