IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

An Efficient Texture Feature Selection and Classification of Mammographic Image using AQPSO

Varsha Baby

Student, Department of Computer Science and Engineering, Karunya University, Coimbatore, Tamil Nadu, India

varshababy08@gmail.com

Abstract

In computer-aided diagnosis systems, Image processing algorithms can be used to extract features directly from digitized mammograms. In general, two classes of features are extracted from mammograms with these algorithms, such as morphological and non-morphological features. Image texture analysis is one of an important technique that represents gray level properties of images used to illustrate non-morphological features. This technique has made known to be a promising technique in analyzing mammographic lesions caused by masses. The texture descriptor namely entropy, energy, sum average, sum variance, and cluster tendency has been analyzed for texture pattern ROI. These textures features are derived from co-occurrence matrices, wavelet and ridgelet transforms of mammographic images. Earlier work used Genetic algorithm and Random Forest algorithm for selection and classification of these features. In order to improve the performance, proposed system uses Adaptive Quantum-behaved Particle swarm optimization for feature selection process. Comparison of AQPSO with Genetic Algorithm can be done experimentally and proves that the proposed system provides better result when compare with existing work.

Keywords: Mammography, Texture extraction, Co-occurrence Wavelet and ridgelet, Adaptive Quantum-behaved Particle Swarm Optimization

Introduction

Breast cancer is the most commonly diagnosed cancer for women in all over the world. It is the second most general and leading cause of cancer death among women. Nearly, One out of ten women could develop breast cancer during their lifetime. In favor of this reason, studies have shown that early detection is the key to improve breast cancer prediction. Mammography is considered as the most effective technology currently available for breast cancer screening. Computer Aided Detection (CAD) with digital or digitized mammograms has proven to be a very useful tool for radiologists. Image processing algorithms can be utilized in computer-aided diagnosis systems to extract features directly from digitized mammograms. Normally, two phases of features are extracted from mammograms with these algorithms, namely morphological and non-morphological features. Texture analysis of image is an important technique that characterizes gray level properties of images used to describe non-morphological features. It has shown to be a promising technique in analyzing mammographic lesions caused by masses. Multiresolution analysis has proved to be useful in mammographic image processing, image enhancement, mass detection, and feature extraction. The general task is to decompose the original image into sub-bands that preserve high and low frequency information. Numerous studies have examined the use of wavelet transform as a multiresolution analysis tool for analysis of texture and classification.

The features used for texture classification are derived from co-occurrence matrices, wavelet and ridgelet transforms of mammographic images is assessed. Especially, a false positive reduction is initiated in computer-aided detection of masses. The data set used in this work consisted of 90 cranio-caudal mammograms, in which 60 images containing a mass, rated as abnormal images, and the rest 30 images with no lesions.

Screen/film mammography has shown to be an effectual assist for radiologists in the early detection of occult breast cancers in asymptomatic women and in the reduction of mortality rates. While it is seen as the most reliable method for early detection of breast cancer, its interpretation is very complex. Sensibility of screening mammography is influenced by image quality and depends on the radiologist's level of expertise. Consequently, some studies show that approximately 10–30% of malignant breast cancers are visible on mammograms and go undetected by radiologists during routine mammographic screening. In addition, it has been observed that only 15–34% of women who suffer a biopsy based on the results of a mammographic examination actually have malignant lesions. In order to improve the accuracy of mammography, double reading of the same screening mammogram has confirmed to increase the sensitivity rate. So the major problems are described below:

- Mammograms with false positive results may direct the normal person to undergo treatment and diagnosis.
- The major problem faced is the false positive.

The main contribution of the proposed work is to validate whether a given sub-image containing a suspicious region contains a real lesion or it is only a region depicting a normal parenchyma. The process of contribution includes as follows:

- To develop a CAD system that classifies the mass region based on the feature selection using AQPSO
- To improve the false positive reduction.

Related Work

In [1] Oliver et al. presented CAD systems.. This system can be broadly categorized into two types – computer- aided detection and computer-aided diagnosis. Computer- aided detection schemes are systems that automatically detects suspicious lesions in mammograms, being used as a localization task. Computer-aided diagnosis systems extend the computer analysis to give way as output the description of a region or the estimated probability of lesion malignancy. This system is focused on the classification task. Image processing algorithms are used in computer-aided diagnosis systems to extract features directly from digitized mammograms. Typically, two classes of features are extracted from mammograms with these algorithms, namely morphological and non-morphological features. Morphological features are aimed to describe information related to the morphology of a lesion, namely lesion size and shape. Image texture analysis is a significant class that represents gray level properties of images used to describe non-morphological features.

In [2] Mousa et.al presented two techniques based on wavelet analysis and fuzzy-neural approaches. These approaches are mammography classifier which is based on globally processed image and mammography classifier based on locally processed image that is region of interest. The framework classifies normal from abnormal, abnormal severity (benign or malignant) and mass for micro calcification. The evaluation of the framework is carried out on Mammography Image Analysis Society (MIAS) dataset.

In [3] Mudigonda et al. presented a method for the detection of masses in mammographic images which employs Gaussian filtering and sub sampling process as preprocessing steps. To segment the mass portions a method is then applied by generating intensity links from the central portions of masses into the surrounding areas in the image. In order to discriminate between TP mass regions and FPs, a method is used for analyzing oriented textural information in mammograms. LDA was used for pattern classification. The texture flow-field methods have been specifically used for estimating features that distinguish masses from FPs.

In [4] Moayedi et.al presented the design and development of an automatic mass classification of mammograms. This consists of three phases. In the first phase, preprocessing is achieved to eliminate the pectoral muscles and to segment regions of interest. The second phase uses contourlet transform as a feature extractor to attain the contourlet coefficients. This phase is ended by feature selection based on the genetic algorithm, resulting in a more discriminative texture feature set. In the final phase, classification is executed based on successive enhancement learning weighted SVM, SVM fuzzy neural network and kernel SVM.

In [5] Do & Vetterli presented the ridgelet transform as a new multiresolution analysis tool that treats effectively with line singularity in two dimensions. The scenario is to map a line singularity into a point singularity using the Radon transform. The ridgelet transform allows for representing singularity along lines in a more proficient way, by the compactness of the representation for a given reconstruction accuracy. This transform has been utilized in texture classification of CT medical and mammographic images.

Previous Work

In the previous work the genetic algorithm is used to select the features.

Feature selection: For the feature selection process, both Matlab and WEKA environments are connected with the aim of integrating the GA and the classifier algorithms. This can be done to find which textural features weight more exhaustively in the classification process of malignant mammographic images. Originally, a random population was generated by having chromosome sizes equal to the set of features extracted from each image. These chromosomes gene were encoded as a binary string, where bit 0 depicts to the absence of a given feature and bit 1 corresponds to the presence of the same feature. Then the features selected by the GA were passed to the WEKA Random Forest classifier algorithm for training and testing purpose. Towards the end of the execution, information about the classifier performance for the used data set were passed back to the GA, which is in turn used the classification area under the ROC curve as the chromosome fitness value. One of the main drawbacks of this evolutionary algorithm is their absence of memory, which can affect the search and convergence ability of the algorithms.

In GA, the concept of memory relies on elitism. Only a stronger operator can propagate accurate solutions in a faster way. So the below given forms the problem statement:

- Lack of memory
- Limitation of the search and convergence ability of the algorithm
- No stronger operator to propagate accurate solutions in a faster way.

Proposed Work

In the proposed work, features of mammographic images has been selected by using AQPSO, is an advanced PSO technique. In this the two extreme of the particles can be shared, then the proposed method adaptively searches their optimum solutions in parallel. This combines the optimizations of an improved adaptive PSO (APSO) with a quantum Particle Swarm Optimization (QPSO).

The four primary methodologies for implementation are described as follows:

PRE Processing

The preprocessing is done for removing noises in the images and along with this cropping is also performed. Initially, the mammograms were pre-processed to excerpt regions of interest (ROI). Next the cropping was performed on the pre-processed images. Then Texture analysis is performed only on special regions within the ROI and not performed on the full ROI. Texture analysis is frequently restricted to the mass region, except the background tissue region or on bands of pixels that is close to the mass margin.

Feature Extraction

Once the mammographic images were pre-processed as described in the last section, the three approaches considered earlier were applied and textural feature vectors were extracted. The following texture descriptors were then calculated to analyze the ROIs texture patterns: entropy, energy, cluster tendency, sum average and sum variance Feature extraction based on the five texture descriptors was implemented for each of the three considered approaches.

The five descriptors are presented as follows:

Entropy: $f_1 = \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j) \log\big(p(i,j)\big)$

Energy: $f_2 = \sum_{i=1}^{M} \sum_{j=1}^{N} \big[\big(p(i,j)\big)\big]^2$

Sum - average: $f_{3=} \sum_{k=1}^{M+N} k p_{x+y}(k)$

Sum - variance: $f_{4=} \sum_{k=1}^{M+N} (k-f_3)^2 p_{x+y}(k)$

Cluster tendency: $f_5 = \sum_{i=1}^{M} \sum_{j=1}^{N} (i+j-2\mu)^2 p(i,j)$

Co-Occurence Matrix

A matrix for each Region of Interest (ROI) was computed for four directions (0^0, 45^0, 90^0, and 135^0) and the five descriptors were computed for each matrix, finally giving a total of 20 descriptors per ROI. Total of 5 features per image is calculated in the co-occurrence matrix. The data set consists of 3 types of images namely normal, benign and cancer each containing 30 images each thus comprises to 90 images altogether. So a total of 450 features were extracted altogether for a total of 90 images.

Wavelet Transform

For each ROI, the 2D-DWT was applied using the Db3 wavelet mother and two resolution levels, which results six detail coefficient matrices and one approximation coefficient matrix.

Ridgelet Transform

In the ridge let transform, an image is converted into two matrices. Each matrix is divided into 61 columns. So a total of 610 features per image is calculated in the ridge let transform. The data set consists of 3 types of images namely normal, benign and cancer each containing 30 images each thus comprises to 90 images altogether. So a total of 54900 features were extracted altogether for a total of 90 images.

Feature Selection

For the feature selection procedure, it was connected both Matlab and WEKA environments with the aim of integrating the AQPSO and the classifier algorithms.

In this we define two phases of particle swarm: attraction and repulsion. It can be demonstrated that when the Creativity parameter satisfies $\beta \leq 1$, the particles will be bound to converge to its local LIP *p*, and some particles will depart from *p* when $\beta > 1$; the larger the β, the more particles will explode.

Attraction Phase: $\beta=\beta_a$, where $\beta_a \leq 1$;

ISSN: 2277-9655
Impact Factor: 1.852

Repulsion Phase: $\beta=\beta_r$, where $\beta_r > 1$

In attraction phase ($\beta=\beta_a$) the swarm is contracting, and subsequently the diversity decreases. When the diversity drops below a lower bound, d_{low}, switching to the repulsion phase ($\beta=\beta_r$) can be done in which the swarm expands. Finally, when the diversity reaches a higher bound, it switches back to the attraction phase. The result of this is an AQPSO algorithm that alternates between phases of exploiting and exploring-attraction and repulsion-low diversity and high diversity, according to the diversity of the swarm measured by

$$\text{Diversity (S)} = \frac{1}{|S|.|L|}\sum_{i=1}^{|S|}\sqrt{\sum_{J=1}^{D}(Pij-\overline{PJ})^2}$$

where S is the swarm, M=S is the population size, L is the length of longest the diagonal in the search space, D is the dimensionality of the problem, p is the j'th value of the i'th particle (pbest).

The Quantum-behaved PSO algorithm with attraction and repulsion phases is called Adaptive Quantum-Behaved Particle Swarm Optimization (APSO) algorithm, which is described as follows.

Algorithm

Initialize population: random x_i
Do
Find out mbest using diversity equation
Measure the diversity of the swarm)
If (diversity<dlow) beta=betaa;
If (diversity>dhigh) beta=betar;
for i=1 to population size M
If $f(x_i)<f(p_i)$ then $p_i=x_i$
$p_g=\min(p_i)$
for d=1 to dimension D
$fi_1=rand(0,1)$, $fi_2=rand(0,1)$
$p=(fi_1*p_{id}+fi_2*p_{gd})/(fi_1+fi_2)$
u=rand(0,1)
if rand(0,1)>0.5
$x_{id}=p-beta*abs(mbest_d-x_{id})*(\ln(1/u))$
else
$x_{id}=p+beta*abs(mbest_d-x_{id})*(\ln(1/u))$
Until termination criterion is met

Classification

The features selected by the AQPSO were passed to the WEKA Random Forest classifier algorithm to train it and test it.

The Random Forest is defined as a classifier constructed from a collection of classification trees. It is a concept of regression trees induced by bootstrap samples of a training data set, with random features selected in the induction tree process. In this method, each tree is constructed in the following way:

Step 1: Data is withdrawn from a training set through a random sampling process with bootstrap, where 2/3 of the data are used for growing the tree.

Step 2: A random number of features is selected from the training set and the one with the largest number of information's is used to split the node.

Step 3: The growing task continues until no node can be created for lack of information.

Step 4: The error rate is estimated using the 1/3 of the data left, by predicting their classes.

The decision algorithm of the Random Forest method works as follows:

1. Select the number of growing trees, represented by the parameter N, and an integer m not greater than the number of features passed to the classifier.
2. For i from 1 to N train the forest.
3. Randomly select a bootstrap sample of the data. The data that are not selected in this step are named as out-of-bag.
4. Grow a random tree, where at each node the best set is chosen among m variables randomly selected.
5. Use the tree to predict the out-of-bag data.
6. At the end, use the results obtained from the out-of-bag data to indicate the class.

Experimental Result

In this section the Accuracy value can be estimated for the GA with proposed AQPSO. AQPSO provides better accuracy when compared with GA.

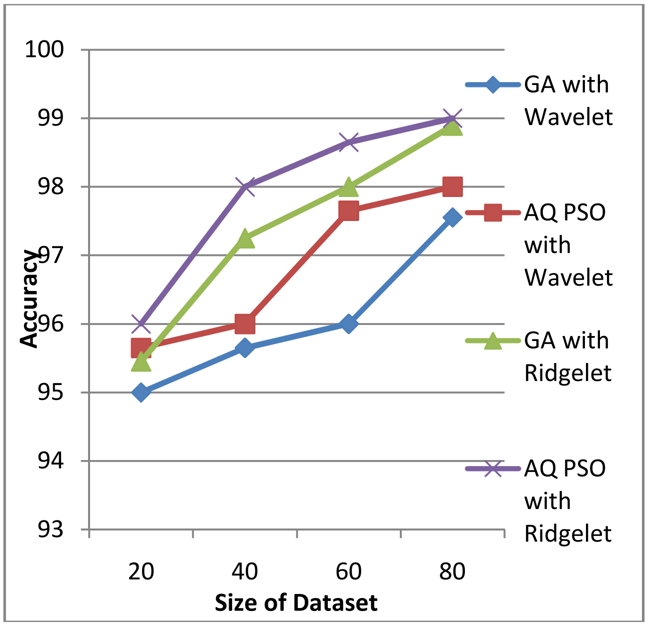

Figure 1: Accuracy comparison graph

The above graph in figure 1 shows that AQPSO of Ridgelet and wavelet shows higher accuracy value in feature selection when compare with GA of Ridgelet and wavelet. Co-occurace value is not considered since it contains only less number of features.

ISSN: 2277-9655
Impact Factor: 1.852

Conclusion

Multiresolution techniques are important in image processing because of its similarity to the human visual system. These system processes less image data by selecting relevant details to carry out visual recognition tasks. In the present work, Adaptive Quantum-behaved Particle Swarm Optimization Algorithm (AQPSO) is proposed for selecting the optimal feature. The proposed approach formulates the quantum-behaved PSO and then gives an adaptive approach of parameter control and propose AQPSO algorithm. Two multiresolution techniques, namely wavelet and ridgelet transforms are used to extract textural features from mammogram images for breast cancer classification. This feature can be selected by using AQPSO. Next, the selected feature can be given as input for Random forest classifier for classifying the normal and abnormal tissues in mammographic images. Experimental result provides better result when compare with the existing work.

References

[1] *Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. IEEE Transactions on Image Processing, 12, 16–28.*

[2] *Llado, X., Oliver, A., Freixenet, J., Martí, R., & Marti, J. (2009).A textural approach for mass false positive reduction in mammography. Computerized Medical Imaging and Graphics, 33, 415–422.*

[3] *Moayedi, F., Azimifar, Z., Boostani, R., & Katebi, S. (2010). Contourlet - based mammography mass classification using the SVM family. Computers in Biology and Medicine, 40, 373–383.*

[4] *Mousa, R., Munib, Q., & Moussa, A. Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Systems with Applications, 28, 713–723*

[5] *Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Transactions on Medical Imaging, 20, 1215–1227*