ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

Automatic classification of prostate cancer using pseudo-gaussian radial basis function neural network

O. Valenzuela[1], I.Rojas[2], F.Rojas[2], L.Marquez[1]
1: Department of Applied Mathematic
2: Department of Computer Architecture and Computer Technology
University of Granada (Spain)

Abstract -Recent advances in multimedia and image processing techniques can be utilized to assist pathologists in this respect. In fact, many investigators believe that automation of prostate cancer analysis increases the rate of early detection. In this paper, we will propose an automatic procedure for prostate cancer light micrograph based on soft-computing technique, for image interpretation, with increased accuracy. We propose a feature subset selection algorithm that selects the most important features, used by a pseudo-gaussian radial basis function neural networks to classify the prostate cancer light micrograph. A high classification rate has been achieved which will reduce the subjective human invention and will increase the diagnostic speed.

1. Introduction

Pathologists, daily, screen large numbers of slides containing cancerous cells manually, which are similar in shape, size or cell structure. Therefore this procedure becomes arduous, difficult and can effect the Pathologist's decision.

In this paper, the classification of prostate cancer (an uncontrolled, malignant, growth of cells in the prostate male gland) and benign tumour are considered for automatic classification. Prostate cancer is the most commonly diagnosed cancer in men [1]. It develops from cells within the gland. Risk factors include age; 75% of cases are in men over 65 years. There are more cases in western countries and 10% of cases can be linked to a person's family history. Dietary and other factors in the environment are also linked to prostate cancer. The normal prostate gland is the size of a walnut in a young man and enlarges with age. Prostate cancer, unlike many other forms of cancer, tends to be slow growing. Eventually it can spread to other organs and tissues, including bones. The exact cause of prostate cancer, like many cancers, is not known. Doctors know, however, that the growth of the cancer is dependent on the male sex hormone testosterone. Hormones can also control the growth of cancer cells as well and this is what testosterone does in prostate cancer.

Because many prostate tumours are slow growing, survival rates are excellent when the disease is detected in its early stages [2][3]. Normal prostate enlargement is not cancerous and is referred to as benign prostatic hypertrophy.

Recent advances in multimedia and image processing techniques can be utilized to assist pathologists in this respect. In fact, many investigators believe that automation of prostate cancer screening analysis increases the rate of early detection. The current procedure assumes that the image is recorded on a X-ray film and that it is interpreted by a human expert. In this paper, we will propose an automatic procedure for prostate cancer light micrograph based on pseudo-gaussian radial basis function neural network, for image interpretation and classification.

The structure of this paper is as follow: In the next section we will present the first step in the automatic procedure: the feature extraction from the prostate cancer light micrograph. In Section 3 we will present a feature selection algorithm, which selects the more important variables to the Neural Network. The structure of the modified Radial Basis Function Neural Network is presented in Section 4. Finally, Section 5 presents the simulation results and Section 6 the main conclusion.

2. Feature Extraction

The implemented feature extraction procedure relies on the texture and morphology (shape description) domains. Four gray level sensitive histogram moments are extracted from the pixel value histogram of each image:
1 and 2.- Average gray level of foreground in enhanced image and of the background:

$$Average_{foreground} = \frac{\sum_{(x,y)\in foreground} X(x,y)}{sum(pixel_{foreground})} \tag{1}$$

3 and 4.- Standard deviation of gray levels of the foreground and of the background in enhanced image:

$$StdDev_{foreground} = \left(\sum_{(x,y)\in foreground} \left[X(x,y) - Average_{foreground} \right]^2 \right)^{1/2} \tag{2}$$

The above four features are based on the fact that the cancer clusters have apparently different gray levels compared to the background tissues.
Two new features are introduced using the gradient vector:
5.- Average of maximum gradient of boundary pixels: This is obtained by calculating the gradient of each boundary pixel's eight connected neighbours and taking the maximum gradient value.
6.- Average of mean gradient of boundary pixels: This feature is obtained by calculating the gradient of each boundary pixel's eight connected neighbours and taking the average of its neighbour's gradient value as its gradient.

The following three features are focus on the morphology of the image:
7.- Compactness:

$$\lambda = \frac{perimeter^2}{4\pi \bullet area} \tag{3}$$

8.- Moment: For a two-dimensional real bounded image f(x; y), the moments M_{pq} (the order ofthe moment is p + q) are defined as

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy \quad \text{for } p,q=0,1,2,... \tag{4}$$

while the central moments are defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-\bar{x}\right)^{p}\left(y-\bar{y}\right)^{q} f(x,y)dxdy$$
$$\text{where } \bar{x}=\frac{M_{10}}{M_{00}} \text{ and } \bar{y}=\frac{M_{01}}{M_{00}} \tag{5}$$

9.- Fourier descriptor: It is a transformation feature for shape representation in digital image processing, which transforms a feature set from time-domain into frequency-domain. First, it transforms the two-dimensional image into a complex representation image, of the form u(n) = (x(n) +jy(n)); n=0, 1,…, N-1. Then by using algorithm called "chain code" to trace the outline of the tumour boundary with the fixed 2kπ period, we can get the Fourier representation of the shape features. The mathematical model of the Fourier descriptor is as follows:

$$u(n) \triangleq \frac{1}{N}\sum_{k=0}^{N-1} a(k)\exp\left(\frac{j2\pi kn}{N}\right), \quad 0 \le n \le N-1$$
$$a(k) \triangleq \frac{1}{N}\sum_{k=0}^{N-1} u(n)\exp\left(\frac{-j2\pi kn}{N}\right), \quad 0 \le k \le N-1 \tag{6}$$

3. Input Feature Selection using Mutual Information

Mutual information is a good indicator of relevance between variables, and has been used as a measure in several feature selection algorithms. However, calculating the mutual information is difficult, and the performance of a feature selection algorithm depends on the accuracy of the mutual information [4]. In this section we use a method of calculating mutual information between input and class variables based on the Parzen window [6]. In the prostate cancer classification problem presented in this paper, the mutual information between the input features X and the class C can be represented as follows:

$$I(X;C) = H(C) - H(C|X) \tag{7}$$

In this equation, because the class is a discrete variable, the entropy of the class variable H(C) can be calculated as:

$$H(X) = -\sum_{x\in\overline{X}} p(x)\log p(x) \tag{8}$$

Where the discrete variable X has $\overline{X}$ alphabets and the probability density function (*pdf*) is p(x)=Pr{X=x), $x \in \overline{X}$. The hard problem is to compute the conditional entropy H(C/X). Because it is not easy to estimate p(c/x), being N the number of classes. By the Bayesian rule, the conditional probability p(c/x) can be written as

$$p(c|x) = \frac{p(x|c)\,p(c)}{p(x)} \tag{9}$$

Using the Parzen Window method [5][6], is it possible to estimate the conditional *pdf* $\hat{p}(x|c)$:

$$\hat{p}(x|c) = \frac{1}{n_c} \sum_{i \in I_c} \frac{\exp\left(-\frac{(x - x_i)^T \sum^{-1} (x - x_i)}{2h^2} \right)}{(2\pi)^{d/2} h^d |\Sigma|^{1/2}} \tag{10}$$

Where $\sum$ is a covariance matrix of a d-dimensional vector of random variable (x-x_i), c=1,…,N; n_c is the number of the training examples belonging to class c; and I_c is the set of indices of the training examples of class c, and we have use the Gaussian window function. Using the Bayesian rule, an the estimate the conditional *pdf* $\hat{p}(x|c)$, the conditional probability is:

$$\hat{p}(c|x) = \frac{\sum_{i \in I_c} \exp\left(-\frac{(x - x_i)^T \sum^{-1} (x - x_i)}{2h^2} \right)}{\sum_{k=1}^{N} \sum_{i \in I_k} \exp\left(-\frac{(x - x_i)^T \sum^{-1} (x - x_i)}{2h^2} \right)} \tag{11}$$

Therefore, using n training samples, the conditional entropy, assuming that each sample has the same probability is:

$$\hat{H}(C|X) = -\sum_{j=1}^{n} \frac{1}{n} \sum_{c=1}^{N} \hat{p}(c|x_j) \log \hat{p}(c|x_j) \tag{12}$$

Where x_j is the *j*th sample of the training data.

4. Defining the neuro-fuzzy system structure using pseudo-gaussian functions

The output of the networks is defined as the weighted average $\tilde{F}^*_{RBF}$ of the radial basis function (instead of the classical weighted sum) with the addition of lateral connections between the radial neurons, as follows:

$$\tilde{F}^*_{RBF}(x_n) = \frac{\sum_{i=1}^{K} w_i \phi_i(x_n, c_i, \sigma_i)}{\sum_{i=1}^{K} \phi_i(x_n, c_i, \sigma_i)} \tag{13}$$

We propose to use a pseudo-gaussian function for the nonlinear function within the hidden unit, which produce good result as presented in [10]. The output of a hidden neuron is computed as:

$$\phi_i(x) = \prod_v \varphi_{i,v}(x^v)$$

$$\varphi_{i,v}(x^v) = e^{-\frac{\left(x^v - c_i^v\right)^2}{\sigma_{i,-}^v}} U(x^v; -\infty, c_i^v) + e^{-\frac{\left(x^v - c_i^v\right)^2}{\sigma_{i,+}^v}} U(x^v; c_i^v, \infty) \qquad (14)$$

$$where \quad : \quad U(x^v; a, b) = \begin{cases} 1 & if\ a \le x^v < b \\ 0 & otherwise \end{cases}$$

The index *i* runs over the number of neurons (K) while *v* runs over the dimension of the input space (v∈ [1,D]). The weights connecting the activation of the hidden units with the output of the neural system, instead of being single parameters, are functions of the input variables. Therefore, the w_i are given by:

$$w_i = \sum_v b_i^v x^v + b_i^0 \qquad (15)$$

where b_i^v are single parameters.

5. Experimental results

The Classification step occurs after feature extraction and selection have been applied. extraction. The images were divided into two sets, the training and validation. The first 72 normal images and 87 cancer images were used for training. The second 35 normal images and 35 cancer images were used for testing. All the images used were acquired as hospital samples captured with a light microscope in combination of a CCD camera with magnification. Fig 1 shows an examples of a benign tumour and a prostate cancer.

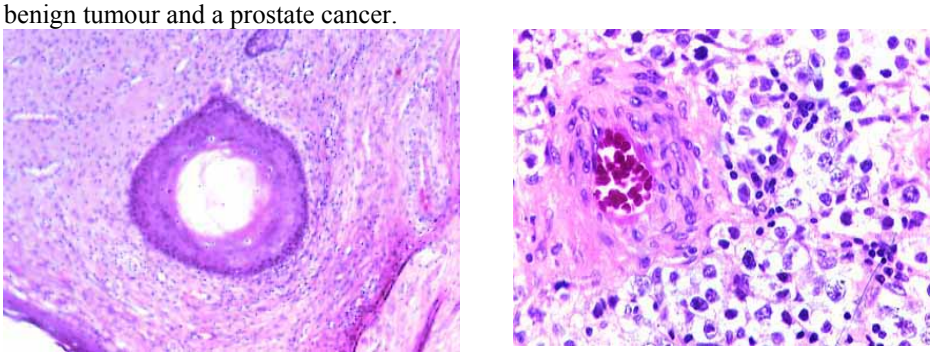

Fig 1 a) Prostate cancer b) Benign tumour

Finally, in Table 1 we present the performace of the proposed bio-inspired system, obtaining very good result for the problem of prostate cancer light micrograph classification.

Classified as / Correct Classes	Cancer	Normal	Classification Rate (%)	Error Rate
Cancer	30	5	85,7	0,142
Benign tumour	3	32	91,4	0,086

Table 1: Recognition error for the proponed algorithm. Confusion matrix

6. Conclusion

In this paper, we introduced a pseudo-gaussian basic function neural networks for prostate cancer diagnosis, based on input feature selection using mutual information. The first phase of the presented methodology consist in extracting different features from a prostate cancer light micrograph. We have used features based from the spatial domain and also from the morphology domain. In order to select the most relevant variables, and input feature selection algorithm using mutual information, based on Parzen Window is used. The final step is to use an artificial neural network to perform the diagnosis. A high classification rate has been achieved which will reduce the subjective human invention and will increase the diagnostic speed.

Acknowledgement
This work has been partially supported by the Projects CICYT TIC2002-11350-52E and TIN2004-01419.

References

[1] www.cancer.org, "Cancer Statistics 2004", downloaded on 4th April, 2004. www.prostate.com/

[2] U Patel, TRUS and prostate biopsy: current status, Prostate Cancer and prostatic diseases. ISSN 1365-7852, vol.7, pp.208-210, Septermber, 2004

[3] M Lujan, A Paez, C Santonja, T Pascual, I Fernandez & A Berenguer,"Prostate cancer detection and tumor characteristics in men with multiple biopsy sessions" Prostate Cancer and prostatic diseases, vol.7, pp.238-242, Septermber, 2004

[4] G.A. Babich and O.I. Camps, "Weighted Parzen Window for Pattern Classification," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 5, pp. 567-570, May 1996.

[5] K. Fukunaga and R.R. Hayes, "The Reduced Parzen Classifier," IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 11, no. 4, pp. 423-425, Apr. 1989.

[6] Nojun Kwak, Chong-Ho Choi," Input Feature Selection by Mutual Information Based on Parzen Window, IEEE Trans.Pattern Analysis and Mach. Intelligence,vol.24, no.12, pp.1667-1671, 2002.

[7] H.Akaike, "Information theory and extension of the maximum likelihood principle", Akademia Kiado, Budapest, pp.267-810, 1973.

[8] C. M. Bishop, " Neural Networks for Pattern Recognition", Clarendon Press, Oxford, 1995.

[9] N.B.Karayiannis, G.Weiqun Mi, "Growing Radial Basis Neural Networks: Merging Supervised and Unsupervised Learning with Network Growth Techniques", IEEE Transaction on Neural Networks, vol.8, no.6, pp.1492-1506, November 1997.

[10] I.Rojas, H.Pomares,J.L.Bernier,J.Ortega,B.Pino,F.J.Pelayo,A.Prieto , "Time series analysis using normalized PG-RBF network with regression weights", *Neurocomputing, vol.* 42 (2002)267-285