

## 방광암종의 등급 결정을 위한 컴퓨터를 이용한 분류기 생성에 관한 연구

최현주 · 윤혜경<sup>1</sup> · 최흥국

인제대학교 정보컴퓨터공학부  
의료영상연구실, <sup>1</sup>인제의대 부산백병원  
해부병리과

접 수 : 2002년 3월 4일  
게재승인 : 2002년 6월 4일

책임저자 : 최 현 주  
우 621-749 경남 김해시 어방동 607  
인제대학교 정보컴퓨터공학부 의료영상연구실  
전화: 055-320-3643  
Fax: 055-322-3107  
E-mail: hjchoi@mitl.inje.ac.kr

\*본 연구는 보건복지부 보건의료기술연구  
개발사업(과제번호 00-PJ1-PG3-20600-  
0025) 지원으로 이루어졌음.

### Study on Creating A Classifier for Grading of Bladder Carcinoma Based on Computerized Method

Hyun Ju Choi, Hye Kyoung Yoon<sup>1</sup> and Heung Kook Choi

Medical Image Technology Laboratory, School of Information and Computer Engineering, Inje University, Gimhae; <sup>1</sup>Department of Pathology, Busan Paik Hospital, College of Medicine, Inje University, Busan, Korea

**Background :** We have described an objective and reproducible classification method for grading malignancy in the Feulgen stained bladder carcinoma. To create an optimized classifier for malignancy grading of histological bladder carcinoma cell images, it is necessary to extract the features that accurately describe the order/disorder of the nuclear variation and to evaluate the significance of the features. Above all, features selection considered about the correlation of features is very important, because the performance of the classification method depends on the selected features. **Methods :** First, we acquired 40 representative histological bladder carcinoma cell images from each of four groups (Grade 1, Grade 2A, Grade 2B, Grade 3) and extracted morphology features, texture features and the texture features of wavelet transformed images. Second, we evaluated the significance of the extracted features using variance analysis. Third, we created classifiers for each selected feature and its combination set using discriminant analysis. Finally, we compared and analyzed the correct classification rate of each classifier. **Results :** The optimized classifier was created from the combination of morphology features, texture features and the texture features of wavelet transformed images. **Conclusions :** We found that the correlation of features is more important than one feature's great significance in grading the malignancy of bladder carcinoma, and we have confirmed that the correct classification rate is determined by feature extraction, feature evaluation and feature selection.

**Key Words :** Bladder Neoplasms-Classification-Multivariate Analysis

의료 영상 분석은 의료 영상이 가지고 있는 정보 즉, 질병에 대한 정보를 객관적 수치로 표현함으로써 측정치에 대한 객관성과 재현성을 부여하고 통계학적 해석을 가능하게 하므로, 질병의 유무나 진행 정도를 파악하여 환자를 진단하고 치료하는 데 있어서 매우 중요한 요소이다.<sup>1</sup> 특히, 암세포 조직 영상에서의 세포핵은 암에 대한 정보를 가지고 있으므로, 이러한 정보를 추출하고 분석하여 분류된 결과는 암의 진행 정도와 환자의 생존율을 예측할 수 있는 하나의 기준으로 사용할 수 있기 때문에 암의 진단과 예후 추정에 유용하다.

하지만 현재 암세포 조직 영상에 대한 분석은 전체 조직 중 일부분만을 선택하여 분석하는 표본 선택의 한계성과 영상 해석

의 주관성으로 인해 어려움을 나타내고 있다. 이러한 어려움의 원인은 첫째는 정성적인 평가로 인한 낮은 재현성이며, 둘째는 세포 조직 영상에 적합하지 않은 일반 영상 분석 방법의 적용으로 인한 분석 결과의 낮은 정확성과 전문가 의견과의 비일치성이다. 이에 보다 객관적이고 재현성이 높은 분석 방법 개발에 대한 연구가 필요한 실정이다.

일반적으로, 암세포 조직 영상 분석은 영상 획득, 전처리, 분할, 특성 추출, 분류기 생성, 결과 분석의 단계를 거쳐 이루어진다.<sup>2,3</sup> 이러한 과정 중에서 가장 중요한 단계는 특성 추출과 분류기 생성에 사용할 특성 선택 단계이다. 그 이유는 전문가의 판단 근거와 추출하고 분석하고자 하는 특성과의 일치성에 대한

문제 즉, 암의 진행 정도를 가장 잘 나타낼 수 있는 유효한 특성 추출이 분류 정확도를 결정하는 필수적 요소이기 때문이다. 그리고 특성은 분별력 면에서 다른 특성과 구별되는 각각의 고유한 특성을 갖는데, 예를 들면 어떤 특성은 특정 부류를 구별하는 데 강한 반면, 다른 부류에 대해서는 비교적 약한 경우가 있다. 그러므로 이러한 특성들의 유의성을 평가하고, 이 분석 결과를 특성 선택에 응용하여, 특성 벡터의 차원을 줄이는 것은 최적의 분류기 생성을 위한 필수적 선행 과제이다. 따라서 암의 진행 정도를 판단하기 위한 최적의 분류기를 생성하기 위해서는, 암의 진행 정도를 판단하는 데 있어서 중요한 요인인 세포핵의 변화에 대한 규칙성과 불규칙성을 가장 잘 나타낼 수 있는 특성을 추출해야 하며, 추출한 특성들에 대한 유의성 평가로 특성들간의 상관관계까지 고려한 특성 벡터를 구성하여 분류기를 생성하여야 한다.

이런 점을 고려할 때, 방광암 세포 조직 영상에서 세포핵의 크기, 모양 및 방향성, 인접 세포와의 거리, 형태 변이 정도에 대한 규칙성과 불규칙성은 암의 진행 및 전이성을 판단하는 데 있어서 중요한 요인이므로,<sup>2,3</sup> 세포핵의 기하학적 특성을 분석하는 형태학적 특성 분석과 염색질 양상의 내부적 구조 변화를 분석할 수 있는 질감 특성 분석이 필요하다. 그러나 지금까지의 질감 분석 방법은 한 가지 해상도에서 분석하는 단점을 가지고 있으므로 다해상도 표현이 가능한 웨이블릿 변환을 적용하면 한 영상을 여러 가지 해상도에서 분석할 수 있다.<sup>4</sup> 그리고 웨이블릿 변환은 공간 영역, 시간 영역에서 표현된 영상을 주파수 영역으로 변환하는 방법 중의 한 가지이므로, 주파수 영역으로의 변환이 가지는 장점을 이용할 수 있다. 즉, 하나의 파형으로 생각하면 분석하기 힘든 임의의 신호를 주파수 영역으로 변환하면, 그 신호를 구성하고 있는 여러 개의 신호로 분류할 수 있으므로, 각각의 신호를 분석하면 원 신호의 특성을 알 수 있다.

본 논문에서 사용하는 Feulgen으로 염색된 방광암 세포조직 영상에 대한 관련 연구로서는 세포핵 DNA의 IOD (Integrated Optical Density) 밝기값에 관한 연구인 T-분류 및 요법에 대한 연구가 있으며,<sup>5</sup> 또한 전체적인 등급에 관한 연구가 있다.<sup>2</sup> 본 연구는 이를 바탕으로 새로운 방법론을 적용하여 암의 진행 정도를 판단하는 데 최적의 분류기를 생성하기 위한 연구로, 형태학적 특성, 질감 특성 그리고 웨이블릿 변환을 적용한 후 분할된 영역에서 질감 특성을 추출하고 추출된 특성들의 유의성을 평가하여 낮은 부류 분별력을 가진 특성은 분별력이 약한 것이므로 제거하고, 높은 부류 분별력을 가진 것을 선택하여 좋은 특성들의 부분집합을 만든 후 각각의 특성, 또는 특성들의 조합에 대하여 분류기를 생성하고, 마지막으로 실험을 통하여 생성된 각각의 분류기에 대한 분류 정확도를 비교 분석함으로써, 병리학자가 결정한 주관적 등급과 컴퓨터 영상처리 및 분석으로 결정한 등급 사이의 상호관계를 고려하여 최적의 분류기를 찾고자 한다.

## 재료와 방법

### 영상 획득

본 연구에서 사용된 실험 영상은 1975년부터 1988년까지 스웨덴 옘살라 대학병원에서 등급 1 (28명), 등급 2A (62명), 등급 2B (44명), 등급 3 (63명)의 방광암 환자에 대한 영상으로부터 각 등급에서 대표적 영상들을 선택하여 회고 연구한 것이다. 물론 WHO (World Health Organization)에서 방광의 유두 모양 종양에 대해서 3등급을 택하고 있지만,<sup>6</sup> 여기에서는 Bergkvist의 4단계의 등급을 사용하였다.<sup>7</sup> 영상획득 과정은 먼저 파라핀에 포매한 방광암 조직을 5  $\mu\text{m}$ 의 두께로 박절하고 5 M HCl을 사용하여 60분 동안 22°C에서 가수분해시켜 유리된 알데히드기를 Schiff시약으로 22°C에서 120분 동안 반응시켜 Feulgen 염색을 시행하였으며,<sup>8</sup> 조직 영상은 400배 배율에서 현미경에 부착된 색영상 카메라를 통해 해상도 128×128인 8비트 그레이 영상으로 디지털화하였다. 분류하고자 하는 4개(등급 1, 등급 2A, 등급 2B, 등급 3)의 그룹에서 각각 10개씩, 총 40개 영상을 획득하였으며(Fig. 1), 그 중 20개는 학습집단으로, 나머지 20개는 생성된 분류기의 분류정확도를 측정하기 위한 시험집단으로 사용하였다. 이 연구에서는 본 연구실(Medical Image Technology Laboratory)에서 개발한 IMAN (Image Analyzer)을 영상 처리 및 분석기로 사용하였다.

### 특성 추출

#### 전처리

색영상 카메라를 통하여 영상을 획득할 때 정보 손실이 발생하므로 오류를 보정하는 전처리과정이 필요하다. 시료가 없는 상태에서 광원에서부터 조사된 빛으로 만든 영상의 화소는 이론적으로 동일한 밝기값을 가져야 하나 실체는 그렇지 못하다. 그것은 조사광의 변동, 광축에서부터 화소까지의 거리 등 광학 기기의 물리적 성질상 발생하는 오차 때문이다. 그 결과 영상 분석 과정에서 얻고자 하는 밝기값의 식별이 장애를 받게 되는데, 이를 shading 현상이라고 한다. 이러한 shading 현상을 보정하기 위하여 시료가 없는 상태의 화면에서 각 화소의 밝기값에 따른 보정치를 산출하였다가 시료 영상의 각 화소의 밝기값에 이 보정치를 더하거나 곱하여 줌으로써 시료의 모든 화소가 일정한 광조사를 받은 것처럼 교정시켜 주었다.

#### 형태학적 특성 추출

암세포 조직 영상에서 세포핵은 어둡고 배경은 밝게 나타나기 때문에 히스토그램을 분석하여 문턱치를 정한 뒤 문턱치 이하의 화소는 1의 값을 주고, 문턱치 이상의 화소는 0의 값을 주어 이진 영상을 생성하였다. 일정한 문턱치 이하의 화소만을 선택하여 밝기값이 높은 배경과 구분할 수 있었으며, 이진 영상으로

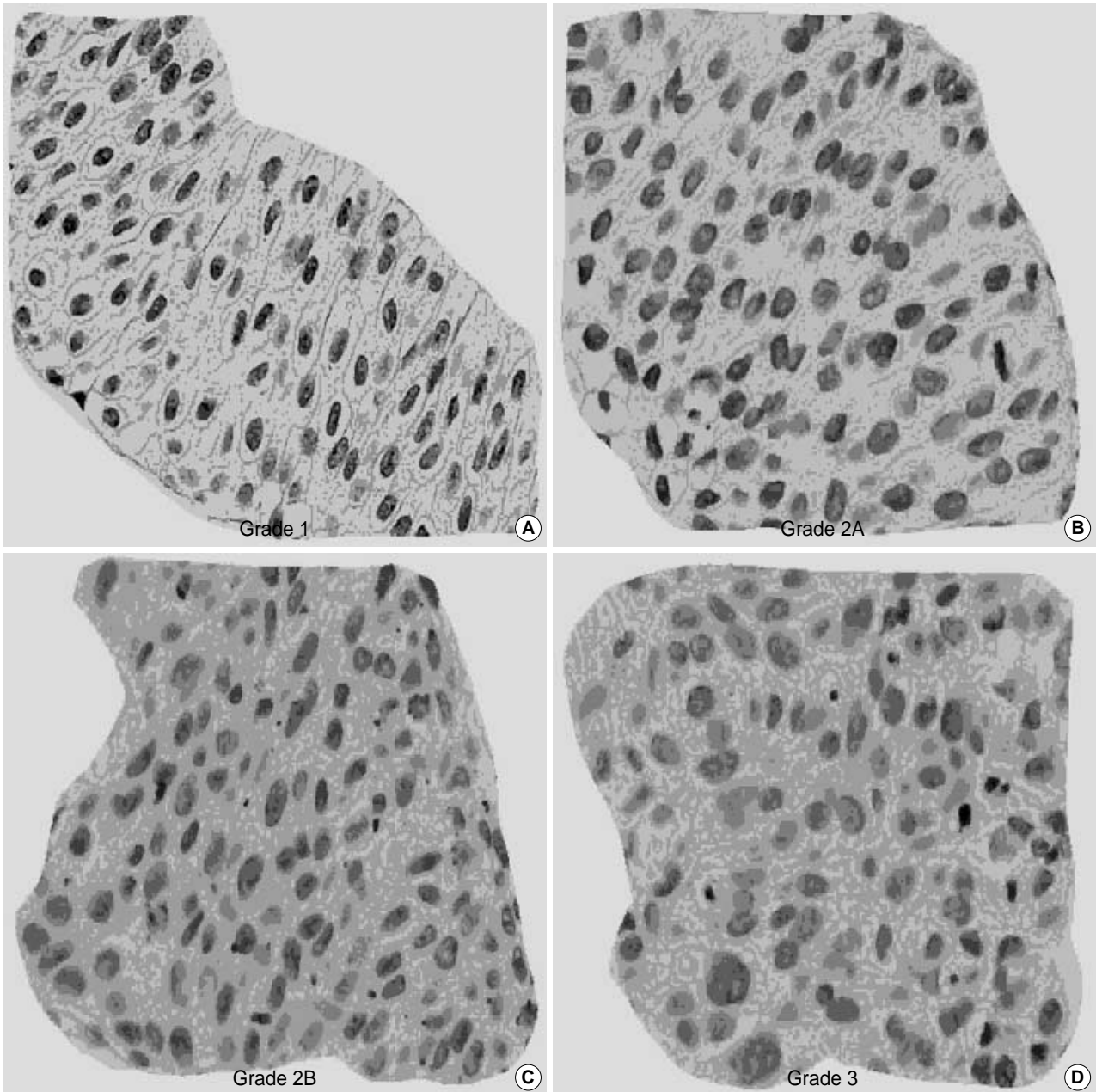


Fig. 1. Digital microscope images of the bladder carcinoma cell stained by Feulgen. In each grade, the region of interest was chosen by pathologists: (A) Grade 1 (upper left), (B) Grade 2A (upper right), (C) Grade 2B (lower left) and (D) Grade 3 (lower right).

변환함으로써 수학적 형태학적 연산 적용이 가능하도록 할 수 있었다. 문턱치의 설정은 반복적 영상 처리 작업의 경험에 바탕을 두고 이루어졌다. 생성된 이진 영상은 잡음과 분할된 세포핵 영역 내의 염색의 정도에 따라 구멍이 생길 수 있으므로, 침식과 팽창 그리고 구멍을 메우는 연산을 반복적으로 수행하여 잡음과 구멍을 제거하였다. 이렇게 생성된 영상에 8-connectivity를 바탕으로 이미지 부호화를 한 후 각각의 객체에 대하여 형태학적 특성을 추출하였다. 추출한 형태학적 특성은 객체의 면적

과 둘레, 그리고 객체가 원형에 얼마나 가까운가를 표현하는 특성인 원형성이다. 추출된 형태학적 특성은 영상을 구성하고 있는 각각의 객체에 대한 특성이므로, 각각의 영상을 구성하고 있는 객체의 수로 나눈 특성의 평균과 표준편차, 분산계수를 각 영상에 대한 형태학적 특성으로 사용하였다. 표준편차는 분산의 절대적 측정치이며 분산계수는 상대적 측정치로서 표준편차를 평균으로 나눈 값이다.

$$Mean = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j)$$

$$SD = \sqrt{\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (f(i, j) - Mean)^2}$$

$$CV = \frac{SD}{Mean} \times 100$$

**질감 특성 추출**

질감 특성 추출 방법으로는 세포 조직 영상이 규칙적인 무늬의 연속으로 표현할 수 있는 기하학적인 규칙성을 가지고 있지 않고 거의 불규칙적인 양상을 보이고 있어 통계적인 모델에 의해 그 특성이 가장 잘 나타날 수 있기 때문에 통계학적 방법을 사용하였다.<sup>9</sup>

먼저, joint probability density function  $f(i, j; d, \theta)$ 를 통해 GLCM (Gray Level Co-occurrence Matrix)를 생성하였다. 이 때 두 화소간의 거리  $d$  값은 1에 대해서 실험하였고, 두 화소 간의 각 값은  $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 에 대해 계산한 후 모두 더한 다음 4로 나누어 평균값에 대한 GLCM으로 실험하였다. 또한 GLCM을 원영상으로부터 바로 계산할 경우, 원영상의 256 gray level에 대한 모든 두 화소간의 밝기값의 분포를 계산하여야 함에 따라 GLCM의 크기가 너무 크고 많은 계산량이 요구되는 단점이 있으므로, 본 실험에서는 이 문제를 해결하기 위하여 히스토그램 평활화를 적용하고 256 gray level를 16 gray level로 축소하여 16x16 GLCM을 생성하여 질감 특성을 추출하였다. 사용된 질감 특성은 Haralick과 Pressman에 의해 정의된 질감 특성으로 angular second moment, entropy, different moment, inverse different moment, correlation, max probability로 그 식은 다음과 같다.<sup>10,11</sup>

아래 식에서  $PM$ 은 GLCM이며  $N$ 은 GLCM의 크기,  $i, j$ 는 임의의 두 화소의 밝기값,  $\mu_x, \mu_y$ 는 행과 열에 대한 평균이며  $V_x, V_y$ 는 행과 열에 대한 분산이다.

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} PM^2$$

$$Entropy = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (PM) \log(PM)$$

$$DM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 PM$$

$$IDM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (1+(i-j)^2)^{-1} PM$$

$$Correlation = \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (ij) PM - \mu_x \mu_y \right) / \sqrt{V_x V_y}$$

$$Max\ probability = \max(PM)$$

**웨이블렛 변환 영역에서 질감특성 추출**

웨이블렛 변환은 최근 들어 신호 처리나 패턴 인식 분야에서 가장 많이 쓰이고 있는 변환 방법 중 하나이다. 웨이블렛 변환 방법은 공간 영역과 주파수 영역에서 그 지역적 특징을 잘 나타내고 다해상도 분석이 가능하므로, 서로 다른 해상도로 영상을 분할하여 개별적으로 부영상을 분석할 수 있다는 특징으로 인해 질감 정보를 얻고자 할 때 많이 쓰이고 있다. 이 특성은 웨이블렛이라고 불리는 기저함수를 이동하고 확장함으로써 저대역 부밴드와 고대역 부밴드로 주파수 영역에 따른 다해상도를 갖게 된다. 웨이블렛 기저함수는 아래 식과 같은데, 여기서  $a$ 는 척도 변수이고  $b$ 는 이동 변수를 나타낸다.

$$CWT_x^a(a, b) = \frac{1}{\sqrt{a}} \int x(t) \psi\left(\frac{t-b}{a}\right) dt$$

2차원 영상의 웨이블렛 변환은 필터 뱅크 구조를 구성하여 수평 방향과 수직 방향으로 1차원 웨이블렛 변환을 반복적으로 수행해서 얻어지는데, L과 H는 각각 저주파 대역 통과 거르개와 고주파 대역 통과 거르개를 이용하는 것임을 의미한다.

본 연구에서는 2단계 웨이블렛 변환을 적용한 후,<sup>4</sup> 분할된 하위 영역 중 LH1, HL1, HH1, LH2, HL2, HH2 하위 영상에서 질감 특성을 추출하였다(Fig. 2). LL 하위 영상은 저해상도 영상으로 원래 영상을 흐려지게 함으로써 생성되는 영상인 반면, 나머지 하위 영상은 원영상의 흐려짐 처리로 인해 손실되는 정보를 성분별로 나타내는 영상이므로, LL 하위 영상을 제외한 나머지 하위 영상으로부터 질감 특성을 추출하였다.<sup>12</sup> 추출된 질감 특성은 angular second moment, different moment, inverse different moment이다.

**분류 방법**

**특성에 대한 유의성 평가**

추출된 특성에 대한 분석을 하는 이유는 암의 진행 정도를 가장 잘 나타내는 특성을 선택하기 위한 것이며, 각 군간에 차이를 의미 있게 설명해 줄 수 있는 특성을 발견하고 군 판별을 위해 어떤 특성들이 가장 높은 기여를 나타내는지를 평가하기 위한 것이다. 추출된 특성에 대한 유의성을 평가하기 위하여 통계 패키지인 SAS (Statistical Analysis System, 8.01, SAS Institute Inc, Cary, NC, USA)을 이용하여 분산분석을 하였다(Table 1-3).<sup>13</sup> 분산분석은 특성치의 변화에 영향을 주는 원인을 찾기 위한 방법으로 특성치의 변동을 요인들에 의한 변동으로 분해하여 오차에 의한 변동보다 큰 영향을 주는 요인이 무엇인지를 살펴서 요인의 유의성을 알아보는 방법이다.<sup>14</sup> 즉, 집단을 구분해 주는 독립변수의 차이에 따른 종속변수의 변화(집단간 분산)와 동일한 독립변수값을 갖는 집단 내에서의 종속변수의 변화(집단내 분산)를 비교하여 통계적으로 유의한 차이가 나타나는지를 검증한다. 분산분석 후 검증된 특성이 통계적으로

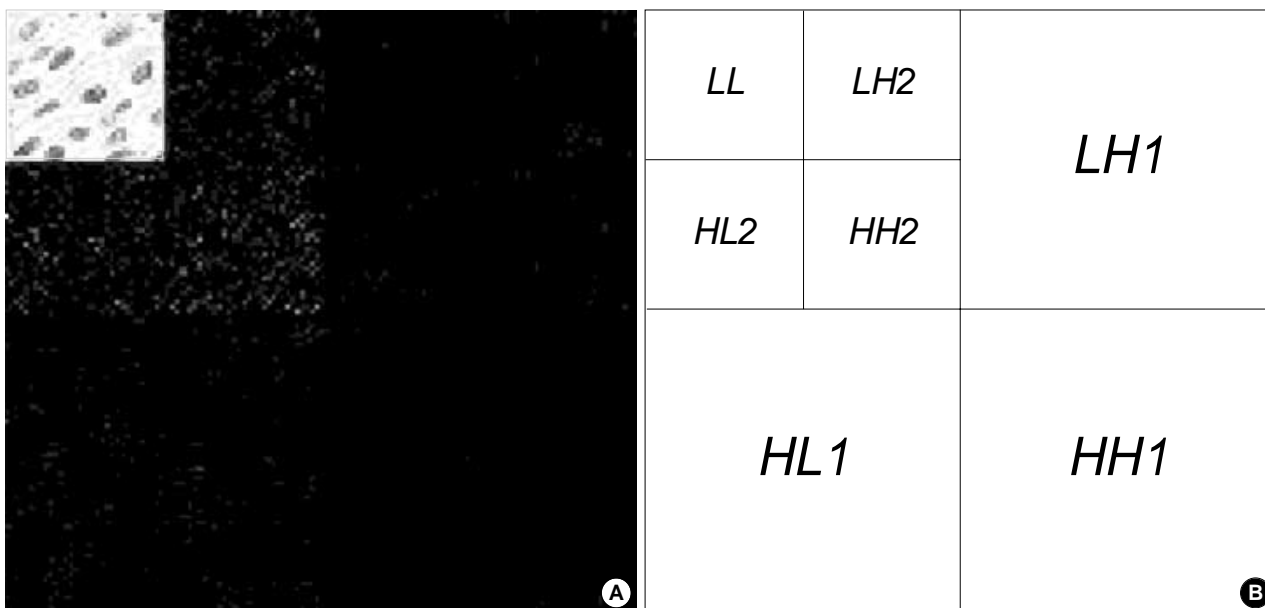


Fig. 2. (A) An example of wavelet transform of a grade 1 image (depth 2), (B) Location of frequency bands in four-band wavelet transformed image. LL (lowpass and lowpass frequency band) represents the low resolution subimage, LH (lowpass and highpass frequency band) represents vertical component of the image, HL (highpass and lowpass frequency band) represents horizontal component of the image and HH (highpass and highpass frequency band) represents diagonal component of the image.

Table 1. The result of ANOVA (analysis of variance) for morphology features

Feature	F-value	P-value
Area-Mean	6.2322	0.0052
Area-SD	3.4857	0.0405
Area-CV	6.2208	0.0053
Perimeter-Mean	5.5086	0.0068
Perimeter-SD	5.7268	0.0074
Perimeter-CV	6.8243	0.0036
Shape-Mean	7.2731	0.0027
Shape-SD	1.3264	0.3006
Shape-CV	1.2089	0.3385

CV: coefficient variance.

Table 2. The result of ANOVA (analysis of variance) for texture features

Feature	F-value	P-value
ASM	1.4222	0.2730
Entropy	23.8211	0.0001
Different Moment	67.3827	0.0001
IDM	73.8296	0.0001
Correlation	3.7105	0.0336
Max Probability	24.4447	0.0001

ASM: angular second moment, IDM: inverse different moment.

유의한지에 대한 평가는 F-통계량(F-value)과 유의확률값(P-value)을 보고 판단한다. F-통계량은 집단간 분산/집단내 분산

Table 3. The result of ANOVA (analysis of variance) for texture features of wavelet transformed images

Band	Feature	F-value	P-value
LH1	ASM	82.8361	0.0001
	DM	82.7446	0.0001
	IDM	108.0646	0.0001
HL2	ASM	120.9602	0.0001
	DM	121.0660	0.0001
	IDM	200.3672	0.0001
HH1	ASM	107.0952	0.0001
	DM	107.0952	0.0001
	IDM	108.9383	0.0001
LH2	ASM	22.8090	0.0001
	DM	41.5194	0.0001
	IDM	40.7377	0.0001
HL2	ASM	22.9354	0.0001
	DM	33.8581	0.0001
	IDM	33.0297	0.0001
HH2	ASM	128.3619	0.0001
	DM	128.3619	0.0001
	IDM	143.6410	0.0001

ASM: angular second moment, DM: different moment, IDM: inverse different moment, LH1: lowpass and highpass (depth 1) band, HL1: highpass and lowpass (depth 1) band, HH1: highpass and highpass (depth 1) band, LH2: lowpass and highpass (depth 2) band, HL2: highpass and lowpass (depth 2) band, HH2: highpass and highpass (depth 2) band.

의 비율로서, 집단간 분산이 크고 집단내 분산이 작다는 것은 그 특성이 집단간의 차이를 잘 나타낸다는 것을 의미하므로 F-

통계량이 클수록 집단간 판별력이 높다고 할 수 있다. 반면에 유의확률값은 데이터를 기초로 귀무가설이 기각될 유의수준을 역산한 확률값이므로 적을수록 통계적 유의성이 높다고 할 수 있다.

### 분류기 생성

전체적으로 각 특성에 대한 F-통계량과 유의확률값을 비교해 보면 웨이블릿 변환 영역에서의 질감 특성이 형태학적 특성이나 질감 특성보다는 더 유의한 특성이라 할 수 있다. 그러나 형태학적 특성 9개, 질감 특성 6개, 웨이블릿 변환 영역에서의 질감 특성 18개에 대한 다변량분산분석의 결과인 F-통계량을 비교해 보면 각각 4.6231, 14.6676, 5.1453으로 질감 특성이 더 유의한 특성이라 할 수 있다. 그러므로 어떤 특성을 선택하고 조합하여 분류기를 생성하느냐에 따라 분류 정확도가 결정된다는 사실을 알 수 있다.

이에 따라 분산 분석의 결과를 바탕으로 유효한 특성을 선택 하기 위하여, 먼저 각각의 특성에서 sequential backward selection 방법을 사용하여 F-통계량이 낮은 특성부터 제거하면서 판별분석에 의해 분류기를 생성하였다. 판별분석에 의해 생성된 분류기란 선택된 특성들의 선형결합으로 이루어진 판별식으로, 분류하고자 하는 각 대상들의 특성을 대입해서 각 대상들이 속하는 집단을 찾아내는 판별 함수이다.<sup>13</sup> 생성된 분류기가 집단간의 차이를 잘 반영하는지에 대한 검증과 특성 선택을 위한 평가는 시험집단에 대한 판별 함수의 오분류율, 학습집단에 대한 오분류율, 그리고 다변량 분산 분석의 결과인 F-통계량을 보고 판단하였다. 단변량분산분석은 단 하나의 종속변수 상에서 집단 평균들 사이의 차이를 검증하는 통계적 기법이지만, 다변량분산분석은 다수의 종속변수를 동시에 다룸으로써 종속변수들 사이의 상관관계까지 고려하는 통계적 기법이므로, 선택된 특성들에 대해 다변량 분산분석을 하여 그들 사이의 상관관계까지 고려한 통계적 유의성을 검증하였다. 형태학적 특성, 질감 특성, 웨이블릿 변환 영역에서의 질감 특성 각각에 의하여 생성된 분류기에 대한 평가 결과를 기반으로, 가장 높은 분류정확도를 가지는 특성만을 선택한 후, 3가지 특성 모두에 대한 조합으로 이루어진 집합에서 동일한 방법을 사용하여 분류기를 생성하고 평가하였다.

## 결 과

### 분석 결과

먼저 형태학적 특성, 질감 특성, 웨이블릿 변환 영역에서의 질감 특성 각각의 특성들만으로 이루어진 분류기에 대한 분류정확도를 평가하였으며, 형태학적 특성, 질감 특성, 웨이블릿 변환 영역에서의 질감 특성 각각에 의하여 생성된 분류기에 대한 평

가 결과를 기반으로, 가장 높은 분류정확도를 가지는 특성들만을 선택한 후, 세 가지 특성 모두에 대한 조합으로 이루어진 특성들의 집합에서 분류기를 생성하고 평가하였다. 다음은 각 실험에 의해 생성된 분류기에 대한 평가 결과이며, 각 실험에서 최종적으로 선택된 4개의 분류기의 분류정확도를 비교하여 방광암 세포 조직 영상 분류를 위한 최적의 분류기를 획득하였다.

### 형태학적 특성에 의해 생성된 분류기에 대한 평가

추출한 9개의 형태학적 특성에 대하여 sequential backward selection 방법을 이용하여 F-통계량이 낮은 특성을 하나씩 제거하면서 실험하여 학습집단과 시험집단에 대한 오분류율과 다변량 분산 분석의 결과인 F-통계량을 비교하였다.

평가 기준에 대한 우선순위는 시험집단에 대한 오분류율을 1순위, 학습집단에 대한 오분류율을 2순위, MANOVA에 의한 F-통계량을 3순위로 하여 평가하였다. 이를 바탕으로 평가할 때 시험집단에 대한 최소 오분류율은 45%, 학습집단에 대한 최소 오분류율은 5%로, shape factor의 분산계수를 제외한 나머지 8개 특성을 이용하여 생성된 분류기가 가장 높은 분류정확도를 나타내었다(Fig. 3).

### 질감 특성에 의해 생성된 분류기에 대한 평가

다음으로 추출한 6개의 질감 특성에 대하여 sequential backward selection 방법을 이용하여 F-통계량이 낮은 특성을 하나씩 제거하면서 실험한 학습집단과 시험집단에 대한 오분류율과 다변량 분산 분석의 결과인 F-통계량을 비교하였다.

특성의 수가 3개와 4개일 때 시험집단에 대한 오분류율이 20%로 최소를 나타내지만, 학습집단에 대한 오분류율은 특성 수가 3개일 때 0%로 최소를 나타내므로, 3개의 특성을 이용하여 생성된 분류기가 가장 높은 분류정확도를 나타내었다(Fig.

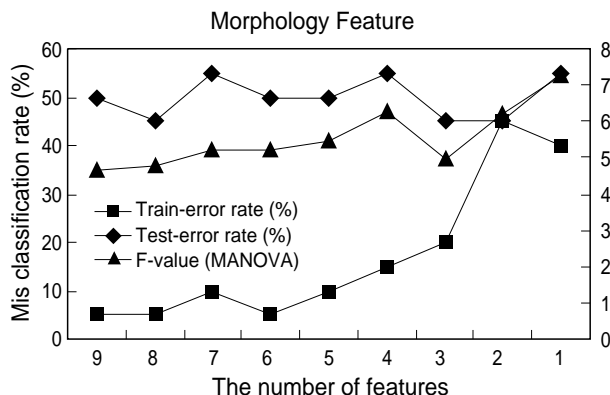


Fig. 3. The experimental result for selecting a significant feature vector from morphology features. Left side represents misclassification rate; Right side represents F-value; X axis represents the number of features.

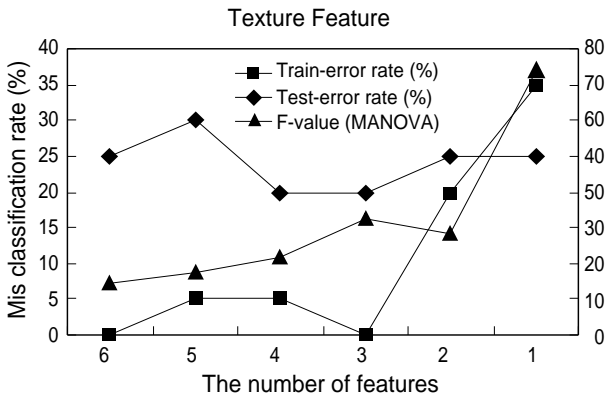


Fig. 4. The experimental result for selecting a significant feature vector from texture features.

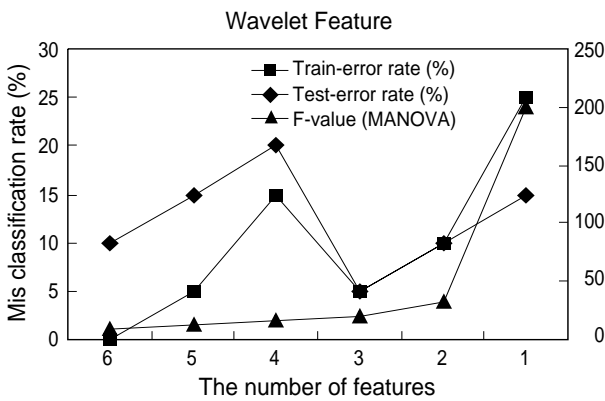


Fig. 5. The experimental result for selecting a significant feature vector from wavelet texture features.

4). 선택된 3개의 특성은 different moment, inverse different moment, max probability이다.

**웨이블렛 변환 영역의 질감 특성에 의해 생성된 분류기에 대한 평가**

웨이블렛 변환 영역에서의 질감 특성은 웨이블렛 변환 적용 후, 생성된 6개의 하위 영역에서 추출한 18개의 질감 특성 중 각각의 영역에서 가장 높은 F-통계량을 가지는 특성을 추출하여 1차적으로 6개의 특성을 선택하였다. 그 이유는 선행 연구에서 높은 분류정확도를 나타내었기 때문이다.<sup>17</sup> 선택된 6개의 특성에 대하여 sequential backward selection 방법을 이용하여 F-통계량이 낮은 특성을 하나씩 제거하면서 실험한 학습집단과 시험집단에 대한 오분류율과 다변량분산분석의 결과인 F-통계량을 비교하였다.

특성의 수가 3개일 때 시험집단에 대한 오분류율은 5%로 최소를 나타내며 특성의 수가 6개일 때 학습집단에 대한 오분류율이 0%로 최소를 나타내지만, 평가 기준의 우선순위와 MANOVA

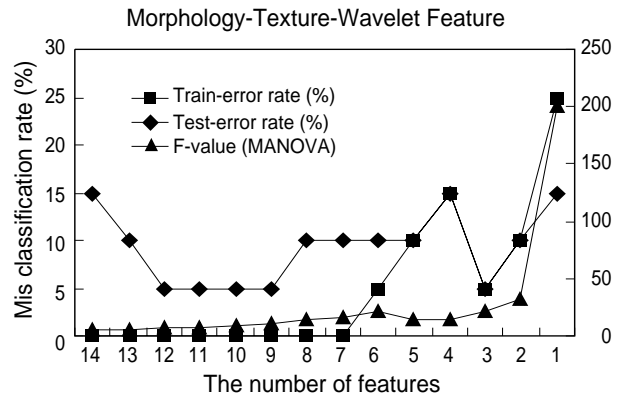


Fig. 6. The experimental result for selecting a significant feature vector from the combination of 3 category features.

VA 결과인 F-통계량을 비교해 볼 때, 3개의 특성을 이용하여 생성된 분류기가 가장 높은 분류정확도를 나타내었다(Fig. 5). 선택된 3개의 특성은 HL1 영역에서의 inverse different moment, HH1 영역에서의 inverse different moment, HH2 영역에서의 inverse different moment이다.

**형태학적 특성, 질감 특성, 웨이블렛 변환 영역의 질감 특성의 조합에 의해 생성된 분류기에 대한 평가**

형태학적 특성, 질감 특성, 웨이블렛 변환 영역에서의 질감 특성 각각에 의하여 생성된 분류기에 대한 평가 결과에서 가장 높은 분류정확도를 나타내는 특성을 선택하여 분류기를 생성하고 평가하였다. 조합한 특성의 수는 형태학적 특성 8개, 질감 특성 3개, 웨이블렛 변환 영역의 질감 특성 3개로 총 14개이다.

선택된 14개의 특성에 대하여 sequential backward selection 방법을 이용하여 F-통계량이 낮은 특성을 하나씩 제거하면서 실험한 학습집단과 시험집단에 대한 오분류율과 다변량분산분석의 결과인 F-통계량을 비교하였다.

특성의 수 12개, 11개, 10개, 9개 모두 시험집단에 대한 오분류율 5%와 학습집단에 대한 오분류율 0%로 최소를 나타낸다. 그러나 MANOVA의 결과인 F-통계량이 특성의 수가 9개일 때 가장 높은 값을 나타내므로, 9개의 특성을 이용하여 생성된 분류기가 가장 높은 분류정확도를 나타내었다(Fig. 6).

선택된 9개의 특성은 형태학적 특성인 area mean, perimeter CV, shape factor mean과 질감 특성인 different moment, inverse different moment, max probability 그리고 웨이블렛 변환 영역의 질감 특성인 HL1, HH1, HH2 영역 각각에서의 inverse different moment이다.

본 연구 결과에 따른 최적의 분류기는 형태학적 특성, 질감 특성, 웨이블렛 변환 영역의 질감 특성 모두를 조합하여 생성한 분류기로서 95%의 분류정확도를 나타내었다. 시험집단에 대한 분류정확도만을 비교하면 웨이블렛 변환 영역의 질감 특성들만

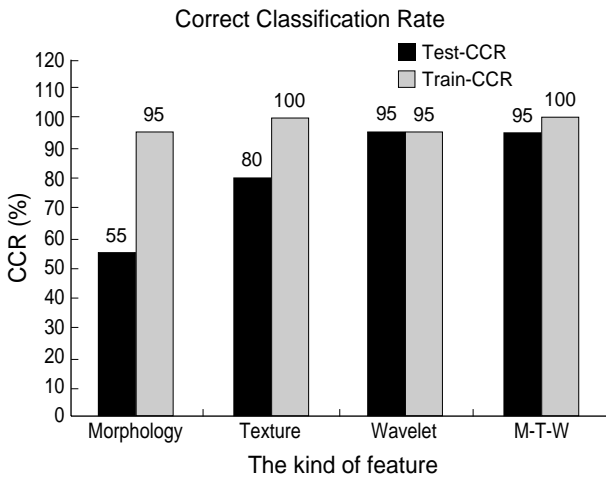


Fig. 7. The estimation result of the classification performance. Vertical axis represents correct classification rate and Horizontal axis represents the number of features. CCR represents correct classification rate.

으로 생성한 분류기 역시 95%의 분류정확도를 가지지만, 평가 기준의 우선순위에 따라 학습집단에 대한 분류정확도까지 평가 하면 세가지 특징 모두를 조합하여 생성한 분류기가 학습집단에 대해서는 100%의 분류정확도를 가지므로 최적의 분류기라 할 수 있다(Fig. 7).

### 고찰

본 연구 결과에 따른 최적의 분류기는 형태학적 특성, 질감 특성, 웨이블릿 변환 영역의 질감 특성을 조합하여 생성한 분류기로서 한 종류의 특성만을 이용하여 생성한 분류기보다 높은 분류정확도를 나타내었다. 이 결과로부터 한 특성의 높은 유의성보다는 여러 특성들간의 상호보완적 관계가 그룹 판별을 위해 높은 기여도를 나타낸다는 것을 알 수 있었다. 그리고 분류기의 분류정확도는 특성 추출과 특성의 평가, 그리고 특성 선택에 의해 결정된다는 사실을 확인할 수 있었다. 이전에 197명의 방광암 환자의 조직에 대한 자료를 이용하여 등급화한 연구에서 최등<sup>2</sup>은 질감 분석을 사용하여 84.3%의 상관성을 보였고, Jarkrans 등<sup>3</sup>은 형태학적 분석으로 73%를 보였으나, 이번 연구에서는 부분적이기는 하지만 95%의 높은 상관성을 얻었다.

본 연구에서 추출한 세포핵의 기하학적 특징을 분석하는 형태학적 특징이나, 염색질 양상의 내부 구조 변화를 분석할 수 있는 질감 특징, 그리고 영상이 가지고 있는 특성을 성분별로 분해하여 분석할 수 있는 웨이블릿 변환 영역에서 추출한 질감 특징은 현재 암세포 조직 영상 분석이 가지고 있는 문제점인 낮은 정확성과 전문가 의견과의 비일치성을 개선할 수 있었다.<sup>16</sup>

그러나 본 연구에 사용된 영상은 방광암 환자에 대한 영상으로부터 각 등급에서 대표적 영상들을 선택하여 수행한 것이므

로, 전체 슬라이드를 대상으로 수행하게 될 자동화 진단에 그대로 적용하기에는 기술적 어려움이 있다. 그러므로 영상 분석 또한 전체 조직 중 일부분만을 선택하여 분석한다는 문제점은 여전히 존재하며 이에 대한 연구가 필요하다.

본 연구는 방광암 세포 조직 영상을 암의 진행 정도에 따라 분류하기 위한 최적의 분류기를 찾고자 하는 연구이므로, 연구 결과에 따라 최종적으로 선택된 특성과 분류기가 방광암이라는 특정 암세포 조직 영상을 분류하는 데 있어서는 적합하지만, 모든 암세포 조직 영상 분류에 적합하다고는 할 수 없다. 그러므로 암의 종류에 따른 분류방법에 대한 연구가 필요한 실정이며 좀 더 표준화된 분류방법이 요구된다. 그리고 암 조직의 영상 특성들을 수치화된 데이터로 추출하였으므로 이에 대한 분류 방법론으로 신경망이론이나 퍼지이론을 적용하여 실험한 결과를 비교 연구하고,<sup>17</sup> 암 조직영상의 캡처 단계에서 색상 밴드별로 읽어들이거나 혹은 세포핵에 대한 삼차원 입체에서의 분석은 더욱 많은 정보를 갖고 실험할 수 있을 것이다.<sup>18,19</sup>

### 참고문헌

- Sowter C, Slavin G, Sowter G, *et al.* Morphometry of bladder carcinoma. morphometry and grading complement each other. *Anal Cell Pathol* 1991; 3: 1-9.
- Choi HK, Vasko J, Bengtsson E, *et al.* Grading of transitional cell bladder carcinoma by texture analysis of histological sections. *Anal Cell Pathol* 1994; 6: 327-43.
- Jarkrans T, Bengtsson E, Choi HK, *et al.* Grading of transitional cell bladder carcinoma by image analysis of histological section. *Anal Cell Pathol* 1995; 8: 135-58.
- Van de Wouwer G, Scheunders P, Livens S, *et al.* Wavelet correlation signatures for color texture characterization. *Pattern Recognition* 1999; 32: 443-51.
- Malmström PU, Busch C, Norlen BJ, *et al.* Progression and survival in bladder cancer: a retrospective study of 232 patients with 5 years follow-up. *Scan J Urol Nephrol* 1987; 21: 185-95.
- Edward CMB, Nellie WS, Jan PAB, *et al.* The value of morphometry and DNA flow cytometry in addition to classic prognosticator in supperficial unriary bladder carcinoma. *Am J Clin Pathol* 1989; 91: 243-8.
- Bergkvist A, Ljungqvist A, Moberger G. Classification bladder tumors based on the cellular pattern. *Acta Chir Scan* 1965; 130: 371-8.
- Schimmelpenning H, Falkmer UG, Hamper K, *et al.* Variation in feulgen stainability of epithelial parenchymal cells extracted from paraffin-embedded salivary gland specimens. *Cytometry* 1990; 11: 475-80.
- Flezar M, Us-Krasovec M, Zganec J, *et al.* Hydrolysis profiles of for-

- malin fixed paraffin-embedded tumor based on IOD (Integrated Optical Density) and nuclear texture feature measurements. *Anal Cell Pathol* 1999; 19: 111-8.
10. Robert MH, Shanmugam K, Dinstein I. Texture feature for image classification. *IEEE Trans. On System, Man, and Cybernetics* 1973; 3: 610-62.
  11. Pressman NJ. Markovian analysis of cervical cell images. *Histochem Cytochem* 1976; 24: 138-44.
  12. Choi HJ, Lee BI, Lee YS, *et al.* Texture analysis of carcinoma cell tissue images based on wavelet transform. *The Proceedings of Signal Processing and System* 2000; 305-8.
  13. Ronald PC, Jeffrey KS. Applied statistics and the SAS programming language. Prentice-Hall Inc., 1997; 629-725.
  14. Richard AJ, Dean WW. Applied multivariate statistical analysis, Prentice-Hall Inc., 1998; 150-80.
  15. Choi HJ, Lee BI, Choi HK. Classification of bladder carcinoma cell tissue sections using texture features and wavelet transform. *The Proceedings of Asia Pacific Association for Medical Informatics-MIC 2000 Conference* 2000; CD.
  16. Ooms E.C.M, Anderson W.A.D, Alons C.L, *et al.* An analysis of the performance of pathologists in the grading of bladder tumors. *Hum Pathol* 1983; 14: 140-3.
  17. Kolles H, Wangenheim A.V, Vince G.H, Niedermayer I, Feiden W. Automated grading of astrocytomas based on histomorphometric analysis of Ki-67 and feulgen stained paraffin sections: classification results of neuronal networks and discriminant Anlysis. *Anal Cell Pathol* 1995; 8: 101-16.
  18. Macville M.V.E, Van der Laak J.A.W.M, Speel E.J.M, *et al.* Spectral imaging of multi-color chromogenic dyes in pathological specimens. *Anal Cell Pathol* 2001; 22: 133-42.
  19. Preston K, Siderits R. New techniques for three-dimensional data analysis in histopathology. *Anal Quanti Cytol Histol* 1992; 14: 398-406.