Indian J.Sci.Res. 14 (1): 257-262, 2017

ISSN: 0976-2876 (Print)
ISSN: 2250-0138 (Online)

FEATURE BASED BREAST CANCER DETECTION USING MAMMOGRAPHIC IMAGES

J. SHINY CHRISTOBEL[a1] AND C. VINOTH KUMAR[b]

[a]Post Graduate Student, Department of ECE, SSN College of Engineering, Chennai, Tamil Nadu, India
[b]Assistant Professor, Department of ECE, SSN College of Engineering, Chennai, Tamil Nadu, India

ABSTRACT

After HIV, the second leading cause of death for women all over the world is breast cancer. Cancer is the tremendous growth of cells that originates in the blood tissue. The tumor may be cancerous (malignant) or non cancerous (benign). Early detection increases the chances of survival and reduces the death rate. This paper presents a novel approach to detect the mammogram images using histogram of oriented gradients (HOG) and classified by support vector machines (SVM) which classifies the breast cancer using support vectors. The above method is generally used for textural pattern analysis and human detection. Hence to increase the classification efficiency a pre-processing technique called CLAHE (Contrast Limited Adaptive Histogram Equalization) is proposed. Graph based computation is introduced for the existing methodology and the outcome is studied. This approach reduces the computation time and increases the accuracy.

KEYWORDS: Breast Cancer, Mammogram Images, Histogram of Oriented Gradient, CLAHE, Morphological Operation

The processing of images by mathematical operations by using any form of signal processing is called image processing. The input can be an image or a series of images such as a photograph or frame. The output of the processing technique may be an image, a set of characteristics, features or the parameters related to the input images or input videos. The image can be a two-dimensional signal or a three-dimensional signal. The input images are processed by applying standard signal-processing techniques to it for detection. The acquisition of images is referred to as imaging.

After HIV, the second leading cause of death for women is breast cancer. All over the world, more than 400 thousand women are affected by breast cancer. Cheng et al., 2010, have analyzed that the breast cancer is type of cancer that develops from the breast tissue. The cancer is an erratic growth of cells that originates in the blood tissue which has three stages: normal, benign and malignant. The benign is less critical and the malignant is the most advanced stage of cancer [Cahoon et.al., 2011]. Early detection reduces the death rate and increases the chance of surviving.

The mammogram image is the X-ray image of the cancer affected breast of the women. Vidhya et al., 2011, describes the pre-processing techniques and its benefits. Since the X-Ray of the human body contains unwanted distortions and noise so it has to be enhanced to acquire the desired result. The mammogram images are pre-processed to remove noise, accentuate or sharpens the image features such as edges, boundaries or contrast. For a specific application, the image is processed so that the result is more suitable than the original image which is the principal objective of image enhancement. The enhancement does not change the information content of data but it increases the changing range of features chosen. Image enhancement methods are based on spatial domain or frequency domain techniques.

The features from the enhanced images are extracted using texture analysis. Texture analysis is the characterization or segmentation of regions in an image by the texture content of an image. It quantifies intuitive qualities such as roughness, smooth, silky or bumpy. The texture content refers to the function of spatial variation in pixel intensities. In this sense, the variations in the intensity values or gray levels are due to the roughness or bumpiness.

After the texture analysis, the images are classified using a classifier. Now-a-days, the usages of classifiers are increasing rapidly in medical diagnosis. There are various classifiers such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naive Baye's (NB) classifier, KNN classifier so on. Generally the classifiers detect the images as normal or abnormal. In this project SVM classifier is used for classification. SVM undergoes two processes for classifications: testing and training. The classifiers are trained using certain values and with the help of those values the images are tested. The accuracy, sensitivity, specificity and Precision-Predicted Value (PPV) are calculated.

The objective of this paper is to perform feature based breast cancer detection from mammographic images by to extracting the features using HOG (Histogram of Oriented Gradients). The extracted features are classified using SVM and also to obtain high accuracy with less computation time in detecting tumors.

[1]**Corresponding author**

The paper is organized as follows: the section 2 deals with the methodology of the cancer detection which includes the mammogram image, its histogram, two pre-processing techniques, texture analysis using HOG and the classification by SVM classifier. The simulation and the performance measures are discussed in section 3 and section 4 followed by the conclusion.

METHODOLOGY

For this paper, the mammogram images are taken from MIAS (mammographic image analysis society) database which is widely used database for breast cancer. The mammogram is an X-Ray of the cancer affected breast. The histograms of the both normal and the cancer input images are obtained. The work flow of the breast cancer detection is shown in Fig.1.

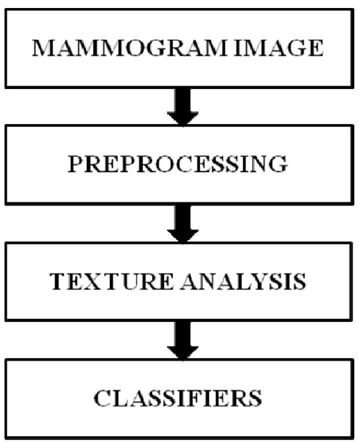

Figure 1: Work Flow of the Breast Cancer Detection

Mammogram Image

Mammogram is the most efficient method for detecting breast cancer. Mammograms play a very important role in early detection of breast cancer and help to decrease the death rate of women. The mammogram image is the X-ray of the affected breast. An X-ray of the breast is scanned with an appropriate device by compressing and flattening the breast.

Pre-Processing

Since the mammogram image contains noise and unwanted distortions it has to be pre-processed. The pre-processing is an essential step to filter the noise and unwanted distortions from the image and to improve the qualities like brightness, sharpness and contrast in images. The pre-processing techniques used in this paper are dilation and CLAHE method.

The dilation is one of the morphological operations that add pixel to the boundaries. It uses a structuring element to process the image [Shameena and Jabbar, 2014]. The structuring element is a binary matrix in which the pixel with value 1 defines the neighborhood. The structuring element is part of the dilation and erosion operations. It is used to probe the input image and can posses any arbitrary shape and size.

After dilation, the image is enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) method. The CLAHE is used because the adaptive histogram equalization (AHE) amplifies the noise present in the image.AHE is the transformation function derived from the histogram of the image. It works well when the distribution of the pixel value is similar throughout the image but the image can contains both darker and brighter regions. The contrast in the image is not enhanced properly. So CLAHE method is used to improve the quality of an image. The CLAHE operates even on the smaller regions called tiles in an image rather than the entire image. The value at which the histogram gets clipped is called clip limit.By using CLAHE, the amplification of noise present in an image is avoided. The resultant image is brighter than the original input image.

Texture Analysis by HOG

Textures are one of the important parameter in the detection of medical images. It identifies the objects and the region of interest (ROI) for various kinds of images [Hemanth and anitha, 2012]. It is important for computer image analysis for classification, detection and segmentation based on intensity and color [Khoshnevisan et.al., 2009 and Hall et.al., 1971]. The texture analysis is the process of extracting features from the enhanced image by a specific method. Image texture is the entity that consists of mutually related pixels or the group of pixels. The information about the arrangement of color or the intensities of the image or the selected region of the image is given by texture analysis. The features like mean, variance, standard deviation, entropy, skewness and kurtosis are extracted using HOG.

The histogram of oriented gradients (HOG) is a global descriptor. Based on the distribution of intensity gradients or edge directions, the shape and appearance within an image are described using histogram of oriented gradients. The intensity is the numerical value of the pixel or the amount of light. The gradient is the directional change in the intensity or the colour of an image and it is used to extract the important information from an image. The image is divided into small regions called cells with 50% overlap and for each cell the histogram of gradient is compiled. The descriptor is the concatenation of these compiled histograms. The cells are collected to form larger regions called blocks and for each block, the intensity measure is calculated. The cells within the blocks are normalized using contrast normalization method by the intensity value. Then the features are extracted from the enhanced image.

The HOG algorithms involve the following steps: the first step is the computation of gradient by Gaussian smoothing followed by discrete derivative masks. The next step is the orientation binning and it is the fundamental nonlinearity of the descriptor [Dalal and Triggs, 2005]. The third step is casting a vote for a pixel. Each pixel has a weighted vote for an edge orientation histogram channel which the value is based on orientation of gradient. The votes are collected into orientation bins over the local spatial regions. The cells can be either rectangular or radial.

The orientation bins are evenly spaced over 0 to 180 degree ("unsigned" gradient) or 0 to 360 degree ("signed" gradient). To reduce aliasing, votes are bilinearly interpolated between the neighboring bins. The interpolation is done in both orientation and position. After the normalized cell responses the final vector from all the blocks in the detection window forms the final descriptor. Thus the features from the cancer and normal images are extracted.

Support Vector Machine (SVM)

In machine learning, support vector machines (SVM) have associated learning algorithms with the supervised learning models. The SVM classifier is used to analyze data, recognize patterns, classification of features and analysis. It is a successful statistical learning method for classification. They rely on support vectors (SV) to categorize the decision between two different classes and it has a hyper plane to categorize the two different classes [Akay, 2007]. The SVM undergoes two processes: training and testing. The training algorithm builds a model that gathers the features extracted from one class into one category and other class in other category. The decision makes SVM as an efficient binary classifier. The extracted features from both normal and cancer images are given as an input to the SVM classifier which detects the mammogram images to be cancerous or normal images.

SIMULATION AND PERFORMANCE MEASURES

Simulation Results

The input image is the mammogram images taken from MIAS database. The input image can be normal and cancerous. The normal and the cancer mammogram image are shown in Fig.2 and Fig.3 respectively.

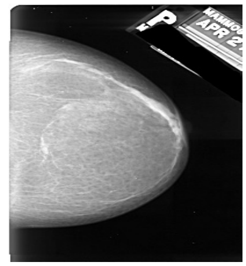

Fig. 2 Normal Image

Fig. 3 Cancer Image

The histogram is obtained for the cancerous image is shown in Fig.4. The histogram is the graphical representation of number of pixels in an image. In histogram, the X-axis represents the tonal variations and Y-axis represents the number of pixels. In the X-axis, the left side denotes the bright black and dark areas, the middle denotes medium gray and right side denotes light and pure white areas.

The output of dilation is shown in Fig.5. The dilation operation usually uses a structuring element for expanding the shapes contained in the input image. The bright regions within the dark regions grow in size and dark regions within the bright regions shrink in size. The small dark spots in images disappear and small bright spots become larger spots.

The enhanced image obtained by CLAHE method in which the quality of the image is increased. The noise in the image is suppressed. The enhanced image is shown in Fig.6.The enhanced histogram is shown in Fig.7. The histogram is enhanced by equalizing the input histogram. The number of pixels is also reduced than the original histogram.

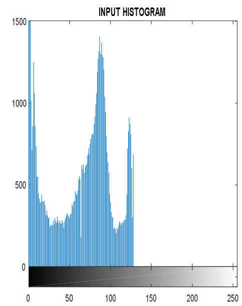

Fig.4 Input Histogram

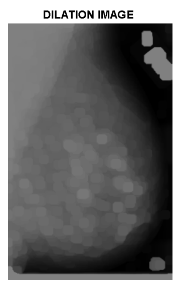

Fig.5 Dilation image

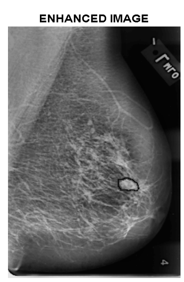

Fig.6 Enhanced Image

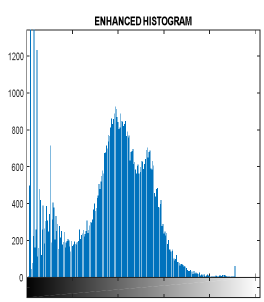

Fig.7 Enhanced Histogram

The HOG image is shown in Fig.8. The Orientation of each cell is obtained. The shape of the cancerous image is visualized with specific directions. The appearance and shape of the cancerous image is visualized by intensity gradients or edge directions.

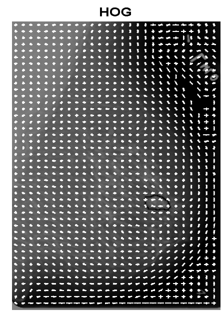

Fig.8 HOG Image

The extracted features like mean, variance, standard deviation, entropy, kurtosis and skewness are extracted from the image and the values are given to the SVM classifier. The SVM classifier is used to classify the input mammogram images as normal or cancerous. The SVM is trained with the extracted features and using the trained values, SVM can detect the normal images and the cancerous images by testing the features.

From the extracted features, the SVM is trained with 80 normal and cancer images. The last 20 normal and cancer images are tested. The images are grouped by representing the cancer images as 1 and the normal images as 0. Among 20 normal images the SVM has detected 17 normal images correctly and among 20 images, 19 images are detected correctly as cancer. The Table 1 shows the classifier output for HOG with enhancement.

Table 1: SVM output

SVM	Training	Testing	Detected
Normal	80	20	17
Cancer	80	20	19

From the decision given by the SVM, number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are calculated.

(i) True Positive (TP): The cancer input is detected as cancer by the SVM classifier.
(ii) True Negative (TN): The cancer image is detected as a normal and labeled as a healthy person by the classifier.
(iii) False Positive (FP): The normal image is detected as a normal and labeled as a healthy person by classifier.
(iv) False Negative (FN): The normal image is detected as a cancer by the SVM classifier.

Performance Measures

To identify that the proposed method detects to be accurate, three parameters are calculated. They are accuracy, sensitivity, specificity and Positive-precision value (PPV). The Accuracy, sensitivity, specificity and PPV are calculated by the below formulae,

Accuracy (%) = [(TP+FP)/ (TP+TN+FP+FN)]

Sensitivity (%) = [TP/ (TP+FN)]

Specificity (%) = [TN/ (FP+TN)]

PPV (%) = [(TP/ (TP+FP)]

The Accuracy, sensitivity, specificity and PPV are calculated using their specific formulae and the values are tabulated in Table 2. The accuracy is 90% and it shows that the HOG using SVM gives better detection. The computation time is also less.

Table 2: Performance Measures (in %)

Accuracy	Sensitivity	Specificity	PPV
90	86.36	5.55	52.77

CONCLUSION

In this paper, an existing and a very popular method called histogram of oriented gradient is used to extract the features from the mammogram image. The mammogram image is pre-processed by CLAHE method and dilation to enhance the image and the quality of the image is improved. After enhancing, the features are extracted using HOG. The locally normalized histogram of gradient orientation features gives very good results for cancer detection, reducing false positive rates. The feature values extracted are trained by the SVM classifier to classify and detect the cancer images with more accuracy and less computation time. To improve the accuracy in future, a filter called non-local means (NLM) which is the most commonly used filter in biomedical images. It can be designed and implemented with the appropriate values. The Non-Local Means (NLM) filter is first applied to mammographic patches. The filter is chosen to enhance the degraded mammographic patches and the feature extraction achieves a robust classification. This may give more accuracy compared to the existing HOG method. By tuning the filter design or enhancing techniques or the features extraction, the accuracy may change in accordance.

REFERENCES

Cheng H.D., Shan J., Ju T., Guo Y. and Zhang L., 2010. “Automated Breast Cancer Detection And Classification Using Ultra Sound Images A Survey”, Pattern Recognition, **43**:299-317.

Cahoon T.C., Sutton M.A. and Bezdek J.C., 2011. “Breast Cancer Detection Using Image Processing Technique”, Institute of Electrical and Electronics Engineers (IEEE), International Conference on Fuzzy Systems, **2**:973-976.

Vidhya M., Sangeetha N., Vimalkumar M.N. and Helenprabha K., 2011. “Early Stage Detection of Cancer in Mammogram Using Statistical Feature Extraction”, 2011 International Conference on Recent Advancements in Electrical, Electronics and Control Engineering, pp.401-404.

Shameena N. and Jabbar R., 2014. “A Study of Pre-Processing and Segmentation Techniques on Cardiac Medical Images”,International Journal of Engineering Research & Technology (IJERT), **3**(4):2278-0181.

Hemanth D.J. and Anitha J., 2012. “Image Pre-Processing and Feature Extraction Techniques for Magnetic Resonance Brain Image Analysis”, Communications in Computer and Information Science, FGCN, pp.349-356.

Khoshnevisan M., Bhattacharaya S. and Smarandache F., 2009. “A Proposed Artificial Neural Network Classifier to Identify Tumour Metastates”, pp.1-14.

Hall E.L., Kruger R.P., Dwyer S.J., Hall D.L., Mclaren R.W. and Lodwick G.S., 1971. “A Survey of Pre-Processing and Feature Extraction Techniques for Radiographic Images”, Institute of Electrical and Electronics Engineers (IEEE) Transactions On Computers, **20**(9):1032-1044.

Dalal N. and Triggs B., 2005. “Histograms of Oriented Gradients for Human’, Proceedings of the 2005 , Institute of Electrical and Electronics Engineers (IEEE) Computer Society Conference on Computer Vision and Pattern Recognition, pp.1-8.

Akay M.F., 2007. “Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis” Expert Systems with Applications 36, pp.3240-3247.

Sahakyan A. and Sarukhanyan H., 2012. "Segmentation of the Breast Region in Digital Mammograms and Detection of Masses", (IJACSA) International Journal of Advanced Computer Science and Application, **3**(2):1-4.

American Cancer Society, 2007. "Cancer Facts & Figures 2007", Archived From The Original on 10 April 2007.

Acroix M., 2006. "Significance Detection And Markers Of Disseminated Breast Cancer Cells", Endocrine-Related Cancer (Bioscientifica), **13**(4):1033–67.

Chaurasia O.P., 2012. "An Approach to Fingerprint Image Pre-Processing", International Journal of Image Graphics and Signal Processing, **6**:29-35.

Vithlani P., 2014. "Pre-processing Techniques in Character Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, **4**(11):601-604.

Ahmadvand A. and Daliri M.R., "A Review on Texture Analysis Methods in Biomedical Image Processing", OMICS J Radiol., **5**(2):1-2.

URL-http://www.mammoimage.org/databases/,MIAS database.