Improved computer recognition of Fuhrman grading system in analysis of clear-cell renal carcinoma

M. Kruk & J. Kurek
Warsaw University of Life Sciences, Warsaw, Poland

S. Osowski
Warsaw University of Technology, Military University of Technology, Warsaw, Poland

R. Koktysz
Military Institute of Medicine, Warsaw, Poland

ABSTRACT: The paper presents the improved system of recognition of the Fuhrman grading in the analysis of clear-cell renal carcinoma. The previous system was described in (Kruk et al. 2014). The proposed improvement is based on the new segmentation method which gives better results. A redundant wavelet transformation is used for denoising the image and enhancing the edges in multiple levels of resolution. The image gradient is estimated using the wavelet transformation. The watershed transform is then applied to the obtained gradient image, thanks to this the segmented regions that do not satisfy specific criteria are removed.

1 INTRODUCTION

A clear-cell renal carcinoma (CC-RCC) is a very common renal disease in adults aged ≥ 50 years. The incidence of renal cell carcinoma has gradually increased each year over the past 30 years. The tumors show a solid, nesting, and tubular pattern of growth with tumor cells arranged in nests and separated from each other by an extensive network of delicate sinusoidal vascular channels. The most popular and widely used system for grading CC-RCC is a nuclear grading system described in 1982 by Fuhrman (Fuhrman et al. 1982).

The Fuhrman grading scheme describes 4 nuclear features for assigning a specific grade: nuclear size, nuclear shape, chromatin pattern, and size of the nucleoli. It defines a scale of 1–4, where grade 1 carries the best prognosis and grade 4 the worst. The grade is strictly correlated with the stage of development of illness. Nuclear grade has been shown to be independent on the tumor type as a prognostic factor, but its value in specific histological subtypes of renal cell carcinoma is still in question.

The grading schema of CC-RCC is based on the microscopic image of the neoplasm cells with hematoxylin and eosin (H&E) staining. Grade 1 tumors have the round, uniform nuclei with inconspicuous or absent nucleoli. Nuclear contours at grade 2 are more irregular than in grade 1; the nuclei are about 15 microns in diameter. They may be visible at high magnification. At grade 3 the nuclear contours are even more irregular in size and shape. The nuclear diameters can approach 20 microns, and the nucleoli are readily seen. The size of cells at grade 4 exceeds 20 micrometers, and the pleomorphic and hyperchromatic nuclei, and prominent nucleoli in a minority of cells, are usually observed.

The main task of this paper is to improve an automatic computer system which was described in (Kruk et al. 2014) that would evaluate the Fuhrman grade of renal cells with the accuracy comparable to that of human experts. To solve the problem, we apply the segmentation algorithm described in (Jung et al. 2002) which gives better results than standard watershed algorithm. To perform classification we use the Support Vector Machine (SVM) classifier based on the data representing over 100 patients. To counteract the problem of biasing the score, we use 3 medical experts. Each expert performed the manual estimation of the Fuhrman grade of cells for the same set of images. In the learning phase of the system we have used only the cells for which class recognition was the same in the assessment of all experts. In this way we incorporate the average knowledge and experience of different experts in the learning process of the classifier, thereby reducing the interobserver and intraobserver variabilities.

2 MATERIALS

All medical materials used in experiments were gathered from Warsaw Military Institute of Health Services. 70 microscopic images representing all 4 grades were prepared and then used to learn the system. In the verification of the established system we used the additional 62 patients of different Fuhrman grade of the illness and analyzed 94 randomly selected images representing these patients.

The microscopic digital images of hematoxylin and eosin (H&E) stained kidney neoplasm cells of

the appropriate tumor area were collected. They were acquired in a magnification of 400× through an Olympus BX-61 microscope and registered with an Olympus DP-72 camera in RGB format at a resolution of 2070 ×1548. The example of such images is presented in Figure 1.
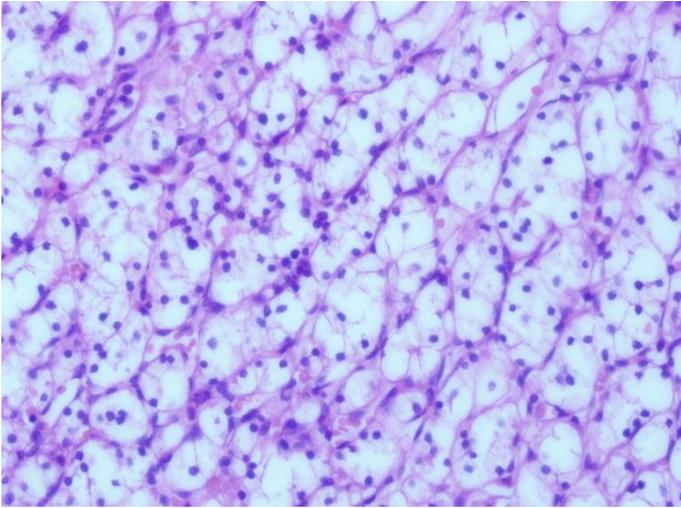

Figure 1. Example of the microscopic image of kidney tissue

3 METHODS

The proposed system contains five main steps: (1) segmentation of nuclei, (2) generation of numerical descriptors, (3) evaluation and selection of descriptors, (4) application of the classifier determining the Fuhrman grade, (5) saving results to the database.

3.1 *The segmentation algorithm*

The previous segmentation algorithm was based on the gradient method of nuclei detection, instead of the often-used automatic thresholding such as Otsu method (Gonzales et al. 2008). However, its main disadvantage is too high sensitivity to the local extrema, hence sometimes the over segmentation has been observed. The proposed new approach to segmentation was described in (Jung et al. 2002) and is based on the wavelet method of image denoising and edge enhancement in multiple resolutions. In this approach, the gradients of the denoised and enhanced image are estimated using the wavelet transform, and then the watershed transform is applied to the obtained gradient image. In these experiments we have applied Daubechies wavelet function db4. and 4 levels of decomposition. Wavelet transformation improves tumor robustness of the watershed transformation. The watershed implementation uses the immersion simulation approach. A post-processing stage is finally applied to remove over-segmented regions with small areas, and to merge

erroneously segmented regions, that are separated by weak borders after denoising and enhancement.

The example of an original image is shown in Fig. 2a and the result of its segmentation is depicted in Fig. 2b.
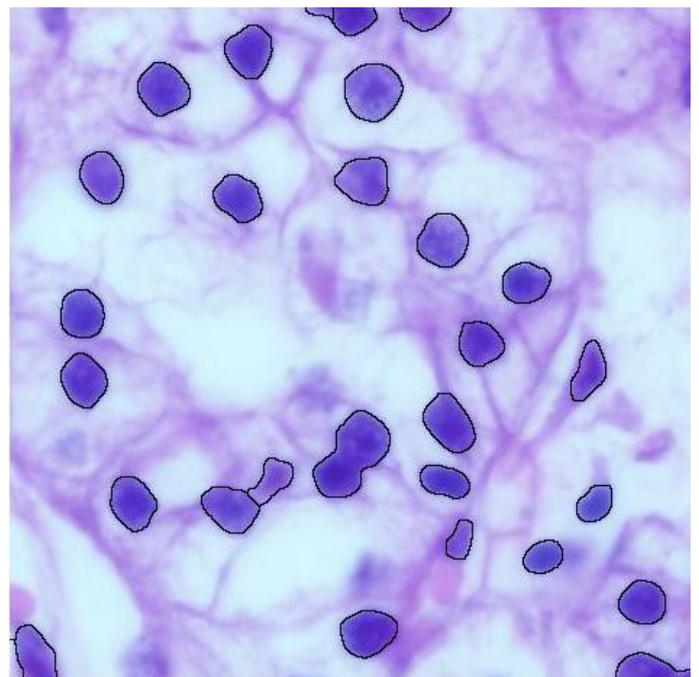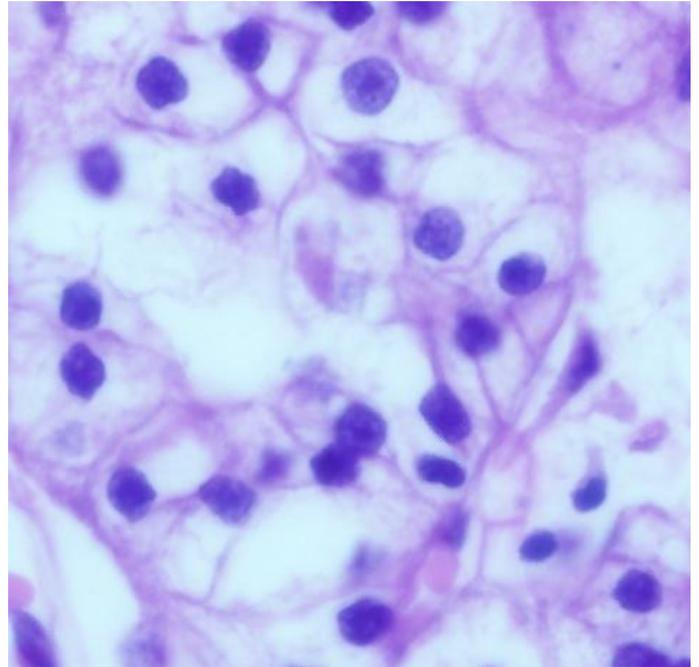

Figure 2. Original image (a) and segmented nuclei (b)

3.2 *The set of numerical descriptors*

All segmented nuclei are described by the numerical features used as the input matrix for classifier. The large set of features has been defined by authors and then feature selection process was applied to evaluate each feature and avoid these features without any value in classification task. The evaluation method was described in next subsection. The numerical fea-

tures used in our experiment are divided into the following groups:

- **Texture Haralick features:** The following 13 texture features described in (Huang 2010) have been generated.
  1. Energy
  2. Contrast
  3. Correlation
  4. Sum of variances
  5. Inverse difference moment
  6. Entropy
  7. Information measures of correlation
  8. Sum average
  9. Sum entropy
  10. Sum variance
  11. Difference variance
  12. Contrast
  13. Difference entropy
- **Morphometric features**
  14. Area
  15. Major axis length
  16. Minor axis legth
  17. Eccentricity
  18. Convex area
  19. Equivalent diameter: $ed = \sqrt{\frac{4*\text{Area}}{\text{pi}}}$
  20. Solidity: $\text{solidity} = \frac{\text{Area}}{\text{Convex area}}$
  21. Perimeter
- **Features describing colors:**
  22. Relative sum of pixels of red components: $Sr = \frac{R}{R+G+B}$
  23. Relative sum of pixels of green components: $Sg = \frac{G}{R+G+B}$
  24. Relative sum of pixels of blue components: $Sb = \frac{B}{R+G+B}$
- **Features from histogram:**
  25. Mean of histogram
  26. Standard deviation of histogram
  27. Kurtosis of histogram
- **Modified features:** Four texture features described in (Gianazza et al. 2010; Young IT et al. 1986) have been defined:
  28. Heterogenity
  29. Homogenity
  30. Clump
  31. Condensation

In this way 31 numerical features have been generated.

3.3 *Features selection*

Feature selection has been performed by applying Fisher measure (Duda et al. 2003):

$$Sf_{c1c2} = |mean(c1) - mean(c2)|/(\sigma_{c1} + \sigma_{c2})$$

where mean($c1$), mean($c2$) are the mean of feature f in the first and second class, and σ_{c1}, σ_{c2} are the standard deviations of the feature in class 1 and 2, respectively.

Fisher measure arranges the features from the most important (the highest value of measure) to the least important. To find their proper population we have to set appropriate value of threshold. On the basis of some introductory experiments we have found the best results corresponding to the value of threshold equal 1.

In the numerical experiments of classification we have applied the strategy one-versus-all (Scholkopf & Smola, 2002). In this approach we have to build 4 classifiers, each responsible for recognition of one Fuhrman scale (Fuhrman scale associated with a class).

Table 1 shows the sets of features selected for each classification system arranged in this way. The numbers in the table indicate the numbering of features described earlier. As we can see the population of selected features depends on the class and extends from 14 (class 1 versus all) to 30 (class 2 versus all).

Table 1. Selected features

Class 1 vs all	2	5	10	11	12	14	15	16	18	19	21
	24	25	26								
Class 2 vs all	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	
	23	24	25	26	27	28	29	30			
Class 3 vs all	1	2	3	4	5	6	8	9	10	11	12	13
	14	15	16	17	18	19	21	22	23	24	25	
	26	27	28	29	30						
Class 4 vs all	2	4	5	6	7	9	10	11	12	13	14
	15	16	17	18	19	20	21	22	23	24	
	25	26	27	28							

3.4 *Classification system*

As a classifier we have used Support Vector Machine of the Gaussian kernel (Scholkpf & Smola, 2002). To deal with a problem of many classes, we applied the one versus all approach. In this approach we train as many local two-class recognition classifiers as is the number of classes. On the basis of the results of all classifiers the final winner for each observation is found by applying the majority competition (Scholkopf & Smola, 2002). In learning the classifier systems we have used 3537 images of nuclei which were earlier separated by the segmentation algorithm and then annotated by human experts. The population of all classes is presented in Table 2.

Table 2. The population of cells of different classes taking part in numerical experiments

Class	1	2	3	4

Number of cells	1193	1167	808	369

The number of cells forming classes differ significantly, from 1193 (class 1) to 369 (class 4). This is due to the availability of cells representing these classes in the available data base of the Military Institute of Health Services.

4 RESULTS

The results of experiments will be given in the numerical and graphical forms. Fig. 3 presents the annotated cells corresponding to the image of Fig. 2. The recognized cells have been described by numbers, where number 1 referes to Fuhrman grade 1, number 2 to grade 2 and number 3 to grade 3.
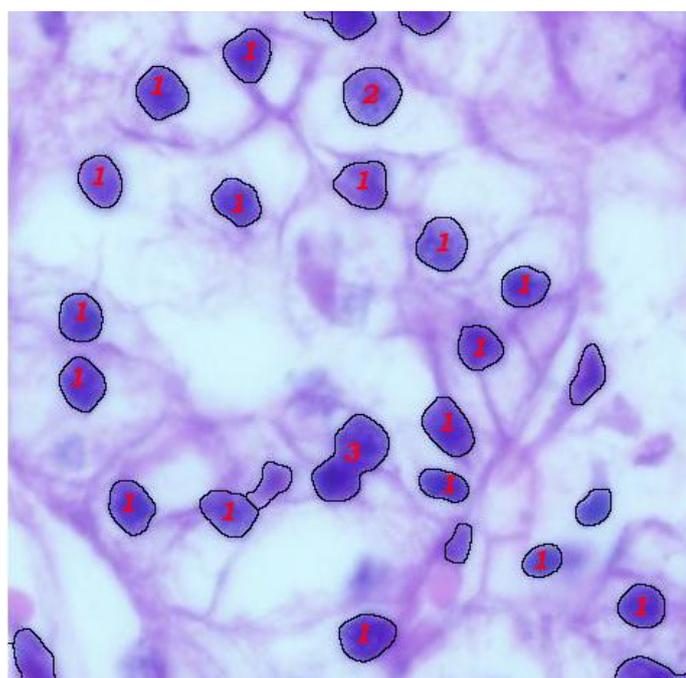

Figure 3. The graphical results of classification system: a)

The numerical results correspond to the accuracy, sensitivity and precision and compare the new values to the previous results presented in (Kruk et all. 2014). The confusion matrices of recognition of all Fuhrman grades of cells in the previous and our improved system is shown in Table 3 and 4. It is easy to observe that most mistakes are made between two adjacent classes. This problem is very hard to omit because the border between two classes is rather fuzzy.

Table 3. Confusion matrix of the Fuhrman grade recognition of the previous system

	Grade 1	Grade 2	Grade 3	Grade 4
Grade 1	0.967	0.025	0.008	0
Grade 2	0.035	0.942	0.015	0.008
Grade 3	0.006	0.048	0.916	0.030
Grade 4	0.010	0.048	0.099	0.843

Table 4. Confusion matrix of the Fuhrman grade recognition of the improved system

	Grade 1	Grade 2	Grade 3	Grade 4
Grade 1	0.978	0.015	0.007	0
Grade 2	0.030	0.951	0.011	0.008
Grade 3	0.005	0.041	0.922	0.032
Grade 4	0.010	0.038	0.091	0.861

It is easy to observe that the improved system has generated better results (higher values of diagonal terms and smaller values of off-diagonal elements). Similar improvement of sensitivity and precision is shown in Table 5.

Table 5. Comparison of class sensitivity and precision of the previous and improved versions of the Fuhrman grade recognition system

Fuhrman Grade	Grade 1	Grade 2	Grade 3	Grade 4
Previous sensitivity	96.7%	94.2%	91.6%	84.3%
Actual sensitivity	97.8%	95.1%	92.2%	86.1%
Previous precision	94.99%	88.01%	88.25%	95.69%
Actual precision	95.4%	89.9%	90.01%	96.21%

Slides of images representing all investigated cases were reevaluated independently by 3 experts blind to the previous scores. Each slide was assigned with respect to Fuhrman grade according to the opinion of the particular expert. The results of expert assessments were compared to the outcome of our computerized automatic system (AS). The comparison was done on the basis of Cohen's kappa methodology (Carletta, 1996). Kappa is widely accepted and used in the field of content analysis. It is interpretable, allows different results to be compared, and suggests a set of diagnostics in cases where the reliability results are not good enough for the required purpose. The results are presented in Table 5.

Table 5. Cohen Kappa values for the results of recognition of 4 Fuhrman grades

	AS	Expert 1	Expert 2	Expert 3
AS	-	0.4341	0.5242	0.4730
Expert 1	0.4341	-	0.3660	0.2499
Expert 2	0.5242	0.3660	-	0.2476
Expert 3	0.4730	0.2499	0.2476	-

In all cases the Kappa values were larger for our improved system than for the inter experts results.

5 CONCLUSION

The paper shows that improvement of segmentation algorithm significant has influence on results of classification system.

The important advantage of the improved system is its repeatability of scores, which is in great contrast with the human expert results, significantly dependent on the particular choice of the expert and his/her mental and physical condition in the time of assessment. Moreover, the our system allows a reduction of time required for the image analysis in comparison to the human expert. Thanks to this the acceleration of the research in this area is possible.

The main task for the future is to find the method to recognize the most important area in whole specimen. Now the main region of interest must be indicated by human expert to further processing.

6 REFERENCES

Carletta J. (1996) Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 22:249-254

Huang PW, Lai YH (2010) Effective segmentation classification for HCC biopsy images. J Pattern Recognition 43:1550-156

Duda RO, Hart PE, Stork P (2003): Pattern classification and scene analysis, Wiley, New York.

Fuhrman SA., Lasky LC., Limas C (1982), Prognostic significance of morphologic parameters in renal cell carcinoma. Am. J. Surg. Pathol. 6:655-663.

Gianazza E, Chinello C, Mainini V, Cazzaniga M, Squeo V, Albo G, Signorini S, Di Pierro SS, Ferrero S, Nicolardi S, van der Burgt YE, Deelder AM, Magni F (2012), Alternations of the serum peptidome in renal cell carcinoma discriminating Benin malignant Sidney tumors, J Proteom 76:125-140

Gonzalez R, Woods R, 2008: Digital Image Processing. New Jersey, Prentice Hall

Huang, M., & Haralick, R. M. (2010). Recognizing Patterns in Texts. Pattern Recognition and Machine Vision, 6, 19.

Jung, C. R., & Scharcanski, J. (2002). Robust watershed segmentation using the wavelet transform. In Computer Graphics and Image Processing,. Proceedings. XV Brazilian Symposium on (pp. 131-137). IEEE.

Kruk M, Osowski S, Markiewicz T, Słodkowska J, Koktysz R, Kozłowski W, Świderski B, (2014), Computer approach to recognition of Fuhrman grade of cells in clear-cell renal cell carcinoma. Analytical and Quantitative Cytology and Histolgy 36(3):147-160

Schölkopf, B., & Smola, A. J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press.

Young IT, Verbeek PW, Mayal BH (1986), Characterization of chromatin distribution in cell nuclei. Cytometry 7:467-474.