Cell-nuclear data reduction and prognostic model selection in bladder tumor recurrence

Dimitris K. Tasoulis [a,b], Panagiota Spyridonos [c],
Nicos G. Pavlidis [a,b], Vassilis P. Plagianakos [a,b],
Panagiota Ravazoula [d], Georgios Nikiforidis [c],
Michael N. Vrahatis [a,b,*]

[a] Computational Intelligence Laboratory, Department of Mathematics, University of Patras,
GR—26110 Patras, Greece
[b] University of Patras Artificial Intelligence Research Center (UPAIRC), University of Patras,
GR—26110 Patras, Greece
[c] Computer Laboratory, School of Medicine, University of Patras, GR—26110 Patras, Greece
[d] Department of Pathology, University Hospital, University of Patras, GR—26110 Patras, Greece

Summary

Objective: The paper aims at improving the prediction of superficial bladder recurrence. To this end, feedforward neural networks (FNNs) and a feature selection method based on unsupervised clustering, were employed.
Material and methods: A retrospective prognostic study of 127 patients diagnosed with superficial urinary bladder cancer was performed. Images from biopsies were digitized and cell nuclei features were extracted. To design FNN classifiers, different training methods and architectures were investigated. The unsupervised k-windows (UKW) and the fuzzy c-means clustering algorithms were applied on the feature set to identify the most informative feature subsets.
Results: UKW managed to reduce the dimensionality of the feature space significantly, and yielded prediction rates 87.95% and 91.41%, for non-recurrent and recurrent cases, respectively. The prediction rates achieved with the reduced feature set were marginally lower compared to the ones attained with the complete feature set. The training algorithm that exhibited the best performance in all cases was the adaptive on-line backpropagation algorithm.

* Corresponding author. Tel.: +30 2610 997374; fax: +30 2610 992965.
  E-mail addresses: dtas@math.upatras.gr (D.K. Tasoulis), spyridonos@med.upatras.gr (P. Spyridonos), npav@math.upatras.gr (N.G. Pavlidis), vpp@math.upatras.gr (V.P. Plagianakos), gnikif@med.upatras.gr (G. Nikiforidis), vrahatis@math.upatras.gr (M.N. Vrahatis).

Conclusions: FNNs can contribute to the accurate prognosis of bladder cancer recurrence. The proposed feature selection method can remove redundant information without a significant loss in predictive accuracy, and thereby render the prognostic model less complex, more robust, and hence suitable for clinical use.

1. Introduction

Transitional cell cancer (TCC) of the bladder is a heterogeneous group of tumors accounting for approximately 90% of neoplasms arising in the urinary bladder [1]. Superficial bladder tumors (stages Ta, TIS and T1) account for 70—80% of neoplasms, while the remaining 20—30% are invasive (T2, T3, T4), or metastatic at the time of initial presentation. Over 60% of patients affected with superficial tumors will have one or more recurrences after initial treatment [2]. If left undetected and untreated, recurrent tumors may progress, thereby worsening the prognosis of the patient [3]. The early detection of recurrence is therefore crucial for patient management. Typically, the detection of recurrence is achieved by regular patient follow-up with cytoscopy and cytological examinations. The relatively low sensitivity of these methods has motivated the development of more effective and non-invasive techniques for monitoring patients [4,5]. The alternatives investigated up to date constitute improvements of cytoanalysis that replace subjective qualitative methods with quantitative assessments [6,7], as well as, enhancements of recurrence predictability at patient level [2]. In [6], high-resolution image cytometry was employed to distinguish with high reliability urothelial neoplasia from normal urothelium, examining only diploid cell nuclei. In [7] the authors report that the measurement of the rapidly hydrolyzed component of DNA present in the nuclei of bladder urothelium offers a highly sensitive and reliable supplement to the qualitative and subjective cytologic procedures currently in practice. In this study we explore means to improve the prognosis of cancer recurrence by examining specimens of transitional cell carcinomas. We further attempt to identify prognostic markers that enable the prediction of real differences in the biological behavior of tumors.

To predict the biological behavior of superficial TCCs, histological grade, clinical stage, urinary cytology, tumor number and tumor size, are commonly used [8]. These factors, however, have been reported to be unsatisfactory for the prediction of tumors and their results have a rather low reproducibility for the prognosis of tumor recurrence [2].

The need to improve prediction at patient level, has motivated an ongoing research on computerized methods based mainly on artificial intelligence approaches [3,9—11]. Previous approaches for the prediction of cancer recurrence require information obtained, either subjectively from pathologists (including stage, grade and tumor size among others) [9—11], or through tedious and costly techniques for evaluating molecular markers [3]. The proposed prognostic system for superficial TCC recurrence exploits quantitative information about cell nuclei appearance. In cancer, cell nuclei undergo significant changes which, if quantified, can allow the diagnosis and potentially the prediction of the course of the disease [12—14]. Van Velthoven et al. [15] applied computer assisted microscopy to Feulgen stained nuclei to characterize the risk of recurrence in patients with superficial TCC. Kaplan—Meier discriminant analysis indicated a cut-off value ($p < 10^{-5}$) for distinguishing between TCC cases with remission for more than 60 months and cases presented with recurrence within 12 months. In this work, a panel of 36 cell nuclei features was estimated from hematoxylin—eosin stained tissue sections from patient biopsies, using quantitative light microscopy.

Artificial neural networks (ANNs), and in particular feedforward neural networks (FNNs), were employed to determine the prognostic information conveyed by nuclear features for the prediction of bladder cancer recurrence. The capabilities of ANNs have attracted considerable attention in numerous clinical, diagnostic and prognostic tasks [16—18]. One of the most widely known applications of ANNs in medicine is the PAPNET system which is designed for the automated cytological screening of cervical smears [19]. Kolles et al. [20] created a system for grading astrocytomas based on immunohistochemically and DNA stained microscopic images. ANN based decision tools for breast cancer diagnosis have been investigated by several researchers [21—23]. Downs et al. [24] studied pruning strategies to design ANNs capable of shifting the balance between sensitivity and specificity, depending on the requirements of the medical application.

To identify feature subsets that convey the most significant prognostic information, a feature selection methodology based on clustering was

employed. We considered two clustering algorithms, fuzzy *c*-means (FCM) [25] and the unsupervised *k*-windows algorithm (UKW) [26,27]. The main contribution of the paper is the identification of appropriate network architectures and training algorithms to construct FNNs capable of predicting cancer recurrence with high accuracy, and in the investigation of the impact of feature selection through cluster analysis on classification accuracy.

2. Material and methods

The proposed methodology was applied on data collected from 127 patients with bladder cancer. Patients were diagnosed and followed-up during the period 1991—2000 at the University Hospital of Patras, Greece. The follow-up period was at least 60 months. Out of the 127 patients, 56 patients had no recurrence during the observation interval. The remaining 71 patients experienced recurrence of the tumor over a time period ranging between 1 month and 5 years. In Table 1, the grade diagnosis and follow-up of the patients is reported, while Fig. 1 depicts the time interval between initial treatment and the first recurrent event for all recurrent cases.
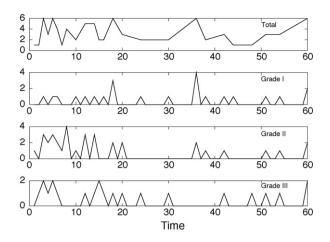
The software used for all the experiments performed, for the training and the evaluation of the FNNs and the clustering algorithms for feature selection was developed under the Linux operating system using the C++ programming language, and the gcc 3.3.4 compiler. Additionally, all the numeric data values for each feature were rescaled in the range $[-1, 1]$.

2.1. Image acquisition and feature extraction

Hematoxylin—eosin tissue sections from the biopsies of the 127 different patients were collected from the Department of Pathology at the University Hospital of Patras. For each slide (tissue section), a pathologist specified the most representative region. From this region, images were acquired (at a resolution of $768 \times 576 \times 8$ bit) using a light

Figure 1 Time interval between initial treatment and the first recurrent event for all patients, and for each grade category.

Zeiss Axiostar Plus microscope connected to an Ikegami color video camera. Histological images of a recurrent bladder TCC, and a non-recurrent bladder TCC are presented in Fig. 2. For each case, 36 features were estimated automatically from morphological and textural nuclear features [28]. Information about nuclear size and shape was captured by 18 morphological features, which constituted measurements of nuclear area, roundness and concavity [28]. The feature of concavity, attempts to measure the severity of concavities, or the indentations of a nucleus [28]. In each case, the mean value, standard deviation, range, skewness and kurtosis of each morphological feature was computed to describe non-uniform modifications of nuclear morphology. Non-uniform alterations in nuclear size and shape have proved to convey significant diagnostic and prognostic information [11,28]. Taking into consideration that only a few malignant cells might occur in a given sample, the maximum value for each morphological feature was estimated, by averaging the three largest values.

The remaining 18 features were textural features that encoded chromatin distribution of the cell nucleus. These features were estimated by means of nuclear histograms and the co-occurrence matrix [29]. Nuclear chromatin-texture quantification has been examined in several studies, and has proved to carry significant diagnostic information in the analysis of pathologic material [30—32]. To quantify texture properties of nuclei, textural features were formed from first order statistics and from spatial gray tone co-occurrence probability matrices [29,33]. The gray level co-occurrence matrix was used for second order texture information extraction from cell nuclei. A co-occurrence matrix *P* is an estimate of the second order joint conditional
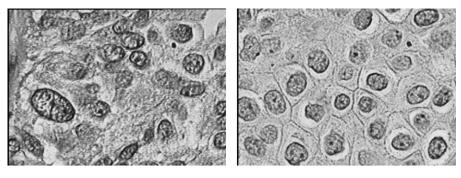
Table 1 Grade diagnosis and follow-up of 127 patients with urinary bladder carcinoma

	Number	Recurrent cases	Non-recurrent cases
Grade I	33	18	15
Grade II	57	35	22
Grade III	37	18	19

Figure 2 Histological images of a recurrent bladder TCC (left) and a non-recurrent bladder TCC (right).

probability density function (PDF) $P(i, j|d, \Phi), \Phi^\circ = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Each $P(i, j|d, \Phi)$ is the probability of transition from gray level i to gray level j, given an inter-sample spacing of d, and the direction is given by the angle Φ. Numerous features can be extracted from co-occurrence matrices and a large number of such features has been proposed [29,33]. Most authors, however, agree that in practice only a few of these are independent. In this work we adopted the following features, which appear to be the most effective in texture discrimination, as it is also reported by other researchers [31,34,35].

Energy (uniformity):

$$E = \sum_i \sum_j (P(i, j|d))^2. \tag{1}$$

Entropy:

$$H = -\sum_i \sum_j P(i, j|d) \log P(i, j|d). \tag{2}$$

Local homogeneity (inverse different moment):

$$L = \sum_i \sum_j \frac{1}{1 + (i - j)^2} P(i, j|d). \tag{3}$$

Inertia:

$$I = \sum_i \sum_j (i - j)^2 P(i, j|d). \tag{4}$$

Correlation:

$$C = \frac{\sum_i \sum_j (i - \mu_x)(j - \mu_y) P(i, j|d)}{\sigma_x \sigma_y}. \tag{5}$$

Cluster shade:

$$CS = \sum_i \sum_j ((i - \mu_x) + (j - \mu_y))^3 P(i, j|d). \tag{6}$$

Cluster prominence:

$$CP = \sum_i \sum_j ((i - \mu_x) + (j - \mu_y))^4 P(i, j|d), \tag{7}$$

where

$$\mu_x = \sum_i i \sum_j P(i, j|d), \tag{8}$$

$$\mu_y = \sum_j j \sum_i P(i, j|d), \tag{9}$$

$$\sigma_x^2 = \sum_i (i - \mu_x)^2 \sum_j P(i, j|d), \tag{10}$$

$$\sigma_y^2 = \sum_j (j - \mu_y)^2 \sum_i P(i, j|d). \tag{11}$$

Each co-occurrence feature was estimated for inter-sample spacing $d = 1$ and 3. Since each gray level is not efficiently employed to describe the cell nucleus, the co-occurrence matrix ($N_g \times N_g$) should have fewer entries than the total number of pixels in the nucleus to avoid a sparse matrix. To reduce the number of gray levels, each nuclear image was re-quantized to 16 levels ($i, j \in N_g, N_g = \{17, 33, 49, 65, 80, 96, 112, 128, 144, 160, 176, 192, 207, 223, 239, 255\}$) via histogram equalization. Thus, sufficient resolution was retained to characterize the texture of each nucleus and $N_g \times N_g$ was small enough compared to the total number of pixels in the nucleus [35,36]. In Tables 2 and 3 the, minimum (min), mean, maximum (max) values, as well as, the standard deviation for each cell nuclei descriptor is reported.

2.2. Supervised training of FNN classifiers

The efficient supervised training of FNNs is a subject of considerable ongoing research and numerous algorithms have been proposed to this end. Supervised training amounts to the global minimization of the network's error function. To select an appropriate training algorithm for the TCC recurrence data, the following methods were considered:

- resilient backpropagation (RPROP) [37];
- improved resilient backpropagation (IRPROP) [38];
- scaled conjugate gradient (SCG) [39];
- backpropagation with variable stepsize (BPVS) [40];
- adaptive on-line back propagation (AOBP) [41].

Other classical neural network training algorithms, such as the backpropagation (BP) training

Table 2 Overview of morphometry cell nuclei descriptors used in the study

Morphometry				
	Min	Mean	Max	S.D.
Area				
Mean	544.06	1025.9	1885.1	299.52
Standard deviation	64.265	339.265	1782.420	230.938
Kurtosis	1.7581	4.1067	23.9840	3.2643
Skewness	− 0.46917	0.76624	4.58530	0.78870
Maximum	674.00	1746.2	5390.70	772.56
Range	226	1486.8	10435	1186.6
Roundness				
Mean	1.0463	1.1638	1.4200	0.0566
Standard deviation	0.031300	0.084055	0.387450	0.044857
Kurtosis	1.8883	5.0353	32.3880	4.0712
Skewness	− 4.17600	0.96778	2.94850	0.91763
Maximum	1.13210	1.3551	2.33400	0.1512
Range	0.13875	0.39229	1.988	0.25768
Concavity				
Mean	0.0357	0.1136	0.1732	0.0301
Standard deviation	0.0295160	0.0482554	0.0776810	0.0084507
Kurtosis	1.57440	2.6949	6.74940	0.60625
Skewness	− 0.76241	0.21314	1.71540	0.37863
Maximum	0.111231	0.204260	0.302060	0.036419
Range	0.123400	0.199437	0.349740	0.03702

algorithm [42], the adaptive BP [40], and the BP with momentum [40] were also tested. However, the classification accuracy obtained through these methods was lower than that of the more advanced methods used here. Regarding network architecture, it has been shown that FNNs with a single hidden layer can approximate any continuous function uniformly on any compact set and any measurable function to any desired degree of accuracy [43,44]. Based on this theorem, we restricted the

Table 3 Overview of texture cell nuclei descriptors used in the study

Texture				
	Min	Mean	Max	S.D.
Densitometry				
Mean density	92.869	141.535	171.584	17.992
Mean variance	7.6913	27.510	50.9261	11.2179
Mean skewness	− 0.551440	0.068525	0.983200	0.278486
Mean kurtosis	2.11332	2.97061	4.57328	0.37377
Co-occurrence features, $d = 1$				
Mean energy	0.0085558	0.0111559	0.0187034	0.0017259
Mean entropy	4.36491	4.82126	5.06320	0.14185
Mean inertia	1091.14	2107.08	3265.34	431.07
Mean local homogeneity	0.158023	0.198898	0.256618	0.019235
Mean correlation	0.556898	0.7178	0.846328	0.0531
Mean cluster Shade	1.8137e + 05	3.5332e + 08	6.6557e + 08	1.4974e + 08
Mean cluster Prominence	1.4885e + 05	2.0678e + 07	3.9880e + 08	7.5839e + 07
Co-occurrence features, $d = 3$				
Mean energy	0.0047239	0.0067035	0.0125843	0.0013918
Mean entropy	4.69565	5.17909	5.43461	0.15972
Mean inertia	3040.2	6131.4	9429.1	1324
Mean local homogeneity	0.075	0.099478	0.141362	0.013314
Mean correlation	0.029414	0.195414	0.461123	0.081314
Mean cluster shade	1.3372e + 05	1.8989e + 08	3.7098e + 08	7.9582e + 07
Mean cluster prominence	− 5.3057e + 04	1.1791e + 07	2.1182e + 08	4.3213e + 07

network topology of FNNs to one hidden layer. The number of nodes in the hidden layer was determined experimentally.

To evaluate the performance of FNNs the leave-one-out (LOO) method was employed [45]. According to this method, an FNN is trained using all the patterns from the training set excluding one. The excluded pattern is subsequently used to assess the classification ability of the network. This process is repeated excluding a different pattern of the training set each time until all patterns of this set are excluded once. The LOO method is typically used in cases where the size of the training set is so small that cross validation and early stopping using a separate validation set result in a training set of insufficient size. In general, the optimal choice of network architecture, and number of training epochs for a particular training algorithm is a very hard task. To this end, single hidden layer networks with the number of hidden neurons ranging from three to ten were considered. For each network architecture and training algorithm the number of training epochs was varied from 100 to 1000 with a step size of 100. Finally, for each combination of training algorithm, network architecture and number of training epochs, 100 simulations were performed.

A comparative evaluation of the performance of FNNs with well-known and widely used classifiers, namely the Bayesian classifier, probabilistic neural networks and the k-nearest neighbor classifier, was carried out.

2.3. Feature selection through cluster analysis

An important issue, in any classification task is to determine the features that significantly contribute to classification accuracy. This procedure is called dimension reduction, and it has been extensively studied in the context of classification [46]. The problem of high dimensionality is often tackled by asking the user to specify the subspace (a subset of the dimensions) of interest. However, user-identification of subspaces is error-prone. Another approach is to apply a dimensionality reduction method to the dataset. Methods such as principal component analysis [46], optimally transform the original data space into a lower dimensional space by forming dimensions that are linear combinations of given attributes. The new space has the property that distances between points remain approximately the same as before the transformation. These techniques can be successful in reducing the dimensionality while retaining a sufficiently large portion of the information content of the original data. However, the new dimensions are difficult to interpret. Moreover, to compute the new

subset of dimensions, information from all the original dimensions is required. Other commonly used techniques are random projections to subspaces [47,48] and the selection of a subset of attributes through clustering [49,50]. In this paper we perform dimension reduction through two clustering algorithms, namely UKW [27], and FCM [25]. The algorithms are applied over the entire dataset to detect clusters of features. Feature selection is performed by selecting from each feature cluster one representative feature. In particular, the feature with the minimum Euclidean distance to the cluster center is selected as the most representative feature of the cluster.

2.3.1. The unsupervised k-windows clustering algorithm

The k-windows clustering algorithm uses a windowing technique to discover the clusters present in a dataset. Assuming that the dataset lies in d dimensions, the algorithm initializes a number of d-dimensional windows (d-ranges) over the dataset. Subsequently, it iteratively moves and enlarges these windows, in order to capture all the patterns of a cluster within a window. The movement and enlargement procedures are guided by the number of points that lie within a window, and are terminated when they do not increase this number significantly. The final set of windows is the result of the algorithm.

A fundamental issue in cluster analysis, independent of the particular clustering technique applied, is the determination of the number of clusters present in a dataset. For instance well-known and widely used iterative techniques, such as the k-means algorithm [51], as well as the FCM algorithm, require from the user to specify the number of clusters present in the dataset prior to the execution of the algorithm. On the other hand, the UKW algorithm approximates the number of clusters through an extension of the original algorithm. For a comprehensive description of the algorithm and an investigation of its capability to automatically identify the number of clusters present in a dataset, refer to [26,52–56].

2.3.2. The fuzzy c-means clustering algorithm

This algorithm [25], considers each cluster as a fuzzy set. It initializes a number, c, of prototype vectors (*centroids*), p^j, over the dataset, that represent the centers of the clusters. It then computes the degree of membership of each data vector, x^i, to each cluster through the following membership function:

$$\mu_j(x^i) = \left(\sum_{l=1}^{c} \left(\frac{\|x^i - p^j\|}{\|x^i - p^l\|} \right)^{1/r-1} \right)^{-1}, \qquad (12)$$

Table 4 Best results for all the considered training algorithms (complete feature set)

	Min	Mean	Max	S.D.
RPROP 36-3-2 for 300 epochs				
Specificity	83.93	89.52	94.64	2.09
Sensitivity	87.32	91.55	95.77	1.67
Overall accuracy	87.40	90.65	93.70	1.38
AUC	0.905	0.932	0.961	0.012
iRPROP 36-3-2 for 400 epochs				
Specificity	83.93	89.27	94.64	2.16
Sensitivity	84.51	90.68	95.77	2.10
Overall accuracy	86.61	90.05	92.91	1.54
AUC	0.889	0.921	0.953	0.014
SCG 36-5-2 for 200 epochs				
Specificity	75.00	85.34	92.86	3.19
Sensitivity	83.1	90.00	94.37	2.26
Overall accuracy	83.46	87.94	92.13	1.75
AUC	0.880	0.917	0.947	0.012
BPVS 36-4-2 for 400 epochs				
Specificity	85.71	89.80	92.86	1.35
Sensitivity	88.73	92.94	94.37	1.26
Overall accuracy	89.19	91.56	93.70	0.96
AUC	0.913	0.931	0.963	0.009
AOBP 36-7-2 for 500 epochs				
Specificity	87.50	90.43	94.64	1.23
Sensitivity	90.14	92.83	94.37	1.08
Overall accuracy	89.76	91.77	93.70	0.84
AUC	0.913	0.936	0.954	0.008

which assumes values in the interval [0,1], where $r \in (1, \infty)$ determines the fuzziness of the partition. If r tends to 1_+, then the resulting partition asymptotically approaches a crisp partition. If on the other hand, r tends to ∞, the partition becomes a maximally fuzzy partition. Next, the c prototypes are updated using the equation:

$$p^j = \frac{\sum_{i=1}^{n}[m_j(x^i)]^r x^i}{\sum_{i=1}^{n}[m_j(x^i)]^r}. \tag{13}$$

This procedure is executed iteratively until the change in the distortion measure, d:

$$d = \sum_{j=1}^{c}\sum_{i=1}^{n}[m_j(x^i)]^r \cdot ||x^i - p^l||^2, \tag{14}$$

drops below a user defined threshold.

3. Presentation of experimental results

The performance of the different FNNs was evaluated in terms of specificity, sensitivity and overall classification accuracy. Furthermore, a receiver operating characteristic (ROC) curve analysis, and in particular the area under the curve (AUC) was employed to measure the general predictiveness of each classifier. The AUC has an appealing statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance, which is equivalent to the Wilcoxon test of ranks [57]. Table 4, reports the best results obtained for each method with the 36-dimensional feature vector. More specifically, the table depicts the minimum (min), mean, maximum (max) and the standard deviation (S.D.), over 100 simulations. In Fig. 3 the performance of each classifier is illustrated through ROC curves. In particular, for each classifier, the ROC curves that correspond to the minimum, mean and maximum AUC in the 100 simulations are shown. Note that the ROC curve that corresponds to the mean AUC depicts the ROC curve of the classifier whose AUC is closer to the mean AUC over the 100 simulations. All the training algorithms tested exhibited remarkably high classification accuracy, with BPVS and AOBP slightly outperforming the other methods. This is a clear indication of the concise and proper feature extraction techniques used.

Subsequently, we employed the UKW and the FCM clustering algorithms to identify the most informative features for TCC recurrence prediction. The application of the UKW algorithm over the feature space for the 127 cases produced 11 clusters of features. The parameters of the UKW algorithm used
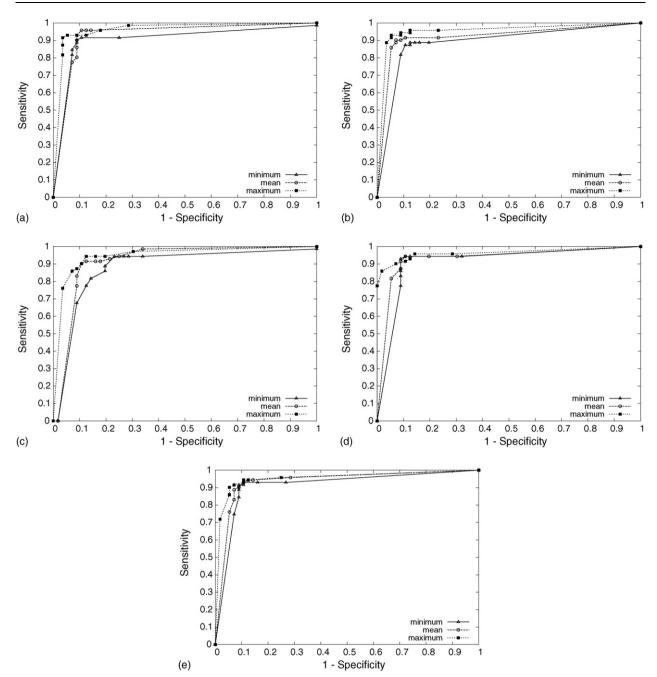
Figure 3   ROC curves: (a) RPROP, (b) iRPROP, (c) SCG, (d) BPVS and (e) AOBP (complete feature set)

were: The values of the parameters $\{\theta_e, \theta_m, \theta_c, \theta_v\}$ were set to $\{0.8, 0.1, 0.2, 0.02\}$ for the UKW algorithm [55].

As previously mentioned, from each cluster the feature closest to the corresponding cluster center was selected. From the eleven features five were textural descriptors. In detail, one was derived from cell nuclei histogram (mean variance) and four were obtained from the co-occurrence matrices (mean energy, mean entropy, mean cluster shade and mean cluster prominence). The remaining features

describe non-uniform distributions of nuclear size and shape (maximum of area, skewness of area, maximum concavity, skewness of concavity, skewness of roundness and kurtosis of roundness). These features were used as an input to the FNN classifiers. To determine the appropriate network architecture and number of epochs for each training algorithm we employed the same experimental setting as in the complete feature case. The results obtained are summarized in Table 5, while the corresponding ROC curves are illustrated in Fig. 4. As in the full feature

case, the best performing methods were BPVS and AOBP. A comparison of the results exhibited in Tables 4 and 5, suggests that for RPROP, iRPROP and SCG, the performance of the classifiers trained on the reduced feature set is worse than that of classifiers trained on the complete feature case with respect to all performance measures. The classifiers trained through BPVS and AOBP, on the other hand, achieved the same performance in the two cases with respect to the AUC measure. Regarding network architecture, the best results of the RPROP and the iRPROP methods were obtained by networks having the same number of hidden units as in the complete feature case. SCG was the only algorithm that required less hidden units. BPVS and AOBP yielded FNNs with more hidden units that also required more epochs to train.

To evaluate the results obtained through the UKW algorithm we also tested the FCM algorithm. The FCM algorithm requires the explicit determination of the number of clusters, c, present in the dataset. As this quantity is a priori unknown, we experimented with all integer values in the range [2,30]. Furthermore, since the result of the algorithm depends on its initialization, 10 executions were performed for each value of c. The fuzziness of the partition, r, had the fixed value 2, and the

algorithm was considered to converge when the change d in the distortion measure, was less than 0.001. As in the case of the UKW algorithm, the features that were selected at each execution of the algorithm were the ones closest to the cluster centers. The final subset of features for each value of c was determined by the features that were selected most frequently in the 10 experiments. For each subset of selected features 100 experiments were performed with the AOBP algorithm. We chose the AOBP algorithm, since it exhibited a predictable and robust performance. AOBP was allowed to perform 1000 training epochs. The FNN topology selected was $x-10-2$, where x represents the dimension of the reduced feature set. This architecture was selected as it yielded the best results for the UKW case.

The best result for the feature selection scheme using the FCM algorithm was obtained for a feature space of dimension 24. The corresponding total classification accuracy was 92.28%, and in terms of specificity and sensitivity 90.71% and 93.52%, respectively. It should be noted that these classification rates are higher than the full feature case. On the other hand, classification performance was very low when 2 or 25 features were selected. This finding, highlights the importance of the correct

Table 5 Best results for all the methods considered (11-dimensional feature vector)

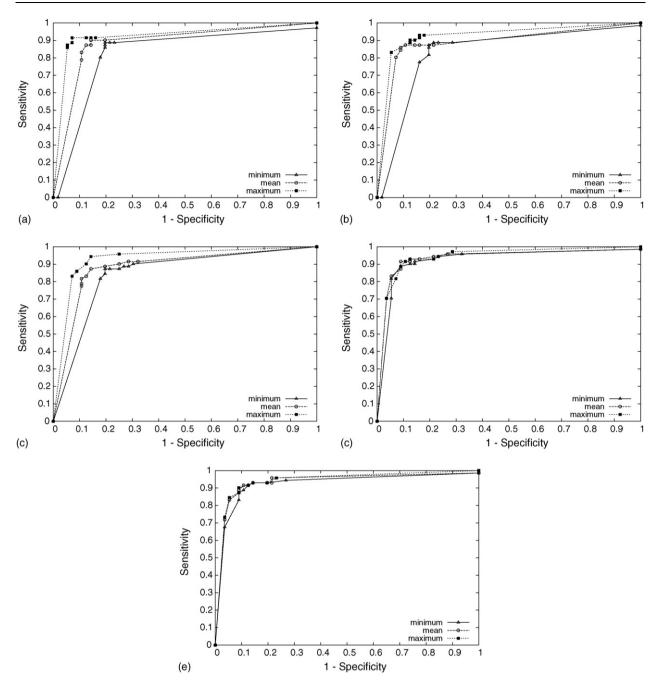|	Min	Mean	Max	S.D.
RPROP 11-3-2 for 300 epochs				
Specificity	78.57	85.44	92.86	2.82
Sensitivity	83.10	88.07	91.55	1.93
Overall accuracy	81.89	86.91	92.13	1.79
AUC	0.825	0.885	0.924	0.018
iRPROP 11-3-2 for 100 epochs				
Specificity	76.78	84.54	91.07	2.82
Sensitivity	83.10	88.68	92.96	1.99
Overall accuracy	83.46	86.85	91.34	1.68
AUC	0.832	0.888	0.925	0.017
SCG 11-3-2 for 200 epochs				
Specificity	71.43	82.75	91.07	3.53
Sensitivity	81.69	87.07	92.96	2.22
Overall accuracy	78.74	85.16	89.76	1.91
AUC	0.844	0.882	0.929	0.016
BPVS 11-10-2 for 800 epochs				
Specificity	85.71	88.71	91.07	1.01
Sensitivity	88.73	90.58	91.55	0.71
Overall accuracy	88.19	89.75	91.34	0.59
AUC	0.924	0.935	0.943	0.004
AOBP 11-10-2 for 1000 epochs				
Specificity	85.71	87.95	91.07	1.06
Sensitivity	90.14	91.41	92.96	0.55
Overall accuracy	88.98	89.88	91.34	0.56
AUC	0.924	0.936	0.942	0.004

Figure 4   ROC curves: (a) RPROP, (b) iRPROP, (c) SCG, (d) BPVS and (e) AOBP (reduced feature set)

choice of the number of clusters and the sensitivity of the overall outcome with respect to this parameter. In Fig. 5 the mean accuracy over the 100 different experiments for each value of the parameter c is exhibited. The vertical line plotted around each value depicts the minimum and the maximum values obtained.

Comparing the feature selection performed by the two clustering algorithms, the first point to note is the ability of the UKW algorithm to provide an estimation of the number of clusters. This fact, does not burden the user with the determination of this critical, for the resulting classification accuracy, parameter. Furthermore, the classification performance obtained by the UKW feature selection algorithm is very close to the best one obtained by the FCM algorithm, when the latter uses twice as many features. Moreover, the estimation of the number of clusters provided by UKW is at a point where the curve of overall accuracy with respect to different values of c employed by the FCM algorithm exhibits a plateau. This is a further indication that the
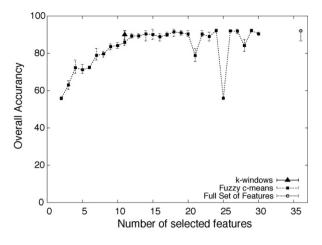
Figure 5 FNN classification performance

Table 6 Classification results employing alternative classification methods

Classifier	Specificity	Sensitivity	Overall accuracy
Bayesian	57.14	83.09	71.65
PNN	57.14	76.05	67.72
KNN	71.42	85.92	79.53

estimated cluster number provided by the UKW algorithm is close to the minimum number of features necessary to describe the phenomenon satisfactorily.

For a comparative evaluation, we also conducted experiments with other classification schemes. Table 6, reports the performance of the Bayesian classification scheme [58], probabilistic neural networks (PNNs) [59], and the *k*-nearest neighbor classifier (KNN) [58], on the same task. To highlight the superior performance of FNNs, for each classifier the feature vector employed was the one that yielded the best results among all possible combinations of features.

4. Discussion and concluding remarks

The prognostic characterization of TCC is a critical component of clinical management for the survival and cure of patients. Patients with non-invasive TCC of the urinary bladder are often observed without tumor progression, but many of them experience recurrence of disease. Currently, a number of conventional clinicopathological factors are used to predict the disease outcome. Histological grade and clinical stage are the most frequently employed, while other predictive indicators used

include tumor size, tumor location and patient age. Several researchers have attempted to model TCC bladder cancer recurrence. Naguib et al. [10] employed FNNs trained through the error backpropagation algorithm to predict bladder cancer recurrence from clinical and pathological information, exhibiting specificity and sensitivity of 76% and 55%, respectively. Patient follow-up ranged from 8 months to 24 years. Lynn et al. [9] also considered FNNs to predict the early recurrence of bladder tumor. The FNNs employed clinical and pathological variables to predict recurrence at first check cystoscopy (3 months) with sensitivity and specificity of 75% and 73%, and recurrence within the first year with sensitivity and specificity of 77% and 71%, respectively. Neural networks based on self organizing maps and the radial basis function algorithm have been used by Qurechi et al. [3] to predict recurrence within 6 months. The network exhibited a sensitivity of 70% and a specificity of 80% using clinicopathological and molecular markers. In a recent work [11], histopathological parameters evaluated by pathologists, in conjunction with nuclear features were used as an input to an artificial neural network to predict tumor recurrence. The system exhibited an accuracy of 74.5% and 71.1% for recurrence and non-recurrence cases, respectively, for a follow-up period of at least 5 years.

In this work, we analyzed the prognostic value of quantitative descriptors of nuclear morphometry by means of efficient classification engines based on FNNs. The choice of FNNs was guided by our previous experience [60], as well as, by their capability to address highly complex problems effectively and efficiently. Our approach managed to achieve high predictive rates for both recurrent and non-recurrent cases. Including all the 36 variables as input to the FNN the mean correct prediction rates reached 90.43% and 92.83% for the non-recurrent and recurrent cases, respectively. The mean overall accuracy was 91.77%, and the mean area under the ROC curve was 0.936.

Reducing the feature space dimensionality is of particular importance, since the removal of redundant information, renders the predictive model more robust, less complex and consequently more suitable for clinical use. Moreover, from the viewpoint of a medical expert it is important to have an insight to the information used by the prognostic system. Understanding the quantification of a certain number of descriptors might help pathologists to reconsider and improve their prognostic criteria. To identify a lower dimensional subspace that satisfactorily explains the original data, we used a feature selection method based on clustering. For this

task, the UKW and FCM clustering algorithms were examined. The UKW algorithm reduced drastically the feature space dimensionality by identifying a feature vector of eleven features. When this feature vector was used as input to the FNNs the mean overall accuracy was 89.88%. The accuracy for the non-recurrent and recurrent cases was 87.95% and 91.41%, respectively, and the mean area under the ROC curve was 0.936. The results obtained through alternative classification methods like the probabilistic neural networks, the Bayesian classifier and k-nearest neighbor classifiers, indicate that FNNs are more suitable for identifying the complex non-linear relationships between dependent and independent variables. Considering the training methods used, the AOBP method followed by BPVS proved to be the most efficient. To the best of our knowledge, the classification accuracies reported in Tables 4 and 5 are the highest reported in the literature.

In conclusion, the advantages of the proposed prognostic model for superficial TCC urinary bladder cancer recurrence are the following: (i) it is highly accurate; (ii) it uses objective, quantitative information of the cell nuclei; (iii) it does not require specialized staining protocols, or the use of costly biomolecular markers; and (iv) it is capable of automatically selecting the most informative predictive variables.

Acknowledgments

References

[1] Kelloniemi E, Rintala E, Finne UHTFG, Stenman P. Tumor-associated trypsin inhibitor as a prognostic factor during follow-up of bladder cancer. Adult Urol 2003;62(2):249—53.

[2] Witjes J, Kiemeney L, Wheeless L, Grossman H. The value of histopathological prognostic factors in superficial bladder cancer: Do we need more. Urol Oncol 2000;5(5):185—90.

[3] Qureshi K, Naguib R, Hamdy F, Neal D, Mellon K. Neural network analysis of clinicopathological and molecular markers in bladder cancer. J Urol 2000;163:630—3.

[4] Badalament R, Kimmel M, Gay H, Cibas E, Whitmore Jr W, Herr H, et al. The sensitivity of flow cytometry compared with conventional cytology in the detection of superficial bladder carcinoma. Cancer 1987;59(12):2078—85.

[5] Devonec M, Darzynkiewicz Z, Claps MK, Collste L, Whitmore Jr W, Melamed M. Flow cytometry of low stage bladder tumors: correlation with cytologic and cystoscopic diagnosis. Cancer 1982;49(1):109—18.

[6] Gschwendtner A, Hoffmann-Weltin Y, Mikuz G, Mairinger T. Quantitative assessment of bladder cancer by nuclear tex-ture analysis using automated high resolution image cytometry. Mod Pathol 1999;12(8):806—13.

[7] Gschwendtner A, Mairinger T. Quantitative assessment of bladder carcinoma by acid labile dna assay. Cancer 1999;86(1):105—13.

[8] Oosterlink W, Lobel B, Jakse G, Malmstrom P-U, Stokle M, Sternberg C. Guidelines on bladder cancer. Eur Urol 2002;41:105—12.

[9] Lynn N, Shapiro J, Lynn M, Waymont B, Inglis J, Philp N. Use of artificial neural network to predict the early recurrence of superficial bladder cascer. Eur Urol Suppl 2003;2(1):193.

[10] Naguib R, Qureshi F, Hamdy K, Neal D. Neural network analysis of prognostic markers in bladder cancer. p. 1007—9In: Proceedings of the 19th international conference-IEEE/EMBS; 1997.

[11] Spyridonos P, Cavouras D, Ravazoula P, Nikiforidis G. A computer-based diagnostic and prognostic system for assessing urinary bladder tumour grade and predicting cancer recurrence. Med Inform Internet Med 2002;27(2):111—22.

[12] Miller J, Geisler J, Manahan K, Geisler H, Miller G, Zhou Z, et al. Nuclear size, shape and density in endometrial carcinoma: relationship to survival at over 5 years of follow-up. Does analyzing only cells occupying the $g_0 - g_1$ peak and useful information? Int J Gynecol Cancer 2004;14(1):138—44.

[13] Weyn B, Tjalma W, Wouver GVD, Daele AV, Scheunders P, Jacob W, et al. Validation of nuclear texture, density, morphometry and tissue syntactic structure analysis as prognosticators of cervical carcinoma. Anal Quant Cytol Histol 2000;22(5):373—82.

[14] van de Wouwer G, Weyn B, Scheunders P, Jacob W, van Marck E, van Dyck D. Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei. J Microsc 2000;197(1):25—35.

[15] van Velthoven R, Petein P, Oosterlinck W, Wilde T, Mattelaer J, Hardeman M, et al. Identification by quantitative chromatin pattern analysis of patients at risk for recurrence of superficial transitional bladder carcinoma. J Urol 2000;164:2134—7.

[16] Mobley B, Schechter E, Moore W, McKee P, Eichner J. Neural network predictions of significant coronary artery stenosis in men. Artif Intell Med 2005;34(2):151—61.

[17] Papik K, Molnar B, Schaefer R, Dombovari Z, Tulassay Z, Feher J. Application of neural networks in medicine—a review. Med Sci Monitor 1998;4(3):538—46.

[18] Street W. A neural network model for prognostic prediction. In: Shavlik J, editor. Machine learning: proceedings of the 15th international conference. San Francisco, CA: Morgan Kaufmann; 1998. p. 540—46.

[19] Boon M, Kok L, Nygaard-Nielsen M, Holm K, Holund B. Neural network processing of cervical smears can lead to a decrease in diagnostic variability and an increase in screening efficacy: a study of 63 false-negative smears. Mod Pathol 1994;7(9):957—61.

[20] Kolles H, Von-Wangenheim A, Vince GH, Niedermayer I, Feiden W. Automated grading of astrocytomas based on histomorphometric analysis of ki-67 and feulgen stained paraffin sections. classsification results of neuronal and discriminant analysis. Anal Cell Pathol 1995;8(2):101—16.

[21] Downs J, Harrison R, Cross S. Evaluating a neural network decision-support tool for the diagnosis of breast cancer. In: Barahona P, Stefanelli M, Wyatt J, editors. Lecture notes in artificial intelligence, vol. 943. Springer-Verlag; 1995.

[22] Sharkey A, Sharkey N, Cross S. Adapting an ensemble approach for the diagnosis of breast cancer. In: Niklasson L, Bodn M, Ziemke T, editors. Eighth international conference on artificial neural networks. Springer-Verlag; 1998.

[23] Wolberg W, Street W, Mangasarian O. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Anal Quant Cytol Histol 1995;17(2):77—87.

[24] Downs J, Harrison R, Kennedy R, Cross S. Application of the fuzzy artmap neural network model to medical pattern classification tasks. Artif Intell Med 1996;8(4):403—28.

[25] Bezdek JC. Pattern recognition with fuzzy objective function algorithms Kluwer Academic Publishers; 1981.

[26] Tasoulis DK, Vrahatis MN. Unsupervised distributed clustering. In: The IASTED international conference on parallel and distributed computing and networks; 2004.

[27] Vrahatis MN, Boutsinas B, Alevizos P, Pavlides G. The new k-windows algorithm for improving the k-means clustering algorithm. J Complexity 2002;18:375—91.

[28] Spyridonos P, Ravazoula P, Cavouras D, Berberidis K, Nikiforidis G. Computer-based grading of heamatoxylin-eosin stained tissue sections of urinary bladder carcinomas. Med Inform Internet Med 2001;26(3):179—90.

[29] Harralick R, Shanmugam K. Textural features for image classification. IEEE Trans Syst Man Cybernet 1973;3(6):610—21.

[30] der Poel HV, Schaafsma HE, Vooijs PG, Debruyne FM, Schalken JA. Review article. Quantitative light microscopy in urological oncology. J Urol 1992;148:1—13.

[31] van Velthoven R, Petein M, Zlotta A, Oosterlinck W, Meijden A, Zandona C, et al. Computer-assisted chromatin texture characterization of feulgen-stained nuclei in a series of 331 transitional bladder cell carcinomas. J Pathol 1994;173:235—42.

[32] Walker RF, Jackway PT, Lovell B. Cervical cell classification via co-occurrence and markov random field features. In: Proceedings of digital image computing: techniques and applications; 1995.

[33] Ohanian P, Dubes R. Performance evaluation for four classes of textural features. Pattern Recogn 1992;25(8):819—33.

[34] Choi H-K, Vasko J, Bengtsson E, Jarkrans T, Malmstrom U, Wester K, et al. Grading of transitional cell bladder carcinoma by texture analysis of histological sections. Anal Cell Pathol 1994;6:327—43.

[35] Walker R, Jackway P, Longstaff I. Improving co-occurrence matrix feature discrimination. In: Maeder A, Lovell B, editors. Digital image computing: techniques and applications. Brisbane, Australia; 1995, p. 643—8.

[36] Chen P, Pavlidis T. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. Comp Graphics Image Process 1979;10:172—82.

[37] Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proceedings of the IEEE international conference on neural networks; 1993.

[38] Igel C, Hüsken M. Improving the Rprop learning algorithm. In: Bothe H, Rojas R, editors. Proceedings of the second international ICSC symposium on neural computation (NC 2000). ICSC Academic Press; 2000.

[39] Moller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993;6:525—33.

[40] Magoulas GD, Vrahatis MN, Androulakis GS. Effective backpropagation training with variable stepsize. Neural Netw 1997;10(1):69—82.

[41] Magoulas GD, Plagianakos VP, Vrahatis MN. Adaptive stepsize algorithms for on-line training of neural networks. Nonlinear Anal TMA 2001;47(5):3425—30.

[42] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. Parallel distributed processing: explorations in the microstructure of cognition: foundations. MIT Press; 1986. p. 318—62.

[43] Hornik K. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2:359—66.

[44] White H. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. Neural Netw 1990;3:535—49.

[45] Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell 1991;252—64.

[46] John GH, Kohavi R, Pfleger K, Irrelevant features and the subset selection problem. In: International conference on machine learning; 1994, p. 121—29, journal version in AIJ, available at http://citeseer.nj.nec.com/13663.html.

[47] Dasgupta S. Experiments with random projection. In: 16th conference on uncertainty in artificial intelligence (UAI 2000); 2000.

[48] Drineas P, Frieze A, Kannan R, Vempala S, Vinay V. Clustering in large graphs and matrices. In: SODA: ACM-SIAM symposium on discrete algorithms (a conference on theoretical and experimental analysis of discrete algorithms); 1999.

[49] Aggarwal CC, Procopiuc C, Wolf JL, Yu PS, Park JS. Fast algorithms for projected clustering. In: 1999 ACM SIGMOD international conference on management of data. NY, USA: ACM Press; 1999.

[50] Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. In: SIGMOD '98: proceedings of the 1998 ACM SIGMOD international conference on management of data. NY, USA: ACM Press; 1998. p. 94—105.

[51] Hartigan JA, Wong MA. A k-means clustering algorithm. Appl Stat 1979;28:100—8.

[52] Alevizos P, Boutsinas B, Tasoulis DK, Vrahatis MN. Improving the orthogonal range search k-windows clustering algorithm. In: Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence. Washington, DC; 2002. p. 239-45.

[53] Alevizos P, Tasoulis DK, Vrahatis MN. Parallelizing the unsupervised k-windows clustering algorithm. In: Wyrzykowski R, editor. Lecture notes in computer science, vol. 3019. Springer-Verlag; 2004.

[54] Tasoulis DK, Alevizos P, Boutsinas B, Vrahatis MN. Parallel unsupervised k-windows: an efficient parallel clustering algorithm. In: Malyshkin V, editor. Lecture notes in computer science, LNCS 2763. Berlin, Heidelberg: Springer-Verlag; 2003

[55] Tasoulis DK, Vrahatis MN. Novel approaches to unsupervised clustering through the k-windows algorithm. In: Sirmakessis S, editor. Knowledge mining: series studies in fuzziness and soft computing, vol. 185. Berlin, Heidelberg: Springer-Verlag; 2005.

[56] Tasoulis DK, Vrahatis MN. Unsupervised clustering on dynamic databases. Pattern Recogn Lett 2005;26(13):2116—27.

[57] Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29—36.

[58] Friedman N, Kohavi R. Data mining tasks and methods: classification: Bayesian classification. In: Handbook of data mining and knowledge discovery. Oxford University Press, Inc.; 2002. p. 282—8.

[59] Specht DF. Probabilistic neural networks. Neural Netw 1990;3(1):109—18.

[60] Tasoulis DK, Spyridonos P, Pavlidis NG, Cavouras D, Ravazoula P, Nikiforidis G, Vrahatis MN. A neural network approach to the comparative evaluation of physician's subjective assessment versus quantitative nuclear features in grading urine bladder tumors. In: Howlett VRJ, Jain LC, editors. Lecture notes in computer science, (LNAI), vol. 2774. Berlin, Heidelberg: Springer-Verlag; 2003.