

|                                                                                            |                                                                                                       |                                                    |
|--------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|----------------------------------------------------|
|           | (19) 대한민국특허청(KR)<br>(12) 공개특허공보(A)                                                                    | (11) 공개번호 10-2014-0134922<br>(43) 공개일자 2014년11월25일 |
| <hr/>                                                                                      |                                                                                                       |                                                    |
| (51) 국제특허분류(Int. Cl.)<br>C12Q 1/04 (2006.01) G01N 33/48 (2006.01)<br>G01N 33/574 (2006.01) | (71) 출원인<br>인하대학교 산학협력단<br>인천광역시 남구 인화로 100, 인하대학교 (용현동)                                              |                                                    |
| (21) 출원번호 10-2013-0054968                                                                  | (72) 발명자<br>이주홍<br>경기 성남시 분당구 돌마로486번길 7, 202동 401호 (서현동, 효자촌동아아파트)<br>송재원<br>인천 남동구 석산로72번길 9, (간석동) |                                                    |
| (22) 출원일자 2013년05월15일<br>심사청구일자 2013년05월15일                                                | (74) 대리인<br>특허법인 정안                                                                                   |                                                    |

전체 청구항 수 : 총 14 항

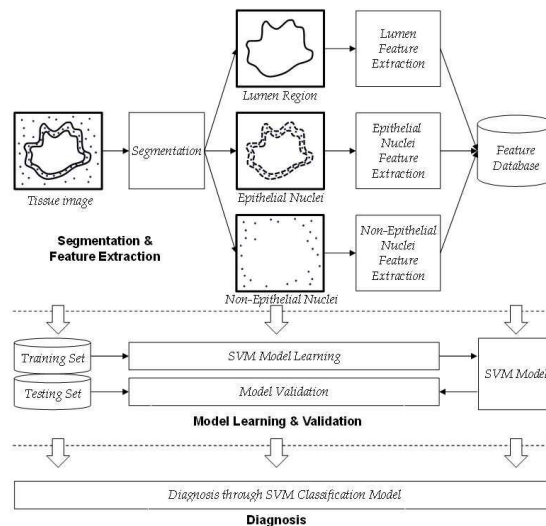
(54) 발명의 명칭 형태학적 특징 기반 췌장선암 등급 구분 방법

### (57) 요약

주어진 조직세포 영상으로부터 형태학적 특징을 추출하고, 추출된 특징을 기반으로 췌장선암을 진단하는 형태학적 특징 기반 췌장선암 등급 진단 방법에 관한 것으로서, 상기 조직세포 영상으로부터 형태학적 특징을 추출하는 제1 단계; 지도학습 기반의 분류기를 이용하여 진단 모델을 구성하는 제2 단계; 및, 상기 진단 모델을 이용하여, 상기 조직세포 영상의 췌장선암을 진단하는 제3 단계를 포함하고, 상기 진단 모델은 입력되는 형태학적 특징에 대하여 췌장선암 여부 또는 등급을 결정하는 구성을 마련한다.

상기와 같은 췌장선암 등급 진단 방법에 의하여, 형태학적 특징들을 추출하고, 추출된 특징으로 진단 모델을 학습시킴으로써 췌장선암을 정확하게 판별하고 각 병기들을 구분할 수 있다.

**대표도** - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 2010-0024216

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 일반연구자지원사업

연구과제명 병리진단 지식을 활용한 온톨로지 기반의 지능형 암 진단 시스템

기 여 율 1/1

주관기관 인하대학교 산학협력단

연구기간 2010.09.01 ~ 2013.08.31

---

## 특허청구의 범위

### 청구항 1

조직세포 영상으로부터 형태학적 특징을 추출하여 췌장선암을 진단하는 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서,

상기 조직세포 영상으로부터 형태학적 특징을 추출하는 제1 단계;

지도학습 기반의 분류기를 이용하여 진단 모델을 구성하는 제2 단계; 및,

상기 진단 모델을 이용하여, 상기 조직세포 영상의 췌장선암을 진단하는 제3 단계를 포함하고,

상기 진단 모델은 입력되는 형태학적 특징에 대하여 췌장선암 여부 또는 등급을 결정하는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

### 청구항 2

제1항에 있어서, 상기 제1 단계는,

(a) 조직세포 영상에서 내강 영역을 세그멘테이션하는 단계;

(b) 상기 조직세포 영상에서 핵들을 식별하는 단계;

(c) 식별된 핵들을 상피 세포 핵과 비상피 세포 핵으로 구분하는 단계;

(d) 상기 내강 영역에 대하여 내강 특징을 추출하는 단계; 및,

(e) 상기 상피세포 핵들로부터 상피세포 특징을 추출하는 단계를 포함하는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

### 청구항 3

제1항에 있어서,

상기 진단 모델은 SVM(Support Vector Machine) 분류기를 이용하는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

### 청구항 4

제1항에 있어서,

상기 제2 단계에서, 상기 형태학적 특징들의 벡터(이하 특징벡터)와 진단결과와 등급의 쌍으로 구성된 학습데이터로 상기 분류기를 학습하여 상기 진단 모델을 구성하는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

### 청구항 5

제2항에 있어서,

상기 (a)단계에서, 상기 내강 영역을 씨드 영역 양성(SGR) 방법에 의하여 식별하여 세그멘테이션하는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

## 청구항 6

제5항에 있어서, 상기 (a)단계는,

- (a1) 상기 조직세포 영상에 중간값 필터링과 배경상관 알고리즘을 적용한 후 최대 엔트로피 문턱치를 적용하여 이진 영상을 획득하는 단계;
- (a2) 상기 이진 영상으로부터 방향 누적맵을 생성하는 단계;
- (a3) 상기 방향 누적맵의 지역 최대점들을 후보 씨드점으로 선정하는 단계; 및,
- (a4) 상기 후보 씨드점들을 SRG 방법의 초기 파라미터로 이용하여 내강 영역의 경계를 구하는 단계를 포함하는 것을 특징으로 하는 형태학적 특징 기반 채장선암 등급 진단 방법.

## 청구항 7

제2항에 있어서,

상기 (b)단계에서, 상기 조직세포 영상에 k평균에 기반한 문턱치 처리를 한 후, 와더셰드(Watershed) 알고리즘을 적용하여 핵을 분리하는 것을 특징으로 하는 형태학적 특징 기반 채장선암 등급 진단 방법.

## 청구항 8

제2항에 있어서,

상기 (c)단계에서, 상기 내강 영역의 경계에 인접한 핵들을 상피세포 핵으로 구분하고, 상기 식별된 핵 전체에서 상피세포 핵을 제외한 나머지를 비상피 세포 핵으로 구분하는 것을 특징으로 하는 형태학적 특징 기반 채장선암 등급 진단 방법.

## 청구항 9

제2항에 있어서, 상기 (d)단계는,

- (d1) 상기 내강 영역의 경계(이하 오리지널 경계)로부터 이상적인 경계를 구하는 단계;
- (d2) 상기 오리지널 경계와 상기 이상적인 경계에서 서로 수직하는 점들의 거리를 이용하여 이형성 진폭 시그니처를 구하는 단계; 및,
- (d3) 상기 이형성 진폭 시그니처로부터 내강 특징을 추출하는 단계를 포함하는 것을 특징으로 하는 형태학적 특징 기반 채장선암 등급 진단 방법.

## 청구항 10

제9항에 있어서,

상기 (d1)단계에서, 상기 이상적인 경계는 상기 오리지널 경계의 볼록껍질을 상기 오리지널 경계의 내부 영역의 크기로 스케일링하여 구하는 것을 특징으로 하는 형태학적 특징 기반 채장선암 등급 진단 방법.

## 청구항 11

제9항에 있어서,

상기 (d3)단계에서, 상기 이형성 진폭 시그니처는 상기 이상적인 경계의 둘레 길이에 대한 둘레 상의 포인트(이하 이상적인 경계 포인트)의 이형성 진폭으로 나타내고, 상기 이형성 진폭은 상기 이상적인 경계 포인트에서 수직에 위치하는 오리지널 경계 상의 포인트 사이의 부호를 가진 거리인 것을 특징으로 하는 형태학적 특징 기반

췌장선암 등급 진단 방법.

#### 청구항 12

제11항에 있어서,

상기 이형성 진폭은 다음 [수식 2]의 이형성 진폭 함수  $A(t)$ 에 의해 구해지는 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

[수학식 4]

$$A(t) = \text{sgn}_t(q) \times \min_{p_t \in B^I, q \in B^O} (\text{distance}(p_t, q))$$

such that  $p_t \perp q$

단,  $t$ 는 상기 이상적인 경계에 있는 포인트들의 순서를 가리키는 인덱스 변수이고,  $\text{sgn}_t(q)$ 는  $p_t \in B^I$ 에서 수직 위치에 있는 점  $q \in B^O$  사이의 거리에 대한 부호를 나타내는 함수로서 점  $q$ 가 점  $p_t$ 의 바깥 경계에 위치해 있으면 +1, 안쪽 경계에 위치해 있으면 -1을 반환한다.

#### 청구항 13

제11항에 있어서,

상기 내강 특징은 이형성 진폭들에 대한 제곱 평균의 제곱근(RMSAA), 이형성 진폭 시그니처의 변곡점에서의 변화량의 합(TSAV), 이형성 비율(AtypiaRatio), 이형성 영역의 개수(#AtypiaRegions) 중 어느 하나 이상인 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

#### 청구항 14

제11항에 있어서,

상기 상피세포 특징은 상피 세포질 길이(CytoplasmLength) 및 상피세포의 세포질 길이에 대한 표준편차(CytoplasmLength<sub>SD</sub>) 중 어느 하나 이상인 것을 특징으로 하는 형태학적 특징 기반 췌장선암 등급 진단 방법.

### 명세서

#### 기술분야

[0001] 또한, 본 발명은 주어진 조직세포 영상으로부터 형태학적 특징을 추출하고, 추출된 특징을 기반으로 췌장선암을 진단하는 형태학적 특징 기반 췌장선암 등급 진단 방법에 관한 것이다.

#### 배경기술

[0002] 전통적으로 병리진단은 광학 현미경을 통해 주어진 조직 슬라이드를 수동으로 검토하고 진단의 경험과 지식에 의해 주관적으로 진단을 수행한다. 이러한 전통적인 병리진단은 몇 가지 문제점이 있다. 먼저 광학 현미경을 통한 스크리닝과정은 고배율에서 진행되기 때문에 매우 많은 시간과 노력을 필요로 한다 [비특허문헌 1], [비특허문헌 2]. 또한 진단의 개인적인 능력은 실제 진단에 결정적인 영향을 준다. 즉, 정량적 특징 분석에 의한 객관화된 진단이 아니라 주관적인 진단이 이루어지기 때문에 같은 증상에 대해 의사들마다 다양한 의견을 보일 수

있다[비특허문헌 3]-[비특허문헌 5].

- [0003] 디지털 병리학(Digital pathology)의 등장은 이러한 병리진단 환경을 변화시키고 있다. 디지털 병리학은 글래스 슬라이드(glass slide)를 디지털화함으로써 편리한 스크리닝 환경과 정량적 특징분석을 통한 객관적 진단을 가능하게 했다[비특허문헌 6]-[비특허문헌 8]. 그럼에도 불구하고 조직영상의 스크리닝을 통한 병리진단은 수동적으로 수행되어지기 때문에 많은 시간과 노동력을 필요로 한다. 따라서 주관적 판단에 따른 문제가 여전히 존재한다.
- [0004] 최근 이러한 문제점을 극복하기 위해 디지털 병리학과 CAD(Computer Aided detection) 기술을 결합하려는 연구가 점차 증가하고 있다. 많은 자동화 진단과 관련된 연구들은 종양의 진단뿐 아니라 발견된 종양의 등급(grading)까지 커버하고 있다. 병리적 등급(grading)은 질병의 진행을 식별하고 환자에 대한 올바른 처방을 위해 매우 가치가 있다[비특허문헌 9]. 실제로 병리 의사들은 환자의 베스트한 처방을 위해 병기를 구분에 대한 많은 훈련을 하는 것으로 알려져 있다 [비특허문헌 7]. 현재 많은 연구들은 전립선 암과 유방암을 대상으로 진행되고 있다.
- [0005] 먼저 전립선 암 진단과 관련된 CAD연구들은 대부분 Gleason에 의해 개발된 글리슨 등급 시스템(Gleason grading system)[비특허문헌 10]에 기반한다. Tabesh et al. [비특허문헌 11]는 전립선암(prostate cancer)을 진단하고 글리슨 등급(Gleason grading)에 기반하여 저(low)와 고(high)로 병기를 구분하는 머신러닝 기반의 자동화 시스템을 제안하였다. 그들은 종양(tumour)의 진단과 병기 구분을 위해 주어진 병리영상의 전반적 수준(global (image) level)과 조직학적 개체 수준(historical object level)에서 색(color), 텍스처(texture), 그리고 형태학적 특징(morphometric features)들을 추출하고 그들을 이용하여 분류기(Gaussian, k-NN, SVM과 같은) 학습하였다.
- [0006] Naik et al. [비특허문헌 12]은 중간 글리슨 등급(intermediate Gleason grades)를 구분하기 위한 시스템을 제안했다. 그들은 저수준(low-level) 정보를 이용하는 베이저안 분류기(Bayesian classifier)를 사용하여 후보 분비선 영역(candidate gland region)을 식별하고, 실험 도메인(Empirical domain) 정보를 이용하여 분비선(gland)으로 식별된 참거짓(false positive) 영역들을 제외하였다. 이후, 식별된 분비선으로부터 형태학적 특징들을 추출하고, SVM을 통해 Gleason grade 3, Gleason grade 4, 그리고 양성(benign)을 구분하였다. Huwang 과 Lee는 [비특허문헌 13] 감별 박스 계측(differential box-counting)과 엔트로피 기반 프랙탈 차원 추정(tropy-based fractal dimension estimation) 기법을 사용하여 Gleason grading을 위한 특징들을 추출하고, Bayesina, k-NN, SVM분류기를 이용하여 전립선 암 이미지(prostate cancer image)를 4등급으로 구분하였다. 이 외에도 많은 연구들이 전립선 종양(postate tumour)을 진단과 Gleason Grading 시스템을 자동화하기 위해 진행되고 있다 [비특허문헌 14]-[비특허문헌 17].
- [0007] 다음으로 유방암과 관련된 CAD연구들은 다음과 같다. Anderson et al. [비특허문헌 18] 은 비외과적 유방 질환(non-invasive breast lesions)의 양성(benign)인 DH(ductal hyperplasia)와 악성(malignant)인 DCIS(ductal carcinoma in situ)구분 문제를 연구하였다. 그들은 이 연구에서 지식가이드 머신 비전(knowledge-guided machine vision)을 사용하여 유방 유관(breast duct)을 자동으로 세그먼트하고, DH와 DCIS를 구분하는데 중요한 증식성 병소의 유관 패턴(duct pattern)을 측정하기 위해 유관체(duct cribriformity)와 구조적 복잡도(architectural complexity)와 같은 측정도구(measure)들을 제안하였다.
- [0008] Bilgin et al. [비특허문헌 19]은 그래프 이론 기술(graph theoretical techniques)에 기반하여 유방암(breast cancer)을 진단하였다. 그들은 k-means 알고리즘을 이용하여 주어진 조직세포 영상을 세그먼트하고, 세포들의 위치좌표를 사용하여 셀 그래프(cell-graph)를 생성하였다. 주어진 조직 영상을 양성(benign), 침습성(invasive), 비침습성(non-invasive)으로 분류하는 SVM모델은 생성된 cell-graph로부터 계산된 정량적 척도(quantitative metric)들에 의해 학습된다. Basavanhally et al. [비특허문헌 20]는 HER2+ 유방암(breast cancer)의 예측 인자(viable prognostic indicator)로서 알려진 백혈구 침범(lymphocytic infiltration)의 정도(extent)를 식별하고 등급 결정(grading)하는 시스템을 제안했다. 그들은 우선 림프구(lymphocytes)를 영역 양성(region grwowing)과 마코프 랜덤필드 알고리즘(Makov random field algorithm)을 이용하여 식별하였다. 이후, 식별된 림프구를 이용하여 구조적 특징(architectural features)들을 추출하고, 그 특징들에 의해 학습된 SVM 분류기를 이용하여 LI의 정도(extent)를 저(low), 저(midium), 고(high) 등급으로 구분하였다. 그 밖에 유방암 진단과 grading을 위한 연구들이 많이 진행되고 있다 [비특허문헌 21]-[비특허문헌 24].
- [0009] 많은 병리 CAD연구들은 각 질병에 대한 병리적 특징들과 그들을 양적으로 측정하기 위한 방법들을 고안하는데 많은 노력을 기울인다. 왜냐하면 병리진단은 병소와 질병의 유형에 따라 다양한 진단 방법이 존재하기

때문이다. 현재 유방과 전립선 종양을 대상으로 하는 병리진단 CAD 외에 결장(colonic)[비특허문헌 25], 방광(bladder)[비특허문헌 26], 신경아세포종(neuroblastoma) [비특허문헌 27][비특허문헌 28][비특허문헌 29], 여포성 림프종(follicular lymphoma)[비특허문헌 30][비특허문헌 31] 종양(tumour)들을 대상으로 몇몇 연구들이 진행되고 있다.

[0010] 아직 병리진단 CAD 연구는 아직 초기 단계이며 몇몇 소수 부위만을 대상으로 하고 있다. 다양한 종양진단을 위한 많은 방법들이 있다. 따라서 병리학적(pathological) CAD에 대한 보다 많은 연구들이 필요하다.

## 선행기술문헌

### 비특허문헌

- [0011] (비특허문헌 0001) [비특허문헌 1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," Biomedical Engineering, IEEE Reviews in, vol. 2, pp. 147 -171, 2009.
- (비특허문헌 0002) [비특허문헌 2] A. E. Tutac, D. Racoceanu, T. Putti, W. Xiong, W.-K. Leow, and V. Cretu, "Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images," in International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008, 2008, vol. 2, pp. 107 -112.
- (비특허문헌 0003) [비특허문헌 3] S. M. Ismail, A. B. Colclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. O'Sullivan, J. M. Summerell, and R. G. Newcombe, "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," BMJ, vol. 298, no. 6675, pp. 707-710, Mar. 1989.
- (비특허문헌 0004) [비특허문헌 4] A. Andron, C. Magnani, P. G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, and B. Terracini, "Malignant mesothelioma of the pleura: interobserver variability," J. Clin. Pathol., vol. 48, no. 9, pp. 856-860, Sep. 1995.
- (비특허문헌 0005) [비특허문헌 5] C. Demir and B. Yener, "Automated Cancer Diagnosis Based on Histopathological Systematic Images: A Systematic Survey." Rensselaer Polytechnic Institute, 2005.
- (비특허문헌 0006) [비특허문헌 6] S. Al-Janabi, A. Huisman, and P. J. Van Diest, "Digital pathology: current status and future perspectives," Histopathology, vol. 61, no. 1, pp. 1-9, 2012.
- (비특허문헌 0007) [비특허문헌 7] A. Madabhushi, "Digital pathology image analysis: opportunities and challenges," Imaging in Medicine, vol. 1, no. 1, pp. 7-10, Oct. 2009.
- (비특허문헌 0008) [비특허문헌 8] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi, and J. R. Gilbertson, "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," Human Pathology, vol. 40, no. 8, pp. 1057-1069, Aug. 2009.
- (비특허문헌 0009) [비특허문헌 9] J. I. Epstein and G. J. Netto, Biopsy Interpretation of the Prostate, Fourth. Lippincott Williams & Wilkins, 2007.
- (비특허문헌 0010) [비특허문헌 10] D. F. Gleason and G. T. Mellinger, "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging," J. Urol., vol. 111, no. 1, pp. 58-64, Jan. 1974.
- (비특허문헌 0011) [비특허문헌 11] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images," IEEE Transactions on Medical Imaging, vol. 26, no. 10, pp. 1366 -1378, Oct. 2007.
- (비특허문헌 0012) [비특허문헌 12] S. Naik, S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, Gland Segmentation and Computerized Gleason Grading of Prostate Histology by Integrating Low-, High-level and Domain Specific Information. MIAAB Workshop: , 2007.

- (비특허문헌 0013) [비특허문헌 13] P.-W. Huang and C.-H. Lee, "Automatic Classification for Pathological Prostate Images Based on Fractal Analysis," IEEE Transactions on Medical Imaging, vol. 28, no. 7, pp. 1037 -1050, Jul. 2009.
- (비특허문헌 0014) [비특허문헌 14] A. W. Wetzel, "Evaluation of prostate tumor grades by content-based image retrieval," 1999, vol. 3584, pp. 244-252.
- (비특허문헌 0015) [비특허문헌 15] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszewski, "AUTOMATED GRADING OF PROSTATE CANCER USING ARCHITECTURAL AND TEXTURAL IMAGE FEATURES," in 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007, 2007, pp. 1284 -1287.
- (비특허문헌 0016) [비특허문헌 16] S.-K. Tai, Y.-C. Wu, C.-Y. Li, Y.-J. Jan, and S.-C. Lin, "Computer-assisted Detection and Grading of Prostatic Cancer in Biopsy Image | www.libsearch.com - TEST," Mar-2010. [Online]. Available: <http://www.libsearch.com/view/1046167>. [Accessed: 19-Jul-2012].
- (비특허문헌 0017) [비특허문헌 17] Y. Peng, Y. Jiang, S.-T. Chuang, and X. J. Yang, "Computer-aided detection of prostate cancer on tissue sections," Appl. Immunohistochem. Mol. Morphol., vol. 17, no. 5, pp. 442-450, Oct. 2009.
- (비특허문헌 0018) [비특허문헌 18] N. H. Anderson, P. W. Hamilton, P. H. BarTels, D. Thompson, R. Montironi, and J. M. Sloan, "COMPUTERIZED SCENE SEGMENTATION FOR THE DISCRIMINATION OF ARCHITECTURAL FEATURES IN DUCTAL PROLIFERATIVE LESIONS OF THE BREAST," The Journal of Pathology, vol. 181, no. 4, pp. 374-380, 1997.
- (비특허문헌 0019) [비특허문헌 19] C. Bilgin, C. Demir, C. Nagi, and B. Yener, "Cell-Graph Mining for Breast Tissue Modeling and Classification," in 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, 2007, pp. 5311 -5314.
- (비특허문헌 0020) [비특허문헌 20] A. N. Basavanahally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology," IEEE Trans Biomed Eng, vol. 57, no. 3, pp. 642-653, Mar. 2010.
- (비특허문헌 0021) [비특허문헌 21] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008, 2008, pp. 496 -499.
- (비특허문헌 0022) [비특허문헌 22] W. H. Wolberg, W. N. Street, D. M. Heisey, and O. L. Mangasarian, "Computer-derived nuclear features distinguish malignant from benign breast cytology," Hum. Pathol., vol. 26, no. 7, pp. 792-796, Jul. 1995.
- (비특허문헌 0023) [비특허문헌 23] R. Fernandez-Gonzalez, T. Deschamps, A. Idica, R. Malladi, and C. O. de Solorzano, "Automatic segmentation of histological structures in mammary gland tissue sections," Journal of Biomedical Optics, vol. 9, no. 3, pp. 444-453, May 2004.
- (비특허문헌 0024) [비특허문헌 24] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, Breast Cytology Diagnosis Via Digital Image Analysis. 1993.
- (비특허문헌 0025) [비특허문헌 25] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray, "Fractal analysis in the detection of colonic cancer images," IEEE Transactions on Information Technology in Biomedicine, vol. 6, no. 1, pp. 54 -58, Mar. 2002.
- (비특허문헌 0026) [비특허문헌 26] H. K. Choi, T. Jarkrans, E. Bengtsson, J. Vasko, K. Wester, P. U. Malmström, and C. Busch, "Image analysis based grading of bladder carcinoma. Comparison of object, texture and graph based methods and their reproducibility," Anal Cell Pathol, vol. 15, no. 1, pp. 1-18, 1997.

- (비특허문헌 0027) [비특허문헌 27] M. N. Gurcan, J. Kong, O. Sertel, B. B. Cambazoglu, J. Saltz, and U. Catalyurek, "Computerized pathological image analysis for neuroblastoma prognosis," AMIA Annu Symp Proc, pp. 304-308, 2007.
- (비특허문헌 0028) [비특허문헌 28] O. Sertel, J. Kong, H. Shimada, U. V. Catalyurek, J. H. Saltz, and M. N. Gurcan, "Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development," Pattern Recognition, vol. 42, no. 6, pp. 1093-1103, Jun. 2009.
- (비특허문헌 0029) [비특허문헌 29] B. B. Cambazoglu, O. Sertel, J. Kong, J. Saltz, M. N. Gurcan, and U. V. Catalyurek, "Efficient processing of pathological images using the grid: computer-aided prognosis of neuroblastoma," in Proceedings of the 5th IEEE workshop on Challenges of large applications in distributed environments, New York, NY, USA, 2007, pp. 35-41.
- (비특허문헌 0030) [비특허문헌 30] K. Belkacem-Boussaid, M. Pennell, G. Lozanski, A. Shana'ah, and M. N. Gurcan, "Effect of pathologist agreement on evaluating a computer-aided assisted system: Recognizing centroblast cells in follicular lymphoma cases," in 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010, pp. 1411-1414.
- (비특허문헌 0031) [비특허문헌 31] O. Sertel, G. Lozanski, A. Shana'ah, and M. N. Gurcan, "Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation," IEEE Trans Biomed Eng, vol. 57, no. 10, pp. 2613-2616, Oct. 2010.
- (비특허문헌 0032) [비특허문헌 32] G. Kloppel, R. H. Hruban, D. S. Longnecker, G. Adler, S. E. Kern, and T. J. Partanen, "Ductal adenocarcinoma of the pancreas," in World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the digestive System, 3th ed., International Agency for Research on Cancer, 2000.
- (비특허문헌 0033) [비특허문헌 33] R. H. Hruban, P. Boffetta, C. Iacobuzio-Donahue, Y. Kato, S. E. Kern, D. S. Klimstra, G. Kloppel, A. Maitra, G. J. A. Offerhaus, and M. B. Pitman, "Ductal adenocarcinoma of the pancreas," in WHO Classification of Tumours of the Digestive System, 4th ed., vol. 3, International Agency for Research on Cancer, 2010, pp. 217-224.
- (비특허문헌 0034) [비특허문헌 34] C. J. Yeo, J. L. Cameron, T. A. Sohn, K. D. Lillemoe, H. A. Pitt, M. A. Talamini, R. H. Hruban, S. E. Ord, P. K. Sauter, J. Coleman, M. L. Zahurak, L. B. Grochow, and R. A. Abrams, "Six hundred fifty consecutive pancreaticoduodenectomies in the 1990s: pathology, complications, and outcomes.," Ann Surg, vol. 226, no. 3, pp. 248-260, Sep. 1997.
- (비특허문헌 0035) [비특허문헌 35] R. Adams and L. Bischof, "Seeded region growing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 6, pp. 641-647, Jun. 1994.
- (비특허문헌 0036) [비특허문헌 36] J. W. Song, J. H. Lee, T. S. Park, S. J. Chun, and J. H. Choi, "Mucinous Cystadenoma Classification System using Automated Epithelial Tissue Detection," International Conference on Machine Vision, 2010.
- (비특허문헌 0037) [비특허문헌 37] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images," Anal Cell Pathol, vol. 25, no. 1, pp. 1-36, 2003.
- (비특허문헌 0038) [비특허문헌 38] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," Comput. Vision Graph. Image Process., vol. 41, no. 2, pp. 233-260, Feb. 1988.
- (비특허문헌 0039) [비특허문헌 39] A. Z. Chitade and S. K. Katiyar, "Color based image segmentation using K-means clustering," International Journal of Engineering Science and Technology, vol. 2, no. 10, pp. 5319-5325, Oct. 2010.
- (비특허문헌 0040) [비특허문헌 40] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Prentice Hall, 2008.
- (비특허문헌 0041) [비특허문헌 41] C. Pan, C.-X. Zheng, and H.-J. Wang, "Robust color image

- segmentation based on mean shift and marker-controlled watershed algorithm," in 2003 International Conference on Machine Learning and Cybernetics, 2003, vol. 5, pp. 2752 - 2756 Vol.5.
- (비특허문헌 0042) [비특허문헌 42] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679-688, Oct. 2006.
- (비특허문헌 0043) [비특허문헌 43] F.-L. Chung, T.-C. Fu, V. Ng, and R. W. P. Luk, "An evolutionary approach to pattern-based time series segmentation," IEEE Transactions on Evolutionary Computation, vol. 8, no. 5, pp. 471 - 489, Oct. 2004.
- (비특허문헌 0044) [비특허문헌 44] "Efficient Time Series Matching by Wavelets," in Proceedings of the 15th International Conference on Data Engineering, Washington, DC, USA, 1999, p. 126-.
- (비특허문헌 0045) [비특허문헌 45] M. Kosmahl, U. Pauser, M. Anlauf, and G. Kloppel, "Pancreatic ductal adenocarcinomas with cystic features: neither rare nor uniform," Mod. Pathol., vol. 18, no. 9, pp. 1157-1164, Sep. 2005.
- (비특허문헌 0046) [비특허문헌 46] M. Kallergi, "Evaluation Strategies for Medical- Image Analysis and Processing Methodologies," in Medical Image Analysis Methods, vol. 18, L. Costaridou, Ed. CRC Press, 2005.
- (비특허문헌 0047) [비특허문헌 47] J. K. Udupa, V. R. LaBlanc, H. Schmidt, C. Imielinska, P. K. Saha, G. J. Grevera, Y. Zhuge, L. M. Currie, P. Molholt, and Y. Jin, "Methodology for evaluating image-segmentation algorithms," pp. 266-277, May 2002.
- (비특허문헌 0048) [비특허문헌 48] T. Allen, Particle Size Measurement: Volume 1: Powder sampling and particle size measurement. Springer, 1996.
- (비특허문헌 0049) [비특허문헌 49] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images," Analytical Cellular Pathology, vol. 25, no. 1, pp. 1-36, Jan. 2003.
- (비특허문헌 0050) [비특허문헌 50] W. . Rasband, "ImageJ," 2012-1997. [Online]. Available: <http://imagej.nih.gov/ij/>. [Accessed: 25-Aug-2012].
- (비특허문헌 0051) [비특허문헌 51] V. N. Vapnik, Statistical Learning Theory, 1st ed. Wiley-Interscience, 1998.
- (비특허문헌 0052) [비특허문헌 52] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Min. Knowl. Discov., vol. 2, no. 2, pp. 121-167, Jun. 1998.
- (비특허문헌 0053) [비특허문헌 53] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 27:1-27:27, May 2011.
- (비특허문헌 0054) [비특허문헌 54] D. L. Verbyla and J. A. Litvaitis, "Resampling methods for evaluating classification accuracy of wildlife habitat models," Environmental Management, vol. 13, no. 6, pp. 783-787, Nov. 1989.
- (비특허문헌 0055) [비특허문헌 55] H. J. Ader, G. J. Mellenbergh, and D. J. Hand, Advising on research methods : a consultant's companion. Huizen, Netherlands: Johannes van Kessel Pub, 2008.
- (비특허문헌 0056) [비특허문헌 56] J. A. Swets, "Measuring the accuracy of diagnostic systems," Science, vol. 240, no. 4857, pp. 1285-1293, Jun. 1988.
- (비특허문헌 0057) [비특허문헌 57] M. Greiner, D. Pfeiffer, and R. D. Smith, "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests," Prev. Vet. Med., vol. 45, no. 1-2, pp. 23-41, May 2000.
- (비특허문헌 0058) [비특허문헌 58] L. J. Williams and H. Abdi, "Fisher's least significance difference (LSD) test," in Encyclopedia of Research Design, 1st ed., SAGE Publications, Inc, 2010, pp. 491-494.
- (비특허문헌 0059) [비특허문헌 59] R.L. Graham, "An efficient algorithm for determining the convex

hull of a finite planar set", Info. Proc. Lett. 1, 132-133 (1972).

## 발명의 내용

### 해결하려는 과제

- [0012] 본 발명의 목적은 상술한 바와 같은 문제점을 해결하기 위한 것으로, 주어진 조직세포 영상으로부터 형태학적 특징을 추출하고, 추출된 특징을 기반으로 췌장선암을 진단하는 형태학적 특징 기반 췌장선암 등급 진단 방법을 제공하는 것이다. 즉, PDAC(pancreatic ductal adenocarcinomas, 췌장관 선암종 또는 췌장선암)를 식별하고 그들의 병기를 구분하기 위하여, PDAC의 진단과 등급(grading)을 위한 새로운 형태학적 특징들을 추출하고, 추출된 특징을 이용하여 췌장선암의 등급을 진단하는 방법을 제공하는 것이다.
- [0013] 특히, 본 발명의 목적은 췌장선암에 대한 병리진단에서 주로 관심을 두는 부위인 내강과 상피세포로 구성되는 관을 대상으로, 내강과 상피세포들을 세그멘테이션하고, 식별된 오브젝트들로부터 췌장선암을 진단하기 위해 고안된 형태학적 특징들 추출하는 형태학적 특징 기반 췌장선암 등급 진단 방법을 제공하는 것이다.
- [0014] 췌장선암 진단 시 주로 검토되는 부위는 내강과 상피세포로 구성되는 관 부위이다. 따라서 본 발명은 각 오브젝트들(내강, 상피세포, 비상피세포)을 주어진 조직포 영상으로부터 세그멘테이션하고, 진단을 위한 형태학적 특징들을 추출하고, 이후, 추출된 특징들을 이용하여 진단을 위한 SVM모델을 학습시키고 췌장선암의 진단 및 병기를 구분하는 췌장선암 등급 진단 방법을 제공하는 것을 목적으로 한다.

[0015]

### 과제의 해결 수단

- [0016] 상기 목적을 달성하기 위해 본 발명은 조직세포 영상으로부터 형태학적 특징을 추출하여 췌장선암을 진단하는 형태학적 특징 기반 췌장선암 등급 진단 방법에 관한 것으로서, 상기 조직세포 영상으로부터 형태학적 특징을 추출하는 제1 단계; 지도학습 기반의 분류기를 이용하여 진단 모델을 구성하는 제2 단계; 및, 상기 진단 모델을 이용하여, 상기 조직세포 영상의 췌장선암을 진단하는 제3 단계를 포함하고, 상기 진단 모델은 입력되는 형태학적 특징에 대하여 췌장선암 여부 또는 등급을 결정하는 것을 특징으로 한다.
- [0017] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 제1 단계는, (a) 조직세포 영상에서 내강 영역을 세그멘테이션하는 단계; (b) 상기 조직세포 영상에서 핵들을 식별하는 단계; (c) 식별된 핵들을 상피 세포 핵과 비상피 세포 핵으로 구분하는 단계; (d) 상기 내강 영역에 대하여 내강 특징을 추출하는 단계; 및, (e) 상기 상피세포 핵들로부터 상피세포 특징을 추출하는 단계를 포함하는 것을 특징으로 한다.
- [0018] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 진단 모델은 SVM(Support Vector Machine) 분류기를 이용하는 것을 특징으로 한다.
- [0019] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 제2 단계에서, 상기 형태학적 특징들의 벡터(이하 특징벡터)와 진단결과의 등급의 쌍으로 구성된 학습데이터로 상기 분류기를 학습하여 상기 진단 모델을 구성하는 것을 특징으로 한다.
- [0020] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 (a)단계에서, 상기 내강 영역을 씨드 영역 양성(SGR) 방법에 의하여 식별하여 세그멘테이션하는 것을 특징으로 한다.
- [0021] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 (a)단계는, (a1) 상기 조직세포 영상에 중간값 필터링과 배경상관 알고리즘을 적용한 후 최대 엔트로피 문턱치를 적용하여 이진 영상을 획득하는 단계; (a2) 상기 이진 영상으로부터 방향 누적맵을 생성하는 단계; (a3) 상기 방향 누적맵의 지역 최대점들을 후보 씨드점으로 선정하는 단계; 및, (a4) 상기 후보 씨드점들을 SRG 방법의 초기 파라미터로 이용하여 내강 영역의 경계를 구하는 단계를 포함하는 것을 특징으로 한다.
- [0022] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 (a3)단계에서, 상기 방향 누적맵에 문턱치 처리를 적용하여 후보 씨드점을 구하는 것을 특징으로 한다.
- [0023] 또, 본 발명은 형태학적 특징 기반 췌장선암 등급 진단 방법에 있어서, 상기 (b)단계에서, 상기 조직세포 영상에 k평균에 기반한 문턱치 처리를 한 후, 와더셰드(Watershed) 알고리즘을 적용하여 핵을 분리하는 것을 특징

으로 한다.

[0024] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 (c)단계에서, 상기 내강 영역의 경계에 인접한 핵들을 상피세포 핵으로 구분하고, 상기 식별된 핵 전체에서 상피세포 핵을 제외한 나머지를 비상피 세포 핵으로 구분하는 것을 특징으로 한다.

[0025] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 상피세포 핵들의 집합  $N^E$ 를 다음 [수식 1]에 의해 구하는 것을 특징으로 한다.

[0026] [수식 1]

$$N^E = \bigcup_{p \in B^0} \left\{ n \mid \arg \min_{n \in N} (distance(centroid(n), p)) \right\}$$

[0027] 단,  $B^0$ 는 내강 영역의 경계이고, centroid(-)는 식별된 핵 n의 중심 포인트를 반환하는 함수이고, distance( $\cdot, \cdot$ )는 주어진 두 점들 사이의 유클리디언 거리를 반환하는 함수.

[0029] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 (d)단계는, (d1) 상기 내강 영역의 경계(이하 오리지널 경계)로부터 이상적인 경계를 구하는 단계; (d2) 상기 오리지널 경계와 상기 이상적인 경계에서 서로 수직하는 점들의 거리를 이용하여 이형성 진폭 시그니처를 구하는 단계; 및, (d3) 상기 이형성 진폭 시그니처로부터 내강 특징을 추출하는 단계를 포함하는 것을 특징으로 한다.

[0030] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 (d1)단계에서, 상기 이상적인 경계는 상기 오리지널 경계의 볼록껍질을 상기 오리지널 경계의 내부 영역의 크기로 스케일링하여 구하는 것을 특징으로 한다.

[0031] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 (d2)단계에서, 상기 이형성 진폭은 이상적인 경계 상의 포인트에서 수직에 위치하는 오리지널 경계 상의 포인트 사이의 거리로 표현되, 오리지널 경계 상의 포인트가 이상적 경계의 바깥 경계 또는 안쪽 경계에 있는지에 따라 부호를 달리하는 거리로 표현되는 것을 특징으로 한다.

[0032] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 (d3)단계에서, 상기 이형성 진폭 시그니처는 상기 이상적인 경계의 둘레 길이에 대한 둘레 상의 포인트(이하 이상적인 경계 포인트)의 이형성 진폭으로 나타내고, 상기 이형성 진폭은 상기 이상적인 경계 포인트에서 수직에 위치하는 오리지널 경계 상의 포인트 사이의 부호를 가진 거리인 것을 특징으로 한다.

[0033] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 이형성 진폭은 다음 [수식 2]의 이형성 진폭 함수 A(t)에 의해 구해지는 것을 특징으로 한다.

[0034] [수식 2]

$$A(t) = \text{sgn}_t(q) \times \min_{p_t \in B^I, q \in B^0} (distance(p_t, q))$$

such that  $p_t \perp q$

[0035] 단, t는 상기 이상적인 경계에 있는 포인트들의 순서를 가리키는 인덱스 변수이고,  $\text{sgn}_t(q)$ 는  $p_t \in B^I$ 에서 수직 위치에 있는 점  $q \in B^0$  사이의 거리에 대한 부호를 나타내는 함수로서 점 q가 점  $p_t$ 의 바깥 경계에 위치해 있으면 +1, 안쪽 경계에 위치해 있으면 -1을 반환한다.

[0037] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 내강 특징은 이형성 진폭들에 대한 평균의 제공근(RMSAA), 이형성 진폭 시그니처의 변곡점에서의 변화량의 합(TSAV), 이형성 비율(AtypiaRatio), 이형성 영역의 개수(#AtypiaRegions) 중 어느 하나 이상인 것을 특징으로 한다.

[0038] 또, 본 발명은 형태학적 특징 기반 채장선암 등급 진단 방법에 있어서, 상기 상피세포 특징은 상피 세포질 길이

(CytoplasmLength) 및 상피세포의 세포질 길이에 대한 표준편차(CytoplasmLength<sub>sp</sub>) 중 어느 하나 이상인 것을 특징으로 한다.

### 발명의 효과

[0039] 상술한 바와 같이, 본 발명에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법에 의하면, 추출된 형태학적 특징들로 췌장선암을 정확하게 판별하고 각 병기들을 구분할 수 있는 효과가 얻어진다.

[0040] 특히, 대표적인 통계적 분류기법인 SVM모델을 학습하고 분류성능을 평가한 결과, 본 발명에 의해 추출된 형태학적 특징들이 정상과 췌장선암을 구분하는데 있어서 94.38%의 분류 정확도도를 보였으며, 췌장선암에 대한 1 등급(Grade1)과 2 등급(Grade2)의 병기 기분에 있어서도 97.03%의 분류 정확도를 보이고 있다.

### 도면의 간단한 설명

- [0041] 도 1은 본 발명을 실시하기 위한 전체 시스템의 구성도.
- 도 2는 본 발명의 대상인 췌장선암의 조직세포 영상으로서, (a) 정상관, (b) 등급 1(well-differentiated), (c) 등급 2(moderately differentiated), (d) 등급 3(poorly differentiated)의 영상의 예시도.
- 도 3과 도 4는 본 발명의 일실시예에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법을 설명하는 도식화된 도면과 흐름도.
- 도 5는 본 발명의 일실시예에 따른 췌장선암의 등급을 위한 형태학적 특징 추출 방법을 설명하는 흐름도.
- 도 6은 본 발명에 따른 내장 영역 식별 과정을 설명하는 각 과정의 예시 영상.
- 도 7은 본 발명에 따른 식별된 핵의 집합의 예시 영상.
- 도 8은 본 발명에 따른 상피-핵 집합 구하는 알고리즘.
- 도 9는 본 발명에 따른 주요 표기를 나타낸 표.
- 도 10은 본 발명에 따른 오리지널 내장 경계와 이상적인 내장 경계의 영상.
- 도 11은 본 발명에 따른 이형성-진폭 측정 과정을 보여주는 영상 및 그래프.
- 도 12는 본 발명에 따른 도 1의 영상에 대한 이형성-진폭 시그니처.
- 도 13은 본 발명에 따른 도 12의 등급 2에 대한 이형성 시그니처의 일부분과 PIP를 나타낸 그래프.
- 도 14는 본 발명에 따른 (a) 등급 2의 조직세포 영상과, (a) 이형성 영역의 영상.
- 도 15는 본 발명에 따른 정상관과 등급 1의 상피 세포의 CytoplasmLength 측정 결과(파란색 라인) 영상.
- 도 16은 본 발명의 실험에 따른 디지털 슬라이드의 정보를 나타낸 표.
- 도 17은 본 발명의 실험에 따른 획득된 실험 영상의 개수를 나타낸 표.
- 도 18은 본 발명의 실험에 따른 분류 실험에서 사용된 형태학적 특징들에 대한 표.
- 도 19는 본 발명의 실험에 따른 분류 실험들에 사용된 특징 집합들에 대한 심별과 차원에 대한 표.
- 도 20은 본 발명의 실험에 따른 SVM 분류기를 학습하고 평가하기 위한 학습 데이터와 테스트 데이터 집합의 개수에 대한 표.
- 도 21은 본 발명의 실험에 따른 각 특징 집합에 대한 정상관과 PDAC를 분류하기 위한 평가 결과 표.
- 도 22는 본 발명의 실험에 따른 각 특징 집합에 대한 정상관과 PDAC에 대한 분류 정확도의 비교 그래프.
- 도 23은 본 발명의 실험에 따른 케이스 1의 분류기를 위한 ROC 곡선과 AUC 값의 비교 그래프로써, (a) ROC curve and AUC of CNF; (b) ROC curves and AUC values for CLF, PLF, and ALF; (c) ROC curves and AUC values for CEF, PEF, and AEF; (d) ROC curves and AUC values for CDF, PDF, and ADF; (e) ROC curves and

AUC values for CTF, PTF, and ATF에 대한 그래프.

도 24는 본 발명의 실험에 따른 각 특징 집합에 대한 등급 1과 등급 2 사이를 구별하는 평가 결과표.

도 25는 본 발명의 실험에 따른 각 특징 집합에 대한 등급 1과 등급 2에 대한 분류 정확도의 비교 그래프.

도 26은 본 발명의 실험에 따른 케이스 2의 분류기를 위한 ROC 곡선과 AUC 값의 비교 그래프로써, (a) ROC curve and AUC of CNF; (b) ROC curves and AUC values for CLF, PLF, and ALF; (c) ROC curves and AUC values for CEF, PEF, and AEF; (d) ROC curves and AUC values for CDF, PDF, and ADF; (e) ROC curves and AUC values for CTF, PTF, and ATF에 대한 그래프.

도 27은 본 발명의 실험에 따른 F-test의 결과와 비상피세포 특징에 대한 통계표.

도 28은 본 발명의 실험에 따른 F-test의 결과와 내강 특징에 대한 통계표.

도 29는 본 발명의 실험에 따른 F-test의 결과와 상피세포 특징에 대한 통계표.

도 30은 본 발명의 실험에 따른 비상피세포 특징의 LSD-test와 F-test에 대한 표.

도 31은 본 발명의 실험에 따른 내강 특징의 LSD-test와 F-test에 대한 표.

도 32는 본 발명의 실험에 따른 상피세포 특징의 LSD-test와 F-test에 대한 표.

### 발명을 실시하기 위한 구체적인 내용

[0042] 이하, 본 발명의 실시를 위한 구체적인 내용을 도면에 따라서 설명한다.

[0043] 또한, 본 발명을 설명하는데 있어서 동일 부분은 동일 부호를 붙이고, 그 반복 설명은 생략한다.

[0044] 먼저, 본 발명을 실시하기 위한 전체 시스템의 구성에 대하여 도 1을 참조하여 설명한다.

[0045] 도 1에서 보는 바와 같이, 본 발명에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법은 컴퓨터 단말(20) 상의 프로그램 시스템(30)으로 실시될 수 있다. 이때, 프로그램 시스템을 췌장선암 등급 진단 시스템(30)이라 부르기로 한다.

[0046] 즉, 췌장선암 등급 진단 시스템(30)의 각 기능들은 컴퓨터 프로그램으로 구현되어 컴퓨터 단말(20)에 설치되어, 사용자에 의해 명령 등을 컴퓨터 단말(20)의 입력장치를 통해 입력받거나, 추출된 형태학적 특징, 진단 결과 등을 컴퓨터 단말(13)의 출력장치를 통해 출력한다. 한편, 췌장선암 등급 진단 시스템(30)에서 필요한 데이터들은 컴퓨터 단말(20)의 하드디스크 등 저장공간에 저장되어 이용된다.

[0047] 췌장선암 등급 진단 시스템(30)에 입력되는 데이터는 조직세포 영상(10)으로서, 특히, 췌장 부분을 촬영한 영상 또는 이미지이다.

[0048] 다른 실시예로서, 췌장선암 등급 진단 시스템(30)은 마이크로 프로그램으로 구성되어 마이크로프로세서에 의해 구동되는 하나의 전용 IC칩으로 실시되거나, ASIC(주문형 반도체) 등 하나의 전자회로로 구성되어 실시될 수 있다. 즉, 소프트웨어 형태, FPGA 칩이나 여러 개의 회로소자로 구성된 전자회로의 형태로 구성될 수도 있다. 그 외 가능한 다른 형태도 실시될 수 있다. 그러나 이하에서 설명의 편의를 위해 컴퓨터 단말(20)에 구현된 췌장선암 등급 진단 시스템(30)으로 설명하기로 한다.

[0049] 다음으로, 본 발명에 따른 형태학적 특징 추출 방법을 설명하기에 앞서, 특징 추출의 대상인 췌장선암(PDAC, 췌장관 선암종)에 대하여 도 2를 참조하여 먼저 설명한다.

[0050] 췌장암은 소화기 암으로 인한 사망 원인 중 결장암 다음가는 암종(neoplasm)으로서, 서구에서는 전체 암으로 인한 사망 원인 중 5위를 차지하고 있다[비특허문헌 32]. 이중 췌장선암은 전체 췌장 종양(neoplasm)들 중 85~90%를 차지하고 있다. 주로 80%가 60~80세 환자들에게서 나타나며, 40세 이하 사람들에게서는 매우 드물게 나타난다. 성별로는 여성이 남성보다 발병률이 약간 더 높게 나타나고 있다. 인종별로는 흑인이 가장 높은 발병률을 보인다[비특허문헌 33]. 췌장선암은 치료에 있어서 수술적 치료가 최선의 치료방법으로 알려져 있다. 그러나 췌장선암의 조기진단은 어렵기 때문에 완치를 기대할 수 있는 근치적 절제술의 대상이 되는 환자는 약 5~22%정도이다. 따라서 수술 전 병기의 정확한 결정은 치료방침과 예후를 결정하는 핵심적인 요소가 된다[비특허문헌

34].

- [0051] 췌장선암은 관의 구조적인 특징과 핵의 형태학적인 특징들에 의해 등급 1(Well-differentiated carcinomas), 등급 2(Moderately differentiated carcinomas), 등급 3(Poorly differentiated ductal adenocarcinomas)들로 구분된다.
- [0052] 먼저 등급 1(Well differentiated carcinomas)들은 관유사(duct-like) 구조를 가지며, 중간 크기의 분비선종양(neoplastic gland)들로 결합된다. 일반적으로 통모양(Tubular) 혹은 체(Cribriform) 형태를 보이며, 관에서 불규칙한 유두상을 보일 수 있다. 세포의 유사분열(mitotic activity)은 적게 나타난다. 세포는 원주(columnar) 경향을 보이며, 점액을 생성한다. 그리고 상피세포 중 몇몇은 극성을 잃어 버린다.
- [0053] 등급 2(Moderately differentiated carcinomas)는 대개 중간 크기의 관유사 모양(duct-like)과 통모양(tubular) 구조의 혼합된 형태를 보이며, 결합조직형성 스트로마(Desmoplastic stroma)에 존재한다. 관의 모양은 불완전하게 형성된 분비선(gland)들이 일반적이다. 등급 1과 비교해서 핵의 크기가 크고, 유사분열(mitotic activity)이 보다 빈번하게 나타나고 점액의 생성은 감소한다.
- [0054] 등급 3(Poorly differentiated ductal carcinomas)은 빈번하게 발생하지는 않는다. 이것은 작은 불규칙한 분비선(gland)과 솔리드 튜머 셀 시트(solid tumour cell sheet)와 네트(net)들의 혼합된 형태를 보인다. 일반적으로 관의 형태는 크고 관유사 모양(duct-like)구조와 관내 종양 요소(intraductal tumour component)들은 잘 나타나지 않는다. 세포에서는 괴사(necrosis)와 출혈(haemorrhage)이 발생하며, 활발한 유사분열(mitotic activity)을 보인다. 암종(Carcinoma)이 점점 진행되면서 분비선(gland)과 주변 조직은 작은 종양세포들의 군집에 의해 침습된다.
- [0055] 도 2는 췌장선암의 등급 1과 등급 2, 그리고 등급 3에 대한 조직세포 영상을 보여주고 있다.
- [0056] 위에서 설명하듯이 등급 3은 잘 발생하지 않으며, 형태학적으로도 등급 1과 등급 2에서 확인한 차이를 보인다. 따라서 본 발명에서는 정상관의 조직과, 등급 1, 그리고 등급 2의 감별진단 하는 것을 주 목표로 한다.
- [0057] 다음으로, 본 발명의 일실시예에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법을 도 3과 4를 참조하여 설명한다. 도 3과 4는 본 발명에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법을 흐름으로 도식화한 도면과 흐름도이다.
- [0058] 도 3 및 도 4에서 보는 바와 같이, 본 발명에 따른 형태학적 특징 기반 췌장선암 등급 진단 방법은 다음과 같이 3개의 단계로 구성된다: (a) 조직세포의 영상처리 및 특징추출 단계(S100), (b) 진단 모델 구성 및 검증 단계(S200), (c) 학습된 모델을 이용한 진단 단계(S300).
- [0059] 첫 번째 단계인 영상처리 및 특징 추출 단계에서는 진단시 필요한 정량적인 특징들을 추출하기 위해, 주어진 조직세포 영상을 3부분으로 세그멘테이션 하고, 각 부위의 특징들을 추출되고 특징 데이터베이스에 저장된다. 두 번째 단계는 모델 학습 및 검증 단계로서 첫 번째 단계에서 구성된 특징데이터베이스를 이용하여 진단모델을 학습하고 검증한다. 그리고 마지막 단계는 생성된 진단 모델을 이용하여 조직세포 영상에 대한 진단을 수행한다.
- [0060] 먼저, 본 발명의 일실시예에 따른 췌장선암의 등급을 위한 형태학적 특징 추출 방법을 도 5를 참조하여 설명한다. 형태학적 특징 추출 방법은 앞서 조직세포의 영상처리 및 특징추출 단계(100)에 대한 구체적인 방법이다.
- [0061] 도 5에서 보는 바와 같이, 췌장선암의 등급을 위한 형태학적 특징 추출 방법(또는 과정)(S100)은 내강영역 세그멘테이션 단계(S10); 핵 식별 단계(S20); 상피세포 핵 구분 단계(S30); 내강 특징 추출 단계(S40); 및, 상피세포 추출 단계(S50)로 구성된다.
- [0062] 본 과정은 주어진 조직세포 영상을 세가지 타입의 오브젝트들로 식별한다. 3가지 오브젝트 타입은 관을 구성하는 내강과 상피 핵이고, 나머지 하나는 비-상피 핵이다. 도 3b는 조직 이미지를 3가지 오브젝트 타입으로 영상을 세그멘테이션하는 과정을 보여준다. 각 오브젝트 타입의 식별 방법은 이하에서 설명한다.
- [0063] 먼저, 내강(Lumen) 세그멘테이션 단계(S10)에 대하여 설명한다.
- [0064] 본 발명에서 내강 영역 식별은 씨드 영역 양성("Seeded" region growing, SRG) 방법[비특허문헌 35]을 이용하였

다. SGR(seeded region growing) 방법은 초기 시작점을 요구한다. 이전 연구[비특허문헌 36]에서 내강 영역의 후보 씨드점(candidate seed points)을 찾음으로써 SRG 방법을 이용한 내강 영역 식별을 자동화 하였다.

[0065] 본 발명의 씨드 영역 양성(SGR) 단계(S10)는 구체적으로, (a1) 이진 영상 획득 단계(S11); (a2) 방향 누적맵 생성 단계(S12); (a3) 후보 씨드점 선별 단계(S13); 및, (a4) 후보 씨드점으로 SRG 방법을 적용하여 내강 영역의 경계를 구하는 단계(S14)로 구성된다.

[0066] 우선 SRG가 보다 쉽게 적용할 수 있도록 주어진 영상에 중간값 필터링(median filtering)과 배경 상관(background correlection)[비특허문헌 37] 알고리즘을 적용한 후 최대 엔트로피 문턱치(maximum entropy treshold)[비특허문헌 38]을 적용하여 이진 영상 A를 만든다(S11).

[0067] 이후, 이진 영상 A로부터 씨드점(seed point)을 얻기 위해 방향 누적맵(Direction Cumulative Map) H(A)을 생성한다(S12). 이 맵은 전처리된 이진 영상 A의 각 방향(left, right, up, and down)에 대해 흰색 픽셀들만을 누적하고, 그 누적 값들의 제곱근의 합을 구함으로써 각 내강 영역의 중심에 대응되는 위치에서 큰 값을 가지게 된다.

[0068] 이때, H(A)의 지역 최대점(local maximum point)들은 SRG 알고리즘의 후보 씨드점으로 사용된다(S13). 그러나, H(A)로부터 바로 후보 씨드점을 구하게 되면 불필요하게 작은 영역에 대해서도 후보 씨드점들이 생성된다. 이러한 문제를 해결하기 위해 H(A)의 작은 값들 문턱치(threshold)로 처리 한  $H_T(A)$ 로부터 후보 씨드점들을 구한다. 문턱치 처리 방법은 오프(Otsus) 방법이 사용되었다.

[0069] Otsu 방법은 다양한 문턱치 처리 방법들 중 일반적으로 사용되는 문턱치 처리 방법이다. 이 방법은 클래스 간 분산을 최대화하는 지점을 문턱치 값으로 선택한다. 즉, Otsu 방법은 두 클래스간 분산을 최대화하도록 문턱치 값을 결정하는 방법이다. 영상의 각 픽셀 값이 선택된 문턱치 값 이상이면 1, 미만이면 0으로 할당한다. 불필요한 작은 영역들을 제거하기 위해 Otsu 방법을 이용한다.

[0070] 이제 얻어진 후보 씨드점들을 SRG 알고리즘의 초기 파라미터 이용하여 내강 영역을 식별한다(S14). 이때 식별된 내강 영역의 경계를  $B^0$ 로 표현한다.

[0071] 도 6는 지금까지 설명된 내강 영역 식별에 대한 과정을 시각적으로 보여준다. 도 6(a)는 내강 경계를 식별하기 위하여 전처리된 이진 영상이고, 도 6(b)는 영상을 위한  $H_T(A)$ 와 후보 씨드점(노란점)을 나타내고 있다. 이때, 이 영상은 시각화를 위해 0 ~ 255로 스케일링되었다. 도 6(c)는  $H_T(A)$ 를 위한 3차원 그래프를 나타낸 것이고, 도 6(d)는 후보 씨드(녹색선)에 의해 식별된 내강 경계를 나타내고 있다.

[0072] 이 단계에서, 식별한 내강 영역의 경계를 오리지널 내강 경계라 부르기로 한다.

[0073] 다음으로, 상피 핵과 비-상피 핵의 식별하는 과정을 설명한다.

[0074] 본 과정에서 주어진 조직 이미지(또는 조직세포 영상)의 핵들을 식별하고(S20), 그들을 상피 핵과 비-상피 핵으로 구분한다(S30). 그 과정은 다음과 같다.

[0075] 먼저 핵 식별 단계(S20)를 설명한다.

[0076] 본 단계는 주어진 조직 영상에 있는 모든 핵들을 식별한다(S20). 먼저 주어진 조직 영상에서 핵을 식별하는데 불필요한 부분인 세포질과 같은 부분을 k평균(k-means)에 기반한 컬러 문턱치 처리(color thresholding)[비특허문헌 39]에 의해 제거한다.

[0077] 바람직하게는, 이후 문턱치 처리된 이미지에 있는 핵들에 존재하는 홀(hole)들을 홀채움(Hole-Filling) 알고리즘[비특허문헌 40]을 적용한다. 핵에 홀이 존재하는 경우는 잘 발생하지 않는다. 가끔 염색이 잘 되지 않은 핵들(핵이 부분적으로 연하게 염색되는 경우)은 문턱치 처리에 의해 홀이 발생할 수 있다. 이때 이 부분들을 메우기 위해 홀 채움 알고리즘을 적용한다. 이 과정은 생략해도 되나, 발생할지 모르는 경우를 대비하기 위한 과정이다.

[0078] 마지막으로 와더셰드(Wathershed) 알고리즘[비특허문헌 41]을 적용하여 핵(nuclei)들을 분리한다. 이때 분리된 핵들을 핵집합 N으로 정의한다. 와더셰드 알고리즘은 영상을 분할하는 방법 중의 하나이다. WaterShed 알고리즘은 객체들이 인접해 있는 경우 다른 영상 분할 알고리즘들에 비해 효과적으로 객체들을 분할하는 방법으로 알려

져 있다. 따라서 객체들의 중첩이 빈번한 세포영상에서 핵들을 분리하는데 가장 많이 사용되는 방법이다.

[0079] 도 7(a)는 식별된 핵들의 집합  $\mathcal{N}$ 을 보여준다. 도 7(b)는 이하의 단계에서 수행하여, 핵집합  $\mathcal{N}$ 에서 선택된 상피 핵 집합  $\mathcal{N}^E$  (빨간색으로 마크됨)을 보여준다.

[0080] 다음으로, 상피 세포 핵과 비-상피세포 핵의 구분 단계(S30)를 설명한다. 본 단계에서는 식별된 핵들을 상피 핵 집합과 비상피세포 핵 집합으로 분리한다.

[0081] 상피세포는 내강을 에워싸고 있다. 따라서 상피세포 핵은 식별된 내강 경계  $B^0$ 에 인접한 핵들을 식별된 핵 집합  $\mathcal{N}$ 으로부터 선별함으로써 선택될 수 있다. 상피 세포 핵 집합은  $\mathcal{N}^E$ 와 표기되고 다음과 같이 정의된다.

[0082] [수학식 1]

$$\mathcal{N}^E = \bigcup_{p \in B^0} \left\{ n \mid \arg \min_{n \in \mathcal{N}} (\text{distance}(\text{centroid}(n), p)) \right\}$$

[0083]

[0084] 여기서  $p$ 는 내강 경계를  $B^0$ 를 구성하는 포인트이고,  $\text{centroid}(-)$ 는 식별된 핵  $n$ 의 중심 포인트를 반환하는 함수이다. 그리고  $\text{distance}(\cdot, \cdot)$ 는 주어진 두 점들 사이의 유클리디언 거리를 반환하는 함수이다.

[0085] 즉, 핵 집합  $\mathcal{N}$ 으로부터 상피 핵의 분리는  $B^0$ 에 속한 각 포인트  $p$ 에서 가장 가까운 핵을 선택하고, 선택된 핵을 상피-핵 집합  $\mathcal{N}^E$ 에 포함시킴으로써 구현된다. 도 8의 알고리즘 1은 이 과정을 보여준다.

[0086] 비상피 세포 핵은 식 2와 같이 식별된 핵 집합에서 식별된 상피세포 핵을 제외시킴으로써 구해진다.

[0087] [수학식 2]

$$\mathcal{N}^{NE} = \mathcal{N} - \mathcal{N}^E$$

[0088]

[0089] 도 9의 표는 본 발명에서 사용되는 주요 표기들을 요약한다.

[0090] 다음으로, 형태학적 특징을 추출하는 과정을 설명한다. 형태학적 특징 추출 과정은 세그먼트된 내강 영역에 대하여 내강 특징을 추출하고(S40), 구분된 상피세포 핵들로부터 상피세포 특징을 추출한다(S50).

[0091] 조직 영상에서 PDAC를 진단하기 위한 주요 부위는 관(duct)이다. 앞에서 설명된 것처럼 PDAC는 관의 분화도에 따라 등급 1, 등급 2, 그리고 등급 3으로 분류된다. 관은 내강과 그것을 에워싸는 상피세포로 구성된다. 따라서 세그먼트된 내강과 상피세포 핵들로부터 PDAC진단을 위한 형태학적 특징들을 추출하기 위한 방법들에 대해서 설명한다.

[0092] 먼저, 내강 특징(Lumen Features) 추출 단계(S40)에 대하여 설명한다.

[0093] PDAC에서 관은 유두상을 보이고, 상피세포는 극성을 잃어 불완전한 분비선(gland) 형태를 보인다. 병기가 진행됨에 따라 관의 모양은 복잡하고 이형성 정도는 점점 더 심해진다. 본 단계에서는 관의 이형성을 표현하는 방법과 이형성을 정량적으로 측정하는 형태학적 특징 추출방법들에 대해 설명한다.

[0094] 내강 특징 추출 단계(S40)는 이상적인 경계를 구하는 단계(S41); 이형성 진폭 시그니처를 구하는 단계(S42); 및 이형성 진폭 시그니처로부터 내강 특징을 추출하는 단계(S43)로 구성된다.

[0095] .

[0096] 관의 이형성(Atypia of Duct)의 표현에 대하여 설명한다.

[0097] 일반적으로, 정상적인 관의 내강은 이형성이 거의 나타나지 않기 때문에 볼록껍질(convex hull)과 같이 보인다. 그에 비해 PDAC의 관은 병기가 진행됨에 따라 내강 경계의 복잡한 변화와 함께 보다 복잡한 관의 이형성을 보인다.

다. 이러한 관점으로부터 우리는 주어진 판의 이상적인 내강 경계를 추정하고, 오리지널 내강 경계와의 비교를 통해 오리지널 내강의 이형성을 표현한다.

[0098] 먼저, 이상적인 내강 경계를 구한다(S41).

[0099] 이상적인 내강 경계는, 오리지널 내강 경계로부터 볼록껍질(또는 볼록껍질 경계)을 구하되, 오리지널 내강의 크기로 볼록껍질을 스케일링하여 구한다.

[0100] 본 단계에서 오리지널 내강의 경계는  $B^0$ 로 표현하고 추정되는 이상적인 내강 경계는  $B^I$ 로 표현한다.  $B^0$ 와  $B^I$ 는 각각의 내강 경계를 구성하는 포인트들의 시퀀스이다. 이상적 내강 경계를 추정하기 위한 절차는 다음과 같다.

[0101] 우선 오리지널 내강경계  $B^0$ 의 볼록껍질을 구한다. 이때 구한  $B^0$ 의 볼록껍질은  $B^C$ 로 표기한다. 볼록껍질이란 주어진 점들을 포함하는 가장 작은 볼록(convex)다각형으로 정의된다. 즉, 최외각의 한점을 기준으로 해서(점들을 감싸는 최소 사각형상의 한점) 그 자신을 뺀 모든 점들이 왼편이나 일직선상에 놓일수 있는 점을 찾는다. 다시 이 찾은 점을 기준으로 나머지 점들에 대해서 동일한 조건을 만족하는 점들을 차례로 찾으면, 주어진 점들을 감싸는 최소의 점들의 모임을 구할 수 있고, 이것이 볼록껍질(convex hull)을 구성한다.

[0102] 이때,  $B^C$ 를 경계로 하는 영역은  $B^0$ 를 경계로 하는 영역크기 때문에, 이상적인 내강계  $B^I$ 는  $B^C$ 를 경계로 하는 영역을  $B^0$ 가 경계로 하는 영역의 크기로 줄인 영역의 경계로서 구한다.

[0103]  $B^C$ 를 경계로 하는 영역을  $B^0$ 가 경계로 하는 영역으로 줄이기 위한 스케일링 계수  $s$ 는 식 3과 같다. 즉,  $B^0$ 와  $B^C$ 의 면적을 이용하여 이상적인 내강 경계  $B^I$ 를 구하기 위한 스케일링 계수를 구한다.

[0104] [수학식 3]

$$s = \sqrt{\frac{Area(R^0)}{Area(R^C)}}$$

[0105]

[0106] 여기서  $R^0$ 는  $B^0$ 를 경계로 하는 영역의 의미하고  $R^C$ 는  $B^C$ 를 경계로 하는 영역을 의미한다.  $Area(\cdot)$ 는 영역의 크기를 반환하는 함수이다. 내강  $B^I$ 는  $R^C$ 의 중심(center)에 관해서  $s$ 로 스케일된 영역의 경계를 구성하는 포인트들의 집합이다.

[0107] 도 10은 오리지널 내강 경계  $B^0$ (녹색)와 이상적인 내강 경계  $B^I$ (붉은색)를 보여준다.

[0108] 다음으로, 이형성 진폭 시그니처(Atypia-Amplitude Signature)를 구한다(S42)

[0109] 본 단계에서는 오리지널 내강 경계와 이상적인 내강 경계를 이용하여 내강의 이형성을 표현하는 1-D 시그니처를 제안한다. 제안된 1-D 시그니처는 이상적인 내강과 오리지널 내강 사이의 이형성 진폭을 측정함으로써 오리지널 내강에 존재하는 이형성을 시각화한다. 이때, 이형성 진폭은 이상적인 내강과 오리지널 내강 사이의 부호를 가지는 수직 거리로서 이형성-진폭 함수  $A(t)$ 에 의해 측정된다.

[0110] 식 4는 이형성 진폭함수의 정의를 보여준다.

[0111] [수학식 4]

$$A(t) = \text{sgn}_t(q) \times \min_{p_t \in B^I, q \in B^0} (\text{distance}(p_t, q))$$

such that  $p_t \perp q$

[0112]

[0113] 여기서  $t$ 는 이상적인 내강 경계에 있는 포인트들의 순서를 가리키는 인덱스 변수이고,  $\text{sgn}_t(q)$ 는  $p_t \in B^1$ 에서 수직 위치에 있는 점  $q \in B^0$  사이의 거리에 대한 부호를 나타내는 함수로서 점  $q$ 가 점  $p_t$ 의 바깥 경계에 위치해 있으면 +1, 안쪽 경계에 위치해 있으면 -1을 반환한다.

[0114] 즉, 상기 이형성 진폭은 이상적인 경계 상의 포인트에서 수직에 위치하는 오리지널 경계 상의 포인트 사이의 거리이되, 부호를 가진 거리로 표현된다.

[0115] 도 11(a)는  $p_t \in B^1$ 에서 이형성-진폭 측정과정을 보여준다. 이 예에서  $q$ 는  $p_t$ 의 바깥 경계에 있는 점이므로 양의 값으로  $A(t)$ 가 측정된다.

[0116] 본 발명에서 이형성-진폭 함수  $A(t)$ 를 이용한 1-D 시그니처를 이형성 진폭 시그니처로 명명한다. 이형성 진폭 시그니처(Atypia-Amplitude Signature)는 이상적인 내강 경계의 원점( $p_0$ )에서  $p_t$ 까지의 둘레 길이  $L(t)$ 를  $x$ 좌표 그리고  $B^1$ 의 포인트  $p_t$ 에서 구한 이형성 진폭  $A(t)$ 를  $y$ 좌표로 하는 포인트 ( $L(t)$ ,  $A(t)$ )에 의해 그려진다(도 12(b)). 여기서 원점( $p_0$ )에서 점  $p_t$ 까지의 둘레 길이를 표현하는  $L(t)$ 는 식 5와 같이 각 포인트 사이의 거리 합으로 정의된다.

[0117] 도 12의 테이블은 도 1의 (a), (b), (c)에 대한 이형성 진폭 시그니처를 보여준다.

[0118] 다음으로, 내강 영역에 대한 관의 이형성 측정 특징들을 추출한다(S43).

[0119] 본 단계에서는 이전 단계에서 구한 이상적 내강 경계와 이형성-진폭 시그니처를 이용하여 내강의 이형성을 정량적으로 측정하는 특징들을 설명한다. 제안되는 특징들은 RMSAA(Root Mean Square Atypia Amplitude), TSAV(Total Sum of Atypia Volatility), 이형성 비율(AtypiaRatio), 이형성 영역의 개수(#AtypiaRegions)이다.

[0120] 먼저, RMSAA(Root Mean Square Atypia Amplitude)에 대하여 설명한다.

[0121] RMSAA는 이전 단계에서 구한 이형성 진폭 시그니처로부터 측정한다. 이것은 이형성 진폭  $A(t)$ 들에 대한 제곱 평균(the average squared atypia amplitude)의 제곱근(square root)이다.  $A(t)$ 는 추정된 이상적 내강 경계  $B^1$ 에서 오리지널 내강 경계  $B^0$ 사이의 수직 거리(called by atypia-amplitude)로서 오리지널 내강  $B^0$ 가 추정된 정상 내강  $B^1$ 로부터 얼마나 이탈되었는지를 표현한다.

[0122] 따라서 측정된  $A(t)$ 는 회귀 분석에서 관찰(observation)과 추정된 함수 값 사이의 차이를 나타내는 나머지(residual or error)로 해석될 수 있다. 마찬가지로 MSAA 회귀식의 평균 에러를 측정하는 MSE(Root Mean Square Error)[비특허문헌 42]에 대응된다. RMSAA는 식 6으로 정의된다.

[0123] [수학식 6]

$$\begin{aligned} RMSAA &= \sqrt{MSAA} \\ &= \sqrt{\frac{1}{m} \sum_{t=1}^m A(t)^2} \end{aligned}$$

[0124]

[0125] 여기서  $m$ 은  $B^1$ 에 있는 포인트들의 수이다.

[0126] TSAV(Total Sum of Atypia Volatility)에 대하여 설명한다.

[0127] 웨장선암은 병기가 진행될 수록 내강은 복잡해지기 때문에 내강의 경계변화도 더 복잡하게 변한다. TSAV는 병기

의 진행에 따라 복잡해 지는 내강의 경계변화를 측정한다. TSAV의 측정은 이형성-진폭 시그니처의 주요 변곡점을 찾고, 그 지점들에서의 변화량의 합으로 계산된다.

[0128] 본 발명에서는 이형성-진폭 시그니처의 주요 변곡점을 찾기 위해 PIP 알고리즘[비특허문헌 3],[비특허문헌 4]를 이용하였다. PIP 알고리즘은 시계열 데이터(Time Series Data)로부터 주요한 트렌드를 표현하는 중요한 포인트들을(PIP) 발견하는 방법이다.

[0129] 본 발명에서는 이형성-진폭 시그니처에 존재하는 모든 변곡점들을 찾기 위해, 고정된 개수의 PIP를 탐지하는 기존 PIP알고리즘을 수정하였다.

[0130] 도 13은 도 12에 있는 등급 2에 대한 이형성 시그니처의 일부분과 수정된 PIP 알고리즘에 의해 식별된 PIP들을 보여준다.

[0131] 도 13에서, PIP에서의  $AV_i$  인  $p_i$ 는  $a_i$ 와  $b_i$  사이의 각도( $\theta_i$ )로 측정된다.

[0132] TSAV는 식별된 PIP들의 각 지점에서 AV를 측정하고, 그 값들의 총합으로 계산된다. 이때,  $i$ 번째 PIP <sub>$i$</sub> 에서  $AV_i$ 는 두 벡터  $a_i = PIP_i - PIP_{i-1}$ 와  $b_i = PIP_{i+1} - PIP_i$ 에 사이의 사이 각( $\theta_i$ )로 정의된다. 도 13은 PIP <sub>$i$</sub> 에서  $AV_i$ 가 어떻게 측정되는지를 보여준다.

[0133] [수학식 7]

$$TSAV = \sum_{i=2}^{m-1} AV_i$$

$$= \sum_{i=2}^{m-1} \left( \arccos \left( \frac{\mathbf{a}_i \cdot \mathbf{b}_i}{\|\mathbf{a}_i\| \|\mathbf{b}_i\|} \right) \right)$$

$$(\mathbf{a}_i = p_i - p_{i-1}, \mathbf{b}_i = p_{i+1} - p_i)$$

[0134] 여기서  $m$ 은 이형성 진폭 시그니처로부터 식별된 PIP들의 개수이다.

[0136] 이형성 비율과 이형성 영역의 개수에 대하여 설명한다.

[0137] 관샘압(채장선압, 취장관샘압)은 병기가 진행됨에 따라 관모양은 복잡해 지고 뚜렷한 유두상을 보인다. 병기가 진행된 실제 내강 영역  $R^0$ 은 이상적인 관 영역  $R^1$ 을 벗어나거나 침습한다. 따라서 본 단계에서는 원래 관 영역과 이상적인 관 영역에 집합 연산(set operation)을 적용하여 이형성 영역들을 구하고, 원래 관에서의 이형성 비율과 이형성 개수를 카운팅 한다.

[0138] 이 두 특징을 구하기 위해 이형성 영역들의 집합  $AR = \{ar_1, ar_2, \dots, ar_m\}$ 을 구한다.  $AR$ 은  $(R^1 \cup R^0) - (R^1 \cap R^0)$ 를  $B^1$ 에 의해 나누지는 영역들의 집합으로 정의된다. 도 14는 등급 2에서 식별된 이형성 영역들을 보여준다.

[0139] 도 14에서, (a)는 등급 2의 조직 세포 영상이고, (b)는 (a)의 영역  $R^0$ 와  $R^1$ 에 의해 생성된 이형성 영역이다.

[0140] 이제 획득된 이형성 영역들로부터 실제 관에서 이형성 영역이 차지하는 비율(AtypiaRatio)과 이형성 영역의 개수(#AtypiaRegions)를 구한다.

[0141] [수학식 8]

$$AtypiaRatio = \frac{\sum_{i=1}^m Area(ar_i)}{Area(R^0)}$$

[0142]

[0143] [수학식 9]

$$\#AtypiaRegions = \left| \left\{ ar \mid ar \in AR, Area(ar) > Threshold \right\} \right|$$

[0144]

[0145] 여기서  $Area(\cdot)$ 는 주어진 영역의 크기를 리턴하는 함수이다.  $m$ 은  $AR$ 집합의 엘리먼트 개수를 의미한다.  $AtypiaRatio$ 는 관이 전반적으로 얼마나 많이 찌그러져 있는지를 나타낸다.  $\#AtypiaRegions$ 는 실제 관 영역이 이상적인 관 영역을 벗어나거나 침습하는 이형성 영역들의 개수를 카운팅함으로써 유두상 정도를 표현한다. 이때 유두상을 표현하기에는 너무 작은 영역들은 쓰레숄딩을 적용하여 카운팅에서 제외한다.

[0146] 바람직하게는, 쓰레숄딩 값(또는 문턱치)은  $300 \mu m^2$ 으로 설정한다.

[0147] 다음으로, 상피세포 핵들로부터 상피세포 특징(Epithelial Cell Feature)을 추출한다(S50).

[0148] 관을 구성하는 두 번째 요소는 상피세포들이다. 많은 경우 관샘암에서 상피세포들은 점액소(mucin)을 생성하며, 원주형 모양(columnar) 경향과, 핵의 극성을 잃어버리는 특징을 보인다[비특허문헌 33][비특허문헌 45]. 본 단계에서 식별된 상피세포를 이용하여, 이러한 특징들을 추출하는 방법을 설명한다.

[0149] 먼저, 상피 세포질 길이(Epithelial Cytoplasm Length)의 특징에 대하여 설명한다.

[0150] 일반적으로 정상조직의 관은 입방형 상피세포들로 에워싸여있다. 반면, PDAC의 관은 풍부한 세포질로 인해 원주형 상피세포로 에워싸여있다. 원주형세포의 핵은 달걀 모양을 가진다. 풍부한 세포질을 가지는 원주형 상피세포의 세포질 길이는 입방형 상피 세포에 비해 길다.

[0151] 따라서 상피세포의 세포질 길이의 측정은 상피세포가 원주형 세포인지 아닌지를 표현한다. 종래 논문[비특허문헌 36]에서 상피세포들의 세포질 길이를 측정하는 세포질 길이(Cytoplasm Length) 특징을 사용한다. 세포질 길이 특징은 오리지널 내강 경계  $B^0$ 와 상피 핵  $n$ 사이의 수직 거리이다. 식 16은 세포질 길이 특징  $CytoplasmLength$ 의 수식을 나타낸다.

[0152] [수학식 10]

$$CytoplasmLength(n) = distance(centroid(n) - q) \\ \text{such that, } centroid(n) \perp q$$

[0153]

[0154] 여기서  $m$ 은 집합  $N^E$ 의 엘리먼트 수이고  $B^0$ 의 점  $q$ 는  $centroid(n)$ 와 수직 위치에 있는 점이다.

[0155] 도 15은 정상과 등급 1의  $CytoplasmLength$  측정 결과를 보여준다. 도 15에서 붉은 점은 상피 핵이고, 녹색 선은 식별된 내강 경계이다. 그리고 식별된 내강 경계와 상피 핵 사이의 파런선은 측정된  $CytoplasmLength$ 이다.

[0156] 다음으로, 세포질 길이 특징( $CytoplasmLength$ )의 표준편차에 대하여 설명한다.

[0157] 본 특징은 채장선암에서 나타는 특징 중 하나인 핵 극성의 손실(the loss of nuclear polarity)를 측정한다. 이 특징이 나타나는 상피 세포들은 내강을 따라 불규칙하게 배열되기 때문에 세포질 길이의 편차가 크다. 반면, 정

상의 경우는 극성 손실이 발생하지 않게 때문에 상피 세포들은 내강을 따라 규칙적으로 배열되며 세포질 길이의 편차는 매우 작다.

[0158] 따라서 상피세포의 세포질 길이에 대한 표준편차를 구함으로써 핵 극성의 손실(the loss of nuclear polarity)을 측정한다.

[0159] [수학식 11]

$$CytoplasmLength_{SD} = \sqrt{\frac{1}{m} \sum_{n \in N^*} (CytoplasmLength - CytoplasmLength(n))^2}$$

[0160]

[0161] 다음으로, 진단 모델 구성 및 검증 단계(S200)에 대하여 설명한다.

[0162] 진단 모델은 지도학습 기반의 분류기를 이용하여 구성된다. 바람직하게는 SVM(Support Vector Machine) 분류기를 이용한다. 각 특징들과 췌장선암의 진단결과(또는 등급)를 쌍으로 하는 학습데이터를 이용하여 학습시켜 진단모델을 구성한다. 그리고 교차 검증을 통해 검증한다.

[0163] 진단 모델은 췌장선암과 병기를 구분하는 '분류기'이다. 즉, 췌장선암의 진단은 곧 췌장선암 여부를 분류하는 것과 같다. 컴퓨터를 활용하여 분류문제를 다루는 방법으로서, 지도학습 기반의 방법들을 이용한다. 바람직하게는, 통계학 기반 지도학습 방법인 SVM(Support Vector Machine)을 이용한다.

[0164] 지도학습 방법은 학습데이터들을 이용하여 진단모델 또는 분류모델(분류기)을 학습시킨다. 여기서 학습데이터는 특징벡터( $\mathbf{x}$ )와 클래스 레이블( $y$ )의 쌍(pair)의 원소로 구성되는 집합이다. 기호로 학습데이터를  $D$ 를 표현하면 다음과 같다.

[0165]  $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}.$

[0166] 특징벡터( $\mathbf{x}$ )는 앞서 구한 췌장선암의 특징들을 벡터화한 것이다.

[0167] 클래스 레이블  $y$ 는 특징벡터  $\mathbf{x}$ 의 실제 클래스 레이블을 가리키는 것으로서, 췌장선암의 진단결과(또는 진단결과와 등급)를 구분한 것이다. 예를 들어 정상과 췌장선암 분류 문제의 경우  $y$ 는 '정상' 혹은 '췌장선암'의 값을 가질 수 있으며, 췌장선암 병기 구분 문제의 경우  $y$ 는 'Grade 1'과 'Grade 2'의 값을 가질 수 있다.

[0168] 이때 진단 모델의 학습은 SVM 학습시 결정되는 학습 파라미터들 ( $C^*, \gamma^*$ )에 의해 많은 영향을 받는다(여기서,  $C$ 는 모델 파라미터이고,  $\gamma$ 는 커널 파라미터이다). 물론, 학습데이터가 가장 중요하지만, 학습데이터를 학습시키는데 이들 파라미터들이 중요한 역할을 한다. 구체적인 예는 도 20을 참조한 본 발명의 실험에 기재된 것을 참조한다.

[0169] 학습된 진단 모델의 검증은  $k$ 겹 교차 검증( $k$ -fold cross validation)을 통해 수행된다.  $k$ 겹 교차 검증은 학습데이터  $D$ 를  $k$ 개의 데이터 집합으로 분할하고  $k-1$ 개의 데이터 집합을 SVM모델을 학습하는데 이용한다. 그리고 남은 한 개의 데이터 집합을 진단모델을 성능을 검증(평가)하는데 사용한다. 이 과정을  $k$ 번 수행하게 된다. 즉, 분할된 데이터집합을 이용하여  $k$ 개의 진단모델을 생성하고,  $k$ 번의 진단모델 성능평가의 평균으로 모델을 검증하게 된다.  $k$ 겹 교차 검증을 위한 학습은 SVM 학습에서 얻어진 학습파라미터 ( $C^*, \gamma^*$ )를 이용한다.

[0170] 다음으로, 학습된 모델을 이용한 진단 단계(S300)에 대하여 설명한다.

[0171] 학습된 진단모델(또는 진단 분류기)을 이용하여 주어진 임의의 조직 영상을 진단하는 과정이다. 즉, 진단 모델(또는 진단 분류기)에 특징벡터를 입력하여, 진단 모델에 의해 나오는 결과값(또는 진단결과와 등급)이 진단결과이다. 즉, 진단 모델은 클래스를 구분하는 하나의 함수와 같다.

[0172] 다음으로, 본 발명의 효과를 실험을 통해 구체적으로 설명한다.

[0173] 본 발명의 실험을 위하여, 21개의 정상 조직 슬라이드와 26개의 PDAC조직 슬라이드를 영남대학교 병리학 교실로

부터 제공받았다. 제공받은 조직 슬라이드들은 Hematoxylin & Eosin 방법으로 염색되었다. 그리고 이들 조직 슬라이드들은 ScanScopeCS 시스템을이용하여 20X배율의 디지털 슬라이드로 변환하였다. 각 디지털 슬라이드 이미지는 획득한 조직에 따라 가변적이다. 도 16는 디지털 슬라이드의 정보를 보여준다.

[0174] 본 발명의 특징들의 평가를 위해, 우리는 획득된 디지털 슬라이드로부터 하나의 관이 포함되도록 실험이미지를 생성했다. 실험 영상은 24bit tiff포맷으로 저장되었고 각 영상의 크기는 관의 크기에 따라 가변적이다. 진단에서 중요한 이슈는 inter- and intra- observer variability 문제이다. 이 문제들은 부정확하고, 비일관적이고, 그리고 바이어스된 진단결과를 초래할 수 있다[비특허문헌 46]. 이러한 문제는 진단 시스템을 학습하고 성능을 평가를 위한 전문가의 ground truth 데이터에서도 역시 나타난다. ground thruth의 variability 문제를 줄이기 위해 잘 알려진 방법은 여러 전문가들이 참여하여 ground thruth를 구성하는 것이다[비특허문헌 47]. 본 발명에서, ground through 평가를 위해 영남대의 병리학교실의 3 명의 병리의사의 도움을 받았다. 디지털 슬라이드로부터 생성된 각 관 영상은 3명의 병리학자들의 협의를 거쳐 조심스럽게 각 클래스를 레이블링 하였다.

[0175] 도 17는 ground thruth 평가를 위해 전문가들에 의해 레이블된 실험영상의 개수를 보여준다.

[0176] 주어진 조직세포 영상을 세 부분(내강, 상피 핵, 비-상피 핵)으로 세그멘테이션하고, 각 부위에서 기존의 전통적인 형태학적 특징들과 함께 관샘암에 특화된 새로운 특징들을 추출하였다. 도 18은 취창선암의 진단을 위해, 실험에서 이용된 특징들이다. 세그먼트된 각 부위에서 추출되는 특징들을 "\*"로 표시하였다. 도 18에서 음영 처리된 행들은 본 발명에서 제안된 관샘암에 특화된 특징들이고, 나머지 행들은 기존의 형태학적 특징들[비특허문헌 5], [비특허문헌 24], [비특허문헌 48], [비특허문헌 49]이다. 상피 핵과 비-상피핵의 경우 하나의 조직세포 영상에 여러 개의 오브젝트들이 있기 때문에 각 오브젝트별로 특징값들을 구하고 평균화함으로써 각 영상의 특징 값들을 표현하였다. 실험영상의 특징추출을 위한 실험 환경은 AMD Athlon II 3GHz CPU와2G RAM, Windows7-64bit에서 진행되었다. 기존 특징추출 방법들과 제안된 특징추출 방법들은 JAVA언어 기반의 공개용 이미지 분석 패키지인 Image J[비특허문헌 50]을 통해 구현 되었다

[0177] 관샘암(PDAC) 진단을 위해 본 발명에서 제시된 형태학적 특징들의 우수성을 입증하기 위해, 기존 특징들과 제안된 특징들에 의해 학습된 분류기 들의 분류성능을 비교하였다. 본 발명에서, 분류기의 학습 바업은 통계적 학습 기법으로서 잘 알려진 SVM[비특허문헌 51] 을 이용하였다. SVM은 VC차원과 경험적 리스크를 최소화하기 때문에 일반적으로 성능이 좋은 것으로 알려져 있다[비특허문헌 52]. 실험은 다음 두 경우에 대한 분류결과를 측정하였다. 첫째는 정상과 관샘암 분류이고, 둘째는 관샘암의 등급 1과 등급 2의 병기 분류이다. 본 발명의 특징들이 분류능력 향상에 얼마나 기여하는지를 측정하기 위해 기존 특징들과 본 발명의 특징들을 조합하여 분류기를 학습하였고, 각 분류기의 분류결과들의 비교를 통해 본 발명의 특징들을 평가하였다. 도 19은 분류 실험들에 사용된 특징 집합들에 대한 심벌과 차원을 보여준다.

[0178] 실험 데이터는 두 실험(Normal vs PDAC와 등급1 vs 등급 2)을 위해 도 19의 각 특징 집합에 따라 생성되었다. 즉, 첫 번째 PDAC진단 실험에서 13개의Data Set 들이 다음과 같이 생성되었다 D(CNF), D(CLF), D(PLF), D(ALF), D(CEF), D(PEF), D(AEF), D(CDF), D(PDF), D(ADF), D(CTF), D(PTF), and D(ATF). 여기서 D(·)는 인자로주어진 특징 feature set으로 구성되는 Data set을 의미한다. D(·)는  $D(\cdot) = \{(x(\cdot)_1, y_1), \dots, (x(\cdot)_m, y_m)\}$ 처럼 표기 된다. 여기서  $x(\cdot)_i$ 는 인자로 주어진 특징 set을 가지는 i번째 특징 벡터이다. 클래스 레이블  $y_i$ 는 $x(\cdot)_i$ 의 클래스 레이블로서 Normal class를 의미하는 -1 이나 PDAC클래스를 의미하는 1 값을 가진다. 마찬가지로 두 번째 PDAC의 병기구분 실험에서도 15개의 Data Set이 생성되었다. 이때 클래스  $y_i$ 가 -1이면 등급 1을 의미하고 1이면 등급2를 의미한다.

[0179] 특징 집합에 따른 SVM의 분류성능을 평가하기 위해 특징 집합별로 생성된 Data set으로부터 training set과 test set을 구성하였다. 학습과 평가를 위한training set과 test set의 비율은 60:40으로 설정하였다. 첫 번째 실험(normal vs PDAC)에서, PDAC실험 데이터는 160개의 PDAC 데이터(등급 1의 데이터와 등급 2의 데이터)로 부터 80개 데이터를 샘플링 함으로써 구성되었다. 정상 샘플의 수가 80개이기 때문에, 분류기의 공평한 평가를 위해 우리는 PDAC 샘플 수를 80개로 제한 하였다.

[0180] 다음 도 20은 실험 1(Classification between the normal and PDAC)과 실험2(Classification between the normal and PDAC)에서 사용되는 training data와 test data의 개수를 보여준다.

[0181] 본 발명에서, SVM분류기는 soft-margin 방법과 RBF커널[비특허문헌 51]이 이용되었다. 따라서 모델 파라미터 C

와 커널 파라미터인  $\gamma$ 가 필요하다. 최적의 분류 파라미터 ( $C^*, \gamma^*$ )쌍은 10 cross validation 방법을 이용하여 트레이닝 집합의 정확도를 최대화하는 쌍을  $(C, \gamma) \in \{10^{-1}, 10^{-0.5}, \dots, 10^4\} \times \{2^{-5}, 2^{-4.5}, \dots, 2^0\}$ 을 그리드 검색을 통해 선택하였다[비특허문헌 53]. 본 발명에서 사용된 실험 영상의 수는 매우 부족하다. 따라서 생성된 분류기의 정확도는 바이어스(biase)될 수 있다[비특허문헌 54]. 통계적으로, 이 문제를 해결하기 위해, 부트스트랩 기법[비특허문헌 55]이 사용된다. 각 특징 집합에 대한 분류기의 unbiased 평가를 위해 부트스트랩 기법을 사용한다. 우선, 부트스트랩 평가를 위해 주어진 특징집합(테이블 6)에 대응되는 데이터 집합  $D(\cdot)$ 로부터 10개의 훈련 집합들과 10개의 테스트 집합들을 생성했다. 10개의 훈련 집합과 테스트 집합들은  $D(\cdot) = \{(L(\cdot)_1, T(\cdot)_1), \dots, (L(\cdot)_{10}, T(\cdot)_{10})\}$ ,  $\forall L(\cdot)_i, T(\cdot)_i \subset D(\cdot)$  ( $1 \leq i \leq 10$ )로 표기 된다. 여기서,  $L(\cdot)_i$  과  $T(\cdot)_i$ 는 각각 트레이닝 데이터와 테스트 데이터를 의미한다. 따라서 각 특징 집합에 대한 분류 성능은 10개의 트레이닝 데이터 집합과 테스트 데이터집합에 대해 개별적으로 최적화된 분류기 평가 결과들의 평균으로 측정된다.

[0182] 실험에서 분류기의 성능평가를 위해 사용된 평가항목은 다음과 같다: rue Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Sensitivity (SN), Specificity (SP), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Accuracy (ACC). TP, TN, FP, FN에 대한 설명은 각 케이스의 실험에서 설명된다. 그 외 나머지 성능평가 항목들은 다음과 같이 정의된다.

[0183] [수학식 12]

$$SN = \frac{TP}{TP + FN}$$

[0184]

[0185] [수학식 13]

$$SP = \frac{TN}{TN + FP}$$

[0186]

[0187] [수학식 14]

$$PPV = \frac{TP}{TP + FP}$$

[0188]

[0189] [수학식 15]

$$NPV = \frac{TN}{TN + FN}$$

[0190]

[0191] [수학식 16]

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

[0192]

[0193] 다음으로, 실험 결과에 대하여 설명한다.

[0194] 먼저, 케이스 1(정상관 vs. PDAC)에 대하여 설명한다.

[0195] 도 21와 도 22는 Normal과 PDAC구분을 위해 각 특징집합에 의해 학습된 분류기의 부트스트랩 평가를 보여준다. 도 21에 있는 괄호 안 수치는 10개의 bootstrap evaluation결과에 대한 표준편차를 나타낸다. 우선 내각 오브젝트로부터 추출된 특징집합에 의해 학습된 분류기들의 비교를 살펴 보자. 이 비교에서, PLF에 의해 학습된 분류기의 정확도는 91.56이었다. 그것은 CLF(73.44%)에 학습된 분류기보다 약 18%정도 높은 정확도를 보였다. CLF와

PLF의 조합으로 구성된 ALF 특징 집합으로 학습된 분류기의 정확도는 87.35로서 PLF에 비해 오히려 정확도가 낮아졌다. 이러한 결과는 제안된 PLF 특징 집합이 PDAC를 진단하는데 보다 적절함을 보여준다. 또한 PLF와 CLF의 조합을 통한 추가적인 성능 향상이 없음을 보여준다. 이것은 Area, Perimeter와 같은 간단한 형태학적 특징들이 PDAC의 복잡한 특성을 표현하는데 적합하지 않음을 보여준다.

[0196] 다음으로 상피 핵으로 부터 추출된 특징 집합들의 분류결과를 살펴보자. 먼저 기존 형태학적 특징들의 집합인 CEF는 72.66의 정확도로 내강의 CLF와 비슷한 분류 성능을 보였다. PEF와 AEF는 87.50%의 동일한 분류 정확도를 보였다. 하지만 AEF와 함께하는 분류기의 경우 분류 정확도의 표준편차가 1.47로서 3.21의 표준편차를 가지는 PEF에 비해 좀더 안정 적인 성능을 보이고 있다. AEF가 제안된 PEF와 기존 CEF의 조합인것을 고려했을 때, AEF의 정확도는 제안된 PEF에 기인한다고 볼 수 있다.

[0197] 또한, 주목할만한 것은 PEF의 특징 차원이 2차원이라는 것이다. 이러한 결과는 PEF가 PDAC를 식별하는데 매우 적절함을 보여준다. 또한 비용 측면에서도 매우 효과적임을 보여준다. PDAC진단에서, 관은 진단시 중요한 관심 영역이다. 관은 내강과 상피세포로 구성된다. 따라서, 내강과 상피 세포로 부터 추출된 특징들의 조합이 의미 있을 것이다. 이를 위해, 3가지 특징 집합을 준비했다. CDF(CLF+CEF), PDF(PLF+PEF), and ADF(CDF+PDF). 그리고 이들을 이용하여 분류 성능을 평가했다. 이 실험에서, 분류기의 정확도는 PDF와 함께 94.38%를 보였다. 이것은 각 오브젝트에서 최고의 분류정확도를 보였던 PLF와 AEF(or PEF)에 비해 약 3~7% 높게 정확도가 측정되었다. 이것은 내강과 상피 세포 특징들의 조합이 PDAC진단에 도움이 되는 것을 보여준다.

[0198] 다음으로, 세 객체로부터 추출된 특징들의 조합을 사용하는 CTF, PTF, ATF의 실험 결과들은 관 객체에서 수행되었던 실험결과에 의존됨을 볼 수 있다. 따라서, 이것은 모든 객체들의 특징을 조합을 통한 성능개선은 없음을 보여준다. PTF의 경우는 PDF에 CNF특징 집합을 추가함으로써, 오히려 PDF의 분류 정확도에 비해 약 2%정도 정확도가 낮아졌다.

[0199] 다음으로 각 특징 집합에 의해 학습된 분류기들에 대해 ROC분석을 수행하였다. ROC분석은 진단 검사의 정확성을 나타내고 비교하는 기준으로 의학분야에서 널리 사용되고 있다. ROC분석은 TP rate(Sensitivity)와 FP rate(1-Specificity)에 의해 그려지는 ROC 커브를 통해 분석된다. 이 분석에서 ROC커브의 아래의 면적(AUC; Area under the ROC curve)는 진단의 정확도를 의미한다. Swets는 AUC 수치 따라 비정보적(non-informative)(AUC=0.5), 덜 정확한(less accurate)( $0.5 < AUC \leq 0.7$ ), 중등도의 정확한(moderately accurate)( $0.7 < AUC \leq 0.9$ ), 매우 정확한(highly accurate)( $0.9 < AUC < 1$ ), 그리고 완벽한 검사(perfect tests)(AUC=1)로 분류했다[비특허문헌 57],[비특허문헌 58]. 즉, ROC커브는 좌 측 모서리에 가까울수록 좀더 정확한 진단 되었다고 해석할 수 있다. 도 23은 각 객체 타입 별로 특징집합들에 의해 학습된 10개 분류기들의 평균화된 ROC그래프와 AUC값을 보여준다.

[0200] ROC분석에서는 PDF에 의해 학습된 분류기의 AUC값이 0.96으로 가장 높게 나타났다. 이것은 이전의 분류기 정확도 분석과 마찬가지로 내강과 상피 핵 특징의 조합이 의미가 있음을 보여주고 있다. 그 다음으로는 PEF에 의해 학습된 분류기의 AUC 값이 0.94로 내강 보다 0.93의 AUC값을 가지는 PLF보다 높게 측정되었다. CNF특징을 제외하고 제안된 PDF나 PEF가 포함되는 특징집합들에 의해 학습된 분류기들은 AUC값이 0.9이상으로써 매우 정확한 진단을 수행한다고 해석 할 수 있다.

[0201] 전체적으로 본 발명의 특징 집합을 포함하는 실험들은 기존 특징 집합과 함께하는 분류기들 보다 좋은 성능을 보였다. 앞에서 언급했듯이, 관이 PDAC진단에 중요한 영역임을 보였다. PLF와 PEF로 구성된 PDF는 분류기의 성능 향상을 이끌었다. 또한 제안된 내강과 상피세포로부터 추출된 PLF와 PEF는 기존 특징 집합들 CLF와 CEF, CNF에 비해 높은 성능을 보였다.

[0202] 다음으로, 케이스 2(등급1 vs. 등급2)에 대하여 설명한다.

[0203] 본 단계에서, 특징집합에 따라 PDAC의 병기 등급 1과 등급 2의 구분을 수행한다. 케이스 1의 실험과 마찬가지로, 등급 1과 등급 2을 구분을 위한 10개의 훈련 집합과 테스트 집합을 생성하고, 분류 정확도를 평가했다. 도 24와 도 25은 분류기의 성능평가 결과를 보여준다. 등급 1과 등급 2 사이의 병기를 구분하기 위한 각 특징 집합의 분류 결과들은 케이스 1(Normal vs PDAC)을 구분하는 실험들보다 낮은 분류 정확도를 보였다.

[0204] 우선적으로, 내강 객체에 대한 실험에서 PLF(77.03%)는 CLF(57.97%)보다 약 19%의 높은 정확도를 보였다. 특히, PLF의 specificity (79.69)는 CLF (45.94)보다 약73%의 성능 이득(performance gain)을 보여준다. CLF와 PLF의 조합인 ALF와 함께 하는 실험에서, 분류정확도는 68.44%를 보였다. 이는 PLF로 학습된 분류기의 성능보다 오

히려 낮아짐을 보여준다.

[0205] 다음으로, 상피-핵으로 부터 추출된 특징집합과 함께하는 실험들은 전체적으로 케이스 1(normal vs pdac)의 분류정확도에 비해 낮은 성능을 보여준다. PDAC의 Grad1과 등급2에서 상피세포들은 원주형 모양을 보이며 핵의 극성 손실 특성을 보인다. 따라서 이 특징들이 등급 1과 등급 2를 구분하는 것은 어렵다. 그럼에도 불구하고, PEF(70.78%)의 분류 정확도는 CEF(56.41%)보다 14%높았다. AEF를 사용하는 실험은 PEF에 비해 낮게 측정되었다. 이것은 CEF와 PEF를 통한 성능 개선이 없음을 보여준다.

[0206] 관 특징 집합들을 사용하는 실험들은 내강 객체로부터 추출된 특징집합들의 실험 결과와 동일하다. PDF 을 사용한 Case 1(Normal vs PDAC)실험과 달리, PLF와 PEF를 통한 성능개선은 없었고 PLF와 동일한 실험 결과를 보였다. 이것은 PDF를 통한 분류기의 성능 향상은 PEF가 아니라 전적으로 PLF에 의존됨을 보여준다. 세 오브젝트들로부터 추출된 특징 집합들의 조합(CTF, PTF, ATF)에 대한 실험들은 관으로부터 추출된 특징집합의 실험결과와 동일한 결과를 보였다.

[0207] 도 26는 각 객체 타입 별로 특징집합들에 의해 병기 구분 분류기들의 평균화된 ROC그래프와 AUC값을 보여준다. 전체적으로 Case1에 비해 많은 성능차이를 보였다. CNF, CLF, CEF등과 같은 기존 형태학적 특징 집합을 이용한 분류기의 경우 AUC값은 각각 0.61, 0.45, 0.61로 less accurate 진단 결과를 보였다. 반면 PLF와 PEF와 같은 본 발명의 특징 집합들을 이용한 분류기들은 각각 0.79와 0.7로 한단계 높은 moderately accurate test결과를 보여주고 있다.

[0208] 이번 병기 구분 실험들에서 PLF와 함께하는 분류기의 성능이 가장 좋은 성능을 보였다. 정상과 PDAC를 구분하는 Case1의 실험과 달리 PEF와 PLF결합을 통한 성능향상은 보여주지 못했다. 본 실험에서 특이 할만한 것은 Lumen 오브젝트의 특징집합들의 실험결과가 duct 오브젝트와 Tissue오브젝트의 특징집합들의 실험결과와 동일하다는 것이다. duct오브젝트 특징집합과 tissue오브젝트의 특징집합은 lumen 오브젝트로부터 추출된 특징들에 다른 오브젝트로부터 추출된 특징들을 결합으로 구성되어 있다. 이는 내강 오브젝트로부터 추출된 특징들이 분류기의 성능에 직접적으로 영향을 주고 있음을 보여주며 내강부위가 진단 시 필요한 정보를 가장 많이 포함하고 있음을 의미한다.

[0209] 전체적으로 등급1과 등급2의 병기 구분실험들은 본 발명의 특징들과 기존 특징들 모두에서 정상과 PDAC를 구분하는 실험에 비행 낮은 성능을 보였다. 이것은 PDAC에서 나타나는 특성이 등급 1과 등급2 병기들에서 나타나기 때문이다. 또한 본 발명의 특징들은 우선적으로 체장 선암과 정상조직을 분류하도록 설계되었기 때문에 비슷한 형태학적 특징들(characteristics)이 포함되는 체장 선암의 병기 진단에는 한계를 보여주고 있다. 하지만, 본 발명의 특징들은 기존 특징들에 비해 높은 분류 정확도를 보였다.

[0210] 다음으로, 추출된 특징의 적합성 실험에 대하여 설명한다.

[0211] 즉, 세그먼트된 세부분들(non-epithelial cells, epithelial cells, lumen)로부터 추출된 특징들이 체장선암을 진단하기 위해 얼마나 적합한지를 보이기 위해 특징들에 대한 통계적 분석을 수행한다. 먼저 추출된 특징이 집단에 적합할 특징일 경우, 특징 값은 집단간에 구분되는 차이를 보일 것이라 가정했다. 추출된 특징들이 세 집단 사이에서 차이가 있는지를 통계적으로 보이기 위해, 각 특징들에 대해 분산분석(ANOVA; Analysis of variance)을 수행하였다. 추출된 특징이 차이가 있는지를 검정하기 위한 가설은 다음과 같다.

[0212] [수학식 17]

$$H_0 : \mu_N = \mu_{G1} = \mu_{G2}$$

[0214] 여기서,  $\mu_N$ ,  $\mu_{G1}$ ,  $\mu_{G2}$ 는 각각 Normal, 등급 1, 등급 2에 대한 모 평균을 의미한다. ANOVA의 유의성 검정은 F-통계량에 의해 검정된다. 도 27, 도 28, 도 29은 세그먼트된 세부위(비상피세포, 상피세포, 내강)로부터 얻은 특징들에 대한 F-test를 통한 검정 결과이다.

[0215] 세 객체 타입에 대한 특징들의 F-test결과는 대부분의 특징들이 집단간에 차이가 존재함을 통계적으로 보여주고 있다. 내강 오브젝트에서 추출된 특징들에 대한 F-test경우 Roundness and Solidity 특징을 제외하고 모든 특징들이 그룹들(normal, 등급1, and 등급2) 사이에 차이가 있음을 보여주고 있으며, 상피-핵 오브젝트로부터 추출된 특징들의 F-test 결과는 모든 추출된 모든 특징들이 그룹들 사이에 차이가 있음을 보여주고 있다. 비-상피핵

에서는 MinorAxis과 skewness를 제외하고 모든특징들이 그룹들간 차이가 있음을 보여주었다.

[0216] 위의 귀무가설이 기각된 특징들(즉, 세집단간에 차이가 존재하는)에 대해 어느 집단들의 비교에서 차이가 존재하는지를 찾아내기 위해 다중범위 검정을 수행하였다. 본 발명에서는 여러 가지 다중범위검정방법들 중 가장 흔히 사용되는Fisher's LSD(LSD(Least Significant Difference) test[비특허문헌 58] 이유했었다. LSD는 비교하고자 하는 두 모집단의 평균들,  $\mu_i$ 와  $\mu_j$ 간에 차이가 있는지를 검정함으로써, 위의 영가설이 어디서 기각되었는지를 보여준다. 본 발명에서는 3개의 group들, normal, 등급1, and 등급2, 이 있으므로 총 <sup>(3)</sup><sub>(2)</sub>개 쌍들간에 대해 차이가 있는지를 검정한다. 각 쌍들을 검정하기 위한 영가설은 다음과 같다.

[0217] [수학식 18]

$$\begin{aligned} H_{01} : \mu_N &= \mu_{G1} \\ H_{02} : \mu_N &= \mu_{G2} \\ H_{03} : \mu_{G1} &= \mu_{G2} \end{aligned}$$

[0218]

[0219] LSD-test는 Eq. 17의 영가설을 기각한 특징들에 대해서만 수행된다. LSD-test 결과는 도 30 내지 도 32의 테이블과 같다.

[0220] 도 30 내지 도 32에서 음영처리된 행은 LSD-test에 대한 Eq. 30의 세 영가설을 모두 기각한 특징들이다.

[0221] 그리고 표에서 '-'로 표시된 특징들은 F-test에서 Eq.29의 가설을 reject하지 못했기 때문에 LSD-test가 수행되지 않았음을 표시한다. 우선, 내강 객체로부터 추출된 특징들에 대한 LSD-test에서 Circularity 와 AspectRatio를 제외한 12개의 특징들이 Eq. 18의 모든 영가설을 기각했다. 이들 중 8개는 기존 특징들이고, 4개는 본 발명의 특징들이다. 비록 기존의 많은 특징들이 집단간(Normal, 등급1, 등급2)사이의 차이를 보임에도 불구하고 제안된 PLF를 사용하는 실험의 결과가 Case 1과 Case 2에서 CLF에비해 좋은 성능을 보였다.

[0222] 상피 핵 객체에서는, 제안된 CytoplasmLength 와 CytoplasmLength<sub>sp</sub>특징들을 포함하여 5개의 특징들이 Eq. 18의 영가설을 기각했다. 비-상피 핵의 경우는Perimeter, Width, MajorAxis, and Fereter's Diameter만이 영가설들(Eq. 30)을 모두 reject했다.

[0223] LSD-test에서 Eq.18의 영가설 H<sub>03</sub>(등급 1과 Grae 2사이의 유의한 차이를 검정하기 위한)는 상피 핵과 비-상피핵들로 부터 추출된 많은 특징들에서 기각되지 않았다. F-test에서 영가설(Eq.29)을 기각한 특징들 중 LSD-test의 H<sub>03</sub> of Eq. 30을 기각하지 못한 특징은 Circularity 와 AspectRatio 단 2개의 특징이다. 그러나 F-test에서 모든 특징들이 영가설(Eq.29)을 기각했던 epithelial nucleus 타입의 경우, 14개의 특징들 중 단 5개만이 LSD-test에서 H<sub>03</sub> of Eq. 30을 기각했다. 비-상피 핵 타입에서는 단 4개의 특징들만이 LSD-test에서 H<sub>03</sub> of Eq. 30을 reject했다.

[0224] 이 LSD-test결과는 내강이 PDAC진단 및 병기구분에 있어서 가장 중요한 오브젝트임을 보여주고 있다. 또한 이 결과들은 병기구분을 위한 Case 2의 실험들이 정상과 PDAC를 진단하는 Case1실험에 비해 성능이 낮은 이유가 통계적으로 설명된다.

[0225] 본 발명은 관샘암을 자동으로 진단하기 위한 방법들을 설명하였다. 관샘암 진단은 주로 관의 분화도와 관을 에워싸고 있는 상피세포를 관찰함으로써 수행되어 진다. 본 발명은 주어진 조직세포 영상을 내강과 상피세포, 그리고 비상피세포로 세그멘테이션하였다. 세그멘테이션된 관과 상피세포로부터 전통적인 특징들과 함께 관샘암을 진단하는데 특화된 새로운 형태학적 특징들을 제안하였다. 관샘암의 경우 관은 정상조직의 관에 비해 모양이 더 복잡하고 경계의 굴곡 변화가 매우 심하다. 본 발명에서는 이러한 관점으로부터 관의 변동성과 관의 이형성 비율을 측정하기 위한 특징들을 제안했다. 관의 변동성을 직관적으로 표현하기 위해 이형성 진폭을 이용한 1-D 시그너처로 관을 표현하였다. 이 이형성 진폭 시그너처로부터 이형성 진폭의 변동 폭을 측정하기 위한 RMSAA특징과, 변곡점에서의 변화량의 측정을 위한 PIP기반의 TSAV특징을 제안하였다. 그리고 이상적 내강과 오리지널 내강 영역들을 사용하여, 관의 찌그러진 정도를 측정하기 위한 AtypiaRatio와 관의 유두상 정도를 측정하기 위한 #AtypiaRegions특징들이 제안되었다. 또한 세그멘트된 상피세포로부터 세포질 길이, 극성 정도등을

측정하기 위해 CytopasmLength와 CytoplasmLength<sub>SD</sub>특징들이 사용되었다. SVM을 이용한 분류결과와 특징들에 대한 통계적 분석은 본 발명의 특징들이 관샘암을 진단과 등급1과 등급등급 구분에 적합한 특징임을 보여 주고 있다.

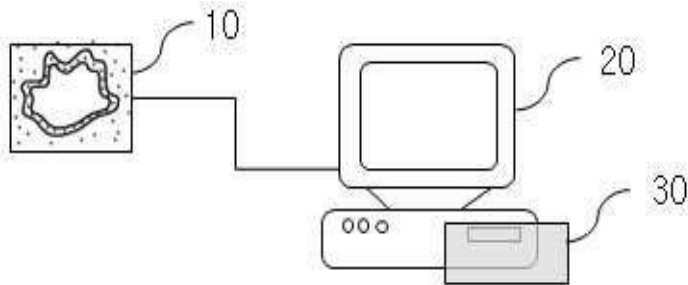
[0226] 이상, 본 발명자에 의해서 이루어진 발명을 상기 실시 예에 따라 구체적으로 설명하였지만, 본 발명은 상기 실시 예에 한정되는 것은 아니고, 그 요지를 이탈하지 않는 범위에서 여러 가지로 변경 가능한 것은 물론이다.

### 부호의 설명

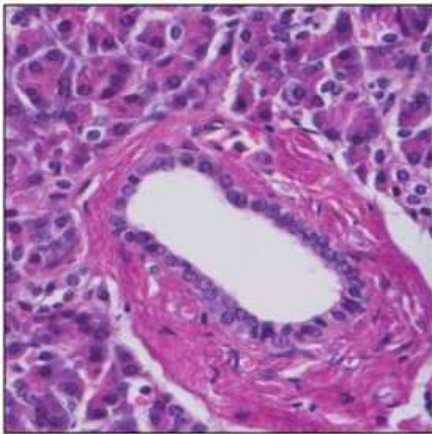
[0227] 10 : 조직세포 영상                      20 : 컴퓨터 단말  
30 : 프로그램 시스템

### 도면

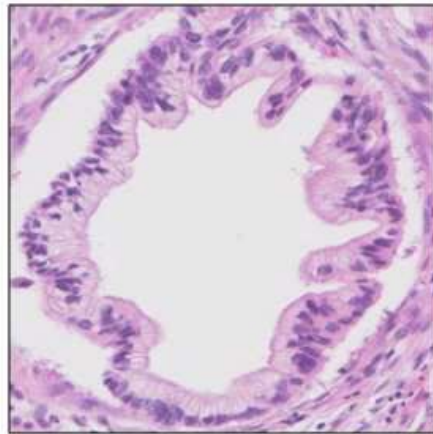
#### 도면1



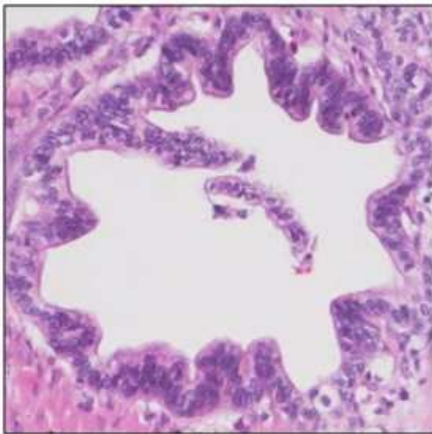
도면2



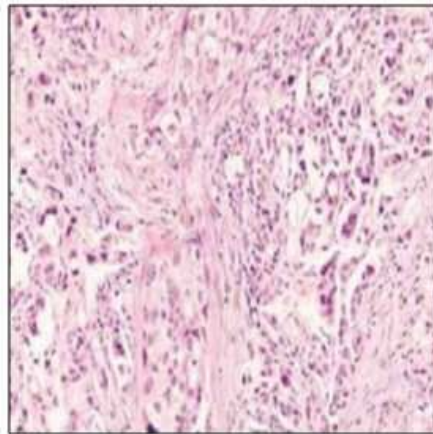
(a)



(b)

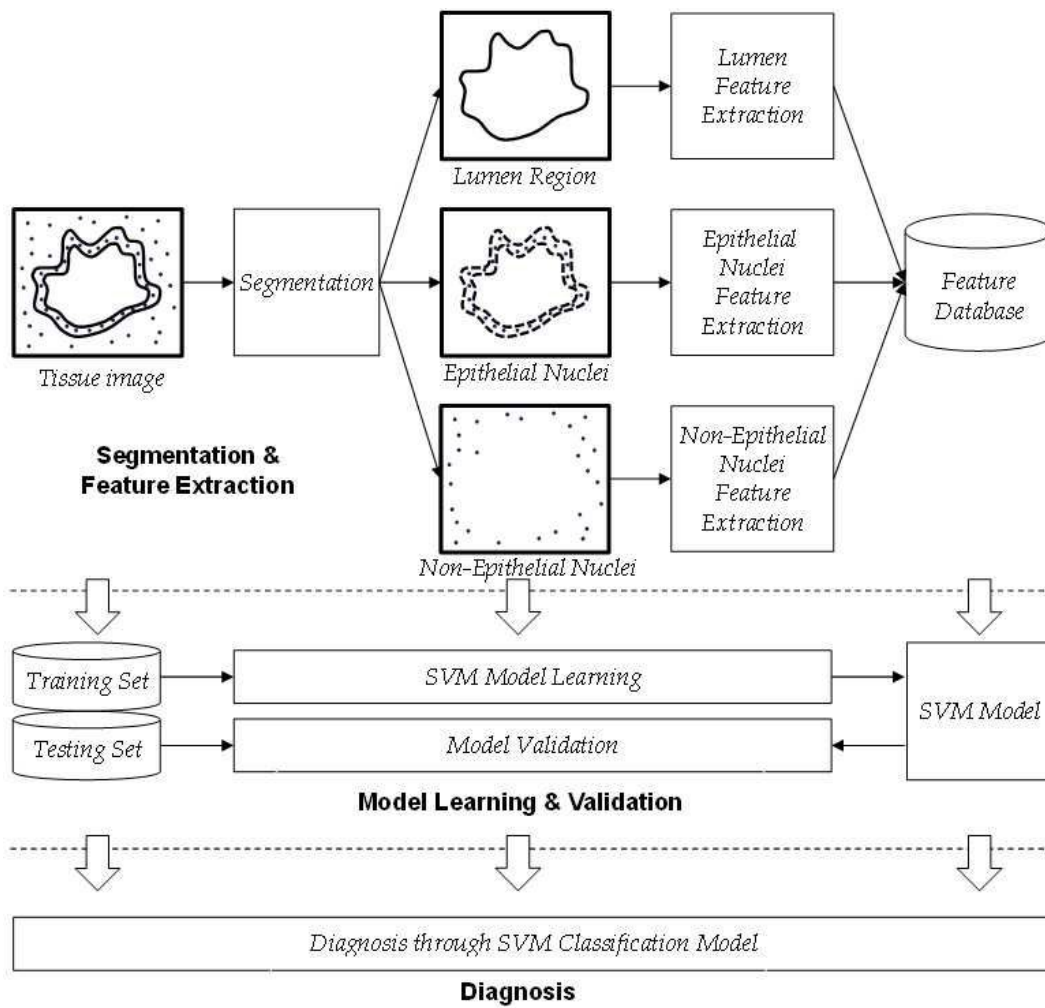


(c)

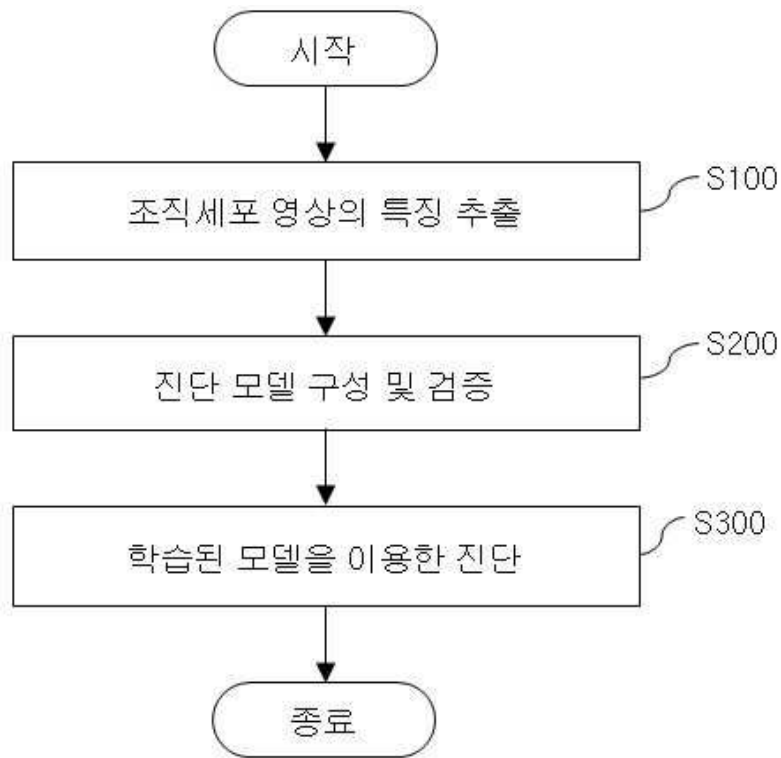


(d)

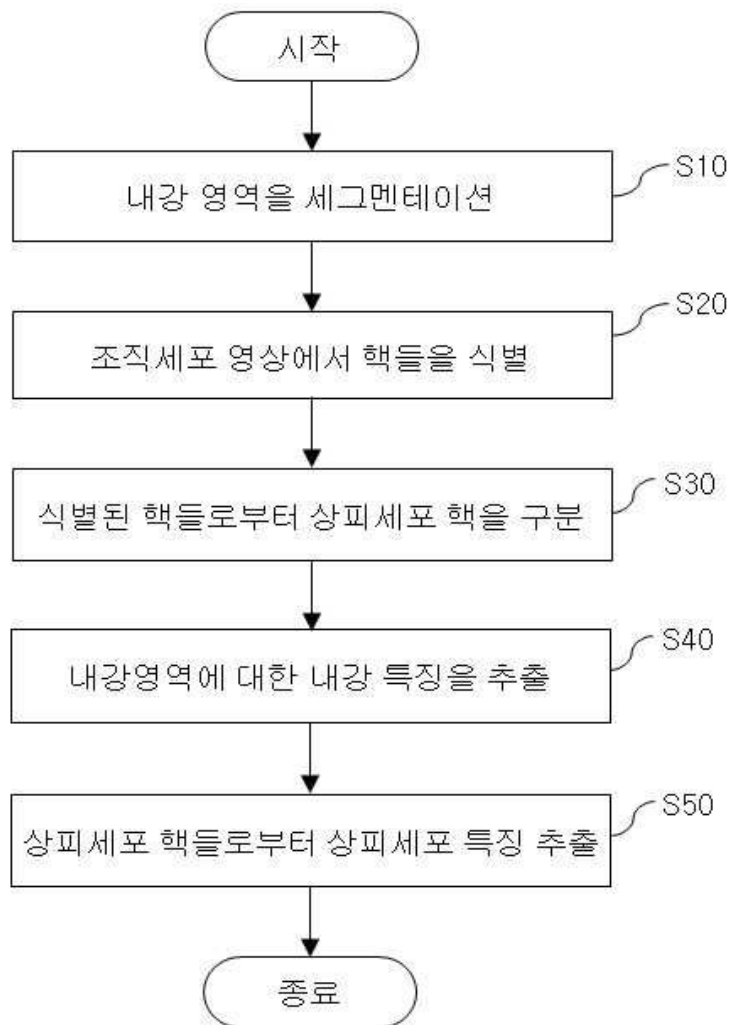
도면3



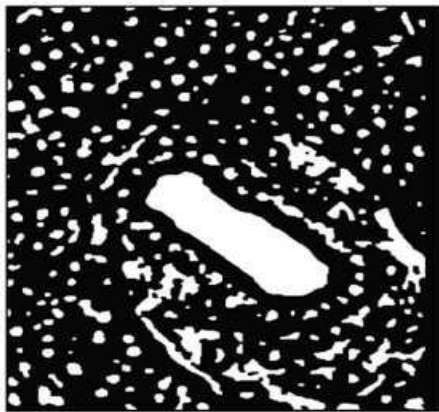
도면4



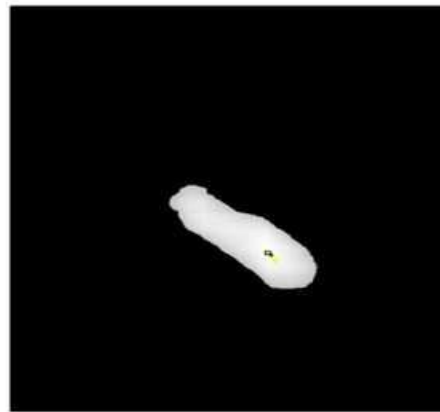
도면5



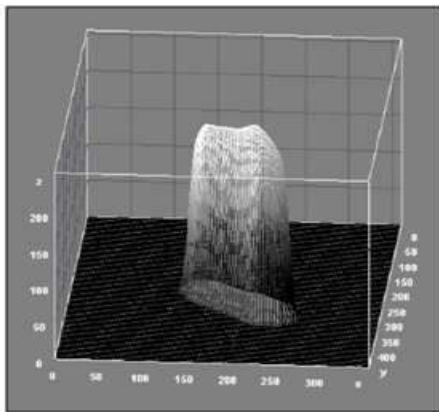
도면6



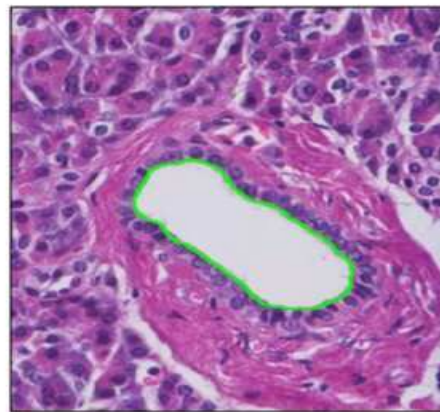
(a)



(b)

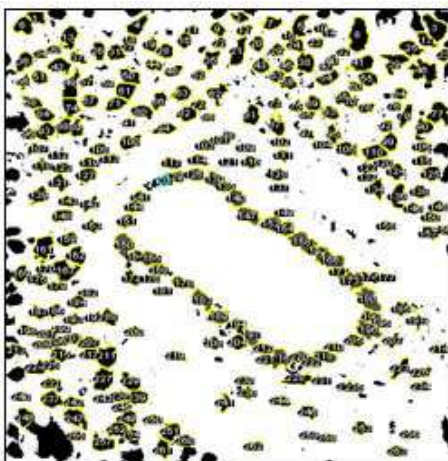


(c)

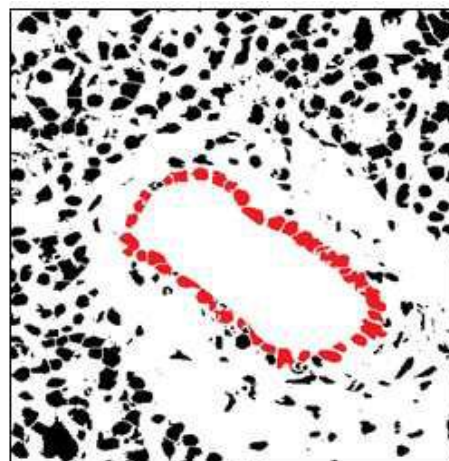


(d)

도면7



(a)



(b)

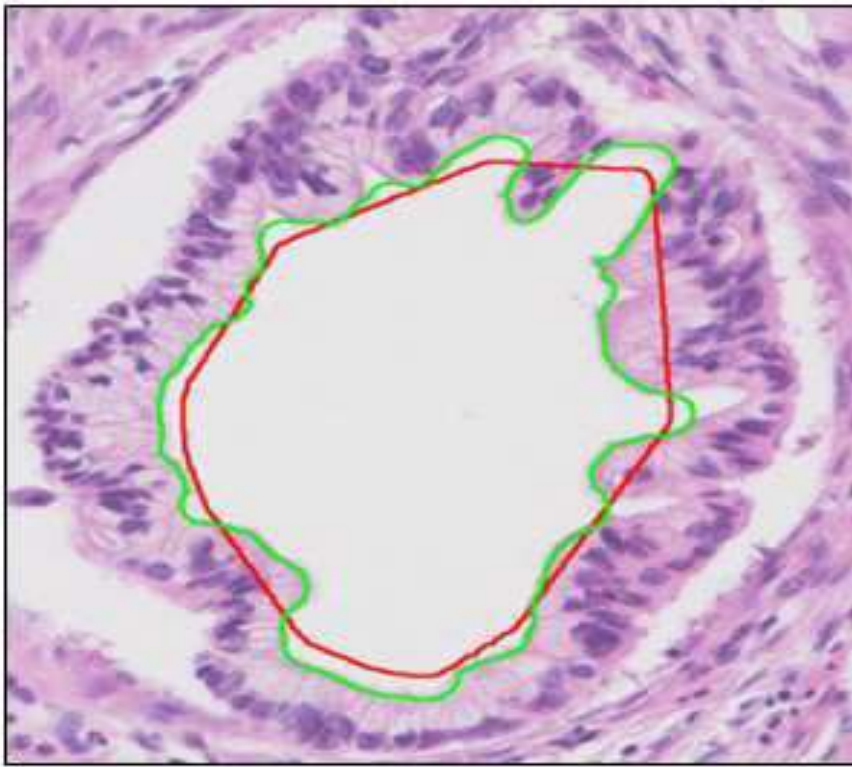
도면8

| Algorithm 1. Selection_Epithelial_Nuclei( $\mathcal{N}$ ). |                                                        |
|------------------------------------------------------------|--------------------------------------------------------|
| <b>Input:</b> A set of nuclei $\mathcal{N}$                |                                                        |
| <b>Output:</b> A set of epithelial nuclei $\mathcal{N}^E$  |                                                        |
| <b>Method :</b>                                            |                                                        |
| 1                                                          | $\mathcal{N}^E \leftarrow \{ \}$                       |
| 2                                                          | <b>for all</b> point $p$ in $B^O$ <b>do</b>            |
| 3                                                          | $n \leftarrow \text{getNearestNuclus}(p, \mathcal{N})$ |
| 4                                                          | $\mathcal{N}^E \leftarrow \mathcal{N}^E \cup n$        |
| 5                                                          | <b>end for</b>                                         |
| 6                                                          | $\mathcal{N}^E$                                        |

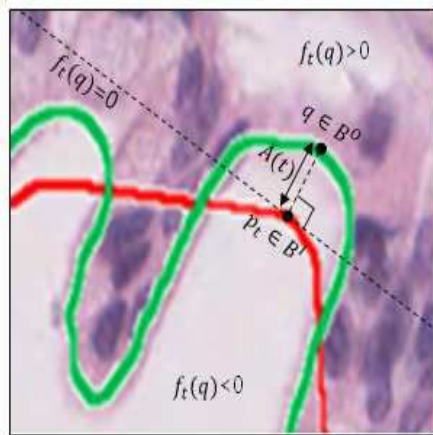
도면9

| Symbol                          | Description                                                                                                                                                                                                                                                                                                               | Symbol                   | Description                                                                                                                                                                                                                         |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $p$                             | The $p$ is a pair of numbers that are $x$ and $y$ coordinates. $p = (x, y)$ .                                                                                                                                                                                                                                             | $\mathbf{A}$             | The matrix representation for preprocessed binary image.                                                                                                                                                                            |
| $H(\mathbf{A})$                 | The Direction Cumulative Histogram for $\mathbf{A}$                                                                                                                                                                                                                                                                       | $H_T(\mathbf{A})$        | The thresholded $H(\mathbf{A})$                                                                                                                                                                                                     |
| $B^O$                           | A sequence of points that consist of the original lumen boundary.                                                                                                                                                                                                                                                         | $B^C$                    | A sequence of points for the convex hull of $B^O$                                                                                                                                                                                   |
| $B^I$                           | A sequence of points for the ideal lumen boundary that is estimated from $B^O$ .                                                                                                                                                                                                                                          | $\mathcal{N}$            | A set of nuclei. $\mathcal{N} = \{n_1, n_2, \dots, n_m\}$                                                                                                                                                                           |
| $\mathcal{N}^E$                 | A set of epithelial nuclei. $\mathcal{N}^E \subset \mathcal{N}$                                                                                                                                                                                                                                                           | $\mathcal{N}^{NE}$       | A set of no-epithelial nuclei. $\mathcal{N}^{NE} = \mathcal{N} - \mathcal{N}^E$                                                                                                                                                     |
| $\text{distance}(\cdot, \cdot)$ | The Euclidean distance function.                                                                                                                                                                                                                                                                                          | $\text{centroid}(\cdot)$ | The function returning center point of given object.                                                                                                                                                                                |
| $A(t)$                          | The atypia-amplitude function.                                                                                                                                                                                                                                                                                            | $L(t)$                   | The perimeter from $p_0$ (start point) to $p_t$ ( $t^{\text{th}}$ point).                                                                                                                                                           |
| $AR$                            | A set of atypia regions. $AR = \{ar_1, ar_2, \dots, ar_m\}$ .                                                                                                                                                                                                                                                             | $R^O$                    | The region surrounded by $B^O$                                                                                                                                                                                                      |
| $R^C$                           | The region surrounded by $B^C$                                                                                                                                                                                                                                                                                            | $R^I$                    | The region surrounded by $B^I$                                                                                                                                                                                                      |
| $\text{Area}(\cdot)$            | The function returning area of given region.                                                                                                                                                                                                                                                                              | $D(\cdot)$               | The data set for given feature set. $D(\cdot) = \{(\mathbf{x}_1(\cdot), y_1), \dots, (\mathbf{x}_m(\cdot), y_m)\}$ , $\mathbf{x}(\cdot)$ is feature vector for given feature set and $y_i$ is class label for $\mathbf{x}(\cdot)$ . |
| $\mathcal{D}(\cdot)$            | The data set is a set of pairs $(\mathcal{L}(\cdot), \mathcal{T}(\cdot))$ for bootstrap evaluation. $\mathcal{D}(\cdot) = \{(\mathcal{L}(\cdot)_1, \mathcal{T}(\cdot)_1), \dots, (\mathcal{L}(\cdot)_{10}, \mathcal{T}(\cdot)_{10})\}$ , $\mathcal{L}(\cdot)$ and $\mathcal{T}(\cdot)$ are learning and testing data set. |                          |                                                                                                                                                                                                                                     |

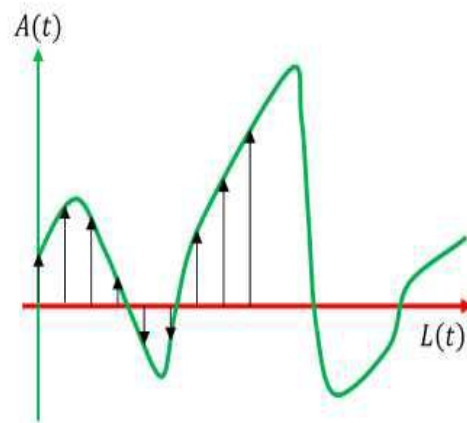
도면10



도면11

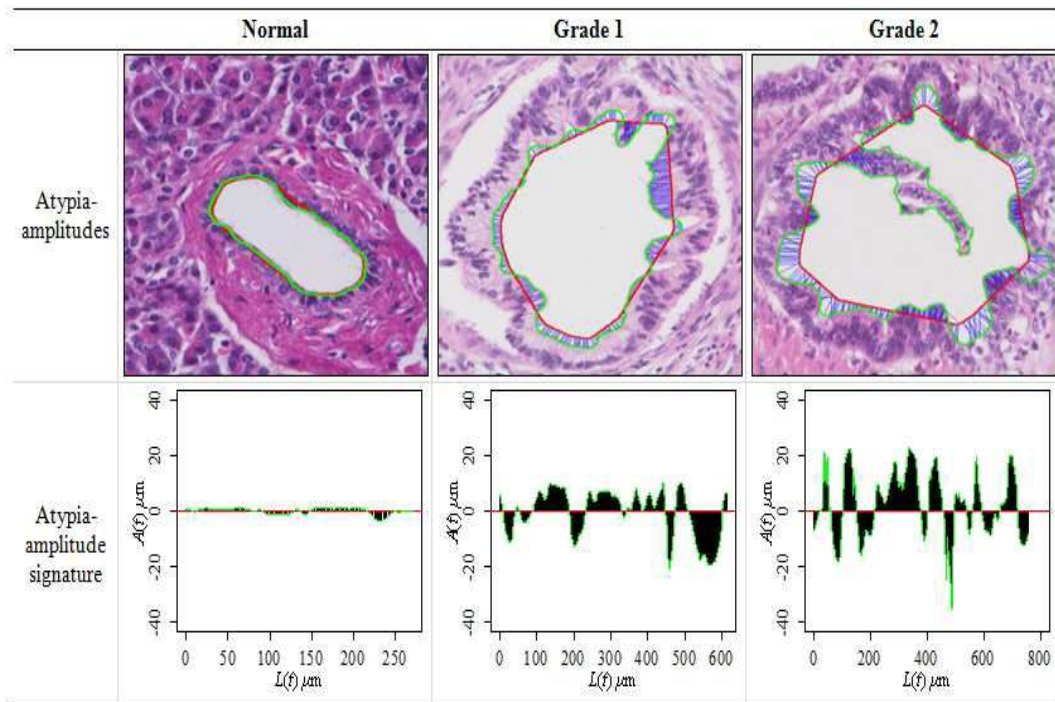


(a)

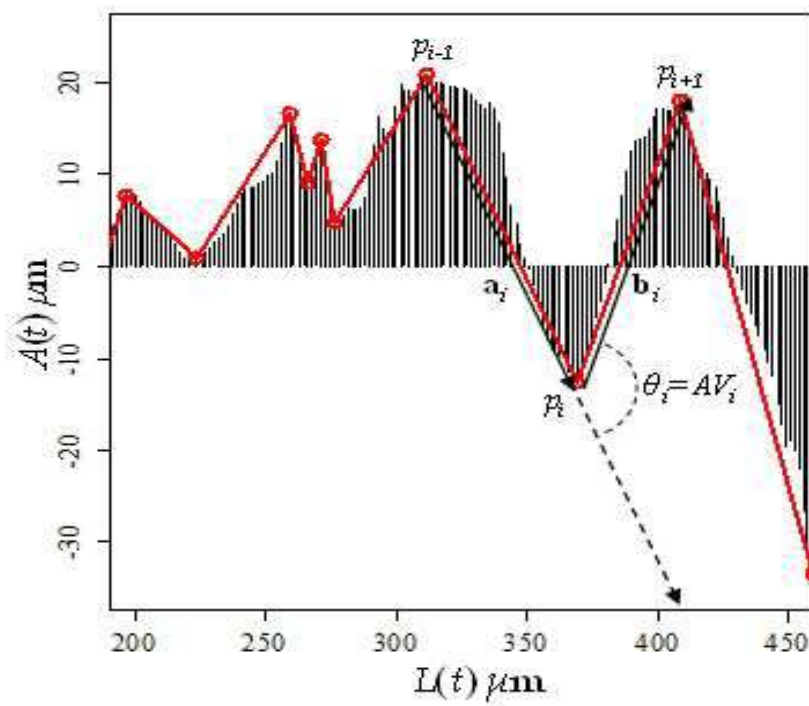


(b)

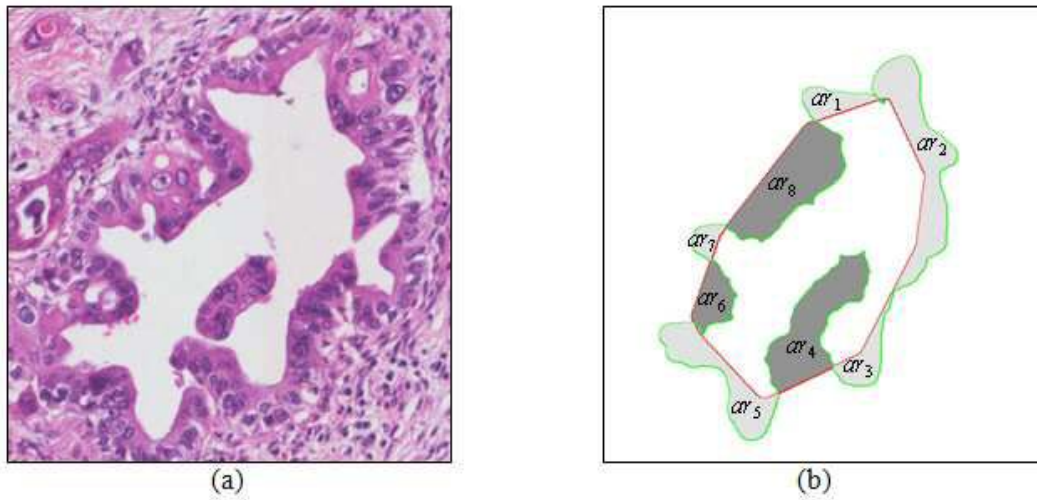
도면12



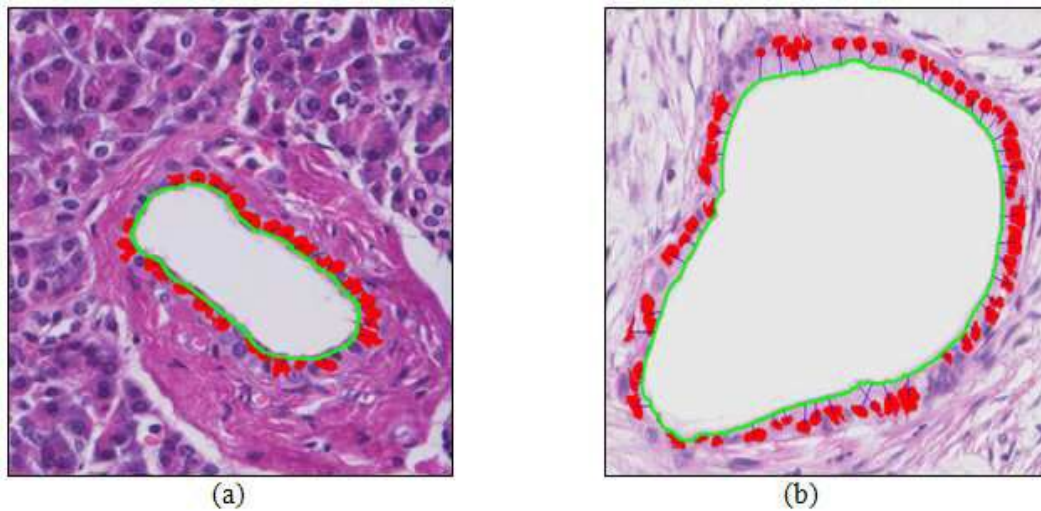
도면13



도면14



도면15



도면16

|                    |                                  |
|--------------------|----------------------------------|
| Digital Slide size | Variable size                    |
| Image Resolution   | 0.492 $\mu\text{m}/\text{pixel}$ |
| Image Type         | SVS/JPEG2000                     |
| Image Channels     | 3                                |
| Image Bit Depth    | 8 bits                           |
| Magnification      | 20X                              |
| Organization       | tiled                            |
| Tile Width         | 240 pixels                       |
| Tile Height        | 240 pixels                       |

도면17

| Type   | Number of Images |
|--------|------------------|
| Normal | 80               |
| Grade1 | 80               |
| Grade2 | 80               |

도면18

| No | Feature                             | Description                                                                              | Lumen | EN <sup>1</sup> | NEN <sup>2</sup> |
|----|-------------------------------------|------------------------------------------------------------------------------------------|-------|-----------------|------------------|
| 1  | <i>Area</i>                         | Area of selection in square pixels                                                       | *     | *               | *                |
| 2  | <i>Perimeter</i>                    | The length of the outside boundary of the selection                                      | *     | *               | *                |
| 3  | <i>Width</i>                        | Width of selection                                                                       | *     | *               | *                |
| 4  | <i>Height</i>                       | Height of selection                                                                      | *     | *               | *                |
| 5  | <i>Major axis</i>                   | Major(primary) axis of the best Fit Ellipse                                              | *     | *               | *                |
| 6  | <i>Minor axis</i>                   | Minor(secondary) axis of the best Fit Ellipse                                            | *     | *               | *                |
| 7  | <i>Circurity</i>                    | $4\pi(\text{area}/\text{perimeter}^2)$ . A value of 1.0 indicates a perfect circle.      | *     | *               | *                |
| 8  | <i>Feret's Diameter</i>             | The longest distance between any two points along the selection boundary.                | *     | *               | *                |
| 9  | <i>Aspect Ratio</i>                 | The ratio of the width of the selection to its height.                                   | *     | *               | *                |
| 10 | <i>Skewness</i>                     | The third order moment about the mean                                                    | *     | *               | *                |
| 11 | <i>Roundness</i>                    | $\text{Compactness}/4\pi$ . The measure of the sharpness of particle's edge and corners. | *     | *               | *                |
| 12 | <i>Solidity</i>                     | $\text{Area}/\text{ConvexArea}$                                                          | *     | *               | *                |
| 13 | <i>RMSAA</i>                        | Root-Mean-Squared Atypia- Amplitude                                                      | *     |                 |                  |
| 14 | <i>TSAV</i>                         | Total Sum of Atypia Volatility for PIPs                                                  | *     |                 |                  |
| 15 | <i>AtypiaRatio</i>                  | The Ratio of atypia region                                                               | *     |                 |                  |
| 16 | <i>#AtypiaRegions</i>               | The number of atypia regions for identifying papillary                                   | *     |                 |                  |
| 17 | <i>CytoplasmLength</i>              | The cytoplasm length of epithelial nucleus.                                              |       | *               |                  |
| 18 | <i>CytoplasmLength<sub>SD</sub></i> | The standard-deviation of cytoplasm length                                               |       | *               |                  |

EN<sup>1</sup>: Epithelial Nuclei, NEN<sup>2</sup>: Non-Epithelial Nuclei

도면19

| Object           | Symbol | Feature Set                                                            | Dimension |
|------------------|--------|------------------------------------------------------------------------|-----------|
| NEN <sup>1</sup> | CNF    | Classical Non-epithelial nuclei Features                               | 12        |
|                  | CLF    | Classical Lumen Features                                               | 12        |
| Lumen            | PLF    | Proposed Lumen Features                                                | 4         |
|                  | ALF    | <i>CLF+PLF</i>                                                         | 16        |
| EN <sup>2</sup>  | CEF    | Classical Epithelial nuclei Features                                   | 12        |
|                  | PEF    | Proposed Epithelial nuclei Features                                    | 2         |
|                  | AEF    | <i>CEF+PEF</i>                                                         | 14        |
| Duct             | CDF    | Classical Duct Features ( <i>CLF+CEF</i> )                             | 24        |
|                  | PDF    | Proposed Duct Features ( <i>PLF+PEF</i> )                              | 6         |
|                  | ADF    | <i>CDF + PDF</i>                                                       | 30        |
| Tissue           | CTF    | Classical Features extracted from three objects ( <i>CLF+CEF+CNF</i> ) | 36        |
|                  | PTF    | Proposed Features extracted from three objects ( <i>PLF+PEF+CNF</i> )  | 18        |
|                  | ATF    | All Features( <i>CLF+CEF+PLF+PEF+CNF</i> )                             | 42        |

NEN<sup>1</sup>: Non-Epithelial Nuclei, EN<sup>2</sup>: Epithelial Nuclei,

도면20

| Experiment         | Class  | Number of Training Data | Number of Testing Data |
|--------------------|--------|-------------------------|------------------------|
| Case 1:            | Normal | 48                      | 32                     |
| Normal vs PDAC     | PDAC   | 48                      | 32                     |
| Case 2:            | Grade1 | 48                      | 32                     |
| Grade 1 vs Grade 2 | Grade2 | 48                      | 32                     |

도면21

| Object | Feature Set | TN     | FP     | FN     | TP     | SN(%)  | SP(%)   | PPV(%) | NPV(%) | ACC(%) |
|--------|-------------|--------|--------|--------|--------|--------|---------|--------|--------|--------|
| NEN    | CNF         | 26.50  | 6.30   | 5.50   | 25.70  | 82.82  | 80.32   | 81.07  | 82.50  | 81.56  |
|        |             | (1.58) | (2.11) | (1.58) | (2.11) | (4.94) | (6.60)  | (5.17) | (4.22) | (3.95) |
| Lumen  | CLF         | 27.30  | 12.30  | 4.70   | 19.70  | 85.32  | 61.57   | 69.38  | 81.72  | 73.44  |
|        |             | (2.21) | (3.27) | (2.21) | (3.27) | (6.92) | (10.21) | (4.22) | (6.25) | (3.13) |
|        | PLF         | 29.30  | 2.70   | 2.70   | 29.30  | 91.57  | 91.57   | 91.99  | 91.88  | 91.56  |
|        |             | (1.77) | (2.00) | (1.77) | (2.00) | (5.52) | (6.26)  | (5.45) | (4.70) | (3.14) |
| EN     | ALF         | 27.50  | 3.60   | 4.50   | 28.40  | 85.94  | 88.75   | 88.96  | 86.36  | 87.35  |
|        |             | (0.85) | (2.46) | (0.85) | (2.46) | (2.66) | (7.68)  | (6.92) | (2.12) | (3.57) |
|        | CEF         | 25.10  | 10.60  | 6.90   | 21.40  | 78.44  | 66.88   | 70.72  | 75.75  | 72.66  |
|        |             | (1.66) | (2.84) | (1.66) | (2.84) | (5.20) | (8.86)  | (4.97) | (3.99) | (3.84) |
| Duct   | PEF         | 27.50  | 3.50   | 4.50   | 28.50  | 85.94  | 89.07   | 89.52  | 86.67  | 87.50  |
|        |             | (1.51) | (2.95) | (1.51) | (2.95) | (4.72) | (9.23)  | (6.80) | (3.25) | (3.21) |
|        | AEF         | 27.00  | 3.00   | 5.00   | 29.00  | 84.38  | 90.63   | 90.55  | 85.73  | 87.50  |
| Tissue |             | (2.05) | (1.89) | (2.05) | (1.89) | (6.42) | (5.89)  | (5.01) | (4.53) | (1.47) |
|        | CDF         | 27.30  | 12.30  | 4.70   | 19.70  | 85.32  | 61.57   | 69.38  | 81.72  | 73.44  |
|        |             | (2.21) | (3.27) | (2.21) | (3.27) | (6.92) | (10.21) | (4.22) | (6.25) | (3.13) |
|        | PDF         | 29.80  | 1.40   | 2.20   | 30.60  | 93.13  | 95.63   | 95.78  | 93.50  | 94.38  |
| Tissue |             | (1.40) | (1.51) | (1.40) | (1.51) | (4.37) | (4.71)  | (4.34) | (3.80) | (2.35) |
|        | ADF         | 27.50  | 3.60   | 4.50   | 28.40  | 85.94  | 88.75   | 88.96  | 86.36  | 87.35  |
|        |             | (0.85) | (2.46) | (0.85) | (2.46) | (2.66) | (7.68)  | (6.92) | (2.12) | (3.57) |
| Tissue | CTF         | 26.90  | 11.70  | 5.10   | 20.30  | 84.06  | 63.44   | 70.34  | 80.92  | 73.75  |
|        |             | (2.28) | (3.74) | (2.28) | (3.74) | (7.13) | (11.70) | (5.32) | (6.11) | (3.44) |
|        | PTF         | 29.80  | 2.50   | 2.20   | 29.50  | 93.13  | 92.19   | 92.74  | 93.26  | 92.66  |
| Tissue |             | (1.32) | (2.22) | (1.32) | (2.22) | (4.11) | (6.95)  | (6.15) | (3.76) | (3.21) |
|        | ATF         | 27.50  | 3.60   | 4.50   | 28.40  | 85.94  | 88.75   | 88.96  | 86.36  | 87.35  |
|        |             | (0.85) | (2.46) | (0.85) | (2.46) | (2.66) | (7.68)  | (6.92) | (2.12) | (3.57) |

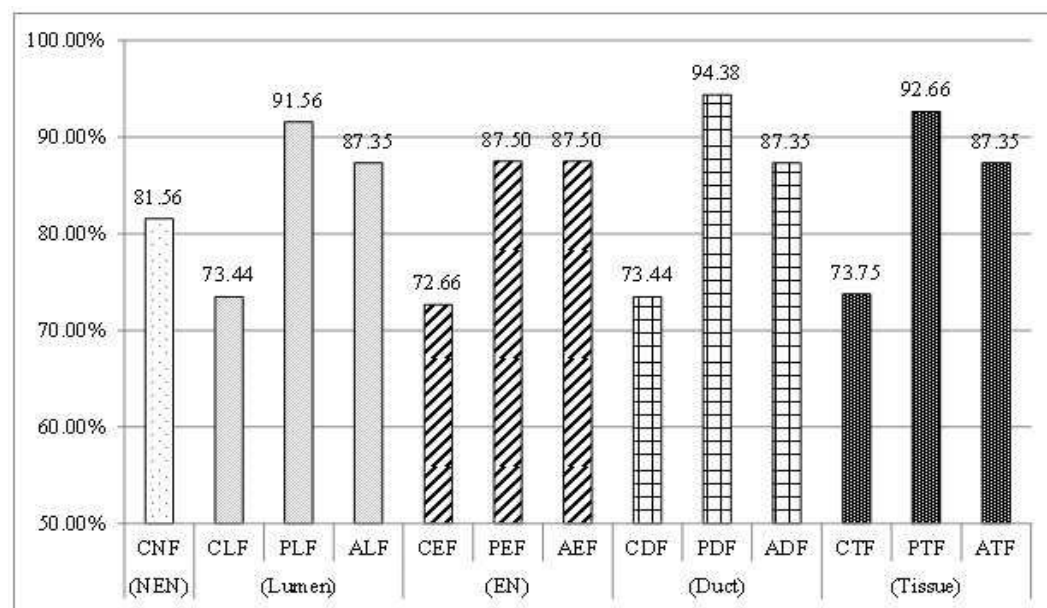
TP(True Positive): the number of PDACs that are correctly classified as PDACs.

FP(False Positive): the number of Normals that are incorrectly classified as PDACs.

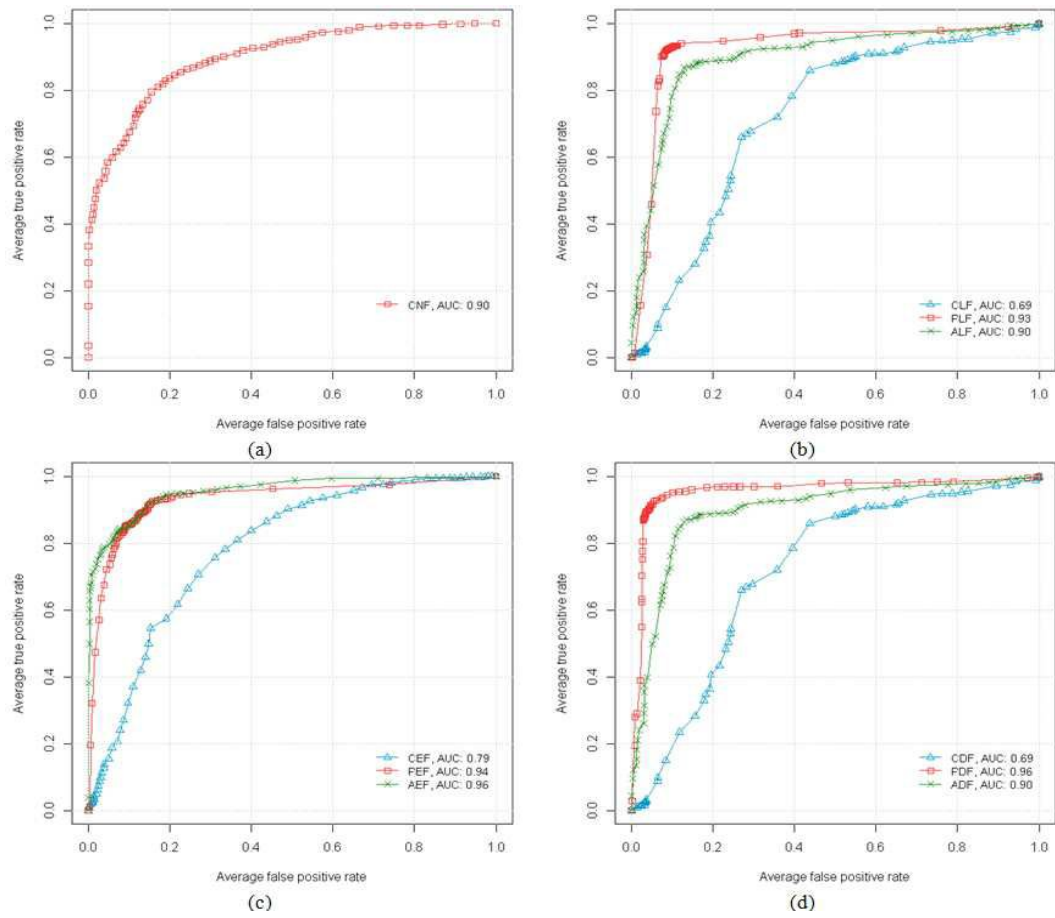
FN(False Negative): the number of PDACs that are incorrectly classified as Normals.

TN(True Negative): the number of Normals that are correctly classified as Normals..

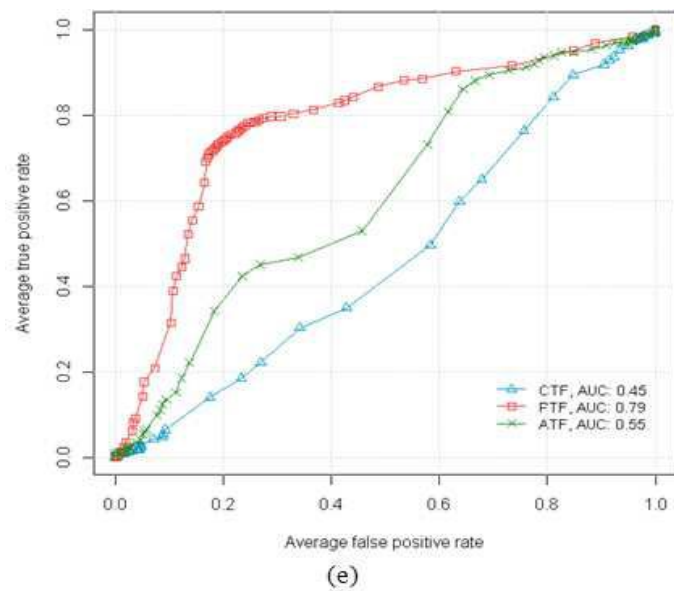
도면22



도면23a



도면23b



도면24

| Object | Feature Set | TN     | FP     | FN     | TP     | SN(%)   | SP(%)   | PPV(%)  | NPV(%)  | ACC(%) |
|--------|-------------|--------|--------|--------|--------|---------|---------|---------|---------|--------|
| NEN    | CNF         | 19.60  | 11.70  | 12.40  | 20.30  | 61.25   | 63.44   | 64.35   | 63.29   | 62.34  |
|        |             | (6.26) | (4.97) | (6.26) | (4.97) | (19.56) | (15.52) | (8.93)  | (9.49)  | (7.78) |
| Lumen  | CLF         | 22.40  | 17.30  | 9.60   | 14.70  | 70.00   | 45.94   | 58.39   | 66.13   | 57.97  |
|        |             | (5.99) | (6.95) | (5.99) | (6.95) | (18.70) | (21.70) | (10.63) | (16.21) | (5.23) |
|        | PLF         | 23.80  | 6.50   | 8.20   | 25.50  | 74.38   | 79.69   | 79.65   | 75.40   | 77.03  |
|        |             | (1.32) | (3.95) | (1.32) | (3.95) | (4.11)  | (12.35) | (10.15) | (4.96)  | (6.83) |
|        | ALF         | 25.20  | 13.40  | 6.80   | 18.60  | 78.75   | 58.13   | 67.05   | 73.46   | 68.44  |
|        |             | (2.90) | (6.31) | (2.90) | (6.31) | (9.06)  | (19.72) | (9.88)  | (5.85)  | (7.39) |
| EN     | CEF         | 18.60  | 14.50  | 13.40  | 17.50  | 58.13   | 54.69   | 56.45   | 57.03   | 56.41  |
|        |             | (4.30) | (4.14) | (4.30) | (4.14) | (13.44) | (12.95) | (7.17)  | (9.10)  | (7.53) |
|        | PEF         | 22.20  | 8.90   | 9.80   | 23.10  | 69.38   | 72.19   | 71.89   | 70.04   | 70.78  |
|        |             | (1.14) | (2.92) | (1.14) | (2.92) | (3.55)  | (9.14)  | (6.94)  | (4.08)  | (5.26) |
|        | AEF         | 22.40  | 11.60  | 9.60   | 20.40  | 70.00   | 63.75   | 65.95   | 68.16   | 66.88  |
|        |             | (2.32) | (2.27) | (2.32) | (2.27) | (7.25)  | (7.10)  | (5.42)  | (5.84)  | (5.50) |
| Duct   | CDF         | 22.40  | 17.30  | 9.60   | 14.70  | 70.00   | 45.94   | 58.39   | 66.13   | 57.97  |
|        |             | (5.99) | (6.95) | (5.99) | (6.95) | (18.70) | (21.70) | (10.63) | (16.21) | (5.23) |
|        | PDF         | 23.80  | 6.50   | 8.20   | 25.50  | 74.38   | 79.69   | 79.65   | 75.40   | 77.03  |
|        |             | (1.32) | (3.95) | (1.32) | (3.95) | (4.11)  | (12.35) | (10.15) | (4.96)  | (6.83) |
|        | ADF         | 25.10  | 13.20  | 6.90   | 18.80  | 78.44   | 58.75   | 67.25   | 73.40   | 68.60  |
|        |             | (2.88) | (6.21) | (2.88) | (6.21) | (9.02)  | (19.42) | (9.75)  | (5.88)  | (7.31) |
| Tissue | CTF         | 22.40  | 17.30  | 9.60   | 14.70  | 70.00   | 45.94   | 58.39   | 66.13   | 57.97  |
|        |             | (5.99) | (6.95) | (5.99) | (6.95) | (18.70) | (21.70) | (10.63) | (16.21) | (5.23) |
|        | PTF         | 23.80  | 6.50   | 8.20   | 25.50  | 74.38   | 79.69   | 79.65   | 75.40   | 77.03  |
|        |             | (1.32) | (3.95) | (1.32) | (3.95) | (4.11)  | (12.35) | (10.15) | (4.96)  | (6.83) |
|        | ATF         | 25.10  | 13.20  | 6.90   | 18.80  | 78.44   | 58.75   | 67.25   | 73.40   | 68.60  |
|        |             | (2.88) | (6.21) | (2.88) | (6.21) | (9.02)  | (19.42) | (9.75)  | (5.88)  | (7.31) |

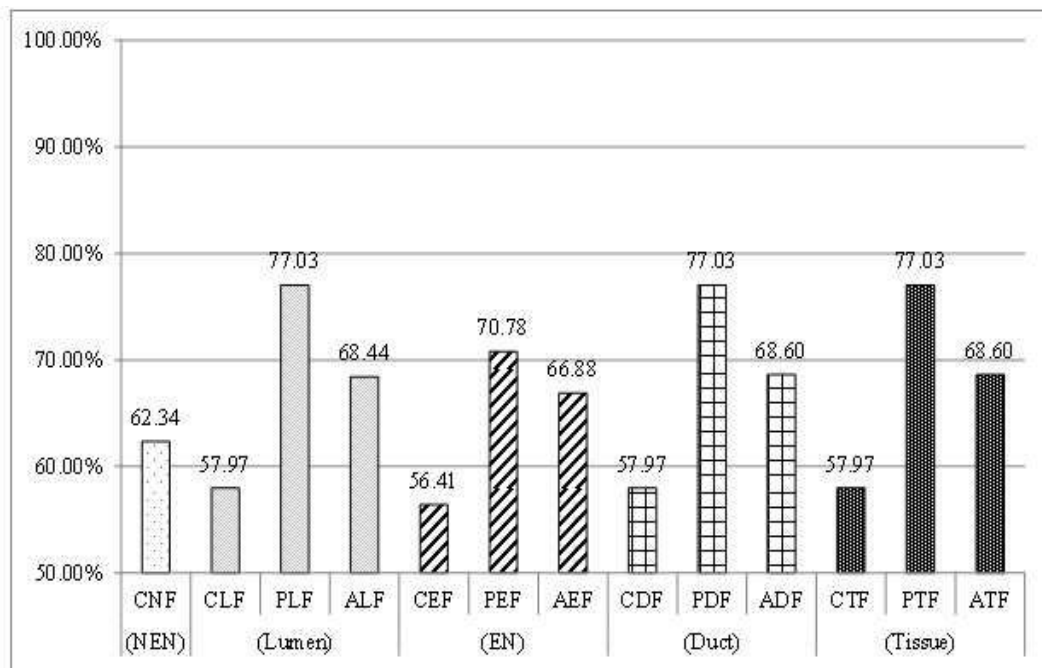
TP(True Positive): the number of Grade2s that are correctly classified as Grade2s.

FP(False Positive): the number of Grade1s that are incorrectly classified as Grade2s.

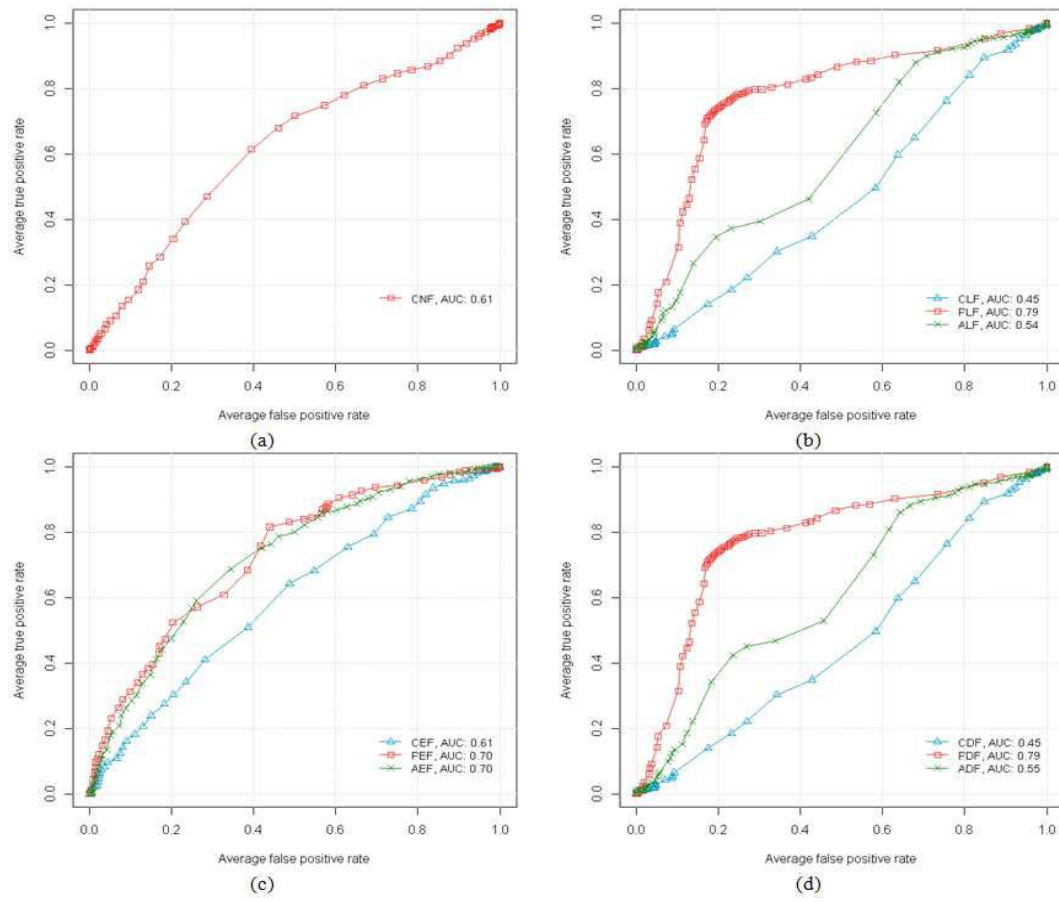
FN(False Negative): the number of Grade2s that are incorrectly classified as Grade1s.

TN(True Negative): the number of Grade1s that are correctly classified as Grade1s..

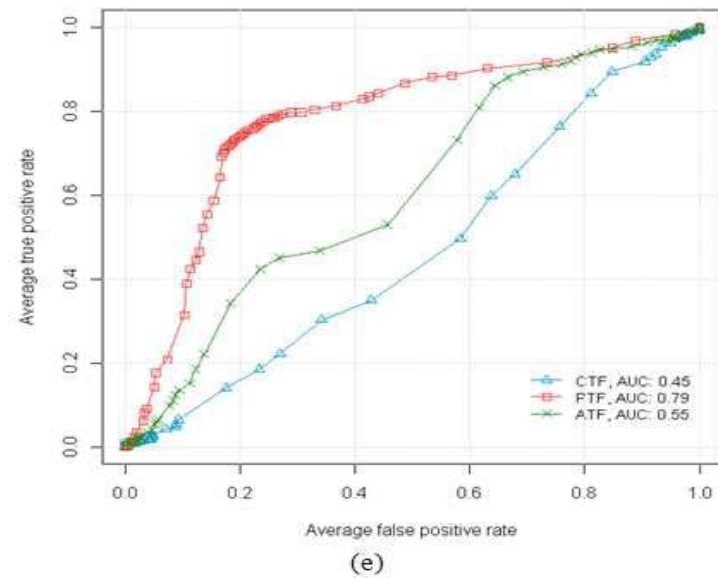
도면25



도면26a



도면26b



도면27

| Lumen<br>Features           | Normal ( $df=79$ )          | Grade 1( $df=79$ )          | Grade 2( $df=79$ )          | F-tests ( $\mu_N=\mu_{G1}=\mu_{G2}$ ) |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|---------------------------------------|
|                             | $\bar{N}$ (C.I)             | $\bar{G1}$ (C.I)            | $\bar{G2}$ (C.I)            | F-value                               |
| Area( $\mu m^2$ )           | 3.68E+01 ( $\pm 1.57E+00$ ) | 4.09E+01 ( $\pm 1.67E+00$ ) | 3.84E+01 ( $\pm 1.57E+00$ ) | 1.13E+01*                             |
| Perimeter( $\mu m$ )        | 2.66E+01 ( $\pm 6.35E-01$ ) | 2.95E+01 ( $\pm 7.09E-01$ ) | 2.84E+01 ( $\pm 6.67E-01$ ) | 3.30E+01*                             |
| Width( $\mu m$ )            | 7.50E+00 ( $\pm 1.85E-01$ ) | 8.48E+00 ( $\pm 2.34E-01$ ) | 7.97E+00 ( $\pm 2.18E-01$ ) | 3.64E+01*                             |
| Height( $\mu m$ )           | 7.92E+00 ( $\pm 2.20E-01$ ) | 8.46E+00 ( $\pm 2.57E-01$ ) | 8.31E+00 ( $\pm 2.41E-01$ ) | 9.47E+00*                             |
| MajorAxis( $\mu m$ )        | 8.58E+00 ( $\pm 1.76E-01$ ) | 9.68E+00 ( $\pm 1.78E-01$ ) | 9.28E+00 ( $\pm 1.83E-01$ ) | 6.78E+01*                             |
| MinorAxis( $\mu m$ )        | 5.37E+00 ( $\pm 1.20E-01$ ) | 5.28E+00 ( $\pm 1.35E-01$ ) | 5.17E+00 ( $\pm 1.28E-01$ ) | 4.17E+00                              |
| Circularity                 | 6.67E-01 ( $\pm 1.08E-02$ ) | 6.01E-01 ( $\pm 1.17E-02$ ) | 6.10E-01 ( $\pm 1.24E-02$ ) | 6.63E+01*                             |
| Feret's Diameter( $\mu m$ ) | 9.63E+00 ( $\pm 2.02E-01$ ) | 1.09E+01 ( $\pm 2.28E-01$ ) | 1.04E+01 ( $\pm 2.12E-01$ ) | 5.80E+01*                             |
| Skewness                    | 3.35E-01 ( $\pm 1.03E-01$ ) | 4.23E-01 ( $\pm 2.02E-01$ ) | 4.00E-01 ( $\pm 1.52E-01$ ) | 5.85E-01                              |
| AspectRatio                 | 1.66E+00 ( $\pm 3.58E-02$ ) | 1.98E+00 ( $\pm 6.07E-02$ ) | 1.91E+00 ( $\pm 5.62E-02$ ) | 7.55E+01*                             |
| Roundness                   | 6.50E-01 ( $\pm 1.14E-02$ ) | 5.73E-01 ( $\pm 1.25E-02$ ) | 5.84E-01 ( $\pm 1.21E-02$ ) | 8.30E+01*                             |
| Solidity                    | 8.38E-01 ( $\pm 4.40E-03$ ) | 8.19E-01 ( $\pm 4.58E-03$ ) | 8.19E-01 ( $\pm 4.81E-03$ ) | 3.96E+01*                             |

C.I: Confidence Interval

\*: It indicates features whose null hypothesis was rejected with  $F\text{-value} > F_{0.01}(2,237)$

도면28

| Lumen<br>Features           | Normal ( $df=79$ )          | Grade 1( $df=79$ )           | Grade 2( $df=79$ )   | F-tests ( $\mu_N=\mu_{G1}=\mu_{G2}$ ) |
|-----------------------------|-----------------------------|------------------------------|----------------------|---------------------------------------|
|                             | $\bar{N}$ (C.I)             | $\bar{G1}$ (C.I)             | $\bar{G2}$ (C.I)     | F-value                               |
| Area( $\mu m^2$ )           | 3.78E+04 ( $\pm 9.35E+03$ ) | 6.77E+04 ( $\pm 1.23E+04$ )  | 1.19E+05 (2.54E+04)  | 4.00E+01*                             |
| Perimeter( $\mu m$ )        | 1.49E+03 ( $\pm 1.78E+02$ ) | 2.09E+03 ( $\pm 1.74E+02$ )  | 2.73E+03 (2.69E+02)  | 6.05E+01*                             |
| Width( $\mu m$ )            | 1.86E+02 ( $\pm 2.64E+01$ ) | 2.73E+02 ( $\pm 2.93E+01$ )  | 3.66E+02 (4.70E+01)  | 4.48E+01*                             |
| Height( $\mu m$ )           | 1.79E+02 ( $\pm 1.99E+01$ ) | 2.39E+02 ( $\pm 2.31E+01$ )  | 3.05E+02 (2.78E+01)  | 4.88E+01*                             |
| MajorAxis( $\mu m$ )        | 2.29E+02 ( $\pm 2.94E+01$ ) | 3.36E+02 ( $\pm 3.23E+01$ )  | 4.36E+02 (5.01E+01)  | 5.10E+01*                             |
| MinorAxis( $\mu m$ )        | 1.83E+02 ( $\pm 2.09E+01$ ) | 2.42E+02 ( $\pm 2.01E+01$ )  | 3.20E+02 (2.99E+01)  | 5.70E+01*                             |
| Circularity                 | 1.86E-01 ( $\pm 1.10E-03$ ) | 1.82E-01 ( $\pm 2.70E-03$ )  | 1.83E-01 (2.59E-03)  | 5.91E+00*                             |
| Feret's Diameter( $\mu m$ ) | 5.31E+02 ( $\pm 6.42E+01$ ) | 7.53E+02 ( $\pm 6.43E+01$ )  | 9.84E+02 (1.00E+02)  | 5.87E+01*                             |
| Skewness                    | 2.15E-01 ( $\pm 1.71E-01$ ) | -7.27E-01 ( $\pm 1.91E-01$ ) | -9.88E-01 (1.71E-01) | 8.79E+01*                             |
| AspectRatio                 | 1.25E+00 ( $\pm 5.58E-02$ ) | 1.41E+00 ( $\pm 1.12E-01$ )  | 1.38E+00 (1.10E-01)  | 5.68E+00*                             |
| Roundness                   | 8.15E-01 ( $\pm 3.22E-02$ ) | 7.47E-01 ( $\pm 4.62E-02$ )  | 7.60E-01 (4.53E-02)  | 5.21E+00                              |
| Solidity                    | 1.00E+00 ( $\pm 0.00E+00$ ) | 1.00E+00 ( $\pm 0.00E+00$ )  | 1.00E+00 (0.00E+00)  | 1.00E+00                              |
| AtypiaRatio                 | 5.11E-02 ( $\pm 1.09E-02$ ) | 2.72E-01 ( $\pm 6.89E-02$ )  | 4.69E-01 (7.09E-02)  | 9.22E+01*                             |
| #AtypiaRegions              | 7.50E-02 ( $\pm 1.39E-01$ ) | 2.48E+00 ( $\pm 5.56E-01$ )  | 4.84E+00 (6.39E-01)  | 1.61E+02*                             |
| RMSAA                       | 1.03E+00 ( $\pm 2.60E-01$ ) | 8.85E+00 ( $\pm 2.97E+00$ )  | 1.88E+01 (3.63E+00)  | 7.51E+01*                             |
| TSAV(rad)                   | 1.95E+02 ( $\pm 1.88E+02$ ) | 8.65E+03 ( $\pm 3.45E+03$ )  | 2.61E+04 (6.45E+03)  | 6.80E+01*                             |

\*: It indicates features whose null hypothesis was rejected with  $F\text{-value} > F_{0.01}(2,237)$

도면29

| Lumen Features                | Normal ( $df=79$ )          | Grade 1( $df=79$ )          | Grade 2( $df=79$ )          | F-tests ( $\mu_N=\mu_{G1}=\mu_{G2}$ ) |
|-------------------------------|-----------------------------|-----------------------------|-----------------------------|---------------------------------------|
|                               | $\bar{N}$ (C.I)             | $\bar{G1}$ (C.I)            | $\bar{G2}$ (C.I)            | F-value                               |
| Area( $\mu m^2$ )             | 4.22E+01 ( $\pm 2.67E+00$ ) | 5.26E+01 ( $\pm 2.82E+00$ ) | 4.83E+01 ( $\pm 2.94E+00$ ) | 2.43E+01*                             |
| Perimeter( $\mu m$ )          | 2.81E+01 ( $\pm 1.09E+00$ ) | 3.20E+01 ( $\pm 9.63E-01$ ) | 3.10E+01 ( $\pm 9.88E-01$ ) | 2.81E+01*                             |
| Width( $\mu m$ )              | 8.04E+00 ( $\pm 3.11E-01$ ) | 9.11E+00 ( $\pm 2.51E-01$ ) | 8.65E+00 ( $\pm 2.72E-01$ ) | 2.58E+01*                             |
| Height( $\mu m$ )             | 8.16E+00 ( $\pm 3.38E-01$ ) | 9.19E+00 ( $\pm 3.17E-01$ ) | 8.92E+00 ( $\pm 2.59E-01$ ) | 2.10E+01*                             |
| MajorAxis( $\mu m$ )          | 8.87E+00 ( $\pm 2.99E-01$ ) | 1.01E+01 ( $\pm 2.55E-01$ ) | 9.66E+00 ( $\pm 2.42E-01$ ) | 3.73E+01*                             |
| MinorAxis( $\mu m$ )          | 5.92E+00 ( $\pm 2.01E-01$ ) | 6.45E+00 ( $\pm 2.10E-01$ ) | 6.13E+00 ( $\pm 2.25E-01$ ) | 1.10E+01*                             |
| Circularity                   | 6.87E-01 ( $\pm 1.95E-02$ ) | 6.47E-01 ( $\pm 1.69E-02$ ) | 6.33E-01 ( $\pm 1.42E-02$ ) | 1.86E+01*                             |
| Feret's Diameter( $\mu m$ )   | 9.94E+00 ( $\pm 3.52E-01$ ) | 1.13E+01 ( $\pm 2.97E-01$ ) | 1.09E+01 ( $\pm 2.83E-01$ ) | 3.46E+01*                             |
| Skewness                      | 3.42E-01 ( $\pm 1.19E-01$ ) | 1.79E-01 ( $\pm 8.28E-02$ ) | 1.86E-01 ( $\pm 9.83E-02$ ) | 5.84E+00*                             |
| AspectRatio                   | 1.51E+00 ( $\pm 4.77E-02$ ) | 1.60E+00 ( $\pm 4.40E-02$ ) | 1.62E+00 ( $\pm 4.14E-02$ ) | 1.23E+01*                             |
| Roundness                     | 6.88E-01 ( $\pm 1.73E-02$ ) | 6.56E-01 ( $\pm 1.41E-02$ ) | 6.50E-01 ( $\pm 1.29E-02$ ) | 1.29E+01*                             |
| Solidity                      | 8.48E-01 ( $\pm 7.22E-03$ ) | 8.41E-01 ( $\pm 7.98E-03$ ) | 8.32E-01 ( $\pm 6.66E-03$ ) | 8.87E+00*                             |
| CytoplasmLength( $\mu m$ )    | 6.16E+00 ( $\pm 8.20E-01$ ) | 1.29E+01 ( $\pm 1.16E+00$ ) | 1.59E+01 ( $\pm 1.01E+00$ ) | 1.72E+02*                             |
| CytoplasmLength <sub>SD</sub> | 2.10E+00 ( $\pm 4.06E-01$ ) | 5.75E+00 ( $\pm 7.69E-01$ ) | 8.47E+00 ( $\pm 7.17E-01$ ) | 1.68E+02*                             |

\*: It indicates features whose null hypothesis was rejected with  $F\text{-value} > F_{0.01}(2,237)$

도면30

| Non-Epithelial Nuclei Features | LSD-test               |                        |                         | LSD-value |
|--------------------------------|------------------------|------------------------|-------------------------|-----------|
|                                | $ \bar{N} - \bar{G1} $ | $ \bar{N} - \bar{G2} $ | $ \bar{G1} - \bar{G2} $ |           |
| Area( $\mu m^2$ )              | 4.08.E+00*             | 1.66.E+00              | 2.42.E+00*              | 2.23.E+00 |
| Perimeter( $\mu m$ )           | 2.89.E+00*             | 1.81.E+00*             | 1.09.E+00*              | 9.34.E-01 |
| Width( $\mu m$ )               | 9.75.E-01*             | 4.68.E-01*             | 5.07.E-01*              | 2.97.E-01 |
| Height( $\mu m$ )              | 5.41.E-01*             | 3.92.E-01*             | 1.49.E-01               | 3.34.E-01 |
| MajorAxis( $\mu m$ )           | 1.10.E+00*             | 7.04.E-01*             | 3.99.E-01*              | 2.49.E-01 |
| MinorAxis( $\mu m$ )           | -                      | -                      | -                       | -         |
| Circularity                    | 6.66.E-02*             | 5.68.E-02*             | 9.75.E-03               | 1.62.E-02 |
| Feret's Diameter( $\mu m$ )    | 1.22.E+00*             | 7.66.E-01*             | 4.57.E-01*              | 2.98.E-01 |
| Skewness                       | -                      | -                      | -                       | -         |
| AspectRatio                    | 3.24.E-01*             | 2.60.E-01*             | 6.40.E-02               | 7.24.E-02 |
| Roundness                      | 7.64.E-02*             | 6.60.E-02*             | 1.04.E-02               | 1.67.E-02 |
| Solidity                       | 1.92.E-02*             | 1.88.E-02*             | 3.46.E-04               | 6.40.E-03 |

\*: It indicates that the absolute pairwise difference is greater than LSD-value.

도면31

| Lumen<br>Features                 | LSD-test               |                        |                         |           |
|-----------------------------------|------------------------|------------------------|-------------------------|-----------|
|                                   | $ \bar{N} - \bar{G1} $ | $ \bar{N} - \bar{G2} $ | $ \bar{G1} - \bar{G2} $ | LSD-value |
| Area( $\mu\text{m}^2$ )           | 2.99E+04*              | 8.14E+04*              | 5.15E+04*               | 2.39E+04  |
| Perimeter( $\mu\text{m}$ )        | 6.01E+02*              | 1.25E+03*              | 6.46E+02*               | 2.95E+02  |
| Width( $\mu\text{m}$ )            | 8.69E+01*              | 1.80E+02*              | 9.27E+01*               | 4.93E+01  |
| Height( $\mu\text{m}$ )           | 6.05E+01*              | 1.26E+02*              | 6.56E+01*               | 3.32E+01  |
| MajorAxis( $\mu\text{m}$ )        | 1.07E+02*              | 2.08E+02*              | 1.01E+02*               | 5.34E+01  |
| MinorAxis( $\mu\text{m}$ )        | 5.91E+01*              | 1.37E+02*              | 7.80E+01*               | 3.34E+01  |
| Circularity                       | 3.90E-03*              | 3.18E-03*              | 7.25E-04                | 3.13E-03  |
| Feret's Diameter( $\mu\text{m}$ ) | 2.22E+02*              | 4.54E+02*              | 2.32E+02*               | 1.09E+02  |
| Skewness                          | 9.41E-01*              | 1.20E+00*              | 2.61E-01*               | 2.48E-01  |
| AspectRatio                       | 1.63E-01*              | 1.34E-01               | 2.90E-02                | 1.34E-01  |
| Roundness                         | -                      | -                      | -                       | -         |
| Solidity                          | -                      | -                      | -                       | -         |
| AtypiaRatio                       | 2.21E-01*              | 4.17E-01*              | 1.97E-01*               | 7.99E-02  |
| #AtypiaRegions                    | 2.40E+00*              | 4.76E+00*              | 2.36E+00*               | 6.90E-01  |
| RMSAA                             | 7.82E+00*              | 1.77E+01*              | 9.92E+00*               | 3.77E+00  |
| TSAV(rad)                         | 8.45E+03*              | 2.59E+04*              | 1.74E+04*               | 5.88E+03  |

\*: It indicates that the absolute pairwise difference is greater than LSD-value.

도면32

| Epithelial<br>Nuclei Features     | LSD-test               |                        |                         |           |
|-----------------------------------|------------------------|------------------------|-------------------------|-----------|
|                                   | $ \bar{N} - \bar{G1} $ | $ \bar{N} - \bar{G2} $ | $ \bar{G1} - \bar{G2} $ | LSD-value |
| Area( $\mu\text{m}^2$ )           | 1.04.E+01*             | 6.19.E+00*             | 4.26.E+00*              | 3.91.E+00 |
| Perimeter( $\mu\text{m}$ )        | 3.92.E+00*             | 2.92.E+00*             | 1.00.E+00               | 1.41.E+00 |
| Width( $\mu\text{m}$ )            | 1.07.E+00*             | 6.07.E-01*             | 4.66.E-01*              | 3.89.E-01 |
| Height( $\mu\text{m}$ )           | 1.02.E+00*             | 7.60.E-01*             | 2.65.E-01               | 4.27.E-01 |
| MajorAxis( $\mu\text{m}$ )        | 1.21.E+00*             | 7.96.E-01*             | 4.19.E-01*              | 3.71.E-01 |
| MinorAxis( $\mu\text{m}$ )        | 5.30.E-01*             | 2.19.E-01              | 3.12.E-01*              | 2.95.E-01 |
| Circularity                       | 3.96.E-02*             | 5.35.E-02*             | 1.38.E-02               | 2.36.E-02 |
| Feret's Diameter( $\mu\text{m}$ ) | 1.36.E+00*             | 9.36.E-01*             | 4.22.E-01               | 4.34.E-01 |
| Skewness                          | 1.64.E-01*             | 1.56.E-01*             | 7.53.E-03               | 1.40.E-01 |
| AspectRatio                       | 9.12.E-02*             | 1.11.E-01*             | 1.95.E-02               | 6.18.E-02 |
| Roundness                         | 3.19.E-02*             | 3.78.E-02*             | 5.90.E-03               | 2.07.E-02 |
| Solidity                          | 7.74.E-03              | 1.65.E-02*             | 8.74.E-03               | 1.02.E-02 |
| CytoplasmLength( $\mu\text{m}$ )  | 6.78.E+00*             | 9.78.E+00*             | 3.00.E+00*              | 1.40.E+00 |
| CytoplasmLength <sub>SD</sub>     | 3.65.E+00*             | 6.37.E+00*             | 2.72.E+00*              | 9.05.E-01 |

\*: It indicates that the absolute pairwise difference is greater than LSD-value.