

# 결정 트리를 이용한 기저 세포암 특징 추출

\*박아론, 백성준, 원용관, 김동국  
전남대학교 전자컴퓨터공학부

e-mail : dodenet@daum.net, tozero@chonnam.ac.kr, ykwon@chonnam.ac.kr,  
dkim@chonnam.ac.kr

## Feature Extraction of Basal Cell Carcinoma with Decision Tree

\*Aaron Park, Seong-Joon Baek, Yonggwan Won, Dong Kook Kim  
The School of Electronics and Computer Engineering  
Chonnam National University

### Abstract

In this study, we examined all peaks of confocal Raman spectra as peaks are the most important features for discrimination between basal cell carcinoma (BCC) and normal tissue (NOR). 14 peaks were extracted from these peaks using decision tree. For dimension reduction, frequently selected 4 peaks were chosen. They are located at 1014, 1095, 1439, 1523cm<sup>-1</sup>. These peaks were used as an input feature of the multilayer perceptron networks (MLP). According to the experimental results, MLP gave classification error rate of about 6.5%.

### I. 서론

일반적으로 피부암은 기저 세포암과 편평상피 세포암 (Squamous Cell Carcinoma)으로 분류한다[1]. BCC는 보다 자주 나타나는 피부 종양이지만 주위의 암이 아닌 조직과의 구별하기가 쉽지 않다. BCC는 일반적으로 조직 샘플을 병리학적으로 검사하여 검출하지만 이 방법은 처리가 복잡하고 각 병리학자의 경험에 의한 주관적인 판단에 따르게 되고 때로는 정확한 생체 검사를 위해 피부 조직을 과도하게 필요로 한다[2].

본 논문에서는 공초점 라만 스펙트럼의 피크 (peak)가 BCC 검출에 가장 중요한 특징이므로 각 데이터의

\* 본 연구는 산업자원부 지방기술혁신사업(RTI04-03-03) 지원으로 수행되었음.

모든 피크를 조사하였다. 조사한 데이터의 피크는 결정 트리 (Decision Tree)를 사용하여 보다 변별력있는 상위 N개 피크로 특징벡터를 구성하였다. 최종적으로 이 피크들을 MLP를 이용하여 분류하였다.

### II. 실험방법

공초점 라만 분광기를 이용한 조직 샘플의 스펙트럼은 피부 표면으로부터 수직인 방향 30-40μm 간격으로 서로 다른 지점에서 측정하였다[2]. 이와 같은 방식으로 10명의 환자로부터 본 연구에서 사용한 216개의 라만 스펙트럼을 얻었다.

측정 후 스펙트럼은 원래의 구간 443-1757cm<sup>-1</sup>보다 변별력 있는 801-1757cm<sup>-1</sup> 구간을 Clipping Window를 사용하여 얻은 다음 [0, 1] 범위로 정규화하였다. 스펙트럼에서 피크를 검색하기 전에 불필요한 피크 검색을 줄이기 위해 모든 구간에 대하여 다음의 식을 이용하여 스무딩하였다.

$$\hat{x} = \frac{1}{25} \sum_{n=-12}^{12} x(n)$$

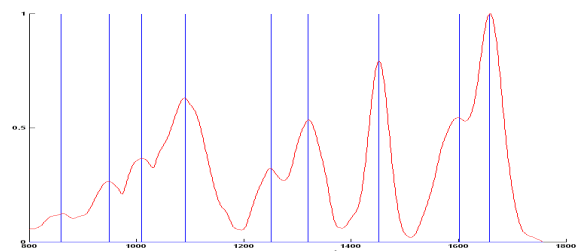


그림 1. 공초점 라만 스펙트럼에서 검색된 피크.

그림 1은 스무딩한 후 검색된 피크를 나타낸다. 그림에서 알 수 있듯이 전 구간이 매끄러워졌으며 중요한 모든 피크들이 검색되었다.

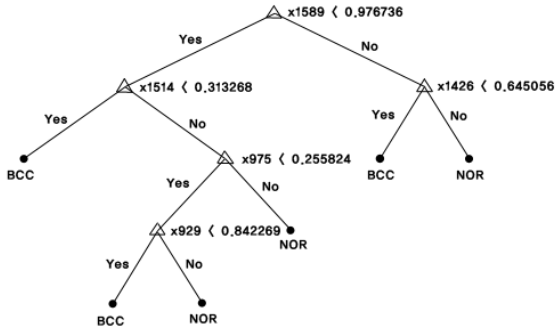


그림 2. 훈련용 패턴을 사용하여 결정 트리에서 추출된 중요 피크의 예.

모두 216개의 스펙트럼에서 386개의 피크가 검색되었고 그 중 변별력 있는 피크를 확인하기 위해 결정 트리를 이용하였다. 그림 2는 결정 트리를 이용하여 14개 피크들을 추출한 결과를 나타낸다. 그림의 x1589 < 0.976736에서 x1589는 피크이고 0.976736은 그 피크에서의 경계값이다. 표 1은 결정 트리에서 추출된 결과를 보여준다. 표에서 보듯이 빈도수가 높은 피크인 1014, 1095, 1439, 1523cm<sup>-1</sup>를 추출하였다.

표 1. 결정 트리에서 추출된 피크와 빈도.

피크	846	929	975	1014	1095	1246	1262
빈도	1	1	1	6	7	1	1
피크	1267	1426	1439	1514	1523	1553	1585
빈도	1	1	8	1	8	1	1

분류 실험에서는 이렇게 추출된 4개의 피크 성분만을 MLP 입력으로 사용하였다. MLP는 4개의 은닉 유닛과 하나의 출력 유닛으로 구성하고 NOR 클래스에 대해 -1, BCC 클래스에 대해 +1이 출력되도록 오류 역전파(Error Back Propagation) 알고리즘을 이용하여 훈련시켰다. MLP의 유닛 연산은  $o_k = f(net_k)$ , 입력 유닛은  $net_k$ , 활성화 함수를  $f()$ 라고 하면 다음 식을 따른다[3].

$$net_k = \sum_i \omega_{ik} o_i + bias_k,$$

$$f(net_k) = \frac{2}{1 + e^{-2net_k}} - 1$$

### III. 결론 및 토론

표 2에는 4개의 중요한 피크를 입력으로 사용한 MLP 분류 결과를 요약하였다. 그 결과에 따르면 sensitivity가 92.1%이고 specificity가 95.0%임을 볼 수 있다. MLP는 초기값에 따라 그 성능이 달라지므로 본 연구에서는 20회 실험의 결과를 평균하였다.

표 2. 4개의 중요한 피크만을 사용한 분류 결과 (\*는 전문 병리학자의 결정임).

	MLP	
	BCC	NOR
BCC*	92.1	7.9
NOR*	5.0	95.0

이 실험의 결과로 공초점 라만 스펙트럼의 피크 중 1014, 1095, 1439, 1523cm<sup>-1</sup>에 형성된 피크가 변별력이 있음을 확인하였다. 비록 이 실험의 결과가 이전의 연구보다 분류율이 향상되지 않았더라도 Principal Component Analysis를 사용하여 추출한 전역 특징이 아닌 지역 특징만으로 6.5% 미만의 분류 오류율을 보인다는 것은 피크 검색과 결정 트리를 이용한 피크 추출의 의미를 잘 보여준다고 할 수 있다.

본 연구에서는 피크점과 크기만으로 실험을 하였지만 향후 연구에서는 피크폭과 같은 다른 요소를 고려하며 실험할 예정이다. 또한 이 실험을 통하여 얻어진 결과를 이용하고 몇 가지 특징 모델링을 도입하여 분류 오류율을 더욱 감소시키고자 한다.

### 참고문헌

- [1] A. Jijssen, T. C. B. Schut, F. Heule, P. J. Caspers, D. P. Hayes, M. H. Neumann and G. J. Puppels, "Discriminating Basal Cell Carcinoma from its Surrounding Tissue by Raman Spectroscopy," *Journal of Investigative Dermatology*, vol. 119, pp. 64-69, July 2002.
- [2] J. Choi, J. Choo, H. Chung, D.-G. Gweon, J. Park, H. J. Kim, S. Park and C.-H. Oh, "Direct observation of spectral differences between normal and basal cell carcinoma(BCC) tissues using confocal Raman microscopy," *Biopolymers*, vol. 77, pp. 264-272, April 2005.
- [3] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Jone Wiley & Son, Inc. 2001,