BREAST CANCER DETECTION AND CLASSIFICATION BY USING CLUSTERING METHODS WITH SUPPORT VECTOR MACHINE

G. Anusha[1], K. Shirisha[2], K. Varshini[2], K. Shravani[2], K. Akshaya[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

Abstract: Breast cancer is one of the leading diseases for women in the world. It is ranked second among all types of cancers, to cause death in women. So, the early detection is necessary to reduce the mortality rate. The early detection of breast cancer and treatment leads to an increase in the survival rate of women. Mammography is a standard radiological screening technique, which is used for checking of breast cancer in women when there are no symptoms. Few machine learning (ML) approaches like support vector machine (SVM) utilized for the detection and classification of breast cancer in the literature. However, SVM produced the results with inaccurate classification. Thus to overcome this problem, the research work is implemented with the Advanced deep learning based Probabilistic neural networks (PNN) classification mechanism. Initially, k-means clustering based segmentation approach is used for efficient detection the region of breast cancer. Finally, to archive the maximum efficiency of the system, PNN developed for classification of breast cancer with the gray level co-occurrence matrix (GLCM) based Texture features; discrete wavelet Transform (DWT) based low level features and Statistical Color features respectively. Thus, the research work can be effectively used for classification of Benign and Malignant breast cancers. The simulation analysis shows that the proposed method shows better qualitative and quantitative analysis compared to the state of art approaches.

Key words: MRI Medical Images, DWT, Clustering Process SVM.

1.Introduction:

Breast cancer is one of the leading diseases, that reflects an uncontrolled growth of abnormal cells in the breast. Due to the breast anomalies properties and the nature of the human visual perception, it is natural that, sometimes the abnormalities are missed or miss classified. As a result, unnecessary biopsies are taken. In breast, normal cells grow and divide at a particular time but in case of the cancerous cells, the cell growth is continuous and uncontrolled as shown in Figure 1.

Most of the research work is focused on using optimization techniques to develop a Classification [8] and Diagnosis [9] of breast cancer from Mammographic images. The detection & classification of irregularities in Mammographic images are considered for investigation in this paper. Poor noise-to-signal

ratio is a drawback in Mammographic images. The anatomically distinct structures are often seen with a very low contrast. Reliable standard image processing technique[10] is needed for its computation.Modification in image content is done in a highly controlled and reliable way without any compromise in clinical decision-making, but the presence of artifacts leads to 10 – 25% of tumors being missed by radiologists. Basic noise removal filters [11] can not be applied on Mammographic images as they are not able to remove the artifacts effectively. If we use those fundamental filters then, image get corrupted and enhancement operation will not work.

The early detection of breast cancer and treatment leads to an increase in the survival rate of women. Mammography is a standard radiological screening technique, which is used for checking of breast cancer in women when there are no symptoms. Few machine learning (ML) approaches like support vector machine (SVM) utilized for the detection and classification of breast cancer in the literature. However, SVM produced the results with inaccurate classification. Thus to overcome this problem, the research work is implemented with the Advanced deep learning based Probabilistic neural networks (PNN) classification mechanism. Initially, k-means clustering based segmentation approach is used for efficient detection the region of breast cancer. Finally, to archive the maximum efficiency of the system, PNN developed for classification of breast cancer with the gray level co-occurrence matrix (GLCM) based Texture features; discrete wavelet Transform (DWT) based low level features and Statistical Color features respectively. Thus, the research work can be effectively used for classification of Benign and Malignant breast cancers. The simulation analysis shows that the proposed method shows better qualitative and quantitative analysis compared to the state of art approaches.

2.Related Work

Topic :image segmentation via adaptive k-mean clustering and knowledge-based morphological operations

Image segmentation remains one of the major challenges in image analysis. In medical applications, skilled operators are usually employed to extract the desired regions that may be anatomically separate but statistically indistinguishable. Such manual processing is subject to operator errors and biases, is extremely time consuming, and has poor reproducibility. We propose a robust algorithm for the segmentation of three-dimensional (3-d) image data based on a novel combination of adaptive k-mean clustering and knowledge-based morphological operations. The proposed adaptive k-mean clustering algorithm is capable of segmenting the regions of smoothly varying intensity distributions. Spatial constraints are incorporated in the clustering algorithm through the modeling of the regions

by gibbs random fields. Knowledge-based morphological operations are then applied to the segmented regions to identify the desired regions according to the a priori anatomical knowledge of the region-of-interest. This proposed technique has been successfully applied to a sequence of cardiac ct volumetric images to generate the volumes of left ventricle chambers at 16 consecutive temporal frames. Our final segmentation results compare favorably with the results obtained using manual outlining. Extensions of this approach to other applications can be readily made when a priori knowledge of a given object is available.

3.Proposed system

The proposed research work majorly focusing on detection of following breast cancers such as Malignant and Benign, respectively. The detailed operation of the breast cancer detection and classification approach is presented Database is trained from the collected images of "International Breast Imaging Collaboration " Archive. BCP is one of the biggest available collections of quality controlled dermoscopic images. The dataset consisted of 266 benign and 200 malignant images. All the images are trained using the PNN network model with GLCM features, statistical and texture features. And random unknown test sample is applied to the system for detection and classification, respectively.

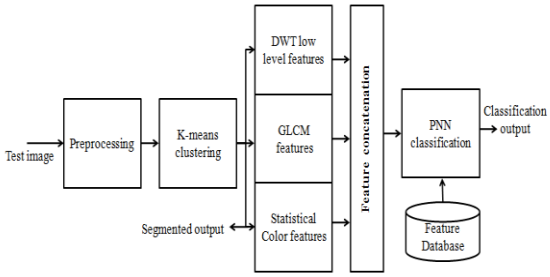

Fig . Breast cancer detection and classification.

3.1 Wavelets Method

The Wavelet change is a change of this sort. It gives the time-recurrence portrayal. (There are different changes which give this data as well, for example, brief time Fourier change, Wigner dispersions, and so on.)

Wavelet change is equipped for giving the time and recurrence data at the same time, consequently giving a period recurrence portrayal of the picture. DWT utilizes two arrangements of capacities, called scaling capacities and wavelet capacities, which are related with low pass and high pass channels, separately. than □ . The picture can consequently be sub sampled by 2, basically by disposing of each other example. This comprises one degree of decay and can numerically be communicated as pursues:

$$y_{high}[k] = \sum_{n} x[n] \cdot g[2k - n]$$

$$y_{low}[k] = \sum_{n} x[n] \cdot h[2k - n]$$

where y high[k] and y low[k] are the yields of the high pass and low pass channels, individually, in the wake of sub sampling by 2.

3.2 K-mean algorithm

In the proposed method, we combine histogram statistics and k-means clustering to track the tumor objects in MR brain images. It's the on the techniques for the clustering concept in the data mining process and is very famous algorithm for the K-means clustering, because it is similar or simpler and easier in computation of an efficient K-means clustering algorithm it is the simplest unsupervised learning algorithms that solve the well known clustering problems. It's the K-means algorithm is an unsupervised clustering algorithm that classified in the input data points into multiple classes based on their intrinsic distance from other dataset points of his cluster The flowchart of our method is shown as fig.1.

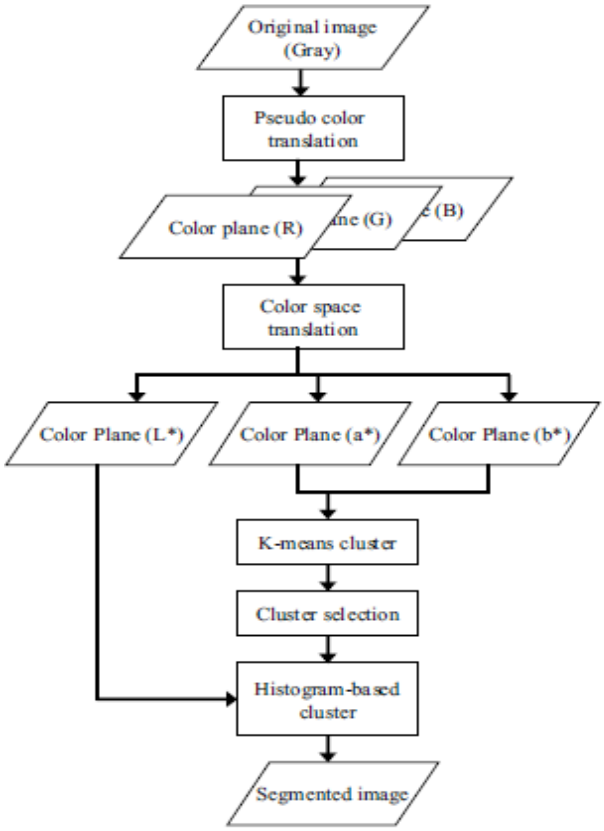

fig:1. Present algorithm for k-means

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space similar feature vectors into as in glecluster and for grouping data points with dissimilar feature vectors into different clusters. let the feature vectors derived from l clustered data set {xi1,2,…,l} . The generalized algorithm initial task cluster

centroids c □ {c j j1,2,……,k} by randomly selecting k feature vectors from x. a pixel can be classified into one or the two classes. An image f(x,y) can be segmented into two classes using a gray value threshold t so that

$$g(x,y)=\begin{cases}1 & \text{if } f(x,y)>T\\ 0 & \text{if } f(x,y)\le T,\end{cases} \quad (1)$$

$$J_m=\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}^{m}\left\|x_i-c_j\right\|^2 \quad , \quad 1\le m<\infty$$

Where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ‖*‖ is any norm expressing the similarity between any measured data and the center.

4. SVM The classification

SVM The classification of hand gesture images is performed using the SVM classifier by employing one-against-all (OAA) approach. Support Vector Machine is a powerful machine learning technique for classification and regression. It is a supervised learning machine where its support vectors and kernels are employed for many learning tasks. By choosing the appropriate kernel functions, different tasks could be performed in various domains. A support vector machine constructs a separating hyper plane in a high dimensional space. SVM is used to classify the group of test data as one of the ten gestures, depending upon the feature values. In this research, classification is done for ten categories of gesture images. Therefore, ten binary SVM models are created where each SVM model is trained to distinguish one class of images from the remaining nine classes. For example, the SVM classifier for class one data (number zero) is assigned +1 and the remaining nine classes (numbers one, two, three, four, five, six, seven, eight, nine) are assigned as -1. Other SVM classifiers are constructed on the same way. Ten SVMs are trained independently for classifying ten classes of hand gestures. When the test or query image is given, it is classified based on the trained SVM models This can be rewritten as: We can put this together to get the

$$\vec{w}\cdot\vec{x}_i+b\ge 1, \text{ if } y_i=1$$

or

$$\vec{w}\cdot\vec{x}_i+b\le -1, \text{ if } y_i=-1.$$

To for "The and that solve this problem determine our classifier, An easy-to-see but important consequence of this geometric description is that max-margin hyper plane is completely determined by those which lie nearest to it. These are called support vectors.

5.Extraction Stage

Feature extraction is the process of getting useful information from the Image/character image. The information will be used to generate modules to train the classifier and to be used for classification purposes. In general there are two

categories of features extracted, structural and statistical features. Choosing the right feature extraction method might be the most important step for achieving a high recognition rate . However, in some cases the combination of several features extraction types could be a wise decision to enhance the overall recognition performance.

6.RESULT

The experiments are done using MATLAB R 2013a tool. BCP is one of the biggest available collections of quality controlled dermoscopic images. For the implementation of the proposed method, spatial domain, and frequency domain of 30 dermoscopic breast lesion images (266-benign and 200-Malignant) have been obtained respectively by applying rotations at different angles. Train images of each label have been used to train the PNN architecture

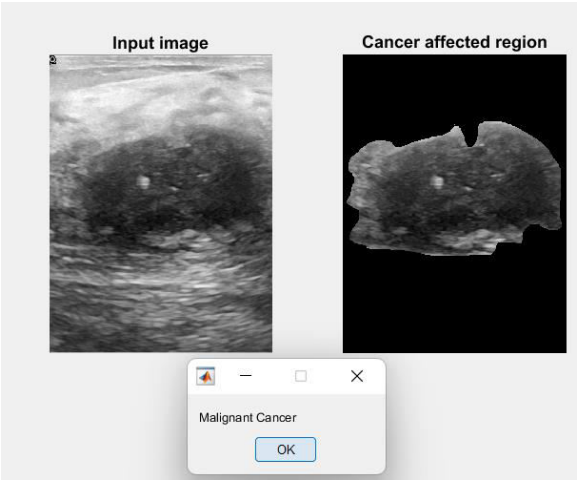

Fig. 6.1: Segmented output images

including denoising and enhancing to increase the SNR. How to combine segmenting with preprocessing procedure is our work in future.

Metric	method	Test 1	Test 2	Test 3	Test 4
Accuracy	EXISTING	0.9157	0.78099	0.85796	0.47765
	PROPOSED	0.99985	0.99715	0.99999	0.99999
Sensitivity	EXISTING	0.70588	0.90024	0.9166	0.83857
	Proposed	0.99931	0.99198	1	1

For evaluating the performance measure the proposed method is implemented with the two types of segmentation methods, they are Active contour (AC) and k-means clustering, respectively. For performing this comparisons Accuracy, Sensitivity, parameters are calculated, respectively.

Feature scope

As a future work, network performance can be analyzed with different Types Of Neural Network.

REFERNCE

[1] chen, c. W, luo, j., parker, k. J., “image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with

biomedical applications,"*ieee transactions on image processing*, vol. 7, no, 12,dec. 1998, pp. 1673-1683.

[2] christine, c., thomas, f., "a study of efficiency and accuracy in the transformation from rgb to cielab color space," *ieee transactions on image processing*, vol. 6, no. 7, july 1997, pp. 1046-1048.

[3] dhawan, a. P., "a review on biomedical image processing and future trends," computer methods and programs in biomedicine, vol. 31, no.3-4, 1990, pp.141-183.

[4] gonzalez, r. C.; woods, r. E., *digital image processing, 2nd ed.*, prentice-hall, englewood cliffs, nj, 2002.

[5] ng, h. P., ong, h. H., foong, k. W. C., goh, p. S., nowinski, w. L., "medical image segmentation using kmeans clustering and improved watershed algorithm,"ieee southwest symposium on image analysis and interpretation, 2006, pp. 61-65.

[6] tsai, c. S., chang, c. C., "an improvement to image segment based on human visual system for object-basedcoding," fundamental informaticae, vol. 58, no. 2, 2004,