International Journal of Engineering & Technology, 7 (4.6) (2018) 108-111

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Detection of Masses in Mammograms using Bayesian Method and Machine learning

Priyadarshini Chatterjee[1*], Ch.Mamatha[2], T. Jagadeeswari[3], Katha Chandra Shekhar[4]

[1]*Department of Computer Science and Engineering, Vardhaman College of Engineering*
[2]*Department of Information Faculty, Vardhaman College of Engineering*
[3]*Department of Computer Science and Engineering, Vardhaman College of Engineering*
[4]*Department of Computer Science and Engineering MLR Institute of Technology, Hyderabad*
**Corresponding author E-mail: jinipriya@gmail.com*

Abstract:

Every 100th cases in cancer we come across are of breasts cancer cases. It is becoming very common in woman of all ages. Correct detection of these lesions in breast is very important. With less of human intervention, the goal is to do the correct diagnosis. Not all the cases of breast masses are futile. If the cases are not dealt properly, they might create panic amongst people. Human detection without machine intervention is not hundred percent accurate. If machines can be deeply trained, they can do the same work of detection with much more accuracy. Bayesian method has a vast area of application in the field of medical image processing as well as in machine learning. This paper intends to use Bayesian probabilistic in image segmentation as well as in machine learning. Machine learning in image processing means application in pattern recognition. There are various machine learning algorithms that can classify an image at their best. In the proposed system, we will be firstly segment the image using Bayesian method. On the segmented parts of the image, we will be applying machine learning algorithm to diagnose the mass or the growth.

Keywords: *masses in mammograms, Bayesian method, machine learning*.

1. Introduction:

In recent years, machine learning has become one of the key research area's. Machine learning is one of the core research area among all the computer science research areas in which is linked up with artificial intelligence in providing deep learning. In which we come across many things like automatic translations, collaborative filtering, and classification of objects, pattern recognition and many more.

Machine learning can be used in making critical decisions among many areas like medical diagnosis, stock trading etc. Machine learning uses initial techniques like unsupervised learning which used to apply training on data which as to be considered as input as well output in order to predict future results which takes part in clustering. Another is supervised learning I order to find hidden patterns from inputs which is a predictive model based approach used in classification and regression.

Mammograms is functionally used in medical image segmentation process which is used in examine the scan reports taken for identifying breast cancer. Early detection, diagnosis and analysis of best cancer is done using mammography. Image segmentation suing mammograms will be very useful in finding regions of breast cancer for better diagnosis. The analysis of mammograms can be done by double threshold based approach. The outline of the segmented image to the original report or images helps the physicians in finding regions related to breast cancer. Generalization can include both in x-ray related mammograms and bio medical images for image segmentation. Generally mammograms are the images with are generated by mammography in which provides two view of an images. The ease of mammograms with screen is to provide an advantage of lowering the risk over breast cancer in which helps in earlier detection. Mammograms can be used for detection of breast cancer but not all as it depends on many of the factors.

Edged based segmentation approach is one of the most used method in identifying the boundary or mass contour with in doubtful region or the region of interested portion in scanned mammograms. This is applied by doing stretching or contrast enlarging among mammograms by using or applying noise filtering techniques to eliminate noise. Each pixel of region is examined by co-occurrence matrix of pixel in which energy image texture is obtained. Image segmentation is very important technique with respect to medical images and decisions. It is an outstanding approach or advancement in the field of medical science. In few situations it remains unsolved in biological problems. Bayesian classifies acts an effective approach in order it provides many improvements in biological computer or machine vision boundaries which undergo many image processing techniques.

Bayesian methods are general methods used for analyzing the huge amount of data and variety of data. Bayesian methods generally used for mathematical tools which are provides an approach to statistical work. The basic concepts of Bayesian methods can be used for image analysis where this approach can lead to many benefits in interpretation because it concentrates on prior knowledge. Bayesian knowledge mainly revolves throughout the probability of posterior. As we are very familiar with baye's theorem which states that the probability of posterior data

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

is proportional to product of prior probability and livelihood of present work of new data. In many situations there is no sufficient data in order to process the results where data is not enough to process a solution to the problem stated.

In recent years, the area of medical biological imaging which has undergone many comprehensive challenges and advances. Accession of medical images into different effects like 2D and 3D rotations along with higher dimensions are more common in clinical based research applications. In this area we require experience, practice, intuition and deep insight is certainly required for diagnosing clinical imaging in which when compared with recent years there was huge improvement and development among image processing, image acquisition and enhancement techniques over computational areas.

2. Related Work:

An approach towards Bayesian method in medical imaging can be done using image segmentation. Bayesian classifier is well organized which can be used in refining image segmentation. The discipline of medical image encounters many revolutionary advances across past two years. Acquisition of medical images can done in 2D, 3D or any other higher dimensions . the techniques related to image acquisition can be done in 3D ultra sound imaging, magnetic resonance imaging(MRI) , computed tomography(CT) , diffusion weighted imaging(DWI) , functional MRI (fMRI) and many other medical related diagnosis [1].

Breast cancer is world's second common diseases among cancer affected patients after lung cancer which causes many deaths within less span of time among women. In early 1980's mammograms plays an effective role in identifying cancer infected regions. Screening mammograms is used in detection of cancer regions of affected patients who don't have any symptoms which is used for physician to detect in early stages [2][3].

Mammograms screening is functional examined using Image segmentation which acts in an effective manner of identifying cancer related tissues in early stages. In which, image segmentation is an approach related in partitioning the images into component objects and separate parts which also involve with various approaches [4-9]. Classification of particle is one of the important components in multivariate analysis which takes place in statistical methods can be considerably used in extracting data or information from micrographs within the single particles. This can be done by making use of Bayesian gibbs algorithm for sampling considered in classification [10]. Image processing methods which is used in processing and analyzing single particle is the study of unordered macrographs which collaboratively used in combining huge assortment of images of macromolecules in identical [11]. well.

Many clinical scans can be accommodated using data density distributions where the probability of distinctive values can be initiated using fractional components from two or more peculiar components of tissues which can be considered as partial measuring [12]. The spatial correlation provides a contemporary view on data results which results in reducing degree of freedom in an effective number [13]. Where over flattering can be achieved by resulting using renormalized images of histograms [14]. Magnetic resonance image processing is been accommodated using biological tissues. In the view accordingly with traditional computer vision will be considered as relatively simple for any given protocols. Bayesian probability can be applied for any tissues representation in consideration with frequency based approach [15].

Bayesian approaches are used in assessing the degree of atrophy to accommodate solutions for wide range of problems related in freaking diseases in order to predetermine elaboration on few diseases associated with priors, learning and Bayesian risks [16][17]. Image subtraction is one of the frequent tool used in analyzing the pair of images within changes to accommodate interpretation and surveillance of bio medical data [18-21].

An approach for automatic image segmentation with respect to heterogeneous data or information related to huge assortment of images which accommodates the gap between bottoms up affinity image segmentation methods along with top down relative model based scenarios. The Bayesian theory provides formulation in calculating affinities into multiple levels of image segmentation which incorporates with weighted aggregation algorithm [22]. The methods of Bayesian statistics can be applied in analyzing fMRI data. Where the structure related with noise in fMRI location based time sequence is very difficult to understand as the data can be auto correlated as stated in many point of view. The model based approaches like cross correlation analysis and candidate test is commonly used approaches in functional magnetic resonance based imaging applications. Whereas machine learning and artificial intelligence research areas methods are easy to implement and gives very effective results in data analysis. Self organizing mapping based approaches are tremendously adaptive in analyzing bio medical and fMRI data sets [23].

3. Proposed System:

The proposed system firstly aims to segment the entire image into parts and sub-parts using Bayesian Classifier. Bayesian Classifier is selected as it is an efficient tool to partition the image into subparts. The accuracy of using Bayesian Classifier is higher than any other classifier available in the field of image segmentation. Before we apply Bayesian Classifier, we need to suppress the background noises of the image. There are various kinds of filters available to suppress the background noise. Median filters are used in our proposed system to suppress the background noise as a part of preprocessing stage. Firstly the median of the kernel is calculated by sorting every pixel in the kernel. This is the most critical part. It is used to remove multiplicative as well as speckle noise. Median filter is so used that it reduces the variances of intensities in the image. After the noise removal, the second step is enhancement of the image using adaptive histogram equalization. The maximum grey level is computed. Labeling the maximum grey label as 'L'. Then denoting the letter X as number of rows of the image and Y as number of columns of the image to compute [(L-1)/X*Y]. Following this step is the calculation of cumulative value of the pixels. With this cumulative value, the next step is to multiply [(L-1)/X*Y] to find the new scaled value. Then substituting with the new scaled value with the closest brightness value of the pixel to obtain the new scaled value.

The next step as per Figure 1 is to apply Bayesian method to segment the image. The breast masses have variable size and properties. Normal methods of segmentation are not useful for proper segmentation and division of the image into appropriate parts. The Bayesian method helps to find out the pdf of a set of variable. Bayesian model is a directed acyclic graph. The nodes of the graph represent a variable in each domain and the edge between two nodes says about the relationship between the variables. Bayesian method is also called posterior probability and it can be calculated if prior probability is known. The below mentioned equation is used to calculate the posterior probability.

4. $$P\,(Z_l/Y) = \frac{P(Zl)P(Y|Zl)}{\sum_{k=1}^{n} P(ZK)P(Y|ZK)}$$

For each node in the graph there exists a conditional probability. The Bayesian network represents a joint probability distribution. Given any data x, the posterior probability given by the Baye's law is

5. $$P\,(l|\,x) \propto p(x\,|l)\,p(l)$$

The calculations of the priori along with Gaussian priori, is used for reconstruction of the data.

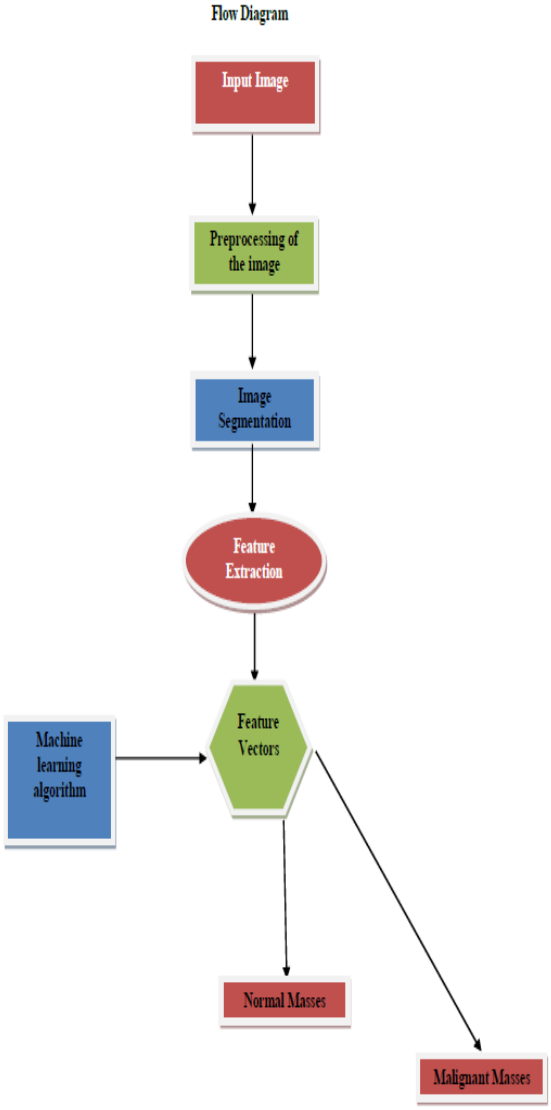


Figure 1: Flow Diagram

After image segmentation, to make the classification more productive, a machine learning algorithm is used. Now the segmented image will be exposed to feature extraction, resulting in feature vectors. These feature vectors will be computed with a machine learning activity, to distinguish between normal and malignant masses. The proposed algorithm consists of two parts i.e maximum posteriori probability generation using Bayesian network and maximum posteriori probability estimation using Markov random field. Using some vital information about the pixels, the pixels will be divided into classes. Bayesian Network is then used to transform the maximum posteriori probability of the pixels to image domain. In the second phase, taking into consideration some knowledge about the image and also considering the distance between individual pixels, Markov random field method is used to do a reasonable segmentation. The segmentation generates the classes consisting pixels of highest probability, second highest probability and third highest probability. The pixels having highest probability are considered to be abnormal. Firstly we us the machine learning algorithm to classify these abnormal pixels as malignant or into another category where the probability of cancerous growth is almost negligible. Then we model the system with set of weights (maximum length of the lesion, water flow inside the tumor, age of the patient). The steps in chronicle order, if followed will help to differentiate the classes are:

1. Labeling: set of images with the correct answer, in our case “correct length”,” shape”,” mass” of the tumor.
2. Teaching: the machine learning algorithm is given the labeled data of step 1.
3. Confirmation: set of input images given during training.
4. Test: a set of images are given to see if the algorithm has discovered some new features during teaching, whether it can apply to improve performance. In our proposed system, we have supplied some images, which are diagnosed as malignant to test our proposed system. The result was accurate.

Our machine learning system follows supervised machine learning algorithm (k-nearest neighbor with k=1) as the training phase is supplied with labeled data.

The shape is given a weight of 5. There are five shapes to be considered at the time of feature extraction, oval, round, not regular, as lobule and distorted.

Another feature the machine learning system considers is the length of the mass. This is given a weight of 5. If the mass is round with no irregularity towards border and as per the correct length specified i.e for our experiment we consider the correct length to be of one inches, the feature is classified as benign. The proposed system is as per Figure 1. All the measures of our proposed system is considered after consulting the radiologist of TATA MEDICAL RESEARCH CENTRE, Kolkata.

Nonetheless, all machine learning algorithms are based on nodes, layers, and an activation function along with some weights. Important part of the algorithm is in identifying the activation function and the weights.

4. Result Analysis

After the step of feature extraction, we categorize the masses into three categories as per table 1.

Table: 1 Results after Feature Extraction

Category	Suggestion	Result
Category0	Negative	The breasts are normal
Category1	Medium abnormal	Probable of being malignant
Category 2	Highly Abnormal	Highly malignant

The dataset consists of 6 patients. These patients have undergone biopsies to confirm whether the lesions were strictly malignant. The proposed system was implemented on these set of data. The image set was divided for training and testing phase. The data set was tested with Naïve Bayes model with Laplace smoothing. All these functions are inbuilt in MATLAB. The computational time and accuracy was noted. The time it took to get the end result was 17 minutes. Then the same data set was analyzed with knn.cv function of MATLAB. The computational time and accuracy was again noted. This time it was found to be 65% more accurate than the previous one. The entire output in the second case was achieved in 7 minutes.

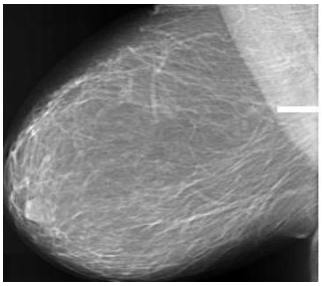


Figure 2: Mammogram after preprocessing

In Figure 2, the lesion is significantly seen after removal of background noise and enhancing the image. Followed this step

is image segmentation and feature extraction. The image looks like Figure 3.

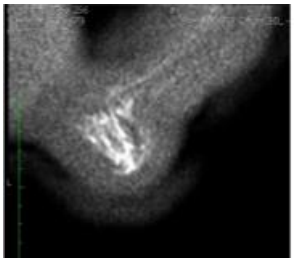


Figure 3: After Segmentation and Feature Extraction

5. Conclusion.

In the experimental phase the results of our proposed system was compared with Naïve Bayes and Laplace smoothing. Our system gave 65% accurate result in 7 minutes and Naïve Bayes took around 17 minutes to generate the output. Bayesian network is a recent finding and shows best result when it is used in case of unknown data or uncertainty. As Bayesian network are represented with graphs, they give a much better relationship at the time of feature extraction.

Though Bayesian outputs are not as useful as inputs to machine leaning algorithms, still we use this approach in our proposed system, keeping in mind the accuracy of Bayesian results. The proposed system is robust and time efficient as the Bayesian approach can do the classification in a lesser time than the other available classifiers in the market.

References:

[1] M.C.Jobin Christ, K.Sasikumar, and R.M.S.Parwathy "Application of Bayesian Method in Medical Image Segmentation" TECHNIA – International Journal of Computing Science and Communication Technologies, VOL. 2, NO. 1, July 2009. (ISSN 0974-3375), pp 353-355.

[2] Benjamin O. Anderson, ―Breast Cancer—Thinking Globally‖, Science, Vol. 343, Issue 6178, pp. 1403, 28 March 2014

[3] National Breast Cancer Foundation (NBCF), Texas, ―Mammogram‖, URL:ttp://www.nationalbreastcancer.org/diagnostic-mammogram.

[4] Alasdair McAndrew, ―An Introduction to Digital Image Processing with Matlab Notes for SCM2511 Image Processing 1‖, School of Computer Science and Mathematics Victoria University of Technology, Ch. 7, pp. 137-141, Semester 1, 2004.

[5] B. K. Gayathri and P. Raajan, ―A Survey of Breast Cancer Detection Based on Image Segmentation Techniques‖, 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, 2016, pp. 1-5. Doi: 10.1109/ICCTIDE.2016.7725345.

[6] Alaa A. Hefnawy, ―An Improved Approach for Breast Cancer Detection in Mammogram based on Watershed Segmentation‖, International Journal of Computer Applications (0975 – 8887), Vol. 75, No.15, August 2013.

[7] Abdul Qayyum and A. Basit, ―Automatic breast segmentation and cancer detection via SVM in mammograms,‖ 2016 International Conference on Emerging Technologies (ICET), Islamabad, 2016, pp. 1- 6. Doi: 10.1109/ICET.2016.7813261.

[8] M. Mustafa, N. A. Omar Rashid and R. Samad, ―BreAst Cancer Segmentation Based On GVF snake,‖ 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, 2014, pp. 928-931. Doi: 10.1109/IECBES.2014.7047647.

[9] R. B. Dubey, S. Bhatia, M. Hanmandlu and S. Vasikarla, ―Breast Cancer Segmentation Using Bacterial Foraging Algorithm,‖ 2013 10th International Conference on Information Technology: New Generations, Las Vegas, NV, 2013, pp. 541-545. Doi: 10.1109/ITNG.2013.88

[10] Montserrat Samso, Michael J. Palumbo, Michael Radermacher, Jun S. Liu,c and Charles E. Lawrencea, A Bayesian method for classification of images from electron micrographs ,August 2001, and in revised form 28 January 2002.

[11] Radermacher,M.,1988.Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J. Electron Microsc. Tech. 9, 359–394.

[12] M. Pokric´, N.A. Thacker, M.L.J. Scott, A. Jackson, Multi-dimensional medical image segmentation with partial voluming, Proc. MIUA, Birmingham (2001) 77–81.

[13] K.J. Friston, P. Jezzard, R. Turner, Analysis of functional MRI timeseries, Hum. Brain Mapp. 1 (1994) 153–171.

[14] M. Pokric´, N.A. Thacker, M.L.J. Scott, A. Jackson, Multi-dimensional medical image segmentation with partial voluming, Proc. MIUA, Birmingham (2001) 77–81.

[15] R. Guillemaud, J.M. Brady, Estimating the bias field of MR images, IEEE Trans. Med. Imag. 16 (3) (1997) 238–251.

[16] H. Jeffreys, Theory of Probability, Oxford University Press, 1939.

[17] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.

[18] D. Murray and A Basu, Motion Tracking with an Active Camera, IEEE Trans. Pattern Analysis and Machine Intell., 16(5), 1994, 449-459.

[19]S. Rowe and A. Blake, Statistical Mosaics for Tracking, Image and Vision Computing, 14(8), 1996, 549-564.

[20] D. Koller, J. Weber and J. Malik, Robust Multiple Car Tracking with Occlusion Reasoning, Proc. ECCV 1994, J-O. Ekhlund (Ed), Stockholm, pp189-196, 1994.

[21] A. Baumberg and D. Hogg, Learning Flexible Models from Image Sequences, Proc. ECCV 1994, J-O. Ekhlund (Ed), Stockholm, pp299-308, 1994.

[22] Jason J. Corso, Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha, and Alan Yuille, Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification, IEEE Transactions on medical imaging, vol. 27, no. 5, May 2008 .

[23] E. Bullmore, M. Brammer, S. C. R. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham, "Statistical methods of estimation and inference for functional MR image analysis," Magn. Reson. Med., vol. 35, pp. 261–277, 1996.