Classifying Breast Lesions to Reduce False-Negatives

Github: https://github.com/aakashthumaty/Comp562MLFinalProject

JACOB CHANDRAN, ANDREW MCKINNON, AAKASH THUMATY, ROSS DIBETTA

1 ABSTRACT

Breast cancer detection methods dependent on visual assessment of fine needle aspiration lead to high variability in the diagnosis of malignant breast tissue. This paper explores different classification methods and their ability to facilitate consistency and accuracy in tissue characterization.

2 INTRODUCTION

Cancer is one of the leading causes of death in the United States each year. According to the United States Cancer Statistics Working Group, female breast cancer has the highest rate of incidence among women in the US at 124.2 per 100,000 women in 2016. This is more than the next three highest incidence rates combined.

If a mass is found during a clinical examination of a lump in the breast, often the next step in the examination is a recommendation for a Fine Needle Aspiration (FNA) cytology. This procedure involves drawing fluid from the mass in question and analyzing it under a microscope [Mendieta 2018]. The fluid is analyzed by a trained professional's eye to determine the nature of the lesion and plan treatment. At this phase a classification is made; the lump is deemed either cancerous (malignant) or non-cancerous (benign). There is a non-zero chance that a false-negative diagnosis occurs, or a non-cancerous determination when cancer actually exists, by this test examiner. This false-negative diagnosis can occur in up to 8% of classifications [Dey and Mitra 2017].

Further, the ability to detect breast cancer in its early stages allows administration of simpler treatments with fewer risks and reduces the mortality rate by up to 25% [Khairunnahar et al. 2019]. The use of a predictive model based on the cellular characteristics of the FNA test could reduce the occurrence of false-negative classifications.

This paper constructs four predictive binary classifiers to limit the number of false-negatives. An accurate model that checks the potential mistakes made by the human eye will help reduce the occurrence of false-negatives, and will increase the likelihood of a patient to initiate early treatment of breast cancer, thus reducing the likelihood of mortality.

3 BACKGROUND

3.1 Input Data

The data examined in this project was a set obtained from the Wisconsin Breast Cancer Database. It consists of 566 data points of which 354 were labeled benign, and 212 malignant. Each data point contained 30 features. These features were derived from FNA breast mass images that describe characteristics of the cell nuclei using a single linear formulation algorithm [Bennett and Mangasarian 1992]. All features contain numeric data.

3.2 Methods

Given that the question of correctly classifying tumors is a question of binary classification, we identified four binary classification techniques to compare in classifying samples as benign or malignant. These classification techniques are the following:

(1) K Nearest Neighbors
(2) Support Vector Machine
(3) Decision Tree
(4) Logistic Regression

Due to the relatively small size of our data set (especially in comparison to other data sets in the medical imaging field) we used a 5-fold cross validation method to robustly test the accuracy of each of the classifiers listed above. Furthermore, for classifications performed using k nearest neighbors and decision tree classification we used an iterative parameter tuning technique to arrive at the optimal parameters for each model.

During our analysis and model creation we also explored the possibility of dimension reduction using principle component analysis (PCA). Again, we used an iterative testing approach to arrive at the optimal number of principle components for our model by testing principle components with i=5 through 30 components at increments of 5.

4 DATA ANALYSIS

After performing five-fold cross validation for all of the classification techniques listed above, we arrived at the following results:

Classification Technique	**% accuracy**
K Nearest Neighbors (KNN)	86.31
Support Vector Machine (SVM)	62.74
Decision Tree (DT)	91.92
Logistic Regression (LR)	95.27

As shown in the table above, logistic regression most accurately predicted benign and malignant tumors. Both logistic regression and decision tree classifiers had accuracy of over 90%, while support vector machine and k nearest neighbors performed significantly worse. The poor performance of support vector machines in this binary classification problem is unsurprising due to the high ratio of features to observations. The confusion matrix for the logistic regression classifier selected from one of our k-fold results can be

Author's address: Jacob Chandran, Andrew McKinnon, Aakash Thumaty, Ross DiBetta.

2019. XXXX-XXXX/2019/12-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

seen below: **diagnosis: 0 = malignant | diagnosis: 1 = benign**

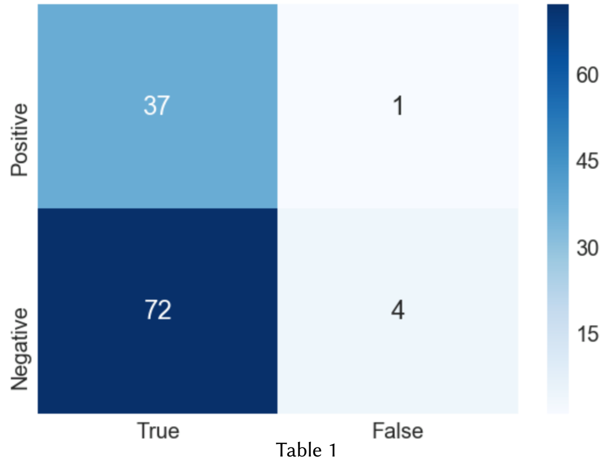

Table 1

The performance of this model was very strong, resulting in one false positive and four false negatives. Our false negative rate is a 56% improvement on doctor classifications. Additionally, taking into consideration the fact that our model's false negatives may not overlap with a doctor's false negatives suggests that a doctor augmented by our model would see an increase in accurate diagnoses.

Further results of interest from the analysis of our logistic regression classifier are the features most highly correlated with tumor classification. Fig 1, featuring "concavity_worst", illustrates the negative correlated relationship between the outcome (benign) and the feature. Inversely, Fig 2 highlights the positive correlation between radius_mean and the outcome (benign).

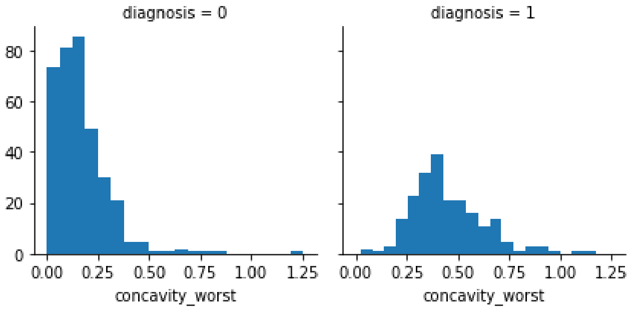

Fig. 1

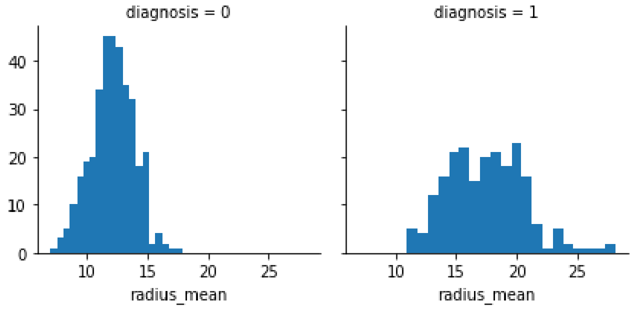

Fig. 2

Interpreting these features at a high level, concavity_worst which is the "worst" or largest mean value for severity of concave portions of the contour for the FNA image and radius_mean which is the mean of distances from center to points on the perimeter for the FNA image were most useful features from our data set in determining tumor classification.

Contrary to the prior two graphs analyzed, the following graphs (Fig. 3, Fig. 4) represent the features with lowest correlation to outcome (benign) resulting from our logistic regression classifier:

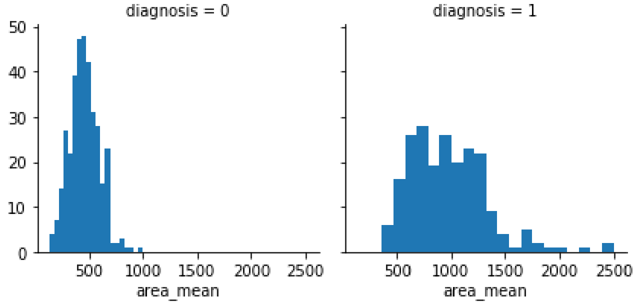

Fig. 3

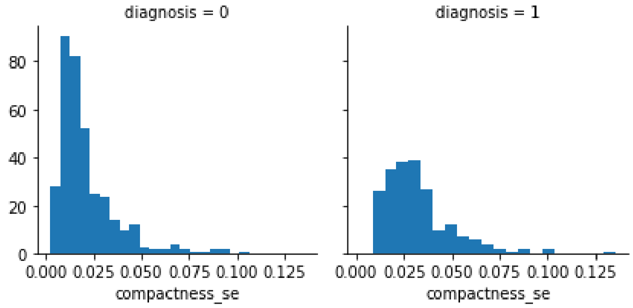

Fig. 4

We also tested whether a principle component analysis of our training data with 5, 10, 15, 20, 25, and 30 principle components would result in similar or better performance for each of our models. We selected the number of components to use in each model by testing accuracy using the number of principle components that resulted in the highest accuracy and compared this to accuracy achieved with no dimensional reduction. A comparison of the results of each model with and without PCA is shown below:

Model	% accuracy without PCA	% with PCA
KNN	86.31	92.45
LR	95.27	95.08
DT	91.92	91.04
SVM	62.74	62.74

As shown in these results, PCA did prove helpful in dimension reduction resulting in similar or better accuracy than the models achieved without using PCA. The optimal number of principle components for each model is shown in the table below:

Model	Principal Components
KNN	5
LR	20
DT	10
SVM	5

5 CONCLUSIONS AND FUTURE WORK

Our results show that a logistic regression model produces the highest accuracy in classifying a breast lesion as malignant or benign. Feature reduction to latent space using PCA yielded increased accuracy in classification using k-nearest neighbors and maintained model accuracy in the other classifiers. This indicates that dimension reduction may prove to be useful in reducing computational power and memory usage while maintaining accuracy in datasets with much larger feature sets. While these results can best be used in practice to reduce the number of false-negative diagnoses administered, further work can be done.

A clear limitation to the performance of our model in practice is the size of the dataset. Acquiring similarly formatted data that is classifiable by our model may prove to be difficult. Therefore, construction of a Neural Network to analyze the unaltered images produced by the FNA cytology would be a logical next step as this removes the use of the single linear transformation algorithm needed to currently process the images into the data format we ingest. Comparison of this new method against our results is vital. While processing images without an intermediary algorithm to convert image data into numeric features will improve usability in industry, sacrificing model accuracy could lead to more misdiagnoses and should be avoided.

REFERENCES

2019. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2018 submission data (1999-2016): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. www.cdc.gov/cancer/dataviz.

Kristin P. Bennett and O. L. Mangasarian. 1992. Robust Linear Programming Discrimination of Two Linearly Inseparable Sets. *Optimization Methods and Software* 1 (1992).

P. Dey and S. Mitra. 2017. Fine-needle aspiration and core biopsy in the diagnosis of breast lesions: A comparison and review of the literature. *CytoJournal* (2017).

Laila Khairunnahar et al. 2019. Classification of Malignant and Benign Tissue with Logistic Regression. *Informatics in Medicine Unlocked* 16 (2019).

Milton Mendieta. 2018. Organ identification on shrimp histological images: A comparative study considering CNN and feature engineering. *Ecuador Technical Chapters Meeting (ETCM) 2018 IEEE Third* (2018), 1–6.