Advances and Applications in Mathematical Sciences
Volume 18, Issue 9, July 2019, Pages 901-908

CLASSIFICATION OF MASSES AS MALIGNANT OR BENIGN USING SUPPORT VECTOR MACHINE

A. M. SOLANKE, R. MANJUNATH and D. V. JADHAV

Department of Electrical and Electronics Engineering
JAIN (Deemed-to-be-University)
Bangalore, Karnataka, 560069, India

Joint Director, Technical Education
Amaravati, Maharashtra, 444601, India
E-mail: anjali.solanke1@gmail.com
manju_r_99@yahoo.com
dvjadhao@yahoo.com

Abstract

Cancer is leading cause of death. There are several types of cancers such as lung, breast, rectal, stomach and skin. According to World Health Organisation breast cancer is the main cause of death among women worldwide. Awareness of the disease, availing medical facilities for treatment and accurate diagnosis may save the lives of women. Among all screening techniques, mammography is most recommended technique by doctors and radiologists. Reading and analysis of mammogram is important part of treatment. Computer aided detection (CAD) techniques are used as a helping assistant for analysing mammograms. Mammograms are analysed for detection of calcifications, masses, architectural distortion and bilateral asymmetry. In this paper, mammogram is classified as normal or abnormal. Further abnormal mammogram is analysed for mass detection. Mass is classified as benign mass or malignant mass using Support Vector Machine classifier.

1. Introduction

According to World Health Organisation (WHO) cancer cases are increasing every year. Cancer cases and death statistics is shown in Table1. New cancer cases of Breast and lung are 11.6% which is more as compared to other cancer cases such as rectal and stomach in 2018. Number of deaths due to breast cancer are 6.6% worldwide [1]. Deaths

2010 Mathematics Subject Classification: 68U10, 62H35, 68U07.
Keywords: Breast cancer, Computer Aided Detection, Mass, Support Vector Machine.
Received January 4, 2019; Accepted January 25, 2019

due to breast cancer can be reduced further by awareness and correct diagnosis. There are variety of imaging modalities like ultrasound, mammography, Magnetic Resonance Imaging and thermograph [2]. Mammography is widely recommended by doctors for screening of breast at the age of forty and above [3]. According to Breast Imaging Reporting and Data System (BIRADS), masses micro calcifications, architectural distortion and bilateral asymmetry are indicators of cancer on mammogram. Masses are of tumour like structure and micro calcifications are small spots of calcium. Micro calcifications are tiny spots where as area of mass is large than micro calcifications [4]. Cancerous mass is called as malignant and non cancerous mass is called as benign. Mass detection is based on their characteristics such as size, shape and boundary. Generally benign masses are circle shaped and malignant masses are having irregular shape with rough boundary. Masses may hide in dense tissues which cannot be visualised by human eye. To improve the detection rates CAD techniques are used. Image processing is used in the CAD techniques for various steps involved in it [5].

Table 1. Cancer statistics of year 2018[6].

Type of cancer	Cancer cases (Lakh)	Deaths due to cancer (Lakh)
Lung	20.9	176,61,007
Breast	20.9	6,27,000
Rectal	18	8,62,000
Stomach	10.3	7,83,000

1.1 Related work:

Detection of mass is based on variation in characteristics of benign and malignant masses. Shape of mass is one of the key features for mass detection. Some of the shape features considered to distinguish between benign and malignant mass are major axis length, perimeter, eccentricity, solidness, orientation, bounding box, extent, convex area, filled area, pixel list and Euler number, continuity, curvature, irregularity, eccentricity, circularity and circular density [7, 8, 9, 10]. Geometrical shape features such as compactness, Normalized Distance Moment, compactness, Normalized Radial Length, Fourier Features and relative gradient orientation based features are used for classification of benign and malignant masses. Normalized distance moment is zero for smooth circular shape and increases with roughness. Compactness gives value zero for circle and one for irregularity [11]. Another approach for shape determination is calculation of luminance variation from border towards centre [12]. Further boundary of region of interest is considered for mass detection with the features such as mean of kurtosis, standard deviation, mean of entropy, standard deviation entropy of wavelets and Gradient texture features [13]. Along with shape and boundary of masses texture features are instrumental in detection of malignant masses. Texture features are broadly classified as statistical, structural and spectral. Statistical texture features are computed by Gray Level Co occurrence Matrix (GLCM) [14]. To improve the accuracy of detection combined GLCM and optical density features are extracted [15]. BIRADS classification is referred by radiologists for detection of breast cancer. Masses are classified in 2, 3, 4 and 5 class as per BIRADS classification scheme on the basis of textural features. BIRADS 2 denotes that mass is 100% benign. BIRADS 3 stands for abnormality but probably benign, BIRADS 4 stands for biopsy recommended (10 to 50% malignant) and BIRADS 5 stands for 98% malignancy [16]. Further abnormality patient age, assessment rank and subtlety value are selected as features along with shape, margin and density features for detection of masses [17]. In our proposed work we have developed user friendly graphic user interface based mass detection system. The system initially classifies input mammogram as normal or abnormal. Further mass is detected on abnormal mammogram and classified as malignant or benign mass.

2. Proposed Work

Mass and micro calcifications are the two prominent signs of breast cancer. Mass detection is challenging for dense breasts. Mass detection is carried out by pre-processing, segmentation, feature extraction and classification as shown in figure1.

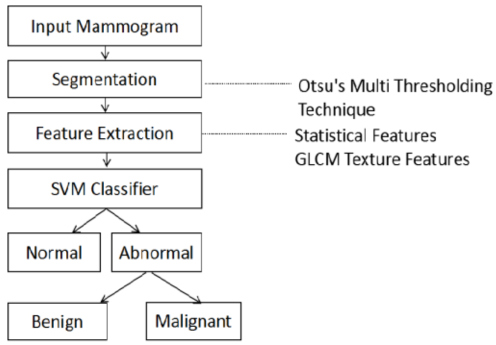

Figure 1. Flow of mass detection using CAD.

Mammography is a low x ray dose image with poor contrast, noise and labels. In pre-processing stage input image is converted to gray scale and unnecessary parts such as labels and noise on mammogram is removed. In this paper histogram equalization is used for image enhancement. Histogram equalization distributes gray levels uniformly throughout the image. Histogram equalization is given by equation 1.

$$S_k = T(r_k), \tag{1}$$

Where S_k is equalized processed image, T is transform, r is input image and k is intensity range. Boundary extraction is given by equation 2. Pre-processed image A is first eroded with structuring element B. Eroded image is subtracted from image A [18].

$$\delta(A) = A - (A - B). \tag{2}$$

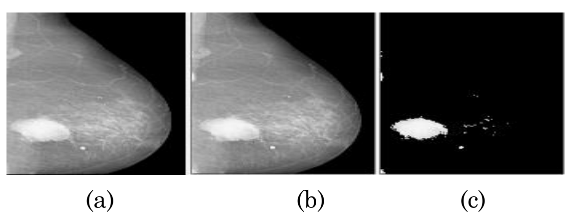

Figure 2. Mass detection steps using CAD (a) Input image (b) Pre-processed image (c) Segmented image.

Mass intensity values are different than normal breast tissues therefore thres holding technique is used for segmentation [19]. Segmentation is carried out by using multi threshold technique and resulted image is as shown in figure 2c). GLCM, Statistical and DWT features are extracted from segmented image. L. D. is low pass and H. D is high pass filter in Discrete Wavelet Transform. Columns are down sampled first and then rows are down sampled. Decomposition of DWT is as shown in figure 3.

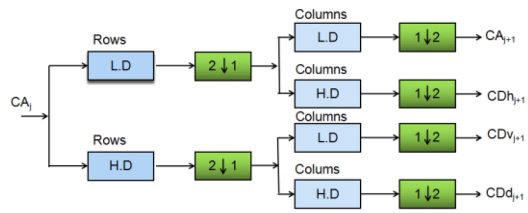

Figure 3. Decomposition of Discrete Wavelet Transform.

DWT decomposition is resulted into four matrix coefficients. CA is approximation coefficient matrix where as CD_h, CD_v and CD_d are horizontal, vertical and diagonal coefficients matrices. CA contains maximum information which is further decomposed by db4 and decomposition is carried out till third level [20]. Principal Component Analysis (PCA) is used for dimensionality reduction of resulted matrix coefficients of DWT. There are fourteen GLCM features determined by Haralick. Here only four features are extracted as texture features such as contrast, correlation, homogeneity and energy [21]. Other nine extracted features are mean, standard deviation, variance, entropy, root mean square, smoothness, skewness and inverse difference moment. Mass is classified as benign or malignant based on these extracted features. There are many classifiers available such as support Vector Machine (SVM), K-Nearest Neighbour (KNN), decision tree and random forest. Here SVM is used for classification of masses. SVM analyses large classification data. Data is separated in two classes by creating hyper plane in high dimensional space. Margin between the classes must be large so that generalization error will be less. The performance of SVM largely depends on the kernel and discriminant function which is given by equation 3.

$$g(X) = \sum_{i=1}^{L_s} \alpha_i d_i K(X_i, X) + \alpha. \tag{3}$$

Support vectors are denoted by x_i, total number of support vectors is L_s, and Class is indicated by d_i [22]. Mass detection is evaluated by Accuracy [23] which is represented by equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Where *TP*, *TN*, *FP*, *FN* stands for true positive, true negative, false positive and false negative respectively. SVM classifies the mammogram as normal or abnormal. Abnormal mammogram is analysed for mass detection. Mass is detected and classified as benign or malignant mass with 80% accuracy.

Conclusion

Database consists of total 25 mammogram images. SVM classifier is used for classification of mass as benign or malignant with the help of thirteen features. Graphic User Interface is designed for user friendly access and proper visualisation of each step of proposed algorithm. Proposed system achieved 80% accuracy. In future work large database will be used for better accuracy.

Acknowledgement

The authors thank SKN hospital Pune for providing database.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal, Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, C A Cancer J. Clin. (2018), 394-424.

[2] L. Wang, Review Early Diagnosis of Breast Cancer, Sensors (Basel) Vol.17 (2017), 1-20.

[3] S. K. Moore, Better Breast Cancer Detection, Journal of Medical Imaging, IEEE spectrum (2001), 50-54.

[4] A. Mencattini, Salmeri M, Lojacono R, Frigerio M and F. Caselli, Mammographic image enhancement and de-noising for breast cancer detection using dyadic wavelet processing, IEEE Transaction on Instrumentation Meas. 57(7) (2008), 1422-1429.

[5] M. Niemeijer, M. Loog, M. David A., M. A. Viergever, M. Prokop and B. Ginneken, On Combining Computer-Aided Detection Systems, IEEE Transactions on medical imaging, Vol. 30 (2011), 215-223.

[6] https://www.who.int/

[7] C. Balleyguier, S. Ayadi, K. V. Nguyen, D. Vanel and C. Dromain, Robert, BIRADS classification in mammography, European Jour. of Radiology 61 (2007), 192-194.

[8] W. Borges, Sampaio, E. M. Diniz, A. Correa, Silv, A. C. Paiva and M. Gattass: Detection of masses in mammogram images using CNN, geostatistic functions and SVM, journal of Computers in Biology and Medicine, Science Direct, (2011), 653-664.

[9] R. D. Ghongade and D. G. Wakde, Computer-aided Diagnosis System for Breast Cancer Using RF Classifier, Proceedings of IEEE WiSPNET, conference, (2017), 1068-1072.

[10] A. Makandar and B. Halalli, Classification of Mass Type Based on Segmentation Techniques with Support Vector Machine Model for Diagnosis of Breast cancer, International Conference on Data Management, Analytics and Innovation (ICDMAI), (2017), 81-86.

[11] L. Xiaoming and Jinshan Tang, Mass Classification in Mammograms Using Selected Geometry and texture Features, and a New SVM-Based Feature Selection Method, IEEE systems journal 8(3) (2014), 910-920.

[12] A. Mencattini, G. Rabottino, M. Salmeri and R. Lojacono, Assessment of a Breast Mass Identification Procedure Using an Iris Detector, IEEE Transactions on Instrumentation and Measurement 59 (2010), 2505-2512.

[13] N. R. Mudigonda, R. M. Rangayyan and J. E. Leo Desautels, Gradient and Texture Analysis or the Classification of Mammographic Masses, IEEE Transactions on medical imaging 19(10), (2000), 1032-1043.

[14] I. Christoyianni, E. Dermatus and G. Kokkinakis, Fast detection of masses in computer aided Mammography, IEEE signal processing magazine, (2000), 54-64.

[15] S. Tai, Zih-Siou Chen, Wei-Ting Tsai, Chin-Peng Lin and Li-li Cheng, A Mass Detection System in Mammograms Using Grey Level Co-occurrence Matrix and Optical Density Features, Advances in Intelligent Systems & Applications, SIST 21 (2013), 369-376.

[16] F. Chokri and M. H. Farida, Mammographic mass classification according to Bi-RADS lexicon, IET Comput. Vis. 11(3) (2017), 189-198.

[17] P. Mc Leod and B. Verma,Variable Hidden Neuron Ensemble for Mass Classification in Digital Mammograms, IEEE Computational Intelligence Magazine 8(1) (2013), 68-76.

[18] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Upper Saddle River, New Jersey, Prentice Hall, (2002).

[19] J. Bozek, M. Mushra, K. Delac and M. Grgic, A Survey of Image Processing Algorithms in Digital Mammography, Journal of Recent Advances in Multimedia Signal Processing and Communications. Vol. 231, Springer, (2009), 631-652.

[20] S. Lahamage and H. Patil, Breast Cancer Classification with Statistical Features of Wavelet Coefficient of Mammograms, IJSCERT 3(6) (2014), 920-926.

[21] R. M. Haralick, K. Shanmugam and I. D. Dinstein, Textural features for image classification, IEEE Trans. Syst, Man, Cybern 6 (1973), 610-621.

[22] D. Shrivastav and L. Bhambhu, Data classification using support vector Machine, Journal of Theoretical and Applied Information Technology 12, (2010), 1-7.

[23] R. M. Rangayyan, F. J. Ayres and J. E. Leo Desautels, A Review of Computer Aided Diagnosis of Breast Cancer: Toward the Detection Subtle Signs, Journal of Franklin Institute, (2007), 312-348.