

# Computer Approach to Recognition of Fuhrman Grade of Cells in Clear-cell Renal Carcinoma

Michał Kruk, \_\_\_\_, Stanisław Osowski, \_\_\_\_, Tomasz Markiewicz, \_\_\_\_,  
Janina Słodkowska, \_\_\_\_, Robert Koktysz, \_\_\_\_, Wojciech Kozłowski, \_\_\_\_, and  
Bartosz Swiderski, \_\_\_\_

**OBJECTIVE:** To present a computerized system for recognition of Fuhrman grade of cells in clear-cell renal carcinoma on the basis of microscopic images of the neoplasm cells in application of hematoxylin and eosin staining.

**STUDY DESIGN:** The applied methods use combined gradient and mathematical morphology to obtain nuclei and classifiers in the form of support vector machine to estimate their Fuhrman grade. The starting point is a microscopic kidney image, which is subject to the advanced methods of preprocessing, leading finally to estimation of Fuhrman grade of cells and the whole analyzed image.

**RESULTS:** The results of the numerical experiments have shown that the proposed nuclei descriptors based on different principles of generation are well connected with the Fuhrman grade. These descriptors have been used as the diagnostic features forming the inputs to the classifier, which performs the final recognition of the cells. The average discrepancy rate between the score of our system and the human expert results, estimated on the basis of over 3,000 nuclei, is below 10%.

**CONCLUSION:** The obtained results have shown that the system is able to recognize 4 Fuhrman grades of the cells with high statistical accuracy and agreement with different expert scores. This result gives a good perspective to apply the system for supporting and accelerating the research of kidney cancer. (*Anal Quant Cytopathol Histopathol* 2014;36:000–000)

**Keywords:** cell classification, clear-cell renal carcinoma, Fuhrman grading, image processing, renal carcinoma cells.

A clear-cell renal carcinoma (CC-RCC) belongs to the most common renal malignancy in adults aged  $\geq 50$  years.<sup>1</sup> Renal cell carcinoma attacks males approximately twice as often as it does females, and is most commonly diagnosed in patients in their early 60s. 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 sep-

---

From Warsaw University of Life Sciences; Warsaw University of Technology; Military University of Technology, Warsaw; and Military Institute of Medicine, Warsaw.

\_\_\_\_ Kruk is \_\_\_\_\_, Warsaw University of Life Sciences.

\_\_\_\_ Osowski is \_\_\_\_\_, Warsaw University of Technology, and \_\_\_\_\_, Military University of Technology.

\_\_\_\_ Markiewicz is \_\_\_\_\_, Warsaw University of Technology, and \_\_\_\_\_, Military Institute of Medicine.

\_\_\_\_ Słodkowska is \_\_\_\_\_, Military Institute of Medicine, Warsaw.

\_\_\_\_ Koktysz is \_\_\_\_\_, Military Institute of Medicine, Warsaw.

\_\_\_\_ Kozłowski is \_\_\_\_\_, Military Institute of Medicine, Warsaw.

\_\_\_\_ Swiderski is \_\_\_\_\_, Warsaw University of Life Sciences.

Supported by the National Science Centre (Poland), grant #2012/07/B/ST7/01203 for 2013–2016.

*Financial Disclosure:* The authors have no connection to any companies or products mentioned in this article.

0884-6812/14/3600-0000/\$18.00/0 © Science Printers and Publishers, Inc.

*Analytical and Quantitative Cytopathology and Histopathology*®

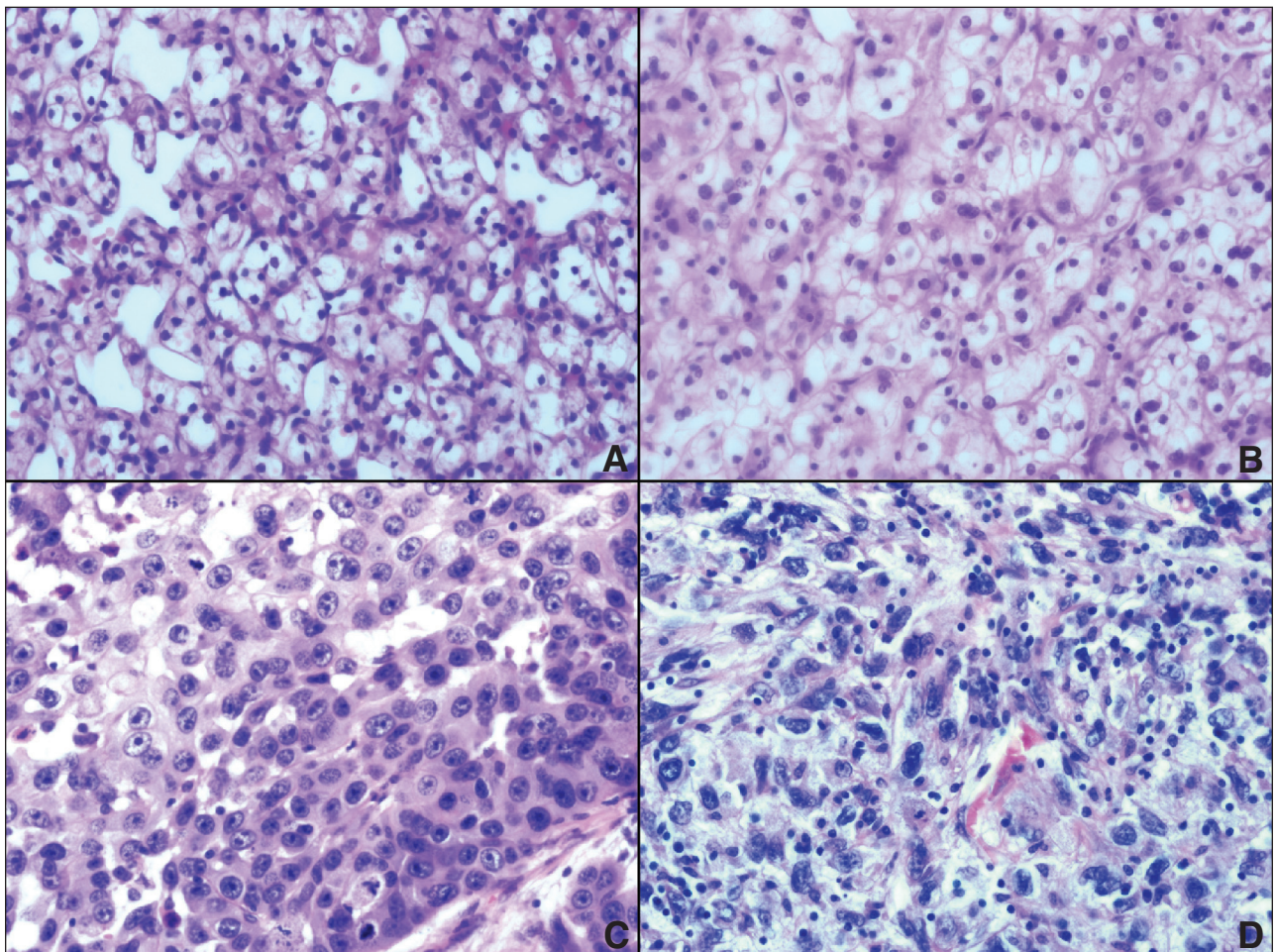
arated from each other by an extensive network of delicate sinusoidal vascular channels. The tumor cells are generally large and polygonal and show the distinct cell membranes. The cytoplasm is optically clear in most cells. The nuclei are small, round to slightly oval, and regular. They can be recognized on the basis of one or even a few nucleoli.

The degree of the nuclear atypia is used for grading these carcinomas. The most popular and widely used system for grading CC-RCC is a nuclear grading system described in 1982 by Fuhrman.<sup>2</sup> The Fuhrman grading scheme utilizes 4 nuclear features for assigning a specific grade: nuclear size, nuclear shape, chromatin pattern, and size of the nucleoli. It is defined on 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. Figure 1 presents typical examples of cells corresponding to these 4 different grades.

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



**Figure 1** The typical histological appearance of different grades of CC-RCC: (A) grade 1, (B) grade 2, (C) grade 3, and (D) grade 4 (H&E stain,  $\times 400$ ).

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 simplest method for assessing the Fuhrman grade is based on the observation of nucleoli at different magnifications. If nucleoli are visible at a magnification of 400 $\times$ , it indicates Fuhrman grade 1. If they are visible at a magnification of 200 $\times$ , this indicates Fuhrman grade 2, and so on.<sup>2,3</sup>

More advanced systems utilize a few nuclear features for assigning a specific grade to the cells: nuclear size, nuclear shape, chromatin pattern, and even number and size of nucleoli. However, for distinguishing between grade 1 and 2 and between grade 2 and 3, careful evaluation of each component of the grading scheme is needed. Interobserver variability exists due to the mostly subjective assignment of the nuclear grade. The difficulty in accurate nuclear grading arises due to the heterogeneity of nuclear features in the same tumor and due to small changes between the neighboring parts of the tumor.

A study by Perroud et al<sup>4</sup> shows the existence of clusters in the multidimensional space, associated with different grades of Fuhrman, confirming the possibility of an automatic grade recognition. Gianazza et al<sup>5</sup> have shown the application of proteomics analysis combined with the support vector machine and the genetic algorithm to recognizing different stages of CC-RCC, achieving a sensitivity of recognition of the malignant stage from the others in the range of 90%.

Bektas et al<sup>6</sup> have presented the results of a study on intraobserver and interobserver variability of the Fuhrman grading system. They show that the results of recognition change significantly when the same images have been assessed by two different experts (the mean interobserver variability for the four-tiered system evaluated using  $\kappa$ -statistics equals  $\kappa=0.416$ ). The intraobserver variability is also very high as determined by the same expert checking the same cases at two different times (the mean  $\kappa$ -value for four-tiered system equals 0.486). This demonstrates that there is an urgent need to develop a computerized system based on the samples assessed independently by many experts. Such a system will incorporate their knowledge and experience into the evaluation procedure.

However, up to now there are no published solutions directed to such an automatic process of recognition of the Fuhrman grade on the basis of the

computerized image analysis of slides representing CC-RCC. There are only some scarce works related to the application of artificial intelligence methods to the classification of cells in different pathological cases, e.g., hepatocellular carcinoma<sup>7</sup> or colon cells.<sup>8</sup> However, each type of pathology has its own specific nature and needs a special approach. The pilot study on the research of CC-RCC and a discussion of problems associated with this task was presented at the European Congress on Telepathology and has been published.<sup>9</sup>

The main task of this project is to develop an automatic computer system that would evaluate the Fuhrman grade of renal cells with the accuracy comparable to that of human experts. To solve the problem, we have applied the support vector machine (SVM) classifier based on the data representing over 100 patients. To counteract the problem of biasing the score, we have used 3 medical experts (authors of this paper). Each expert was responsible for 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 assignments were 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.

We herein present the computerized solution, which is able to perform this grading without intervention of the human operator. The solution includes (1) the application of image filtering techniques and mathematical morphology to extract the cells (nuclei) from the whole image, (2) generation and selection of diagnostic features, and (3) application of the classifier for the final estimation of the Fuhrman grade of CC-RCC cells.

### **Materials and Methods**

We used a database from the Military Institute of the Health Services that was created for research purposes and was prepared between the years 2005–2013. Seventy CC-RCC cases representing all 4 Fuhrman grades were evaluated and used in the learning phase of the system. In the verification of the established system we used an additional 62 patients of different grades of the illness and analyzed 94 randomly selected images representing those patients.

Digital images of H&E-stained kidney neoplasm cells of the appropriate tumor area were collected.



They were acquired at a magnification of  $400\times$  through an Olympus BX-61 microscope and taken with an Olympus DP-72 camera in RGB format at a resolution of  $2070\times 1548$ .

The appropriate fields representative for the tumor areas were selected by the experts. In the numerical experiments of the system learning we collected the cells from 105 histological slides which were assessed in the same way by all experts. The base of 3,446 microscopic images of nuclei extracted from these slides was created and annotated by the medical experts. Table I presents the population of cells belonging to all Fuhrman grades (classes).

### Algorithm for Estimation of Fuhrman Grading of Nuclei

To achieve the automatic grading system of cells we had to solve the following tasks: the segmentation of the individual cells (nuclei) from the image using the mathematical morphology and color-based operations, the preprocessing stages which lead to the extraction of the nuclei of carcinoma, the generation of features of the extracted cells, the selection of the most important features, and the final recognition stage relying on the classification of data. According to this description the algorithm is composed of the following stages: (1) reading the image and segmentation of the nuclei, (2) generation of the numerical descriptors treated as the features characterizing the nuclei, (3) evaluation of the most important class discriminative features, (4) application of the classifier for the determination of the Fuhrman grade of cells, and (5) saving the results to the data base.

#### Segmentation of Nuclei from Image

The first step of the image processing<sup>10</sup> (after reading it from the proper database) is the segmentation of the nuclei. The investigated base of the images was created from the microscopic photos taken in  $2,070\times 1,548$  resolution and saved in TIFF (tagged image file format). In further processing we considered only the red component of the RGB image. This choice was done based on the observation that the chromatin is best visible in this component and

at the same time the background (disturbing process) is the least visible. It is well demonstrated in Figure 2.

In these experiments 70 slides of the kidneys, each representing different patients, were used. Over 3,000 nuclei (cells) were segmented and saved. Among those cells all 4 grades of Fuhrman were present.

The first step is the smoothing and filtering of the image, i.e., removal of the small particles that disturb the image.<sup>9</sup> It is an important operation, allowing us to avoid a lot of unimportant details in the edge detection process by using the gradient method. This operation was implemented in 3 steps: (1) the histogram equalization enhancing the contrast of the image (transformation of the values of an intensity for better visibility of chromatin), (2) the morphological erosion at circular structural element of 5-pixel size (found in experiments to be optimal), and finally (3) the application of the Gaussian filter. In the last operation we applied the window size  $50\times 50$  and  $\sigma=4$ , found as optimal. After all these operations we attained a smooth image.

The next step is the edge detection—the crucial part of the nuclei segmentation.<sup>10,11</sup> The accuracy of this step affects the whole result of the segmentation. To counteract the different luminosities of parts of the image arising from H&E staining, we decided to apply the gradient method of the nuclei detection instead of the often-used automatic thresholding, e.g., the Otsu method.<sup>10,11</sup>

Representing the image intensity at the position  $(x,y)$  by  $I(x,y)$ , the discrete gradient components in  $x$  and  $y$  directions are described as follows<sup>9,11</sup>:

$$\nabla_x I(x,y) = \frac{I(x+h,y) - I(x-h,y)}{2h} \quad (1)$$

$$\nabla_y I(x,y) = \frac{I(x,y+h) - I(x,y-h)}{2h} \quad (2)$$

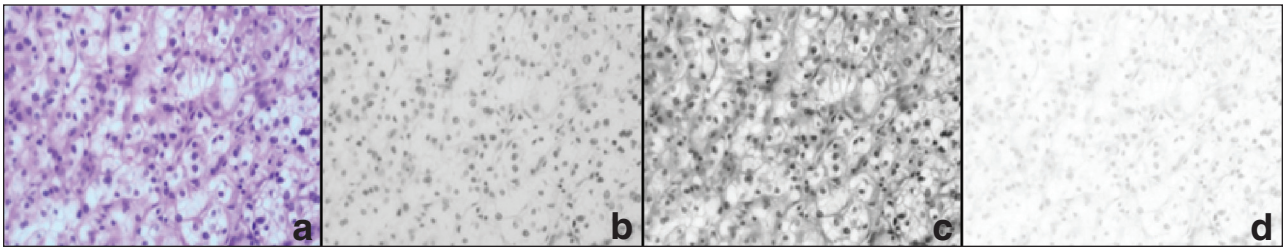
In our implementation we applied  $h=2$  pixels. On the basis of these two values we form one scalar discrimination function, expressed in the form.

$$F(x,y) = \left( \frac{1}{\nabla_x I(x,y)^2 + \nabla_y I(x,y)^2} \right) \quad (3)$$

The discrimination function of the defined above gradient was found to be significantly resistant to the noise of the background. This can be well illus-

**Table I** The Cell Population Included in the Experiments

Class	1	2	3	4
No. of cells	1,164	1,133	786	363



**Figure 2** The original RGB image (a) and its views at red (b), green (c) and blue (d) components.

trated on the exemplary subimage representing the small part of the nucleus as is shown in Figure 3.

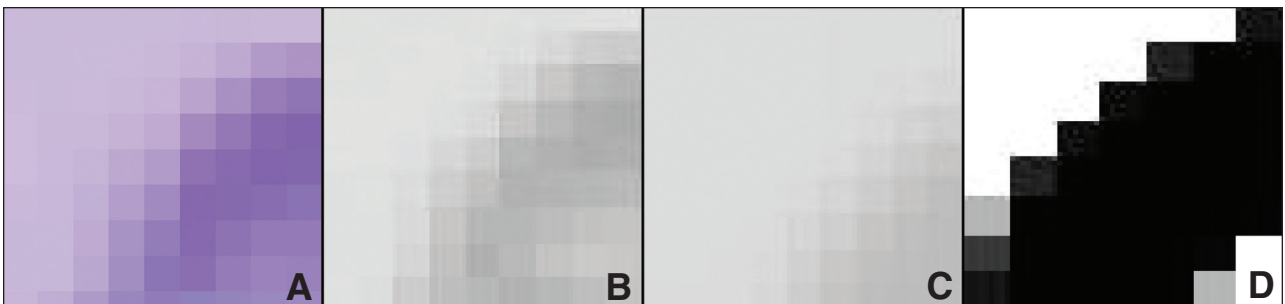
After performing the task of the edge detection, we prepare the binary mask, representing the nuclei discovered in the image. In this step few morphological operations are used: smoothing, filling the holes created in the nuclei by the gradient, and clearing the image by applying the morphological operations of closing and opening. Figure 4 illustrates the succeeding steps of this stage in the image processing.

After performing all of these preprocessing steps, the image may still contain some groups of glued nuclei that should be also separated. This was done by using the watershed algorithm.<sup>11,13</sup> The crucial point in this step is to choose the proper size of the window, responsible for finding the local minima. In our experiments we have found the size of 20 pixels as the most universal for the images under analysis. Figure 5 shows the exemplary result of the application of the watershed algorithm to the image of Figure 2a.

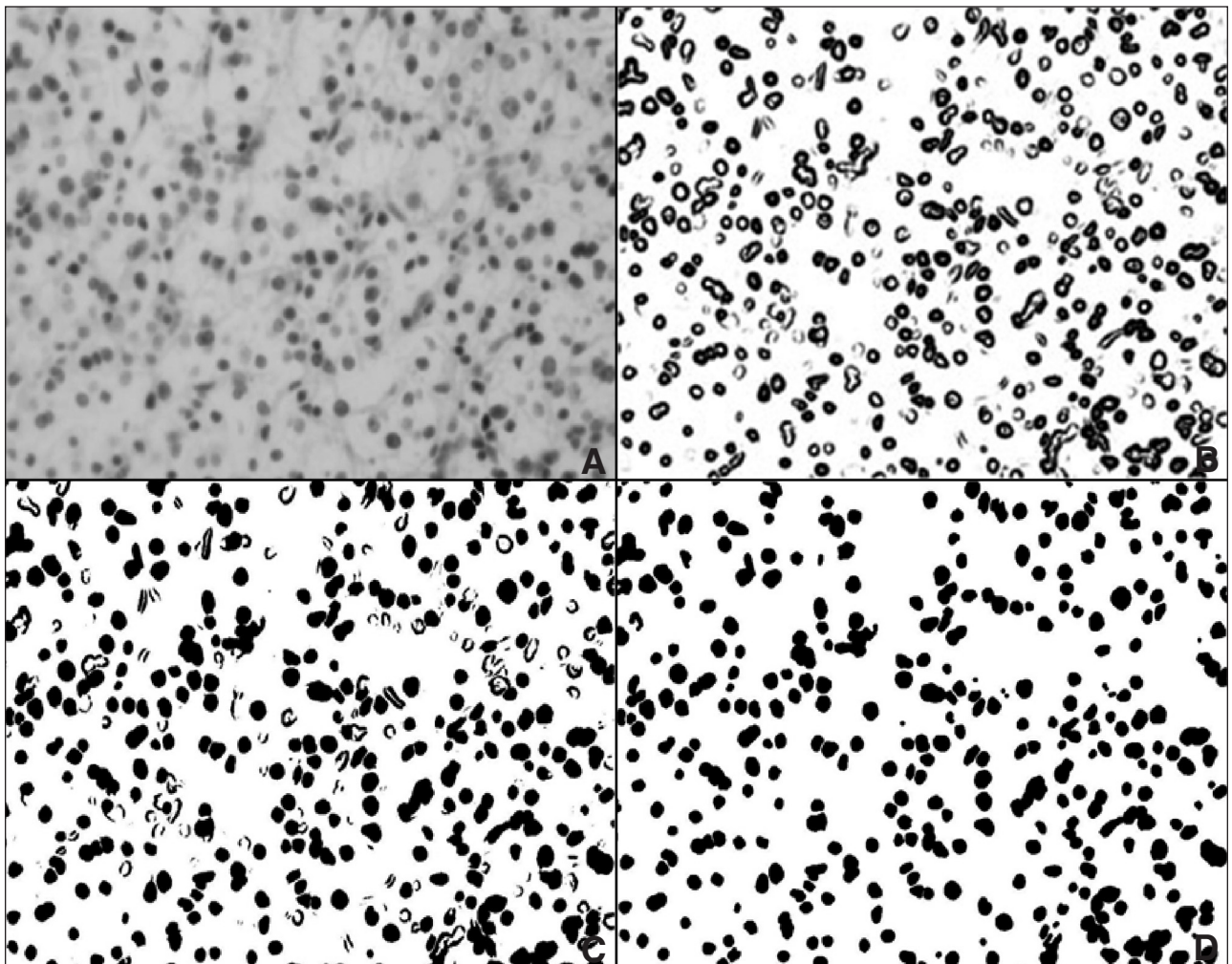
In an automatic nuclei localization the structures not representing the nuclei might be extracted also.

According to the medical experts the nuclei should have similar shape factors, which may differ slightly from the circular shape. All segmented structures must therefore undergo the test of circularity. We estimated the lengths of the major and minor axis for each structure. The extracted structure is accepted as the nucleus if the ratio between its major and minor axis does not exceed the value of 1.5. This value was found in experiments as the most optimal, leading to the best identification of nuclei in the slides.

After segmentation of the image, all nuclei are saved to the file as the separate objects for further processing. Examples of the automatically segmented nuclei corresponding to different Fuhrman grades are shown in Figure 6. They are depicted in such a way as to preserve the proportion of sizes as much as possible. It is easy to observe the variability among the cells corresponding to different Fuhrman grades. They differ by the size, texture, presence or absence of small nucleoli, the details of color, the homogeneity of the structure, etc. From the medical diagnostic point of view the most important issue is the recognition between the groups



**Figure 3** Illustration of the gradient approach to the edge detection. (A) Original subimage. (B–D) Transformed structure: (B) original image in red component, (C) image after smoothing, and (D) gradient for of image.



**Figure 4** Illustration of the following steps of the image processing: (A) image after smoothing, (B) gradient form, (C) image after filling the holes, (D) the binary mask of segmented nuclei after final morphological operations of closing and opening.

of nuclei belonging to the first 2 and the last 2 Fuhrman grades.

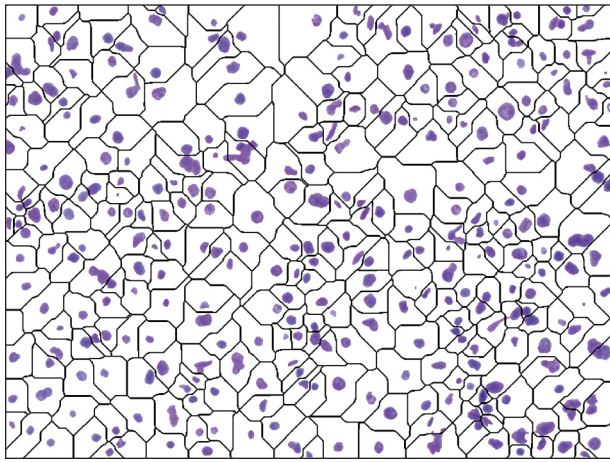
#### *Generation of Numerical Descriptors of Nuclei*

To build an automatic classification of the cells, we must generate the numerical descriptors well representing the nuclei. They have been created in such a way as to characterize the texture, the geometrical shape, and the color. The first group of descriptors are related to the texture.<sup>14,15</sup> We have applied the Haralick approach. The descriptors are based on the statistics related to gray-level co-occurrence matrices.<sup>14</sup> The co-occurrence matrices (the square matrices with dimension  $N$ , where  $N$  is the total number of the gray levels in the image) focus on the distri-

bution and the relationship among the gray levels of the neighboring pixels of the image. The  $[i,j]$ th element of this matrix is produced by counting the total occurrences of the pixel with the value  $i$  adjacent to the pixel with the value  $j$  and then subsequently dividing the whole matrix by the total number of such comparisons that are made. The adjacency is defined to take place in horizontal, vertical, left and right diagonal directions, in a two-dimensional square pixel image, and as a result we can calculate 4 of this kind of matrices. The texture features are then created on the basis of averaging all of these 4 co-occurrence matrices.

From these averaged matrices many Haralick texture descriptors of the statistical nature can be





**Figure 5** Illustration of the final step of the watershed algorithm application.

evaluated. In this application we have limited our consideration to the following<sup>14</sup>:

1. energy
2. contrast
3. correlation



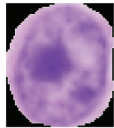



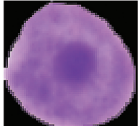




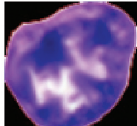


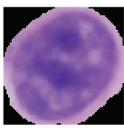

4. sum of variances
5. inverse difference moment
6. entropy
7. information measures of correlation 1 and 2
8. sum average
9. sum entropy
10. sum variance
11. difference variance
12. contrast
13. difference entropy.

The additional set of descriptors has been generated on the basis of 4 texture features defined in the paper.<sup>16</sup> They describe the distribution of the nuclear chromatin while considering the heterogeneity, granularity, condensation, and margination of the chromatin in the cell nuclei. The following 4 descriptors have been defined using these measures:

14. heterogeneity
15. margination
16. clump
17. condensation.

The next set of descriptors is related to the geometrical parameters of the nuclei. They include the following<sup>12</sup>:

18. area

Fuhrman 1	Fuhrman 2	Fuhrman 3	Fuhrman 4
			
			
			
			

**Figure 6** The examples of the segmented nuclei of different Fuhrman grades.

19. major axis length
20. minor axis length
21. eccentricity
22. convex area
23. equivalent diameter defined by

$$ed = \sqrt{\frac{4 \cdot \text{area}}{\pi}} \quad (4)$$

24. solidity described by the expression

$$\text{solidarity} = \frac{\text{area}}{\text{convex area}} \quad (5)$$

25. perimeter.

The color descriptors were defined in a relative form as the ratio of the sum of intensities of the particular color (R, G, B) to the sum of the intensities of all colors. They include:

26. ratio of the pixel intensities of the red (R) component

$$S_R = \frac{R}{R+G+B} \quad (6)$$

27. ratio of the pixel intensities of the green (G) component

$$S_G = \frac{G}{R+G+B} \quad (7)$$

28. ratio of the pixel intensities of the blue (B) component

$$S_B = \frac{B}{R+G+B} \quad (8)$$

The last set of descriptors is based on the histogram of the image. These descriptors represent:

29. mean of histogram
30. standard deviation of histogram
31. kurtosis of histogram.

In this way 31 numerical descriptors, candidates for the diagnostic features, have been generated for all nuclei. Among them there were more or less class discriminative features. The natural step is therefore evaluation of their importance as the diagnostic features at the recognition of the classes. To select the most important we have applied the forward and the backward stepwise regression called stepwisefit.<sup>11</sup>

It is a systematic method for adding and removing terms (each representing the column of the input data represented by the matrix  $\mathbf{X}$ ) from the multilinear model on the basis of their statistical significance in the regression. The procedure begins with an initial model and then compares its clas-

sification power of the incrementally larger and smaller models. At each step of the processing the p value of F-statistics is estimated to test the models with and without the potential term. If the term is not currently in the model, the null hypothesis is that the term would have a zero coefficient after adding it to the model. If there is a sufficient evidence to reject the null hypothesis, the term is added to the model. On the other side, if a term is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is an insufficient evidence to reject the null hypothesis, the term is removed from the model. The process terminates when no single step improves the model. The method is locally optimal since different initial models or a different sequence of steps may lead to different solutions.

In practical implementation of the method we used a small part of the data (around 20% of the total number of samples representing all Fuhrman grades) for finding the optimal values of parameters *penter* and *remove* controlling the stepwisefit procedure. We apply the results of this procedure in the classification, accepting this set of parameters, which leads to the smallest misclassification error.

#### Support Vector Machine Classification System

As the classifier we have used the SVM of the Gaussian kernel.<sup>16</sup> It is a feedforward network of one hidden layer (the kernel function layer) known for excellent generalization ability. In contrast to other neural solutions the learning problem of SVM is formulated as the task of the separation of learning vectors into two classes denoted by either  $d_i = 1$  (one class) or  $d_i = -1$  (the opposite class), with the maximal separation margin. This margin is formed in the learning stage according to the assumed value of the regularization constant  $C$  providing some robustness of the classifier to the noise.

The great advantage of the SVM is the unique formulation of the learning problem leading to the quadratic program with linear constraints, which is easy to solve. The SVM of the Gaussian kernel has been used in our application. The hyperparameters  $\sigma$  of the Gaussian function and the regularization constant  $C$  have been adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one on the validation data sets. The optimal values of these parameters were  $\gamma = 0.3$  and  $C = 10,000$ . The SVM classifier needs application of the carefully selected set of input signals forming the diagnostic features, hence



it cooperates strictly with a stepwise regression presented in the previous section.

To deal with a problem of many classes, we applied the one against one approach.<sup>18</sup> In this approach we train many local two-class recognition classifiers (all two-class combinations), on the basis of which the final winner is selected. At  $M$  classes we have to train  $M(M-1)/2$  two-class SVM networks. In the simplest case, the majority voting of all trained classifiers is applied to find the final winner at the processing of the actual input vector  $x$ . In our case, at the existence of 4 classes we have to train 6 two-class SVM networks, and the majority dictates the class membership of the analyzed sample.

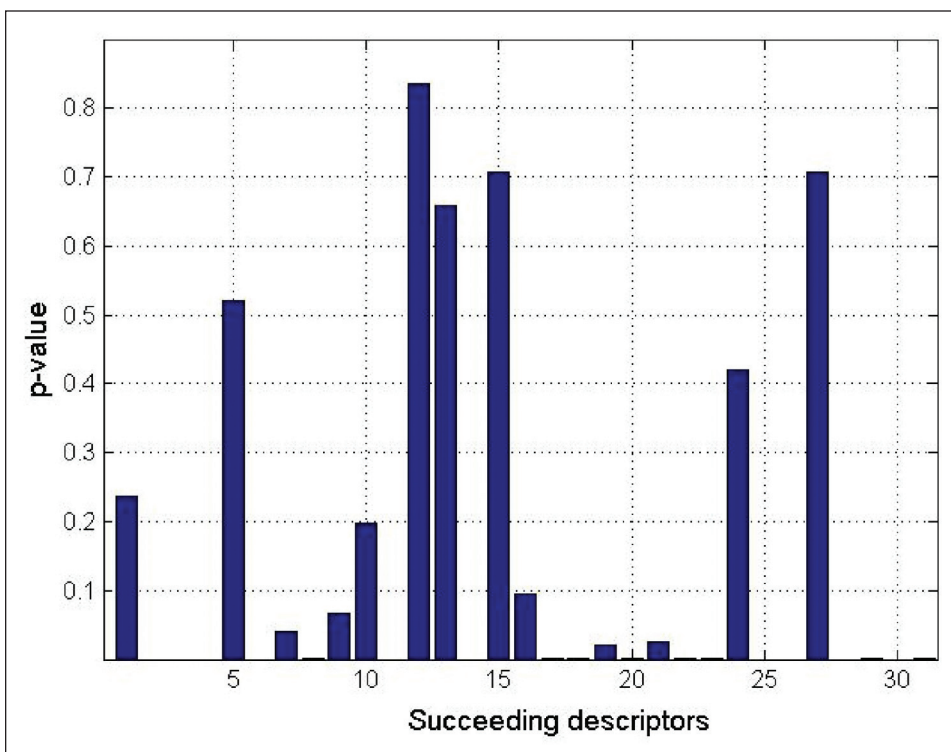
### Results of Numerical Experiments of Cell Recognition

In the numerical experiments of cell recognition we have used the base of several thousand images of the nuclei, each annotated by 3 experts responsible for the same set of images. In further experiments we have used only the cells of the class memberships that were the same in the assessment of all experts. Their population was 3,446. Then each image

of the nuclei was preprocessed to create the numerical descriptors according to the given definitions. As a result each cell was represented by the set of 31 descriptors.

The first step is to assess the discriminative ability of the potential features and choose only these of the highest discriminative power. The multilinear stepwise regression built in the form of the stepwise fit function of Matlab<sup>12</sup> was used. Since it is of the local optimality, we have repeated it many times at different values of *penalty* and *remove* parameters, using 20% of the available data. The results of the feature selection in each run were applied to the SVM classifier with the Gaussian kernel. As an optimal we regarded the set leads to the best classification results. It was found by assuming the control parameter's equal: *penalty* = 0.06 and *remove* = 0.10.

The estimated p values for testing whether the particular feature is important or not in the recognition are presented in Figure 7. A small estimated p value corresponding to the particular descriptor is evidence of its significance in the recognition process. At the values of the parameter *penalty* = 0.06 and *remove* = 0.10 (found as the best), we identified



**Figure 7**  
The estimated p values of the succeeding descriptors in the stepwise fit procedure.

8 descriptors (1, 5, 10, 12, 13, 15, 24 and 27) that exceeded the allowed level of the p value. They have not been included in the optimal set of features used in the recognition. The best selected set of features was formed by the descriptors of the following numbers: 2, 3, 4, 6, 7, 8, 9, 11, 14, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 28, 29, 30, 31. We can observe that the selected features belong to all families of descriptors (texture, geometry, color and histogram). This result proves that the assumed wide spectrum of descriptors generated on the basis of different principles was the right direction of research.

It is interesting to compare the selected set of features with their Fisher discriminative values (individually for each feature  $f$  at recognition of  $i$ th from  $j$ th class) expressed by the formula

$$S_{ij}(f) = \frac{|E(f_i) - E(f_j)|}{\sigma(f_i) + \sigma(f_j)}, \quad (9)$$

where  $E(f_i)$  and  $\sigma(f_i)$  represent the expected value

and standard deviation, respectively, of the feature  $f$  in  $i$ th class. Table II presents the values of this measure for each feature at all combinations of classes. The last column of the table presents the average value of these discriminative measures over all combinations of classes.

Bold notations denote the features that have been found by the stepwise regression as the least significant in the recognition process. As we can see, the application of the Fisher measures to the individual features has produced a bit different sequence of features. According to the Fisher measure some features (for example 6, 26, 28) have low discrimination ability. However, they have been included in the set of the best features by the stepwise regression. It confirms the known fact that a feature not having the best individual discrimination ability may perform a meaningful role in recognition while cooperating with others.

After selection of the best set, we started the

**Table II** Values of the Fisher Measures of the Individual Descriptors at the Recognition of All Combinations of Classes

Descriptor	$S_{12}$	$S_{13}$	$S_{14}$	$S_{23}$	$S_{24}$	$S_{34}$	Average( $S_{ij}$ )
1	1.4485	0.2934	0.4985	1.0483	0.4784	0.2975	<b>0.6618</b>
2	2.598	2.5782	2.6053	0.3625	0.5449	0.2506	1.4439
3	0.5855	1.2401	0.2568	1.6512	0.5848	0.4554	0.7793
4	1.8640	1.0078	0.9729	0.4738	0.2328	0.0807	0.7630
5	0.1809	0.2529	0.2226	0.0018	0.0246	0.0136	<b>0.0891</b>
6	0.9037	0.5347	0.4492	0.2665	0.2465	0.0278	0.3791
7	2.5702	1.3740	1.2943	0.9676	0.4469	0.2717	1.2218
8	1.3405	0.2179	0.4243	1.0124	0.4622	0.2842	0.6029
9	1.9265	0.7389	0.8979	1.0607	0.4316	0.2369	0.8797
10	0.8431	0.7349	0.6388	0.0612	0.0230	0.0536	<b>0.3918</b>
11	3.4951	2.8560	2.5298	0.4690	0.0263	0.2420	1.6024
12	0.7912	1.0704	0.9008	0.3066	0.2749	0.0568	<b>0.5676</b>
13	0.1988	0.2948	0.2682	0.4728	0.2692	0.0416	<b>0.2417</b>
14	0.5147	0.8957	0.9472	0.4029	0.4948	0.2860	0.5729
15	1.5180	0.0145	0.2513	1.4239	0.7353	0.2880	<b>0.6878</b>
16	2.1764	0.0209	0.3610	2.0457	1.0564	0.2700	0.9877
17	0.9286	1.0408	0.7652	1.8752	1.2362	0.0241	0.9777
18	2.3543	5.3372	6.2648	3.0204	3.8134	1.4817	3.6946
19	1.4795	3.2479	3.3696	1.6723	1.9467	0.6830	2.0552
20	1.7538	2.9551	2.9392	1.2369	1.4107	0.4808	1.7998
21	0.6757	0.3035	0.6042	0.9187	0.9384	0.2659	0.6171
22	2.3515	5.3264	6.2900	3.0220	3.9289	1.5772	3.7490
23	1.6592	3.0528	3.2165	1.4488	1.6414	0.5541	1.9088
24	0.0134	0.0150	0.0530	0.0114	0.0433	0.0297	<b>0.0270</b>
25	1.6745	3.2754	3.3920	1.5542	1.8682	0.6840	2.0567
26	0.2583	0.7604	0.4840	0.5291	0.2714	0.0713	0.3928
27	1.1353	0.0662	0.4947	1.0980	0.2724	0.3907	<b>0.5742</b>
28	0.6959	0.4781	0.5630	0.2743	0.0508	0.2405	0.3488
29	2.7911	5.5818	6.8892	2.8509	4.2597	1.8660	4.0225
30	3.4573	4.9093	5.8892	1.5808	2.8874	1.5194	3.3746
31	0.8901	1.2974	1.6032	2.0897	1.9312	0.4351	1.3758

Bold numbers denote the features that have been found by the stepwise regression as the least significant in the recognition process.

experiments of testing the presented solution. The remaining 80% of the database was split into 10 approximately equal parts containing randomly selected samples. We applied the tenfold cross validation procedure in which 9 parts of data were used in learning and the last one was used in testing. In each run the testing part was exchanged and their results of testing added to the others. The results of the computerized system have been compared to the actual expert assessments of the grade. The final results in the form of a confusion matrix are depicted in Table III. The rows represent the percentage of the real class memberships (the agreed expert scores), and the columns represent the actually recognized classes. Each of the off-diagonal elements of the matrix mean misclassification.

The average of the diagonal terms of the confusion matrix represents the estimated weighted accuracy of the cell recognition system. Its value is equal to 93.3%. There were 192 misclassifications out of 2,866 samples. This means that the total average percentage error of the cell recognition was equal to 6.7%. Observe that the most misclassifications happened between the neighboring classes. There are only single misclassifications between the extreme classes. On the basis of the results presented in Table III, we can determine the other important measures of the quality of our system in the form of the sensitivity and the class precision. We can separately determine their values for the recognition of each class (the Fuhrman grade) with respect to the rest. These values for each grade are depicted in Table IV.

The lowest value of sensitivity corresponds to the fourth grade. The main reason for this is the smallest number of representatives of this class in the database used in the experiments. This has resulted in some biasing of the applied classifiers.

In medical practice quite often there is a need to recognize only two classes in the introductory assessment of CC-RCC (class 1 = the first two grades, and class 2 = grade 3 and 4 taken together). If we an-

**Table III** Confusion Matrix of the Recognition of the Cell Fuhrman Grade

	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 IV** Class Sensitivity and Precision of the Method in Determination of All Fuhrman Grades

Fuhrman grade	1	2	3	4
Sensitivity	96.7%	94.2%	91.6%	84.3%
Precision	94.99%	88.01%	88.25%	95.69%

alyze such an arrangement of classes, the confusion matrix is as shown in Table V.

In this case the relative misclassification error of the cells' recognition is equal to only 2.35%. The sensitivity for determination of class 1 is now equal to 98.9%, and 96.5% for class 2. The respective values of the class precisions are 96.58% and 98.87%.

### Results of Recognition of Grade of Tumor Cases

The results of the analyzed nuclei are used to draw the final conclusion regarding the total Fuhrman grade for the investigated tumor case, representing the particular patient. In summary, the following procedure of the whole analysis is applied:

- Extract all nuclei existing in the slide image of the particular patient and apply the proposed procedure leading to their classification to one of four Fuhrman grades.
- Determine the total number of cells of the image in each class of the Fuhrman grade of all analyzed images of the considered case and their proportion to the total number of all cells.
- The final Fuhrman grade of the investigated tumor case is the highest grade that is represented by >10% of cells in the consecutively analyzed images.

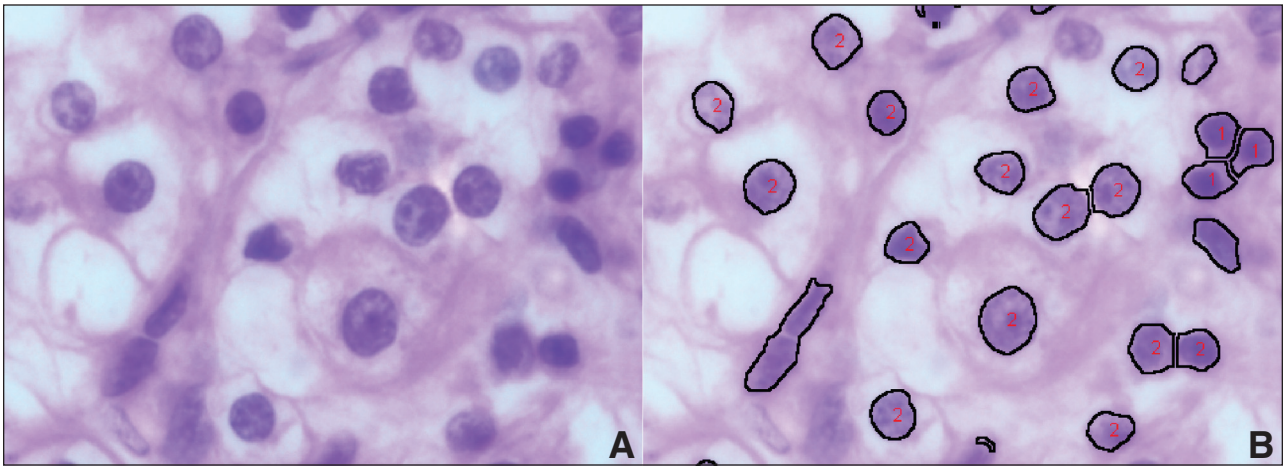
The exemplary result of such a procedure for one image of CC-RCC is illustrated in Figure 8. It shows the enlarged fragment of the image, which was assigned by the human expert to grade 2. In this example 20 objects were segmented. Fifteen of them were recognized by our algorithm as grade 2, three as grade 1, and two (two structures of oblong shape) were not classified as the nuclei. The image as a whole should therefore be assigned to grade 2.

The important problem in the computerized

**Table V** Confusion Matrix for the Recognition of 2 Classes of the Fuhrman Grades (2-Tiered System)

	Grade 1 + 2	Grade 3 + 4
Grade 1 + 2	0.989	0.011
Grade 3 + 4	0.035	0.965





**Figure 8** The enlarged fragment of (A) the analyzed original image of CC-RCC and (B) the class-annotated cells (H&E stain, magnification 1,500 $\times$ ).

recognition of Fuhrman grade is the lack of a gold standard that can be used in learning. As was shown in the paper by Bektas,<sup>6</sup> the interobserver variability in the assessment of Fuhrman grade on the basis of the image analysis is quite large. In our solution the training of the automatic system relied on three experts who assessed the common set of images in the same way. Thanks to this the system is able to incorporate the diagnostic knowledge from different sources, reducing in this way the interobserver variability.

To check the performance of our system in the recognition of the Fuhrman grade for the cases representing different patients, we performed additional experiments using the earlier trained system. We used a newly acquired set of images that did not take part in the learning process. It was composed of 94 images representing 62 patients of different Fuhrman grade. The mean patient age was 62.58, with a standard deviation of 12.69 (range, 32–81 years). Forty-five percent of patients were male and the rest were female. Slides of all cases were rescored independently by 3 experts blinded to the prior scores. Each slide was assigned to a 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.<sup>19</sup> We compared our results with all three experts and also the results among the experts. They are presented in Table VI in recognition of all 4 grades.

We observed the satisfied agreement between the results of our system and the experts' results. They were better than the level of agreement among the experts. The average value of kappa for our system was 0.4508, while for the three experts it was only 0.2878. It is evident that our system has learned the proper knowledge from the experience of different experts.

It is interesting to compare the interobserver agreement of the individual classes related to the case of best kappa of our automatic system (AS and expert 2) and the best kappa for two experts (expert 1 and 2). In the first case we got 64.8% (grade 1), 80.6% (grade 2), 76.9% (grade 3), and 50.0% (grade 4). The appropriate values for two experts (1 and 2) of the best kappa were 71.4% (grade 1), 60.0% (grade 2), 40.7% (grade 3), and 25.0% (grade 4). The small values of agreement seen at grade 4 is a result of a very small number of such cases among the analyzed data (in all cases their estimated number was not higher than 5).

We also compared the grade estimation results in

**Table VI** Cohen Kappa Values for the Results of Recognition of 4 Fuhrman Grades (4-Tiered System)

	AS	Expert 1	Expert 2	Expert 3
AS	—	0.4246	0.5049	0.4230
Expert 1	0.4246	—	0.3660	0.2499
Expert 2	0.5049	0.3660	—	0.2476
Expert 3	0.4230	0.2499	0.2476	—

the application of only a two-tiered system. In this case the grades 1, 2 and 3, 4 were combined together, forming two cumulated classes 1+2 as class 1 and 3+4 as class 2 (a two-tiered system). The results in the form of the Cohen kappa values for this case are shown in Table VII.

This time we also got good agreement of our results and the experts' scores. The average value of kappa for our system is 0.5689. The mean value for the three experts was only 0.4240.

### Conclusion

This paper has presented a new approach to the assessment of the Fuhrman grade of CC-RCC cells. The starting point of the analysis is the microscopic image of neoplasm cells of the kidney, stained with H&E. The first step is an automatic extraction of the nuclei existing in the image. This problem was solved by applying the gradient method to the nuclei detection in combination with the morphological operations and the watershed algorithm.

In the next phase the images are converted to the numerical descriptors, forming the potential features that may be used as the input signals in the final classification stage. We generated them on the basis of the texture, geometry, histogram, and color representation. In this way 31 potential features were created. In the following phase their class discrimination ability was checked by applying the multilinear stepwise fit regression. According to the results of that stage, 23 important diagnostic features were selected.

In the last step the selected diagnostic features were used in the class recognition of the individual cells applying the SVM. The final stage of checking of the developed system's efficiency was performed in the tenfold cross validation approach using the database of >3,000 nuclei. The obtained results have shown that the system is able to recognize 4 Fuhrman grades of the cells with an average accuracy of 93.3%.

**Table VII** Cohen Kappa Values for the Results of Recognition of Two Cumulated Fuhrman Grades (Two-tiered System)

	AS	Expert 1	Expert 2	Expert 3
AS	—	0.6299	0.6284	0.4585
Expert 1	0.6299	—	0.5015	0.4310
Expert 2	0.6284	0.5015	—	0.3394
Expert 3	0.4584	0.4310	0.3394	—

The advantage of machine learning is the automatic inclusion of some "statistical" approach to the potentially contradicting verdicts of experts. In this way different points of view represented by the experts have been taken into account at the final decision of the computer. The probability of making the correct decision has been increased in this way. This conjecture was confirmed by the results of the experiments performed on the additional set of data, in which we compared the agreement of the outcome of our system and the assessment results of 3 independent experts. This comparison shows that the Cohen's kappa index of our system is better than the kappa values among the independent experts.

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

### References

- Bostwick DG, Cheng L: Urologic Surgical Pathology. Philadelphia, Mosby Elsevier, 2008, pp 00-00
- Fuhrman SA, Lasky LC, Limas C: Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol* 1982;6:655-663
- Kontak JA, Campbell SC: Prognostic factors in renal cell carcinoma. *Med Biol Engineer Comput* 2003;30:467-480
- Perroud B, Ishimaru T, Borowsky AD, Weiss RH: Grade-dependent proteomics characterization of kidney cancer. *Mol Cell Proteomics* 2009;8:971-985
- 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: Alterations of the serum peptidome in renal cell carcinoma discriminating benign malignant kidney tumors. *J Proteom* 2012;76:125-140
- Bektas S, Bahadir B, Kandemir NO, Barut F, Gul AE, Ozdamar SO: Intraobserver and interobserver variability of Fuhrman and modified Fuhrman grading system for conventional renal cell carcinoma. *Kaohsiung J Med Sci* 2009;25: 596-600
- Huang PW, Lai YH: Effective segmentation classification for HCC biopsy images. *J Pattern Recognition* 2010;43:1550-1563
- Kruk M, Osowski S, Koktysz R: Recognition and classification of colon cells applying the ensemble of classifiers. *Comput Biol Med* 2009;39:156-165
- Kruk M, Osowski S, Kozłowski W, Koktysz R, Markiewicz T, Słodkowska J: Computer-assisted Fuhrman grading system for the analysis of clear-cell renal carcinoma: A pilot study. *Przełąd Elektrotechniczny* 2013;89:268-271

10. Gonzalez R, Woods R: Digital Image Processing. New Jersey, Prentice Hall, 2008, pp 00-00
11. Soille P: Morphological Image Analysis: Principles and Applications. Second edition. Berlin, Springer, 2003, pp 00-00
12. Matlab user manual – Image processing toolbox. Natick, MathWorks, 2012
13. Vincent L, Soille P: Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transaction on Pattern Analysis and Machine Intelligence 1991;13: 583-598
14. Tae-Yun K, Hyun-Ju C, Soon-Joo C, Heung-Kook C: Study on texture analysis of renal cell carcinoma nuclei based on the Fuhrman grading system. Enterprise networking and Computing in Healthcare Industry, 2005 HEALTHCOM 2005 Proceedings of 7th International Workshop, 2005, pp 384-387
15. Wagner T: Texture analysis. In Handbook of Computer Vision Application. Edited by B Jahne, H Haussecker, P Geisser. Boston, Academic Press, 1999, pp 275-309
16. Young IT, Verbeek PW, Mayal BH: Characterization of chromatin distribution in cell nuclei. Cytometry 1986;7:467-474
17. Scholkopf B, Smola A: Learning with Kernels. Cambridge, MIT Press, 2002, pp 000-000
18. Hsu CW, Lin CJ: A comparison methods for multiclass support vector machines. IEEE Trans Neural Networks 2002;13: 415-425
19. Carletta J: Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 1996;22:249-254