

ARTIFICIAL NEURAL NETWORK BASED CLASSIFICATION FOR MAMMOGRAPHIC MICROCALCIFICATION CLUSTERS

Chandrakala Honnutagi¹, Chidanand S. Kusr²

¹M. Tech Student, Dept. of Computer Science Engineering,

BLDEA's V. P. Dr. P. G. Halakatti College of Engineering & Technology, Vijaypur, Karnataka, India

²Professor, Dept. of Computer Science Engineering,

BLDEA's V. P. Dr. P. G. Halakatti College of Engineering & Technology, Vijaypur, Karnataka, India

Abstract – The presence of microcalcification cluster is a sign of a breast cancer, it is difficult and time consuming for specialists to classify microcalcifications into malignant and benign. In this experiment, a method to classify microcalcification clusters from mammograms and extraction of location of tumor is proposed. Artificial Neural Network with image and data processing techniques was employed to implement an automated microcalcification cluster classification. Operator-assisted classification methods are impractical for large amounts of data and are also non-reproducible. MRI contains a noise caused by operator performance which can lead to serious inaccuracies in classification. Hence, the Probabilistic Neural Network (PNN) was applied for the purposes. Decision making was performed in two stages for classification: feature extraction using the principal component analysis and the PNN and K-Means clustering to extract the location of the tumor. Probabilistic Neural Network gives fast and accurate classification and is a promising tool for classification of the tumors.

Keywords: Breast Cancer, Image processing, MRI, Classification, Mammography, PNN, PCA algorithm, K-Means.

1. INTRODUCTION

Breast cancer is said to be the most common disease in cancer affecting women throughout the world. In Europe, one in six deaths of women are by breast cancer. In U. S., a woman has 12.20% chances of getting breast cancer throughout her life, nearly 1 in 8 women [1]. Mammography is effective and most reliable methods to detect breast cancer in early stage. In many countries, mammography screening programs are conducted based on population for breast cancer detection [2-3]. In a woman's breast tissue, breast calcifications are small calcium deposits which are developed. Microcalcifications are said to be small calcium deposits in breast tissue which appear as small bright spots in mammograms. Microcalcifications are a sign of breast

cancer but not actually result of cancer. The radiological definition of a microcalcification cluster is an area of 1 cm² that contains no fewer than three microcalcifications. The spatial resolution of mammography is very high, so mammography is able to detect microcalcifications at early stage.

1.1 Problem Statement

As lot of data is available in medicine field and clinical documents, it is a challenging task to determine a particular pattern from a huge amount of data. In order to overwhelm these problems, we have designed a system that extracts particular information of our interest from clinical notes that is Medical Resonance Imaging images of breast cancer. In this project, we are interested classification of microcalcification in mammograms. To extract features of the mammographic images, we need sophisticated clinical image processing techniques. There are many and different methods and tools available. Choosing and applying a right method at a right time is a challenging task. So, all such problems are solved by using our proposed method. We have different tools like PCA feature extractor, PNN classifier etc. First we use PCA algorithm to extract the features of the microcalcification and then apply PNN classifier to classify the tumor. Further we use K-means clustering algorithm to locate the tumor in the mammogram.

2. Techniques used for Feature extraction, Classification and Location extraction

In this work there are three techniques to be used. They are as follows.

2.1 Principle Component Analysis (PCA)

It is a path to identify patterns in data, and the data is expressed in such a way that their differences and similarities are highlighted. PCA is of powerful tool that analyses data with high dimensions and luxury of

graphical representation is not available, this data can be hard. The main advantage of PCA is that once the patterns of data has been found, the data can be compressed, compression that is by reducing the number of dimensions, without much loss of information. This technique is used in image compression, as we will see in a later section.

MATHEMATICAL BACKGROUND REQUIRED FOR PCA

- Mean
- Standard Deviation
- Variance
- Covariance
- Covariance matrix.
- Eigen vector
- Eigen value

STEPS TO BE FOLLOWED IN PCA

1. Get some data
2. Subtract the mean
3. Calculate the covariance matrix
4. Calculate the eigenvectors and eigenvalues of the covariance matrix
5. Choosing components and forming a feature vector
6. Deriving the new data set

2.2 Probabilistic Neural Networks (PNN)

In a PNN, the operations are organized into a multilayered feed forward network with four layers:

- Input layer
- Hidden layer
- Pattern layer/Summation layer
- Output layer

PNN is commonly used in problems of classification. When an input is taken, the first layer computes the distance from the input vector to the training input vectors [7-9]. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a complete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and 0 (negative identification) for non-targeted classes.

2.3 K-Means Clustering

K-Means can be thought of as an algorithm relying on hard assignment of information to a given set of partitions. At every pass of the algorithm, each data value is assigned to the nearest partition based upon some similarity parameter such as Euclidean distance of intensity. The partitions are then recalculated based on these hard assignments. With each successive pass, a data value can switch partitions, thus altering the values of the partitions at every pass. In a nearly identical setup, a second K-Means implementation was made using the grayscale intensity of each pixel rather than the RGB color to partition the image. The starting partitions were chosen as equally spaced values from 0 to 255. The number of partitions used, as in the color segmenting setup, was varied in order to increase the number of results available and to study the effects of varying this parameter.

3. Proposed System

Clinical documents contain large amount of data. These data contains essential information like X-rays, MRI images, CT scans etc. In our proposed method, we are only interested to extract essential information like MRI images and tumor from them. To extract the information from mammograms, we are using some different kinds of methods and tools.

The proposed system consists of three main algorithms. The below flow diagram fig. 1, of the system shows how the system works.

The steps involved in classification and location extraction are showed in the below figure. The input given to the system is an MRI mammographic image. This image is taken as input while image acquisition process. The PCA algorithm takes this image as input and processes the image to extract the features. This algorithm calculates the Eigen value and Eigen vector of the image matrix. It forms a feature vector which represents the features of that image. The task of the PCA algorithm terminates after giving Feature vector as the output. This feature vector is taken as input by the PNN classifier to classify the tumor in the mammographic image. The working of PNN starts by taking feature vector of the input image as input. The feature vectors of the trained images are extracted in the next step which is called as the training phase. The feature vector of input image and of the trained images are computed/ tested, which is said to be testing phase. After the computation gets over, the classification of the input image is emitted out. The classification may be normal stage, benign stage, or malignant stage. If the input image

is not mammographic image then the output will be "It is not a mammographic image".

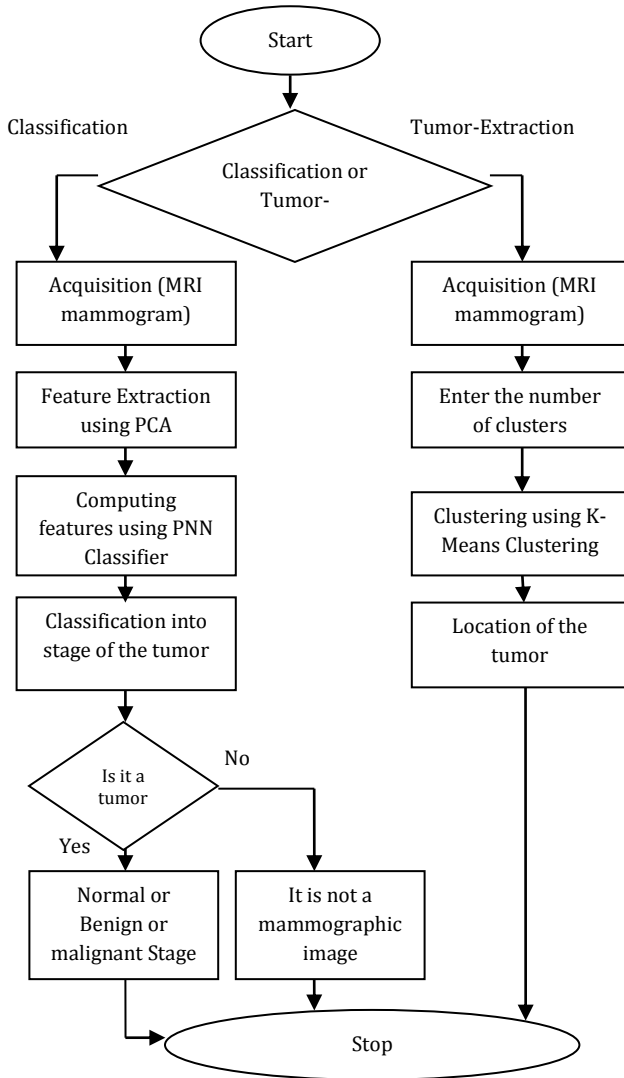


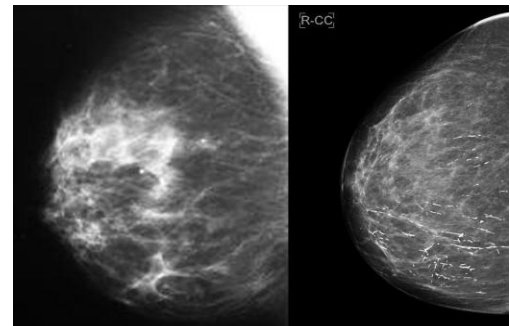
Fig-1: Flow diagram of the proposed system.

To find the location of the tumor in the input image, K-Means clustering algorithm is taken into consideration. The algorithm clusters the images into 4 clusters in our experiment. The location of the tumor will be shown in one of the 4 cluster images.

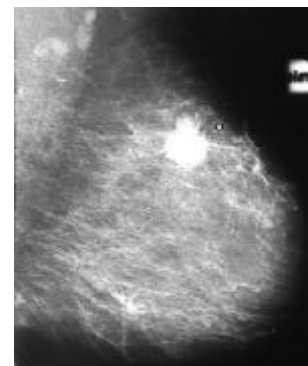
4. Results and Discussion

All experiments in this paper are performed on digitized MRI database provided by Medical Resonance. The proposed method extracts features and classifies the

input mammogram into its respective tumor stage (i.e. normal, malignant, benign stage). Our aim was to classify and locate the tumor so that proper treatment can be provided to the patient. We therefore started from image acquisition further continuing with feature extraction using PCA algorithm. Next the output of this is fed to PNN classifier which computes the features of stored images and classifies into stage of the tumor.



(a)



(b)

Fig-2 (a), (b): Input mammographic images

Feature extraction is done by the PCA algorithm. The features of the image are feature vector containing Eigen value and Eigen vector of that image. After clicking on the feature extraction push button a dialogue box will arrive which says "Features are extracted and stored in the database".

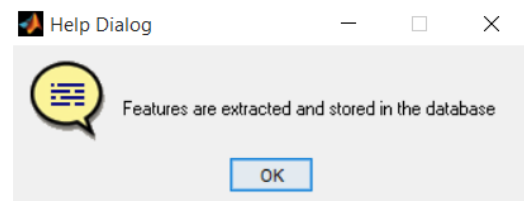


Fig-3: Dialogue box of Feature extraction

After the feature extraction, computing push button is clicked. Here the input image is computed with trained images in the database. After the completion of computation a dialogue box appears along with the image.

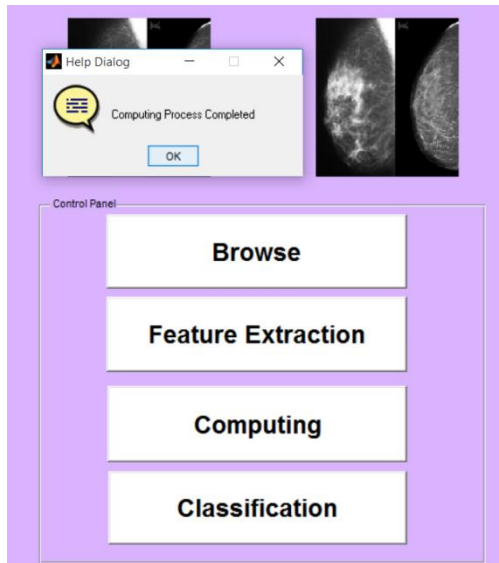
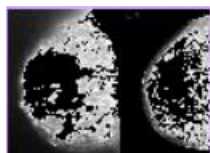


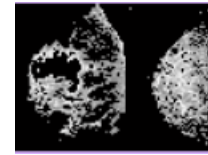
Fig-4: Display after computing

After the computing push button, there is classification push button to be clicked. This push button displays the name of the stage in which the tumor in the input mammographic image is. If the image will be not mammographic image then it will give output as "It is not a mammographic image".

After the classification the location of the tumor is to be extracted. To extract the tumor from the image tumor extraction push button is used. When we click on tumor extraction it opens the file where the input image is present, the image must be selected from the file. The selected image will be displayed in the first cluster. Along with selection we have to enter the number of clusters to be divided into. On typing the number of clusters; the image will be divided and displayed in four clusters that are axes 3, axes 4, axes 5, and axes 6. In the last cluster the location of the tumor will shown clearly with white pixel, it may also be called as the tumor was stated developing from that point.



(a)



(b)



(c)



(d)

Fig-5: Clusters of the location of the input mammographic image (a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4

The clusters show the images in which the location of the tumor is present. The location of the tumor in the sense the starting position of the tumor is extracted. The white pixels in third and fourth clusters indicate the originating position of the tumor. Thus the location of the tumor is extracted using the K-Means clustering algorithm.

5. Conclusion

In this proposed method, we have built a system to classify the tumor and extract the location of the tumor from the mammographic MRI images. The free formed clinical documents contains large amount of data, among which is the data in image form, particularly MRI images. We have designed a system in such a format that, it classifies the tumor into its stages and extracts the location of it. The proposed system contains four important parts: image browser, Feature extractor, tumor classifier, location extractor. We have used PCA algorithm to extract the features of the mammographic MRI images. PNN classifier is used to compute the features of the images and classify the tumor into respective stage. By using the PCA and PNN together the performance has been increased. To extract the location of the tumor, K-Means clustering algorithm is used. This helps in extraction of the position at which the tumor has grown. Thus by this work the effectiveness of the classification along with location extraction of the tumor has been increased.

References

- [1] Zhili Chen and Harry Strange, "Topological Modeling and Classification of Mammographic Microcalcification Clusters", IEEE transactions on biomedical engineering, Vol. 62, No. 4, pp. 1203-1214, 2015
- [2] Karsten Juhl Jorgensen, Per-Henrik Zahl et al, "Breast cancer mortality in organised mammography screening in Denmark: comparative study", BMJ, c1241, doi:10.1136/bmj.c1241, 2010
- [3] Karsten Juhl Jorgensen and Peter C Gotzsche, "Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends", BMJ, b2587, doi:10.1136/bmj.b2587, 2009
- [4] Molly L. Flexman, Michael A. Khalil et al, "Digital optical tomography system for dynamic breast imaging", Journal of Biomedical Optics, Vol. 16, No. 7, 2011
- [5] Kath Moser, Julietta Patnick et al, "Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data", BMJ, b2025, doi:10.1136/bmj.b2025, 2009
- [6] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights", Transportation Research Part C, Elsevier, Vol. 19, pp. 387-399, 2011
- [7] Marius Ene, "Neural network-based approach to discriminate healthy people from those with Parkinson's disease", Annals of the University of Craiova, Vol. 35, pp. 112-116, 2008
- [8] Pantelis Georgiadisa, Dionisis Cavourasb et al, "Improving brain tumor characterization on MRI by probabilistic neural networks and non-linear transformation of textural features", Elsevier Ireland, doi:10.1016/j.cmpb.2007, 2007
- [9] S. Mishra, C. N. Bhende et al, "Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network", IEEE transactions on power delivery, Vol. 23, No. 1, pp. 280-287, 2008