Heart Health

A Kaggle based project

Sprint #3

Spring 2016

Quick preview

1. Image processing and data mining
2. Web and database design

Image Processing & Data Mining

by: Yida Zhang & Shuo lin

Quick review

1. Dataset preprocess

2. Calculate ROIs

ROI is a circle-shaped region the left ventricle

3. Calculate Areas

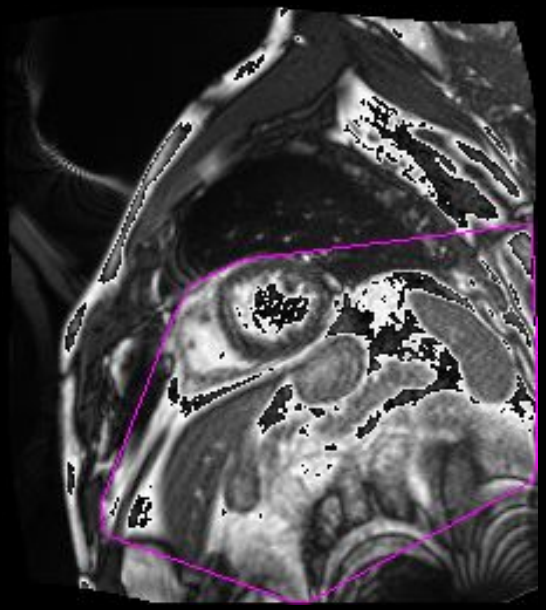
Locate the LV Blood Pool

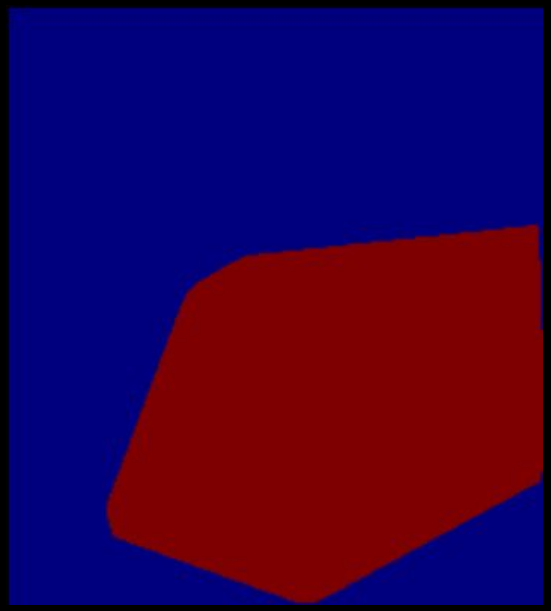
4. Calculate Total Volume

EF = (EDV - ESV) / EDV

Problems we met

- the segmentation picked up a few adjacent structures



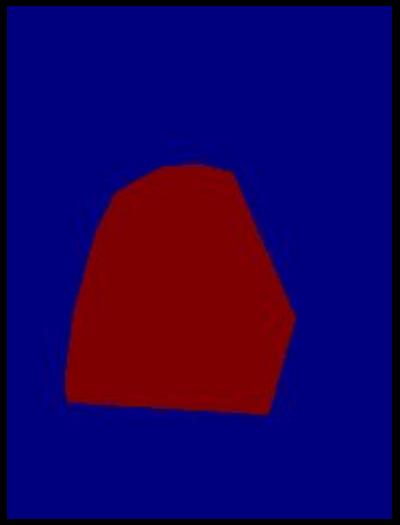
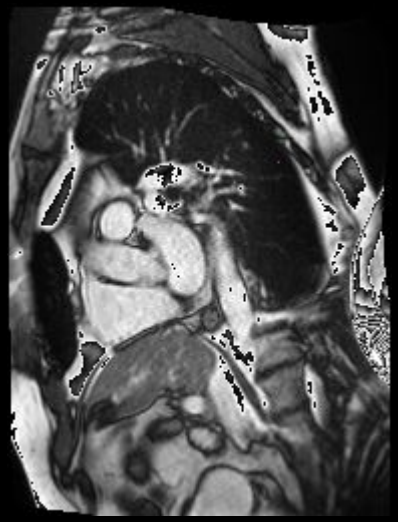
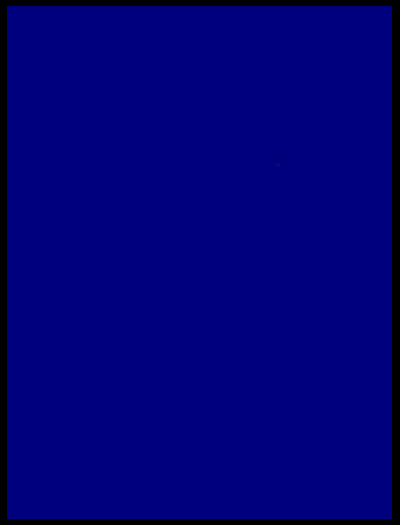



```
"esv": 1088.2866524285325,
"ef": 0.72423149350058824,
"edv": 3946.3775840221051,
```


Dataset	Actual EDV	Actual ESV	Predicted ED	Predicted ES
1	246.7	108.3	0	0
2	137.2	54.6	170.77874	71.065505
3	99.3	32.7	84.212982	23.712354
4	154.5	57.7	182.77331	115.50918
5	235.5	83.3	0	0
6	317.9	225.3	585.71509	266.908825
7	138	64.9	185.79631	57.242091
8	305.5	158.3	373.77546	130.899111
9	152.2	61.4	247.59362	132.709245
10	219.3	105.2	186.33365	100.045947
11	121.7	45.4	0.000018	0.000007
12	93.5	27.8	264.08122	111.095189
13	155.6	56.4	128.97516	63.306379
14	175.2	50.1	0	0
15	236.2	124.1	218.05495	138.35449
16	204.3	117.6	180.84524	104.298053
17	179.4	75.4	399.65892	258.367642
18	174.3	107.9	194.45428	105.736059
19	194.9	115.3	175.76019	99.284685
20	162.3	72.4	249.40962	163.711677
21	182.7	95.9	115.52259	57.738675
22	124.2	44.1	548.5116	232.094493

Problems we met

- Some sample may not be well processed through our algorithm:
  - There is no sharp edge between the LV (ROI) and the remain area.

Possible Methods

- Change the parameters that will affect the region of interest of the model
  - STD_MULTIPLIER, NUM_BINS

- To improve volume accuracy
  - Each slice will have a max value, and a min value during 30 time series
  - Assuming a patient has a series of images with 10 slices
  - We can create a 1x20 matrix.
  - Design a function to map this 1x20 matrix into a EF value

- To improve the EF accuracy
  - We have the distribution of true EF
  - We calculate distribution of predicted EF using the present algorithm
  - Design a function to map the predicted EF to true EF

How to find hidden pattern inside data!

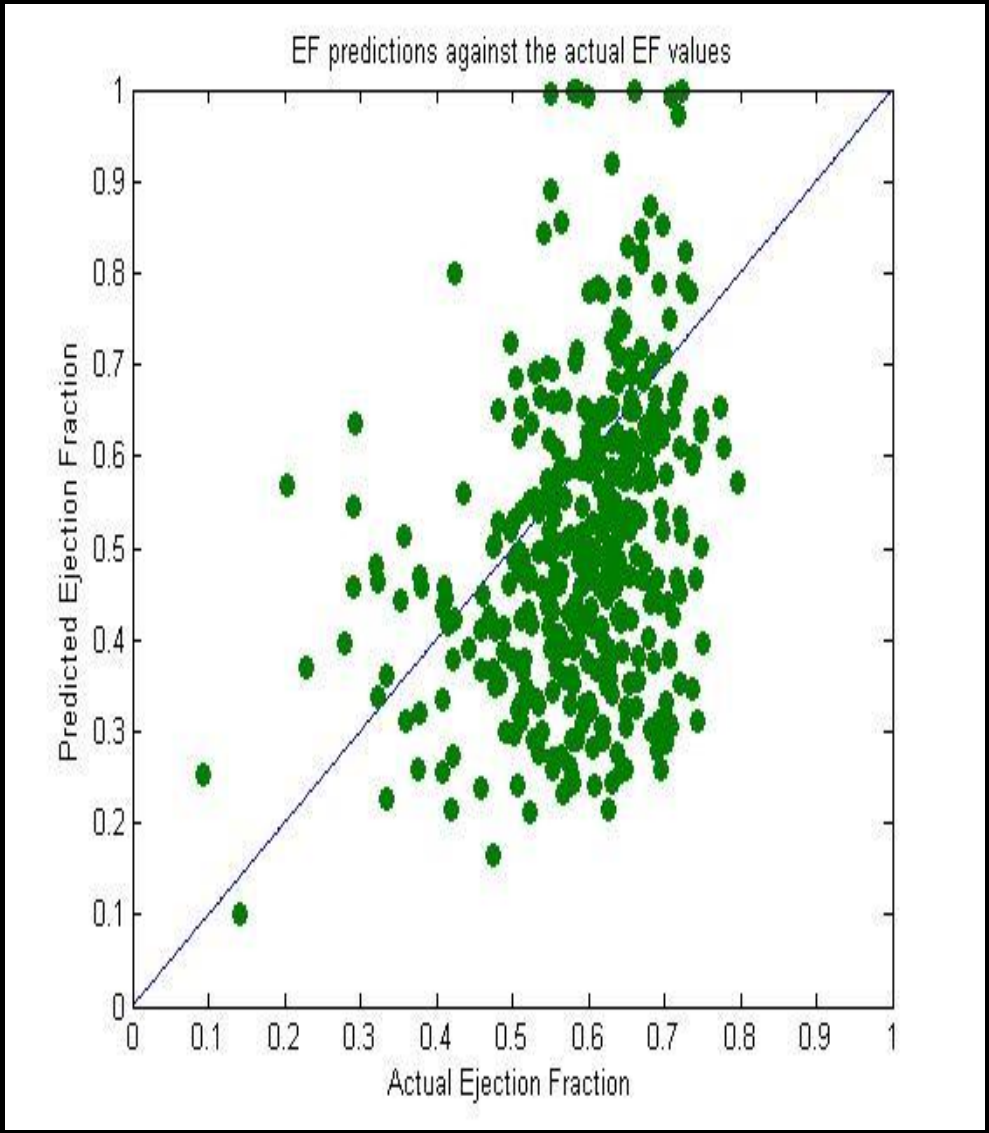
by: Ganyu Lian

Procedure

1. Clean the data set

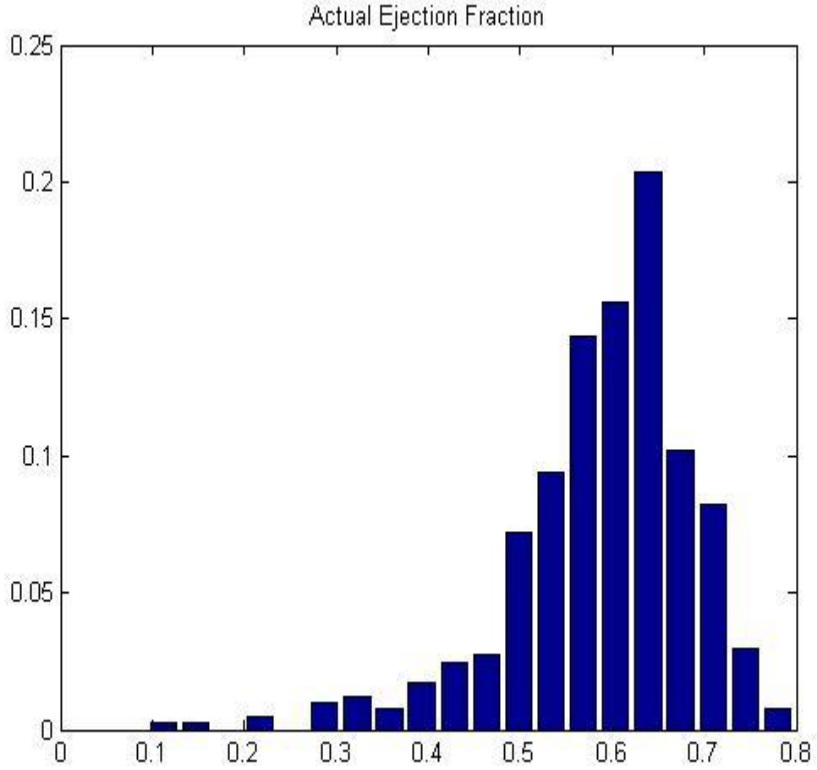
2. Come up with a model to map our prediction to the truth which has minimum cost:
eg. f(predict) argmin(f) cost

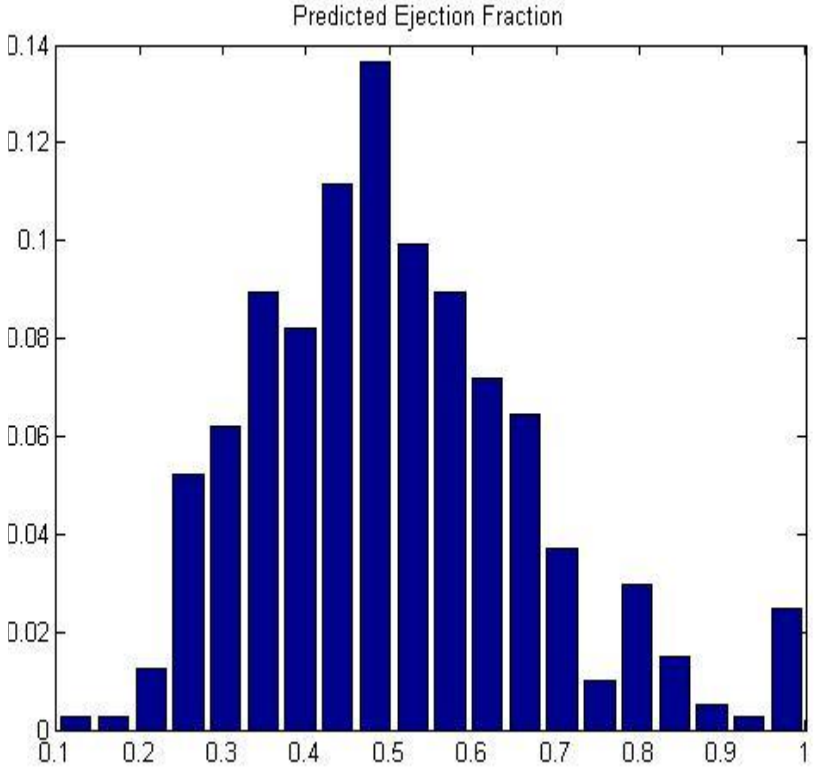
3. Due to lack of data points, we need to do cross-validation of our model.



Method Results Comparison

	Original	Robust	Gaussian	Spline
Mean Absolute Error	60.7385	28.4581	150.589	150.7813
Mean Square Error	14.0939	3.8493	61.192	60.865
MAE Improvement	1	2.1343	0.4033	0.4028
MSE Improvement	1	3.6614	0.2303	0.2316





So the mapping from right side to left side is nothing but a Gaussian shift and normalization.

$f(x)=(x-\mu x)/\sigma^2=kx+b$

Correct Prediction Rate= 0.9107(error within 0.15)

Web and Database design

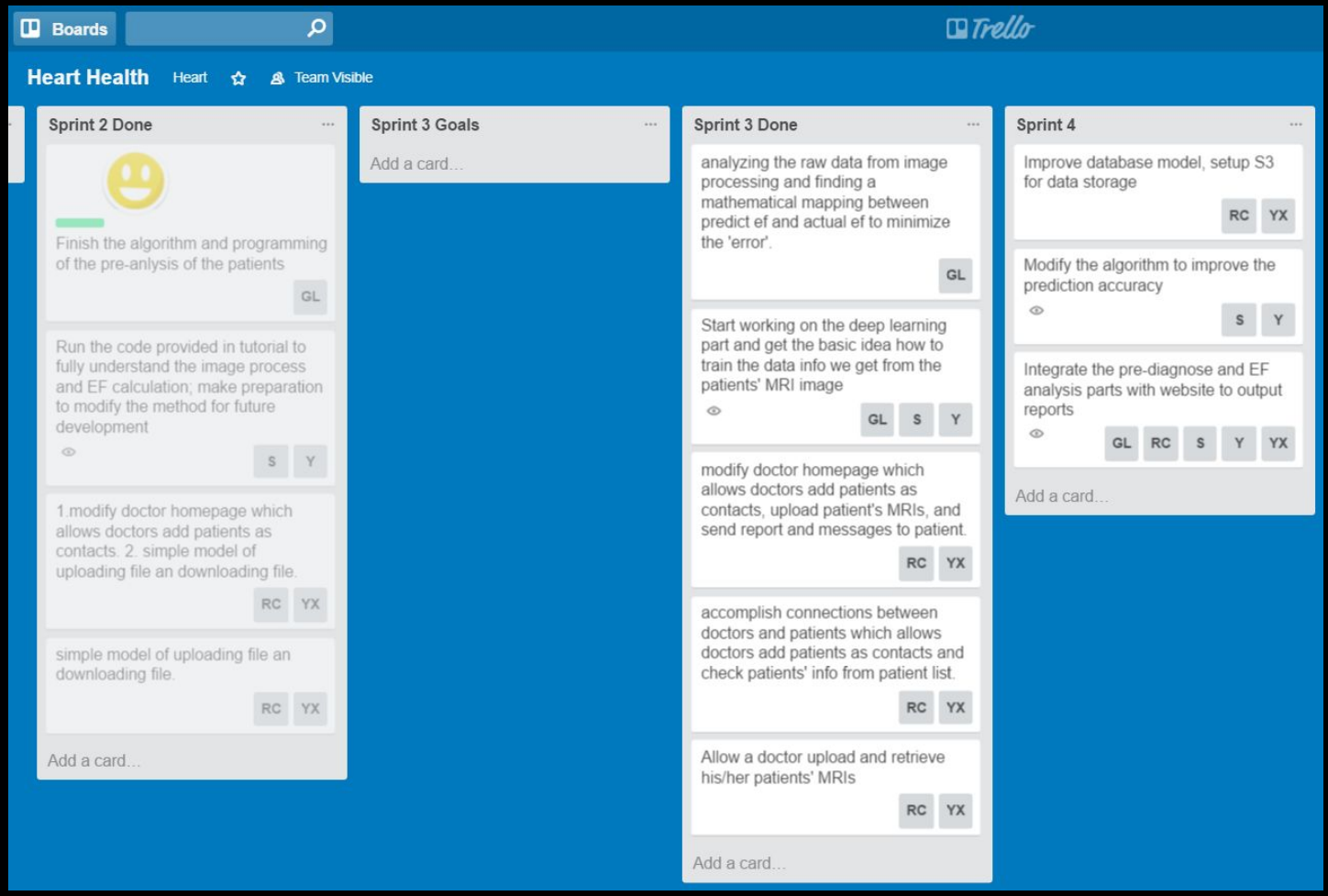
by: Yixuan & Ruojin

Demo：
http://52.36.104.200/heart/

Final Sprint goals

1. Complete pre-diagnose feature on website
   (Challenge: run matlab file on server)
2. Further explore the tutorial and improve the accuracy of the algorithm
3. Finish the data mining model
4. Integrate the image processing and data mining part with Website and generate reliable reports
5. Improve database models; utilize S3 storage on aws
6. Run automated tests on our website

Trello Board



Thank you!

Q & A