



# Semantic Image Classification and Segmentation of Cancer MRI Images

**Visual Analytics - ITCS 5122**

Prof : Jianping Fan, PhD

TA: Tinghao Feng

GROUP MEMBERS:

LIKHITH AALLA-801075869

ABISHEK SURYA-800985589

SHARMISTHA HALDAR-801061221

YESWANTH TADIMALLA-801083447

# Index

- **Abstract**
- **Introduction**
- **Dataset**
- **Procedure for classifier training**
  - i. **Data preprocessing**
  - ii. **Feature extraction**
  - iii. **Classifier training**
  - iv. **Visualizing data distribution**
  - v. **User-system interaction**
  - vi. **Image prediction**
- **Image segmentation**
  - i. **Canny Edge detection**
  - ii. **Extracting objects from image**



## *Abstract*

---

The aim of this project is to train the SVM classifier with various cancer images in order to predict an unlabeled cancer image to its respective category.

---

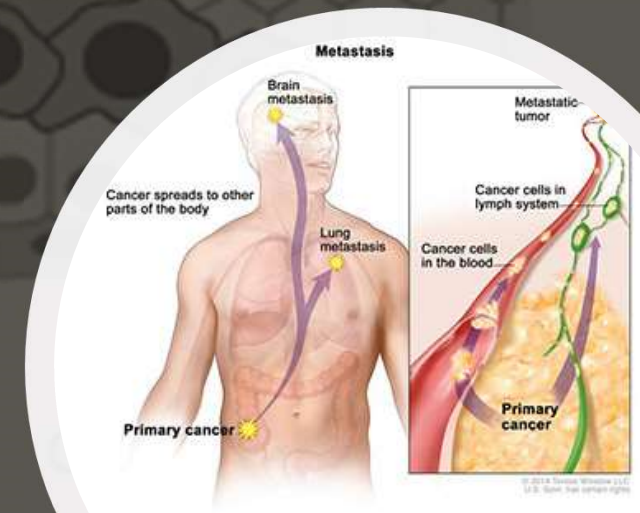
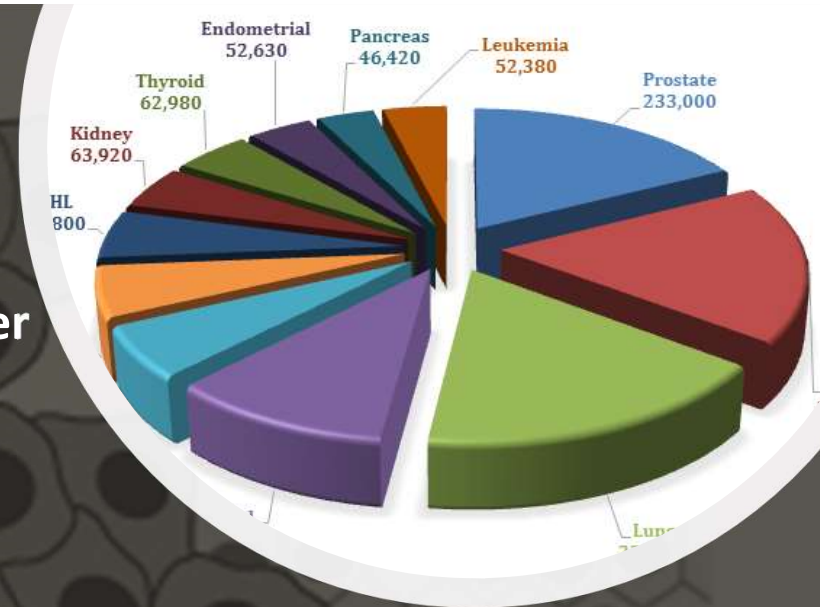
We are using SVM Classifier and skimage library as a base to train our system for Semantic image classification & Visual Analytics

---

The main cancer types being considered here are Glioblastoma Multiforme(Brain), Squamous lung cancer (Lung), Parathyroid cystic (Neck) and Clear cell Kidney carcinoma (kidney).

# Introduction

- Cancer is a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body.
- The main reason that cancer is so serious is its ability to spread to most any part of the body, although different types of cancer are more likely to spread to certain areas than others.
- The most common sites where cancer spreads are the Bone, Brain, Lung, Kidney/Liver, and Neck (Neural).
- Hence for this project we have focused on 4 types of cancers namely : Glioblastoma Multiforme(Brain), Squamous lung cancer (Lung), Parathyroid cystic (Neck) and Clear cell Kidney carcinoma (kidney).

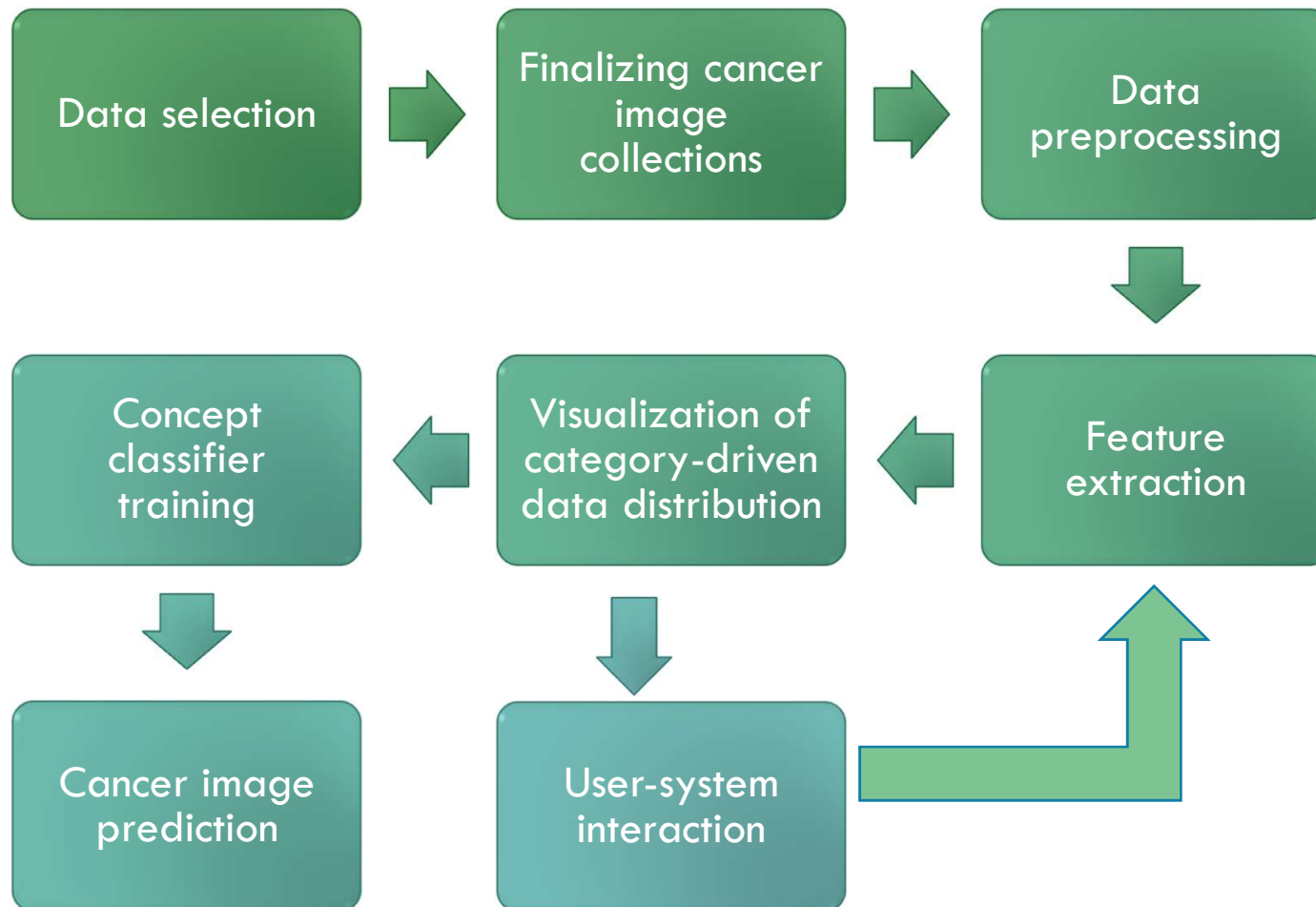




# Data set

- The Cancer Imaging Archive (TCIA)—an open-source, open-access information resource to support research, development, and educational initiatives utilizing advanced medical imaging of cancer.
- TCIA is a service which de-identifies and hosts a large archive of medical images of cancer accessible for public download. The data collection is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive (TCIA).
- <http://www.cancerimagingarchive.net>
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057.

# Procedure



# Feature extraction

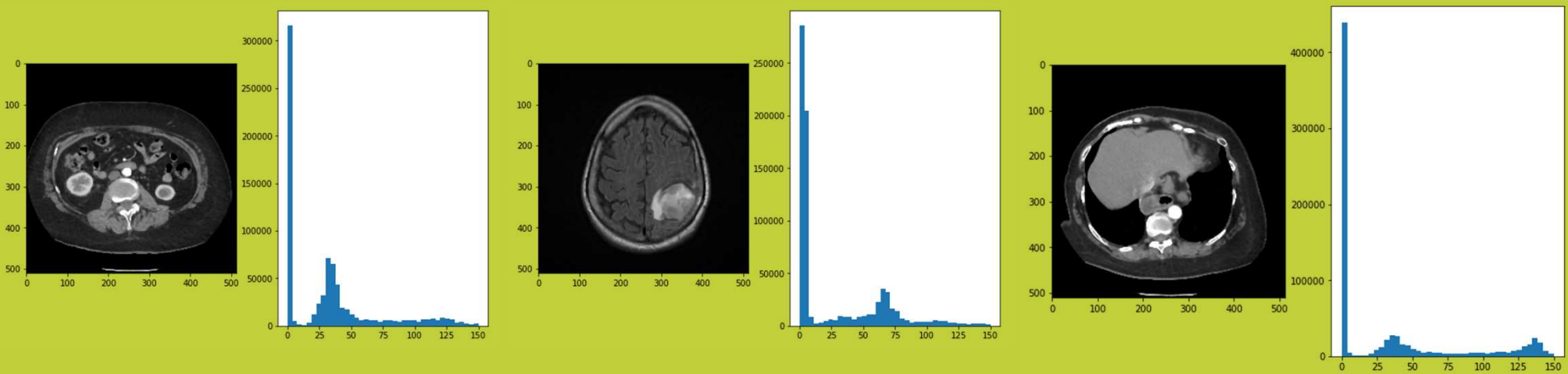
We identified the important features in each image based on which we classified the cancer images.

We extracted following features:

- Area
- Perimeter
- Convex\_area
- Mean\_intensity
- Max\_intensity
- Solidity
- Orientation
- Eccentricity
- Equivalent\_diameter
- Euler\_number
- Extent
- Filled\_area
- Major\_axis
- Minor\_axis

Area	Perimeter	Convex_area	Mean_intensity	Max_intensity	Solidity	Orientation	Eccentricity	Equivalent_diameter	Euler_number	Extent	Filled_area	Major_axis	Minor_axis
248040	2020.818326	250347	0.118741079	1	0.990784791	0.056272512	0.267081988	561.9736085	1	0.978693182	248040	584.1768711	562.955951
245074	2020.374675	247200	0.111001799	1	0.991399676	0.004281526	0.268390617	558.603534	1	0.980860976	245074	581.2785357	559.9515229
244409	2020.303607	246865	0.115093673	0.996078431	0.990051243	-0.014932605	0.252621435	557.8451433	1	0.978199443	244409	579.6385913	560.8381279
245214	2017.569589	247450	0.128791914	1	0.990963831	0.010155963	0.256560595	558.763064	1	0.979414302	245214	580.4363237	561.0080451
244751	2008.818326	246945	0.117733597	0.992156863	0.991115431	-0.043728306	0.25474425	558.2353015	1	0.977565024	244751	579.9158857	560.7835692
246292	2008.770599	248305	0.137100009	1	0.991893035	0.083596799	0.265703118	559.9899231	1	0.983719964	246292	582.7264766	561.7803159
246124	2026.617316	249747	0.118071634	1	0.985493319	0.024471395	0.252858353	559.7989011	1	0.971133207	246124	581.4717859	562.575877
244923	2017.883297	247468	0.125801602	0.992156863	0.989715842	0.009468242	0.251386664	558.4314184	1	0.978252013	244923	580.0837516	561.4553723
248120	2008.048773	249964	0.105714076	1	0.992622938	0.098374487	0.259873282	562.0642275	1	0.979008838	248120	584.0363914	563.970494

Calculating the threshold value for the histogram using threshold\_otsu



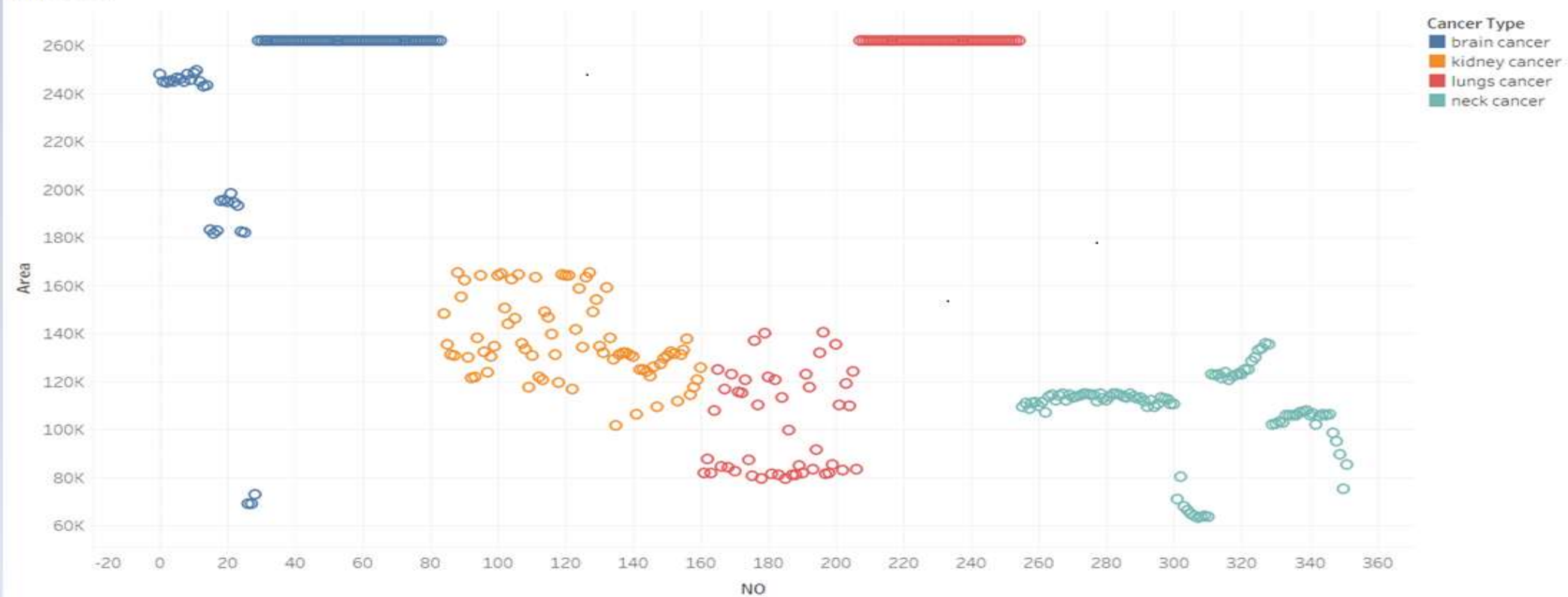


# Classifier training

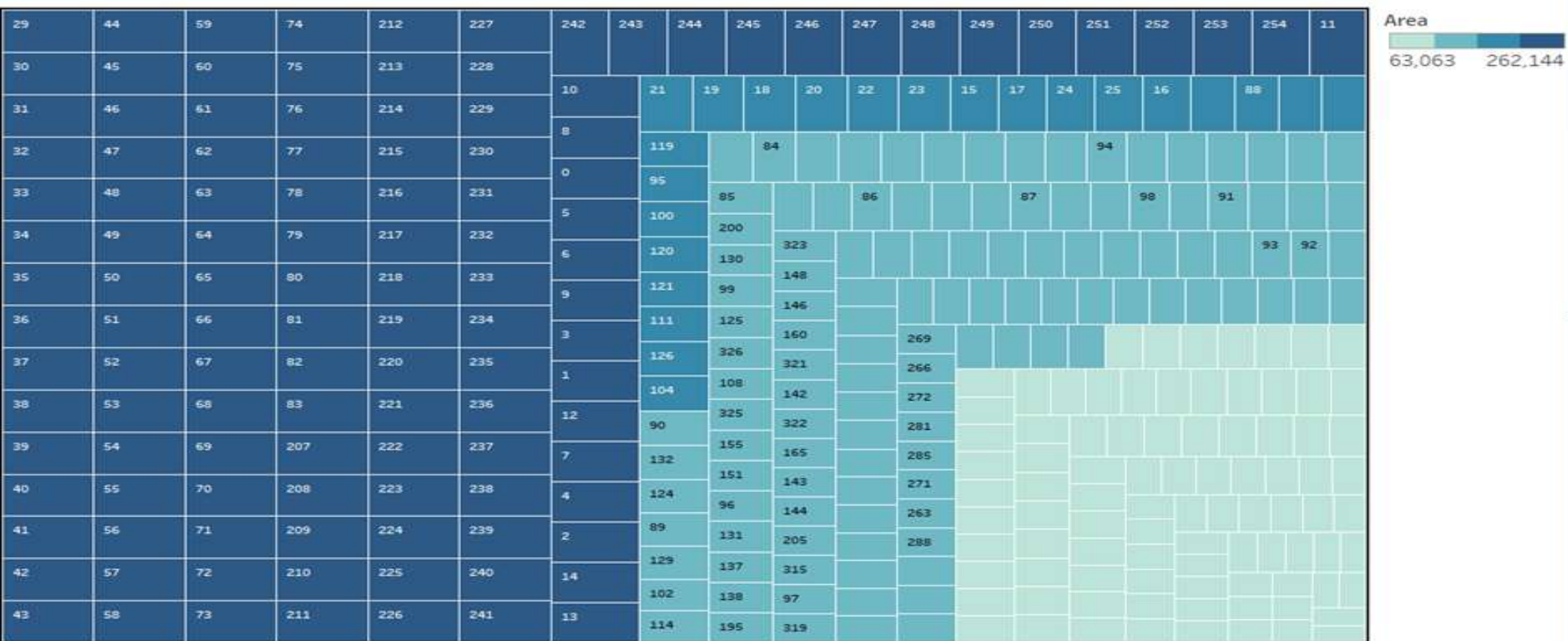
- Image classification uses the quantitative spectral information contained in an image, which is related to the composition or condition of the target surface. There are several core principles of image analysis that pertain specifically to the extraction of information and features from remotely sensed data.
- The algorithm used was SVM, we compared the features of the images to identify the different parts of the body and classify the type of cancer present.
  - I. Importing libraries
  - II. Importing the Dataset
  - III. Exploratory Data Analysis
  - IV. Data Preprocessing
  - V. Training the Algorithm
  - VI. Making Predictions
  - VII. Evaluating the Algorithm
  - VIII. The Evaluation Results

# Visualization of data distribution (Tableau)

Sheet 3



# Area occupied by different types





# Visualization of data distribution (Tableau)



# Cancer image prediction

After training the classifier we have used 30 images to predict the type of cancer. We got the following readings using confusion matrix and classification report.

- Precision of 85%
- Recall of 73%
- F1-score of 73%

```
[[5 0 0 0]  
 [0 4 0 0]  
 [0 0 5 0]  
 [3 0 5 8]]
```

	precision	recall	f1-score	support
brain cancer	0.62	1.00	0.77	5
kidney cancer	1.00	1.00	1.00	4
lungs cancer	0.50	1.00	0.67	5
neck cancer	1.00	0.50	0.67	16
avg / total	0.85	0.73	0.73	30

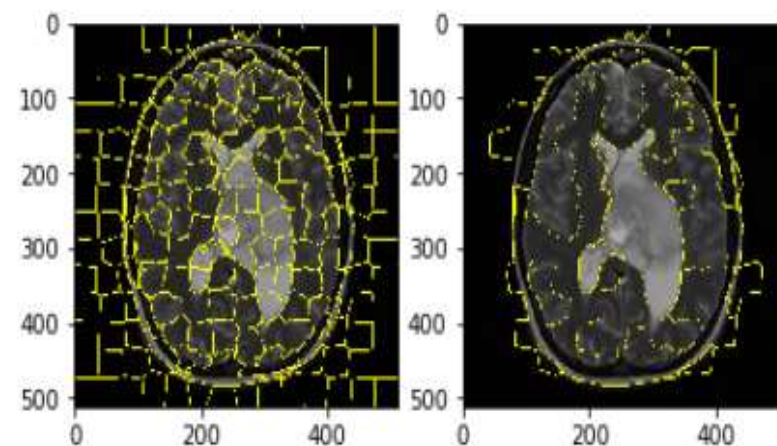
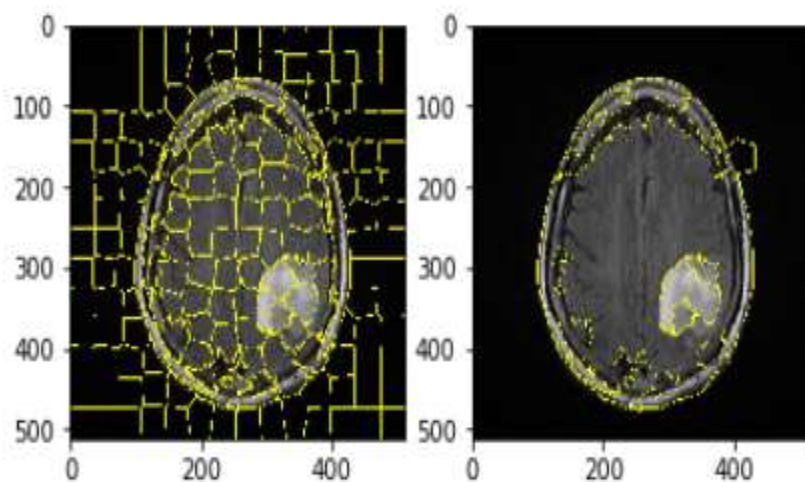


# Image segmentation

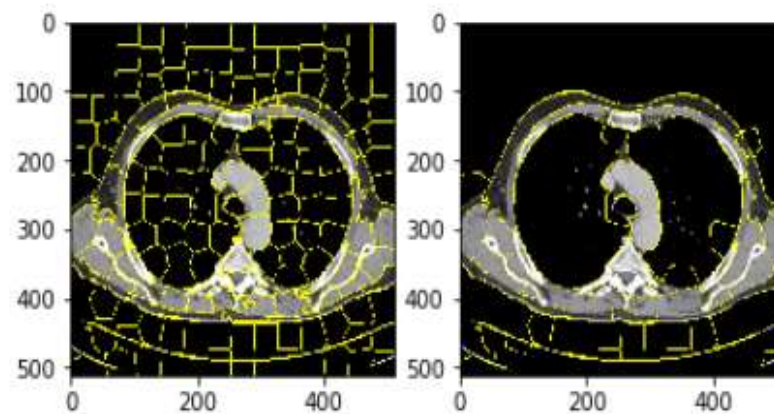
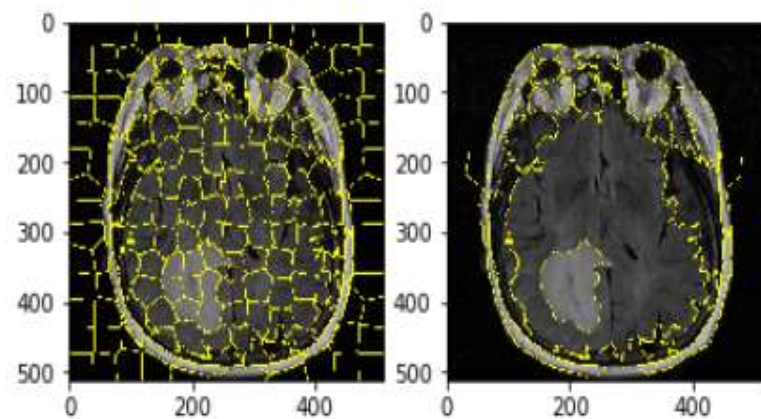
SLIC (Simple Linear Iterative Clustering) is a segmentation algorithm which clusters pixels in both space and color. Therefore, regions of space that are similar in color will end up in the same segment.

SLIC is a superpixel algorithm, which segments an image into patches (superpixels) of neighboring pixels with a similar color. SLIC also works in the Lab colorspace. The compactness parameter controls the relative importance of the distance in image- and color-space.

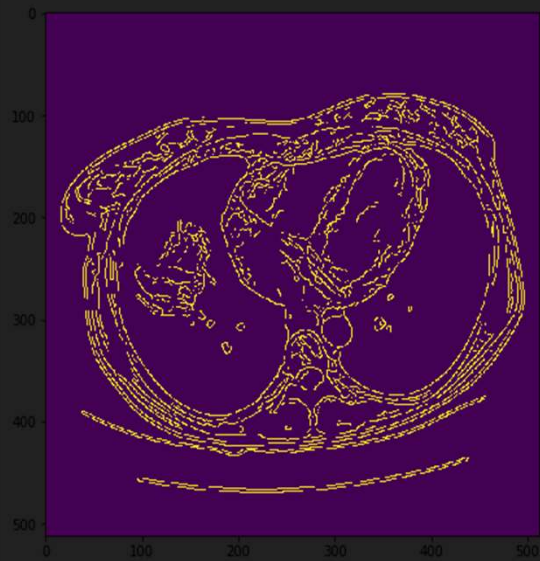
After the super-pixel segmentation (which is also called oversegmentation, because we end up with more segments than we want to), we can add a second clustering step to join superpixels belonging to the same region.



Segments = 200  
Compactness = 20



# Canny edge detector

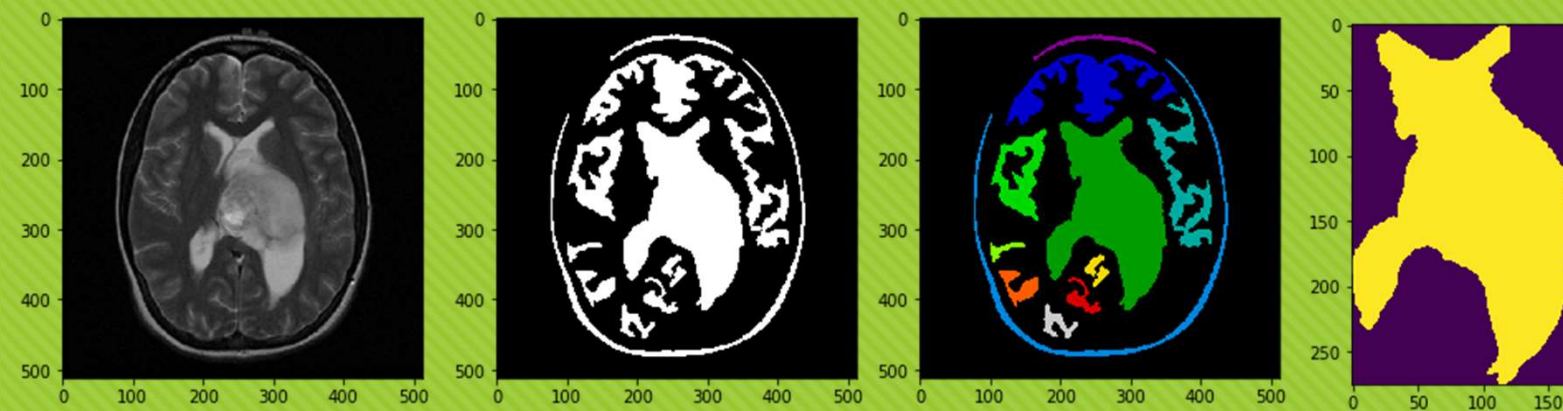




## Image cleaning

If we use the denoising + thresholding approach, the result of the thresholding is not completely what we want: small objects are detected, and small holes exist in the objects. Such defects of the segmentation can be amended, using the knowledge that no small holes should exist, and that blobs have a minimal size.

Utility functions to modify binary images are found in the morphology submodule. Although mathematical morphology encompasses a large set of possible operations, we will only see here how to remove small objects.

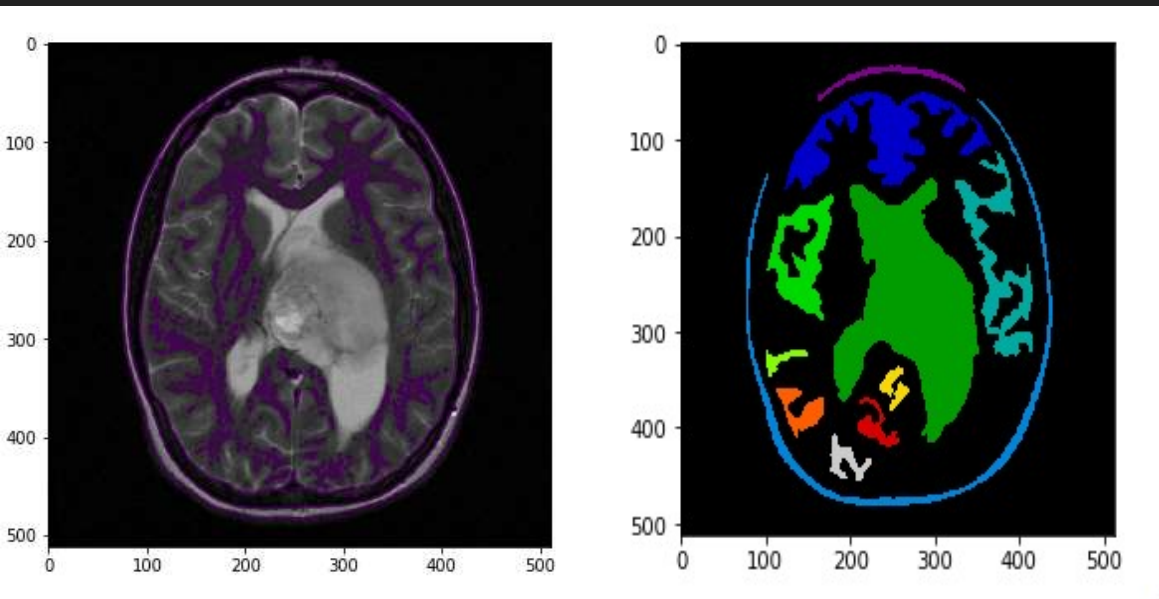


## Measuring region properties

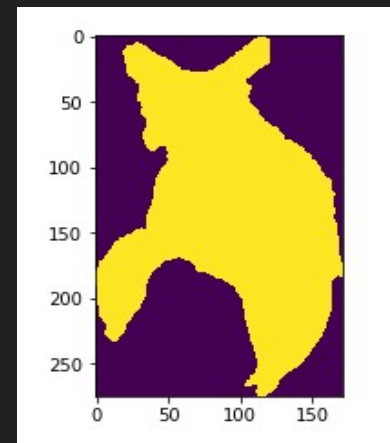
The segmentation of foreground (objects) and background results in a binary image. In order to measure the properties of the different sections, one must first attribute a different label to each section (identified as a connected component of the foreground phase). Then, the utility function `measure.regionprops` can be used to compute several properties of the labeled regions.

Properties of the regions can be used for classifying the objects, for example with `scikit-learn`.

# Extracting objects from image



Measuring region properties of the affected region



Area = 24480  
Perimeter = 1006.67



**Thank you!!**