

Rough Wavelet Hybrid Image Classification Scheme

Hala S. Own¹, Aboul Ella Hassanien²

¹*Department of Solar and Space Research, National Research Institute of Astronomy and Geophysics, El-Marsad Street, P. O. Box 11421 Helwan, Egypt*

halaown@gmail.com

²*IT Department, FCI, Cairo University, Giza, Egypt*

aboitcairo@gmail.com

Abstract

This paper introduces a new computer-aided classification system for detection of prostate cancer in Transrectal Ultrasound images (TRUS). To increase the efficiency of the computer aided classification process, an intensity adjustment process is applied first, based on the Pulse Coupled Neural Network (PCNN) with a median filter. This is followed by applying a PCNN-based segmentation algorithm to detect the boundary of the prostate image. Combining the adjustment and segmentation enable to eliminate PCNN sensitivity to the setting of the various PCNN parameters whose optimal selection can be difficult and can vary even for the same problem. Then, wavelet based features have been extracted and normalized, followed by application of a rough set analysis to discover the dependency between the attributes and to generate a set of reduct that contains a minimal number of attributes. Finally, a rough confusion matrix is designed that contain information about actual and predicted classifications done by a classification system. Experimental results show that the introduced system is very successful and has high detection accuracy.

Keywords

Prostate cancer, rough sets, wavelet transforms, feature selection, feature reduction, classification.

1. Introduction

Transrectal Ultrasound (TRUS) is a key imaging technology widely used in diagnosis, treatment and follow up studies; it provides the anatomical description and volume of the prostate needed for disease treatment and detection of cancer. Unlike other methods such as X-rays Computed Axial Tomography

scan (CAT-scan) which are invasive due to radioactivity, TRUS is a minimally invasive, cost effective and safe technology used as a pre-operative evaluation tool in prostate cancer patients [2,6]. However, prostate cancer detection with TRUS is highly operator dependent; since the pathologist will look at the image in order to diagnosis. Human assessment is more subjective than objective and is limited due to the limitation of the resolution power of the ultrasound imaging and current accuracy of visual interpretation for identifying the presence of cancer is also limited.

The accurate detection of the prostate boundary in ultrasound images is crucial for some clinical applications, such as the accurate placement of the needles during the biopsy, accurate prostate volume measurement from multiple frames, and constructing anatomical models used in treatment planning and estimation of tumor border. These images are the result of reflection, refraction and deflection of ultrasound beams from different types of tissues with different acoustic impedance [5]. In ultrasound images the contrast is usually low and the boundaries between the prostate and background are fuzzy. Also, speckle and weak edges make the ultrasound images inherently difficult to segment. Furthermore, the quality of the image depends on the type and particular settings of the machine [10]. All these factors make the analysis of ultrasound images challenging. Therefore, there is a need to develop computer-aided diagnostic tools for the purpose of more accurate and efficient diagnosis by reducing the error rate resulting from the previous facts. This paper introduces a computer-aided system that classifies and detects prostate cancer using wavelet transform (WT) and rough sets (RS) in conjunction with Pulse Coupled Neural Networks (PCNN). A PCNN is first applied to the original prostate image for intensity adjustment. WT is a well-known image analysis methodology [4, 9]. It is known that the use of wavelets provides a richer feature space. It provides

approximation with excellent time and frequency resolution. The produced wavelet coefficient yields a smaller set of more robust features, which can improve the probability of correct classification [8, 11].

Rough set theory has been used to discover the dependency between attributes and generates a set of rules that contain a minimal set of attributes. It is used to perform quantitative measurements of relevant features that can discriminate between cancer and non cancer images for the purpose of early detection of cancer [14].

2. Related work

Significant progress has already been made in the use of computer- aided techniques in detecting prostate cancer.

Qu, *et al.*, [12] use discrete wavelet transform in discriminate analysis when the number of variables is much more the number of the observation. The method used a data compression stage followed by feature selection method based on Mahalanobis distance in order to select the wavelet coefficient with the highest discriminatory power. They perform feature reduction by thresholding wavelet coefficients. Juan *et al.*, [15] introduce a computer- aided diagnosis method with feature extraction depending on gray level dependence matrices. The evaluation was performed using a k-nearest neighbor classifier. Valdes *et al.*, [19] propose a texture segmentation technique using fractal features to classify cancerous regions in prostate ultrasound images. Pitts *et al.*, [7] investigate the application of grey level occurrence matrix techniques for the interpretation of prostate cancer and identification of corresponding features on the color images.

Feleppa *et al.*, [2] use a new tissue type method (TTI) based on spectrum analysis of radiofrequency (RF) echo signals and a neural network classifier to distinguish between cancerous and non-cancerous tissues.

Recently, a combination of rough sets and wavelets have been used in medical applications in different fields of image processing like denoising, segmentation and classification as introduced in [6,10,20].

However, the previous techniques do not consider the advantage of the multiresolution nature of wavelet transform with rough set theory as a data reduction and powerful classifier technique for early detection of prostate cancer. In this work, we will integrate the strength of wavelet multiresolution representation ability of wavelet transform to provide a sensible

decomposition of the data into different scales series with the principle advantages of rough set.

3. Wavelet Transform

Wavelets mean small waves that cut up data into different frequency components and transfer each component with different resolution that is matched to its scale. The main idea of wavelet analysis is to see both coarse and detail data without a heavy computational penalty. The goal of most modern wavelet researches is to create a set of basis functions and transform them in order to give information.

The wavelet transform (WT) decomposes a signal $f(t)$ by performing inner products with a collection of analysis functions $\{\Psi_{(a,b)}\}$ which are scaled and translated versions of the wavelet ψ ; i.e.

$$W(a,b) = \langle f, \Psi_{(a,b)} \rangle = \int_{-\infty}^{+\infty} f(t) \overline{\Psi_{(a,b)}(t)} dt \quad (1)$$

$$\Psi_{(a,b)}(t) = a^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

Where $W(a,b)$ is the wavelet coefficient of the function $f(t)$ that refers to the degree of similarity between the basis functions (wavelet) and the original signal at the current scale. The amplitude of the WT therefore tends to be maximum at those scales and locations where the signal most resembles the analysis template.

The continuous wavelet transform is a reversible transform, $f(t)$ can be restored using the formula

$$f(t) = \frac{1}{C} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} W(a,b) \left(\frac{t-b}{a}\right) \frac{dad b}{a^2} \quad (3)$$

where C is a constant that depends on the choice of the wavelet. Reconstruction is only possible if C is defined. This condition is called the admissibility condition, which restricts the class of functions that can be wavelet [4].

3.1 Multi-Resolution Analysis

A multiresolution analysis involves a scaling function $\phi(x)$, this function is chosen to have compact support. $\phi(x)$ has dilated and translated versions [3,12,18]. Larger scales correspond to dilated signals

and small scales correspond to compressed signals. The scaling function $\varphi(x)$ takes the form:

$$\varphi(x) = 2^{-\frac{j}{2}} \varphi(2^{-j}x - n) \quad j, n \in \mathbb{Z} \quad (4)$$

The wavelet transform of images tends to be very sparse with a few large scale of wavelet coefficients. The image can be expanded in terms of two-dimensional wavelets. Due to the down sampling operation, each image is decomposed into four sub-images. The size of each sub-image is only a quarter of the original image. Wavelet scaling function, gives a series of pictures of the signal, each at a resolution differing by a factor of two from the previous resolution. In one direction, these successive images approximate the signal with greater and greater precision, approaching the original. Figure (1) demonstrates that the low pass sequences of images can be mapped onto a quad tree structure. Each pixel at resolution J corresponds to four child pixels at resolution J-1. Thus, a pixel at resolution J is defined are parent of their corresponding child pixels at resolution J-1. In turn, the pixels at resolution J-1 are parents to their corresponding child pixels at resolution J-2.

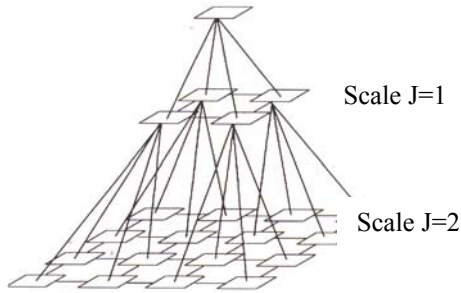


Figure 1. Multiscale decomposition illustration

3.2 PCNN

The Pulse-Coupled Neural Network (PCNN) was derived from studies on the cat's eye. It was Eckhorn who first made a model of the cat's visual cortex [1]. There are several compartments in the Eckhorn neuron. It has two input compartments, linking and feeding. The feeding compartment (F) receives both an external and a local stimulus, whereas the linking compartment (L) only receives a local stimulus. The feeding and linking are combined to form the membrane voltage (U_m). This is then compared to a local threshold θ .

Figure 2 shows the layout structure of PCNN and its component.

The PCNN model is described by the following equations:

$$F_{ij}[n] = e^{\alpha} F_{ij}[n-1] + S_{ij} + V_F \sum M_{ijkl} Y_{kl}[n-1] \quad (5)$$

$$L_{ij}[n] = e^{\alpha} L_{ij}[n-1] + V_L \sum W_{ijkl} Y_{kl}[n-1] \quad (6)$$

$$U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n]) \quad (7)$$

$$Y_{ij}[n] = \begin{cases} 1 & \text{if } U_{ij}[n] > \theta_{ij}[n-1] \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

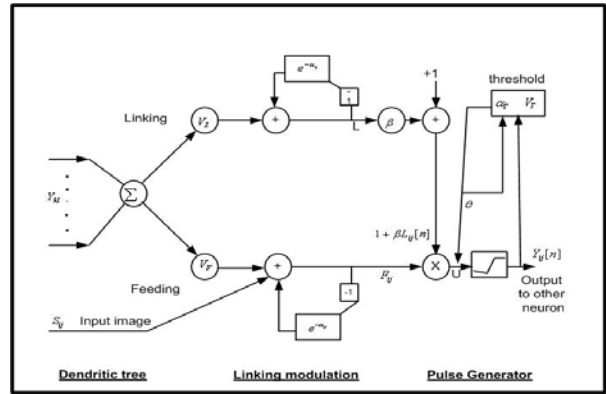


Figure 2. the layout structure of PCNN and its component

The indices i, j refer to the i -th, j -th neuron ($1 \leq i \leq a, 1 \leq j \leq b$), the α terms is decay constant, S_{ij} is the input stimulus (the intensity of pixel x, y), the V 's are the two respective potentials, M and W are the two synaptic weight sets, and the Y terms refer to the output of neurons from the previous iteration $n-1$. β is the linking strength of the two components. The state U is compared to a dynamic threshold θ to form the output Y of pixel (i, j) .

As summarized in equation (5), when a neuron fires ($Y > \theta$), the threshold increases by a large constant amount V_θ . The neuron is thus prevented from firing for a while, until θ decays (according to the decay constant α_θ) sufficiently for the value of Y to exceed θ once again. To calculate their current values, the threshold, and both the feeding and linking compartments retain a memory of their previous state.

3.3 Rough Sets

Basically, rough set theory [9,14] deals with the approximation of sets that are difficult to describe with the available information. In a medical application, a set of interest could be the set of patients with a certain disease or outcome. In rough set theory, the data is collected in a table, called the decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, we assume we are given a set of examples with a class label to indicate the class to which each example belongs. We call the class label the decision attributes, while the rest of the attributes are called the condition attributes. Rough sets theory defines three regions based on the equivalent classes induced by the attribute values, lower approximation, upper approximation and boundary. Lower approximation contains all the objects which are classified surely based on the data collected, and upper approximation contains all the objects which can be classified probably, while the boundary is the difference between the upper approximation and the lower approximation. So, we can define a rough set as any set defined through its lower and upper approximations. The upper and lower approximation are defined by the following equations, respectively.

$$\underline{PY} = \bigcup \{x : x \in \mu / IND(P), X \subseteq Y\} \quad (9)$$

$$\overline{PY} = \bigcup \{x : x \in \mu / IND(P), X \cap Y \neq \{\phi\}\} \quad (10)$$

where μ is the universe.

Indiscernibility notion is fundamental to rough set theory. Informally, two objects in a decision table are indiscernible if one cannot distinguish between them on the basis of a given set of attributes. Hence, indiscernibility is a function of the set of attributes under consideration. For each set of attributes we can thus define a binary indiscernibility relation, which is a collection of pairs of objects that are indiscernible. An indiscernibility relation partitions the set of cases or objects into a number of equivalence classes. An equivalence class of a particular object is simply the collection of objects that are indiscernible to the object in question. One of the nice features of rough sets theory is that rough sets can tell whether the data is complete or not based on the data itself. If the data is incomplete, it suggests more information about the objects needed to be collected in order to build a good classification model. On the other hand, if the data is complete, rough sets theory can also determine whether there are redundant information in the data and can find the minimum data needed for the

classification model. This property of rough sets theory is very important for applications where the domain of the knowledge is very limited or data collection is very expensive/laborious because it makes sure the data collected is just good enough to build a good classification model without sacrificing the accuracy of the classification model or wasting time and effort to gather extra information about the objects [14].

4. Overall CAD Scheme

The general agreement of computer-diagnosis analysis scheme is making useful computer-generated information available to physicians for decision support rather than trying to make a computer act like a diagnostician. The architecture of the proposed CAD system is illustrated in Figure 3. It is composed of four fundamental building phases: pre-processing, feature extraction, rough set data analysis and discrimination. In the first phase of the investigation, an intensity adjustment process maps an image's intensity values to a new range is first used to generate an image with enough bright intensity around the prostate. In the second phase, the set of features relevant to the prostate image are extracted, normalized and represented in a database as vector values. The third phase is rough set data analysis. It is done by computing the minimal number of necessary attributes, together with their significance, and generating the sets of rules.

Finally, rough set confusion matrix is designed for discrimination to test whether the images are cancer or non cancerous; it is dependent on the type of generated rules. These three phases are described in detail in the following sections along with the steps involved and the characteristic feature for each phase.

4.1 Pre-processing Phase: Intensity Adjustment

To increase the efficiency of the prediction if the image contains a cancer or not, a pre-processing stage should be considered to enhance the quality of the input prostate images before feature extraction and classification. The median filter with PCNN [1] is used to reduce noise in an image. It operates one pixel in the image at a time, and looks at its closest neighbors to decide whether or not it is representative of its surroundings. To begin with, one should decide the size of the window that the filter operates on within the image. The size could for example be set to three, which means that the filter will operate on a centered pixel surrounded by a frame of 3×3 neighbors. Then the filter sorts the pixels contained in the image area surrounded by the window. The central pixel will be

replaced by the median; the middle value of the ranking result. The advantage of the median filter, compared with other smoothing filters of similar size, is that it performs noise reduction with considerably less blurring. Thus, the filter also preserves the edges in an image.

The median filter works especially well for random noise. The algorithm works as follows: it firstly finds out the central position of the noised pixel according to the firing pattern, and then removes the noise from the image using a median filter. Initially the threshold of all of the neurons are set to zero, and at the first iteration all the neurons are activated, or output a pulse, which means all neurons receive the maximal linking input at the next iteration. So the proper set of the PCNN's parameters will make the neurons corresponding to noised pixels with high intensity fire before it's neighborhood at the second iteration. According to the current firing pattern the concrete position of noised pixels can be found out. Then the noised pixels can be removed using a 3×3 median filter. The removal of noised pixels with low intensity is same as the removal of noised pixels with high intensity if the intensity is inverted. Just because this algorithm can find the concrete positions of those noised pixels and apply median operation only on the noised regions, its ability to keep the details of the image is strong.

4.2 Feature Extraction Phase

Feature extraction plays an important role in classification process. Badly implemented feature extraction or an improper feature leads to poor classification results even by using the best possible classifier. The previous proposed texture features in medical images are identified by the matrix indexed by grey levels. This matrix is called co-occurrence matrix [15].

In this work the features will be defined in the wavelet transform domain. In the case of images, we consider the image an $N \times N$ square matrix of signal values. We decompose the image row-wise for every row using one dimension wavelet decomposition algorithm. The resulting two matrices are transposed and processed row-wise again to obtain four $\frac{N}{2} \times \frac{N}{2}$ square matrices. This procedure can be repeated any number of time to represent the resolution level we want to reach. The total number of coefficients after decomposition always equals the number of wavelet coefficients at level 1.

Since cancer creates discontinuities in the tissue, in order for cancer to survive, it develops its own blood supply system which is different than the supply system of normal tissue [1]. Therefore, the variance due to the edge is higher when the continuity breaks down, so cancerous tissue is expected to have more edges. The activity level of this area is high. The activity level can be measured as the entropy.

The traditional entropy measure is a global quantity calculated for the whole image, and then it is not appropriate for quantifying the distribution of the information at different scales of resolution. Strack et. al. [16] introduced the concept of multiresolution into an entropy measure. He considers the information contained in the image as the summation of information at different resolution levels. We have introduced a new feature within a wavelet transform domain [13]. It represents the contribution weight of coefficient in the multiresolution contrast entropy. We will use multiresolution entropy instead of entropy. The new feature is called Multiresolution Local Contrast Entropy (MLCE).

Let w_j be the detail coefficients of observed data over an 1 wavelet transform. The multiresolution entropy is:

$$ME(x, y) = \sum_{j=1}^l \sum_{k=1}^N \frac{w_j(k)^2}{2\sigma_j^2} \quad (11)$$

N , l and σ are the number of data, the resolution level, and the noise variance estimated in the finest scale, respectively.

We claim that the most significant feature is the probability of the multiresolution local contrast entropy (MLCE) for each coefficient. The MLCE is calculated by:

$$MLCE(x, y) = \frac{ME(x, y)}{\sum_{i=1}^N \sum_{j=1}^N ME(i, j)} \quad (12)$$

It assigns a value to the wavelet coefficient according to its contribution weight to the local contrast entropy. It will be calculated as follows:

- 1- Perform the discrete wavelet packet decomposition on the signal to the fifth level of resolution. We obtain 32 subbands.
- 2- From each subband at the fifth level of resolution, compute the average multiresolution local contrast entropy of each coefficient as

$$v_i = \frac{1}{n_i} \sum_{j=1}^{n_i} MLCE_j, i=1, 2, \dots, 32 \text{ \& } n_i = 128 \quad (13)$$

Where v_i is the i^{th} feature in a wavelet packet feature vector. To identify prostate cancer, texture features are needed as a discriminative measurement for the samples to measure the homogeneity, the coarseness, the periodicity and the linearity of texture. The rest of the features will be calculated using the coefficient matrix produced to the fifth level of resolution. We use the texture features used in [15,17] to identify cancerous tissue. Then each features value is calculated by the average of each feature value over all the levels. The set of selected features can be seen in Table 1.

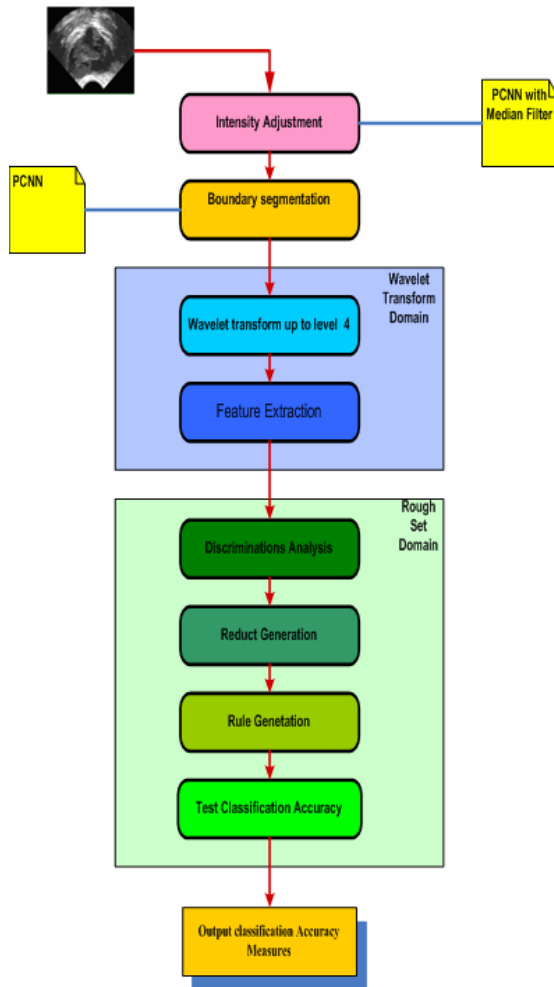


Figure 3. Rough Wavelet /hybrid Architecture

4.3 Rough Sets Data Analysis Phase: Feature Reduction

The basic philosophy of rough sets is to reduce the attributes in the data set based on the information content of each attribute or collection of attributes such that there is a mapping between similar objects and a corresponding decision class. In general, not all of the information contained in a data set is required: many of the attributes may be redundant in the sense that they do not directly influence which decision class a particular object belongs to. One of the primary goals of rough sets is to eliminate attributes that are redundant. Rough sets use the notion of the lower and upper approximations of sets in order to generate decision boundaries that are employed to classify objects.

We introduce a reduct algorithm based on the degree of dependencies and the discrimination factors. The main steps of the reduct generation algorithm are provided below:

Reduct Generation Algorithm

Input: information table (ST) with discretized real valued attribute.

Output: reduct sets

$$R_{final} = \{r_1 \cup r_2 \cup, \dots, \cup r\}$$

1. For each condition attribute $c \in C$, compute the correlation factor between c and the decisions attributes D if the result is positive then define the attribute as a relevant attribute
2. Cumulatively divide the set of relevant attributes into a different variable set.
3. Calculate the classification quality for each previous data set.
4. Choose the set with highest classification accuracy as an initial reduct set.
5. For each attribute in reduct set produced in step 4, calculate the degree of dependences between the decisions attribute and that attribute.
6. Merge the attributes produced in the previous step with the rest of conditional attributes then calculate the discrimination factors for each combination to find the highest discrimination factors that produced, then add the highest discrimination factors combination to the final reduct set.
7. Repeat step 6 until all attributes in initial reduct set are processed.
8. The result sets is the final reduct sets.

Table 1. The set of selected features

No	Name	Description
1	AMLCE	Average Multiresolution local contrast entropy
2	RMSE	Root mean square error
3	ASM	Angular second moment
4	CON	Contrast
5	DIS	Dissimilarity
6	COR	Correlation
7	IDM	Inverse difference moment
8	SD	Slandered deviation
9	ADWC	Average of Detailed Wavelet Coefficient
10	NSSDWC	Sum of Square of Detailed Wavelet Coefficient

Based on the generated reduct sets, we extract a set of generated rules, which will be used for classification. The total number of generated rules is 265. The two level reduct techniques we use ensure that the reduct set produced is the minimal set of attributes. Therefore, generated rules with these minimal set of attributes will be visible for the analyst, making possible the extraction of more information from the rules possible. Moreover, the prediction will be highly accurate, which affects the treatment decision of the patients.

5. Experimental Results

Our dataset consists of 212 TRUS cancer and non cancer images. A number of experiments are carried out to determine if the feature extraction and the classification schemes proposed can detect the malignancy in a TRUS image. All the experiments are implemented in MATLAB. According to our CAD system our experiments go through the following stages:

Stage 1: Intensity Adjustment and Boundary Segmentation

Due to the differences in conditions under which acquisition occurred, the images are not at the same level of contrast. A pre-processing step is then required. We apply PCNN with median filter to enhance the contrast of the images to the same level. The success of the application of PCNNs to image segmentation depends on the proper setting of the various parameters of the network, such as the linking

parameter β , thresholds θ , decay time constants α_θ , and the interconnection matrices M and W. Proper

setting of the parameters is especially important when intensity significantly varies across a single segment. The PCNN segmentation works as follows: An input gray-scale image is composed of $M \times N$ pixels. This image can be represented as an array of $M \times N$ normalized intensity values. Then the array is fed in at the $M \times N$ inputs of the PCNN. If initially all neurons are set to 0, the input results in activation of all of the neurons at the first iteration. The threshold of each neuron, Θ , significantly increases when the neuron fires, then the threshold value decays with time. When the threshold falls below the respective neuron's potential (U), the neuron fires again, which again raises the threshold. The process continues creating binary pulses for each neuron. While this process goes on, neurons encourage their neighbors to fire simultaneously in a way that is supported through interconnections. The firing neurons begin to communicate with their nearest neighbors, which in turn communicate with their neighbors. The result is an autowave that expands from active regions. Thus, if a group of neurons is close to firing, one neuron can trigger the group. Due to linking between neurons, the pulsing activity of invoked neurons leads to the synchronization between groups of neurons corresponding to subregions of the image that have similar properties, and produces a temporal series of binary images. This synchronization results in image segmentation figure 4 shows the original, enhanced and segmented images respectively. Table 2 shows the PCNN parameter values used in this application.

Table 2. PCNN parameters values.

PCNN parameters		Values
β	Linking strength	0.2
α_F	Feeding decay time constant	0.0 01
α_L	Linking decay time constant	1
α_Θ	Thresholding decay time constant	10
V_F	Inherent feeding potential	0.0 1
V_L	Inherent linking potential	1
V_Θ	Inherent thresholding dynamic potential	2
N	Number of iterations	5

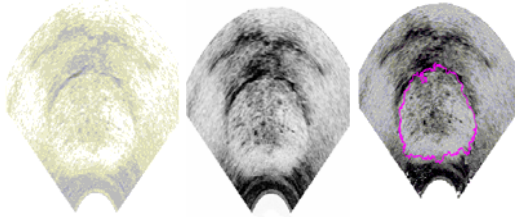


Figure 4. Original and it is enhanced and segmented results.

Stage 2: Features Reduction

Based on adjusted and segmented images produced in stage 1, a set of features predefined in table 1 are extracted from each image. We then apply our reduction algorithm on these features.

Table 3. Correlation between each condition attribute and decision attribute

#	Attribute	Correlation
1	AMLCE	0.567
2	RMSE	0.306
3	ASM	-0.455
4	CON	0.234
5	DIS	-0.123
6	COR	-0.258
7	IDM	-0.536
8	SD	0.445
9	ADWC	0.344
10	NSSDWC	0.489

As can be seen from table 3, the attributes AMLCE, RMSE, CON, SD, ADWC, AND NSSDWC have a positive correlation coefficient with the decision class. Therefore, this makes the AMLCE, RMSE, CON, SD, ADWC, and NSSDWC a primary minimal set of attributes. As for the final phase of reduct, we calculate for each attribute produced in the previous phase the degree of dependencies between the decision attribute and that attribute.

Table 4. Discrimination analysis for the first round

Attribute	Discrimination Factor
AMLCE	0.43
RMSE,	0.0
CON,	0.0
ADWC	0.21
SD	0.14
NSSDWC	0.32

From Table 4, we see that AMLCE has the highest discrimination factor, so we choose it as the first reduct attribute for that phase. The next step is to merge the AMLCE with the rest of the conditional attributes to find the highest discrimination factors that are produced. Table 5 represents the results of this round.

Table 5. Discrimination analysis for the second round

Attributes	Discrimination Factor
AMLCE, RMSE	0.302
AMLCE, CON	0.311
AMLCE, ADWC	0.354
AMLCE, SD	0.382
AMLCE, NSSDWC	0.381

Because AMLCE, SD has a highest discrimination factor, we choose the second attribute to be SD. By continuously applying the same approach, we get the following results, shown in Table 6.

Table 6. Discrimination analysis for the third round

Attributes	Discrimination Factor
AMLCE, SD, RMSE	0.381
AMLCE, SD, CON	0.384
AMLCE, SD, ADWC	0.389
AMLCE, SD, NSSDWC	0.392

Because AMLCE, SD, NSSDWC have a highest discrimination factor, we choose the third attribute to be NSSDWC.

Table 7. Discrimination analysis for the fourth round

Attributes	Discrimination Factor
AMLCE, SD, NSSDWC, CON	0.397
AMLCE, SD, NSSDWC, ADWC	0.399
AMLCE, SD, NSSDWC, RMSE	0.396

We can see from Table 7, that the set containing AMLCE, SD, NSSDWC, ADWC has a highest discrimination factor.

Table 8. Discrimination analysis for the sixth round

Attributes	Discrimination Factor
AMLCE, SD, NSSDWC, ADWC, CON	0.401
AMLCE, SD, NSSDWC, ADWC, RMSE	0.401

From Table 8, we can observe that we have two minimal reduction sets:

R1= {AMLCE, SD, NSSDWC, ADWC, CON}

R2={ AMLCE, SD, NSSDWC, ADWC, RMSE }.

Stage 3: Classification Performance Measures

A natural use of a set of rules is to measure how well the ensemble of rules is able to classify new and unseen objects. To measure the performance of the rules is to assess how well the rules do in classifying new cases. So we apply the rules produced from the training set data to the test set data.

The assessment of the classification results has been made using two methods. The first is the classical one is to split the data into training and testing sets. The first set is used to compute the discriminating function, while the second is classified using this function. The results are compared with the labels of these samples. The second method is known as the cross-validation test. In this approach the data set is randomly divided into *M* disjoint sets of equal size $\frac{N}{M}$, where *N* is the total number of samples. The classifier is trained *m* times, each time with a different set used as a validation set. The estimated performance is simply the mean of these *m* errors.

A Confusion matrix (CON) is constructed to represent the results of the two classifiers. The CON is initialized at 0, and then it is gradually modified as follows: If the *i*th test image, known to be cancerous image is classified as cancerous, then the true cases is increased by one. If the *i*th test image, known to be cancerous is classified as non cancerous then the number of false cases is increased by one. This process continues until all test images are classified to obtain the final confusion matrix. Then each element in CON is divided by the number of images of the target class. Table 9 shows the results.

Table 9, Comparison between two classification methods

Classifier type	Classification Error
Classical method	6.4%
Cross-validation method	5.1%

Table (9) shows that classification error rates for both classifier types are very low

Cross-validation showed more accurate classification results since it uses all the available data

for training; however, it requires more computational time.

It is well- known that implementing feature selection improves the accuracy of a classifier. Actually, the degree of improvement will depend on many factors such as the type of classifier, the effectiveness of the feature selection and the quality of the feature.

The effectiveness of our new introduced feature (AMLCE) is demonstrated by measuring the classification accuracy with and without this feature. The quantitative results are given in Table 10. From table we can see that the classification error increase by 2.7% when we exclude AMLCE

Table 10, Comparison of features importance

Feature	Classification Error
All features	5.1%
Removing AMLCE	7.8%

Yet another test was performed for measuring classification performance between our proposed approach and MLP neural network and k-nearest neighbor. Figure 5 shows a plot of the overall classification error of two approaches compared with our proposed approach based on rough set. It shows that the rough sets approach is much better than MLP and k-nearest neighbor.

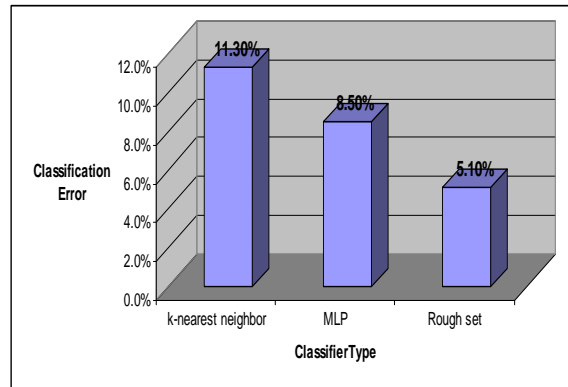


Figure 5. comparing of classification errors between MLP and k-nearest neighbor against our rough set based approach

6. Conclusion

We have developed advanced image analysis methods for detecting a cancer in TRUS images with a high degree of accuracy. Feature extraction techniques are used to derive parameters for the cancerous and non-cancerous image samples. A rough set theory is

used to develop models for distinguishing between cancer and non-cancer image samples. A number of experiments are undergone in order to determine if the feature extraction and classification scheme proposed can detect the cancer in a TRUS image.

7. References

- [1] Aboul Ella Hassanien, "Pulse coupled Neural Network for Detection of Masses in Digital Mammogram", *Neural Network World Journal*, Vol. 2/06, pp.129-141,2006.
- [2] E.J. Feleppa, J. Ketterling, P. Lee;S. Urban, ; A.Kalisz, ; C.R. Porter, ; G. Kutcher, ; F.Arias-Mendoza, " New Developments in Tissue-type Imaging (TTI) for Guiding Prostate Biopsies and for Planning and Monitoring Treatment of Prostate Cancer", *IEEE Ultrasonic Symposium*, Volume 2, , pp. 834 - 837 Vol.2, 23-27 Aug. 2004 .
- [3] M. Gao, P. Bridgman, and S. Kumar, "Computer Aid Prostate Cancer Diagnosis Using Image Enhancement and JPEG2000", *Proceedings of SPIE Annual Meeting*, San Diego, CA, USA, August 2003. http://attila.sdsu.edu/~kumar/Papers/SPIE_prostate.pdf
- [4] J. Goswami, and A. K. Chan, "Fundamentals of Wavelets, Theory, Algorithms and Applications", John Wiley& Son, Inc. 1999.
- [5] S. Hirano, X. Sun, and S. Tsumoto, "Dealing with Multiple Types of Expert Knowledge in Medical Image Segmentation: A Rough Set Style Approach", *IEEE Inter. Conf. on Fuzzy systems*, Vol. 2, pp. 884-889, 12-17 May 2002.
- [6] J. Haizhai, X. Wang, and S. Zhang, "Rough-Neural Image Classification Using Wavelet Transform", *IEEE Proceedings of the sixth International Conf. On Machine Learning and Cybernetics*, Hong Kong ,pp. 3045 – 3050, 19-22 August 2007.
- [7] G. Houston, S. B. Permkumar, and D. E. Pitts, "Prostate Ultrasound Image Analysis: Localization of Cancer Lesions to Assist Biopsy", *IEEE Eight Symposium on Computer Based Medical Systems*, pp. 94-101, Lubbock, TX, USA, 1995.
- [8] A. Kumar, M.Srinivasan, S. Annadurai, "A Wavelet Based Image Denoising Using Statistical Sampler for Bayesian Estimator", *TENCON 2003. Conference on Convergent TechnologiesforAsia-PacificRegion Vol 1* 1, pp. 21 - 25, 15-17 Oct. 2003.
- [9] S. Mallat, "Wavelet for a Vision", *Proceeding of the IEEE*, Vo. 84, pp. 604-614, Apr 1996.
- [10] A. Mohabey, and A.K. Ray, "Rough Set Theoty Based Segmentation of Color Images", *IEEE Trans. On Geoscience and Remote Sensing*, Vol. 40, Issue 11, pp. 2495-2501, Nov. 2002.
- [11] D.E. Nelson, J.A. Starzyk, and D.D. Ensley, "Iterated Wavelet Transformation and Signal Discrimination for HRR Radar Target Recognition", *IEEE Trans. On Systems, Man, and Cybernetic-part A, System and Human*, Vol. 33, No. 1, pp. 52-57, Jan. 2002.
- [12] Y. Qu, et al , "Data Reduction Using a discrete Wavelet Transform in Discriminant Analysis of Very High Dimensionality Data", *Biometrics* Vol. 59, pp, 143-151, March 2003.
- [13] H. S. Own, A. Hassanien, "Image Registration Based in Multiresolution Local Contrast Entropy in Wavelet Transform Domain", *IEEE 14th Inter. Conf. In Digital Signal Processing*, Vol. 2, pp. 889-892, 1-3 July, Greece 2002.
- [14] Pawlak Z., "Rough Sets- Theoretical aspect of Reasoning about Data" Kluwer Academic Publishers, 1991.
- [15] J.C. Perez-Cortes, A. Juan, E. Vallada, " Textural Analysis of Prostate Cancer In Ttransrectal Ultrasound Images", *Biosignal* 2002. <http://prhlt.iti.es/papers/2002/Perez02.pdf>
- [16] M.A. Roula, J. Diamond, A. Bouridane, P. Miller, and A. Amira, "A Multispectral Computer Vision System for Automatic Grading of Prostatic Neoplasia", *IEEE Inter. Symposium on Biomedical Imaging*, pp. 193- 196, 2002.
- [17] U. Scheipers, A. Pesavento, H. Ermert, H. J-Sommerfeld, M. Garcia-Schurmann, K. Kuhne, T. Senge, and S. Philippau, " Ultrasound Multifeature Tissue Characterization for The Early Detection of Prostate Cancer", *IEEE Ultrasonics Symposium*, pp. 1265-1268, 2001.
- [18] J. L. Starck, F. Murtagh, R. Gastaud "New Entropy Measure on the Wavelet Transform and Noise Modeling, *IEEE Trans. On Circuits and Systems-II: Analog and Digital Signal Processing* Vol. 45, No, 8, pp. 1118-1124, August 1998.
- [19] J. J. Valdes, L. C. Molina, S. Espinosa, " Behavior Analysis of Fractal Features for Texture Description in Digital Images: An Experimental Study", *15th Conf. Pattern Recognition*, Vol. 3, pp. 905-908, 3-7 Sept. 2002.

[20] C. Bo, G. Zexun, and Y. Yand, L. Xiaosong, "Application of the Rough Set Image Median Denoising", IEEE eight ACIS Inter. Conf. on Software Engineering, Artificial Intelligence, Networking, and Parallel Distribution Computing , Vol 1, pp. 75 – 78, 30 July – 1 Aug 2007.

Author Biography



Hala Shawky Own received her B.Sc. with honors in 1988 and M.Sc. degree in 1995, both from Ain Shams University, Faculty of Science, Pure Mathematics and Computer Science Department, Cairo, Egypt. On 2003, she received her Ph.D. degree in the field of image processing, from Masnoura university, Faculty of Science, Computer Science Department, Mansoura, Egypt. Currently, she is a researcher at National Research Institute of Astronomy and geophysics (NRIARG), Helwan, Egypt. Her research interests are in the areas of, rough set theory, wavelet theory, medical image analysis, astronomical image analysis.