

EVALUATION OF TEXTURAL FEATURE EXTRACTION FROM GRLM FOR PROSTATE CANCER TRUS MEDICAL IMAGES

R.Manavalan¹ and K.Thangavel²

¹Department of Computer Science and Applications, KSR College of Arts and Science,
Tiruchengode, Namakkal, Tamilnadu, India

²Department of Computer Science, Periyar University, Salem 11

ABSTRACT

Ultrasound imaging is one of the promising techniques for early detection of prostate cancer. There are five steps involved in processing the ultrasound images such as pre-processing, segmentation, feature extraction, feature selection, and classification. In this paper, the Transrectal Ultrasound (TRUS) images are pre-processed with M3 filter. Then it is segmented by using DBSCAN clustering after applying morphological operators, in order to extort the prostate region. It is proposed to extract the features by using Gray Level Run Length Matrix (GLRLM) for different direction from the segmented region. To classify the images into benign or malignant, Support Vector Machine (SVM) is adapted to evaluate the performance of the proposed method through classification. Over 5500 digitized TRUS images of prostate are used for the experimental analysis. The results obtained using the classification showed that texture features based on GLRLM by using combined directions ($\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$) distinguish between malignant and benign on TRUS images, with highest accuracy 85% where as sensitivity and specificity of 82% and 100% respectively.

KEYWORDS: SVM, GLRLM, Prostate, DBSCAN, M3-Filter, Sensitivity, Specificity, Accuracy

I. INTRODUCTION

In Early detection of cancers such as prostate cancer, breast cancer, lungs cancer, cervical cancer and blood cancer, digital image plays a vital role. Prostate cancer is now most frequently diagnosed with one in every eleven men [19]. It is in the second position in cancer-related cause of death only for male [16]. To diagnosis and prognosis, Ultrasound imaging method is recommended in general. Because of the result of reflection, refraction and deflection of ultrasound waves from different types of tissues with different acoustic impedance, an accurate detection of region of interest in ultrasound image is crucial.

Normally, the contrast in ultrasound image is very low and boundary between region of interest and background are more uncertain [15]. Different types of noises affect the mages. The ultrasound image with speckle noise and weak edges make the identification of prostate region more difficult. Hence, the analysis of ultrasound image is more challenging one.

In General, there are two common use of ultrasound medical imaging: (i) To guide the oncologist in the biopsy procedure and (ii) To establish the volume of the prostate. This method is used in diagnosing for more than 50 years. To classify the proteomic into benign or malign, the statistical textural features are extracted from the segmented region. This paper emphasis the extraction of textural features using GRLM for various directions and Support Vector Machine (SVM) is adopted for classification. An overview of the proposed system is presented below.

II. OVERVIEW OF THE SYSTEM

There are five stages in the proposed system consists, such as acquisition of Transrectal Ultrasound (TRUS) image of prostate, pre-processing, segmentation, feature extraction and classification. For automatic detection of prostate tumour in TRUS image, Computer Aided Diagnosis (CAD) system is developed. It can provide the valuable viewpoint and accuracy of earlier prostate cancer detection. The overview of the proposed system is depicted in figure 1.

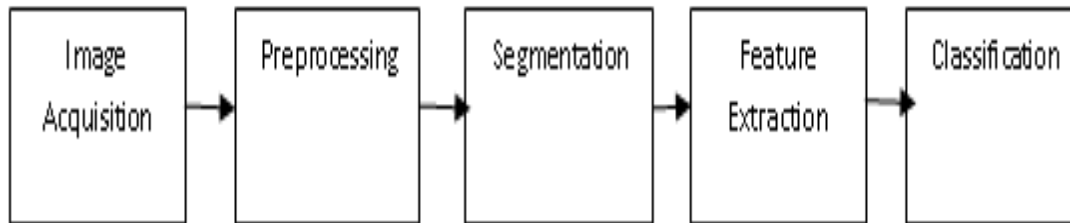


Figure 1: CAD System

The subsequent sections describe each stage in detail. The rest of the paper is organized as follows: section 2 deals about Image Acquisition and pre-processing of original TRUS medical image of prostate for noise removal and enhance the contrast of the image. The DBSCAN clustering with morphological operators for locating the Region of Interest (ROI) from TRUS prostate medical image is described in section 3. The feature extraction through GLRLM for the classification of prostate cancer is illustrated in section 4. The SVM classifier is discussed in section 5. The experimental analysis and discussion are presented in the section 6. Section 7 concludes this work with further directions.

III. IMAGE ACQUISITION AND PREPROCESSING

Among the different medical imaging modalities, Ultrasound imaging is a widely used technology for prostate cancer diagnosis and prognosis [20]. Less than 30 minutes is required for each case of image acquisition processes. The TRUS images of the prostate gland are acquired while the TRUS probe is supported and manipulated by the TRUS Robot which uses joystick located next to the console. By rotating the probe about its axis with minimum displacement and deformation, the entire prostate is scanned. The surgeon can manually adjust the probe depth. Exact recording of images and corresponding TRUS frame coordinates were obtained. The gathered information is used in offline segmentation for the ultrasound image of the prostate gland. Via frame grabber, Image acquisition is done from ultrasound device. Using a series of TRUS images, image of the prostate can be generated [21]. Because of poor image contrast, speckle noise, and missing or diffuse boundaries in the TRUS images, the ultrasound images are very difficult to segment.

The speckle noise is commonly found in the ultrasound medical images and hence M3-Filter is used to generate a despeckled image [14]. Data dropout noise is common and this is generally referred to as speckle noise [18]. An error in data transmission causes this noise. The corrupted pixels are either set to the maximum value. The information of edges is needed for the proper segmentation. The imtophat filter is used to create required edges once the noise is removed to enhance the contrast of the image. The following section discusses the segmentation of this enhanced image using DBSCAN clustering.

IV. IMAGE SEGMENTATION

Enhanced images are changed into binary image where each pixel is restricted to a value of either 0 or 1 to simplify the image segmentation task. This process is accomplished by local adaptive thresholding method [22]. Morphological operators such as opening and closing are applied on these binary images [2, 9]. In TRUS medical images, isolating prostate from other white region in the threshold image is a difficult task. By using opening and closing operators, the thresholded image can be reconstructed with a large disk structuring element in order to isolate the object with high

likelihood to be part of the prostate [9]. Pixels of the thresholded images are grouped by using DBSCAN with the aim of separating background from TRUS image to target possible prostate [7, 17].

By clustering, it takes a binary (segmented) image, and delineates only significantly important regions. The outcome that expected is desired boundary of the TRUS prostate image. A sample image and its segmented image are provided in figure 2(a) and figure 2(b) respectively. The DBSCAN Algorithm for segmentation is detailed in figure 3.

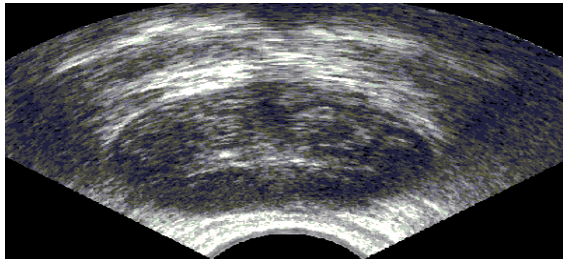


Figure 2(a) sample image



Figure 2(b) segmented image

DBSCAN Algorithm

INPUT: Enhanced TRUS prostate image

OUTPUT: Segmented image which contains only prostate

Step1: Set epsilon (eps) and minimum points (minPts).

Step2: Starts with an arbitrary starting point that has not been visited and then finds all the neighbor points within distance eps of the starting point.

Step3: If the number of neighbors is greater than or equal to minPts, a cluster is formed. The starting point and its neighbors are added to this cluster and the starting point is marked as visited. The algorithm then repeats the evaluation process for all the neighbors recursively.

Step4: If the number of neighbors is less than minPts, the point is marked as noise.

Figure 3: DBSCAN Algorithm

After fixing the two parameters eps and minPts, DBSCAN starts to cluster data points from an arbitrary point q [23]. It begins by finding the neighborhood of point q (all points that are directly density reachable from the point q). For an image, we start with left-top pixel (not necessarily a corner pixel, any arbitrary pixel can be chosen for first iteration) as our first point in the subimage. We look for first pixel satisfying the core pixel condition as a starting (seed) point. If the neighborhood is sparsely populated, i.e. it has fewer than MinPts points, and then the point q is labelled as noise [24]. Otherwise, a cluster is initiated and all points in neighborhood of the point q are marked by new cluster. Next the neighborhoods of the neighborhood q are examined iteratively to check if they can be added into the cluster [12]. If a cluster cannot be expanded further, DBSCAN chooses another arbitrary unlabeled point and repeats the process to form another cluster. This procedure is iterated till all pixels in the image have been identified as noise or with a cluster of prostate pixels.

V. FEATURE EXTRACTION

Texture analysis is normally grouped into four categories: model-based, statistical-based, structural-based, and transform-based methods. Model-based methods are based on the concept of predicting pixel values based on a mathematical model [14]. Statistical methods describe the image using numerical analysis of pixel intensity values. Transform methods generally perform some kind of

modification to the image, obtaining a new “response” image that is then analyzed as a representative proxy for the original images. Structural approaches seek to understand the hierarchical structure of the image. Texture is one of the important characteristics used in identifying an object in an image and to discriminate the images [1]. The texture coarseness or fineness of an image can be interpreted as the distribution of the elements in the matrix [10, 11]. In this paper, segmented image (ROI) is utilized to construct GLRLM using the various directions, and then seven features are extracted from each GLRLM for the classification of TRUS prostate image.

5.1. Gray Level Run Length Matrices (GLRLM)

The GLRLM is based on computing the number of gray-level runs of various lengths [3,4]. A gray-level run is a set of consecutive and collinear pixel points having the same gray level value. The length of the run is the number of pixel points in the run [3, 4]. The Gray Level Run Length matrix is constructed as follows:

$$R(\theta) = (g(i, j) | \theta), 0 \leq i \leq Ng, 0 \leq j \leq Rmax;$$

where Ng is the maximum gray level and $Rmax$ is the maximum length. The following textural features are measured for the TRUS medical image of prostate. Let $p(i, j)$ be the number of times there is a run of length j having gray level i . There are five Run Length Matrix based features computed for 4 directions of run ($0^\circ, 45^\circ, 90^\circ, 135^\circ, \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ run). $i=1, \dots, Ng$ and $j=1, \dots, Nr$.

Table 1: Features and their Formulae

S. No.	Features	Formulae
1	Short Run Emphasis (SRE)	$\frac{1}{n} \sum_{i,j} \frac{p(i, j)}{j^2}$
2	Long Run Emphasis (LRE)	$\frac{1}{n} \sum_{i,j} j^2 p(i, j)$
3	Grey Level Non-uniformity (GLN)	$\frac{1}{n} \sum_i \left(\sum_j p(i, j) \right)^2$
4	Run Length Non-uniformity	$\frac{1}{n} \sum_j \left(\sum_i p(i, j) \right)^2$
5	Run percentage (RP)	$\sum_{i,j} \frac{n}{p(i, j) j}$
6	Low Grey Level Run Emphasis (LGLRE)	$\frac{1}{n} \sum_{i,j} \frac{p(i, j)}{i^2}$
7	High Grey Level Run Emphasis (HGLRE)	$\frac{1}{n} \sum_{i,j} i^2 p(i, j)$

VI. SVM CLASSIFIER

Support vector machine is based on modern statistical learning theory. It is a useful method for classification. It is a leading method for solving non-linear problem. The SVMs were introduced by Vladimir Vapnik [25] and his colleagues. The earliest mention was in (Vapnik, 1979). Classification task usually involves training and testing data which consist of some data instances [5,6]. One target value and several attributes are contained into each instance of training set. Given the attributes of the testing set, the goal of SVM is to produce a model which predicts target value of data instances. Classification in SVM is an example of Supervised Learning. Whether the system is performing in the right way or not is indicated with the help of known labels. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. A

step in SVM classification involves identification as which are intimately connected to the known classes [5].

SVM use the training data to create the optimal separating hyperplane between the classes. The optimal hyperplane maximizes the margin of the closest data points. A good separation is achieved by the hyperplane that has the largest distance to the nearest training features of any class (so-called functional margin). Maximum-margin hyperplane and margins for a SVM trained with samples from two classes. Samples on the margin are called the support vectors [8]. SVM divides the given data into decision surface. Decision surface divides the data into two classes like a hyper plane. Training points are the supporting vector which defines the hyper plane. The basic theme of SVM is to maximize the margins between two classes of the hyper plane (Steve, 1998).

Basically, Support Vector Machines are used to solve binary classification problems only. They are extended to handle multi-class problems also [6]. The idea is to decompose the problem into many binary-class problems and then combine them to obtain the prediction. To allow multi-class classification, SVM uses the one-against-one technique by fitting all binary sub classifiers and finding the correct class by a voting mechanism. The "one-against-one" approach (Knerr et al., 1990) is implemented in SVM for multiclass classification. If K is the number of classes, then $K(K - 1)/2$ binary classifiers are constructed and trained to separate each pair of classes against each other, and uses a majority voting scheme (max-win strategy) to determine the output prediction. For training data from the i th and the j th classes, we solve the following two-class classification problem [13].

Let us denote the function associated with the SVM model of $\{c_i, c_j\}$ as:

$$g(x)_{ij} = \text{sign}(f(x)_{ij})$$

An unseen example, x , is then classified as:

$$f(x) = \arg \max_i \sum_{i=1}^K \sum_{j=1 \wedge i \neq j}^K V_{ij}(x)$$

$$\text{where } V_{i,j}(x) = \begin{cases} 1 & \text{if } g_{ij}(x) = 1 \\ 0 & \text{if } g_{ij}(x) = -1 \end{cases}$$

Each feature set is examined using the Support Vector Machine classifier. In classification, we use a voting strategy: each binary classification is considered to be a voting where votes can be cast for all data points x in the end point is designated to be in a class with the maximum number of votes. In case that two classes have identical votes, though it may not be a good strategy, now we simply choose the class appearing first in the array of storing class names. The objective of any machine capable of learning is to achieve good generalization performance, given a finite amount of training data, by striking a balance between the goodness of fit attained on a given training dataset and the ability of the machine to achieve error-free recognition on other datasets. With this concept as the basis, support vector machines have proved to achieve good generalization performance with no prior knowledge of the data. The optimal separating hyperplane can be determined without any computations in the higher dimensional feature space by using kernel functions in the input space. Commonly used kernels include:

Linear Kernel: $K(x, y) = x \cdot y$

Radial Basis Function (Gaussian) Kernel: $K(x, y) = e^{-(\|x - y\|^2 / 2\sigma^2)}$

Polynomial Kernel: $K(x, y) = (x \cdot y + 1)^d$

VII. EXPERIMENTAL ANALYSIS AND DISCUSSION

The proposed method has been implemented using MATLAB. In order to evaluate this work, experiments conducted over 5500 image with normal and abnormal cases. The total of 5500 TRUS medical images of prostate measuring 250x250 pixels is transformed into a Grey Level Run Length matrix (GLRLM) along the directions of $(0^\circ, 45^\circ, 90^\circ, 135^\circ, \{0^\circ, 45^\circ, 90^\circ, 135^\circ\})$ producing 5 run length matrices. From each GLRLM, 7 texture features are extracted. Extracted features are fed into SVM to construct the prediction model. The performance of the prediction was evaluated in terms of sensitivity, specificity, accuracy, the respective formulæ are given in Table 2.

Table 2.Formula for Measures

Measures	Formula
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
Accuracy	(TP+TN)/(TP+FP+TN+FN)

Accuracy measures the quality of the classification. It takes into account true and false positives and negatives. Accuracy is generally regarded with balanced measure whereas sensitivity deals with only positive cases and specificity deals with only negative cases. TP is number of true positives, FP is number of false positives, TN is number of true negatives and FN is number of false negatives. A confusion matrix provides information about actual and predicted cases produced by classification system. The performance of the system is examined by demonstrating correct and incorrect patterns. They are defined as confusion matrix in Table 3.

Table 3.Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TP - predicts cancer as cancer.
 FP - predicts cancer as normal.
 TN - predicts normal as normal.
 FN - predicts normal as cancer.

The higher value of both sensitivity and specificity shows better performance of the system. In order to quantify the influence of the direction, the GLRLM is constructed with direction choices for feature extraction. The results of performance measures for each feature set have been tabulated in table 4.

Table 4: results of statistical parameters for various directions

Direction	sensitivity	Specificity	Accuracy
GLRLM_000	0.75188	1	0.82945
GLRLM_045	0.800901	1	0.816667
GLRLM_090	0.779528	0.997886	0.821667
GLRLM_135	0.735294	1	0.814
GLRLM_ALL	0.824956	1	0.85

From the computational results we can notice that the features extracted from among GLRLM, the combined direction $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ give better results using three quantitative performance measures with higher sensitivity and specificity rate. The results of individual performance measures for various directions are depicted in the figure 5(a) – 5(c).

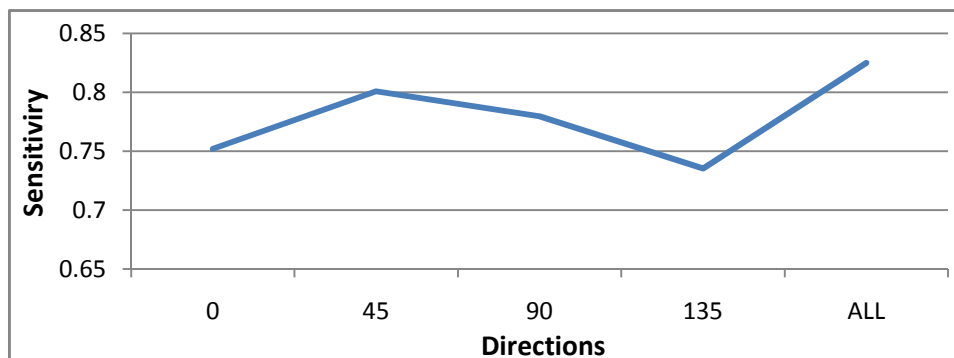


Figure 5(a): performance of sensitivity for various directions

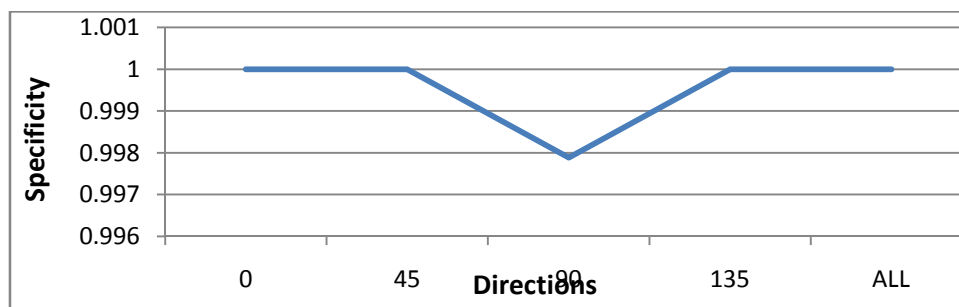


Figure 5(b): specificity for various directions

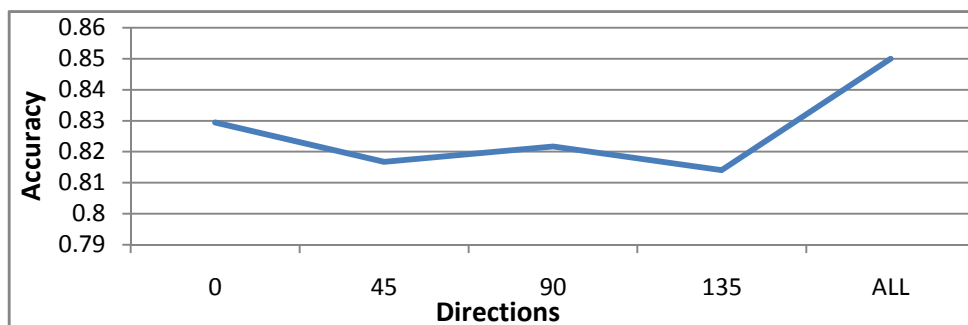


Figure 5(c): performance of accuracy for various directions

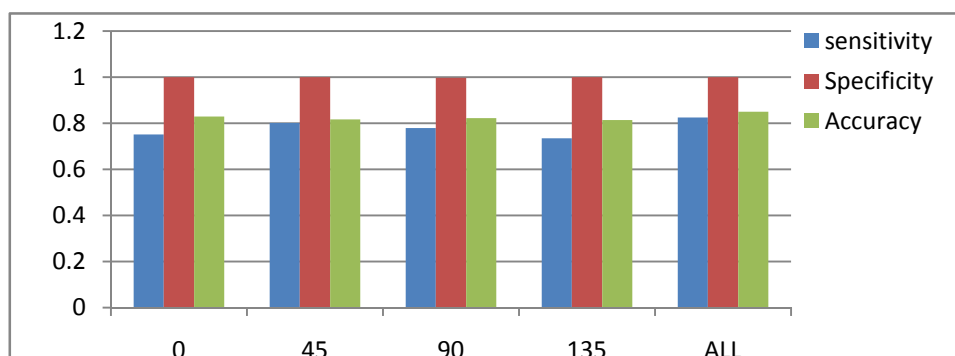


Figure 6: Relative performance measures for various directions

The performance of relative performance measures is depicted in figure 6 and observed that the features extracted from the combination of all directions ($\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$) producing the best one from the result of classification. It achieves 85% of the accuracy, 82% of the sensitivity and 100% of the specificity.

VIII. CONCLUSION

TRUS imaging is one of the best methods for detecting prostate cancer. In some cases, radiologists face difficulties in directing the tumours. To assist the oncologist and improve the accuracy of detection, the method presented in this paper. The proposed approach effectively addresses the feature extraction problem TRUS image analysis of prostate and can also be applied to other image analysis applications. The proposed method of feature extraction from GLRLM with various directions are compared and analyzed. The performance of the texture is measured using quantitative performance measures. The performance of proposed method is tested with ultrasound image regard to prostate. The computational results showed that texture features based on GLRLM by using combined directions ($\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$) distinguish between malignant and benign on TRUS images, with accuracy levels higher than texture features based on others. The maximum accuracy rate for normal and cancer classification is achieved 85%. The GLRLM features combined with other statistical

features to improve the results of classification of TRUS prostate images may be considered for future work.

REFERENCES

- [1] Amadasun, M. and King, R., (1989) 'Textural features corresponding to textural properties', IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 5, pp. 1264 - 1274.
- [2] Anil K. Jain. (1989) Fundamentals of Digital Image Processing: Prentice Hall.
- [3] K.Thangavel, M.Karnan, R.Sivakumar and A. Kaja Mohideen " Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms" AIML Journal, Volume (5), Issue (3), pp.29-40., September, 2005.
- [4] Jong Kook Kim, Jeong Mi Park, Koun Sik Song and Hyun Wook Park. "Texture Analysis and Artificial Neural Network for Detection of Clustered Microcalcifications on Mammograms" IEEE, pp.199 – 206, 1997.
- [5] Devendran V et. al.,(2008) 'Texture based Scene Categorization using Artificial Neural Networks and Support Vector Machines: A Comparative Study,' ICGST-GVIP, Vol. 8, Issue IV, pp. 45-52.
- [6] Duda, R., Hart, P. And Stork, D. (2001), 'Pattern Classification': John Wiley and Sons.
- [7] Ester, M., Kriegel, H.P., Sander, J., and Xu, X.(1996) 'A density-based algorithm for discovering clusters in large spatial databases with noise'. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, Portland: AAAI Press. pp. 226-231.
- [8] Fatima Eddaoudi, and Fakhita Regragui.:(2011) 'Masses Detection Using SVM Classifier Based on Textures Analysis' Applied Mathematical Sciences, Vol. 5, No. 8, pp. 367 – 379.
- [9] Gonzalez, R. and Woods, R. (2002). Digital Image Processing. 3rd Edn., Prentice Hall Publications, pp. 50-51.
- [10] Haralick, R.M. , Shanmugan, K.. and Dinstein, I.(1973) 'Textural Features for Image Classification', IEEE Tr. on Systems, Man, and Cybernetics, Vol SMC-3, No. 6, pp. 610-621.
- [11] Haralick, R.M. (1979) 'Statistical and Structural Approaches to Texture', Proceedings of the IEEE, Vol. 67, No. 5, pp. 786-804.
- [12] Massimiliano Pavan and Marcello Pelillo: (2007) 'Dominant Sets and Pairwise Clustering', IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 29, No. 1, pp. 167 – 172.
- [13] Metehan Makinaci:(2005) ' Support Vector Machine Approach for Classification of Cancerous Prostate Regions', World Academy of Science, Engineering and Technology. pp. 166 – 169.
- [14] Mohamed, S.S. and Salama M.M. (2005) 'Computer Aided diagnosis for Prostate cancer using Support Vector Machine' Publication: Proc., medical imaging conference, California, SPIE Vol. 5744, pp. 899-907.
- [15] Thangavel, K., Manavalan, R. and Laurence Aroquiaraj . I. (2009), Removal of Speckle Noise from Ultrasound Medical Image based on Special Filters: Comparative Study, ICGST-GVIP Journal, ISSN 1687-398X, Volume (9), Issue (III), pp. 25-32.
- [16] Scardino, PT.,(1989) 'Early detection of prostate cancer'. ; Urol Clin North Am., Vol. No 16(4) , pp: 635-55.
- [17] Sinan Kockara, Mutlu Mete, Bernard Chen, and Kemal Aydin:(2010). 'Analysis of density based and fuzzy c-means clustering methods on lesion border extraction in dermoscopy images', From Seventh Annual MCBIOS Conference: Bioinformatics Systems, Biology, Informatics and Computation Jonesboro, AR, USA. pp. 19 - 20.
- [18] Sinha, G.R., Kavita Thakur and Kowar, M.K. (2008). "Speckle reduction in Ultrasound Image processing", Journal of Acoustical. Society of India, Vol. 35, No. 1, pp. 36-39.
- [19] Yanong Zhu, Stuart Williams, and Reyer Zwiggelaar (2006), 'Computer Technology in Detection and Staging of Prostate Carcinoma: A review, Medical image Analysis' Vol. No.10, pp 178-199
- [20] Cancer Facts and Figures. American Cancer Society. [Online] <http://www.cancer.org>, 2002.

- [21] Huber, J.S. Qiyu Peng Moses, W.W. Reutter, B.W. Pouliot, J. Hsu, I.C (2011). "Development of a PET-Transrectal Ultrasound Prostate Imaging System" Nuclear Science, IEEE Transactions pp. 674 – 681.
- [22] Xiao-Ping Zhang, and Mita D. Desai.(2001) "Segmentation of Bright Targets Using Wavelets and Adaptive Thresholding" , IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 10, NO. 7, pp 1020 – 1030
- [23] Sait Suer, Sinan Kockara and Mutlu Mete(2011). "An improved border detection in dermoscopy images for density based clustering" biomedcentral, (Suppl 10):S12 pp. 1-10.
- [24] Xin Wang and Howard J. Hamilton DBRS (2003) " A Density-Based Spatial Clustering Method with Random Sampling" Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, Vol 2637/2003, 569.
- [25] Muntean, M. Vălean, H. Ileană, I. and Rotar, C.(2010). "Improving classification with Support Vector Machine for unbalanced data" Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference ,Vol.3 pp. 1 – 6.

Authors

R. MANAVALAN Obtained M.Sc.,Computer Science from St.Joseph's College of Bharathidasan University,Trichy, Tamilnadu, India, in the year 1999, and M.Phil., in Computer Science from Manonmaniam Sundaranar University, Thirunelveli, Tamilnadu, India in the year 2002. He works as Asst.Prof & Head, Department of Computer Science and Applications, KSR College of Arts and Science, Thiruchengode, Nammakal, Tamilnadu, India. He pursues Ph.D in Medical Image Processing. His areas of interest are Medical image processing and analysis, soft computing, pattern recognition and Theory of Computation.



THANGAVEL KUTTIANNAN received the Master of Science from Department of Mathematics, Bharathidasan University in 1986, and Master of Computer Applications Degree from Madurai Kamaraj University, India in 2001. He obtained his Ph. D. Degree from the Department of Mathematics, Gandhigram Rural University in 1999. Currently he is working as Professor and Head, Department of Computer Science, Periyar University, Salem. His areas of interests includes medical image processing, artificial intelligence, neural network, fuzzy logic, data mining, pattern recognition and mobile computing

