Combining Fuzzy Cognitive Maps with Support Vector Machines for Bladder Tumor Grading

Elpiniki Papageorgiou, George Georgoulas, Chrysostomos Stylios[1], George Nikiforidis[2], and Peter Groumpos

Laboratory for Automation and Robotics, Department of Electrical and Computer Engineering, University of Patras, Rion 26500, GREECE
Tel.: +30 2610 997293; Fax: +30 2610 997309
epapageo@ee.upatras.gr, georgoul@ee.upatras.gr, groumpos@ee.upatras.gr
[1] Department of Communications, Informatics and Management, TEI of Epirus, 47100 Artas, Greece
Tel.: +30 6974638830
stylios@teleinfom.teiep.gr
[2] Computer Laboratory, School of Medicine, University of Patras, Rion 26500, Greece

Abstract. Fuzzy Cognitive Map (FCM) is an advanced modeling methodology that provides flexibility on the system's design, modeling, simulation and control. This research work combines the Fuzzy Cognitive Map model for tumor grading with Support Vector Machines (SVMs) to achieve better tumor malignancy classification. The classification is based on the histopathological characteristics, which are the concepts of the Fuzzy Cognitive Map model that was trained using an unsupervised learning algorithm, the Nonlinear Hebbian Algorithm. The classification accuracy of the proposed approach is 89.13% for High Grade tumor cases and 85.54%, for tumors of Low Grade. The results of the proposed hybrid approach were also compared with other conventional classifiers and are very promising.

Keywords: Fuzzy Cognitive Maps, Support Vector Machine, statistical learning, bladder tumor grading.

1 Introduction

For most tumor types, including bladder tumors, malignancy characterization and classification is expressed in pathological terms based on the morphology of tissues as viewed through a light microscope. Due to the subjective nature of this classification several approaches based on computer assisted methods have been used, trying to increase the diagnostic and/or prognostic value of tumors' characterization [1], [2]. Two main approaches have been reported in the literature: Data-driven models -such as statistical classifiers, Support Vector Machines (SVMs), artificial Neural Networks, Decision Trees, [3-9] and the more knowledge oriented models such as belief networks [10]. In this work, we explore the potential of a workable advanced system based on Fuzzy Cognitive Maps (FCMs) combined with SVMs to assist tumour characterization.

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part I, LNAI 4251, pp. 515 – 523, 2006.

The proposed system is based on FCMs, which is a soft computing methodology. FCMs are appropriate to explicit the knowledge and experience accumulated for years on the decision process of human experts [11], [12]. The flexibility of FCMs in system design and modeling, as well as their learning properties [13], [14], make their choice attractive for a variety of modeling and decision support tasks.

In a previous work, an FCM model has been proposed for the characterization of tumours' grade based on the analysis of conventional histopathological criteria [15], [16]. In this research work we are proposing the combination of FCMs with SVMs. SVMs have gained great attention and have been used extensively and, most important, successfully in the field of pattern recognition [17-21]. This hybrid approach has been proven quite successful and the results are very promising.

This paper is structured as follows; sections 2 and 3 briefly introduce the fundamentals of FCMs and SVMs respectively. Section 4 describes the developed FCM for the tumor grading model as well as the integration of SVMs in the whole procedure. Section 5 presents the implementation and the experimental results and in section 6 some conclusions and ideas for future work are discussed.

2 Fuzzy Cognitive Maps Background and Description

The FCM structure is similar to recurrent artificial neural networks, where concepts are represented by neurons and causal relationships by weighted links connecting the neurons. Concepts reflect attributes, characteristics, qualities and senses of the system. Interconnections among concepts of FCM signify the cause and effect relationship that a concept has on the others. These weighted interconnections represent the direction and degree with which concepts influence the value of the interconnected concepts. The directional influences are presented as all-or-none relationships, so the FCMs provide qualitative as well as quantitative information about these relationships. In [11], an analytical description of FCMs is presented.

Generally, the value of each concept is calculated, computing the influence of other concepts to the specific concept, by applying the following calculation rule:

$$A_i^{(k+1)} = f\Big(\sum_{\substack{j=1 \\ j\neq i}}^{N} A_j^{(k)} \cdot e_{ji}\Big) \tag{1}$$

where $A_i^{(k+1)}$ is the value of concept C_i at time $k+1$, $A_j^{(k)}$ is the value of concept C_j at time k, e_{ji} is the weight of the interconnection between concept C_j and concept C_i and f is the logistic sigmoid threshold function.

The methodology for developing FCMs is based on a group of experts who are asked to define concepts, describe relationships among concepts, use IF-THEN rules to justify their cause and effect suggestions among concepts and infer a linguistic weight for each interconnection [11]. Every expert describes each interconnection with a fuzzy rule; the inference of the rule is a linguistic variable, which describes the relationship between the two concepts according to every expert and determines the grade of causality between the two concepts.

The inferred fuzzy weights are aggregated and through the defuzzification method of Center of Area (CoA) an overall linguistic weight is produced, which is transformed to a numerical weight e_{ji}, belonging to the interval [-1, 1] and representing the overall suggestion of experts [11].

3 Support Vector Machines Background

Support Vector Machines (SVMs) are learning systems that are trained using an algorithm based on optimization theory [17-21]. For real life problems, given l observations $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, the SVM solution finds the hyperplane in feature space that keeps both the empirical error small and maximizes the margin between the hyperplane and the instances closest to it. This can be done by minimizing:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{2}$$

subject to

$$y_i((\mathbf{w}\cdot\phi(\mathbf{x}_i))+b) \geq 1-\xi_i,\ \xi_i \geq 0 \quad i=1,...,n \tag{3}$$

where ξ_i are slack variables, which are introduced to allow the margin constraints to be violated, and $\boldsymbol{\varphi}(\cdot)$ is the nonlinear mapping from the input space to the feature space. Parameter C controls the trade off between maximizing the margin and minimizing the error and it is usually determined through a cross-validation scheme [19], [21-22].

The class prediction for an instance $\mathbf{x}$ is given by:

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{l} y_i\alpha_i \left\langle \boldsymbol{\varphi}(\mathbf{x}_i)\cdot\boldsymbol{\varphi}(\mathbf{x})\right\rangle + b\right) \tag{4}$$

where the coefficients a_i are calculated by maximizing the Lagrangian:

$$\sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j \left\langle \boldsymbol{\varphi}(\mathbf{x}_i)\cdot\boldsymbol{\varphi}(\mathbf{x}_j)\right\rangle \tag{5}$$

subject to $\sum_{i=1}^{l} y_i\alpha_i = 0$ and $0 \leq \alpha_i \leq C$, $i = 1,2,...,l$

The points for which a_i>0, are called Support Vectors and are the points lying closest to the hyperplane. If the nonlinear mapping function is chosen properly, the inner product in the feature space can be written in the following form:

$$\left\langle \boldsymbol{\varphi}(\mathbf{x}_i)\cdot\boldsymbol{\varphi}(\mathbf{x}_j)\right\rangle = K(\mathbf{x}_i,\mathbf{x}_j) \tag{6}$$

where K is called the inner-product kernel [17-21].

One of the most commonly used kernel is the Radial Basis Function (RBF) kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}_i\|^2\right) \tag{7}$$

where the width σ^2 is specified a priori by the user and is common for all kernels [22].

4 Fuzzy Cognitive Map Grading Model

Histopathologists with deep knowledge and great clinical experience were our experts whom we asked to develop and construct the FCM model for tumor grading using the methodology presented in section 2 and described analytically in [15]. Experts defined the main histopathological features (concepts) that determine the final grade characterization. More specifically, eight well documented in the bibliography histopathological criteria (features) essential for tumour grading (Table1) [23-24] were selected. In every day practice, each tissue section (patient slide) is evaluated retrospectively by histopathologists using these features. These considered features are the causative variables or factors of the tumour grading system that have been selected by experts to construct the FCM for tumour grading [15], [16].

The FCM tumor grading model was developed consisting of the following 9 concepts: Concept C_1 represents the cell distribution, C_2 represents the cell size, C_3 the cell number, C_4 the cytoplasm, C_5 the nuclei, C_6 the nucleoli, C_7 the necrosis, C_8 the mitoses and C_9 the degree of tumour grade.

Table 1. Main factors for grading

Histological feature	**Assessment Possible**
Cell distribution	Even, clustered
Cell size	Uniform, pleomorphic
Cell number	Numerous, variable
Cytoplasm	Homogeneous, variable
Nuclei	Uniform, irregular, very irregular, bizarre
Nucleoli	Inconspicuous, evident, prominent
Necrosis	Inconspicuous, frequent
Mitosis	Absent-rate, occasional, numerous

In order to use the FCM model histopathologists were asked to examine each tissue section retrospectively and estimate the value of the eight histopathological variables (Table 1); these values were transformed in the range [0, 1], and were assigned to the corresponding concepts.

Furthermore, histopathologists were asked to explain the cause-effect relationships among these concepts using IF-THEN rules that were described in section 2. Thus after defuzzification of the linguistic weights with GoA, the initial weights of the FCM tumor-grading model were determined and given in Figure 1.

The tumor grading procedure is based on the determination of the values of two concepts: concept "Nuclei" which represents the characteristics of nuclear appearance and may be considered as figuring out the nuclear grading concept, and concept "Grade" that characterizes the degree of tumor malignancy. These two concepts are considered the main attributes for the final degree of tumor characterization.

The unsupervised Nonlinear Hebbian Learning algorithm is used to modify the weights of the FCM grading model according to the initial values of concepts for each examined case of urinary bladder tumors. Also, according to the NHL algorithm, experts were asked to select the input and output concepts that determine the triggering process. The concepts C_5 and C_9 were defined as the Decision Output Concepts (DOCs), which determine the tumor grade. All the other concepts have been defined as input concepts.

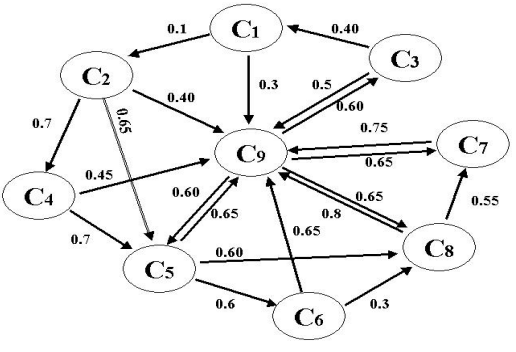

Fig. 1. The FCM tumor grading model consisting of 9 concepts and 21 weight relations

5 Description of the Tumour Grading Procedure Using SVMs

The FCM tumor-grading model can be used, after the development of the FCM model and the determination of the necessary specifications for the implementation of the NHL algorithm. For each case the training procedure is applied and then, through the use of an SVM with RBF kernels, the system determines the grade of the tumor (either "low" or "high").

The experimental data consisted of one hundred twenty-nine tissue sections (slides) each one from a different patient with superficial transitional cell carcinoma were retrieved from the archives of the Department of Pathology of University Hospital of Patras, Greece. Tissue sections were routinely stained with Haematoxylin-Eosin. Every case was reviewed independently by the doctors-experts to safeguard reproducibility. Histopathologists had classified the cases following the World Health Organization (WHO) grading system as follows: eighty-three as Low Grade, and forty-six as High Grade. Because of the restricted number of cases, we used the leave

one out method [22] and the extreme case of multifold cross validation in order to evaluate the performance of the proposed methodology. Therefore, at each experiment we used 128 of the cases to build the model and then the model was tested on the example left out. However in order not to use the same data set for both building the model and estimating its performance [27], for each one of the 129 subsets we employed again the leave one out method within the 128 cases that consisted the training set in order to select the model's parameters. Once the parameters were optimized, the SVM was retrained for the whole training set (128 cases) and its performance was evaluated using the corresponding case that was originally left out. Figure 2 illustrates the integrated method of FCMs with SVMs for tumor grading.

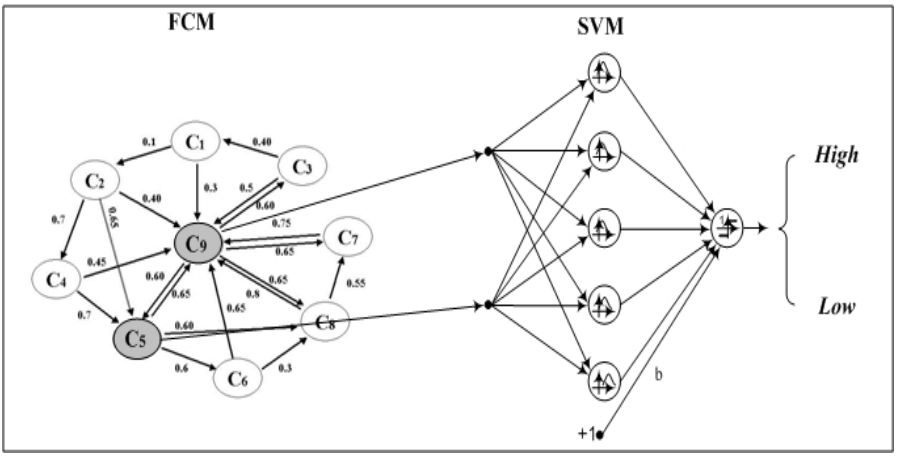

Fig. 2. The process of tumor grading combining FCM with SVM

It must be mentioned that due to the imbalanced nature of the available data, we introduced a modified formulation of the SVM algorithm. We used different error weights C^+ and C^-, so that to penalize more heavily the undesired type of error, and/or the errors related to the class with the smallest population [20], [26]. Therefore, the optimization problem is modified as follows:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C^- \sum_{i:y_i=-1} \xi_i + C^+ \sum_{i:y_i=+1} \xi_i \tag{9}$$

Selecting higher penalty value for the class with the smallest population, which in most cases is the class that needs to be correctly identified, we induce a boundary that is more distant from that class. In our experiments, we tested different values for the error weights C^+ and C^- keeping their ratio equal to 1/(46/83) (the inverse ratio of the corresponding cardinalities).

In previous works, our research group introduced the FCM-grading tool, working exclusively on the qualitative assessments of histopathological variables. The FCM

grading tool had achieved a classification accuracy of 80%, and 87.5% for tumors of low and high grade respectively [16]. Here, this work further extends this research exploring means to improve classification accuracy by utilizing SVM classifier to assist medical diagnosis.

Table 2. Comparative results

Tumor category	FCM using SVMs	FCM-GT	quadratic classifier	linear classifier	9-nn
Low-Grade	85.54%	80 %	86.75%	85.56%	90.36%
High-Grade	89.13%	87.5%	78.26%	84.78%	78.26%
Overall-Accuracy	86.82%	84.5%	83.72%	85.27%	86.06%

The proposed hybrid grading tool, where the FCM-grading tool is combined with the SVM algorithm has achieved a classification accuracy of 85.54% for low grade tumors and 89.13% for high grade tumors. Thus, the introduction of the SVM stage improves the classification accuracy of the FCM-grading tool which had used initially the minimum distance method and only one target concept [16].

For comparison reasons, experiments were also conducted using three conventional classifiers: the k-nearest neighbor, the linear and the quadratic classifiers [27] and the results are summarized in Table 2.

6 Conclusions

In this research effort, a new extended structure is proposed where SVM classifier is added to the FCM-tumor grading model to improve the characterization of urinary bladder tumors. The hybridization of FCM model with SVMs exhibited high performance in correctly classifying tumors into two categories, low-grade and high-grade respectively, utilizing all the available diagnostic information. The proposed method could be considered as an efficient classification procedure able to make decisions with highly diagnostic accuracy.

Furthermore, the tumor characterization procedure based on the integration of the soft computing technique of FCMs and the statistical classification technique of SVMs has been compared with other conventional classifiers and the results are very promising. The current approach yield balanced performance for both classes and outperforms the other methods tested in terms of correctly classified high-grade instances. The hybrid tumor grading tool is fast, easily implemented in clinical practice and performs high accuracy in terms of specificity and sensitivity.

Acknowledgements

This work was supported by the “PYTHAGORAS II” research grant co funded by the European Social Fund and National Resources.

References

1. Parker, S.L., Tony, T, Bolden, S, Wingo, P.A: Cancer Statistics, Cancer Statistics. CA Cancer J Clinics, **47** 5 (1997) 5-27.
2. Ooms, E, Anderson, W, Alons C., Boon M, Veldhuizen R: "Analysis of the performance of pathologists in grading of bladder tumours", Human Pathology, **14** (1983) 140-143.
3. Gil, J, Wu, H, Wang, B.Y.: Image analysis and morphometry in the diagnosis of breast cancer, Microsc. Res. Techniq. **59** (2002) 109-118.
4. Catto J, Linkens D, Abbod M, Chen M, Burton J, Feeley K, Hamdy F: Artificial Intelligence in Predicting Bladder Cancer Outcome: A Comparison of Neuro-Fuzzy Modeling and Artificial Neural Networks. Artif Intell Med **9** (2003) 4172-4177.
5. Choi H, Vasko J, Bengtsson E, Jarkrans T, Malmstrom U, Wester K, Busch C.: Grading of transitional cell bladder carcinoma by texture analysis of histological sections. Anal Cell Pathol **6** (1994) 327-343.
6. Jarkrans T, Vasko J, Bengtsson E, Choi H, Malmstrom U, Wester K, Busch C. Grading of transitional cell bladder carcinoma by image analysis of histological sections. Anal Cell Pathol **18** (1995) 135-158.
7. Michie, D., Spiegelhalter, D. J., Taylor, C.C.: Machine Learning, Neural and Statistical Classification. Englewood Cliffs, N.J.: Prentice Hall (1994). Data available at http://www.ncc.up.pt/liacc/ML/statlog/datasets.html.
8. Podgorelec, V, Kokol, P, Stiglic, B, Rozman, I.: Decision Trees: An Overview and Their Use in Medicine, Journal of Medical Systems, **26** 5 (2002) 445-463.
9. Dreiseit, S, Ohno-Machado, L, Kittler, H, Vinterbo, S, Billhardt, H, Binder, M: A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions, Journal of Biomedical Informatics **34** (2001) 28–36.
10. Antala, P., Fannesa, G., Timmermanb, D., Moreaua, Y, Moor B.D.: Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection, Artificial Intelligence in Medicine **29** (2003) 39–60.
11. Stylios, C. D. and Groumpos, P.P.: Fuzzy Cognitive Maps in Modelling Supervisory Control Systems, J. of Intelligent & Fuzzy Systems, **8**, (2000) 83-98.
12. Papageorgiou, E., Stylios, C., Groumpos, P: An Integrated Two-Level Hierarchical Decision Making System based on Fuzzy Cognitive Maps (FCMs), IEEE Transactions on Biomedical Engineering**, 50** 12 (2003) 1326-1339.
13. Papageorgiou, E.I., Stylios, C.D., Groumpos, P.P: Active Hebbian Learning to Train Fuzzy Cognitive Maps, Int. J. Approx. Reasoning, **37** (2004) 219-249.
14. Papageorgiou, E.I., Groumpos, P.P.: A weight adaptation method for fine-tuning Fuzzy Cognitive Map causal links, Soft Computing Journal, **9** (2005) 846-857.
15. Papageorgiou, E.I., Spyridonos, P., Ravazoula, P., Stylios, C.D., Groumpos, P.P., Nikiforidis, G: Advanced Soft Computing Diagnosis Method for Tumor Grading, Artif Intell Med, **36** (2006) 59-70.
16. Papageorgiou, E.I., Spyridonos, P. Stylios, C.D., Nikiforidis, G., Groumpos, P.P: The Challenge of Using Soft Computing Techniques for Tumor Characterization, *16th Intern. Conf. on Artificial Intelligence and Soft Computing (ICAISC)* Zakopane, Poland, 7-11 June, Lecture Notes in Computer Science 3070, (2004) 1031-1036.
17. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, **2** 2 (1998) 121-167.
18. Vapnik, V. N: The nature of statistical learning theory. New York: Springer-Verlag, 1995.
19. Muller, K. R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B: An Introduction to Kernel-Based Learning Algorithms. IEEE Trans. Neural Networks, **2** 2 (2001) 181-201.

20. Veropoulos, K., Cambell, C. and Cristianini, N.: Controlling the sensitivity of support machines, Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI99), Stockholm, 1999. pp. 55-60.
21. Cristianini, N. and Shawe-Taylor, J: An Introduction to Support Vector Machines. Cambridge University Press, (2000).
22. Haykin, S: *Neural Networks: A Comprehensive Foundation*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall, (1999).
23. Spyridonos, P, Ravazoula, P, Cavouras, D, Berberidis, K, Nikiforidis, G.: Computer-based grading of haematoxylin-eosin stained tissue sections of urinary bladder carcinomas, Med Inform & Intern Med, **26** 3 (2001) 179-190.
24. Bostwick, G, Ramnani, D, Cheng, L.: Diagnosis and grading of bladder cancer and associated lesions, Urologic Clinics of North America, **26** (1999) 493-507.
25. Salzberg, S.L: On Comparing Classifiers: Pitfalls to avoid and a Recommended Approach, *Data Mining and Knowledge Discovery*, **1** (1997) 317-328.
26. Osuna, E.E., Freund, R., and Girosi, F: Support Vector Machines: Training and Applications, MIT, A.I. Memo No. 1602, March (1997).
27. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. 2nd ed. John Wiley & Sons, (2001).
28. Murphy, WM.: Urothelial neoplasia, Pathology and pathobiology of the urinary bladder and prostate, Baltimore, Williams & Wilkins, (1992).