Breast cancer classification using Machine learning – A survey

Akshay Laxmish, Anoop K N, Amit B V, Shiva shankar N M, Mr. Saravana M K

[1]*Student, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.*
[2]*Associate Professor, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.*

---***---

Abstract - *This* work summarizes the experiments performed in order to improve the recall rate for breast cancer detection. Various methods were implemented, and optimizations performed on these methods followed by a comparison of the results. Some methodologies had to be discarded because of very low accuracy. Our final proposed methodology combines techniques like GLCM feature extraction and wavelet transform to detect both mass lesions and microcalcifications with almost the same accuracy as any of the methods used alone. Also, it shows a fair improvement in the recall rate. The feature reduction techniques used help in training the model more efficiently, while reducing the training time and also keeping the same accuracy. The results obtained are, however highly dependent on the dataset.

1. INTRODUCTION

Breast cancer is a type of cancer that develops from the breast tissue. Compared to other diseases or other cancers, breast cancer receives a greater share of resources and attention. Partly because of its relatively high prevalence and long-term survival rates, research is biased towards breast cancer. Some subjects, such as cancer-related weariness, have been considered miniscule except in women with breast cancer.

Breast cancer is the most common detected solid cancer and second foremost cause of cancer demise among women. Based on multiple randomized trials, screening mammography has been shown to decrease breast cancer-related mortality. However, despite this population-based benefit, routine mammography is associated with a high risk of false positive testing and may lead to over diagnosis of clinically insignificant lesions.

According to the federally funded Breast Cancer Surveillance Consortium, the overall sensitivity of digital mammography in the U.S. screening population is 84%, and the overall specificity is 91%.

The recall rate (the number of women who require additional testing after a screening) associated with mammography remains high (approximately 100/1000); yet, the proportion of women recalled who are eventually found to have breast cancer is quite low (5/1000). The impact to women with regards to increased anxiety and potential morbidity associated with false-positive screens and unnecessary downstream diagnostic workup cannot be underscored. False-positive recalls from mammography represent the tipping point for screening recommendations, with the U.S. Preventive Services Task Force recommending less frequent mammography screening due to these potential risks.

In order to improve radiologists' interpretive accuracy, multiple vendors have developed computer-assisted detection software to help radiologists identify subtle suspicious masses and calcifications. The goal of computer-assisted detection is to increase the radiologist's accuracy for detecting cancer and, thus, lead to improved patient outcomes. Unfortunately, population-based data suggests that computer-assisted detection has not had the hoped-for effect for interpretation of screening mammography. Computer-assisted 9 detection to date, developed on test case sets with little potential for dynamic improvement over time, has led to no significant improvement in any performance metric, including sensitivity, specificity, positive predictive value, recall rate and benign biopsy rate.

Given the promise of machine learning and deep learning to aid in improving medical diagnostic accuracy, we present our proposal for an improved model to reduce the recall rate for breast cancer screening, to better classify screening digital mammograms to improve interpretive accuracy and reduce the overall recall rate without increasing the number of false negative exams. An improved diagnostic tool may change the debate surrounding screening mammography and help tip the balance of routine screening towards greater benefit and less harm.

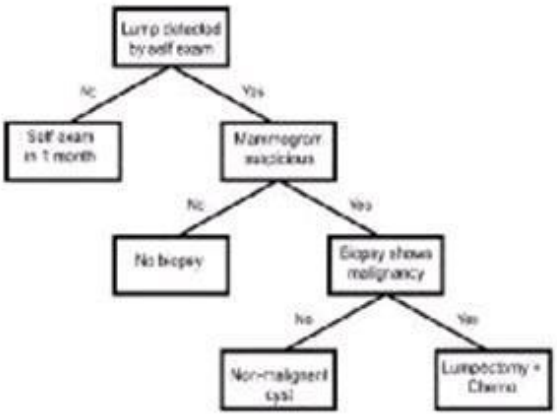
2. TECHNOLOGIES USED

2.1.1 Machine Learning

Machine learning is a sort of artificial intelligence (AI) on the expansion of computer programs that can change when exposed to new data. The procedure of machine learning is similar that of data mining. Both systems examine through data to search for patterns. However, instead of extracting data for human comprehension, as is the case in data mining applications, machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can predict what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets. Machine learning is closely linked to computational statistics, which also focuses on prediction-making using computers. It has robust bonds to mathematical optimization, which brings methods, theory and application domains to the field. Machine learning is occasionally combined with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can be unsupervised and can be used to acquire and create a baseline behavioral profiles for several entities and then used to find meaningful irregularities.

2.1.2 Decision Tree Learning

There are various approaches for machine target value. Decision tree learning is the utmost used techniques for supervised classification learning. For this section, we assume that all the features have finite discrete domains, and there is a single target 18 feature called the classification. Each part of the field of the classification is called a class. A decision tree or a classification tree is a tree in which each interior node is categorized with an input feature. The arcs impending from a node characterized with a feature are labelled with each of the probable values of the feature. Each leaf of the tree is considered with a class or a probable distribution on the classes. To categorize an example, sieve it down the tree, as follows. For respective feature met in the tree, the arc corresponding to the value of the example for that feature is followed. Below is a decision tree structure for mammogram classification.



3. LITERATURE SURVEY

Learning that are present. In this project, we have used decision trees as a machine learning approach to classify images in one of the classification models defined. Decision tree learning customs a decision tree as per a predictive model, which plots observations about an item to conclusions about the item's

3.1 REVIEW PAPER ON CLASSIFICATION OF MAMMOGRAPHY

Concepts Introduced: The effectiveness of dimension reduction and normal distribution transformation in improving the accuracy of classification has been evaluated. The difference in performance of the SVM classifier and the naïve Bayesian classifier was not statistically significant after the transformation. It can eliminate a dimension that is good for discriminating positive cases from negative cases and this unsupervised dimension reduction algorithm improved the classification accuracy.

3.2 A SURVEY OF COMPUTER-AIDED DETECTION OF BREAST CANCER WITH MAMMOGRAPHY (2016)

Concepts Introduced: The paper mentions a technique of using mass to help classify benign and malignant mass and microcalcifications. It introduces the concept of Content -based Image Retrieval (CBIR) – picking similar images from the database. This is to help the radiologist for more comprehensive understanding. It may be not relevant to the challenge, but it has a nice additional feature to have on board.

3.3 A NEW FEATURE EXTRACTION FRAMEWORK BASED ON WAVELETS FOR BREAST CANCER DIAGNOSIS

Concepts Introduced: A two class classification study to identify normal and cancerous breast tissues is conducted. The rotational and scale invariant features are extracted by HOG, DSIFT and LCP descriptors followed by a classification that utilizes SVM, k-NN, Decision trees via 10-fold cross validation. The same procedure was conducted for a three-class classification study and the accuracy was not sufficiently satisfied. An addition to this study was introduced by applying NLM filter to all the images beforehand. The feature extraction and classification is then performed using the previous method on these new images obtained. The accuracy of this framework was noticeably increased.

3.4 COMPUTER-AIDED DETECTION AND CLASSIFICATION OF MICROCALCIFICATIONS IN MAMMOGRAMS

Concepts Introduced: The paper reviews the various available techniques used for detection of microcalcification, all neatly presented in a table format. The findings mentioned in this paper have been very crucial for our project since micro calcifications in the breasts still serve as a major hindrance to achieving a higher accuracy in the prediction of breast cancer using CADs.

3.5 EFFICIENT BREAST CANCER CLASSIFICATION USING IMPROVED FUZZY COGNITIVE MAPS WITH CSO NN (2016)

Concepts Introduced:

The paper introduces three concepts:

- ➢ Fuzzy Cognitive Map:
- ➢ An unsupervised data classification algorithm
- ➢ CAT Swarm Optimization (CSO):
- ➢ Uses seeking and tracking modes
- ➢ Optimal Brain Damage based Neural Network Pruning: Paper claims it is better than General Neural Network in terms of structure, complexity and performance. It does so by eliminating the poor weights, thereby minimizing the cost.
- ➢ The paper achieves two big feats:
- ➢ An accuracy of 98.3 %,
- ➢ A mean square error value of 0.00234

3.6 COMPUTER-AIDED DETECTION AND CLASSIFICATION OF MICROCALCIFICATIONS IN MAMMOGRAMS: A

SURVEY Concepts Introduced:

A detailed study of various methods for enhancement of microcalcification such as conventional enhancement techniques, contrast stretching, histogram equalization method, convolution mask enhancement, region or feature based enhancement with the evaluation of the algorithms has been presented. Comparison of different segmentation methods such as statistical method, region-based approach, mathematical morphology, multiscale analysis and fuzzy approached along with their advantages and disadvantages have been presented. Study of micro classification detection based on its features or statistical texture feature, detection by template matching, Gray level run length method (GLRLM), Gray level difference method (GLDM), Wavelet based method etc..

A complete amalgamation of results from implementation of algorithms designed for different features from various sources has been tabulated.

4. CONCLUSION

As discussed in this paper smart metering plays a major role in water conservation and adopting water conservation methods becomes very important. By implementing smart metering, we can have better control over the usage of water and thus we can reduce the water consumption and we can also automate the billing process.

REFERENCES

[1] Browne, Anthony, "Cancer bias puts breasts first". The Guardian. London (Oct 2001).

[2] Arnst, Catherine, "A Gender Gap in Cancer". Bloomberg Businessweek. ISSN 00077135 (June,2007).

[3] American Cancer Society: http://www.cancer.org/cancer/breastcancer/detailed guide/breastcancer-key-statistics

[4] Oeffinger KC, Fontham ET, Etzioni R, et al. "Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society". JAMA 2015.

[5] Myers ER, Moorman P, Gierisch JM, et al. "Benefits and harms of breast cancer screening: a systematic review". JAMA 2015.

[6] Fenton JJ, Taplin SH, Carney PA, et al. "Influence of computer-aided detection on performance of screening mammography". N Engl J Med 2007.

[7] Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL; "Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-assisted detection". JAMA Int Med 2015.

[8] Khalid Bashir, Anuj Sharma, "Review Paper on Classification on Mammography", Volume 14 Number 4 – Aug 2014.

[9] Cheng, H.D., et al.: "Computer-Aided Detection and Classification of Microcalcifications in Mammograms: A Survey". Pattern Recognition 36(12), 2967-2991 Yanfeng et al., J Health Med Informat 2016.

[10]D. Thuthi Sarabai, Dr. K. Arthi, "Efficient Breast Cancer Classification Using Improved Fuzzy Cognitive Maps with Csonn", ISSN 0973 -4562 Volume 11, Number 4 (2016).

[11] https://en.wikipedia.org/wiki/Convolutional_neural_n etwork

[12]Browne, Matthew, and Saeed Shiry Ghidary. "Convolutional neural networks for image processing: an application in robot vision." Australasian Joint Conference on Artificial Intelligence. Springer Berlin Heidelberg, 2003.

[13]Frakes, William B., and Ricardo Baeza-Yates. "Information retrieval: data structures and algorithms." (1992).

[14]https://www.tensorflow.org/versions/r0.11/tu torials/deep_cnn/index.html

[15] Graham, Benjamin (2014-12-18). "Fractional Max-Pooling". arXiv:1412.6071 [cs.CV].

[16] https://github.com/fchollet/keras/wiki

[17]http://www.breastcancerindia.net/statistics/st at_global.html

[18]Frakes, William B., and Ricardo Baeza-Yates. "Information retrieval: data structures and algorithms." (1992).