Automatic computer estimation of histological Fuhrman grade in kidney cancer

Jarosław Kurek

Michał Kruk

Bartosz Swiderski

The Faculty Of Applied Informatics and Mathematics
ul. Nowoursynowska 166 Warsaw, Poland
E-mail : jaroslaw_kurek@sggw.pl
E-mail : michal_kruk@sggw.pl
E-mail : swidersb@iem.pw.edu.pl

Stanisław Osowski

Warsaw University of Technology
Military University of Technology
ul. Koszykowa 75, Warsaw, Poland
E-mail:sto@iem.pw.edu.pl

Abstract—**The paper presents an automatic approach to assessment of the stage of development of the kidney cancer on the basis of Fuhrman grades. The stage of advancement level of cancer is usually associated with 4 Fuhrman grades. Our approach to Fuhrman grade assessment is composed of few steps. The first one is extraction of the numerical features from the microscopic image of the histological slides of the biopsy of kidney by using mathematical morphology. The next step is the features selection providing descriptors of the best class separating abilities. The last one is application of the automatic classifiers and data mining techniques to assign the actually available samples to one of four classes.**

Keywords—**features selection, neural networks, classification system, kidney cancer, data mining**

I. Introduction

Renal Cell Carcinoma (RCC) accounts for 2% - 3% of all malignant diseases in adults. It is the seventh most common cancer for men and the ninth most common for women [1]. More than 64,000 new cases are diagnosed each year in United States and about 13,500 deaths from RCC annually [1]. RCC consists of a heterogeneous group of tumors with distinct genetic and metabolic defects, as well as histopathologic features. Table I depicts the commonly used classification of RCC from the point of view of histology.

TABLE I.   CLASSIFICATION OF RENAL CELL CARCINOMA

Histology	Frequency	Characteristics
Clear cell	60%-70%	Cells with clear cytoplasm and acinar or sarcomatoid growth pattern.
Papillary	5%-15%	Type I: papillae lined with a layer of tumor cells with scant pale cytoplasm and low-grade nuclei. Type II: abundant eosinophilic cytoplasm and large pseudo stratified nuclei with prominent nucleoli; aggressive subtype.
Chromophobic	5%-10%	Cells are large with finely reticulated eosinophilic cytoplasm, and atypical nuclei with perinuclear halo with solid, tubular or sarcomatoid growth
		pattern; indolent course.
Oncocytic	5%-10%	Benign neoplasms.
Collecting duct	<1%	Papillary or sarcomatoid growth pattern.

In this paper we are focusing on the clear cell (CCRC) type only, since it is the most common kidney cancer. In current clinical practice for this type of cancer, the diagnosis is based on the microscopic image of a neoplasm cells at application of hematoxylin and eosin (H&E) staining. The most popular and widely used system for grading RCC is a nuclear grading system described in 1982 by Fuhrman [2,3]. Fuhrman grade is defined on a scale of 1-4, where grade 1 represents the best prognosis and grade 4 the worst one. Assignment stage of kidney cancer is based on the nucleus shape analysis of the microscopic image. The typical microscopic images representing different stages of advancement of CCRC are depicted in Figure1 and their short description in Table II.
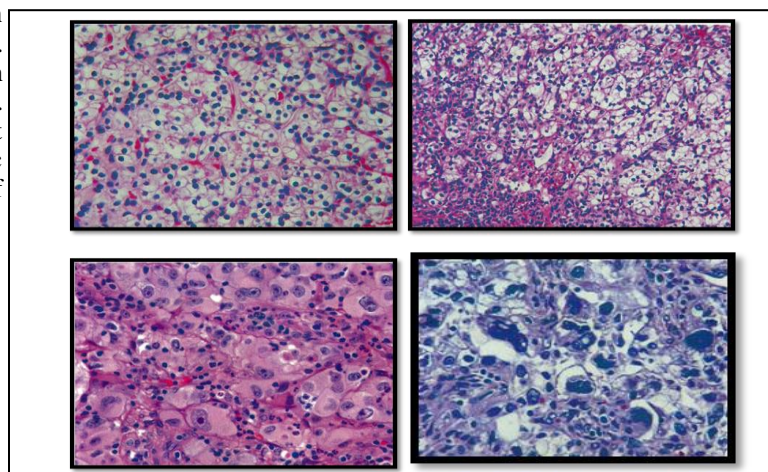

Figure 1. The typical examples of Fuhrman grade clear-cell renal carcinoma: a) grade 1, b) grade 2, c) grade 3, d) grade 4

International Journal of Artificial Intelligence and Neural Networks – IJAINN
Volume 4 : Issue 1   *[ISSN 2250 – 3749]*

Publication Date : 09 January 2014

TABLE II.   FUHRMAN GRADE OPTICAL FEATURES OF MICROSCOPIC IMAGE.

Fuhrman Grades	Optical features
Grade 1	Nuclei round, uniform, approximately 10 μm in diameter; nucleoli inconspicuous or absent
Grade 2	Nuclei sligtly irregular, approximately 15 μm in diameter; nucleoli evident
Grade 3	Nuclei very irregular, approximately 20 μm in diameter; nucleoli prominent
Grade 4	Nuclei bizarre and multilobated, 20 μm or greater in diameter, nucleoli prominent, chromatin clumped

II.   Features generation

A. *Database of microscopic images*

The database used in our experiments consists of 300 real microscopic images registered at 2070x1548 resolution. The images have been acquired from the patients of Warsaw hospitals.

B. *Nuclei segmentation*

The first step of feature generation is to extract the nuclei from the microscopic image of the histological slides of the kidney. It has been achieved by the following morphological operations:

- De-blurring of the image - remove all necessary artifacts from images. Morphological erosion and Gaussian filter of 50x50 windows size have been used.

- Edge detection - it is the most important morphological operation which has huge impact on the next operations and final results. In this step image thresholding based on histogram, according to Otsu method and application of gradient method to nuclei detection have been performed.[5].

- Binary mask - using edge detection and filling holes operations results we have prepared binary mask to cover on original image extracting finally the nuclei from the original image.

- Separating the merged nuclei - from time to time we have observed merged nuclei which we have separated by using the watershed algorithm [16].

After performing all above morphological operations the nuclei existing in the image are extracted.

C. *Database of the segmented nuclei*

Based on nuclei segmentation operation we have prepared database of nuclei, belonging to four classes according to Fuhrman grading scale. Table III presents the composition of database depicting the population of samples belonging to each analyzed class.

TABLE III.   THE DATABASE OF CLASSES OF FUHRMAN GRADE USED IN ANALYSIS

Fuhrman	Nuclei/Sample Count
Grade	
Grade 1	416
Grade 2	424
Grade 3	300
Grade 4	137

D. *Numerical features generation*

Using the extracted nuclei we have generated the following numerical features, characterizing the texture , geometry and distribution of colors.

Texture Haralick descriptors:[6]
1) *angular second moment*
2) *contrast*
3) *correlation*
4) *sum of squares: variance*
5) *inverse difference moment*
6) *sum average*
7) *sum variance*
8) *entropy*
9) *sum entropy*
10) *difference variance*
11) *difference entropy*
12) *info.measure of correlation1*
13) *info.measure of correlation2*

Geometrical descriptors [8]:
14) *area*
15) *major_axis_length - the length (in pixels) of major axis of the ellipse that has the same second-moments as the region*
16) *minor_axis_length - the length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region.*
17) *eccentricity - the eccentricity of the ellipse that has the same second-moments as the region.*
18) *convex_area - the number of pixels in the convex hull, with all pixels within the hull filled in*
19) *equiv_diameter - the diameter of a circle with the same area as the region. computed as sqrt(4*area/pi).*
20) *solidity - the proportion of the pixels in the convex hull that are also in the region. computed as area/convexarea.*
21) *perimeter.*

Descriptors based on RGB statistics:
22) *R1=R/(R+G+B)*
23) *G1=G/(R+G+B)*
24) *B1=B/(R+G+B)*

Descriptors based on histogram:
25) *mean of histogram*
26) *standard deviation of histogram*
27) *kurtosis of histogram*

Mayall's descriptors [7]:
28) *heterogeneity measure*

$$hetero = \frac{N_B + N_W}{N_B + N_G + N_W}$$

where:

N_B - number of pixels in the nucleus labeled black

N_W - number of pixels in the nucleus labeled white

N_G - number of pixels in the nucleus labeled gray

29) homogeneity=1-heterogeneity

With the given window we count the number of black labeled pixels and the number of white labelled pixels and take the aboslute difference diff. Only those mesh windows that fit totaly within the nucleus "row" are used in the computation.

30) clump - refletcs the size distribution of the franules and is given by:

$$clump = \frac{\sum mesh \quad diff}{N_B + N_W}$$

31) condens - reflects the fraction of large granules with respect to total nuclear area.

$$condens = \frac{\sum mesh \quad diff}{N_B + N_G + N_W}$$

III.   Features selection

The next step is to choose the best set of descriptors from all 31 already defined, which will form the features used in automatic classification. The selected features should have the ability of best separation of Fuhrman grades.

To choose the best feature set we have applied the sequential feature selection [8]. This approach selects a subset of features (the columns) from the data matrix X containing all descriptors, that best predict the class defined in the associated vector y. The selection is done sequentially until there is no improvement in prediction of classes [8]. It's possible to apply different classifier e.g. multilinear regression, SVM, KNN, MLP, defining the criterion used to select features and determining the conditions of stop.

Starting from an empty feature set, sequential feature selection creates candidate feature subsets by sequentially adding each of the features not yet selected. For each candidate feature subset, sequential feature selection performs 10-fold cross-validation by repeatedly calling function with different training subsets of X (XTRAIN) and y (ytrain), and test subsets of X (XTEST) and y (ytest). The result of the method is in the form of logical vector, indicating which features are finally chosen. The typical loss measures include sum of squared errors for regression models, and the number of misclassified observations for classification models. Applying sequential feature selection we have obtained the following optimal set of features [8]:

Haralick texture features:
1)   Contrast.
2)   Correlation
3)   Sum Average
4)   Sum Variance
5)   Difference Entropy
6)   Info.Measure of Correlation

Geometrical features:
7)   Minor_Axis_Length
8)   Equiv_Diameter

Features based on RGB statistics:
9)   R1=R/(R+G+B)
10) B1=B/(R+G+B)

Features based on histogram:
11) mean of histogram
12) standard deviation of histogram

Mayall's features :
13) homogeneity

The other descriptors have been found not appropriate in Fuhrman grade recognition.

IV.   Choice of classifier

In our solution we have applied few classifiers performing the same task.

A.   SVM classifier

The assignment of every nuclei sample to one of the 4 Fuhrman classes has been done by applying the Support Vector Machine (SVM) of the Gaussian kernel [8,9,10,11]. The applied classifier is a simple circuit structure of one hidden Gaussian kernel layer and one output linear unit performing the weighted summation. We have chosen the Gaussian kernel due to its universal character and very good performance in comparison to other choices, like linear, polynomial or spline functions. The learning process of SVM network is relatively easy and effective since the whole learning task is simplified to the solution of the quadratic problem with linear constraints. In our experiments we have used the modified Platt algorithm, implementing the sequential optimization [15].

The hyperparameters: gamma of the Gaussian function and the regularization constant C have been adjusted by repeating the learning experiments for the set of its predefined values and choosing the best one on the validation data sets. In the case of four Fuhrman classes we have decided to use "one-against-all" strategy because it was found more accurate in our problem.

B.   k-nearest neighbours classifier

The basic algorithm of k-nearest neighbors classification assigns an input sample vector, which is of unknown class membership, to the majority of classes of its nearest neighbours. In this approach we have used the number of nearest neighbors equal k=4.

C.   Decision tree

A decision tree is a support tool that uses a tree-like graph or model of decisions. At every node we split the data into 2 classes on the basis of some well adjusted thresholds. The subsets which are not split form the terminal nodes (leaves).

Each terminal node assigns the analyzed input pattern to one of four classes corresponding to four Fuhrman grades.

D. *Bootstrap aggregating classifier*

Bootstrap aggregating (bagging) is a machine learning ensemble meta-algorithm, designed to improve the stability and accuracy of machine learning used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree methods, it can be used with any type of methods [14].

In this paper the bagged classification ensemble of trees was used as one of the classifier solution. Construction of bagged classification ensemble is depicted on figure 2.
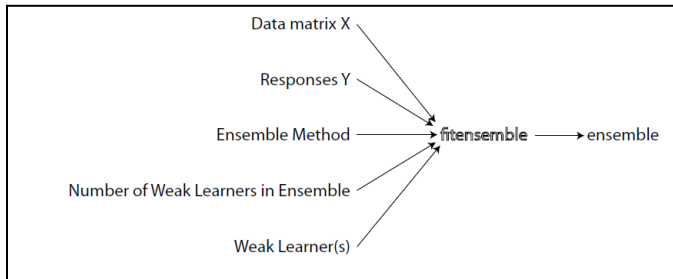

Figure 2.   Structure of bagged classification ensemble.

In our approach we have used 200 trees in this method.

V.   **The numerical results**

On the basis of chosen features we have applied four classifiers: the SVM, k-nearest neighbours, decision tree and the bagged classification ensemble of tree. In the numerical experiments we have applied cross validation. The available data set (1277 samples) was split into two independent parts: 80% samples for learning and 20% for testing. All data belonging to learning and testing sets were chosen in a random way. The experiments have been repeated 10 times at different compositions of both sets. In the experiments we have applied K= 4 neighbors for k-NN and 200 trees for bagged classification ensemble of tree. In the case of SVM we have found the hiperparameters: C=100 and gamma=1 as the best. The results of experiments are presented in table IV below.

TABLE IV.   COMPARATIVE RESULTS OF APPLYING CLASSIFIERS

Classifier	Parameters	Mean error [%]
SVM	C=100, gamma=1	5.10%
k-nn	K=4	10.69%
Decision tree		6.27%
Bagged classification ensemble of trees	K=200	3.50%

VI.   **Conclusions**

On the basis of many performed experiments we have come to conclusion that the best results of recognition (3.5% of the mean relative error) are achieved at application of bagged classification ensemble of trees. Application of the SVM classifier provides recognition of Fuhrman grade with the relative testing error in range of 5.1%. The result of K-nearest neighbors was not satisfying and the best results obtained at k=4 was 10.69% (relative error). However, the training time of decision tree was much shorter than SVM classifier. The ordinary decision tree has resulted in the relative error around 6.27%.

On the basis of the numerical experiments we can conclude that it is possible to build the automatic system which is able to classify the analyzed microscopic image of CCRC to one of four classes of Fuhrman grading scale with the accuracy above 96%.

References

[1]   Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62(1):10-29.

[2]   Fuhrman SA, Lasky LC, Limas C., Prognostic significance of morphologic parameters in renal cell carcinoma. Am J Surg Pathol. Oct 1982;6(7), pp. 655-663.

[3]   Kontak J.A., Campbell S.C., Prognostic factors in renal cell carcinoma, The Urologic clinics of North America. 2003 ; 30 (3), pp. 467-480.

[4]   Allen D. C., Cameron R. I. "Histopathology Specimens: Clinical, Pathological and Laboratory Aspects" , Springer, London 2003, pp.288-289.

[5]   M. Sezgin and B. Sankur (2004). "Survey over image thresholding techniques and quantitative performance evaluation". Journal of Electronic Imaging 13 (1): 146–165. doi:10.1117/1.1631315.

[6]   Robert M. Haralick, "Statistical and structural approaches to texture". Proc. IEEE, vol. 67, no. 5, pp. 786-804, 1979.

[7]   I T Young, P W Verbeek, B H Mayall " Characterization of chromatin distribution in cell nuclei", 10/1986; 7(5): pp.467-74.

[8]   Matlab user manual, MathWorks, Natick, 1999

[9]   L. Platt, L. Fast training of SVM using sequential optimization, (in Scholkopf, B., Burges, B.,& Smola, A., Eds. Advances in kernel methods – support vector learning. Cambridge: MIT Press), 1998. pp. 185-208

[10]   B. Schölkopf, A. Smola, Learning with Kernels, Cambridge, MA, MIT Press, 2002

[11]   V. Vapnik, Statistical Learning Theory, N.Y.: Wiley, 1998.

[12]   J.Milgram M.Cheriet R. Sabourin, "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs? Tenth International Workshop on Frontiers in Handwriting Recognition (2006)

[13]   [1] Breiman, L., J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.

[14]   Breiman, L. "Bagging predictors". Machine Learning 24 (2): 123–140. (1996). doi:10.1007/BF00058655. CiteSeerX: 10.1.1.121.765

[15]   Hsu, C.W., Lin, C.J., A comparison methods for multi class support vector machines, IEEE Trans. Neural Networks, 2002, vol. 13, pp. 415-425

[16]   Soille P., Morphological image analysis, principles and applications, 2003, Springer, Berlin