1st International Conference "From Scientific Computing to Computational Engineering"
1st IC-SCCE
Athens, 8-10 September, 2004

COMPARATIVE EVALUATION OF FEEDFORWARD AND PROBABILISTIC NEURAL NETWORKS FOR THE AUTOMATIC CLASSIFICATION OF BRAIN TUMOURS

Aristotelis Kostopoulos[†], Dimitris Glotsos[††], Panagiota Spyridonos[††], George Nikiforidis[††], Dimitris Sotiropoulos[†], Theodoula Grapsa[†]

[†] Division of Computational Mathematics and Informatics,
Department of Mathematics, University of Patras
26500 Patras, Greece
e-mail: arkostop@math.upatras.gr

[††]Medical image processing and analysis laboratory
Department of Medical Physics, University of Patras
26500 Patras, Greece

Keywords: feedforward NN, probabilistic NN, nonmonotone spectral conjugate gradient, brain tumors, classification.

Abstract. *Brain tumours grading is a crucial step for determining treatment planning and patient management. The grade of a tumour is defined by pathologists after reviewing biopsies under the microscope, a procedure that has been proven highly subjective. In this work, we propose a computer-based system for the automatic classification of astrocytomas that can be used as a second opinion tool for the clinicians contributing to the objectification of the diagnostic process. The system process routine brain tumours biopsies and performs automatic diagnosis of the degree of tumour abnormality (low from high grade tumours) based solely on quantitative information acquired from cell nuclei. We designed the system incorporating two stat-of-art neural network algorithms, namely Perry's nonmonotone spectral conjugate gradient training algorithm for Multi Layer Perceptrons (MLPs) and Probabilistic NN (PNN).* Best performance was obtained using a MLP-NN classifier with 7-hidden neurons topology that discriminating low from high grade tumours with an accuracy of 92.0%. Sensitivity and specificity ranged to 93.1% and 90.5% respectively. The PNN classifier resulted in lower rates (83.3% specificity, 91.5% sensitivity and 89.9% overall accuracy). T*he proposed method is a dynamic new alternative to brain tumour grading since it combines relatively high accuracy rates with daily clinical standards.*

1 INTRODUCTION

Brain tumours grading is a crucial step for determining treatment planning and patient management [1]. Brain astrocytomas are considered as one of the most lethal and difficult to treat forms of cancers [2]. To decide on the degree of tumour abnormality, pathologists review biopsies under the microscope. According to the WHO (World Health Organisation) system, clinicians classify astrocytomas into two degrees of malignancy: low and high grade neoplasms. However, the reviewing procedure has been proven to be a highly subjective task depending both on the experience and skills of the expert [3]. Diagnostic errors are considered as common in daily clinical practice. However, these errors may provoke false treatment planning with adverse effects in patient survival and need to be prevented [4]. In order to facilitate pathologists to more objective decisions, computer-based systems were introduced [5-11].

Automatic tumour grading has been extensively examined [5-10] and until today remains an active research area [11]. Classification systems have been proposed based on algorithms such as the decision trees [5], the nearest-neighbour concept [6], the backpropagation neural networks (NN) [7], discriminant analysis [8], fuzzy logic [9] and support vector machines [10]. However, most of these studies [5-9,11] have been designed based on modifications of the WHO grading system and specialized staining procedures. The staining process aims in highlighting nuclei from surrounding tissue in biopsies. The most accurate the staining process, the easier the identification of malignant properties imprinted in nuclei. However, the staining procedure adopted in daily clinical practice is not as accurate as the specialized procedures utilized by previous studies [5-9,11]. Furthermore,

computer-aided systems are difficult to be incorporated in clinical routine if not using the routine diagnostic standards, namely the WHO classification scheme and the Hematoxylin-Eosin (H&E) staining.

In this work, we propose a computer-based system for the classification of astrocytomas as low or high grade. The system process images from routine H&E stained brain tumours biopsies and perform automatic diagnosis of the degree of tumour abnormality based solely on quantitative information acquired from cell nuclei. We designed the system incorporating two stat-of-art neural network algorithms, namely Perry's nonmonotone spectral conjugate gradient training algorithm for Multi Layer Perceptrons (MLPs) and Probabilistic NN (PNN). We demonstrate that the proposed method is a dynamic new alternative to brain tumour grading since it combines relatively high accuracy rates with daily clinical standards.

2 MATERIALS AND METHODS

One hundred and forty biopsies were prepared with the H&E staining protocol. The material was collected from the Department of Pathology, University Hospital of Patras and Department of Pathology, General Distinct Anticancer Hospital of Piraeus, METAXA. Grading was decided by two pathologists, who classified tumours as low or high-grade according to definitions of the WHO scheme. The H&E staining procedure and the WHO classification are the standards mostly adopted in daily clinical practice [3]. 61 cases were characterized as low-grade and 79 as high-grade. From predefined regions by the experts, images were captured using a light microscopy imaging system described elsewhere [10].

A segmentation algorithm was applied to separate nuclei from surrounding tissue in order to encode tumour malignancy in a set of 40 quantitative nuclear features. Nuclear features have been proven to be powerful descriptors of the degree of tumour abnormality [6].

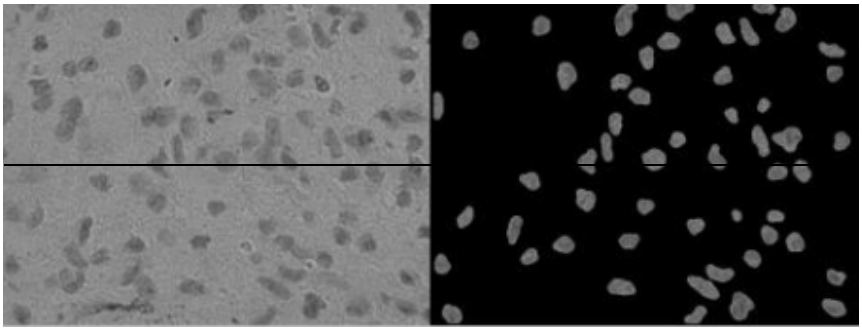

Figure 1. Examples of low (bottom) and high grade (top) brain astrocytomas along with the resulted segmented images

Features extracted, can be categorized into a set of morphological features describing size and shape of nuclei and a set of textural features encoding the chromatin distribution and organization within the nucleus. Morphological features comprised measurements of area, roundness and concavity. For each one of these features the mean value, standard deviation, range, skewness, kurtosis, and maximum value was calculated. Textural features were calculated from the DNA histogram, the co-occurrence and run length matrixes. A more detailed description for calculating these features can be found in [12-13].

In our case, the dimension of the feature vectors is large, namely 40 features. It is theorized that the components of the vectors are highly correlated, since some features such as the nucleus area and perimeter, are strongly related. Thus, it is useful to reduce the dimension of the feature vectors, because this will also reduce computational time. An effective procedure for performing this operation is the Karhunen – Loeve transformation or Principal Component Analysis (PCA). This technique has three majors effects: it orthogonalize's the components of the feature vectors so that they are uncorrelated with each other; it orders the resulting orthogonal components, namely principal components, so that those with the largest variation come first; and it eliminates those components that contribute the least to the variation in the data set [14].

In practice, the algorithm proceeds by first computing the mean of the feature vectors and then subtracting off this mean. Then the covariance matrix is calculated and its eigenvectors and eigenvalues are found. The eigenvectors corresponding to the M largest eigenvalues are retained and the feature vectors are projected onto the eigevectors to give the components of the transformed vectors in the M – dimensional space. In this contribution, we performed PCA, and we reduced the dimensionality of the feature space from 40 to 11. Each of this eleven new features (Scores) was computed as the sum of the coefficients (Principal Components) derived for every original feature (table 1). Subsequently, the input vectors are transformed from the original 40 features to the derived PCA eleven scores. These vectors are now uncorrelated. We used the Matlab version 6.5 routine in order to perform PCA [14-15].

Ninety-one cases (40 low-grade and 50 high-grade) were used to construct each NN based classifier. Forty-nine cases (21 low-grade and 29 high-grade) were used to validate the system's generalization to new data. The system was designed either by employing a multi layer perceptron, trained with Perry's nonmonotone spectral conjugate gradient training algorithm, or a probabilistic neural network. The performance of these two classifiers in the task of discriminating low from high grade tumours is illustrated in table 2.

2.1 Overview of the Multi Layer Perceptron

Multi layer perceptrons (MLPs) are feedforward neural networks with one or more layers of neurons, called hidden layers, between the output layer and the network's input. The output of each neuron is the weighted sum of its inputs passed through an activation function. The neuron is characterized by an internal threshold weight and by the type of the activation function. The MLP can be described by the equations

$$\begin{aligned} net_j^l &= \sum_{i=1}^{N_{l-1}} w_{ij}^{l-1,l} o_i^{l-1} \\ o_j^l &= f\left(net_j^l\right) \end{aligned} \tag{1}$$

where net_j^l is for the j-th neuron in the l-th layer $\left(j = 1,\ldots,N_l\right)$, the sum of its weighted inputs. The weights from the i-th neuron at $\left(l-1\right)$ layer to the j-th neuron at the l-th layer are denoted by $w_{ij}^{l-1,l}$, o_j^l is the output of the j-th neuron that belongs to the l-th layer, and $f\left(net_j^l\right)$ is the j-th's neuron activation function.
A multi layer perceptron is depicted in Figure 2.

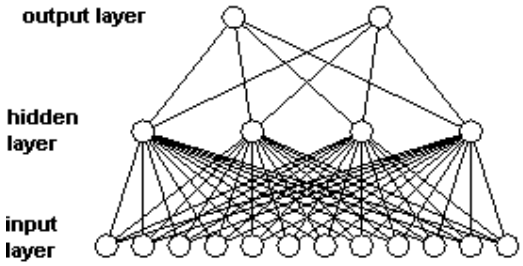

Figure 2. Multi Layer Perceptron

Training an MLP to correctly classify low from high grade tumors is typically realized by adjusting the network weights through a minimization method following an error – correction strategy. This corresponds to minimizing the error function E:

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{N_M}\left(o_{j,p}^M - t_{j,p}\right)^2 \tag{2}$$

where $\left(o_{j,p}^M - t_{j,p}\right)^2$ is the squared difference between the actual output value at the j-th output layer neuron for pattern p and the target output value. The scalar p is an index over input – output pairs. An efficient training algorithm is used, which is called Perry's nonmonotone spectral conjugate gradient with variable learning rate [16, 17]. After training, the MLP is able to discriminate low from high grade tumours by forming hyperplane decision boundaries in the pattern space.

2.2 Overview of Probabilistic Neural Networks

When the data generation function is unknown, non-parametric estimation techniques provide means for estimating the data probability density function (PDF) without any prior assumption on its form. Non-parametric estimation can be realized in a NN from, namely the PNN [18]. The PNN calculates the whole data PDF by adding the PDF estimates based on each individual data sample. The network comprises 4 layers: The input layer has a node for each feature of input data. The pattern layer has one pattern node for each training pattern. Each pattern node forms a product of the weight vector and the given example for classification, where the weights entering a node are from a particular example. After that, the product is passed through the Gaussian activation function (Eq. 4) to the summation layer which receives the outputs from pattern nodes associated with a given class. The output layer has as many nodes as existing classes. In this layer the classification decision is deduced by comparing the output of the PDF estimation for each class.

$$k(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^{d}} \exp\left(\frac{-\|x\|^2}{2\sigma^2}\right), \sigma = spread \tag{3}$$

where k is the activation function, x the pattern example, d the pattern dimension, and σ the spread of the Gaussian kernel. The spread was calculated as $\sigma = \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|x_i - x_j\right\|$, with N the number of data samples.

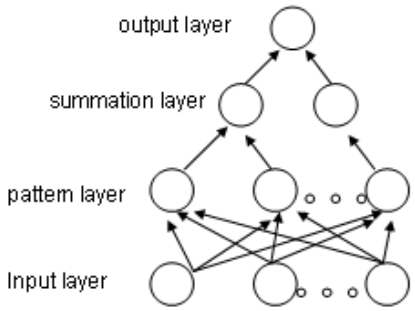

Figure 3. PNN architecture for binary pattern classification problems

3 RESULTS AND DISCUSSION

Eleven PCA-based features were extracted verifying that the original 40-dimensional feature set contained a significant amount of correlated components, such as skewness and kurtosis. Reduction of the input space dimensionality enabled us to optimize computational time. Additionally, the performance of both classifiers was optimized when constructed with the PCA features than with the original 40-features. This might be explained due to the existence of noise within data, which could be generated due to the highly variability of the staining procedure in targeting nuclei. The coefficients used to compute the various PCA scores are illustrated in Table 1.

Score1	Score 2	Score 3	Score 4	Score 5	Score 6	Score 7	Score 8	Score 9	Score 10	Score 11	Features
-0.0065	-0.2878	0.1732	-0.1254	0.0836	0.0031	-0.0845	-0.1410	0.0002	-0.0634	0.0384	**Area**
-0.0595	0.0210	0.0597	-0.0420	-0.0925	-0.0882	-0.1300	-0.2016	0.4566	0.6597	-0.4113	**Roundness**
0.1191	-0.2381	0.0232	-0.1659	-0.0824	-0.2554	0.0042	0.1048	-0.1973	0.0752	-0.0978	**Concavity**
-0.1783	-0.0770	0.1246	-0.2249	-0.0504	0.1103	0.0955	0.1925	-0.1488	0.2139	0.0776	**Density**
-0.0708	-0.2351	-0.0998	0.0943	-0.2304	-0.1057	-0.0720	0.2016	0.0237	-0.0296	0.0561	**Contrast**
0.0861	-0.0757	-0.1013	0.4645	-0.1861	-0.0686	-0.1344	-0.1423	-0.0149	-0.1188	-0.1786	**Skewness**
0.0156	-0.1294	-0.0593	0.3759	-0.0918	-0.1692	-0.1826	-0.1664	-0.2575	-0.0190	-0.3690	**Kurtosis**
0.2926	-0.0018	-0.0551	-0.0536	-0.0712	-0.1085	-0.0289	0.1479	-0.0192	0.0362	-0.0182	**energy***
-0.2971	-0.0542	0.0765	-0.0012	0.0307	0.0425	0.0535	-0.0403	-0.0648	0.0393	-0.0102	**entropy***
-0.2352	0.0647	0.0596	-0.0662	0.1199	-0.1106	0.0607	-0.1183	-0.2312	0.0343	-0.2279	**inertia***
0.2584	-0.1568	-0.0270	-0.0290	-0.1181	0.0106	-0.0676	0.1214	0.1202	-0.0147	0.0327	**local inhomeogeneity***
0.0754	-0.2028	0.0155	0.1632	-0.3197	0.1252	0.0368	0.0522	0.3078	-0.1264	0.1827	**correlation***
-0.1984	-0.1724	0.1057	-0.0040	-0.1332	-0.1216	0.2424	0.1718	0.0834	-0.1082	-0.2105	**cluster prominence***
-0.0039	-0.0094	-0.0152	0.2671	-0.1284	0.2052	-0.2490	-0.2505	-0.4027	0.3780	0.4501	**cluster shade***
0.2781	0.1050	-0.0196	-0.0633	0.0147	0.0164	0.1234	-0.0336	-0.1503	0.0692	-0.1177	**energy****
-0.2722	-0.1410	0.0618	0.0471	-0.0641	-0.0167	-0.0347	0.0701	0.0622	-0.0080	0.0762	**entropy****
-0.2812	-0.0879	0.0221	0.0078	-0.0317	-0.1610	-0.1041	0.1559	-0.0175	0.0448	0.0551	**inertia****
0.2467	-0.0025	0.0510	-0.0849	0.0204	0.2480	0.2888	-0.1881	-0.1402	0.0359	-0.1824	**local inhomeogeneity****
0.1753	-0.0760	0.1145	0.0377	-0.0856	0.3646	0.3992	-0.1843	-0.0265	-0.0109	-0.1400	**correlation****
-0.1713	-0.1897	0.1427	0.0913	-0.1845	0.0836	0.2749	0.0619	-0.0615	0.0350	-0.0641	**cluster prominence****
-0.0956	-0.1557	0.1280	0.1764	-0.3062	0.0187	0.3216	-0.0306	-0.1220	0.1308	0.0359	**cluster shade****

-0.2532	0.1737	0.0094	0.0316	0.0895	-0.0264	0.0821	-0.0797	-0.1260	0.0171	-0.0057	**short run emphasis**
0.2430	-0.1856	0.0143	-0.0528	-0.0894	0.0197	-0.1012	0.0906	0.1148	-0.0146	0.0179	**long run emphasis**
0.0297	-0.2796	0.1649	-0.1519	0.1118	-0.0157	-0.1156	-0.1599	-0.0056	-0.0610	0.0171	**gray level non uniformity**
-0.0619	-0.2690	0.1795	-0.1439	0.0814	-0.0035	-0.0611	-0.1480	-0.0240	-0.0680	0.0566	**run length non uniformity**
0.0985	-0.2157	0.2065	0.0855	0.2725	-0.0304	-0.0331	-0.0002	-0.0055	0.0796	0.0273	**range of area**
0.0750	-0.1072	-0.2808	-0.0717	0.0386	-0.2963	0.2265	-0.0828	0.0009	0.2148	0.1775	**range of roundness**
0.1432	0.1245	0.2514	-0.0399	-0.1441	-0.2823	0.0486	0.0735	-0.1830	-0.0048	0.0460	**range of concavity**
0.0928	-0.2379	0.1894	0.0262	0.2287	0.0021	-0.1109	-0.0566	-0.0382	0.0482	-0.0062	**standard deviation of area**
0.0922	-0.1289	-0.2811	-0.0894	0.0186	-0.2538	0.0995	-0.0243	-0.1019	0.1872	-0.0021	**standard deviation of roundness**
0.1545	0.1167	0.2331	-0.0417	-0.1398	-0.2637	0.0374	0.1494	-0.2378	0.0113	0.0124	**standard deviation of concavity**
0.0843	-0.2520	0.1938	0.0006	0.2195	-0.0009	-0.0882	-0.0771	0.0052	0.0414	0.0386	**maximum of area**
0.0085	-0.1798	-0.3188	-0.0434	0.0701	-0.0282	0.1025	-0.0002	-0.0530	0.0737	0.0241	**maximum of roundness**
0.0981	0.1541	0.2953	-0.0282	-0.1185	-0.2154	-0.0185	0.0336	-0.0805	-0.0045	-0.0005	**maximum of concavity**
0.0758	-0.0037	0.0965	0.4028	0.3589	-0.0427	0.1492	0.2538	0.0631	0.0594	0.0563	**skewness of area**
-0.0699	-0.0877	-0.1339	0.0309	0.1130	-0.1656	0.0137	-0.2610	-0.0674	-0.3930	-0.1832	**skewness of roundness**
0.0182	0.1960	0.2553	0.0876	-0.0599	-0.1783	0.0482	-0.2349	0.1805	-0.0576	0.1645	**skewness of concavity**
0.0452	0.0294	0.0965	0.3618	0.3677	-0.0693	0.2018	0.2730	0.0837	0.1455	-0.0039	**kurtosis of area**
-0.0327	-0.0883	-0.1881	0.0251	0.0844	-0.2756	0.3587	-0.3068	0.1394	-0.0220	0.3107	**kurtosis of roundness**
0.0252	0.1383	0.2600	0.0289	-0.0945	-0.2257	0.0434	-0.2955	0.2298	-0.0596	0.1642	**kurtosis of concavity**

Table 1. The coefficients (Principal Components) used to compute the various scores (*each feature was computed based on the co-occurrence matrix with inter pixel distance d=1, **each feature was computed based on the co-occurrence matrix with inter pixel distance d=3 [12])

The system was designed by either employing the MLP-NN or PNN. Best performance was obtained using a MLP-NN classifier with 7-hidden neurons topology that discriminating low from high grade tumours with an accuracy of 92.0%. Sensitivity and specificity ranged to 93.1% and 90.5% respectively. The PNN classifier resulted in lower rates (83.3% specificity, 91.5% sensitivity and 89.9% overall accuracy). The latter might be explained due to the complicated data probability distribution which with the PNN is approximated by using a Gaussian kernel. However, setting of the PNN classifier was easier and performed significantly faster compared to the MLP. The reason is that the probabilistic network does not need any random initialization or any feedback process.

	System classification accuracy		
Classifier	**Low-grade (specificity)**	**High-grade (sensitivity)**	**Overall accuracy**
MLP-NN (7 hidden neurons topology)	90,5%	93,1%	92,0%
PNN	83,3%	91,5%	89,9%

Table 2. Comparison results of MLP-NN and PNN in their ability to generalize when tested with the 50 new cases

It has to be stressed that the best classification result is not the one with optimum overall accuracy. The diagnostic error is most severe when misclassifying a high-grade tumour as low; these errors might provoke less aggressive tumours therapy endangering patient management. On the other hand, misclassification of low grade tumours is not so severe. Thus, as optimum classification result is considered the maximization of accuracy in recognizing high-grade tumours (sensitivity) with the lower error cost to low grade tumours (specificity). Under this perspective, the best architecture for the MLP-NN classifier consisted of 7 hidden neurons since resulted to the optimum combination of sensitivity and specificity rates (93.1% and 90.5% respectively). This network topology gave 92% overall accuracy. The best sensitivity was accomplished with 8 hidden neurons (96.5%), but this topology was not selected as optimum due to the relative significant cost in low grade tumours identification (85.7%). Figure 4 illustrates the performance variation in sensitivity and specificity for the MLP-NN classifier over different choices on hidden neurons.

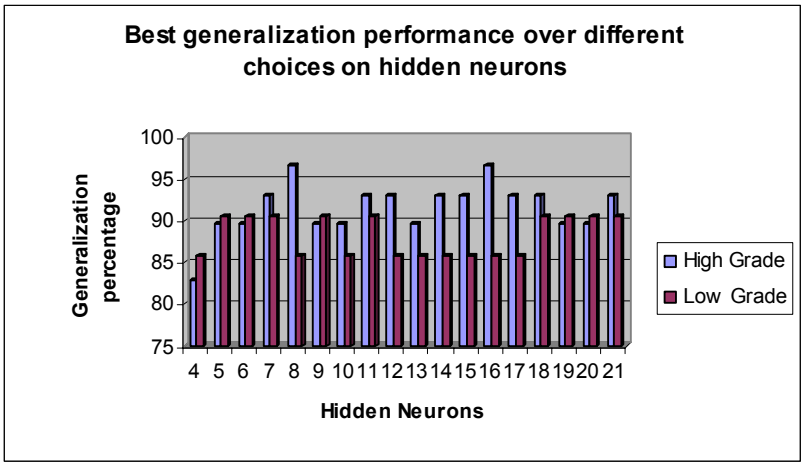

Figure 4. Best performance of the MLP-NN classifier with the respect with the number of hidden neurons for the unseen test set

The same strategic was followed when selecting the optimum performance for the PNN classifier. We tested a range of values from 0.001 to 5 for adjustable parameter σ of the Gaussian kernel. The best performance was obtained for σ=0.05 giving 91.5% and 83.3% sensitivity and specificity relatively.

4 CONCLUSIONS

In this work we propose a computer based system for the automatic identification of the degree of tumour malignancy of brain astrocytomas. The system was constructed to be compatible with daily clinical standards in order to be accessible by pathologists in clinical routine as a second opinion tool. The relatively high classification accuracy of the MLP-NN classifier when tested with new data might be indicative of the system's ability to provide an accurate and objective diagnostic estimation of tumour astrocytic tumours grade. Thus, it could be a valuable tool in the hands of clinicians, preventing and reducing sources of diagnostic errors.

5 REFERENCES

[1] L. M. DeAngelis. Brain tumors. *New England Journal of Medicine*, 344(2), 114-123, 2001.

[2] A. Kaye, D. Walker. Low grade astrocytomas: controversies in management. *Journal of Clinical Neuroscience,* 7(6), 475-483, 2000.

[3] R. A. Prayson, D. P. Agamanolis, M. L. Cohen, M. L. Estes. Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. *Journal of the Neurological sciences*, 175 (1), 33-39, 2000.

[4] W. Coons, P. Jhonson, B. Sceithauer, A. Yates, D. Pearl. Improving diagnostic accuracy and interobserver concordance in the classification and grading of Primary Gliomas. *Cancer,* 79, 1381-93, 1997.

[5] P. Sallinen, S. Sallinen, T. Helen, I. Rantala, E. Rautiainen, H. Helin, H. Kalimo, H. Haapsalo. Grading of diffusely infiltrating astrocytomas by quantitative histopathology, cell proliferation and image cytometric DNA analysis. *Neuropathology and Applied Neurobiology*, 26, 319-331, 2000.

[6] C. Decaestecker, I. Salmon, O. Dewitte, I. Camby, P. Van Ham, J. Pasteels, J. Brotchi, R. Kiss. Nearest-neighor classification for identification of aggressive versus nonaggressive astrocytic tumurs by means of image cytometry-generated variables. *Journal of Neurosurgery*, 86, 532-537, 1997.

[7] J. Martin, McKeown, D. Ramsay. Classification of Astrocytomas and Malignant Astrocytomas by Principal Component analysis and a Neural Net. *Journal of Neuropathology and Experimental Neurology,* 55(2), 1238-1245, 1996.

[8] M. Scarpelli, P. Bartels, R. Montironi, C. Galluzzi, D. Thompson. Morphometrically assisted grading of astrocytomas. *Analytical and quantitative Cytology and Histology*, 16, 351-356, 1994.

[9] N. Belacel, M. Boulassel. Multicriteria fuzzy assignment method: a useful tool to assist medical diagnosis. *Artificial intelligence in medicine*, 21, 201-207, 2001.

[10] D. Glotsos, P. Spyridonos, P. Petalas, D. Cavouras, I. Ravazoula, S. Dadioti, E. Lekka, G. Nikiforidis. Supporting the regular diagnostic procedure followed by the experts in astrocytomas malignancy

grading by means of an automatic classification methodology using Support Vector Machines. *Analytical and Quantitative Cytology and Histology,* 26:2, 77-83, 2004.

[11] N. Reinhold, W. Schlote. Topometric Analysis of Diffuse Astrocytomas. *Analytical and Quantitative Cytology and Histopathology*, 25, 12-18, 2003.

[12] P. Spyridonos, V. Zolota, D. Cavouras, G. Zenebissis, D. Glotsos, G. Nikiforidis. A computer-based image analysis system for classification of astrocytomas according the WHO grading system. In: *Proceedings of the IASTED International Conference on Signal Processing Pattern Recognition & Applications*, 371-374, 2002.

[13] S. Theodoridis, K. Koutroubas, *Feature Generation II. In Pattern recognition*, Academic Press 1999, 233-270.

[14] I. T. Jolliffe. *Principal Component Analysis*, New York: Springer-Verlag, 1986.

[15] Mahthworks, Neural Network Toolbox, http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/prepca.html

[16] D. G. Sotiropoulos, A. E. Kostopoulos and T. N. Grapsa. A Spectral version of Perry's Conjugate Gradient Method for Neural Network Training. In: *Proceedings of 4th GRACM Congress on Computational Mechanics*, 27-29 June, University of Patras, Greece, 2002.

[17] A. E. Kostopoulos, D. G. Sotiropoulos and T. N. Grapsa. A new efficient variable learning rate for Perry's spectral conjugate gradient training method. Accepted for presentation at the *1st International Conference "From Scientific Computing to Computational Engineering"*, 8-10 September, Athens, Greece, 2004.

[18] D. Specht. Probabilistic neural networks, *Neural Networks*, 3(1), 109-118, 1990.