(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0079132
(43) 공개일자 2021년06월29일

(51) 국제특허분류(Int. Cl.)
G16H 30/40 (2018.01) *G06N 20/00* (2019.01)
G16H 30/20 (2018.01) *G16H 50/20* (2018.01)
G16H 50/30 (2018.01) *G16H 50/50* (2018.01)
(52) CPC특허분류
G16H 30/40 (2018.01)
G06N 20/00 (2019.01)
(21) 출원번호 10-2019-0171289
(22) 출원일자 2019년12월19일
심사청구일자 2019년12월19일

(71) 출원인
인제대학교 산학협력단
경남 김해시 인제로 197, 내 (어방동, 인제대학교)
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
최홍국
경상남도 양산시 물금읍 야리로 89 양산물금양우내안애3차센트럴아파트 302-902
수브라타 바타차지
경상남도 김해시 인제로170번길 9-16 목연오피스텔1차 807호
(뒷면에 계속)
(74) 대리인
특허법인세원

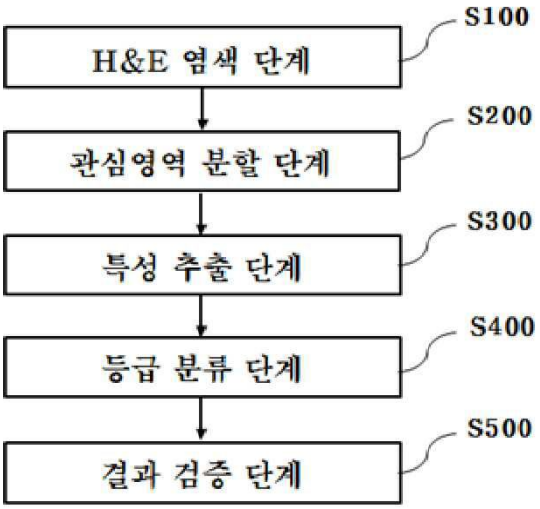
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법**

(57) 요 약

본 발명은 프로세서와 메모리를 구비한 의료 영상 처리 시스템에 의해 수행되는 전립선암의 등급 분류 방법으로서, 주어진 조직 병리 영상 데이터에 대하여 관심영역 추출을 위해 헤마톡실린과 에오신을 이용해 상기 조직 병리 영상을 염색하는 H&E 염색 단계, 상기 염색된 영상에서 버팀질(Stroma), 내부공간(Lumen) 및 세포 핵(Nuclei)을 식별하기 위한 영상을 분할하는 관심영역 분할 단계, 전립선암 등급 분류를 위하여 피셔계수(Fisher coefficient) 및 일원분산분석(One-way Analysis of Variance)을 사용하여 세포 핵과 내부공간 분할 영상에서 형태학적 특성을 추출하는 특성 추출 단계, 상기 형태학적 특성 추출값에 대해 서포트 벡터 머신(SVM)을 이용하여 전립선암의 글리슨 등급을 예측 분류하는 등급 분류 단계를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G16H 30/20 (2018.01)
G16H 50/20 (2018.01)
G16H 50/30 (2018.01)
G16H 50/50 (2018.01)

(72) 발명자
박현균
경상남도 거제시 국산2길 14 석천아트타운 107동 107호
김초희
경상남도 김해시 가락로294번길 18 구산주공3단지 아파트 307동 202호
누안 마두산카
경상남도 김해시 삼안로 255번길 44-16 107호
딕쉬타 프락카쉬
경상남도 김해시 인제로170번길 21-12, 607호
조남훈
서울특별시 강남구 언주로130길 30 동양파라곤 103-301

이 발명을 지원한 국가연구개발사업

과제고유번호	P0002072
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술진흥원
연구사업명	지역혁신클러스터육성(R&D)
연구과제명	충북 공공기관연계 지능정보가속화 개방형 생태계 조성사업
기 여 율	1/1
과제수행기관명	(재)충북테크노파크
연구기간	2018.04.01 ~ 2018.12.31

명 세 서

청구범위

청구항 1

프로세서와 메모리를 구비한 의료 영상 처리 시스템에 의해 수행되는 전립선암의 등급 분류 방법에 있어서,

주어진 조직 병리 영상 데이터에 대하여 관심영역 추출을 위해 헤마톡실린과 에오신을 이용해 상기 조직 병리 영상을 염색하는 H&E 염색 단계,

상기 염색된 영상에서 버팀질(Stroma), 내부공간(Lumen) 및 세포 핵(Nuclei)을 식별하기 위한 영상을 분할하는 관심영역 분할 단계,

전립선암 등급 분류를 위하여 피셔계수(Fisher coefficient) 및 일원분산분석(One-way Analysis of Variance)을 사용하여 세포 핵과 내부공간 분할 영상에서 형태학적 특성을 추출하는 특성 추출 단계, 및

상기 형태학적 특성 추출값에 대해 서포트 벡터 머신(SVM)을 이용하여 전립선암의 글리슨 등급을 예측 분류하는 등급 분류 단계를 포함하여 이루어지는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

청구항 2

청구항 1에 있어서,

상기 관심영역 분할 단계는 K-평균 군집화 알고리즘을 이용하여,

1단계로 생성할 군집 수로 k=3을 지정하고, 2단계로 군집의 중심이 될 k개의 지점을 무작위로 선택하고, 3단계로 유클리디언 거리(Euclidean distance)를 사용하여 각 데이터를 가장 가까운 군집에 할당하며, 4단계로 각 군집에 대한 중심 평균을 계산하여 새로운 군집 중심으로 사용하고, 5단계로 모든 데이터들을 가장 가까운 군집 중심에 다시 할당하며, 6단계로 군집 중심에 변화가 없을 때까지 반복하는 과정을 통하여 결과를 수정하는 과정을 수행하여 내부공간 영상과 세포 핵의 영상을 분할하는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

청구항 3

청구항 2에 있어서,

상기 세포 핵의 분할된 영상에서 겹쳐 있는 세포 핵들을 분리하기 위하여 워터쉐드(WaterShed) 알고리즘을 통한 중복된 세포 핵을 분리하는 과정을 수행하는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

청구항 4

청구항 3에 있어서,

상기 워터쉐드(WaterShed) 알고리즘을 통한 중복된 세포 핵을 분리하는 과정에서 내부 마커를 이용한 마커 선택 워터쉐드 변환 접근방식을 이용하여 과분할(Over-segmentaion) 문제를 해결하는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

청구항 5

청구항 1에 있어서,

상기 특성 추출 단계에서의 세포 핵과 내부공간의 형태학적 특성으로는 면적(Area), 둘레(Perimeter), 주축 길이(Major axis length), 단축 길이(Minor axis length), 원형성(Circularity), 지름(Diameter), 핵에서 핵까지의 거리(Nucleus to nucleus distance), 핵에서 핵까지의 최소 거리(Nucleus to nucleus minimum distance), 세포 핵의 모양(Eccentricity), 빽빽함(Compactness)을 포함하는 것을 특징으로 하는 서포트 벡터 머신을 이용

한 전립선암의 등급 분류 방법.

청구항 6

청구항 1에 있어서,

상기 등급 분류 단계의 서포트 벡터 머신을 통한 분류는 선형(Linear) 분류기와 가우시안(Gaussian) 분류기를 사용한 이진 분류 접근법에 따라 악성 대 양성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 이진화하여 분류하는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

청구항 7

청구항 1에 있어서,

상기 등급 분류 단계 이후, 서포트 벡터 머신 분류의 혼동 행렬(Confusion matrix)을 사용하여 훈련과 시험 과정에 있어서 이진분류의 정확도(Accuracy), 민감도(Sensitively), 특이도(Specificity) 및 매튜의 상관 계수(Matthews' s correlation coefficient)의 성능 측정지표를 통한 분류 결과를 검증 평가하는 결과 검증 단계를 더 포함하여 이루어지는 것을 특징으로 하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법에 관한 것으로서, 더욱 상세하게는 전립선암의 디지털화된 조직병리학 영상에 대한 중요한 특성들을 추출하고 서포트 벡터 머신(Support Vector Machine)을 이용한 기계학습 분석을 통하여 전립선암의 등급을 분류하는 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법에 관한 것이다.

배 경 기 술

[0002] 선암(adenocarcinoma)은 상피조직의 샘 선 구조로부터 형성된 악성 암세포 조직의 한 유형이다. 상기 선암의 일종인 전립선암(prostate cancer)은 두 번째로 흔히 진단되는 암으로서, 남성의 전립선에 발생하는 악성종양이며, 발병률이 높은 고위험군의 병으로 알려져 있다.

[0003] 특히 전립선암은 미국에서 두 번째로 흔한 암 사망 원인으로 알려져 있을 뿐만 아니라, 최근 우리나라도 고령화 사회로 진입하면서 전립선암의 유병률이 증가한 것으로 보이는데, 그 이유는 전립선암의 가장 중요한 위험 인자가 나이로 주로 60대 이후 노인에게 전립선암이 많이 발생하기 때문이다. 하지만 50대 이하의 남성에게서도 많이 발병하고 있는 실정이다.

[0004] 따라서 전립선암은 꾸준한 관리가 필요한 만성 질환으로 자리 잡고 있으며, 이에 대한 전립선암의 정확한 진단이 필요한 시점이다. 또한, 전립선암은 예측하기 어려운 병의 진행 형태를 보이기 때문에 예후를 통해 사전에 예측하는 것이 필요하다.

[0005] 따라서 전립선암은 진단과 치료에 있어 항상 병리학자와 의료전문가들에게 중요한 관심사였다. 보통, 조직병리학에서 암 진단 과정은 생체실험 영상을 악성 및 양성으로 구분하는 것에서 시작된다.

[0006] 전립선암의 분류는 병리학자인 병리학자 도널 글리슨(Donald F Gleason) 박사와 VACURG(Vervetary Administration Cooperative Urolic Research Group)의 회원들에 의해 개발된 글리슨점수(Gleason Score)가 주로 사용되고 있으며, 종양조직의 분화 정도와 세포의 특성에 따라 유형 등급(Grade) 1부터 등급 5으로 나뉘게 된다. 등급 1은 뚜렷한 선 구조를 가지며, 작고 균일한 세포로 이루어져있지만, 등급 5로 갈수록 점점 불규칙적이며, 뚜렷한 분비선이 없어지는 구조를 갖는다. 이러한 등급 체계는 다수의 환자를 대상으로 시험되었으며, 뛰어난 성공 사례로 평가된다

[0007] 그러나 현미경을 통하여 병리학자와 의사들이 생체검사 조직 영상을 판단하여 전립선암을 진단하는 것은 쉽지가 않을뿐더러, 정확한 분류를 하는데 어려움이 따른다.

[0008] 이러한 본 발명의 배경기술로서 대한민국 공개특허공보 10-2018-0091766(2018.08.16.)호에서 전립선암 진단에 관한 정보제공방법이 공지되어 있다.

[0009] 그러나 상기의 종래 기술은 경직장초음파 또는 자기공명영상을 통하여 획득된 영상을 수치화하고 이를 적용한

알고리즘을 이용하여 객관적으로 전립선암에 관한 정보를 제공하는 방법 및 이를 이용한 진단장치를 제공하는 것으로, 알고리즘을 이용하여 객관적으로 전립선암에 관한 정보를 제공하고 있으나, 전립선암의 정확한 등급을 분류하는데 이용하는데는 한계가 있다.

발명의 내용

해결하려는 과제

[0010] 본 발명은 전술한 종래기술의 문제점을 해결하기 위해 도출된 것으로, 본 발명의 목적은, 전립선 등급 영상이미지에 대하여 영상 전처리 후 중요한 특성들을 추출하고 기계학습 중의 지도 학습 모델인 서포트 벡터 머신(Support Vector Machine)분류를 통한 전립선 영상의 등급을 글리슨 등급에 기반하여 일관되고 정확하게 분류하는 SVM을 이용한 전립선암의 정량적 등급 분류 방법을 제공하고자 하는 것이다.

[0011] 본 발명의 다른 목적은 생체검사 조직 영상 처리를 위해 K-평균 군집화에 기초한 색 분할을 제안하고, 조직 영상에서 겹쳐있는 세포 핵을 분리하는 워터쉐드 알고리즘을 구현하며, 마커 선택 접근방식을 이용하여 과분할(Over-segmentaion)을 방지하는 전립선암의 정량적 등급 분류 방법을 제공하고자 하는 것이다.

과제의 해결 수단

[0012] 상기 기술적 과제를 해결하기 위한 본 발명의 일 측면에 따른 프로세서와 메모리를 구비한 의료 영상 처리 시스템에 의해 수행되는 전립선암의 등급 분류 방법은, 주어진 조직 병리 영상 데이터에 대하여 관심영역 추출을 위해 헤마톡실린과 에오신을 이용해 상기 조직 병리 영상을 염색하는 H&E 염색 단계, 상기 염색된 영상에서 버팀질(Stroma), 내부공간(Lumen) 및 세포 핵(Nuclei)을 식별하기 위한 영상을 분할하는 관심영역 분할 단계, 전립선암 등급 분류를 위하여 피셔계수(Fisher coefficient) 및 일원분산분석(One-way Analysis of Variance)을 사용하여 세포 핵과 내부공간 분할 영상에서 형태학적 특성을 추출하는 특성 추출 단계, 상기 형태학적 특성 추출값에 대해 서포트 벡터 머신(SVM)을 이용하여 전립선암의 글리슨 등급을 예측 분류하는 등급 분류 단계를 포함하여 이루어지는 특징이 있다.

[0013] 또한, 본 발명의 상기 관심영역 분할 단계는 K-평균 군집화 알고리즘을 이용하여, 1단계로 생성할 군집 수로 k=3을 지정하고, 2단계로 군집의 중심이 될 상기 k개의 지점을 무작위로 선택하고, 3단계로 유클리디언 거리(Euclidean distance)를 사용하여 각 데이터를 가장 가까운 군집에 할당하며, 4단계로 각 군집에 대한 중심 평균을 계산하여 새로운 군집 중심으로 사용하고, 5단계로 모든 데이터들을 가장 가까운 군집 중심에 다시 할당하며, 6단계로 군집 중심에 변화가 없을 때까지 반복하는 과정을 통하여 결과를 수정하는 과정을 수행하여 내부공간 영상과 세포 핵의 영상을 분할하는 특징이 있다.

[0014] 또한, 본 발명은 상기 세포 핵의 분할된 영상에서 겹쳐 있는 세포 핵들을 분리하기 위하여 워터쉐드(WaterShed) 알고리즘을 통한 중복된 세포 핵을 분리하는 과정을 수행하는 특징이 있다.

[0015] 또한, 본 발명은 상기 워터쉐드(WaterShed) 알고리즘을 통한 중복된 세포 핵을 분리하는 과정에서 내부 마커를 이용한 마커 선택 워터쉐드 변환 접근방식을 이용하여 과분할(Over-segmentaion) 문제를 해결하는 특징이 있다.

[0016] 또한, 본 발명의 상기 특성 추출 단계에서의 세포 핵과 내부공간의 형태학적 특성으로는 면적(Area), 둘레(Perimeter), 주축 길이(Major axis length), 단축 길이(Minor axis length), 원형성(Circularity), 지름(Diameter), 핵에서 핵까지의 거리(Nucleus to nucleus distance), 핵에서 핵까지의 최소 거리(Nucleus to nucleus minimum distance), 세포 핵의 모양(Eccentricity), 빽빽함(Compactness)을 포함하는 것을 특징으로 할 수 있다.

[0017] 또한, 본 발명의 상기 등급 분류 단계의 서포트 벡터 머신을 통한 분류는 선형(Linear) 분류기와 가우시안(Gaussian) 분류기를 사용한 이진 분류 접근법에 따라 악성 대 양성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 이진화하여 분류하는 특징이 있다.

[0018] 또한, 본 발명은 상기 등급 분류 단계 이후, 서포트 벡터 머신 분류의 혼동 행렬(Confusion matrix)을 사용하여 훈련과 시험 과정에 있어서 이진분류의 정확도(Accuracy), 민감도(Sensitively), 특이도(Specificity) 및 매튜의 상관 계수(Matthews's correlation coefficient)의 성능 측정지표를 통한 분류 결과를 검증 평가하는 결과 검증 단계를 더 포함하여 이루어질 수 있다.

발명의 효과

[0019] 전술한 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법에 의한 본 발명은 디지털화된 조직병리학 영상에 대한 등급 분류 시스템을 제공함으로서, 추출된 형태학적 특성을 사용하여 만족스러운 분류 결과를 얻을 수 있다.

[0020] 또한, 본 발명에 의하면 서포트 벡터 머신 이진 분류법에 의한 분류 모델에 따라 악성 대 양성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 분류하는데 있어서, 다른 종래의 문헌들과 비교하여 높은 수준의 분류 결과를 도출할 수 있는 효과가 있다.

[0021] 이에 본 발명은 전립선 영상의 등급을 글리슨 등급에 기반하여 일관되고 정확하게 분류가 가능한 효과를 제공함을 알 수 있다.

도면의 간단한 설명

[0022] 도 1은 본 발명에 따른 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법의 단계를 보여주는 흐름도이고,

도 2는 본 발명에 따른 전립선암의 등급 분류 시스템의 구성을 나타내는 블록도이다.

도 3은 본 발명에 따른 전립선암을 분류하는 데 사용된 조직 병리 슬라이드 이미지의 예시도이고,

도 4는 본 발명에 따른 K-평균 군집화 알고리즘을 적용하여 관심영역을 분할한 영상의 예시도이며,

도 5는 본 발명에 따른 워터쉐드 알고리즘의 단계별 진행 영상을 나타내는 이미지이고,

도 6은 본 발명에 따른 워터쉐드 알고리즘의 과분할 문제 해결 과정의 단계별 영상을 보여주는 예시도이다.

도 7은 본 발명에 따른 이진 분류 방법의 분류방식을 보여주는 예시도이다.

도 8은 본 발명에 따른 세 가지의 이진 분류들의 서포트 벡터 머신 분류 정확도를 비교하는 그래프이고,

도 9는 본 발명에 따른 네 가지 등급 그룹간 서포트 벡터 머신 결과를 비교하는 그래프이다.

도 10은 본 발명에 따른 서포트 벡터 머신을 이용한 전립선암의 등급 분류에 대한 전체 구성을 보여주는 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0023] 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다.

[0024] 따라서 본 명세서에 기재된 실시예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시예에 불과할 뿐이고 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형예들이 있을 수 있음을 이해하여야 한다.

[0025] 이하 본 발명의 바람직한 실시예를 첨부된 도면을 참조하여 상세히 설명하면 아래와 같다.

[0026] 도 1은 본 발명에 따른 서포트 벡터 머신을 이용한 전립선암의 등급 분류 방법의 단계를 보여주는 흐름도이고, 도 2는 전립선암의 등급 분류 시스템의 블록을 나타내는 구성도이다.

[0027] 도 1에서와 같이 본 발명의 전립선암의 정량적 등급 분류 방법은 주어진 조직 병리 영상 데이터에 대하여 H&E 염색 단계(S100), 관심영역 분할 단계(S200), 특성 추출 단계(S300), 등급 분류 단계(S400) 및 결과 검증 단계(S500)의 과정을 통하여 이루어진다.

[0028] 또한, 본 발명은 도 2에서와 같이 의료 영상 처리 시스템(100)을 통하여 수행되는 것으로서, 데이터베이스(DB)와 연동하는 컴퓨터 장치로 이루어지되, 중앙처리장치(CPU)는 예를 들어 인텔(Intel) i5-4660, 메모리 장치(RAM) 24GB 또는 그 이상의 사양일 수 있고, 그래픽처리장치(GPU)는 GTX1060 3GByte의 환경이나 그 이상의 처리 환경에서 이루어졌으며, 영상에서의 세포 핵과 내부공간을 식별하고 체계적인 처리를 수행하기 위해 MATLAB R2018a(The MathWorks, Naticks, MA, USA)를 통하여 본 발명의 일정 방법을 구현하였으며, 또한, 자바로 개발된 기계 학습 소프트웨어 제품군인 웨카(Weka, Waikato Environment for Knowledge Analysis)를 사용하였으며, 통계학적인 분석을 위하여 통계 프로그램 패키지(SPSS, Statistical Package for the Social Sciences) 툴(Tool)을 사용하였다.

[0029] 또한, 의료 영상 처리 시스템(100)은, 중앙처리장치를 포함한 프로세서와 이 프로세서에 연결되는 메모리를 구비할 수 있다. 프로세서는 메모리에 저장된 프로그램을 실행하여 전립선암의 정량적 등급을 분류할 수 있다. 이를 위해 메모리에는 H&E 염색 단계(S100)의 실행을 위한 제1 모듈, 관심영역 분할 단계(S200)의 실행을 위한 제2 모듈, 특성 추출 단계(S300)의 실행을 위한 제3 모듈, 등급 분류 단계(S400)의 실행을 위한 제4 모듈, 및 결과 검증 단계(S500)의 실행을 위한 제5 모듈을 포함할 수 있다. 제1 모듈 내지 제5 모듈은 전립선암의 정량적 등급 분류를 위한 소프트웨어 모듈로서 적어도 일부가 메모리에 저장될 수 있고, 적어도 하나의 프로세서에 탑재되어 각각의 기능 및 동작과 상호 연관된 일련의 동작 및 기능을 수행할 수 있다.

[0030] 도 3은 본 발명의 일실시예에 따른 세포 핵을 검출하고 전립선암을 분류하는 데 사용된 조직 병리 슬라이드 영상 데이터이다.

[0031] 도 3의 (a) 내지 (d)는 헤마톡실린과 에오신으로 염색된 생체검사 영상을 보여주는 것으로서, 도 3의 (a)는 등급 3, (b)는 등급 4, (c)는 등급 5, (d)는 양성의 전체 슬라이드 조직 영상이다.

[0032] 도 3의 (a) 내지 (d)는 주어진 조직 병리 영상 데이터에 대하여 H&E 염색 단계를 수행한 것이다. 그리고 도 3의 (e) 내지 (h)는 상기 (a) 내지 (d)에서 각각 추출한 관심영역의 영상을 나타내고 있는 것이다.

[0033] 또한, 도 3의 (a) 내지 (d)의 전체 슬라이드 크기는 33,584 × 70,352 화소이며, 전체 영상을 40배 확대하여 512 × 512 크기로 추출한 영상은 도 3의 (e) 내지 (h)와 같다.

[0034] 도 3의 영상들은 세포 핵 군집들과 서로 겹치는 세포 핵들이 포함되어 있기 때문에 등급을 분류하는 데에 어려움이 있을 수 있다.

[0035] 이에 본 실시예의 의료 영상 처리 시스템(100)에서는 중요한 역할로서 관심영역 분할 과정(ROI Segmentation)을 수행한다.

[0036] 관심영역 분할 단계(S200)는 영상에서의 세포 핵과 내부공간을 식별하고 체계적인 처리를 수행하기 위해 K-평균 군집화 알고리즘(K-means Clustering)을 적용하여 수행한다.

[0037] 이는 전립선암의 핵 및 내부공간은 조직병리학 영상의 가장 중요한 성분이기 때문에 이를 분할하는 것이 중요하다.

[0038] 본 발명에서는 K-평균 군집화 알고리즘을 적용하여 영상들은 세 개의 군(k = 3)으로 분할된다. 즉 전립선암 영상에서 분할된 구성 요소로는 버팀질(Stroma), 내부공간(Lumen) 및 세포 핵(Nuclei)이 될 수 있다.

[0039] 도 4는 본 발명의 일실시예에 따른 K-평균 군집화 알고리즘을 적용하여 관심영역을 분할한 영상의 예시도로서, 도 4의 (a)는 원본 조직 영상이고, (b)는 내부공간 분할 영상이며, (c)는 세포 핵의 분할 영상이미지이다.

[0040] 도 4에서 볼 수 있듯이 K-평균 군집화 알고리즘은 생체검사 영상에 가장 적합할 수 있다. 본 실시예에서는 도 4의 영상 중 (b)의 내부공간 영상과 (c)의 세포 핵의 영상만을 분할하기 위하여 K-평균 군집화 알고리즘을 적용하였다. 여기서 K-평균 군집화 알고리즘은 대표적인 분리형 군집화 알고리즘 가운데 하나이다.

[0041] 각 군집은 하나의 중심(centroid)을 가지며, 각 개체는 가장 가까운 중심에 할당되고, 같은 중심에 할당된 개체들이 모여 하나의 군집을 형성한다. 이는 사용자가 사전에 군집 수(k)가 정해야 알고리즘을 실행할 수 있다는 것으로서, k는 하이퍼파라메터(hyperparameter)이다. 본 발명에서는 세 개의 군집 수, k = 3으로 분할하도록 하였다.

[0042] 또한, K-평균 군집화 알고리즘은 아래의 [식 1]과 [식 2]와 같이 데이터 할당 단계와 중심점 갱신 단계로 이루어진다.

[0043] 데이터 할당 단계는 [식 1]과 같이 표현될 수 있다.

[0044] [식 1]

[0045] $$\underset{c_k \in C}{\operatorname{argmin}}\, dist(c_k, x)^2$$

[0046] 그리고 중심점 갱신 단계는 [식 2]와 같이 표현될 수 있다.

[0047] $$c_k = \frac{1}{|s_k|} \sum_{x_* \in s_*} x_k$$

[0048] K-평균 군집화 알고리즘은 최종 결과가 나오기 전까지 반복해서 결과를 수정하는데, 다음과 같은 단계로 구성될 수 있다.

[0049] 1단계로, 생성할 군집 수 k개를 지정한다.

[0050] 2단계로, 군집의 중심이 될 k개의 지점을 무작위로 선택한다.

[0051] 3단계로, 유클리디언 거리(Euclidean distance)를 사용하여 각 데이터를 가장 가까운 군집에 할당한다.

[0052] 4단계로, 각 군집에 대한 중심 평균을 계산하여 새 군집 중심으로 사용한다.

[0053] 5단계로, 모든 데이터들을 가장 가까운 군집 중심에 다시 할당한다.

[0054] 6단계로, 군집 중심에 변화가 없을 때까지 반복한다.

[0055] 상기의 과정에 따라 생성된 분할 영상이미지를 통하여 특성 추출 단계, SVM을 이용한 등급 분류 단계(S400)를 수행할 수 있다.

[0056] 이때 K-평균 군집화 알고리즘을 이용하여 세포 핵만 분할한 영상에서 겹치는 세포 핵이 있다는 것이 관찰되어진다.

[0057] 이러한 문제를 해결하기 위하여 본 실시예에서는 겹쳐 있는 세포 핵들을 분리하는 과정을 수행한다.

[0058] 겹쳐진 객체들을 분할하기 위한 여러 방법론들이 있지만 가장 대중화되어 있으며 효과적인 방법으로 알려져 있는 워터쉐드(WaterShed) 알고리즘 방법을 이용할 수 있다. 이전부터 워터쉐드 알고리즘은 등급 분류를 위한 핵계수 및 형상 추출에 사용된 알고리즘이다.

[0059] 워터쉐드 알고리즘에 따르면, $q(x,y)$는 영상 화소 값, M_i 는 국부 최솟값이며 알고리즘의 반복 단계는 다음과 같이 [식 3] 내지 [식 9]를 사용하여 이루어질 수 있다.

[0060] [식 3]

[0061] $$T[n] = \{(x,y) | q(x,y) < n\}$$

[0062] [식 4]

[0063] $$n = \min + 1 \text{ to } n = \max + 1$$

[0064] [식 5]

[0065] $$C_n(M_n) = C(M_i) \cap T[n]$$

[0066] 여기서 $T[n]$은 $q(x,y)$, n은 범람 단계, $C_n(M_i)$은 유역 영역에 있는 점의 좌표 집합이다.

[0067] [식 6]

[0068] $$C_n(M_n) = 1, \text{ at } (x,y); \quad \text{if } (x,y) \in C(M_i) \text{ and } (x,y) \in T[n]$$

[0069] [식 7]

[0070] $$C_n(M_n) = 0, \text{ otherwise}$$

[0071] 위의 두 방정식을 계산 한 후 이진 영상에서 결과를 획득하였다.

[0072] [식 8]

[0073] $$C[n] = \bigcup_{i=1}^{R} C(M_i)$$

[0074] [식 9]

[0075] $$C[\max+1] = \bigcup_{i=1}^{R} C(M_i)$$

[0076] 여기서 $C[n]$는 설정 n에서 홍수 유역 부분의 결합이고, $C[\max+1]$는 모든 유역선의 집합이다. 식 (8)과 식 (9)에 따라 $C[n]$는 $T[n]$의 하위집합이고, $C[n-1]$은 $C[n]$의 하위집합이다. 따라서 $C[n-1]$의 각 연결된 구성요소는 $T[n]$의 정확히 하나의 연결 구성요소에 연결되어 있다.

[0077] 다음으로 워터쉐드 알고리즘은 겹쳐있는 핵을 분리하기 위해 다음의 단계들을 포함할 수 있다.

[0078] 1단계는, 적응 형 임계 값 지정 방법을 사용하여 24비트/픽셀 RGB 컬러 영상을 이진 영상으로 변환한다

[0079] 2단계는, 이진 영상에서 잡음을 제거한다.

[0080] 3단계는. 거리 맵을 생성하기 위해 이진 영상에 유클리디언 거리 변환을 적용한다.

[0081] 4단계는, 거리 맵을 부드럽게 하기 위해 가우스 필터를 사용한다.

[0082] 5단계는, 거리 맵을 평활화 한 후 역 거리 변환을 적용한다.

[0083] 6단계는, 역 거리 변환 영상의 마커를 사용하여 국부 최솟값을 식별한다.

[0084] 7단계는, 마지막으로, 국부 최소화 점을 기반으로 한 워터쉐드 알고리즘을 적용하여 겹쳐있는 모든 객체가 분할될 때까지 반복한다.

[0085] 이상과 같이 중복된 세포 핵을 분리하기 위해 상술된 과정을 통한 워터쉐드 알고리즘을 사용하였다.

[0086] 도 5는 본 발명의 일실시예에 따른 워터쉐드 알고리즘에 필요한 단계별 진행 영상을 보여주는 예시도이다.

[0087] 즉 도 5에 따른 워터쉐드 알고리즘의 단계에서, (a)는 원본에서 분할한 핵 조직 영상이고, (b) 잡음 제거한 이진 영상이며, (c)는 상기 이진 영상에서 유클리디언 거리 변환 적용 영상이고, (d)는 워터쉐드 알고리즘과 색상 사상(Mapping)을 사용한 색상 라벨링 된 핵 결과 영상을 보여주고 있는 것이다.

[0088] 그러나 워터쉐드 알고리즘은 도 5의 (a)에서와 같이 일부 객체가 너무 많은 부분으로 나뉘는 과분할 문제가 발생할 수 있다.

[0089] 도 6은 본 발명에 따른 워터쉐드 알고리즘의 과분할 문제와 이를 해결한 영상을 보여주는 예시도로서, (a)는 과분할된 세포 핵 영상, (b)는 역 거리 변환 결과가 적용된 마커 영상, (c)는 (b) 영상에서 워터쉐드 알고리즘이 적용된 결과 영상, (d)는 잡음과 유역선 제거 후 라벨을 붙인 결과 영상을 나타내고 있다.

[0090] 도 6에 나타나듯이 (a)는 과분할 된 세포 핵 영상으로서 내부 마커를 이용하여 문제를 해결할 수 있다.

[0091] 먼저, 과분할을 제어함과 동시에 분할 결과를 개선하기 위해 마커 선택 워터쉐드 변환이라는 접근방식을 사용한다.

[0092] 상기 접근방식은 각 관심영역의 지역 마커를 설정하여, 상기 지역 마커가 결과 영상의 유일한 국부 최소치들일 수 있도록 거리 맵 영상을 변환한 영상을 만드는 것이다.

[0093] 이에 따라 유클리디언 거리 변환 후 가우스 필터를 적용하여 거리 맵을 매끄럽게 한 후, 도 6의 (b)에서와 같이 내부 마커가 적용된 영상을 반전시켜, 역 거리 변환 결과 영상을 얻어낼 수 있게 된다.

[0094] 그리고 도 6의 (c)는 마커 영상을 사용하여 워터쉐드 알고리즘을 적용된 결과 영상이고, 마지막으로 잡음과 유역선을 제거한 후 결과 영상으로 도 6의 (d)와 같이 각 핵의 중심부에 라벨을 붙였다.

[0095] 이상과 같이 상술되어진 과정을 통하여 관심영역 분할 단계(S200)를 수행하였다.

[0096] 이후 특성 추출 단계(S300)는 병리영상의 분석과 등급 예측에 있어서 전립선암의 모양과 형태학적 특성을 추출하는 단계이다.

[0097] 이는 전립선암의 등급 분류를 위하여는 형태학적 특성과 질감 특성이 가장 중요할 수 있다.

[0098] 표 1은 본 발명에서 사용된 세포핵과 내부공간에서 추출한 특성에 대한 설명으로서, 통계적 분석방식에 따르면

이러한 모든 특성은 통계적으로 매우 유의하다 ($P < 0.001$).

[0099] 이에 본 발명에서는 핵 영상에서 워터쉐드 알고리즘을 수행한 후, 각 영상들의 특성을 추출하여 서포트 벡터 머신 분류방법으로 전립선암을 등급화하였다.

[0100] 또한, 본 발명에서는 분할된 핵 및 내부공간 영상에 영역 기반 및 등고선 기반 방법을 모두 사용하여 형태학적 특성에 대한 데이터를 수집하였다.

[0101] 또한, 본 발명에서는 유의한 특성을 찾기 위해 추출된 모든 특성을 비교하고 피셔계수(Fisher coefficient) 및 일원분산분석(One-way Analysis of Variance)을 사용하여 가장 중요한 특성을 식별하였다.

[0102] 또한, 본 발명에서는 조직 영상 데이터에서 추출하여 선택된 특성 값을 바탕으로 훈련과 시험을 실시하였으며, 총 19개의 특성을 세포 핵과 내부공간에서 추출하였다. 이 중 14개의 유의한 특성들을 서포트 벡터 머신 분류(SVM)에 사용하기 위하여 선정하였다.

[0103] 이에 따른 세포 핵과 내부공간의 형태학적 특성은 면적(Area), 둘레(Perimeter), 주축 길이(Major axis length), 단축 길이(Minor axis length), 원형성(Circularity), 지름(Diameter), 핵에서 핵까지의 평균 거리(Nucleus to nucleus average distance), 핵에서 핵까지의 최소 거리(Nucleus to nucleus minimum distance), 세포 핵의 모양(Eccentricity), 빽빽함(Compactness)이 될 수 있다.

[0104] 표 1은 본 발명에 따른 전립선암의 글리슨 등급 분류에 제안된 특성들을 나타내고 있다.

표 1

[0105]

Feature Type	Feature Description
Nucleus features	Area, perimeter, major axis length, minor axis length, circularity, diameter, compactness, nucleus to nucleus average distance, nucleus to nucleus minimum distance
Lumen features	Area, perimeter, major axis length, minor axis length, eccentricity

[0106] 본 발명의 등급 분류 단계는 세포 핵과 내부공간에 대한 형태학적 특성을 서포트 벡터 머신(SVM) 분류하여 전립선암의 글리슨 등급을 예측하는 단계이다.

[0107] 생체검사 영상을 다양한 글리슨 등급으로 분류하는 것은 매우 어려운 작업이며, 분류 정확도는 분류기와 사용되는 커널 유형에 따라 달라질 수 있다.

[0108] 상기 서포트 벡터 머신은 지도학습 방식이지만 분류 및 회귀 문제에 모두 적용된다. 또한, 마진(Margin)을 최대화하는 반복적인 방식으로 최적의 초 평면(Hyperplane)을 생성할 수 있으며, 여기서 마진은 초 평면으로부터의 거리가 가장 가까운 서로 다른 집단의 데이터 간격이 최대화되는 것을 의미한다.

[0109] 일반적으로 감독되는 학습 접근방식은 다음과 같이 진행한다:

[0110] 훈련 및 시험을 위한 데이터들을 준비하고, 적절한 알고리즘을 선택한 후 훈련 모델에 사용할 특성 값을 선택하고, 모델을 훈련시키고, 정확도를 도출하기 위하여 훈련된 모델을 사용해 평가를 하게된다.

[0111] 즉 본 발명의 등급 분류 단계(S400)를 통하여 서포트 벡터 머신 분류에서는 선형과 가우스 커널을 사용하여 전립선 영상을 양성 대 악성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 분류 하여 각 등급의 차이를 확인할 수 있다.

[0112] 또한, 모델을 훈련시키기 전 정확도를 개선하기 위해 K-겹 교차 검증(K-fold Cross Validation)을 수동으로 조정하였으며, 그 중 정확도가 높게 나온 두 겹 교차 검증(2-fold Cross Validation)으로 설정하여 학습 훈련을 한다.

[0113] 이때 선형 커널 K는 원본 데이터를 사상하여, 아래의 [식 10]과 같이 표현된다.

[0114] [식 10]

[0115] $$K(x) = (x, x' + c)$$

[0116] 여기서 x는 데이터이며 c는 상수이다.

[0117] 서포트 벡터 머신 분류에서의 가우스 커널 함수는 [식 11]과 같이 표현할 수 있다.

[0118] [식 11]

[0119] $$K(x, x') = \exp(-\gamma \|x - x'\|^2), \gamma = \frac{1}{2\sigma^2}$$

[0120] 여기서 x, x' 는 피쳐 벡터(Feature vector)이며, $\|x - x'\|^2$은 두 피쳐 벡터간의 유클리디언 거리이다. γ는 초평면으로 커널의 기울기를 변화시키며, σ는 자유 매개변수(Free parameter)이다.

[0121] 이상과 같이 서포트 벡터 머신(SVM)을 기반으로 하는 지도학습 접근방식을 사용하여 전립선암의 글리슨 등급을 예측할 수 있다.

[0122] 또한, 본 발명에서는 전립선암을 분류하기 위해 동시 분류를 이진으로 분류하는 이진 분류 접근법을 사용하였다.

[0123] 도 7은 본 발명에 따라 제안된 이진 분류 방법의 SVM 분류방식을 보여주는 예시도이다.

[0124] 본 발명의 이진 분류 방법은 각 등급의 분류마다 각기 다른 서포트 벡터 머신 분류기들을 이용하여 등급을 분류하는 것으로서, 도 7에서와 같이 첫 번째 분류에서는 모든 영상 데이터들을 악성 대 양성으로 분류하고, 이후 악성등급에서 등급3 대 등급 4와 5로 분류를 하고, 이어서 등급 4 대 등급 5로 분류하여 모든 등급을 분류하는 방법이다.

[0125] 마지막으로 결과 검증 단계(S500)는 전립선암을 등급 3, 4, 5, 양성으로 등급을 부여하여 각 등급을 분류한 후 결과를 비교분석 하는 단계로서, 전립선 조직 영상에서 세분화된 세포 핵 영상을 분석하기 위한 데이터 분석법을 활용한다.

[0126] 본 발명에서는 영상 데이터의 특성 추출 및 서포트 벡터 머신 분류를 위해 등급 3, 등급 4, 등급 5, 양성으로 등급을 나누었고 총 400개의 영상들이 사용되었다.

[0127] 이 중 240장은 훈련에, 160장은 테스트에 활용되었다. 각 등급에 대한 영상 수는 100장으로 악성 대 양성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 분류되었다.

[0128] 모든 결과는 표 2 내지 4에 나타나 있으며, 훈련(Training)과 시험(Testing)의 정확도는 서포트 벡터 머신 분류의 혼동 행렬(Confusion matrix)들을 사용하여 나타내었다.

[0129] 즉 표 2는 악성 대 양성에 대한 SVM 혼동 행렬의 테스트 정확도를 보여주는 표이고, 표 3은 등급 3 대(vs) 등급 4와 5에 대한 SVM 혼동 행렬의 테스트 정확도를 보여주는 표이고, 표 4는 등급 4 대 등급 5에 대한 SVM 혼동 행렬의 테스트 정확도를 보여주는 것으로서, 각각 두 부분으로 나뉘어 훈련(Training)과 시험(Testing) 과정에 관한 올바른 분류와 잘못된 분류의 결과에 대한 데이터를 보여주고 있다.

표 2

[0130]

Training: 99.2%				Testing: 88.7%			
Train	Malignant	Benign	Data	Test	Malignant	Benign	Data
Malignant	60	0	60	Malignant	34	6	40
Benign	1	59	60	Benign	3	37	40

표 3

[0131]

Training: 91.7%				Testing: 85.0%			
Train	Grade 3	Grade 4+5	Data	Test	Grade 3	Grade 4+5	Data
Grade 3	55	5	60	Grade 3	36	4	40
Grade 4+5	5	55	60	Grade 4+5	8	32	40

표 4

[0132]

Training: 95.0%				Testing: 92.5%			
Train	Grade 4	Grade 5	Data	Test	Grade 4	Grade 5	Data
Grade 4	54	6	60	Grade 4	36	4	40
Grade 5	0	60	60	Grade 5	2	38	40

[0133] 표 5는 정확도와 세 가지의 이진 분류에 사용된 서포트 벡터 머신의 성능 테이블이다.

표 5

[0134]

Groups	Accuracy (%)	Sensitivity (%)	Specificity (%)	MCC (%)
Malignant vs. Benign	88.7	91.8	86.0	70.2
Grade 3 vs. Grade 4, 5	85.0	81.8	88.8	70.3
Grade 4 vs. Grade 5	92.5	94.7	95.0	85.1

[0135] 표 5에서는 정확도(Accuracy), 민감도(Sensitively), 특이도(Specificity), 매튜의 상관 계수 (Matthews's correlation coefficient) 등 4가지 성능 측정지표를 사용했다. 이러한 지표는 true positive(TP), true negative(TN), false positive(FP), false negative(FN)을 사용하여 계산된다. 계산된 정확도뿐만 아니라 다른 측정치에서 또한 100을 곱하여 정규화하도록 하였다. 표 5에 사용된 4가지 성능 지표 유형을 설명하면 다음의 식 (12) 내지 식 (15)와 같다.

[0136] 상기 정확도(Accuracy)는 정확하게 분류된 표본의 비율이라 할 수 있으며, [식 12]와 같이 나타낼 수 있다.

[0137] [식 12]

[0138]
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

[0139] 상기 민감도(Sensitively)는 정확하게 분류된 표본 중 양의 비율로서, [식 13]과 같이 나타낼 수 있다.

[0140] [식 13]

[0141]
$$Sensitivity = \frac{TP}{TP + FN} \times 100$$

[0142] 상기 특이도(Specificity)는 정확하게 분류된 표본 중 음의 비율로서 [식 14]와 같이 나타낼 수 있다.

[0143] [식 14]

[0144]
$$Specificity = \frac{TN}{TN + FP} \times 100$$

[0145] 상기 매튜의 상관 계수(MCC)는 이진 분류에 탁월하며 목표 값과 예측 값 사이의 상관계수로서, [식 15]와 같이 나타낼 수 있다.

[0146] [식 15]

[0147] $$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{((TP+FN)(TP+FP)(TN+FN)(TN+FP))}} \times 100$$

[0148] 이와 같이 표 5는 세 개의 다른 그룹에 대해 제안된 방법의 분류 결과를 보여주고 있으며, 악성 대 양성의 SVM 이진분류 정확도, 민감도, 특수도, 매튜의 상관 계수는 각각 88.7%, 91.8%, 86.0%, 70.2% 임을 알 수 있다.

[0149] 또한, 등급 3 대 등급 4와 5, 분류 정확도, 민감도, 특수도, 매튜의 상관 계수는 각각 85.0%, 81.8%, 88.8%, 70.3%이고, 등급 4 대 등급5는 분류 정확도, 민감도, 특수도, 매튜의 상관 계수는 각각 92.5%, 94.7%, 95.0%, 85.1% 임을 알 수 있다.

[0150] 또한, 표 6은 평가 결과와 세 가지의 이진 분류에 사용된 다층 퍼셉트론의 성능 테이블이고, 도 8은 본 발명에 따른 세 가지의 이진 분류들의 서포트 벡터 머신 분류 정확도 비교 그래프이다.

표 6

[0151]

Groups	Training Accuracy (%)	Testing Accuracy (%)
Malignant vs. Benign	99.0	81.0
Grade 3 vs. Grade 4, 5	98.0	75%
Grade 4 vs. Grade 5	97.5	76.25

[0152] 유효성 확인을 위해 표 6에 나타낸 것과 같이 기계 학습 소프트웨어 Weka에서 다층 퍼셉트론(MLP) 기법을 이용한 전립선암 등급 분류도 실시하였다.

[0153] 다층 퍼셉트론은 입력층, 은층, 출력층 등 최소 3개의 노드로 구성된 피드 포워드(Feed Forward)인공신경망이다. 여기서 각 노드는 입력 노드를 제외한 뉴런이며 비선형 활성화 기능을 사용한다.

[0154] 다층 퍼셉트론은 서포트 벡터 머신과 같은 감독된 학습 기법을 사용한다. 표 5와 표 6에 나타난 결과를 보면 본 발명에서 제안된 서포트 벡터 머신 이진분류가 다층 퍼셉트론보다 더 높은 정확도를 내었으며, 등급 4 대 등급 5의 경우 정확도가 92.5%로 높다는 것을 알 수 있다.

[0155] 또한, 도 8에서 막대 그래프는 서포트 벡터 머신 분류에 사용되는 세 가지 다른 이진 분할에 대한 비교 결과를 보여주고 있다. 이때 세 그룹의 분류 정확도는 서로 유사하며, 가장 높은 정확도는 92.50%로 등급 4 대 등급 5를 분류 하였을 때이다.

[0156] 표 7은 동시 분류(One-Shot Classification)와 이진 분류(Binary Classification)간 결과 비교를 보여주고 있다. 또한, 도 9는 네 가지 등급 그룹간 서포트 벡터 머신 결과 비교결과에 대한 그래프이다.

표 7

[0157]

One-Shot Classification		Binary Classification	
Groups	Accuracy (%)	Groups	Accuracy (%)
Benign	60.0	Benign	92.5
Grade 3	55.0	Grade 3	90.0
Grade 4	85.0	Grade 4	90.0
Grade 5	50.0	Grade 5	95.0
Total	65.5	Total	92.0

[0158] 표 7과 도 9의 SVM 분류 정확도 비교 결과는 동시 분류와 이진 분류 사이에 차이가 있음을 알 수 있다.

[0159] 다중 등급 또는 동시 분류기를 사용하여 데이터를 분류했을 때 양성, 등급 3, 등급 4, 등급 5의 분류 오류는 각각 60%, 55%, 85%, 50%이다.

[0160] 본 발명의 제안된 이진 분류 접근법을 동일한 분류에 적용하면 정확도는 각각 92.5%, 90.0%, 90.0%, 95.0%이다.

[0161] 즉 두 분류기를 동시에 비교하면, 이진 분류를 사용하여 얻은 결과가 다중 등급 분류나 동시 분류를 사용하여

얻은 결과보다 낫다는 것을 알 수 있다.

[0162] 표 8은 다중 퍼셉트론 분류기(MLP)를 사용한 동시 분류와 본 발명의 이진 분류간의 결과 비교표이다.

[0163] 본 발명의 SVM과 MLP 분류방법의 결과를 비교한 결과, 본 발명의 SVM이 MLP 보다 더 나은 결과를 얻었다고 할 수 있다.

[0164] MLP 동시 분류에서는 전체 데이터 군을 동시에 4개의 등급으로 분류한다. 이 경우 한 등급의 오류는 다른 등급의 성능에 영향을 미쳐 분류 정확도에 부정적인 영향을 미친다. 따라서 모델은 정확한 예측을 할 수 없는 단점이 있다.

[0165] 반면에, 본 발명의 이진 분류에서는 전체 데이터 집합을 3개의 등급으로 분류하고 각 단계의 분류는 독립적으로 이루어진다. 이 경우, 처음 단계의 분류에서의 결과가 그 다음 단계의 분류의 결과에 영향을 주지 않는 장점이 있다.

표 8

[0166]

One-Shot Classification		Binary Classification	
Groups	Accuracy (%)	Groups	Accuracy (%)
Benign	37.5	Benign	87.5
Grade 3	67.5	Grade 3	90.0
Grade 4	45.0	Grade 4	75.0
Grade 5	70.0	Grade 5	77.5
Total	55.5	Total	82.5

[0167] 또한, 표 9는 전립선암 등급 분류를 위한 제안된 방법론과 다른 방법론의 정확도를 비교하는 것이다.

[0168] 본 발명의 방법을 사용한 등급 저 위험군 대 고 위험군에 대해 달성한 분류 정확도는 기술된 다른 방법들 보다 높음을 알 수 있다.

[0169] 암 진단에 있어서, 악성 대 양성으로 분류했을 때, 본 발명은 Nir et al(2018), Doyle et al (2006)보다 좋았지만, Tabesh et al(2017)에 비해서는 높지 않았다.

[0170] 이는 Tabesh et al이 조직 영상에서 추출된 다른 유형의 특성, 즉 색 채널 히스토그램, 프랙탈 치수, 프랙탈 코드, 파동, MAGIC을 사용했기 때문이다. 또한, Tabesh et al은 조직 영상에서 상피 핵 물질의 특성을 계산한 반면, 본 발명에서는 생체검사 전립선 조직 영상에 존재하는 모든 세포 핵의 특징을 계산하였기 때문이다.

표 9

[0171]

Authors	Classification Methods	Classes	Accuracy
Tabesh et al. (2007) [4]	kNN	Malignant vs. Benign	96.7%
		Low vs. High Grade	81.0%
Doyle et al. (2012) [5]	Decision Tree (DT)	Grade 3	77.0%
		Grade 4	76.0%
		Grade 5	95.0%
Nir et al. (2018) [6]	SVM	Malignant vs. Benign	88.5%
		Low vs. High Grade	73.8%
Doyle et al. (2006) [7]	Bayesian	Malignant vs. Benign	88.0%
Rundo et al. [8]	Fuzzy C-Means	Multispectral (Tw1 & Tw2)	90.77%
Naik et al. [11]	SVM	Grade 3 vs. Grade 4	95.19%
		Benign vs. Grade 3	86.35%
		Benign vs. Grade 4	92.90%
Albashish et al. (2017) [12]	SVM	Grade 3 vs. Grade 4	88.9%
		Benign vs. Grade 3	97.9%
		Benign vs. Grade 4	92.4%

Nguyen et al. (2012) [13]	SVM	Benign, Grade 3, and Grade 4 carcinoma	85.6%
Proposed	SVM	Malignant vs. Benign	88.7%
		Low vs. High Grade	85.0%
		Grade 4 vs. Grade 5	92.5%
		Grade 3	90.0%
		Grade 4	90.0%
		Grade 5	95.0%

[0172] 도 10은 본 발명에 따른 서포트 벡터 머신을 이용한 전립선암의 등급 분류에 대한 전체 구성을 보여주는 것으로서, 상술한 바와 같이 본 발명은 감독된 학습 방법을 사용하여 디지털화된 조직병리학 영상에 대한 등급 시스템을 제공하고 있으며, 생체검사 조직 영상에 대한 분할 프로세스로는 k-평균 군집화 알고리즘을 사용하여 수행되었고, 겹쳐있는 핵들은 워터셰드 알고리즘을 사용하여 분리하였다.

[0173] 또한, 전립선암 등급과 진단에 대해 형태학적 특성이 선정되었으며, 가우스와 선형 커널은 전립선 조직병리학 영상의 분류에 사용되었다.

[0174] 이러한 커널을 사용하여 결과가 일부 개선된 것을 관찰하고, 훈련과 시험에 사용되는 모델의 성능을 점차 향상시켰다. 커널의 매개변수는 분류 과정에서 중요한 역할을 하며, 보다 나은 분류 정확도를 위해 상수와 감마 변수의 최선의 조합을 선택하였다.

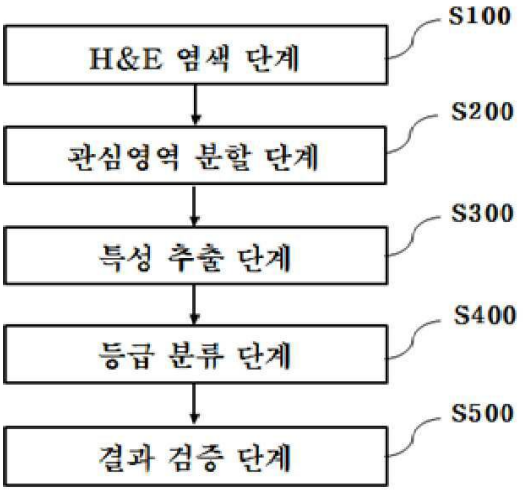
[0175] 또한, 추출된 형태학적 특성을 사용하여 만족스러운 분류 결과를 얻었으며, 이러한 특성은 400배 확대하여 볼 수 있는 하위 이미지에서 추출되었다. 본 발명에서 기술된 정량적 분석은 구현 측면에서 상당히 유연한 것으로서, 전술된 서포트 벡터 머신 이진 분류법은 악성 대 양성, 등급 3 대 등급 4와 5, 등급 4 대 등급 5로 분류하는 데 사용되었음을 알 수 있다.

[0176] 이에 따라 결과 검증을 통하여 다른 문헌들과 비교될 정도로 높은 결과를 도출하였다고 말할 수 있으며, 생체조직 영상에서 추출한 특성에 기초한 정량적 측정치를 산출하였다. 또한, 제안된 방법인 서포트 벡터 머신의 성능을 확인하기 위하여, 다층 퍼셉트론을 사용한 특성 분류를 수행하였고, 이는 동시 분류와 이진 분류 결과를 비교해서 두 가지 분류의 차이점을 보여주고 있음을 알 수 있다.

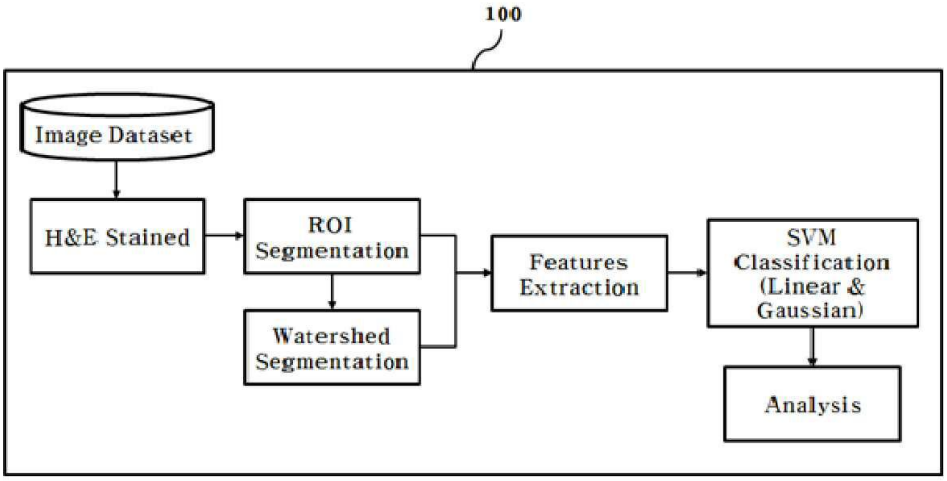
[0177] 전술한 바와 같이 본 발명의 상세한 설명에서는 바람직한 실시예들에 관하여 설명하였지만, 본 발명의 기술분야에서 통상의 지식을 가진 사람이라면 하기의 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음은 이해할 수 있을 것이다.

도면

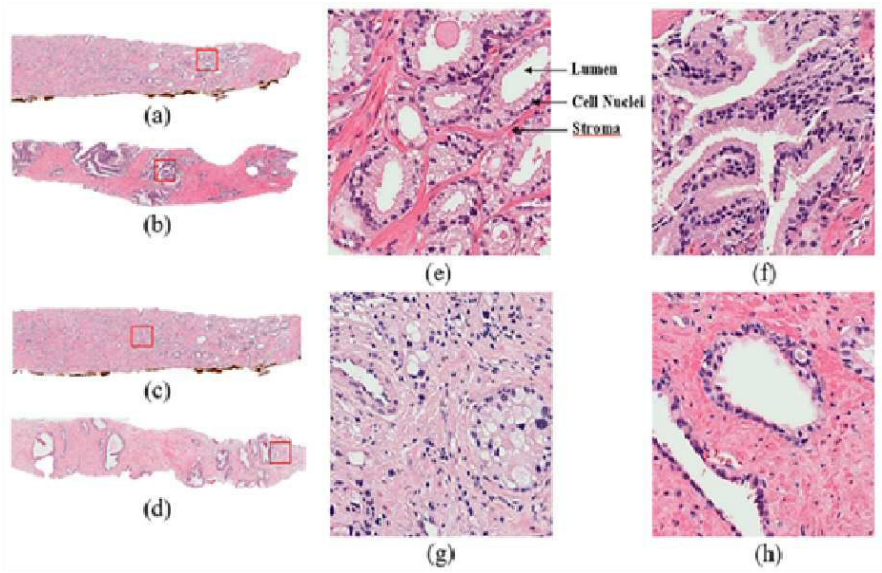
도면1


S100
H&E 염색 단계
S200
관심영역 분할 단계
S300
특성 추출 단계
S400
등급 분류 단계
S500
결과 검증 단계

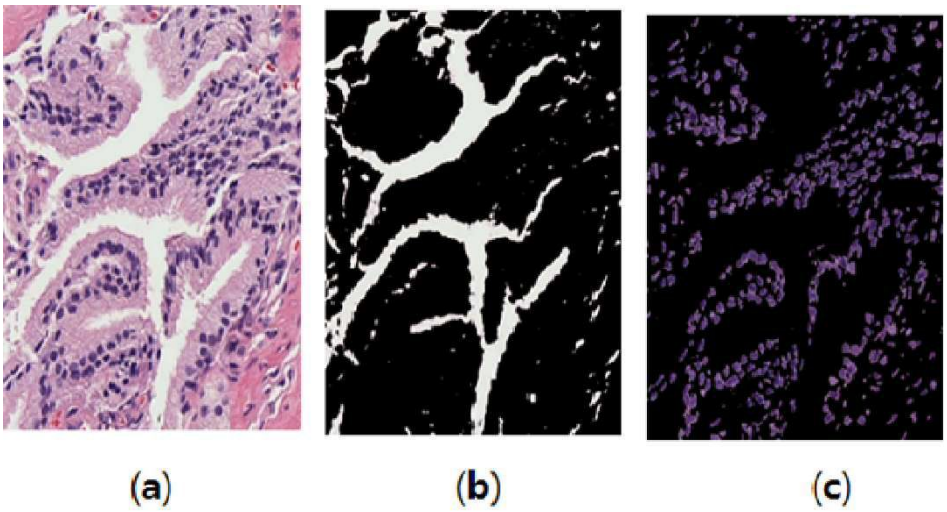
도면2


100
Image Dataset
H&E Stained
ROI Segmentation
Watershed Segmentation
Features Extraction
SVM Classification (Linear & Gaussian)
Analysis

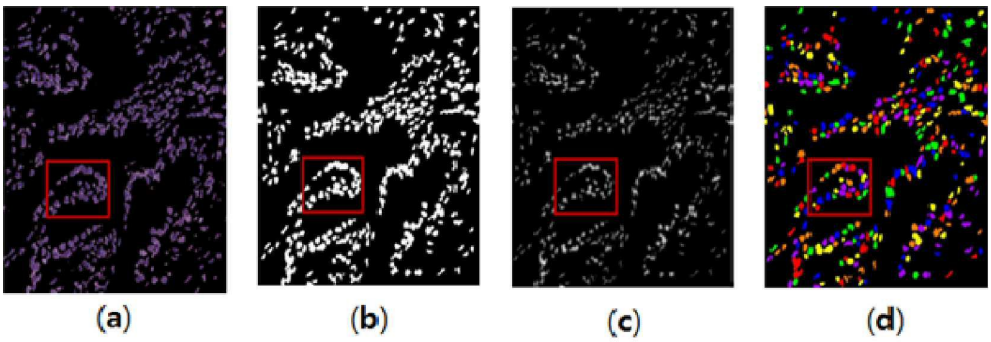
도면3



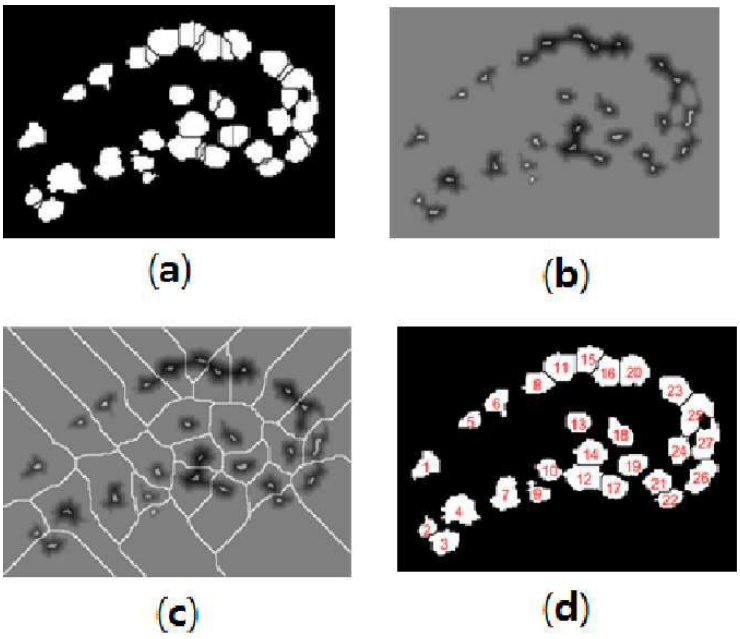
도면4



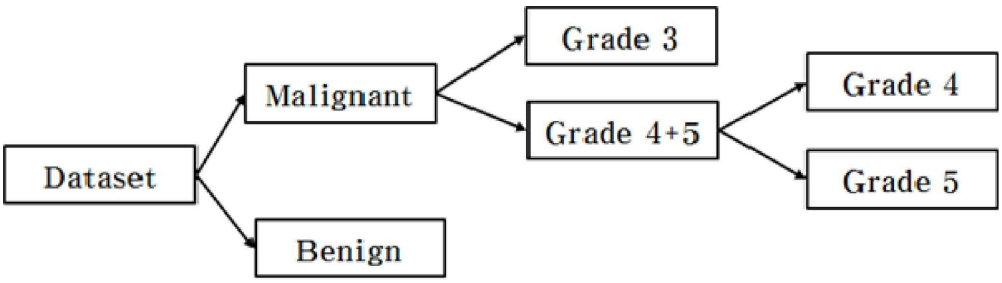
도면5



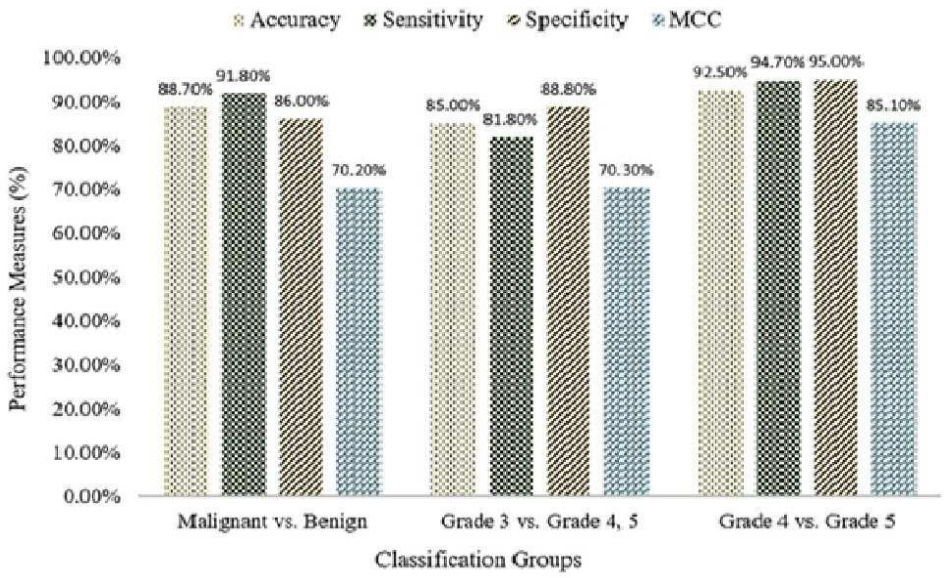
도면6


(a)
(b)
(c)
(d)

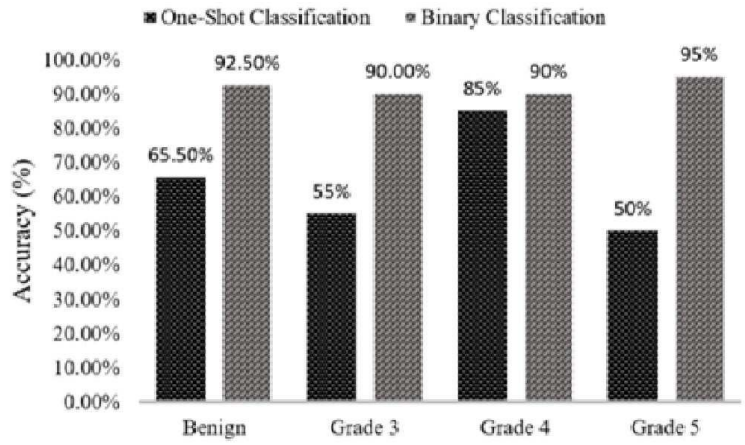
도면7


Dataset
Malignant
Benign
Grade 3
Grade 4+5
Grade 4
Grade 5

도면8


Accuracy
Sensitivity
Specificity
MCC
Performance Measures (%)
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
88.70%
91.80%
86.00%
70.20%
85.00%
81.80%
88.80%
70.30%
92.50%
94.70%
95.00%
85.10%
Malignant vs. Benign
Grade 3 vs. Grade 4, 5
Grade 4 vs. Grade 5
Classification Groups

도면9


One-Shot Classification
Binary Classification
Accuracy (%)
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
65.50%
92.50%
55%
90.00%
85%
90%
50%
95%
Benign
Grade 3
Grade 4
Grade 5

도면10

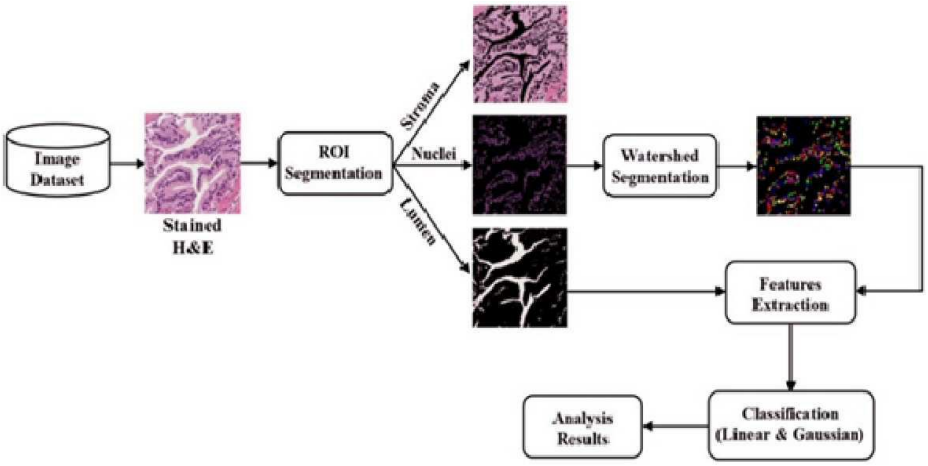


Image Dataset
Stained H&E
ROI Segmentation
Stroma
Nuclei
Lumen
Watershed Segmentation
Features Extraction
Classification (Linear & Gaussian)
Analysis Results