Final Project Presentation

Mammographic Mass Project

Jose C Fernandez EEL 6825 EDGE
Group 19
University of Florida

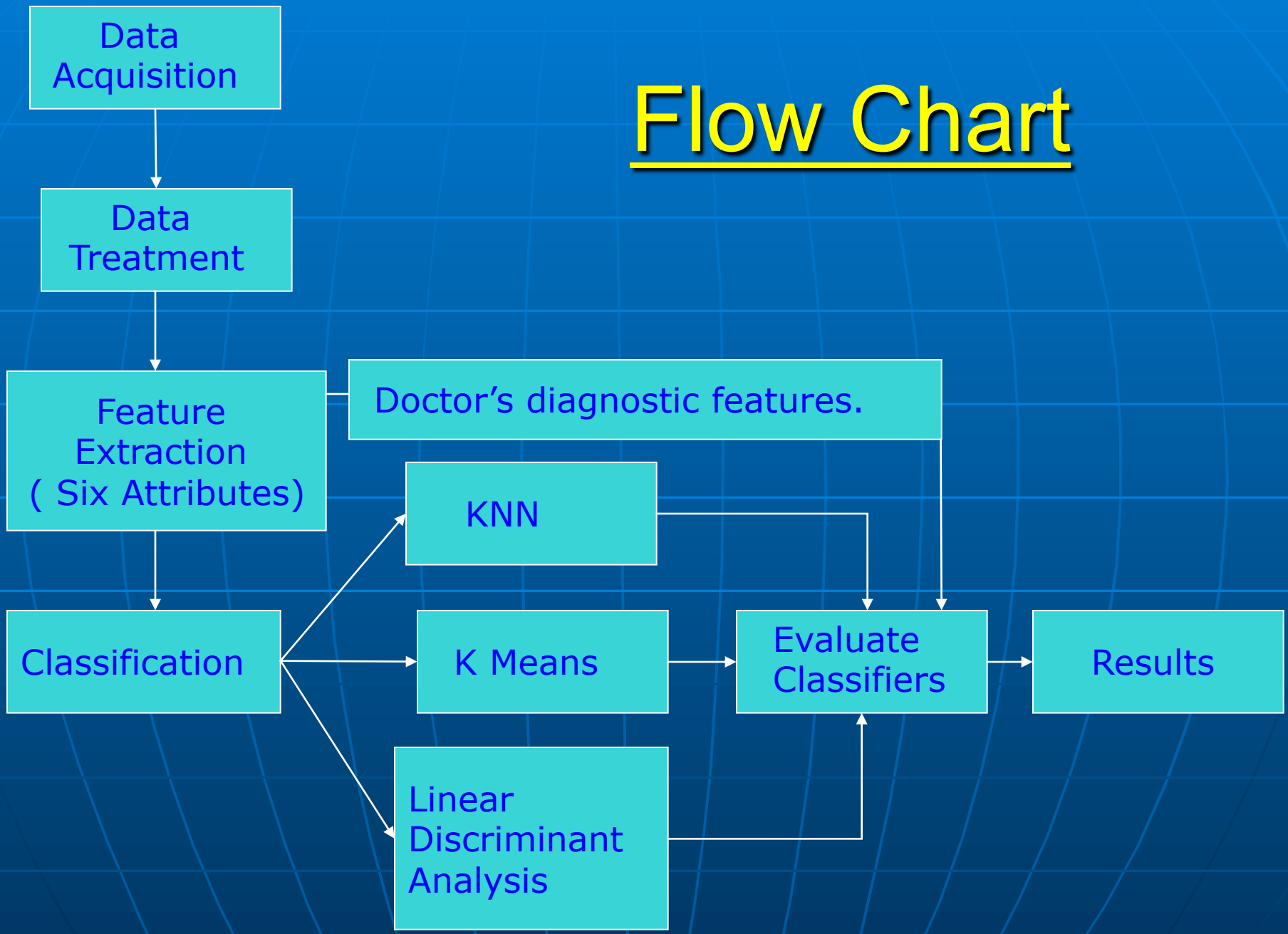
Objective

The objective of this project is to use BI-RADS attributes to successfully discriminate between benign and malignant mammographic masses, to diagnose breast cancer.

Classification Goals

The project needs to be able to predict the presence of malignant tumors in abnormal mammograms based on the typical mass characteristics such as Density, Shape and Margin.

Flow Chart


Data Acquisition
Data Treatment
Feature Extraction (Six Attributes)
Doctor's diagnostic features.
KNN
Classification
K Means
Evaluate Classifiers
Results
Linear Discriminant Analysis

Data Preparation

- Before starting to work with this data set , I noticed that there was a warning about missing fields of data in some samples. I proceeded to do data treatment by Listwise Deletion. This way I can eliminate patients with missing data, achieving unbiased parameter estimates.

- Since the data ranges are known parameters we can eliminate outliers by discarding as well the patients with out of range data that may bias the classification and skew results. For this particular data set, the reduction is less than 5 % , not affecting the outcomes in a significant way.

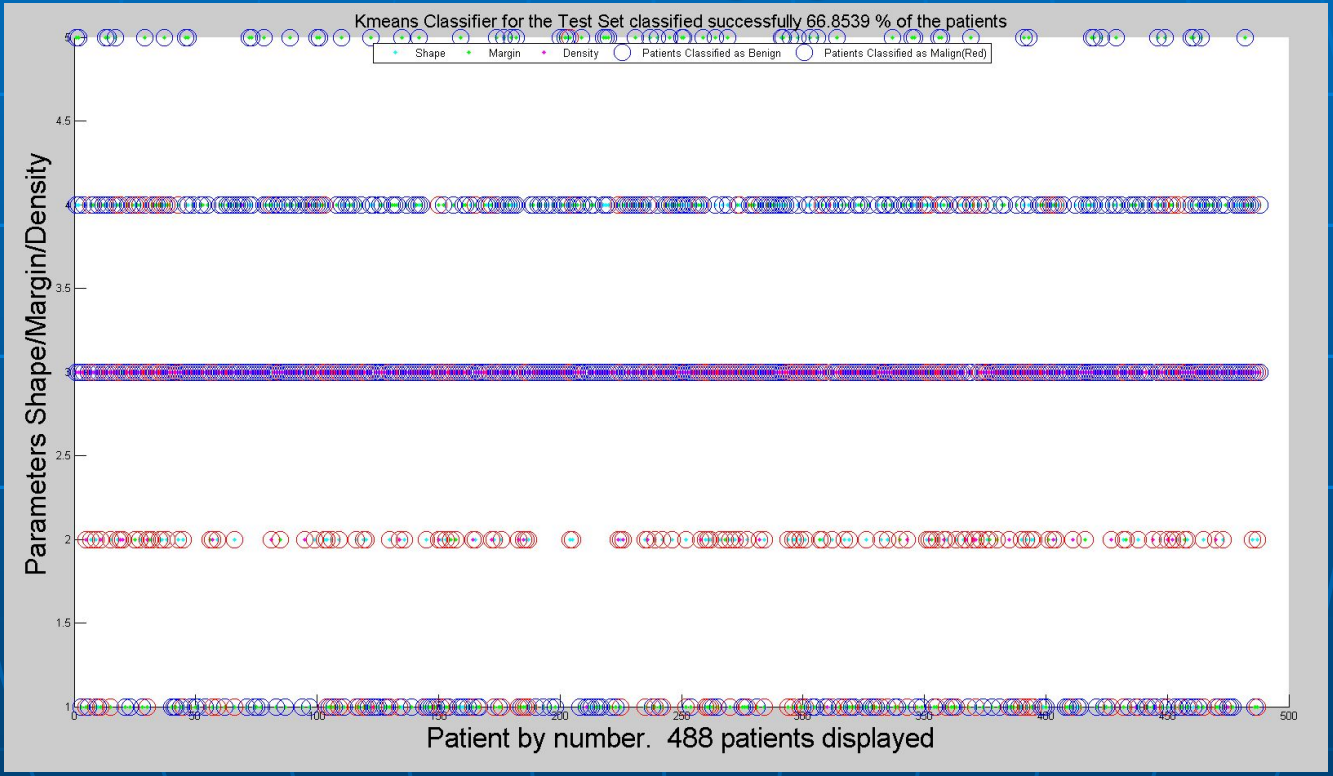
Results and Discussion

- Using Matlab, I implemented three classification methods: K Nearest Neighbor, K Means Clustering, and Linear Discrimination Analysis classification algorithms. After Classifying both training and testing sets, I compared the results of each classification algorithm to the Doctor's results. That provided me with a reference on how well each classifier performed.

- From my results it is easy to infer that due to the nature of the data and the way is distributed some classification methods are very efficient and very close to the Doctor's diagnosis, whereas some others doesn't perform so well. The data means are very close to each other, so the data sets overlap one to each other. Furthermore the covariances for each group have different values, so the data is spread unevenly.

K means Clustering

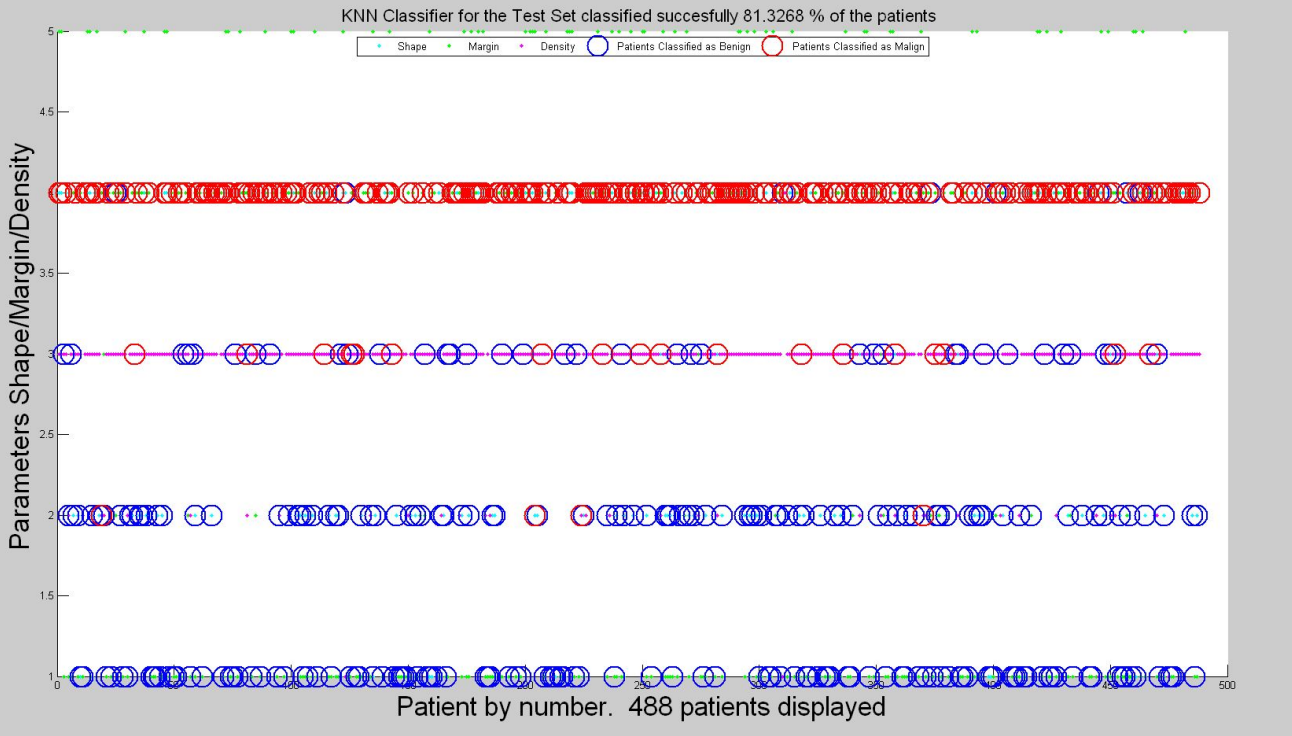
- The K means Clustering diagnostic has an accuracy of 69.03 % of the cases for the training set. The K means Clustering analysis successfully classified 70.97 % of the cases for the data set , converging after 2 iterations.



The patients classified as benign (blue) are overlapped with the malign (red) and vice versa. This overlapping amounts to an almost 30 % of classification errors.

K Nearest Neighbor

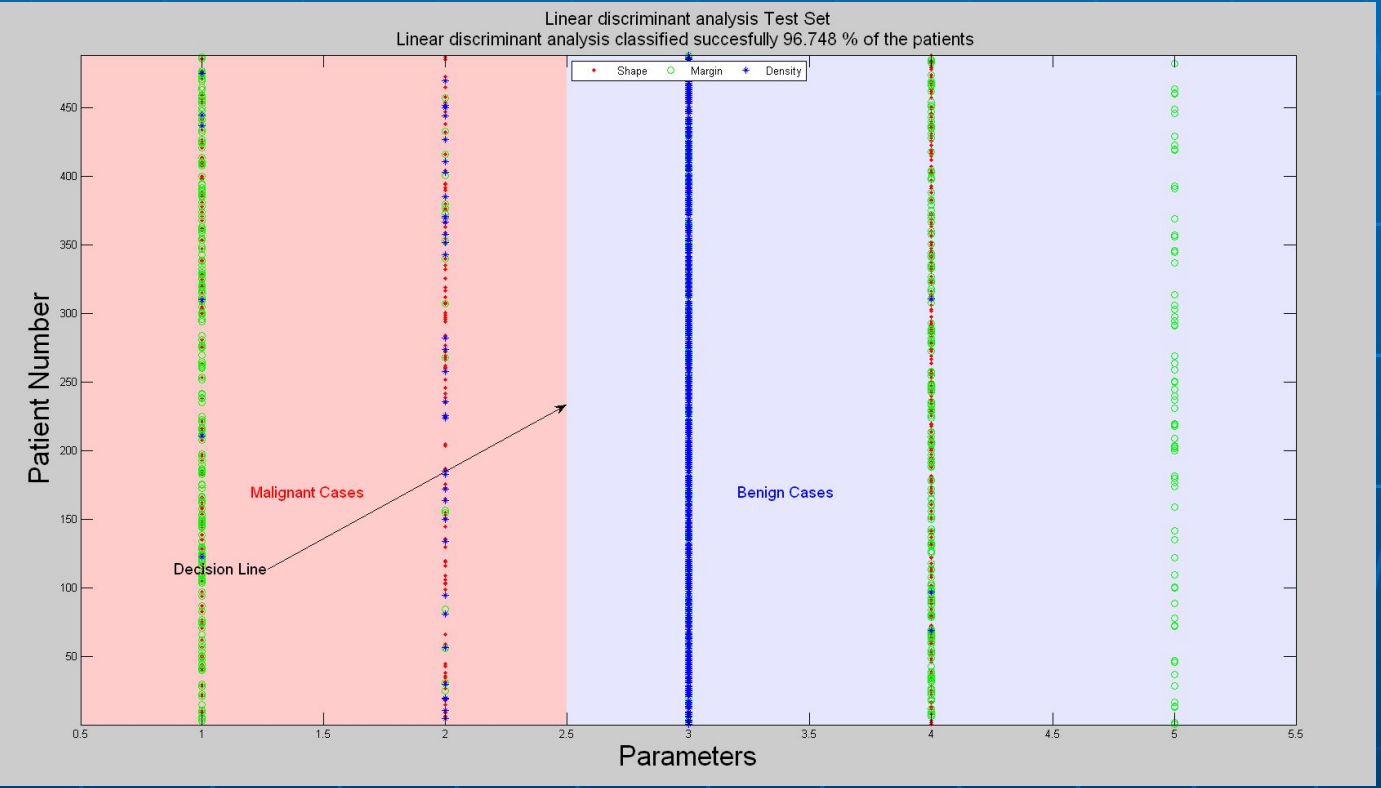
- Due to the factors listed under Results and Discussion the K Nearest Neighbor classifier also performs poorly, yielding around 71.86 % of correct classifications for the Training set and around 81.70 % for the Test Set.



The patients classified as benign (blue) are often overlapped with the malign (red) and vice versa.

Linear Discriminant Analysis

- The data classification using the default linear discriminant analysis (LDA) yielded an accuracy of 93.98 % of the patients for the Training Set and 95.2756 % of the patients for the Test Set out performing the rest of the classifiers.



By reversing the axis , the Malignant Cases are aligned by the left of the Decision line while the Benign cases are on the right of the graph.

Conclusions

By implementing three different classifier algorithms and modifying and optimizing several parameters in each of them, I could successfully classify and predict the malignancy of tumors. Given the readings of Margin, Shape and Density, the best classifier performed very close to the Doctor's diagnostics. I used the BI-RADs assignment and the Severity parameters from the Doctor's diagnostic to correlate the results.

Table of results

Classifier	Performance for the Training Set	Performance for the Test Set
K means Clustering.	69.03%	70.97 %
K Nearest Neighbor.	71.87 %	81.70 %
Linear Discriminant Analysis.	93.98 %	95.23 %

Main References

•Fiona J. Gilbert, F.R.C.R. et al. Single Reading with Computer-Aided Detection for Screening Mammography. Internet: http://www.nejm.org/doi/pdf/10.1056/NEJMoa0803545, October 16, 2008 [April 22, 2012].

•Altman, Alan P. M.D., et al. " Screening Mammogram" Internet: http://www.associnobgyn.com/br_screen.htm National Cancer Institute, December 2010 [April 22, 2012].
M. Elter, R. Schulz-Wendtland and T. Wittenberg (2007) "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process". Medical Physics 34(11), pp. 4164-4172 Internet: http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass [April 22, 2012].

•Margaret M. Eberl, MD, MPH,et al. " BI-RADS Classification for Management of Abnormal Mammograms". Internet: http://www.jabfm.org/content/19/2/161/T1.expansion.html , March–April 2006 , [April 22, 2012].

•Website breast-cancer.ca, Steven B. Halls, MD " Epidermoid Cyst " Internet: http://breast-cancer.ca/miscellaneous-breast-lesions/epidermoid-cyst.htm,17-March-2011 [April 22, 2012].

The rest of the references can be found in the main paper.

END

Thank you for your time, please email any questions to:

Jose C Fernandez joseworks.org
jcarlox74@ufl.edu