Advances and Applications in Mathematical Sciences
Volume 20, Issue 11, September 2021, Pages 2811-2817

A COMPARATIVE ANALYSIS OF FEATURE SELECTION APPROACHES IN DIGITAL MAMMOGRAMS

L. KANYA KUMARI and B. N. JAGADESH

Research Scholar
Department of Computer
Science and Engineering
K L University, Guntur
Andhra Pradesh, India
E-mail: kanyabtech@yahoo.com

Professor
Department of Computer
Science and Engineering
Srinivasa Institute of Engineering
and Technology, Amalapuram
Andhra Pradesh, India
E-mail: nagajagadesh@gmail.com

Abstract

Breast cancer is the main cause of cancer death among the women in the world. Mammogram screening is the best technique to detect the Breast Cancer (BC) in early stages as it is low dosage, most effective and less cost. There are many Computer Aided Diagnosis (CAD) systems are available to detect BC which helps radiologists and physicians to detect the tumor cells. The general steps used by most of CAD systems are like preprocessing, feature extraction, feature selection and classification. This paper gives the description of the feature extraction and feature selection techniques that are used by different researchers which may help the researchers.

I. Introduction

Most of the women in the world are suffering with Breast Cancer (BC). According to American Institute for Cancer Research, there were 2 million new cases in 2018 [1] and suggested to go for screening once in a year after 40

2010 Mathematics Subject Classification: 65D18.
Keywords: Computer Aided Diagnosis, mammogram, breast cancer, feature selection.
Received October 11, 2020; Accepted November 4, 2020

years of age [2]. Women breast consists of lobules, ducts, nipples and fatty tissues. Normally tumors will grow inside the lobules as well as ducts and later form cancer inside the breast [3]. These cancer cells also spread to other parts of the body. The breast tumors are of broad categories such as benign and malignant. Benign tumor cells are non cancerous cells and are not dangerous to life. Malignant cells are life- threatening cells and proper treatment may save the people life. There are number of imaging modalities to detect the cancer cells such as X-ray, Computer Tomography (CT) and Magnetic Resonance Imaging (MRI). Among these modalities, X-ray is the low dose imaging, effective and lost cost imaging technique. These X-ray images are called mammograms [4]. Computer Aided Diagnosis (CAD) systems are available for mammogram classification. The general steps for the CAD systems are preprocessing, feature extraction, feature selection and classification. For preprocessing and feature extraction steps, image processing techniques are used. To select the optimal features, meta-heuristic algorithms are used and machine learning techniques are used for classification. In this paper we explained different feature extraction and feature selection techniques available and applied by different researchers which are playing an important an important role in breast cancer detection.

II. Feature Extraction

Feature extraction is the important step in classifying the mammogram images into normal Vs abnormal and benign Vs Malignant. There are mainly two types of features called local and global features. Texture, detector and statistical features are most likely used local features. Texture features represents low-level feature information in an image, which gives more information that will help to extract the better features [4]. They also provide dimensional and structural information of intensity and color of the image. Feature detector provides information whether a particular feature is available in the image. Structural feature provide the information about the structure of the feature and orientation such as area, convex hull and centroid. Mean, median and standard deviation are also giving some more information on the database and their distribution. These features are called statistical features. The hierarchy of image features is represented in Figure 1.

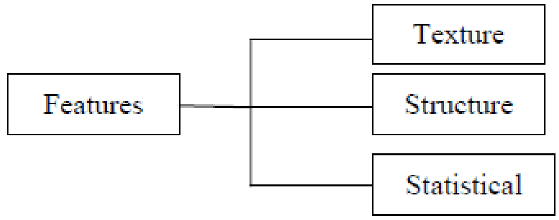

Figure 1. Types of features for breast mage classification.

III. Feature Selection Techniques

The authors designed a method to segment masses from DDSM mammogram images using thresholding and wavelet transform. The extracted features were selected using Genetic Algorithm and achieved 95% sensitivity [5]. The authors constructed an approach to detect micro calcifications in MIAS digital mammograms using Fuzzy C-means clustering technique. This was well defined for clustering the data sets. Particle Swarm Optimization (PSO) was used for selecting the features which avoid the minimal local value. The sensitivity for the approach FCM with PSO was 88.5% [6].

A novel approach for detecting abnormality in MIAS mammogram images was designed [7]. A fractional Fourier transform was used to get the unified time frequency spectrum. These values are given as input to PCA to select the optimal frequency values. The accuracy, specificity and sensitivity obtained were 92.22%, 92.10% and 92.16% respectively to detect the abnormal breasts. The authors [8] constructed an algorithm to find mass lesion from MIAS mammogram images using wavelet transformations. The optimal features were selected using PSO and gained the sensitivity as 94.99%.

The authors constructed an approach to detect abnormalities in digital mammograms using Particle Swarm Optimized Wavelet Neural Network (PSOWNN) [9]. In this approach, features were extracted using Laws Texture Energy measures and optimal features were selected using PSO. The performance measures calculated were accuracy, specificity, Sensitivity and AUC. These values were 93.67% accuracy, 92.105% specificity, 94.17 % and 0.96853 AUC.

In [10], the authors investigated on image preprocessing techniques for obtaining more accurate breast segmentation prior to mass detection, including global equalization transformation, denoising, binarization, breast orientation determination and the pectoral muscle suppression. After performing gray level quantization on the breast images segmented, the presented feature difference matrices could be created by five features extracted from a suspicious region of interest (ROI). Then subsequently, principal component analysis (PCA) was applied to select the features. The sensitivities for MIAS and DDSM were 88% and 86% respectively.

The features were extracted using Artificial Neural Networks (ANN) and Cellular Neural Networks (CNN) segmentation techniques from MIAS and DDSM benchmark datasets [11]. The optimal features were selected from the extracted features by using Genetic Algorithm. They obtained accuracy, sensitivity and specificity as 96.47%, 96.87%, 95.94% for both the datasets. The feature matrix was calculated using GLCM and 2D-DWT [12]. The optimal features were selected using t-test and F-test. The accuracy obtained were 98% and 94.2% for normal Vs abnormal and benign Vs malignant for MIAS dataset where as 98.8%, 97.4% for DDSM dataset respectively.

To extract the features rough set with chain codes was used for MIAS and DDSM dataset [13]. The optimal features were selected using genetic algorithm. The efficiency was evaluated by using Matthew's correlation coefficient (MCC). In their research, the random forest was used and compared with support vector machine (SVM), genetic algorithm support vector machine (GA-SVM), particle swarm optimization support vector machine (PSO-SVM), and decision tree. They achieved the accuracy of 97.73 %, and the MCC value as 0.8668 and 0.8652 for both the databases respectively. A CAD model [14] was designed to extract texture features from MIAS and DDSM datasets using shear let transforms. These features were optimized using T-test statistics and obtained 97% and 100% accuracy for MIAS and DDSM datasets respectively. A new model was constructed [15] to select features from segmented images by using texture, spatial and spectral features. The optimized features were selected based on Sequential Forward Search and achieved the accuracy of 98% for Support Vector Machine classifier where as for general regression neural networks it was 97.8% accuracy. The following table 1 summarizes the techniques applied by the researchers.

Table 1. Summary of mammogram classification techniques.

Reference No	Dataset	Feature Extraction	Feature Section	Results
[5]	DDSM	Thresholding and Wavelet transform	Genetic Algorithm	Sensitivity = 95%
[16]		Texture features	Dragonfly Algorithm	Sensitivity = 98.1% Specificity 97.8%
[6]	MIAS	Fuzzy C-means clustering	PSO	Sensitivity = 88.5%
[7]		fractional Fourier transform	PCA	Accuracy = 92.22% Specificity = 92.10% Sensitivity = 92.22%
[8]		Wavelet transform	PSO	Sensitivity = 94.99%
[9]	Mammogram Screening Centers	Laws Texture Energy measures	PSO	Accuracy =93.67% Specificity =92.105% Sensitivity=94.17 % AUC= 0.96853
[10]	MIAS DDSM	Textual features	PCA	MIAS-Sensitivity=88% DDSM –Sensitivity-86%
[11]		ANN and CNN	GA	Accuracy =96.47% Sensitivity=96.87 % Specificity =95.94%
[12]		GLCM and 2D-DWT	t-test and F-test	MIAS-Normal Vs Abnormal = 98% MIAS-Benign Vs Malignant = 94.2% DDSM Normal Vs abnormal = 98.8.% DDSM-Benign Vs Malignant = 97.4%
[13]		Rough set with chain codes	Genetic algorithm	MIAS- MCC = 0.8668 DDSM- MCC = 0.8652
[14]		Sherlet Transform	t-test statistics	MIAS-Accuracy = 97% DDSM-Accuracy = 100%
[15]	Nijmegen University Hospital (Netherlands) database	Texture, spatial and spectral features	Sequential Forward Search	Accuracy = 98%

The above discussed literature survey provides the information about the breast cancer in the patient which can help the radiologist and physicians in early detection. From the above study it is noticed that most of researchers used feature selection techniques to get good classification accuracy. This will increase the survival time of the patient.

IV. Dataset

There are different formats for different databases. Few image datasets of the JPEG and few images of DICOM format. Most of the researchers have used MIAS and DDSM mammogram image datasets. Some of the researchers have collected the dataset from local hospitals for their experiments. Other mammogram datasets are also available like BCDR and in breast [4] and can be used for the detection of breast cancer.

V. Conclusion

This paper reviewed several mammogram feature extraction and feature selection techniques for better classification accuracy. A review is made on several feature extraction techniques. To extract the features different feature extraction techniques were used by different researchers such as: texture features like energy, correlation, homogeneity, contrast, entropy, intensity based features like mean, variance, skewness, kurtosis and shape based features like area, major axis, minor axis, convex area, diameter, orientation, solidity, perimeter and Fourier descriptors. This paper also reviewed different feature selection algorithms. From the above literature, it is observed that the most of the researches used feature extraction, feature selection and classification steps for mammogram image classification. But to improve the classification accuracy, feature selection is very much needed. Number of optimization techniques were used in the literature. Among all, the mostly used techniques were Genetic Algorithms, Particle Swarm Optimization and Principle Component Analysis. So, keeping all these observations, we can conclude that feature selection is needed to improve the classification accuracy and helps the radiologists and physicians to detect the tumors in early stages.

References

[1] https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics

[2] C. Lewis, Fda sets higher standards for mammography. FDA Consum 33(1) (1999), 13-17.

[3] J. E. Skandalakis, Embryology and Anatomy of the Breast, Springer Berlin, Heidelberg, Germany, (2009).

[4] Abdullah-Al Nahid and Yinan Kong, Involvement of Machine Learning for Breast Cancer Image Classification: A Survey, Hindawi, Computational and Mathematical Methods in Medicine 29 (2017).

[5] Pereira DC, Ramos RP and Zanchetta do Nascimento M, Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm Computer Methods and Programs in Biomedicine 114(1) (2014), 88-101.

[6] J. Dheeba and Tamil Selvi, Bio inspired swarm algorithm for tumour detection in digital mammogram, in Proc First International Conference on Swarm, Evolutionary and Memetic Computing, Dec 16-18; Chennai, India Berlin Heidelberg: Springer (2010), 404-15.

[7] Yu-Dong Zhang, Shui-HuaWang Ge Liu and Jiquan Yang, Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform", Advances in Mechanical Engineering 8(2) (2016), 1-11.

[8] Y. C. Kuo, W.C. Lin, S. C. Hsu and A. C. Cheng, Mass detection in digital mammograms system based on PSO algorithm, IEEE International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan. (2014), 662-8.

[9] JN. Dheeba, A. Singh and S. Tamil Selvi, Computer-aided detection of breast cancer on mammograms: A swarm intelli¬gence optimized wavelet neural network approach, Journal of Biomedical Informatics 49 (2014), 45-52.

[10] C. C. Jen and S. S. Yu, Automatic detection of abnormal mammograms in mammographic images, Expert Systems with Applications 42(6) (2015), 3048-55.

[11] R. Rouhi M. Jafari, S. Kasaei and P. Keshavarzian, Benign and malignant breast tumors classification based on region growing and CNN segmentation, Expert Systems with Applications 42(3) (2015), 990-1002.

[12] Shradhananda Beura, Banshidhar Majhi and Ratnakar Dash, Mammogram classification using two dimensional discrete wavelet transform and ray-level co-occurrence matrix for detection of breast cancer, Neurocomputing 154, (2015).

[13] Min Dong, Lu. Xiangyu, Ma. Yide, Guo. Yanan, Ma. Yurun and J. Keju Wang, An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms *J* Digit Imaging 28(5) (2015), 613-625.

[14] N. Gedik, A new feature extraction method based on multiresolution representations of mammograms, Appl Soft Computing 44 (2016), 128-133.

[15] J. Fu, S. Lee, S. Wong, J. Yeh, A. Wang and H. Wu, Image segmentation feature selection and pattern classification for mammographic microcalcifications, Comput Med Imaging Graph 29(6) (2005), 419-429.

[16] S. Punitha, A. Amuthan and K. Suresh Joseph, Benign and malignant breast cancer segmentation using optimized region growing technique, Future Computing and Informatics Journal 3 (2018), 348-358.