

국내 갑상선암 연구 주제 동향 분석

양지연*, 신승혁*, 허성민^o, 이태경*

^o*금오공과대학교 응용수학과

e-mail: {jyang, shinbaad, cjsm03}@kumoh.ac.kr^o, leetk97@kumoh.ac.kr*

A Study on Research Topics for Thyroid Cancer in Korea

Ji-Yeon Yang*, Seung-Hyeok Shin*, Seong-Min Heo^o, Tae-Gyeong Lee*

^o*Dept. of Applied Mathematics, Kumoh National Institute of Technology.

● 요약 ●

본 논문에서는 국내 갑상선암의 연구 동향을 파악하기 위해 텍스트 중심의 접근법을 제안한다. 국내 갑상선암은 2000년대에 들어서며 발생이 급증하여 과잉진단의 논란을 불러일으켰으나, 다양한 분야의 자정 노력으로 수술 환자가 크게 줄었다. 본 연구에서는 텍스트 마이닝 기술을 사용하여 디비피아에 등록되어 있는 갑상선암 관련 논문의 키워드와 초록을 수집하여 분석하였다. 1980년대는 대부분의 사례보고가 있었고 1990년대에 들어서면서 검진을 통한 조기 진단의 내용이 자주 나타났다. 2000년대에는 여러 장비들을 활용한 검사방법과 미세한 암의 발견에 대한 논의가 증가하였음을 확인 할 수 있었다. 2010년대에 들어서는 환자의 삶의 질에 대한 연구가 많이 이루어졌다. 지난 수십 년 동안 갑상선 암 연구 주제에 대해 뚜렷한 변화가 나타났으며, 향후 연구의 기초자료로 활용될 수 있으리라 기대된다.

키워드: 갑상선암(thyroid cancer), 텍스트 마이닝(text mining).

계층적 군집분석(hierarchical clustering), 소셜네트워크(social network).

I. Introduction

국내 갑상선암 발생은 1990년대 동안 서서히 증가하다가 2000년대 이후 급격히 증가하였다. 2011년 갑상선암 진단 비율은 1993년에 비해 15배나 증가하였고, 2009년부터 2014년까지 국내 암 발생순위 1위를 기록하였다[1].

연구진들은 지난 20-30년간 갑상선암 발생률이 급증한 주요인으로 의료기술 발전으로 인한 과다진단(overdiagnosis)을 꼽았으며 한국이 대표적 사례라고 밝히고 있다. 반면 발생률이 급증한 것에 비해 갑상선암으로 인한 사망률은 크게 변화가 없다고 보고된다[2]. 2016년 11월, 대한갑상선학회에서는 갑상선암 진료 권고안 개정안을 발표하였는데, 과잉진료를 피하기 위해 적극적으로 검사나 치료를 하지 말라는 것이다[3]. 본 연구는 지난 수십 년간 사람들의 관심이 증가한 갑상선암에 초점을 맞추어, 텍스트 마이닝 기법을 사용하여 해당 분야의 연구동향을 파악하고 있다. 본 연구의 구성은 다음과 같다. 제2장에서는 자료수집과 분석방법에 대해 기술하고, 제3장에서는 자료분석 결과를 제시한다. 마지막 4장에서는 본 연구의 결론, 향후 필요한 연구과제에 대해 논의하고 있다.

II. Methods

본 연구는 디비피아에서 "갑상선암"을 키워드로 검색하여 수집된 논문들을 전처리 과정을 통해 초록에서 319개, 키워드에서 181개의 단어들을 얻어 사용되었다.

단어 간 관계를 살펴보기 위해 소셜 네트워크 분석을 실시하였다. 수집된 키워드를 이용하여 연관성을 보이는 중요한 단어들의 관계를 파악하였다. 각 키워드가 네트워크 내의 다른 키워드들과 얼마나 근접한지, 영향력을 가진 중심 키워드가 무엇인지 파악할 수 있다. 또한 계층적 군집분석(hierarchical clustering)을 이용하여 단어들이 어떻게 군집을 이루는지 검토하였다. 분석결과를 덴드로그램으로 시각화하여 연관이 높은 키워드를 파악하였다. 그 후, 이를 바탕으로 각 군집의 특성을 파악하고 연구주제를 추론하였다.

III. Results

소셜 네트워크와 계층적 군집분석을 이용하여 대상 논문 초록에 포함된 키워드 간 유사성을 측정하고 연구주제 네트워크를 구축하며, 상호 연결성이 높은 단어들을 군집화하고자 한다. 논문의 키워드를 대상으로 소셜 네트워크 분석결과 그림 1에 제시되었다.

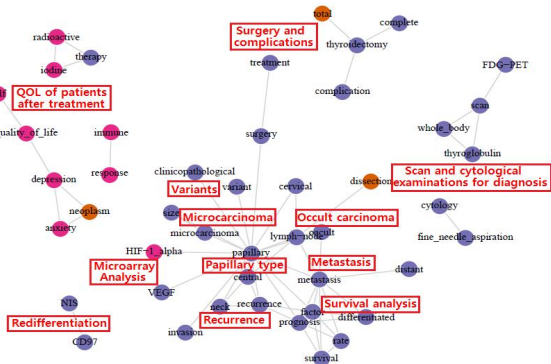


Fig. 1. simple social networks of keywords with cluster annotations

그림 1의 네트워크는 차수가 0인 정점, 기중치가 1 이하인 엣지들을 제거한 46개의 용어를 대상으로 한 간단한 네트워크이다. 각각의 용어에 대해 시대별 등장 빈도를 확인한 후, 가장 많이 나타난 시대를 반영하고자 정점의 색을 달리 했다. 2000년대에는 많은 연구들이 다양한 주제로 진행되던 반면 2010년대에는 환자의 복지, 교육 프로그램, 간호 중재에 대한 연구가 활발하였을 확인하였다. 유두상 암 (papillary), 환자의 삶의 질 (QOL of patients), 소아 갑상선암 (pediatric thyroid cancer), 재분화 (redifferentiation) 등의 주제어 그룹이 나타났다. 특히 환자의 삶의 질 관련 주제가 비교적 큰 독립적인 군집을 형성하고 있다. 이러한 패턴은 그림 2의 덴드로그램에서도 확인할 수 있다. 덴드로그램은 단순 네트워크에 포함된 46개의 용어를 대상으로 코사인 유사도와 위드 링크법을 사용한 결과이다. 이를 보면 환자의 삶의 질이 별도의 군집을 이루고 있으며, 나머지 주제들은 하나의 큰 군집에 속하는데 이 중에는 수술 및 합병증, 재발, 전이, 스캔과 세포 검사, 미소암, 재분화, 생존분석 등 관련 주제들이 포함된다. 갑상선암 연구주제는 2010년이전에는 임상병리학적 특성에만 주된 초점을 두었다면 이후에는 환자의 복지까지 고려하고 있다.

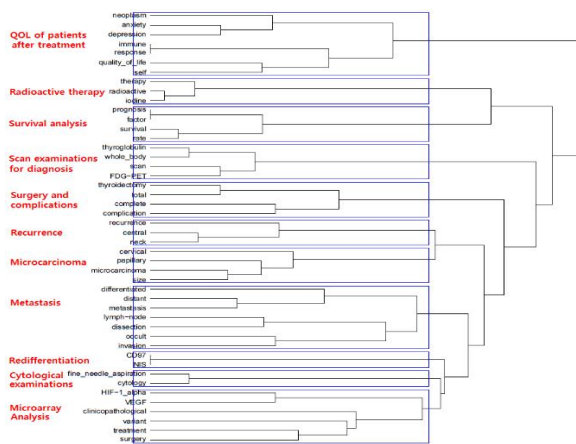


Fig. 2. Dendrogram from hierarchical clustering of keywords with cluster annotations

IV. Conclusions

2010년대 급증한 국내 갑상선암의 발생률의 원인에 대해 기존

연구들은 초음파 검사장비 등 정밀한 장비들의 도입으로 미세한 결절에서도 암세포를 발견할 수 있게 되었고 개인 건강검진이 활성화 되면서 치명적이지 않은 갑상선 이상을 발견하는데 영향을 주었다고 분석하고 있다[2]. 이러한 갑상선암에 대한 임상적 시각, 치료의 변화는 연구 주제에 직접적인 영향을 준 것으로 나타났다. 증례·사례 보고 위주였던 내용은 1990년대에 들어서 건강검진을 통한 갑상선암의 조기 진단이라는 주제가 자주 검토되었다. 2000년대는 많은 연구들이 다양한 주제로 진행되었다. 주목할만한 점은 여러 장비들을 활용한 검사방법과 미세한 암의 발견에 대한 논의가 증가하였다. 2010년대에는 수술과 치료 후 환자의 복지에 관련된 주제가 많이 다루어졌다. 본 연구에서는 관련 문헌 정보를 이용해 주요 연구 분야와 시계열 트렌드를 밝히고 연구 분야 간의 상관관계를 시각화하였다. 이러한 시도는 향후 다른 암, 질병에 확장하여 분석함으로써 연구 활성화 방향의 기초자료로 활용될 수 있을 것이다.

REFERENCES

- [1] K. Jung, Y. Won, H. Kong, and E. Lee, "Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2015," *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, Vol. 50, No. 2, pp. 303-316, Mar. 2018.
- [2] H. Ahn, H. Kim, and H. Welch, "Korea's thyroid cancer epidemic-screening & overdiagnosis," *The New England Journal of Medicine*, Vol. 371, No. 19, pp. 1765-1767, Sep. 2014.
- [3] Korean Thyroid Association. "Revised Korean thyroid association management guidelines for patients with thyroid nodules and thyroid cancer," *Journal of the Korean Society of Radiology*, Vol. 64, No. 4, pp. 389-416, Dec. 2010.