
Breast Cancer Detection Using AI

Ken Chen*

Department of Computer Science and Software Engineering
Auburn University
345 W Magnolia Ave, Auburn, AL 36849
kzc0068@auburn.edu

Donglai Liu

Department of Computer Science and Software Engineering
Auburn University
Address 345 W Magnolia Ave, Auburn, AL 36849
dzl0066@auburn.edu

Junyao Yang

Department of Industrial and Systems Engineering
Auburn University
345 W Magnolia Ave, Auburn, AL 36849
jzy0040@auburn.edu

Abstract

Scientific evidences show that treatments can be more successful when identify breast cancer at early stages of the disease. Despite the existence of screening programs worldwide, correct identify accuracy is still suffering from high rates of false positives and false negatives. In this project, we will try to build an artificial intelligence system that is capable to detect abnormal cells at early stage of the disease. To assess it's performance, multiple comparisons are performed, for example, human vs machine and comparison among AI models. At this moment, we don't know the results and they will be added to this report as our project in progress.

1 Introduction

Breast Cancer's causes are multifactorial and one million women are newly diagnosed with breast cancer every year. Machine learning is a powerful tool and algorithms that facilitate prediction, pattern recognition and classification. In this project, we develop two algorithms of machine learning for breast cancer classification. We use the Wisconsin breast cancer database. The purpose of this project is studying and developing effective machine learning approaches for cancer classification using binary classification data set.

2 Literature Review

Support Vector Machine(SVM), Decision Tree(C4.5), Naive Bayes(NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) datasets are conducted[1]. Naive Bayes(NB) classifier and k nearest neighbor(KNN) are applied for breast cancer classification[2]. More recently,

*

deep learning methods have been researched to solve anomaly detection problems since they do not require explicit feature construction unlike previously methods[3].

3 Methods

Our first task is choosing a baseline algorithm. Based on our literature review, KNN and Naive Bayes algorithm are most well study model in this area. And they are all very basic machine learning algorithm. All other methods we used will be compare to our baseline algorithms on algorithm performance. We are aiming to develop a system that can perform better than our baseline. At this moment, I list the algorithm we are interested.

3.1 Dataset

Wisconsin Breast Cancer (original) datasets will be used.

3.2 K nearest neighbors

The KNN algorithm is used to predict the class. Given N training vector, suppose we have A and Z as training vectors in this bi-dimensional features space, we want to classify c which is feature vector. Classifying c depends on its k neighbors, and the majority vote, k is a positive integer, k is generally smaller than 5. And it's also a parameter we are interested during our experiments.

3.3 Naive Bayes

A Bayesian method is a basic result in probabilities and statistics, it can be defined as a framework to model decisions. In NBC, variables are conditionally independent; NBC can be used on data that directly influence each other to determine a model. From known training compounds, active (D) and inactive (H), Given representation B, the conditional probability distribution $P(B/D)$ and $P(B/H)$ are estimated, respectively. Bayesian classifiers are additionally well adapted for ranking of compound databases all with consideration to probability of activity. Bayesian classifiers use Bayes theorem, which is:

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)} \quad (1)$$

3.4 Neural Network

An ANN is a machine learning algorithm suitable for different tasks including classification, prediction and visualization. Furthermore, an ANN is suitable or multi-disciplinary tasks with the use of multiple types of data which may be unstructured, semi-structured and structured data. Consider that our dataset may not be a linear separate-able space which means linear classifier, like SVM and Perceptron classifier may not perform well. Then ANN will be a good candidate to resolve this issue. Because theoretically, Ann with 2 hidden layers is capable to approximate variety problems.

3.5 Convolutional Neural Network

Convolutional neural network(CNN) are the quickest rising areas of healthcare industry. CNN brings to research on medical imaging is not restricted to deep CNN for extraction of the imaging feature. Indeed, a second field that can support medical research is the use of CNN for synthetic image rendering. We will try to build an neural network with convolutional layer to detect breast cancer cells.

4 Experiments

To be added....

5 Analysis

To be added....

6 Conclusion

To be added....

References

- [1] Hiba Asria. & Hajar Mousannif. & Hassan Al Moatassime.(2016)Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. In The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) *Procedia Computer Science* vol 83, page 1064-1069. ISSN 1877-0509.
- [2] Meriem AMRANE. & Saliha OUKID. (2018) Breast Cancer Classification Using Machine Learning. 2018 *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting 10.1109/EBBT.2018.8391453*.
- [3] Li Shen. & Laurie R. Margolies. (2019) Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Scientific Reports* volume 9, Article number: 12495.