Heart Health

A Kaggle based project

Sprint #2

Spring 2016

Quick preview

1. Image processing
2. Machine learning and data mining
3. Web and database design

Image Processing & Machine Learning

by: Yida Zhang & Shuo lin

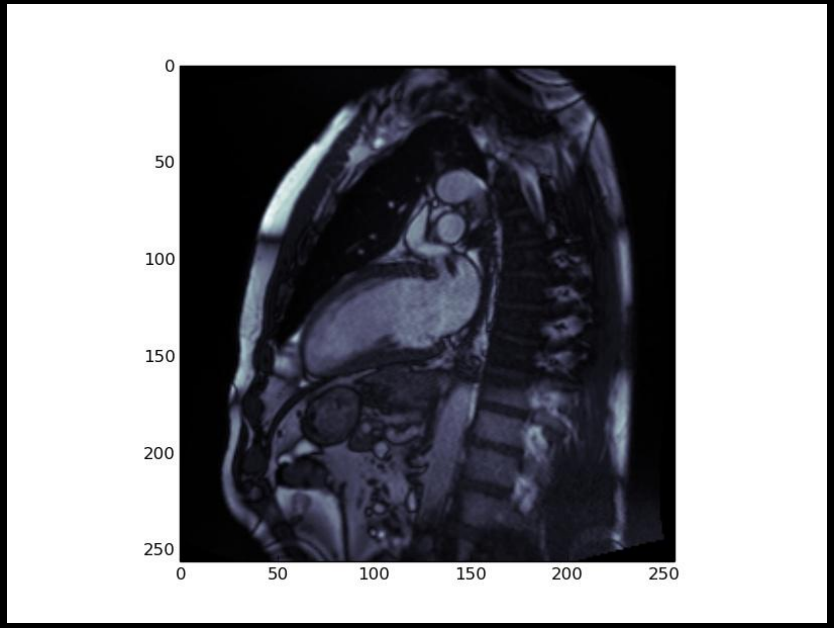
Tools in python

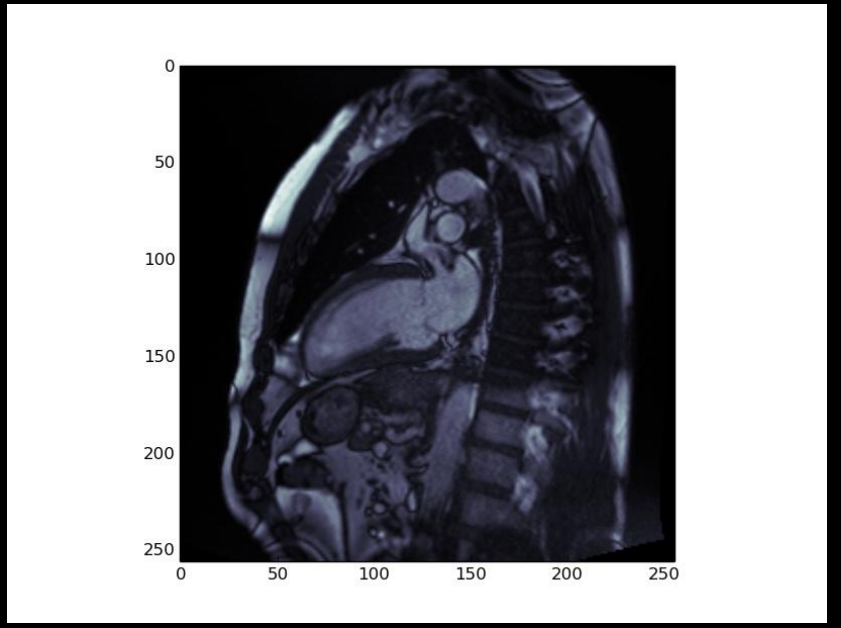
- NumPy : matrix and math operation
- SciPy : more fully-featured versions of the linear algebra modules
- OpenCV: Image processing library
- Matplotlib: Plotting tool
- Pydicom: tools especilly desinged for DICOM

Info. in DICOM images

```
(0008, 0005) Specific Character Set               CS: 'ISO_IR 100'
(0008, 0008) Image Type                           CS: ['ORIGINAL', 'PRIMARY', 'M'
, 'RETRO', 'NORM', 'DIS2D']
(0008, 0013) Instance Creation Time               TM: '144829.031000'
(0008, 0016) SOP Class UID                        UI: MR Image Storage
(0008, 0018) SOP Instance UID                     UI: 1.3.6.1.4.1.9590.100.1.2.23
726949521305755600867233689269279946 3
(0008, 0030) Study Time                           TM: '141855.484000'
(0008, 0031) Series Time                          TM: '144829'
(0008, 0032) Acquisition Time                     TM: '144819.942500'
(0008, 0060) Modality                             CS: 'MR'
(0008, 0070) Manufacturer                         LO: 'SIEMENS'
(0008, 103e) Series Description                   LO: '2ch'
(0008, 1090) Manufacturer's Model Name            LO: 'Aera'
(0008, 1140)  Referenced Image Sequence   3 item(s) ----
   (0008, 1150) Referenced SOP Class UID             UI: MR Image Storage
   ---------
   (0008, 1150) Referenced SOP Class UID             UI: MR Image Storage
   ---------
   (0008, 1150) Referenced SOP Class UID             UI: MR Image Storage
   ---------
(0010, 0010) Patient's Name                       PN: 'NDSB_1'
(0010, 0020) Patient ID                           LO: '1'
(0010, 0030) Patient's Birth Date                 DA: '19000101'
(0010, 0040) Patient's Sex                        CS: 'M'
(0010, 1010) Patient's Age                        AS: '050Y'
(0010, 1040) Patient's Address                    LO: ''
(0010, 2154) Patient's Telephone Numbers          SH: ''
(0018, 0015) Body Part Examined                   CS: 'HEART'
(0018, 0020) Scanning Sequence                    CS: 'GR'
(0018, 0021) Sequence Variant                     CS: ['SK', 'SS']
(0018, 0022) Scan Options                         CS: 'CT'
(0018, 0023) MR Acquisition Type                  CS: '2D'
(0018, 0024) Sequence Name                        SH: '*tfi2d1_15'
```

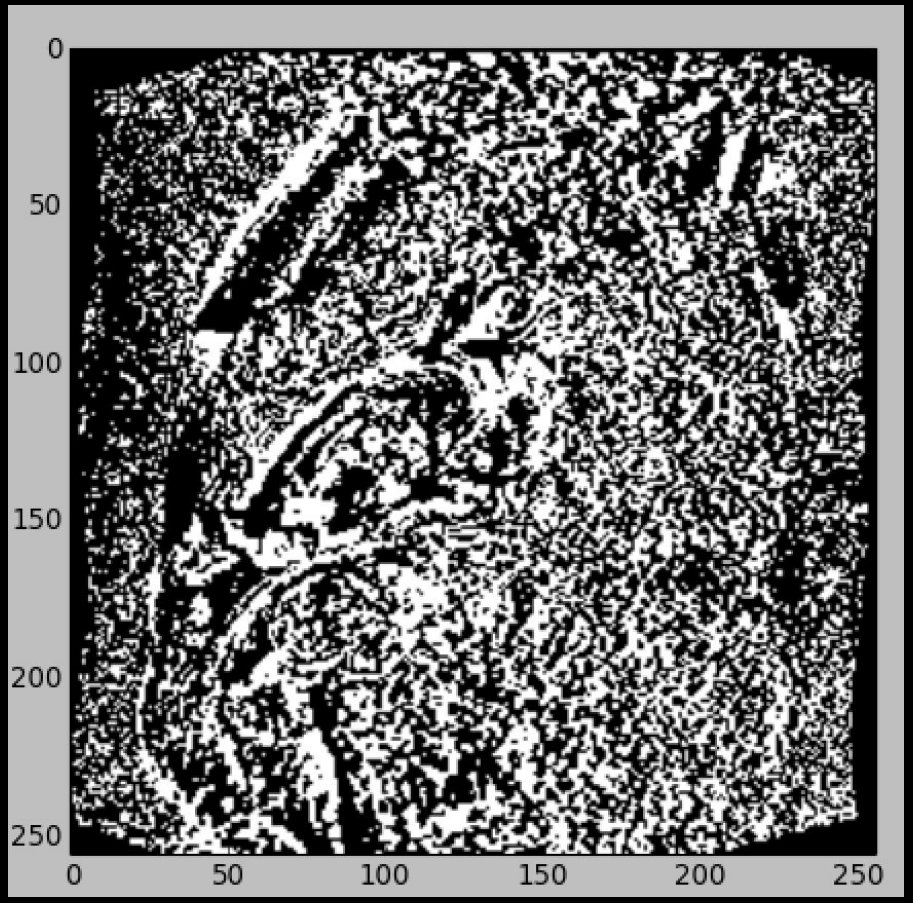
Play with pydicom





Two-chamber images

Plot moving part



Plot the moving part in different images

Kaggle tutorial

1. Load dataset

Combine the data from the different files into one big NumPy array

2. Calculate ROIs

ROI is a circle-shaped region of nonzero pixels which contains the heart

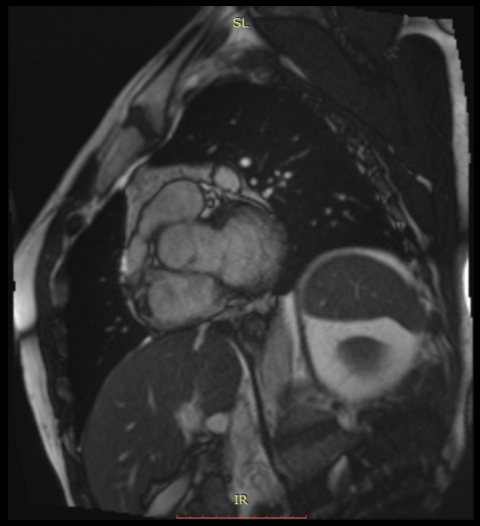
3. Calculate Areas

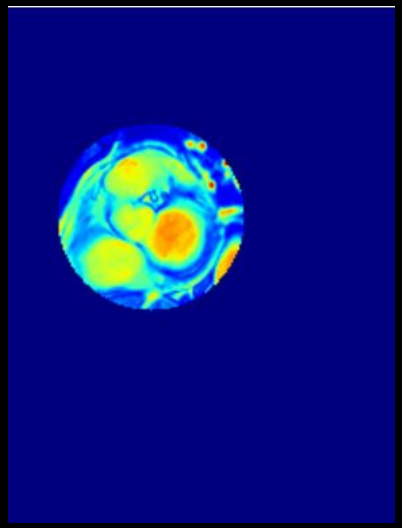
Locate the LV Blood Pool
high intensity: blood
low intensity: myocardial tissue

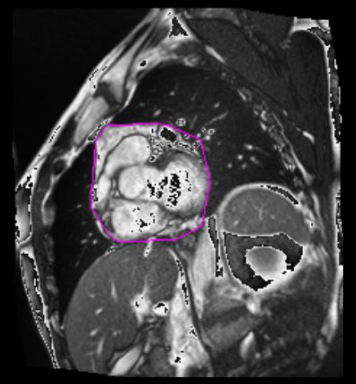
4. Calculate Total Volume

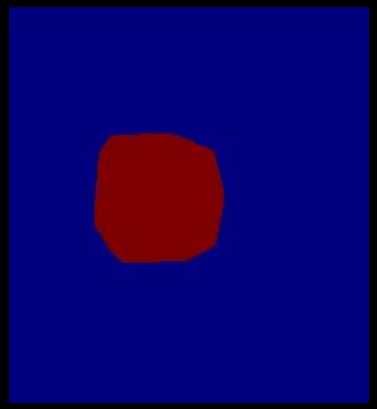
EDV: end-diastolic volume
ESV: end-systolic volume
EF = (EDV - ESV) / EDV

Kaggle tutorial







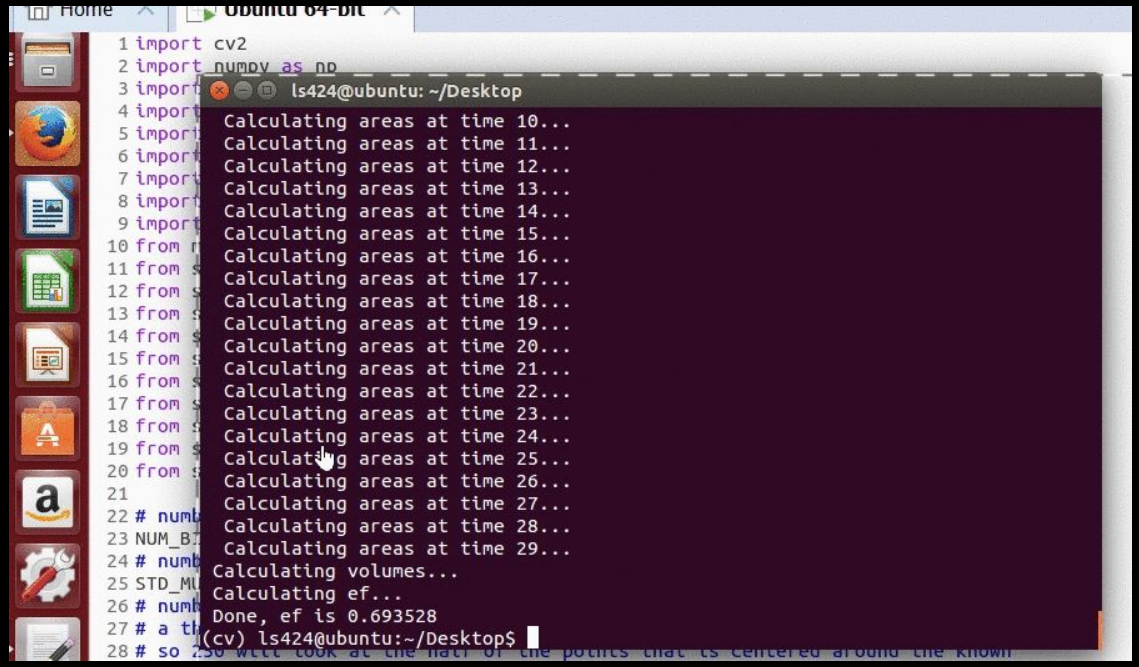


Original Dicom image

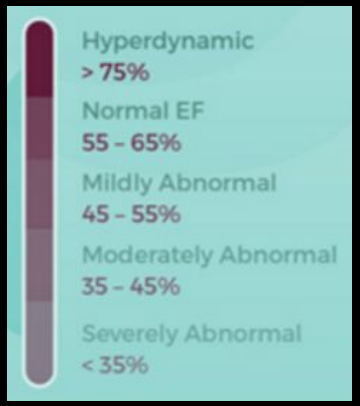
ROI

Blood Pool

Process & Output




```
{
  "esv": 71.065504696759618,
  "ef": 0.58387380777820974,
  "edv": 170.77873497297801,
  "areas": [
    {
      "0": 934,
      "1": 2529,
      "2": 858,
      "3": 793,
      "4": 731,
      "5": 738,
      "6": 640,
      "7": 381,
      "8": 488,
      "9": 239,
      "10": 67
    },
```

2	54.6	137.2

Predicted EF: 0.5838

Actual EF: 0.6020

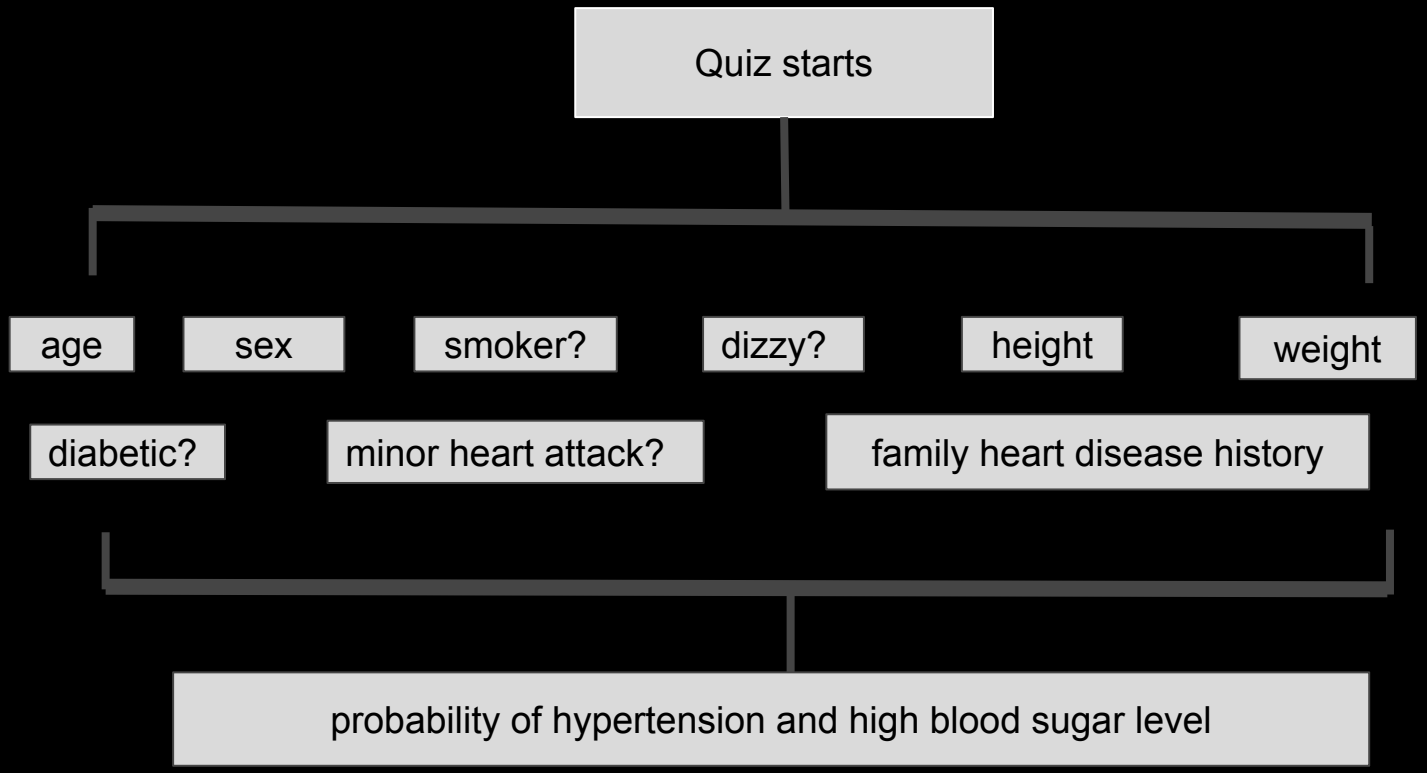
Data Mining

- Pre-diagnose system

by: Ganyu

Pre-diagnose system structure

using big data set of patients and supervised learning to predict two of the most common factors causing heart disease



Quick tutorial for data learning

1. Pre-process the data

2. Find suitable algorithms

3. Compare results and do cross-validation to modify parameters

Data pre-processing

1. Deal with incomplete data

2. Categorical data:
   Eg. ‘High’-’Low’; ‘Male’-’Female’; ‘Yes’_’No’;
   We usually use binary value like 0-1 pair

   For more classes like’High’_’Medium’_’Low’? (0 0 0)

3. Data in different range
   Normalization in to same range(with prior)

4. Data type(single,double)

Popular classification Algorithms

1. Gaussian Discriminative
2. Naive Bayes
3. K-nearest Neighborhood
4. Logistic Regression
5. SVM

Test and Cross-validation

1. Divide all data into training set and testing set.

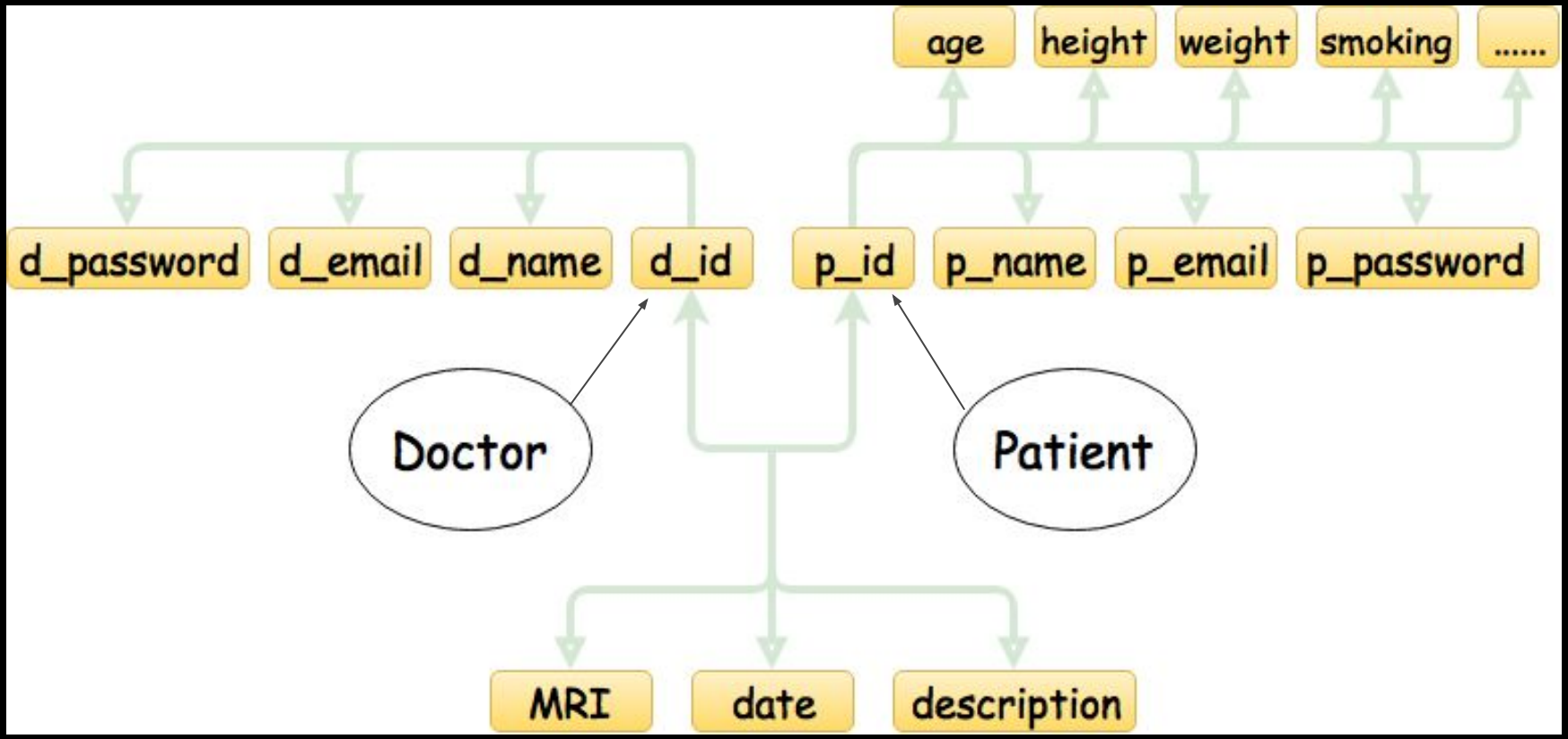
2. Run the algorithm and keep rotate your testing set

3. Revise the parameters with highest CCR(Correct Classification Rate)

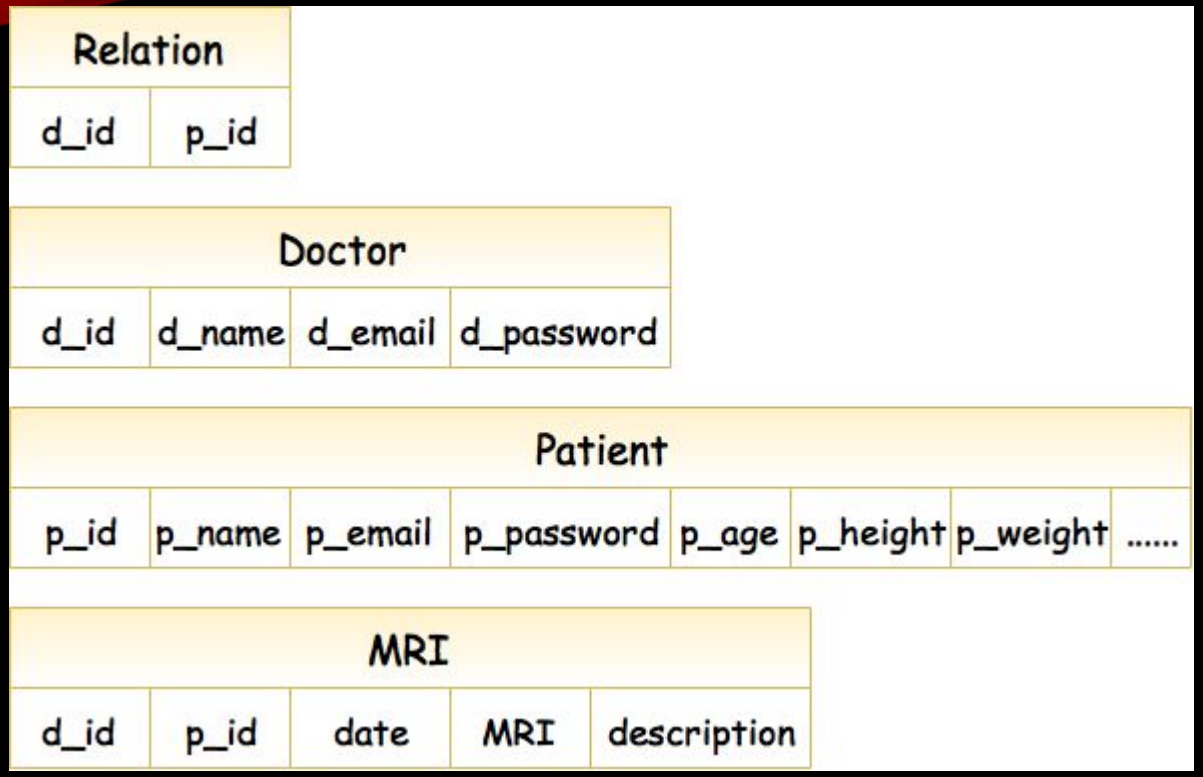
Web and Database design

by: Yixuan & Ruojin

Attributes and their relations



Schemas



Relation	
d_id	p_id

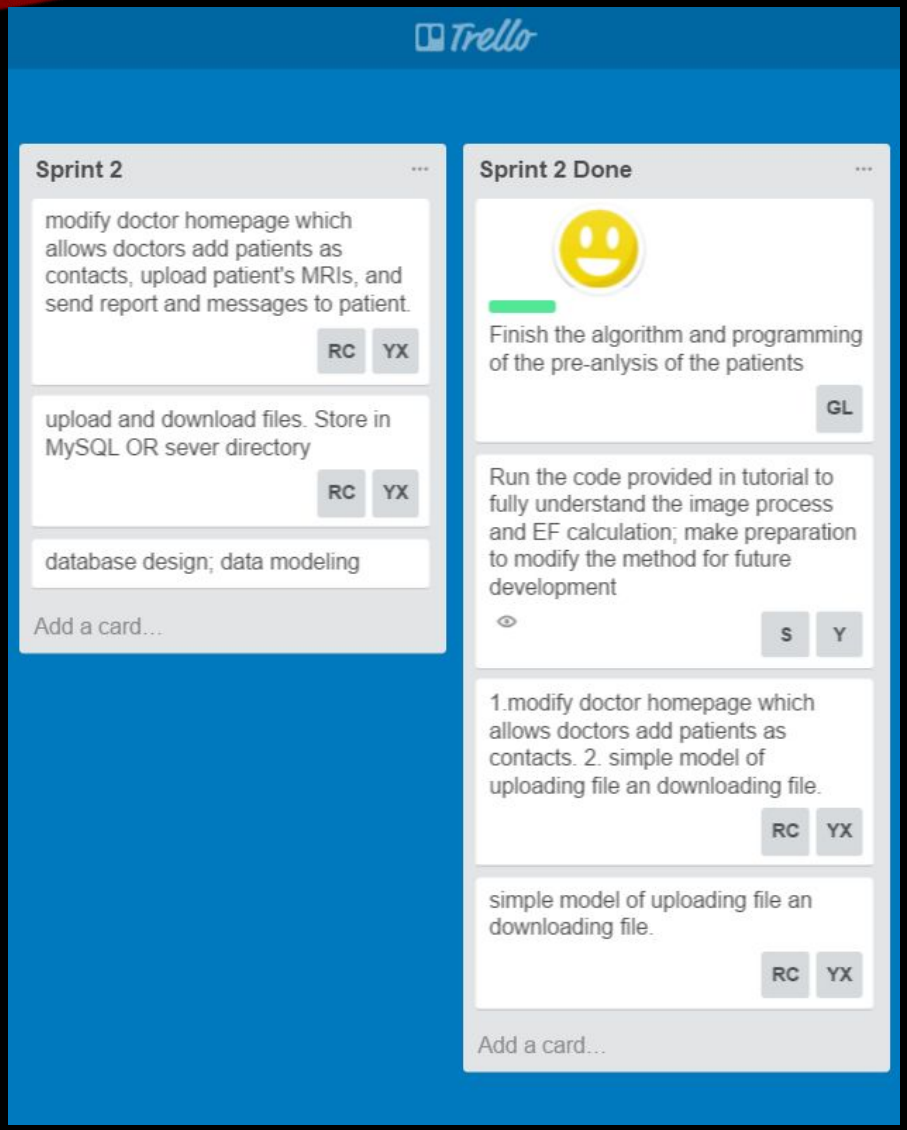
Doctor			
d_id	d_name	d_email	d_password

Patient							
p_id	p_name	p_email	p_password	p_age	p_height	p_weight	.....

MRI				
d_id	p_id	date	MRI	description

1. Includes all needed data and their relations
2. Easy and fast to implement
3. Minimum amount of duplicate tuples
4. Fast and flexible reading and updating of the data

Trello Board



Load file Page:
http://52.36.104.200/LoadFile/

Demo:
http://52.36.104.200/heart/

Sprint 3 goals

1. Implement connections between doctors and patients in backend.
2. Complete pre-diagnose feature in website.
   (Challenge: run matlab file on server)
3. Further explore the tutorial and try to modify its algorithms and Method
4. Based on kaggle tutorial, improve our algorithm
5. Start machine learning part for DICOM images.