

Urinary Bladder Tumor Grade Diagnosis Using On-line Trained Neural Networks

D.K. Tasoulis^{1,2}, P. Spyridonos³, N.G. Pavlidis^{1,2}, D. Cavouras⁴,
P. Ravazoula⁵, G. Nikiforidis³, and M.N. Vrahatis^{1,2*}

¹ Department of Mathematics, University of Patras
GR-26110 Patras, Greece
vrahatis@math.upatras.gr

² University of Patras Artificial Intelligence Research Center (UPAIRC)

³ Computer Laboratory, School of Medicine University of Patras
GR-26110 Patras, Greece

⁴ Department of Medical Instrumentation Technology, TEI of Athens
Ag. Spyridonos Street Aigaleo, GR-12210 Athens, Greece

⁵ Department of Pathology, University Hospital
GR-26110 Patras, Greece

Abstract. This paper extends the line of research that considers the application of Artificial Neural Networks (ANNs) as an automated system, for the assignment of tumors grade. One hundred twenty nine cases were classified according to the WHO grading system by experienced pathologists in three classes: Grade I, Grade II and Grade III. 36 morphological and textural, cell nuclei features represented each case. These features were used as an input to the ANN classifier, which was trained using a novel stochastic training algorithm, namely, the Adaptive Stochastic On-Line method. The resulting automated classification system achieved classification accuracy of 90%, 94.9% and 97.3% for tumors of Grade I, II and III respectively.

1 Introduction

Bladder cancer is the fifth most common type of cancer. Superficial Transitional cell carcinoma (TCC) is the most frequent histological type of bladder cancer [13]. Currently, these tumors are assessed using a grading system based on a variety of histopathological characteristics. Tumor grade, which is determined by the pathologist from tissue biopsy, is associated with tumor aggressiveness. The most widely accepted grading system is the WHO (World Health Organization) system, which stratifies TCCs into three categories: tumors of Grade I, Grade II and Grade III. Grade I tumors are not associated with invasion or metastasis but present a risk for the development of recurrent lesions. Grade II carcinomas are associated with low risk of further progression, yet they frequent recur. Grade III tumors are characterized by a much higher risk of progression

* *Corresponding author.*

and also high risk of association with disease invasion [7]. Although histological grade has been recognized as one of the most powerful predictors of the biological behavior of tumors and significantly affects patients' management, it suffers from low inter and intra observer reproducibility due to the subjectivity inherent to visual observation [12]. Digital image analysis techniques and classification systems constitute alternative means to perform tumor grading in a less subjective manner. Numerous research groups have proposed quantitative assessments to address this problem. H-K Choi et al [3] have developed an automatic grading system using texture features on a large region of interest, covering a typical area in the histological section. The textural based system produced an overall accuracy of 84.3% in assessing tumors grade. In a different study [6], researchers have employed tissue architectural features and classified tumors with an accuracy of 73%. More recent studies have focused on the analysis of cell nuclei characteristics to perform tumor grade classification with success rates that do not significantly exceed 80% [2, 17]. In this study, we present a methodology which improves considerably the level of diagnostic accuracy in assigning tumor grade. The method is based on the application of an ANN as a classifier system. The input data for the ANN, describe a number of nuclear morphological and textural features, that were obtained through an automatic image processing analysis technique. It is worth noting that the prognostic and diagnostic value of these features, has been confirmed [4].

2 Materials and Methods

129 tissue sections (slides) from 129 patients (cases) with superficial TCC were retrieved from the archives of the Department of Pathology of Patras University Hospital in Greece. Tissue sections were routinely stained with Haematoxylin-Eosin. All cases were reviewed independently by the experts to safeguard reproducibility. Cases were classified following the WHO grading system as follows: thirty-three cases as Grade I, fifty-nine as Grade II and thirty-seven as Grade III. Images from tissue specimens were captured using a light microscopy imaging system. The method of digitalization and cell nuclei segmentation for analysis has been described in previous work [17]. Finally, from each case 36 features were estimated: 18 features were used to describe information concerning nuclear size and shape distribution. The rest were textural features that encoded chromatin distribution of the cell nucleus [17]. These features were used as an input to ANN classifier.

3 Artificial Neural Networks

Back Propagation Neural Networks (BPNNs) are the most popular artificial neural network models. The efficient supervised training of BPNNs 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 learning error.

Applications of supervised learning can be divided into two categories: stochastic (also called on line) and batch (also called off line) learning. Batch training can be viewed as the minimization of the error function E . This minimization corresponds to updating the weights by epoch and to be successful it requires a sequence of weight iterates $\{w_k\}_{k=0}^{\infty}$ where k indicates epochs, which converges to a minimizer w^* .

In on line training, network weights are updated after the presentation of each training pattern. This corresponds to the minimization of the instantaneous error of the network $E(p)$ for each pattern p individually. On line training may be chosen for a learning task either because of the very large (or even redundant) training set or because we want to model a gradually time varying system. Moreover, it helps escaping local minima. Given the inherent efficiency of stochastic gradient descent, various schemes have been proposed recently [1, 18, 19]. Unfortunately, on line training suffers from several drawbacks such as sensitivity to learning parameters [16]. Another disadvantage is that most advanced optimization methods, such as conjugate gradient, variable metric, simulated annealing etc., rely on a fixed error surface, and thus it is difficult to apply them for on line training [16].

Regarding the topology of the network it has been proven [5, 20] that standard feedforward networks 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. This implies that any lack of success in applications must arise from inadequate learning, insufficient number of hidden units or the lack of a deterministic relationship between inputs and targets. Keeping these theoretical results we restrict the network topology of ANNs used in this study to one hidden layer.

4 Training Method

For the purpose of training neural networks an on-line stochastic method was employed. For recent proposed on-line training methods as well as application in medical applications see [8, 9, 10, 11, 14] This method uses a learning rate adaptation scheme that exploits gradient-related information from the previous patterns. This algorithm is described in [9], and is based on the stochastic gradient descent proposed in [1]. The basic algorithmic scheme is exhibited in Table 1. As pointed out in Step 4, the algorithm adapts the learning rate using the dot product of the gradient from the previous two patterns. This algorithm produced both fast and stable learning in all the experiments we performed, and very good generalization results.

5 Results and Discussion

To measure the ANN efficiency the dataset was randomly permuted five times. Each time it was split into a train set and a test set. The training set contained about 2/3 of the original dataset from each class. For each permutation the

Table 1. Stochastic On-Line Training with adaptive stepsize

- 0: Initialize weights w^0, η^0 , and meta-stepsize K .
- 1: Repeat for each input pattern p
- 2: Calculate $E(w^p)$ and then $\partial E(w^p)$
- 3: Update the weights:

$$w^{p+1} = w^p - \eta^p \partial E(w^p)$$
- 4: Calculate the stepsize to be used with the next pattern $p + 1$:

$$\eta^{p+1} = \eta^p + K \langle \partial E(w^{p-1}), \partial E(w^p) \rangle$$
- 5: Until the termination condition is met.
- 6: Return the final weights w^* .

Table 2. Accuracy of Grade I,II and III for various topologies

Topology	Grade I	Grade II	Grade III
36-1-3	18.18%	100%	100%
36-2-3	81.81%	100%	100%
36-5-3	90.90%	100%	100%
36-16-3	81.81%	100%	100%

network was trained with the Stochastic On Line method with adaptive step size discussed previously. Two terminating conditions were used: the maximum number of cycles over the entire training set was set to 100, and the correct classification of all the training patterns. Alternatively, the Leave-One-Out (LOO) method [15] was employed to validate ANN classification accuracy. According to this method, the ANN is initialized with the training set including all patterns except one. The excluded pattern is used to assess the classification ability of the network. This process is repeated for all the patterns available and results are recorded in the form of a truth table. The software used for this task was developed under the Linux Operating System using the C++ programming language, and the gcc ver2.96 compiler.

A great number of different ANN topologies (number of nodes in the hidden layer) were tested for the grade classification task. Some of these tests are exhibited in Table 2. Best results were obtained using the topology: 36-5-3.

Table 3 illustrates analytically the ANN performance for each Crossover permutation. The ANN exhibited high classification accuracy for each grade category. It is worth noting that Grade I tumors were differentiated successfully from Grade III tumors. In four out of five Crossovers neither Grade I to III nor Grade III to I errors occurred. As can be seen from Table 3 in one permutation only 1 case of Grade I was misclassified as Grade III. From a clinical point of view, it is important to distinguish low grade tumors, which can generally be treated

Table 3. Crossover Results For the ANNs

Crossover I				
Histological finding	ANN classification			Accuracy(%)
	Grade I	Grade II	Grade III	
GRADE I	10	1	0	90.9
GRADE II	0	19	1	95
GRADE III	0	1	12	92.3
Overall Accuracy				93.2
Crossover II				
Histological finding	ANN classification			Accuracy(%)
	Grade I	Grade II	Grade III	
GRADE I	10	1	0	90.9
GRADE II	0	20	0	100
GRADE III	0	2	11	84.62
Overall Accuracy				93.2
Crossover III				
Histological finding	ANN classification			Accuracy(%)
	Grade I	Grade II	Grade III	
GRADE I	10	1	0	90.9
GRADE II	0	19	1	95
GRADE III	0	1	12	92.3
Overall Accuracy				93.2
Crossover IV				
Histological finding	ANN classification			Accuracy(%)
	Grade I	Grade II	Grade III	
GRADE I	10	0	1	90.9
GRADE II	0	19	1	95
GRADE III	0	0	13	100
Overall Accuracy				95.4
Crossover V				
Histological finding	ANN classification			Accuracy(%)
	Grade I	Grade II	Grade III	
GRADE I	10	1	0	90.9
GRADE II	0	18	2	90
GRADE III	0	0	13	100
Overall Accuracy				93.2

conservatively in contrast to high-grade tumors. The latter often require a more aggressive therapy because of a high-risk cancer progression. Results could also be interpreted in terms of specificity and sensitivity. That is specificity is the percentage of Grade I tumors correctly classified and sensitivity is the percentage of Grade III tumors correctly classified. ANN grade classification safeguarded high sensitivity which is of vital importance for patients treatment course, retaining

Table 4. Leave One out Results For the ANNs

Histological finding	ANN classification			
	Grade I	Grade II	Grade III	Accuracy(%)
GRADE I	30	1	2	90
GRADE II	0	56	3	94.9
GRADE III	0	1	36	97.3
Overall Accuracy				94.06

at the same time high specificity. Another important outcome is that the intermediate Grade II tumors were recognized with high confidence from Grade I and III. This would be particular helpful for pathologist who encounter difficulties in assessing Grade II tumors since some of them fall into the gray zone bordering on either Grade I or Grade III, and the decision is subject to the judgment of the pathologist.

The simplicity and efficiency of the training method enabled us to verify the ANN classification accuracy by employing the LOO method (the whole procedure required 46 seconds to complete in Athlon CPU running at 1500 MHz). It is well known that this method is optimal to test the performance of a classifier when small data sets are available, but this testing procedure is computationally expensive when used with conventional training methods. Classification results employing the LOO method are shown in Table 4. The consistency of the system in terms of high sensitivity (no Grade III to Grade I error occurred) was verified.

In [3], a textural based system produced an overall accuracy of 84.3% in assessing tumors grade. In a different study [6], researchers have employed tissue architectural features and classified tumors with an accuracy of 73%. More recent studies have focused on the analysis of cell nuclei characteristics to perform tumor grade classification with success rates that do not significantly exceed 80% [2, 17]. The ANN methodology proposed in this paper, improved significantly the tumor grade assessment with success rates of 90%, 94.9%, and 97.3%, for Grade I, II and III respectively.

6 Conclusions

In this study an ANN was designed to improve the automatic characterization of TCCs employing nuclear features. The ANN exhibited high performance in correctly classifying tumors into three categories utilizing all the available diagnostic information carried by nuclear size, shape and texture descriptors. The proposed ANN could be considered as an efficient and robust classification engine able to generalize in making decisions about complex input data improving significantly the diagnostic accuracy. The present study extends previous work in terms of the features used and enforces the belief that objective measurements on nuclear morphometry and texture offer a potential solution for the accurate

characterization of tumor aggressiveness. The novelty of this paper resides in the results obtained since they are the highest reported in the literature. Since most Grade I tumors are considered to be good prognosis, while Grade III is associated with bad prognosis, the 0% misclassification of Grade III tumors as Grade I, gives an advantage to the proposed methodology to be part of a fully automated computer aided diagnosis system.

References

- [1] L. B. Almeida, T. Langlois, L. D. Amaral, and A. Plankhov. Parameter adaption in stochastic optimization. *On-Line Learning in Neural Networks*, pages 111–134, 1998. 201
- [2] N. Belacel and M. R. Boulassel. Multicriteria fuzzy assignment method: a useful tool to assist medical diagnosis. *Artificial Intelligence in Medicine*, 21:201–207, 2001. 200, 204
- [3] H.-K. Choi, J. Vasko, E. Bengtsson, T. Jarkrans, U. Malmstrom, K. Wester, and C. Busch. Grading of transitional cell bladder carcinoma by texture analysis of histological sections. *Analytical Cellular Pathology*, 6:327–343, 1994. 200, 204
- [4] C. De Prez, Y. De Launoit, R. Kiss, M. Petein, J.-L. Pasteels, and A. Verhest. Computerized morphonuclear cell image analysis of malignant disease in bladder tissues. *Journal of Urology*, 143:694–699, 1990. 200
- [5] K. Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989. 201
- [6] T. Jarkrans, J. Vasko, E. Bengtsson, H.-K. Choi, U. Malmstrom, K. Wester, and C. Busch. Grading of transitional cell bladder carcinoma by image analysis of histological sections. *Analytical Cellular Pathology*, 18:135–158, 1995. 200, 204
- [7] I. E. Jonathan, B. A. Mahul, R.R Victor, and F.K Mostofi. (transitional cell) neoplasms of the urinary bladder. *The American Journal of Surgical Pathology*, 22(12):1435–1448, 1998. 200
- [8] G. D. Magoulas, V. P. Plagianakos, and M. N. Vrahatis. Global learning rate adaptation in on-line neural network training. In *Proceedings of the Second International Symposium in Neural Computation May 23–26, Berlin, Germany*, 2000. 201
- [9] G. D. Magoulas, V. P. Plagianakos, and M. N. Vrahatis. Adaptive stepsize algorithms for on-line training of neural networks. *Nonlinear Analysis, T. M. A.*, 47(5):3425–3430, 2001. 201
- [10] G. D. Magoulas, V. P. Plagianakos, and M. N. Vrahatis. Hybrid methods using evolutionary algorithms for on-line training. In *INNS-IEEE International Joint Conference on Neural Networks (IJCNN), July 14–19, Washington, D. C., U. S. A.*, volume 3, pages 2218–2223, 2001. 201
- [11] G. D. Magoulas, V. P. Plagianakos, and M. N. Vrahatis. Improved neural network-based interpretation of colonoscopy images through on-line learning and evolution. In D. Dounias and D. A. Linkens, editors, *European Network of Excellence on Intelligent Technologies for Smart Adaptive Systems*, pages 38–43. 2001. 201
- [12] E. Ooms, W. Anderson, C. Alons, M. Boon, and R. Veldhuizen. Analysis of the performance of pathologists in grading of bladder tumors. *Human Pathology*, 14:140–143, 1983. 200
- [13] S. L. Parker, T. Tony, S. Bolden, and P. A. Wingo. Cancer statistics. *Cancer Statistics 1997. CA Cancer J Clin*, 47(5):5–27, 1997. 199

- [14] V.P. Plagianakos, G.D. Magoulas, and M.N. Vrahatis. Tumor detection in colonoscopic images using hybrid methods for on-line neural network training. In G.M. Papadourakis, editor, *Neural Networks and Expert Systems in Medicine and Healthcare (NNEsMED)*, pages 59–64. Technological Educational Institute of Crete, Heraklion, 2001. 201
- [15] S. J. Raudys and A. K. Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 252–264, 1991. 202
- [16] N.N. Schraudolf. Gain adaptation in stochastic gradient descend. 1999. 201
- [17] P. Spyridonos, D. Cavouras, P. Ravazoula, and G. Nikiforidis. Neural network based segmentation and classification system for the automatic grading of histological sections of urinary bladder carcinoma. *Analytical and Quantitative Cytology and Histology*, 24:317–324, 2002. 200, 204
- [18] R.S. Sutton. Adapting bias by gradient descent: an incremental version of delta-bar-delta. In *Proc. 10th National Conference on Artificial Intelligence*, pages 171–176. MIT Press, 1992. 201
- [19] R.S. Sutton. Online learning with random representations. In *Proc. 10th International Conference on Machine Learning*, pages 314–321. Morgan Kaufmann, 1993. 201
- [20] H. White. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3:535–549, 1990. 201