Expert Systems with Applications 36 (2009) 6357–6361

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Early prostate cancer diagnosis by using artificial neural networks and support vector machines

Murat Çınar [a], Mehmet Engin [b,*], Erkan Zeki Engin [b], Y. Ziya Ateşçi [a]

[a] Sifa Hospital Groups, Urology Department, Turkey
[b] Ege Üniversitesi, Mühendislik Fakültesi Elektrik & Elektronik Müh. Bölümü, 35100, Bornova, İzmir, Turkey

ARTICLE INFO

Keywords:
Prostate cancer
Prostate specific antigen
Artificial neural network
Support vector machine

ABSTRACT

The aim of this study is to design a classifier based expert system for early diagnosis of the organ in constraint phase to reach informed decision making without biopsy by using some selected features. The other purpose is to investigate a relationship between BMI (body mass index), smoking factor, and prostate cancer. The data used in this study were collected from 300 men (100: prostate adenocarcinoma, 200: chronic prostatism or benign prostatic hyperplasia). Weight, height, BMI, PSA (prostate specific antigen), Free PSA, age, prostate volume, density, smoking, systolic, diastolic, pulse, and Gleason score features were used and independent sample *t*-test was applied for feature selection. In order to classify related data, we have used following classifiers; scaled conjugate gradient (SCG), Broyden–Fletcher–Goldfarb–Shanno (BFGS), and Levenberg–Marquardt (LM) training algorithms of artificial neural networks (ANN) and linear, polynomial, and radial based kernel functions of support vector machine (SVM). It was determined that smoking is a factor increases the prostate cancer risk whereas BMI is not affected the prostate cancer. Since PSA, volume, density, and smoking features were to be statistically significant, they were chosen for classification. The proposed system was designed with polynomial based kernel function, which had the best performance (accuracy: 79%). In Turkish Family Health System, family physician to whom patients are applied firstly, would contribute to extract the risk map of illness and direct patients to correct treatments by using expert system such proposed.

1. Introduction

Prostate cancer is a disease in which cancer develops in the prostate, a gland in the male reproductive system. Cancer occurs when cells of the prostate mutate and begin to multiply out of control. These cells may spread (metastasize) from the prostate to other parts of the body, especially the bones and lymph nodes. Prostate cancer progresses most frequently in men over fifty. This cancer can occur only in men, as the prostate is exclusively of the male reproductive tract. Prostate cancer is presently the most common type of cancer in men, where it is responsible for more male deaths than any other cancer, except lung cancer.

Prostate cancer diagnosis is complicated by the biological heterogeneity of the disease. There are a lot of treatment options, which has different short and long term risks and complications. These make difficult to choice the treatment for the individual. The patient dilemma has gained increased awareness among urologist. There is an obvious need for decision-making tools that individual patients and physicians can apply to the specific parameters of disease to reach an informed decision (Anagnostou et al., 2003). Prostate cancer is a potentially curable via early diagnosis for many patients. There are usually no clinical findings in early stage. However, it is diagnosed by routine control. The features such as PSA (prostate specific antigen), volume, density and etc., are used to decrease necessity of biopsy. Obesity is significantly associated with a high preoperative PSA velocity, previously shown to be associated with PSA and overall survival after treatment with surgery and radiation therapy (Loeb et al., 2006).

The low positive predictive value of PSA is a major drawback of the marker (Stephan et al., 2006). The pro-forms of PSA (-2, -5, -7 proPSA) and also %free-PSA based ANNs have been suggested to enhance the discrimination between prostate cancer (PCa) and no evidence of malignancy (NEM) by Stephan et al. (2006). They constructed leave-one-out ANN models with the variables PSA, %free-PSA, proPSA volume, and status of digital rectal examination (DRE) and compared them by receiver operating characteristic (ROC) curve analysis. They concluded that proPSA as single parameter did not improve specificity over %freePSA whereas proPSA and %freePSA within an ANN in the PSA range 4–10 mg/l substituted prostate volume and DRE.

Adenocarcinoma of the prostate was probably the first malignancy in which biological serum markers were used diagnosis, determining the response to therapy (Merseburger et al., 2001).

* Corresponding author. Tel./fax: +90 232 388 60 24.
E-mail address: mehmet.engin@ege.edu.tr (M. Engin).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.08.010

Merseburger et al., (2001) assessed serum creatinine as a putative marker for staging/prognosis in localized prostate cancer. Their results showed that the relationship of the creatinine level of PSA recurrence was not significant in the univariate or multivariable analysis. Finally, they concluded that creatinine did not provide independent information for predicting pathologic stage or disease recurrence in patients with early prostate cancer.

Several preoperative nomograms have been developed to predict the risk of prostate cancer (PCa) progression after radical prostatectomy (RP). However, only a few studies showed an accuracy of %70 to predict PSA recurrence within five years of RP, leaving room for improvement (Poulakis et al., 2004). Poulakis et al. (2004) developed and tested an artificial neural network (ANN) for predicting biochemical recurrence based on the combined use of pelvic coil magnetic resonance imaging (pMRI), prostate-specific antigen (PSA) measurement, and biopsy Gleason score in men clinically localized prostate cancer. The predictive ability of ANN was compared with that of logistic regression analysis (LRA), Han tables, and the Katton nomogram using area under ROC analysis. They concluded that ANN was superior to LRA, predictive tables, and nomograms to predict biochemical recurrence accurately by using the pMRI findings.

The philosophy behind many treatment-planning approaches is to design individual patients' plans from scratch for every new patient. The process of adjusting treatment variables and displaying the corresponding dose distribution is repeated until such criteria as dose uniformity within the target region and dose minimization to surrounding critical organs is considered optimized (Wells & Niederer, 1998). Wells and Niederer (1998) developed a medical expert system approach to standardized treatment planning that should lead to improved planning efficiency and consistency. They used a set of artificial neural networks to optimize the treatment variables to the individual patient. They conclude that an expert system approach has the potential of improving the overall efficiency of the planning process by reducing the number of iterations required to generate an optimized dose distribution and to function most effectively should be closely integrated with a dosimetric based treatment planning system.

In this study, we designed a system in order to avoid unnecessary biopsy. Three ANN and SVM algorithms have been used to early diagnose prostate cancer by using prostate volume, density and etc. features. We also investigated that BMI is a risk factor or not for aggressive pathologic findings.

2. Materials and methods

The data from 300 men (mean age: 63 years and range: 43–93 years) were collected at the Urology Department of Bornova Sifa Hospital, Izmir, Turkey since August 2006 to May 2007. The distribution of the patients into pathology classes was prostate adenocarcinoma (100), chronic prostatism or benign prostatic hyperplasia (200). The diagnosis was performed by routine examination methods.

Following features were used in this study; weight, height, BMI, PSA, Free PSA, age, prostate volume, density, smoking, systolic, diastolic, pulse, and Gleason score. BMI is computed as follows:

$$\mathrm{BMI} = [\mathrm{Weight(kg)}]/[\mathrm{Height(m)}]^2 \quad (1)$$

Patients were divided into three groups based upon BMI. BMI was calculated for each patient and the cohort degree was divided into 'normal' (BMI: 18–24.9 kg/m^2), 'overweight' (BMI: 25–29.9 kg/m^2) or 'obese' (BMI $\geqslant$ 30 kg/m^2).

Prostate cancer is most often diagnosed by physical examination or screening blood tests such as PSA test. PSA is presently the most widely used tumour marker and for early detection of prostate cancer (Wingo, Landus, & Ries, 1997). Prostate volume had got with transrectal ultrasonography. We did not add digital rectal examination as input feature, because this feature is relative from doctor to doctor.

Suspected prostate cancer is typically confirmed by removing a piece of the prostate (biopsy) and examining it under a microscope by pathologist at pathology laboratory. Thus, Gleason score grade is derived.

Among the strategies in risk stratification for prognostic groups, two methods can be used; nomograms and ANNs. Nomograms make predictions based on the characteristics of the individual patient. It is a graphical representation of a statistical model, with scale for calculating the prognostic weight of value for each individual variable, with the goal of predicting a particular end point. End points predicted in current prognostic models include pathologic stage (Partin, Yoo, Carter, et al., 1993), disease/progression – free probability (Katton & Scardino, 2002), disease – specific survival (Smaletz, Scher, Small, et al., 2002). However, nomograms do not make treatment recommendations and should not act as a surrogate for physician–patient interactions.

2.1. Artificial neural network (ANN)

ANNs comprise an exploratory approach that may improve predictive modeling and are inspired by the principle of neural networks and contain layers of nodes (Fig. 1). ANN is formed from an input, middle (hidden), and output layer. The layers are richly interconnected by weighed connection lines. Each information is weighted and can increase or decrease the activation of the node. At a given bias the node starts to fire (Gamito, Crawford, & Errejon, 2003). An ANN model must first be "trained" by using cases with known outcomes. It will then adjust its weighting of various input variables over time to refine output data. The performance of the ANN is then evaluated using a validation data set respect to sensitivity and specificity value of the model. ANN can resolve nonlinear complex relations among input variables, without the need for any prior assumptions about these relations.

In this study, back-propagation based multi-layer perceptron (MLP) network was used. Scale-conjugate (SCG), Brodyen–Fletcher–Goldfarb–Shanno (BFGS), and Levenberg–Marquardt (LM) learning algorithms were tested. The most important factor in the MLP structure is the choice of the number of the hidden neurons. The weights are adapted in the learning phase of the network using gradient method. The weights are adapted from cycle to cycle according to the information of gradient of the error function in the gradient method of learning

$$w(k+1) = w(k) + \eta p(k) \quad (2)$$

where η is learning coefficient calculated each cycle and $p(k)$ is search direction vector of minimization in the kth cycle. The following algorithms were used to optimize the learning coefficients.

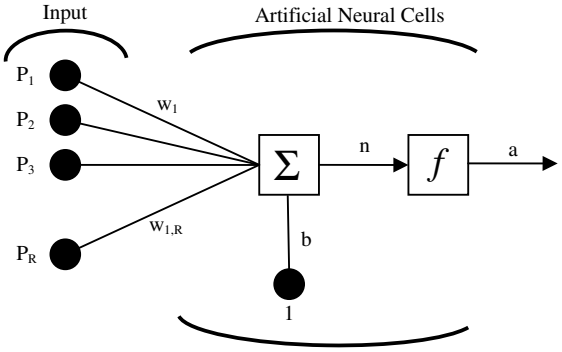

Fig. 1. Basic artificial neural cell structure.

2.1.1. Scaled-conjugate gradient (SCG) algorithm

The basic back-propagation algorithm adjusts the weights in the direction of the negative of the gradient. Although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. A search is performed along conjugate directions, which produces generally faster convergence in the conjugate gradient (CG) algorithms. SCG is fully automated including no user dependent parameters and avoids a time consuming line search (by using a Levenberg–Marquardt approach) (Moller, 1993).

2.1.2. Brodyen–Fletcher–Goldfarb–Shanno (BFGS) algorithm

In Newton's method, the Hessian matrix (second derivatives) of the performance index at the current values of the weights and biases is calculated. Newton's method often converges faster than CG methods. However, it is complex and expensive to compute the Hessian matrix for feed-forward neural networks. There are algorithms that are based on Newton's method, but which do not require calculation of second derivatives. These are called quasi-Newton methods. They update an approximate Hessian matrix at each iteration of the algorithm. The most successful quasi-Newton method is BFGS update. This algorithm requires more computation in each iteration and more storage than the CG methods, although it generally converges in fewer iterations.

2.1.3. Levenberg–Marquardt (LM) algorithm

LM algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feed-forward networks), then the Hessian matrix can be approximated as and the gradient can be computed as where is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard back-propagation technique that is much less complex than computing the Hessian matrix. This algorithm appears to be the fastest method for training moderate-sized feed-forward neural networks (up to several hundred weights).

2.2. Support vector machine (SVM)

SVMs are an attractive approach to data modeling. They combine generalization control with a technique to address the curse of dimensionality. SVM finds a separating hyperplane that separates two classes with a maximal margin. The hyperplane is placed halfway between the two classes in order to maximize the margin (Begg, Palaniswami, & Owen, 2005). A side-effect of the SVM algorithm is that it also identifies support vectors. Support vectors are the data points that are solely responsible for the solution and lie either on or within the margin. If the SVM algorithm was only given the support vectors as data points, the same separating hyperplane would be derived.

The data is not necessarily linearly separable in all cases. This can be due to non-linearities in the underlying process generating the data points or due to noise in the measurements. The SVM algorithm can still find a maximal margin separating hyperplane and accepts the fact that a "few" data points may not be classified correctly and lie within the margin. The SVM is also capable of separating classes of data using non-linear curves. This is done by mapping the input data through a nonlinear mapping to a new space where a linear hyperplane can still separate the data. When the separating hyperplane is mapped back to the original space, it appears as a curve (Fig. 2).

SVMs map the training data non-linearly into a higher-dimensional feature space. This yields a non-linear decision boundary in input space. By the use of a kernel function, it is possible to compute the separating hyperplane without explicitly carrying out the map into the feature space (Joachims, 1998). The kernel mapping provides a unifying framework for most of the commonly employed model architectures, enabling comparisons to be performed.

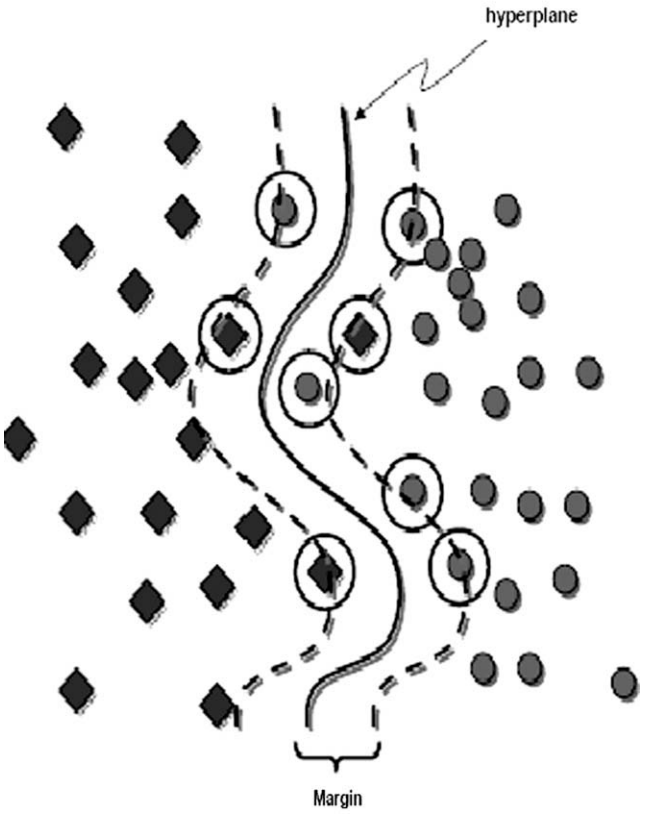

Fig. 2. Classification of data using non-linear curves.

2.3. Statistical test

Independent sample *t*-test was used for feature selection. Statistical analyses were performed by using the Statistical Package for Social Sciences (SPSS). A statistically significant level of $p = 0.05$ was used.

3. Results

The descriptive statistics of preoperative parameters were given in Table 1. Smoking rate for prostate cancer and normal patients was given in Table 2. According to Table 2, smoking rate of patients with prostate cancer is more than smoking rate of normal patients. According to these figures, smoking could be a risk factor for prostate cancer.

From Table 3, we observed that BMI is not affected to discriminate healthy and pathological subjects. However, there is a direct relationship between BMI and aggressive pathologic findings according to Table 4.

The outcome of independent *t*-test was given in Table 5. Since PSA, volume, density, and smoking features were to be statistically significant, they were chosen for classification.

In order to predict success of the classifier, sensitivity (true positive ratio: TPR), specificity (true negative ratio: TNR), and accuracy (true ratio: TR) were calculated by analyzing the figures coming from the applications. The sensitivity was calculated by

Table 1
The descriptive statistics of preoperative parameters: (a) prostate cancer patients, (b) BPH (benign prostatic hyperplasia) patients

Features	*N*	Min.	Max.	Mean	SD
(a)					
Weight	100	53	120	75.4	13.22
Height	100	1.55	1.86	1.69	0.06
BMI	100	0.18	0.42	0.26	0.04
PSA	100	1.03	187	22.40	29.27
Volume	100	14	121	50.75	23.60
Age	100	49	93	68.27	9.15
Density	100	0.03	4.17	0.52	0.69
Smoking	100	0	1	0.69	0.47
Systolic	100	90	170	128.9	17.4
Diastolic	100	50	120	80	9.95
Pulse	100	64	96	79.74	6.02
(b)					
Weight	200	40	135	77.24	13.04
Height	200	1.50	1.90	1.71	0.07
BMI	200	0.14	0.42	0.26	0.41
PSA	200	0	39	5.51	6.57
Volume	200	10	244	62.49	35.91
Age	200	43	92	67.61	8.35
Density	200	0.01	1.11	0.09	0.14
Smoking	200	0	1	0.29	0.46
Systolic	200	90	210	132.83	19.04
Diastolic	200	40	130	81.85	11.49
Pulse	200	52	120	79.02	7.79

N: Number of subjects.

Table 2
Smoking scores

Feature	Prostate cancer		Normal	
	N	(%)	*N*	(%)
Smoking	69	69	58	29
Non-smoking	31	31	142	71

Table 3
(a) BMI frequency of prostate cancer patients, (b) BMI frequency of BPH (benign prostatic hyperplasia) patients

BMI (kg/m^2)	*N*	(%)
(a)		
$\leqslant$24.9	38	38
25–29.9	47	47
$\geqslant$30	15	15
(b)		
$\leqslant$24.9	73	36.5
25–29.9	92	46.0
$\geqslant$30	35	17.5

Table 4
Gleason score

Gleason score	BMI $\leqslant$ 24.9		25 $\leqslant$ BMI $\leqslant$ 29.9		BMI $\geqslant$ 30	
	N	(%)	*N*	(%)	*N*	(%)
<5	0	0	0	0	0	0
5	5	16	3	5	0	0
6	7	22	13	25	4	25
7	9	28	15	29	5	31
8	3	9	6	12	2	13
9	8	25	15	29	5	31

dividing the total of recognized numbers (true positive: TP) to the sum of the true positive (TP) and false negative (FN)

$$\text{Sensitivity}(\%) = \text{TPR} = \frac{\text{TP}}{\text{TP}+\text{TN}} \times 100 \quad (3)$$

Table 5
The outcomes of independent *t*-test

Feature	Sig. (2-tailed)	Mean difference	Std. error difference
Weight	0.269	−1.7750	1.6041
Height	0.059	−0.1465	0.0077
BMI	0.738	−0.0017	0.0050
PSA	<0.001	16.8621	2.1680
Volume	0.003	−11.741	3.9611
Age	0.532	0.660	1.0560
Density	<0.001	0.417	0.0504
Smoking	<0.001	0.400	0.0560
Systolic	0.084	−3.930	2.2670
Diastolic	0.171	−1.850	1.3470
Pulse	0.415	0.725	0.8880

Table 6
The averaged classification results of CV data sets at test phase by using ANN

Algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
SCG	81.0	77.9	79.1
BFGS	79.0	78.8	78.9
LM	76.2	80.8	79.3

Table 7
The averaged classification results of CV data sets at test phase by using SVM

Kernel function	Sensitivity (%)	Specificity (%)	Accuracy (%)
Linear	76.1	79.4	77.2
Polynomial	84.2	74.8	81.1
Gaussian	82.0	72.8	78.9

Sensitivity was also known as the true positive ratio. Specificity was known as the true negative ratio and was calculated as follows:

$$\text{Sensitivity}(\%) = \text{TNR} = \frac{\text{TN}}{\text{TN}+\text{TP}} \times 100 \quad (4)$$

where FP was false positive value. Accuracy was also known as the true ratio and was calculated as follows:

$$\text{Accuracy}(\%) = \text{TR} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}} \times 100 \quad (5)$$

Cross-validation (CV) is a standard test commonly used to test the ability of the classification system using various combinations of the testing and training data sets (Begg et al., 2005). CV is often used for comparing two or more learning ANN models to estimate which model will perform the best on the problem at hand (Subaşi, 2005). CV process was repeated until every data set was included in the testing data set with 10-fold. Each frame of CV data set had 50 prostate cancers and 100 normal whereas their respective training data set included the remaining 50 prostate cancer and 100 normal. The 10-test set scores for each learning model were then averaged.

The obtained classification results of CV data sets by various algorithms at test phase were given in Tables 6 and 7. These results show that support vector machine algorithm gives better result than artificial neural networks learning algorithms. Polynomial based kernel function of SVM gave the best result (sensitivity: %79, specificity: %78.8).

4. Discussion

In this study, we observed that the features vary upon the age and degree of pathology of the patient. Thus, in order to increase

the performance of classification, we have to use lots of data for all ages and pathologies (different stages).

In this work, we have used support vector machine and back-propagation (SCG, BFGS, LM) based MLP for classification. The performance of these networks can be compared with statistical methods and other ANNs models like SOM.

The prostate cancer detection rate showed a linear relationship with age. We realize that SVM and ANNs are as successful as nomograms to predict prostate cancer. In addition, there is no relationship between BMI and prostate cancer. However, BMI can be a risk factor for aggressive pathologic findings.

According to Gleason score grade (Table 4), it founded that early diagnosis of prostate cancer is low for Turkish people. Due to our opinion, the main reason may be that the people do not care about routine health control.

5. Conclusions

In Turkish Family Health System, family physician to whom patients are applied firstly, would contribute to extract the risk map of illness and direct patients to correct treatments by using expert system.

In the future study, we think that enlargement of database and testing new classifiers with diverse features improve the performance.

References

Anagnostou, T., Remzi, M., Lykourinas, M., & Djavan, B. (2003). Artificial Neural Networks for decision-making in urologic oncology. *European Urology, 43*, 596–603.

Begg, R. K., Palaniswami, M., & Owen, B. (2005). Support vector machines for automated gait classification. *IEEE Transactions on Biomedical Engineering, 52*, 828–838.

Gamito, E. J., Crawford, E. D., & Errejon, A. (2003). Artificial neural networks for predictive modelling in prostate cancer. In J. H. Mydlo & J. C. Godec (Eds.), *Prostate Cancer; Science and Clinical Practise* (pp. 167–172). San Diego: Academic Press.

Joachims, T. (1998). *Making large-scale SVM learning practical to be published in advances in kernel methods – support vector learning*. MIT Press.

Katton, M. W., & Scardino, P. T. (2002). Prediction of progression nomograms of clinical utility. *Clinical Prostate Cancer, 1*, 90–96.

Loeb, S., Yu, X., & Nadler, R. B. (2006). Does body mass index affect outcomes following radical prostatectomy. *Journal of Urology*, 175.

Merseburger, A. S., Connelly, R. R., Sun, L., Richter, E., & Moul, J. W. (2001). Se of serum creatinine to predict pathologic and recurrence among radical prostatectomy patients. *Urology, 58*(5), 729–734.

Moller, M. F. (1993). A scale conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*, 525–533.

Partin, A. W., Yoo, J., Carter, H. B., et al. (1993). The use of PSA clinical stage and Gleason score to predict pathological stage in men with localised prostate cancer. *Journal of Urology, 150*, 110–114.

Poulakis, V., Witzsch, U., Vries, R., Emmerlich, V., Meves, M., Altmannsberger, H. M., et al. (2004). Preoperative neural network using combined magnetic resonance imaging variables, prostate-specific antigen, and Gleason score for predicting prostate cancer biochemical recurrence after radical prostatectomy. *Urology, 64*(6), 1165–1170.

Smaletz, O., Scher, M. O., Small, E. J., et al. (2002). Nomogravn for overall survival of patients with progressive metastatic prostate cancer after castration. *Journal of Clinical Oncology, 20*, 3972–3982.

Stephan, C., Meyer, H. A., Kwiatkowski, M., Recker, F., Cammann, H., Loening, S. A., et al. (2006). A (-5, -7) proPSA based artificial neural network to detect prostate cancer. *European Urology, 50*(5), 1014–1020.

Subaşi, A. (2005). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications, 28*, 701–711.

Wells, D. M., & Niederer, J. (1998). A medical expert system approach using artificial neural networks for standardized treatment planning. *International Journal of Radiation Oncology Biology Physics, 41*(1), 173–182.

Wingo, P. A., Landus, S., & Ries, L. A. (1997). An adjustment to the 1997 estimate for new prostate cancer. *Journal of Urology, 80*, 1810–1813.