INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) N175–N185 doi:10.1088/0031-9155/50/15/N02

NOTE

Prostate cancer multi-feature analysis using trans-rectal ultrasound images

S S Mohamed[1], M M A Salama[1], M Kamel[1], E F El-Saadany[1], K Rizkalla[2] and J Chin[2]

[1] Electrical and Computer Engineering Department, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[2] University of Western Ontario, 1151 Richmond Street, Suite 2, London, Ontario N6A 5B8, Canada

Received 16 November 2004, in final form 20 May 2005
Published 20 July 2005
Online at stacks.iop.org/PMB/50/N175

Abstract

This note focuses on extracting and analysing prostate texture features from trans-rectal ultrasound (TRUS) images for tissue characterization. One of the principal contributions of this investigation is the use of the information of the images' frequency domain features and spatial domain features to attain a more accurate diagnosis. Each image is divided into regions of interest (ROIs) by the Gabor multi-resolution analysis, a crucial stage, in which segmentation is achieved according to the frequency response of the image pixels. The pixels with a similar response to the same filter are grouped to form one ROI. Next, from each ROI two different statistical feature sets are constructed; the first set includes four grey level dependence matrix (GLDM) features and the second set consists of five grey level difference vector (GLDV) features. These constructed feature sets are then ranked by the mutual information feature selection (MIFS) algorithm. Here, the features that provide the maximum mutual information of each feature and class (cancerous and non-cancerous) and the minimum mutual information of the selected features are chosen, yeilding a reduced feature subset. The two constructed feature sets, GLDM and GLDV, as well as the reduced feature subset, are examined in terms of three different classifiers: the condensed *k*-nearest neighbour (CNN), the decision tree (DT) and the support vector machine (SVM). The accuracy classification results range from 87.5% to 93.75%, where the performance of the SVM and that of the DT are significantly better than the performance of the CNN.

Table of abbreviations

TRUS	Trans-rectal ultrasound
ROI	Region of interest
GLDM	Grey level dependence matrix

GLDV	Grey level difference vector
MIFS	Mutual information feature selection
CNN	Condensed *k*-nearest neighbour
DT	Decision tree
SVM	Support vector machine
DRE	Digital rectal exam
PSA	Prostate specific antigen
CAD	Computer aided diagnosis
MI	Mutual information
KNN	*k*-nearest neighbour

1. Introduction

One of the highest cancer risks for men is prostate carcinoma. Since it is curable only at an early stage, early detection is highly crucial (Scardino 1989). However, existing diagnostic methods such as the digital rectal exam (DRE), the TRUS imaging system, the analysis of the prostate specific antigen (PSA) and the prostate volume have proven to be unreliable. For example, the common factor between DRE and TRUS is the dependence on the skills of the conducting physician and radiologist. Experienced doctors achieve more accurate prediction rates than inexperienced doctors due to the fact that prior knowledge is correlated to the success of applying these methods. This is the prime motivation for designing a novel system in which trained radiologists can provide feedback to augment knowledge of newer radiologists for decision making. To achieve this, the work in this note relies mainly on improving the computer aided diagnosis (CAD) techniques. When CAD is adopted as a tool in ultrasound imaging, features are constructed from the echoes after the echo processing to identify the tissue type present in the image. Such texture features include the statistical characteristics of the grey level distribution in the image for cancer diagnosis (Bassat *et al* 1993, Scheipers *et al* 2001) and fetal lung maturity (Bhanu *et al* 2002).

The image texture method is adopted in this work and the proposed algorithm is presented in figure 1.

The first stage in this work is identifying the high risk ROIs in the image. At the second stage, the statistical features are constructed from the ROIs. After the most salient features have been selected by the MIFS algorithm, three classifiers are used to determine the quality of the feature sets.

This note is organized as follows: section 2 explains the ROI identification algorithm, section 3 covers the feature construction algorithms, section 4 discusses the MIFS algorithm used for the feature selection and section 5 highlights the classifiers used in this study. The results and discussion are provided in section 6, and finally, the conclusion is given in section 7.

2. TRUS image ROI identification

The images for this investigation obtained from the University of Western Ontario (UWO) are derived from the Aloka 2000 ultrasound machine by using a broadband 7 MHz linear transducer and a field of view of approximately 6 cm. A set of 33 identified TRUS images, where suspected cancer regions are highlighted by an expert radiologist, is used.

The ROI segmentation is a crucial step for prostate cancer diagnosis and is performed by two methods: the first method depends on the contribution of an expert radiologist (Bassat *et al* 1993) who obtains accurate enough results, and the other method requires that

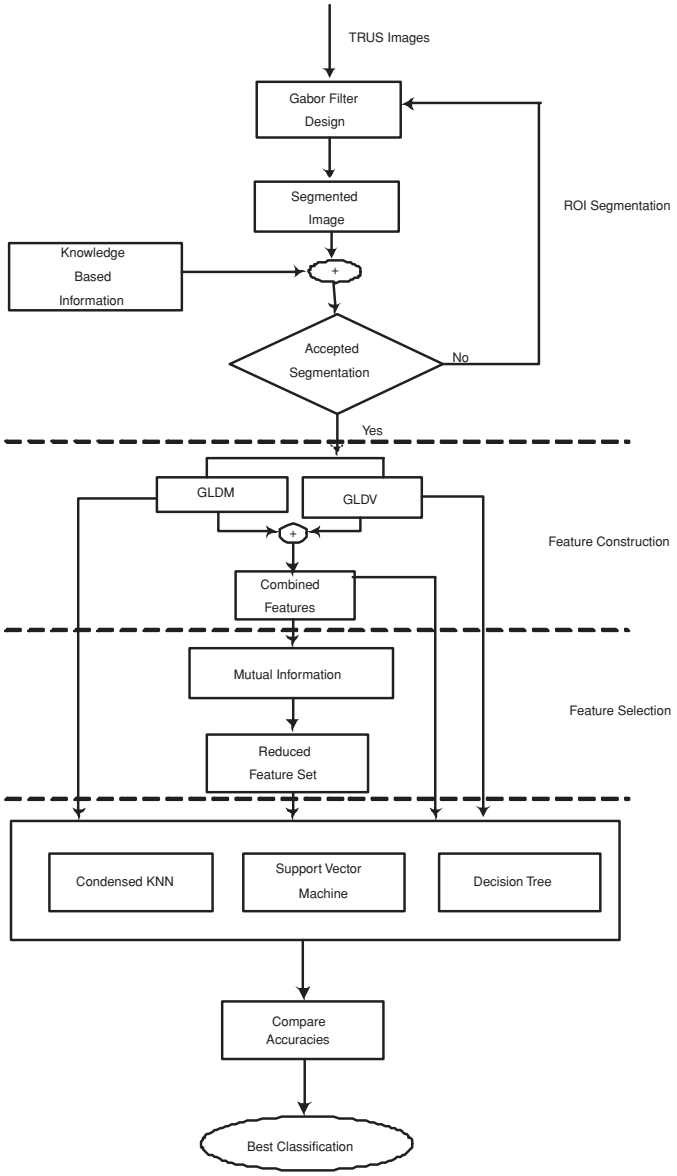

Figure 1. The proposed algorithm.

the image be divided into small square shaped ROIs, which is time consuming (Scheipers *et al* 2001). To assist the radiologist in the decision and to acquire fast and precise results, there is a great need for an automated ROI segmentation algorithm as well. A well-established texture segmentation algorithm based on multi-resolution analysis is used in this work to guarantee accuracy, and proves to be an excellent method for texture segmentation in the field of image processing (Clausi and Jernigan 2000).

This algorithm is applied for the first time to the prostate TRUS image to identify the ROIs. The main advantage is that no prior image assesment is required from the radiologist, that is, outlining the ROIs. Moreover the algorithm is efficient, since it does not require the

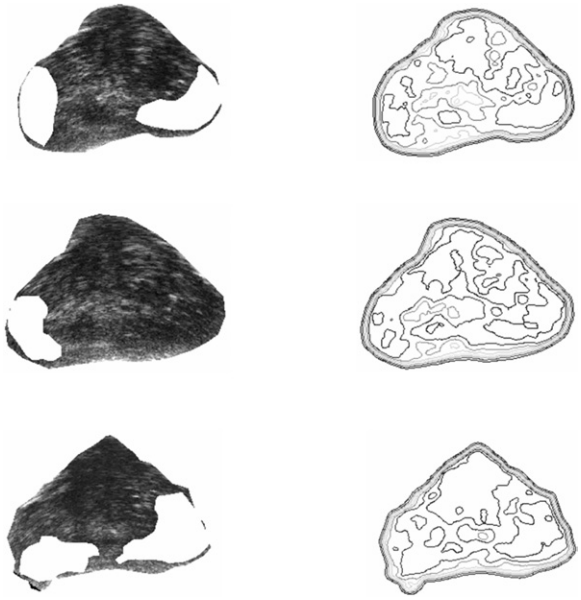

Figure 2. ROIs identified.

examination of the entire image. Furthurmore, the algorithm uses both the frequency domain information and the spatial domain information from the image. This process is achieved by applying the Gabor multi-resolution analysis that is capable of segmenting the image according to the frequency response of the pixels. The pixels that have a similar response are assigned to the same cluster. This process segments the TRUS image into several ROIs. The Gabor function is chosen for its high localization in both the spatial frequency domain and the spatial domain. This algorithm is applied to the available TRUS images and a sample of the professionally segmented TRUS images and their corresponding ROIs are shown in figure 2, where the sensitivity of the recognized ROIs is 88%.

3. TRUS image feature construction

The identified ROIs are subjected to further analysis, where the statistical features are extracted from the ROIs. This step is achieved by checking to see if each pixel is located inside the region under study. The process is repeated for all the ROIs in the image. Due to the fact that the two textures are differentiable by the human eye only if they have different second-order statistics, two second-order texture feature construction methods, the GLDM and the GLDV, are adopted.

3.1. ROI grey level dependence matrix

Some important texture features can be extracted from the identified ROIs by using the GLDM. This second-order statistical approach has been found to be effective in a number of ultrasound image applications such as fetal lung maturity determination (Bhanu *et al* 2002) and prostate cancer diagnosis (Scheipers *et al* 2001). GLDMs are matrices whose elements $p'(i, j)$ are the probabilities of finding a pixel which has a grey-tone i at a distance d and an angle φ from a pixel which has grey-tone j. A set of four features, detailed in Mohamed *et al* (2004), is constructed from the GLDM which are: contrast, entropy, energy, homogeneity.

3.2. ROI grey level difference vector

To overcome the high computational demand of the GLDM and to extract more information about the ROIs' texture, the GLDV is also applied. Its mathematical model is explained in Bhanu *et al* (2002), where it was one of the feature construction algorithms used for fetal long maturity. It is summarized as follows: let $I(x, y)$ be the image intensity function. For any given displacement $d \equiv (\Delta x, \Delta y)$, let p be the probability density of $I_d(x, y)$. If there are m grey levels, the GLDV has the form of m-dimensional vector whose ith component is the probability that $I_d(x, y)$ will have the value i. It is simple to compute p by counting the number of times each value of $I_d(x, y)$ occurs where Δx and Δy are integers. This vector is also used as a texture measure, and features such as contrast, entropy, angular second moment, mean and inverse difference moment are constructed.

4. Texture feature selection

The outputs of the ROI texture feature construction methods are the GLDM and GLDV feature sets that consist of the features of each set, as well as a combined feature set where the features from both methods are combined. This feature set can exhibit some redundancy and correlation between features (the curse of dimensionality) which motivates the use of feature selection techniques (Jain and Zongker 1997) such as MIFS.

Feature selection is used to select a subset of s features from a given set of p features, where $s < p$ without major degradation in the performance of the recognition system. Typically an exhaustive search guarantees a global optimal feature subset; however its time complexity is exponential to the dimension of the feature space, which causes this approach to be impractical even for a small number of features.

The traditional feature selection for classification is classifier-dependent. After a classifier is selected, the discriminatory power of a feature is proportional to the accuracy of the classifier, when that feature is employed. In contrast, classifier-independent feature selection techniques are feature driven regardless of the classifier, and this work is based on the MIFS (Battiti 1994).

MI is a powerful tool for evaluating the information content of each feature with respect to the output class, and with respect to each of the other features. MI is based on the probability distributions of the features and measures the amount of information one random variable contains about another (Battiti 1994). In this note, an intelligent feature selection method by which the best feature subset is chosen regardless of the chosen classifier is adopted. Classifier independence is the most important reason for selecting MI as the foundation of the feature selection method. It can be explained as follows: denote X as a random variable, describing a texture feature and C as a random variable, describing the diagnosis or the class. Then the mutual information $I(C; X)$ is a measure of the amount of information that feature X contains about the diagnosis C. Thus MI provides a criterion for measuring the effectiveness of a feature for the separation of the two classes. Interdependence between feature values and classes is proportional to the value of $I(C; X)$ and the interdependence among the features is denoted by $I(X_1; X_2)$ that should be minimized to avoid selecting two or more similar features. Therefore, the objective is to maximize $I(C; X)$, and minimize $I(X_1; X_2)$. The MI between the feature values and classes can be calculated as follows (Battiti 1994):

$$I(C; X) = H(C) - H(C|X), \tag{1}$$

where the entropy $H(C)$ measures the degree of uncertainty entailed by the classes, the conditional entropy $H(C|X)$ measures the degree of uncertainty entailed by the set of classes C given the set of feature values X. The entropy $H(C)$ depends mainly on the classes, and it

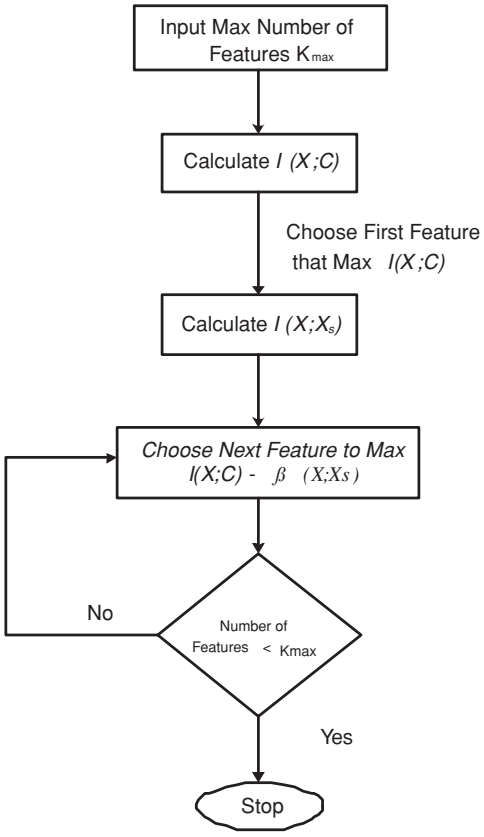

Figure 3. MIFS algorithm.

represents the upper bound for $I(C;\, X)$. The second term in the expression $H(C|X)$ represents the decrease in the certainty. The higher $H(C|X)$ the higher the uncertainty. The MI, $I(C;\, X)$, is maximum when the class is totally dependant on the feature, while it is minimum when the class and the feature are totally independent. The number of selected features should be minimized to get rid of redundancies and increase the efficiency of the classifier. In order to achieve this task, the MI among different features $I(X_1;\, X_2)$ is calculated as follows:

$$I(X_1;\, X_2) = H(X_2) - H(X_2 - X_1). \tag{2}$$

The MIFS algorithm has been proposed by Battiti (1994) and is shown in figure 3. The parameter $\beta > 0$ controls the relative importance of the MI between the candidate features and the previously selected features with respect to the MI of the class.

The MIFS algorithm is used to select three features from the combined set of features, where $\beta = 0.5$ as recommended by Battiti (1994) ($0.5 < \beta < 1$).

5. Classification

After the best representative feature subset has been chosen from the available feature set, it is a crucial step to test those features. The original texture feature sets, the combined feature set and the reduced feature set are examined by using three different classifiers: CNN, DT and SVM, which are summarized in this section.

5.1. Condensed k-nearest neighbour

The CNN classifier is applied in this investigation due to its simplicity and efficiency. The KNN rule assigns an unclassified sample to the same class as the k-nearest stored samples, and the CNN retains the same approach with less time complexity. It uses the consistent subset of the original sample set. A consistent subset is a subset which when used as a stored reference set for the KNN rule correctly classifies the rest of the sample set (Hart 1968).

5.2. Decision tree

Also the DT is a favoured classifier since there are logic rules for the generated classification. Moreover the DT relies at each stage of the training on the information gain. Therefore, the DT is a recursive structure for expressing classification rules. A DT is a system that uses a top-down, divide-and-conquer strategy that partitions a given set of objects into smaller and smaller subsets in steps that correspond to the growth of the tree (Webb 1988). In DT the data are classified to probably cancerous or probably non-cancerous. Construction of leaf nodes and decision nodes that identify the test condition for each of the available features, is called an attribute. All the possible conditions are examined at each stage, and the condition that maximizes the information gain is chosen as the next step in the construction of the decision tree.

5.3. Support vector machines

The SVM is found to be an influential methodology for solving a nonlinear classification problem (Duda *et al* 2001) such as the one described in this note. The SVM has been introduced within the framework of statistical learning theory and structural risk minimization, depending mainly on pre-processing the data to represent patterns in a higher dimensionality space, usually much higher than the original feature space. This is achieved with a suitable nonlinear mapping $\phi(\cdot)$ to a sufficiently high dimension. The data from two classes are consistently separated by a hyper-plane.

6. Results and discussion

The texture feature sets (the GLDM features, the GLDV features, the combined GLDM and GLDV features, and the reduced feature set that results from the MIFS) are tested by the three classifiers, CNN, DT and SVM, and the results are demonstrated in this section. A set of 96 regions obtained from 33 TRUS images that correspond to 21 patients is used in this study; 80 regions are adopted as a training set and a set of 16 regions is employed as the test set. There is a set of parameters defined in Webb (1988) to evaluate these classifiers: *false negative rate, false positive rate*, *sensitivity* and *specificity*. Specificity, sensitivity and overall accuracy are the measures used for testing the different classifiers applied in this work.

6.1. Condensed k-nearest neighbour (CNN)

The results of applying the CNN classifier to the GLDM features, GLDV features and the combined features are summarized in table 1. It is clear from the shown confidence matrices that all the features exhibit a better performance than that of either feature set. However, the results are less satisfactory due to the fact that the performance of the CNN algorithm is limited by the distance between each incidence and in this investigation; features carry information that can be captured only by nonlinear classifiers. Yet the CNN yields acceptable results since

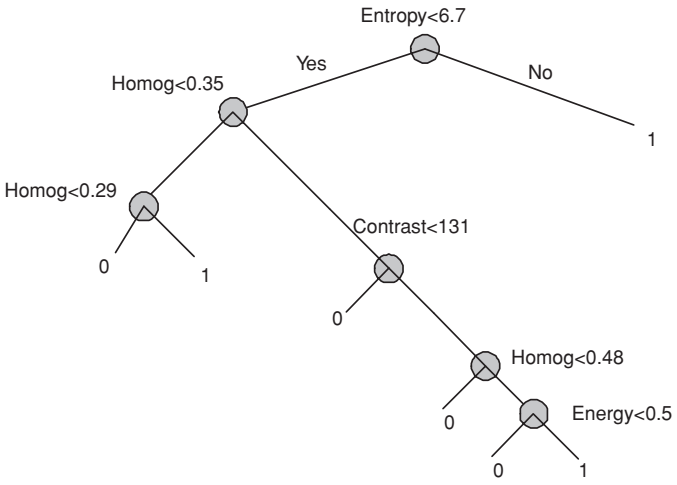

Figure 4. A sample DT classifier.

Table 1. Classification results using CNN.

		Cancer	Non-cancer
GLDM	Cancer	4	2
	Non-cancer	2	8
Sensitivity = 66.67%, specificity 80%, 75% accuracy			
GLDV	Cancer	4	2
	Non-cancer	2	8
Sensitivity = 66.67%, specificity 80%, 75% accuracy			
All	Cancer	4	2
	Non-cancer	1	9
Sensitivity = 66.67%, specificity 90%, 81.25% accuracy			

it generates a piecewise decision boundary which allows the KNN to deal with nonlinearly separable data, especially if a smaller number of neighbours that make it more nonlinear are chosen. This explains why choosing the DT, which relies mainly on the information gain of each feature is selected. However, the CNN can be used as an efficient classifier if the algorithm is applied online during the biopsy process.

6.2. Decision tree

A sample DT that is obtained for the GLDM features is given in figure 4, and the classification results are listed in table 2. From table 2 it is evident that the GLDV feature set exhibits a better specificity than that of the GLDM feature set and the combined feature set with the same sensitivity. Not only is the computational effort decreased since it reduces the non-important biopsies, but also the overall accuracy of the DT classifier is increased.

6.3. Support vector machine (SVM)

The results of applying the SVM to the available feature sets are shown in table 3.

Table 2. Classification results using DT.

		Cancer	Non-cancer
GLDM	Cancer	5	1
	Non-cancer	2	8
83.33% Sensitivity, 80% specificity, 81.25% accuracy			
GLDV	Cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity, 100% specificity, 93.75% accuracy			
All	Cancer	5	1
	Non-cancer	2	8
83.33% Sensitivity, 80% specificity, 81.25% accuracy			

Table 3. Classification results using SVM.

		Cancer	Non-cancer
GLDM	Cancer	5	1
	Non-cancer	1	9
83.33% Sensitivity, 90% specificity, 87.75% accuracy			
GLDV	Cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity, 100% specificity, 93.75% accuracy			
All	Cancer	5	1
	Non-cancer	1	9
83.33% Sensitivity, 90% specificity, 87.75% accuracy			

It is obvious that the SVM achieves approximately the same results as those obtained by the DT classifier for the available feature sets. This makes the DT more favourable for several reasons:

- it provides a clear rule of how the classification results are obtained;
- it depends mainly on the information content of each feature;
- its computational cost is less.

6.4. Feature selection using MI results

The resulting MI between each feature and the classes is depicted in figure 5, where the x-axis represents the feature number and the y-axis represents the mutual information between that feature and the class.

Choosing the feature subset that maximizes the MIFC and minimizes the MIXY according to figure 3, results in three features from the GLDV feature set whose classification results are shown in table 4.

Here it is obvious that the SVM and DT perform much better than the CNN classifier for the same feature subset. However, the chosen feature subset achieves better classification than both the GLDM feature set and the combined features. The subset performs as well as the GLDV features. The MIFS selected features are all GLDV features which are further proof that the GLDV features are more informative than the GLDM features. When the feature sets are combined, the classifier is confused.

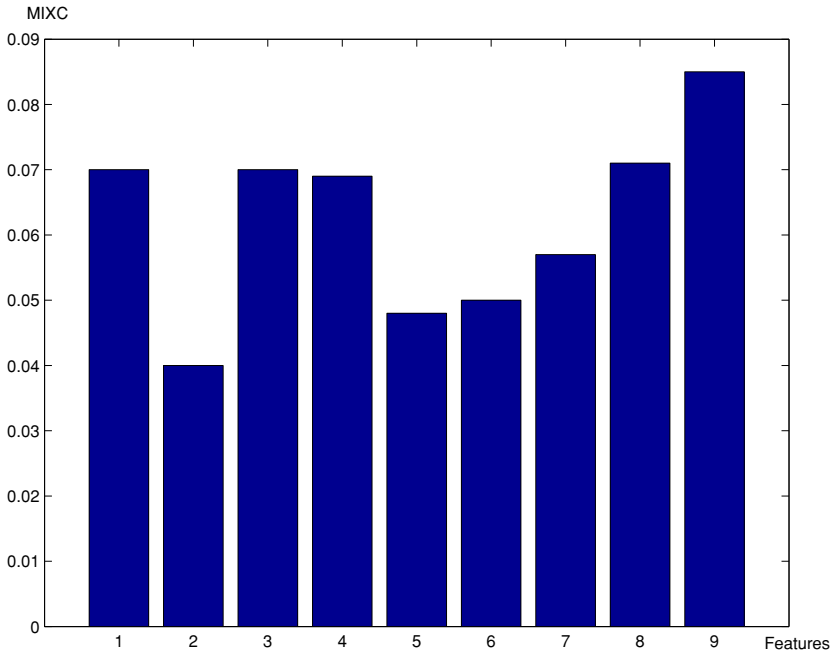

Figure 5. Interdependence between features and classes.

Table 4. Classification results using MI feature subset.

		Cancer	Non-cancer
CNN	Cancer	4	2
	Non-cancer	2	8
Sensitivity = 66.67%, specificity 80%, 75% accurracy			
DT	Cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity, 100% specificity, 93.75% accuracy			
SVM	Cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity, 100% specificity, 93.75% accuracy			

7. Conclusion

This note introduces a comprehensive TRUS tissue characterization algorithm with a multi-feature capability to mimic the expert radiologist in detecting suspicious regions with a high degree of accuracy. In addition, the note offers a comparison between the different feature sets either individually such as the GLDV or the GLDM or as a combination of all the features. In the case of the combined features, the MIFS is applied successfully to extract the most informative features. Moreover, three different classifiers are compared with all the available feature sets, revealing that the combination of the SVM or the DT classifier with the MIFS features has the highest accuracy. This technique can be used as a support for the existing technologies for prostate cancer diagnosis and as a result aiding biopsy planning. The key advantage of the proposed algorithm is the application of multi-resolution texture segmentation

for ROI identification, which leads to substantial time saving either by assisting the radiologist or by eliminating the need for whole image analysis.

References

Bassat O, Sun Z, Mestas J L and Gimenez G 1993 Texture analysis of ultrasound images of the prostate by means of co-occurrence matrices *Ultrason. Imaging* **15** 218–37

Battiti R 1994 Using mutual information for selecting features in supervised neural net learning *IEEE Trans. Neural Netw.* **5** 537–50

Bhanu K N, Ramakrishnan A G, Sureh S and Teresa Chow W P 2002 Fetal lung maturity using ultrasound image features *IEEE Trans. Inf. Technol. Biomed.* **6** 38–45

Clausi D A and Jernigan J E 2000 Designing Gabor filters for optimal texture separability *Pattern Recognit.* **33** 1835–49

Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* (New York: Wiley)

Hart P 1968 The condensed nearest neighbour rule *IEEE Trans. Inf. Theory* **14** 515–6

Jain A and Zongker D 1997 Feature selection: evaluation, application, and small sample performance *IEEE Trans. Pattern Anal. Mach. Intell.* **19** 153–8

Miller P and Astley S 1992 Classification of breast tissue by texture analysis *Image Vis. Comput.* **10** 277–82

Mohamed S S, Salama M M A, Kamel M and Rizkalla K 2004 Region of interest based prostate tissue characterization using least square support vector machine LS-SVM *Lecture Notes in Computer Science* **3212** 51–8

Quinlan J R 1990 Decision trees and decision-making *IEEE Trans. Syst. Man Cybern.* **20** 339–46

Scardino P T 1989 Early detection of prostate cancer *Urol. Clin. North Am.* **16** 635–55

Scheipers U, Lorenz A, Pesavento A, Ermert H, Sommerfeld H-J, Garcia-Schurmann M, Kuhne K, Senge T and Philippou S 2001 Ultrasonic multifeature tissue characterization for the early detection of prostate cancer *IEEE Ultrason. Symp.* **2** 1265–8

Webb E S 1988 *The Physics of Medical Imaging* (Bristol: Adam Hilger)