ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 5, May 2012

Ultra sonogram Images for Thyroid Segmentation and Texture Classification in Diagnosis of Malignant (Cancerous) or Benign (Non-Cancerous) Nodules

Nikita Singh, Alka Jindal

Abstract*- In this work to provide the information about an object clinically in terms of its size, shape and type of image for this image segmentation and classification are important tool in medical image processing. Ultrasound imaging is the best way to prediction of which type of thyroid is there. In this paper, images were distinguishing in two groups Benign (non-cancerous) and Malignant (cancerous) Thyroid Nodules. These images were used in classification analysis. An automatic system is developed that classifies the thyroid images and segments the thyroid nodular area using machine learning algorithms. The classifiers such as SVM, KNN and Bayesian are used. The features like GLCM total 13 features are extracted and these features are used to train the classifiers such as SVM, KNN and Bayesian. The result was compared with the ground truth images obtained from the radiologist and the performance measure such as accuracy is evaluated. It is observed that the SVM gives much better accuracy than KNN and Bayesian.

Keywords- **Thyroid Ultrasound (US) images, GLCM, RBAC, SVM, KNN and Bayesian.**

I. INTRODUCTION

Thyroid gland is a butterfly shaped organ and is compose of two cone-like lobes. Thyroid gland is one of the largest endocrine gland and is located below the skin and muscles at the front of the neck. In modern medicine, various medical images – ultrasound, CT, Scintigraphy, SPECT, MR, PET, X-rays etc play an important role in process of disease diagnosing and treating and have become major evidence to ensure disease [1]. Ultra sonography is the most well accepted imaging modality for the diagnosis and follow-up of thyroid disorder. The advantages of using ultrasonic imaging include its mobility and low cost as well as the ability to measure the dimension of the gland check for the presence of masses or cysts and evaluate the structure and echogenicity of the parenchyma [2]. High-resolution ultrasound (US) is the most sensitive imaging test available for the examination of the thyroid gland, to detect thyroid lesions, accurately calculate their dimensions, and identify the internal structure. A thyroid ultrasound examination provides an objective and precise method for detection of a change in the size of the nodule, used to evaluate the US features, which include size, echogenicity (hypo echoic or hyper echoic), and composition (cystic, solid, or mixed), as well as presence or absence of coarse or, a halo and irregular margins[3]. Segmentation plays an important role in medical imaging to obtain the location of the object of interest as well as to detect the area, volume or the analysis of dynamic behaviour of anatomical structure over time [4]. Thus by segmentation process the affected or the region of interest can be separated from other tissues. To detect the abnormality of thyroid gland, first the location and size of the gland must be segmented. A segmentation algorithm based on localized based [5] method that is basically to select the small region of the thyroid nodule or to segment the local area of the images and to segment the nodule which is gives the information of which type of nodule exist benign and malignant. In digital image processing techniques offer the opportunity for texture description. The thyroid nodule can be characterized by texture description and quantifying properties [2]. The thyroid texture characterization based on statistical parameter could provide an objective diagnostic tool and contribute to the use of computer assisted application in thyroid disorders. The most famous feature extraction technique are presented based on (GLCM) next, classification method which is able to distinguish between a benign (non cancerous) nodule and malignant (cancerous) have been present based on the SVM classifier, KNN classifier and Bayesian classifier. In this project work, we proposed to develop a computer aided diagnosis system of thyroid ultrasound images [9]. In this module, firstly thyroid gland region are segmented from the nodular (noncancerous) region in the normal thyroid nodule images and nodular (with cancerous) region in the abnormal thyroid nodule images. Then the segmented thyroid ultrasound images were use the features extraction techniques and also used SVM, KNN and Bayesian classifier. From this module classified result, the results obtained from radiologist and the performance measures such as accuracy are calculated. The rest of this paper is organized as follows. Methodology in section II introduces a description of the main components of the proposed scheme. Then in section III result and discussion performance Evaluation study on real US thyroid data is presented, demonstrating the effective accuracy. Finally in the last section IV the conclusion of this study is summarized.

II. PROPOSED METHODOLOGY

A. Database- The famous medical imaging is thyroid Ultrasound images. These images are mixed types like some images has nodules, some images has not nodules, some type of images are benign (non-cancerous) and some type of images are malignant (cancerous) nodule. Total 13 Number of thyroid images were used where total 8 cancerous and 5 non-cancerous images was selected in database. These thyroid images provided by internet (Thyroid Images Wilmington Endocrinology PA, Gallery- category-thyroid Ultrasound Images). The format of images was used in JPEG.

B. Software and computer used for analysis- We used MATLAB version 7.7.0 (R2008b) and used image processing Toolbox) For our analysis and used a computer with Intel® Core™i3-350M Processor 2.26 GHz CPU and 3 Gigabyte of memory.

C. Data pre-processing Ultrasound images contain speckle noise and to remove the noise various filters are used and also used histogram equalization produce visual differences and enhanced the contrast between images .The various modules of proposed work are classification of thyroid images and segmentation of thyroid nodular images. The basic steps of the proposed methodology are shown in Fig l.

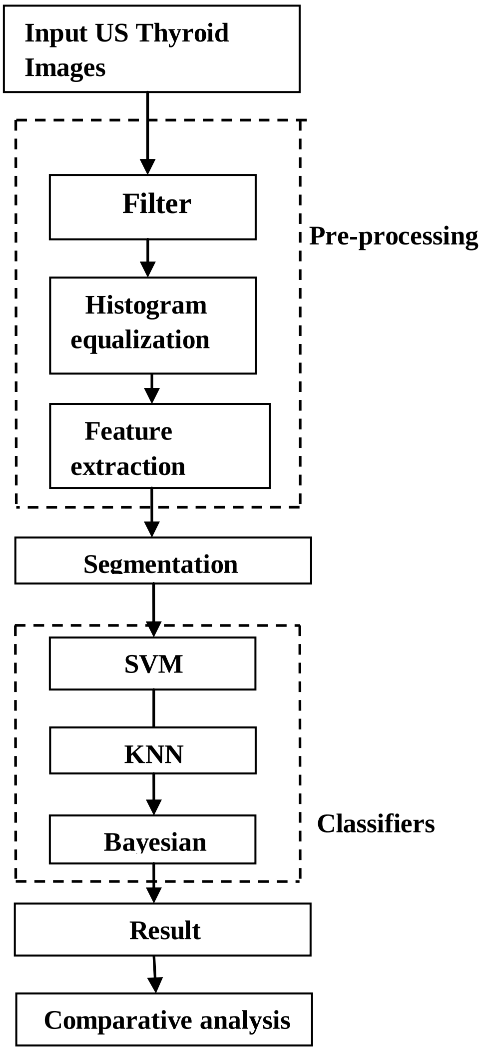

Fig 1 Flow Chart of the Proposed Methodology

D. Texture analysis and feature extraction - A feature used to partition images into region of interest and to classify those region. Its provide information in the spatial arrangement of colours or intensities in an image and the Characterization by the spatial distribution of intensity level in a neighbourhood. Use Gray-co-matrix and extract features from that. GLCM calculates the probability of a pixel with the gray-level value *i* occurring in a specific spatial relationship to pixel with the value *j*. The number of gray levels in the image determines the size of the GLCM [6]. Although there is a function in Matlab Image Processing toolbox that computes parameters Contrast, Correlation, Energy, solidity and Homogeneity, the paper by Haralick suggests the tabulation form where few more parameters that are also computed here [7]. It is easy to add new features based on the GLCM using this code. And also calculate the new formulas which is helps to extract the US features and also helpful to disguising the benign and malignant nodules. There are following feature extraction equations:

$$\text{Correlation} = \frac{\sum_{i=0}^{L}\sum_{j=0}^{L}(ij)P(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (1)$$

$$\text{Difference Entropy} = \sum_{i=0}^{L} P_{x-y}(i)\log(P_{x-y}(i)) \quad (2)$$

$$\text{Difference Variance} = \sum_{i=0}^{L}\left([i - \sum_{j=0}^{L} jP_{x-y}(j)]^2 P_{x-y}(i)\right) \quad (3)$$

$$\text{Sum average} = \sum_{i=2}^{2L} iP_{x+y}(i) \quad (4)$$

$$\text{Sum Entropy} = -\sum_{i=2}^{2L} P_{x+y}(i)\log(P_{x+y}(i)) \quad (5)$$

$$\text{Sum of Squares} = \sum_{i=0}^{L}\sum_{i=0}^{L}(i-\mu)^2 P(i,j) \quad (6)$$

$$\text{Sum Variance} = \sum_{i=0}^{2L}(i - F5)^2 P_{x+y}(i) \quad (7)$$

$$\text{Contrast} = \sum_{n=0}^{L} n^2 \left(\sum_{i=0}^{L}\sum_{\substack{j=0 \\ |i-j|=n}}^{L} P(i,j)\right) \quad (8)$$

$$\text{Energy} = \sum_{i=0}^{L}\sum_{i=0}^{L}(P(i,j))^2 \quad (9)$$

$$\text{Entropy} = -\sum_{i=0}^{L}\sum_{j=0}^{L} P(i,j)\log P(i,j) \quad (10)$$

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 5, May 2012

Local Homogeneity = $\sum_{i=0}^{L}\sum_{j=0}^{L}\frac{P(i,j)}{1+(i-j)^2}$ (11)

ClusterShade= $\sum_{i=0}^{L}\sum_{j=0}^{L}\left(i-E_x+j-E_y\right)^3 P(i,j)$ (12)

ClusterProminence= $\sum_{i=0}^{L}\sum_{j=0}^{L}\left(i-E_x+j-E_y\right)^4 P(i,j)$ (13)

E. Segmentation- We process the image from the region using the segmentation based algorithm localized based active contour (region based) [5] method that is basically to select the small region of the thyroid nodule or to segment the local area of the images and to segment the nodule which is give the information of which type nodule exist benign and malignant.

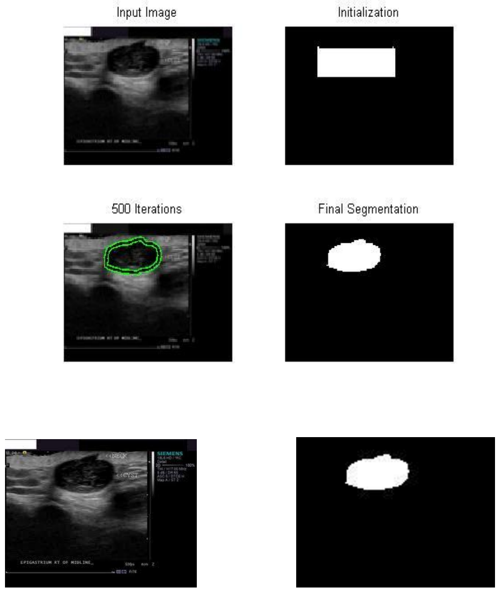

Fig 2 Final Segmentation

In Fig 2 shows the proper procedure of segmentation algorithm the thyroid benign image are segmented with 500 iteration and mask m (30:70, 50:170) = 1 Thyroid lobes are segmented into left and right lobes using region based active contour. RBAC effectively segments the thyroid gland in US images [5]. RBAC algorithm provides robustness against initial curve placement and insensitivity to image noise. RBAC achieves minimize energy:

$$E(\varphi) = \int_{\Omega_x} \delta\varphi(x) \int_{\Omega_x} \mathcal{B}(x,y).F(I(y),\varphi(y))\, dydx + \lambda \int_{\Omega_x} \delta\varphi(x)\|\nabla\varphi(y)\|dx \quad (14)$$

The images are inverted to remove noise from the image. Area, width and depth of thyroid gland are calculated from the segmented thyroid using imtool function in MATLAB.

F: Classifiers- Following types of classification methods are discussed:

1. Support vector machine- Support vector machine (SVM) are basically linear classifiers. SVM is widely accepted classifier, considered very effective for pattern recognition, machine learning and bioinformatics (protein classification and cancer classification) [10]. In SVM, a separator hyper plane between two classes is chosen to minimize the functional gap between two classes, the training data on the marginal sides of this optimal hyper plane called support vector. The Learning process is the determination of those support vectors. For non linearly- separable data, SVM maps the input vector from input space to some normally higher dimension feature space given by kernel function. The kernel function is an important step is successful design of a SVM in specific classification task.

2. K-nearest neighbour- A K-nearest neighbour (KNN) classifier is also in the system proposed, offering a good alternative when simplicity and ease of the training phase are the predominant issues. The KNN method is nonparametric and generally effective classification approach. [10]

3. Bayesian classifier- Bayesian classifiers have been used in many areas of medicine. For example, to built a Bayesian classifier to predict breast cancer. And also given that sonographic features predictive of malignancy have been extensively studied and the sensitivity and specificity of these features for malignancy are readily available [12]. In simple terms, a naïve bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature given the class variable. Also bayes classifier is that it only requires a small amount of training data to estimate the parameters are necessary for classification.

Table 1 Comparative Analysis of Classifiers

Classifiers	Advantages	Disadvantages
SVM	i) Ability to handle large feature spaces, ii) Robust, iii) Deliver unique solution when optimality problem is convex.	i) the lack of transparency of results ii) training is very slow on datasets[11], iii) Not work very well on multiple classes.
KNN	i) Very simple classifier. ii) It allows easy and fast incorporation of new data into an existing trained system.	i) It requires large memory size ii) Heavy computational load.

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 5, May 2012

Bayesian	i) The classification is clear and well founded decision theory [berger,1985] to help one use that output.	i) Information theoretically infeasible. ii) Computationally infeasible. iii) Unautomatic.

G: Performance Measure- Quantitative measurement of classification accuracy is calculated in term of true positive (TP), true negative (TN), false positive (FP), false negative (FN) with respect to the ground truth. Performance metrics calculation:

$$PPV = \frac{TP}{TP + FP} \quad (15)$$

$$NPV = \frac{TN}{TN + FN} \quad (16)$$

$$\text{Specificity } SP = \frac{TN}{TN + FP} \quad (17)$$

$$\text{Sensitivity } SE = \frac{TP}{TP + FN} \quad (18)$$

$$\text{Accuracy} = 100*((TP+TN)/n) \quad (19)$$

$$GM = \text{sqrt}(SP+SE) \quad (20)$$

III. RESULTS

Total no. of 13 images was used. Where 8 images were malignant (cancerous) and 5 images were benign (non-cancerous). And total 13 features were extracted with the help of GLCM. Result shown that the SVM classifier was used these 13 texture features and distinguishing the malignant cancerous nodule and benign noncancerous nodules with up to 84.62%.

Table 2 Shows the TP, TN, FN, FP, Specificity, Sensitivity and GM for SVM, KNN And Bayesian Classifiers

Selected features	SVM	KNN	Bayesian
TP	8	6	0
TN	3	0	5
FN	2	2	8
FP	0	5	0
Specificity	1.00	0.00	1.00
Sensitivity	0.80	0.75	0.00
GM	1.34	0.87	1.00

Table 3 Shows the Percent of Accuracy Using Classification Methods

Classifiers	ACCURACY
SVM	84.62%
KNN	46.15%
Bayesian	38.46%

According to the table 3 SVM shows the best classification accuracy as compare to the others. Fig 3 shows the bar chart. Representing the performance of Bayesian, KNN, and SVM.

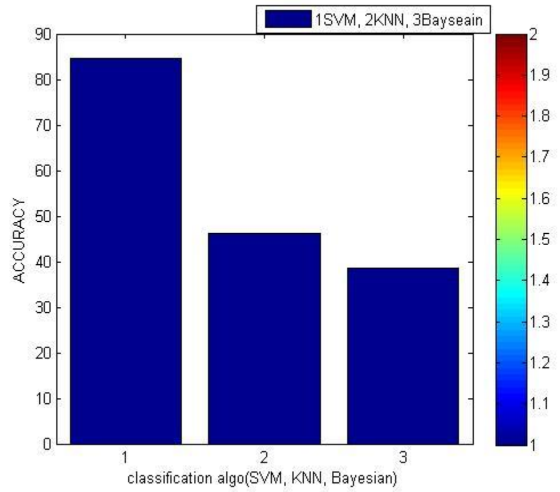

Fig 3 Performance Analysis of Classifiers of Thyroid Using Bayesian, KNN and SVM.

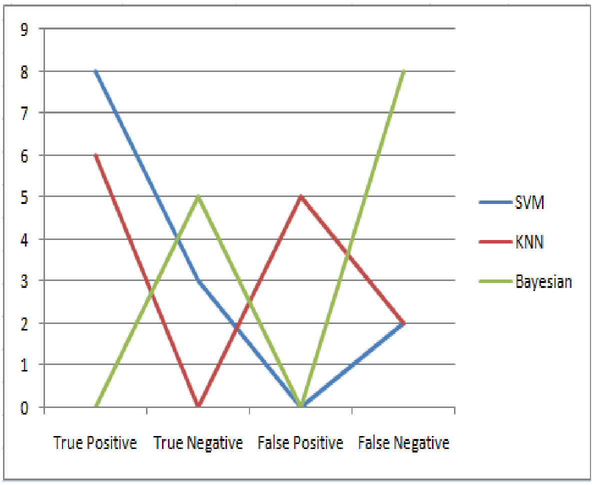

Fig 4 Performance Graph Of Classifiers

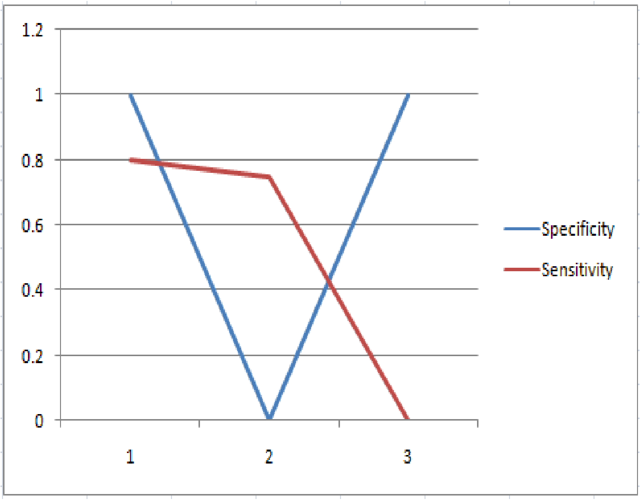

Fig 5 Performance Graph of Sensitivity and Specificity of Classifiers

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 5, May 2012

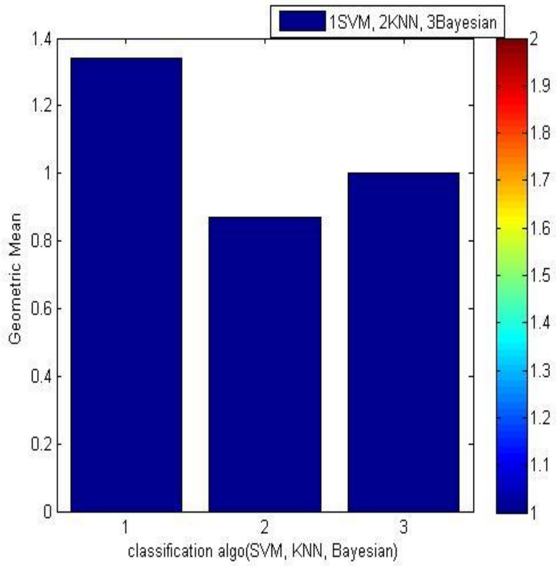

Fig 6 Shows the Percentage of Geometric Mean of Classifiers

From the analysis, it is evident that the classification performance of SVM is higher as compare to KNN and Bayesian.

IV CONCLUSION

The establishment of US as a leading tool in the medical applications is directly associated with the evolution of imaging technology employed in medicine and biology. US images are a widely used tool for clinical diagnosis, although it is time consuming for physicians to manually segment the thyroid nodule. This work proposed the method of classification of thyroid nodule (benign (cancerous) and malignant (non-cancerous) nodules) using the Bayesian, KNN and SVM [8], segmentation of thyroid nodules. From the experimental results, it is concluded that SVM gives the better classification accuracy than KNN and Bayesian. The utilization of new and more efficient classifiers could improve the accuracy performance thyroid region. The features served as input into all classifiers in this study has proven to posses high discriminatory attributes [4] however the generation of more feature may enhance the evaluation procedure accuracy.

REFERENCES

[1] Ms. Nikita Singh, Mrs Alka Jindal "A Survey of Different types of Characterization Technique in Ultra sonograms of the Thyroid Nodules" Published in international journal for computer science and informatics volume 1 issue 4. 2012.

[2] Maria E. Lyra, Nefeli Lagopati, paraskevi charalabatou, Efrosini vasoura, Aristides, costas Georgousopoulos, "Texture characterization in Ultrasound of the thyroid gland". Information Technology and Applications in Biomedicine (ITAB),

[3] Mary C. Frates, Carol B.Benson, "Management of thyroid nodules detected at US: society of radiologists in ultrasound consensus conference statement" radiology 2005.

[4] D.Selvathi, V.S.Sharnitha "Thyroid classification and segmentation in ultrasound imaging using machine learning algorithm" international conference on signal processing, communication, computing and network techniques (ICSCCN) 2011.

[5] Shawn Lankton, Allen Tannenbaum "Localizing Region-Based Active Contours" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 11, NOVEMBER 2008.

[6] Published in final edited form as:

[7] IEEE Trans Image Process. 2008 November; 17(11): 2029–2039.

[8] [8] Chuan-Yu Chang, "Classification of the thyroid nodules using SVM", Neural network, 2008. (IEEE world congress on computational 2008.

[9] D.E Maroulis, M.A Savelonas, S.A Karkanis, D.K. Iakovidis, N.Dimitropoulos, "Computer- aided thyroid nodule detection in ultrasound images,'' proc 18th IEEE Symposium on computer based medical system, 2005.

[10] Eystratios G. Keramidas "TND: A thyroid nodule detection system for analysis of ultrasound images and videos." Springer 2010.

[11] Liujie "SVM-KNN Discriminative Nearest Neighbour Classification for Visual Category Recognition" Proceeding of the 2006 IEEE (CVPR'06).

[12] Yueyi I. Liu "Bayesian classifier for differentiating malignant and benign nodules using sonography features" AMIA 2008 symposium proceeding page-419.

AUTHOR BIOGRAPHY

Nikita Singh is a student of ME CSE (IS) at PEC University of technology, Chandigarh, INDIA. She has received B.Tech Degree in CSE RIET Jaipur Rajasthan, INDIA. Her main research interest includes Image processing.

Alka Jindal is working as Assistant Professor in IT department of *PEC* University of Technology. Her main research interest includes image processing also a reviewer to many international journals.