

## 췌장암 조기진단을 위한 조건부 확률 기반 지능형 진단 방식

장익규 · 정준호 · 고재호<sup>1</sup> · 문현석<sup>2</sup> · 조영호<sup>3</sup>

구미전자정보기술원 전자의료기술연구본부, <sup>1</sup>메디엣주식회사, <sup>2</sup>금오공과대학교 응용화학과, <sup>3</sup>국립암센터의공학과

### Intelligent Diagnosing Method Based on the Conditional Probability for the Pancreatic Cancer Early Detection

IK GYU JANG, JOONHO JUNG, JAE HO KO<sup>1</sup>, HYUN SEOK MOON<sup>2</sup> and YUNG HO JO<sup>3</sup>

Gumi Electronics & Information Technology Research Institute  
Electronic Medical Technology Research Division Convergence Medical Devices Research Center  
<sup>1</sup>MeDioT CO., LTD

<sup>2</sup>Department of Applied Chemistry, Kumoh National Institute of Technology

<sup>3</sup>National Cancer Center Department of Biomedical Engineering

(Manuscript received 6 July 2017 ; revised 6 September 2017 ; accepted 11 October 2017)

**Abstract:** Early diagnosis of pancreatic cancer had been considered one of the important barrier for successful therapy since the five year survival rate after treatment of pancreatic cancer was critically low. Nonetheless, patients often miss the golden time of treatment because they rarely visit the hospital until their symptoms are severe. To overcome these problems, a lot of information about the patient's symptoms should be applied as biomarkers for early diagnosis. For this reason, a biomarker for early detection of pancreatic cancer (CA19-9) has been developed as a diagnostic kit. However, since the diagnosis is not accurate enough, pancreatic symptoms (abdominal pain, jaundice, anorexia, diabetes, etc.) and biomarkers (CA19-9) should be considered together. We develop an intelligent diagnostic system that considers CA19-9 and the incidence of pancreatic cancer for pancreatic symptoms that was determined by studying a large number of patient information. It shows a higher accuracy than one using CA19-9 alone. It may increase the survival rate of pancreatic cancer because it can diagnose pancreatic cancer early.

**Key words:** Pancreatic cancer, Intelligent diagnosis, Bayesian network

### I. Background

췌장암(Pancreatic cancer-PC)은 환자들이 겪는 암중에서 치사율이 가장 높은 질병에 속한다. 췌장암은 15~20% 정도의 절제를 통한 수술이 가능한 것으로 알려져 있으나 생존율이 일년 내에 25% 정도 이고 5년내 5%에 불과할 정도로 매우 공격적인 암 질병이다[1]. 미국에서는 53,070명

의 새로운 췌장암 환자가 2016년에 발생하였고 그 중에 41,780명이 사망한 것으로 기록 되었다[2]. 췌장암은 절제술로 충분히 치료가 가능하지만 대부분의 환자들은 15~20%의 절제수준을 넘어서 상태에서 방문하는 경우가 많아 생존율이 낮다. 그러므로 증상이 경미하게 나타나는 시기에도 조기 진단이 가능한 시스템이 필요하다[3]. 최근에 바이오 마커가 빠르게 환자 개인별 적용이 가능해지면서 어떤 암들은 암표지자를 이용한 발견이 가능해 졌지만 아직 대부분의 경우 환자들의 인터뷰에서 나타나는 증상이 결정적 단서가 되는 경우가 많다[3]. 췌장암의 경우에도 2002년도에 CA19-9(carbohydrate antigen 19-9) 라는 암표지자가 FDA를 통과하여 의료시장에 진입하였지만 일반인이 아닌 췌장암 환자에 대해서만 79~81%의 민감도를 가지고 있고 췌장암,

Corresponding Author : YUNG HO JO  
National Cancer Center Department of Biomedical Engineering, 323 Ilsan-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do 410-769, Korea  
이 연구는 경북창조경제선도기술개발사업의 과제((Project number: SF315010A))의 지원을 받아 수행하였음.

황달과 같은 증상에도 긍정오류(False-Positive)가 발생하는 등의 문제가 관찰되어 조기와 말기에 췌장암을 판별하는 기준으로는 추천되지 않는다[4]. 국내에서도 CA19-9 검사는 임상에서 유용하게 사용되고 있지만 췌장암이 의심되는 일반인의 경우 CA19-9 혈액 검사만으로 췌장암 여부를 확인하기 위한 선별 검사로는 사용할 수 없고 검진 항목에 CA19-9를 추가하여 참고사항으로 도움이 되는 정도이다.

췌장암 조기진단의 핵심은 우선적으로 환자 스스로 ‘혹시 이런 증상이 있으므로 췌장암이 아닐까’ 하는 의심을 갖는 것이다. 암표지자 검사와 영상진단까지 가는 과정에서 이미 수술 가능시기를 지나친 경우가 많으므로 작은 증상에도 의심을 조기에 갖고 이를 확인해 주는 작업이 가장 중요하다. 이러한 수요로 인해 컴퓨터 알고리즘들이 의학적 판단에 보조적으로 이용되는 경우가 많다[5-7]. 이 논문에서는 췌장암에 대한 증상을 통해 지능화된 진단이 가능하도록 베이지안 네트워크를 활용하였다. 이와 유사하게 베이지안 네트워크를 췌장암에 적용한 경우가 있지만[8], 췌장암에 특화되어 암표지자와 함께 증상에 대한 통계를 이용하여 일반인을 대상으로 조기에 발병여부를 지능적으로 판단하는데 직접적으로 활용한 예는 드물다.

대한 증상을 조사하여 도출한 내용이며[9-10] 추가적으로 국립암정보센터의 “2014년 성별 10대암 발생률”(우리나라 2000년 주민등록연앙인구를 표준인구로 사용)를 사용하였다.

표 1은 학습이라는 과정으로 응용되어 췌장암 환자에 대한 통계와 일반인에 대한 통계 레포트를 생성할 수 있고 학습은 췌장암 환자의 의료정보를 보유하고 있는 병원들의 데이터를 수집하여 분류하고 각 통계에 대해서 더 이상 정보가 없을 때까지 누적하여 생성하는 방식이 적용될 수 있다.

위 통계를 바탕으로, 이 논문에서 일반인이 췌장암에 걸릴 확률을  $P(P)$ 라하고 2) 췌장암에 걸렸을 때 특정 증상이 나타날 확률을  $P(S|P)$ 라고 할 때 3) 베이지안 조건부 확률을 통해 반대로 증상이 나타날 때 췌장암일 가능성  $P(P|S)$ 을 구할 수 있다( $P$ : Pancreatic cancer,  $S$ : Symptoms).

사건  $P(P)$ 가 참일 때  $P(S)$ 가 일어날 조건부 확률로 쓰면 다음과 같이 표현할 수 있다.

$$P(S \cap P) = P(S|P)P(P) \quad \text{식 (1)}$$

반대로  $P(S)$ 가 주어졌을 때  $P(P)$ 가 일어날 조건부 확률은  $P(P \cap S) = P(P|S)P(S)$  식 (2)

그러면  $P(S \cap P) = P(P \cap S)$  이므로 다음과 같이 표현할 수 있다.

$$P(P|S)P(S) = P(S|P)P(P) \quad \text{식 (3)}$$

그러므로  $P(P|S) = \frac{P(S|P)P(P)}{P(S)}$  로 표현할 수 있다. 식 (4)

위의 수식에 따라,  $P(P)$ 과  $P(S|P)$ 를 통계로 알고 있다면 반대로  $P(P|S)$ 를 구할 수 있다.

## II. Materials & Methods

이 논문에서는 췌장암 환자들의 증상에 대한 통계를 이용하여 일반인을 추론(진단)하는데 활용 하는 방법을 제안한다. 췌장암 환자의 경우 증상이 나타나는 확률을 다음(표 1)과 같이 조건부 확률표와 수식으로 표현 할 수 있다. 사용된 통계는 동양인 236명(Peking Union Medical College; 중국 베이징)의 2009년부터 2014년까지의 췌장암 환자에

표 1. 췌장암 환자정보를 이용한 조건부 확률표.

Table 1. Conditional probability table based on the information of pancreatic cancer patients.

ITEM	TRUE	FALSE
Probability of pancreatic cancer	11/Hundred thousand(0.011%) [National Cancer Information Center]	99.98%
Abdominal pain	41.7%[4]	58.3%
Nausea and vomiting	9.1%[4]	90.9%
Weight loss	22%[4]	78%
Jaundice	6.8%[4]	93.2%
Newly onset diabetes	28.0%[National Cancer Information Center]	72%
Pancreatic cancer patients(PC = TRUE)	Male	80.3%[4]
	Female	26%[4]
CA19-9 numerical positive	84%[Seoul National University Dept. Hemato Oncology]	16%
Smoking rate	33%[National Cancer Information Center]	67%
Family history of PC	10%[National Cancer Information Center]	90%

확률을 구할 때 하나의 노드 값만이 정해진 경우나 조건부 확률 등을 구할 때는 노드의 조건부 확률을 이용하여 직접 계산으로 구할 수 있지만 여기서는 증상이 여러 개 이므로 노드간 연결을 설정한다. 이 연결을 베이지안 네트워크라 한다.

베이지안 네트워크에서 각 노드의 확률은 조건부 확률표

(CPT, conditional probability table)를 사용해 표현한다. 췌장암 여부를 판단하는 노드의 확률을 부여한다. 췌장암은 통계적으로 10만명중에 11명이 걸리는 질병이므로 다음과 같은 CPT로 표현한다. Abdominal pain(불규칙한 복통은 췌장암 환자일 경우에 증상이 나타날 확률은 41.7%, 나타나지 않을 확률은 58.3%, 췌장암환자가 아닐 경우 증상이

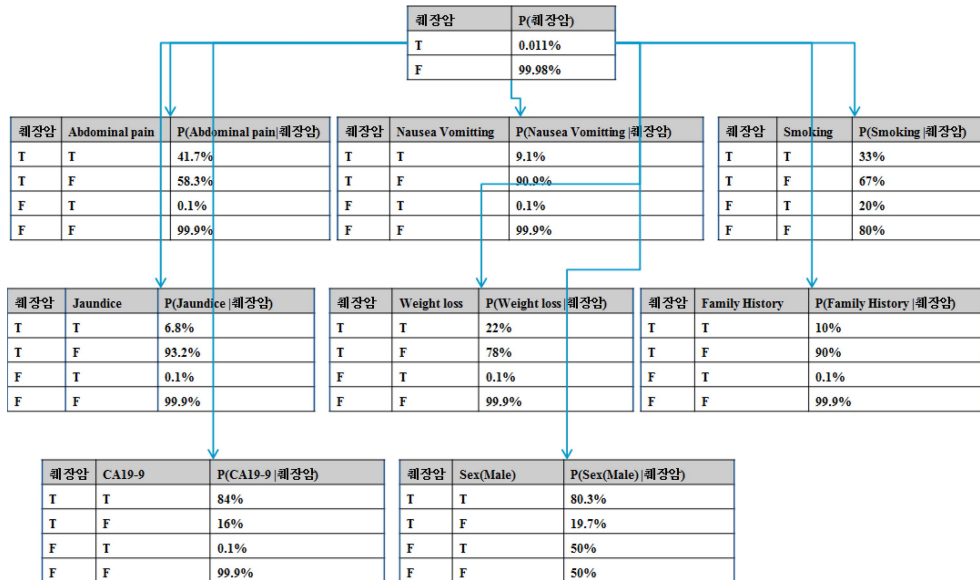


그림 1. 베이지안 네트워크를 활용한 췌장암 진단 노드 생성.  
Fig. 1. Pancreatic cancer diagnosing node using Bayesian network.

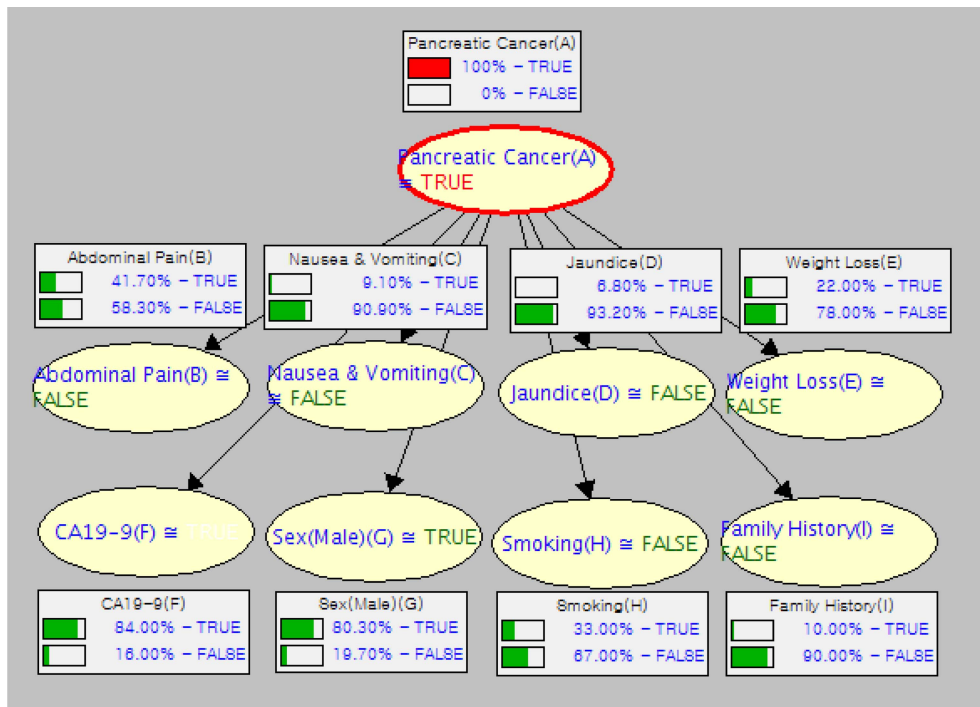


그림 2. 췌장암 환자일 경우 증상이 나타날 확률.  
Fig. 2. Probability of symptom occur rate on pancreatic cancer patients.

표 2. 베이저안 네트워크를 이용한 췌장암 발병 확률 결과.

Table 2. Probability of pancreatic cancer occur rate based on Bayesian network.

symptom	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Abdominal Pain	⊙								⊙				⊙			
Nausea & Vomiting		⊙							⊙	⊙				⊙		
Jaundice			⊙							⊙	⊙					⊙
Weight Loss				⊙							⊙	⊙				
CA19-9					⊙							⊙	⊙	⊙	⊙	⊙
Sex(Male)						⊙										
Smoking							⊙									
Family history									⊙							⊙
Prob. PC(%)	4.39	0.99	0.74	2.36	8.46	0.02	0.02	1.09	80.67	40.5	62.2	95.31	97.47	89.37	86.27	90.24

⊙ = TRUE

나타날 확률은 0.1%, 나타나지 않을 확률은 99.9%로 설정한다. Nausea & Vomiting(식욕부진, 구토) 증상은 췌장암 환자일 경우 증상이 나타날 확률은 9.1%, 나타나지 않을 확률은 90.9%로 설정한다. 이런방식으로 확률표(표 1)을 이용하여 Jaundice(황달), Weight loss(급격한 체중감소), CA19-9수치, 성별, 흡연여부, 가족력이 나타내는 노드의 조건부 확률을 적용한다.

노드가 많아질수록 곱셈의 양이 많아 지므로 효율적인 연구와 향후 확장성을 위해 이 논문에서는 공개 솔루션(SamIam; UCLA; USA)을 사용하여 베이저안 네트워크를 구성하고 각 노드의 확률을 구하는 방법을 사용한다. 노드의 진단과 인과관계를 고려하여 노드를 그림 1과 같이 구성하고 프로그램을 실행한다.

### III. Result

노드를 프로그램에서 구성한 후 노드 생성의 값이 정상 출력되는지 확인하기 위해 췌장암 환자일 경우(Pancreatic cancer = True) 각각의 증상은 어떠한지 살펴본다. 조건부 확률표(표 1)에 나타난 바와 같이 그림 2의 프로그램 결과에서도 Abdominal pain이 41.7%를 나타내고 Nausea와 Vomiting이 9.1%, Jaundice가 6.8%, Weight loss가 22%, CA19-9의 양성 반응 수치가 84%, 성별은 남성이 압도적으로 많은 80.3%, 흡연자는 33%, 가족력은 10%를 가지는 것을 확인할 수 있으며 조건부 확률 표와 일치한다. 이후, 베이저안 네트워크를 이용하여 일반인에게 각각의 증상이 발생할 때 진단 결과는 어떻게 나타나는지 살펴 본다.

표 2에서 보는 바와 같이 다른 증상 없이 Abdominal pain이 True일때 췌장암일 가능성은 4.39%, Nausea & Vomiting가 True일 때 0.99%, Jaundice가 True일 때 0.74%, Weight loss가 True일 때 2.36, 암표지자 CA19-

9가 True일 때 8.46%, 성별이 남성일 때 0.02%, 흡연자일 때 0.02%, 가족력이 있을 때 1.09%로 나타났다. 2가지 증상이 함께 나타난 경우 Abdominal Pain과 Nausea & Vomiting이 True인 경우 80.67%, Nausea & Vomiting과 Jaundice가 True인 경우 40.5%, Jaundice와 Weight loss가 True인 경우 62.2, Weight loss와 CA19-9가 True인 경우 95.31%, CA19-9와 Abdominal pain이 True인 경우 97.47%, CA19-9와 Nausea & Vomiting이 True인 경우 89.37%, CA19-9와 Jaundice가 True인 경우 86.27%, CA19-9와 Family history가 True인 경우 90.24%를 보였다.

### IV. Conclusion

이 논문에서는 췌장암 환자를 통하여 얻어낸 증상의 통계를 바탕으로 반대로 일반인에게 증상이 발생하였을 때 췌장암 여부를 판단할 수 있는 지능형 진단 방식을 베이저안 네트워크를 이용하여 구현하였다. 베이저안 네트워크는 췌장암 발병 여부는 불확실성을 계산해야 하는 대상(A)이며 증상에 대한 통계는 관측하여 값을 알아낼 수 있는 대상(B)으로 생각한다면, A의 확률은 B가 관측된 후 P(A)에서 조건부 확률 P(A|B)로 변화하며, 이 때의 변화를 계산하는 방법을 제공한다.

위의 결과(표 2)에서 보는 바와 같이 증상의 변화에 따라 췌장암 여부에 대한 진단 결과를 살펴볼 수 있었다. 실험 결과에서, 단지 하나의 증상으로는 췌장암 여부를 판단하기 어려울 정도로 낮은 확률이 도출 되었고 적어도 주요 증상(Abdominal pain, Nausea & Vomiting, Jaundice, Weight loss)에 두 가지 이상 나타나면 50% 이상의 췌장암 가능성이 나타나는 것을 확인할 수 있다.

주목할 만한 점은 CA19-9만 True일 경우에는 불과

8.46%의 췌장암 가능성을 보였지만 CA19-9와 Abdominal pain이 함께 True일 경우 97.47%의 높은 값을 보이는데 이것은 각각 True일 경우 8.46%와 4.39%를 보이던 것과는 많은 차이를 보인다. CA19-9와 다른 주요증상이 동시에 True경우에도 90% 내외의 높은 가능성을 보여주는데 이것은 베이지안 네트워크의 조건부 확률의 특성이 반영된 것으로 보이며 의사들이 췌장암 후보 군을 조기 선별하는데 중요한 역할로 작용할 것으로 보인다.

주의해야 할 부분은 췌장암 환자군에서 Abdominal Pain이 41.7%를 보이고 있는데 이 결과는 일부 다른 증상들 보다 훨씬 높은 확률로 베이지안 확률을 이용했을 때 흔한 증상에 해당하는 복통이 있는 사람은 췌장암일 가능성이 높다는 오해를 줄 수 있다. 췌장암 환자가 가지고 있는 복통증상이 다른 복통과는 다른 불규칙한 증상을 보이지만 일반인들은 이 증상을 구별하지 못하기 때문이다. 복통에 대해서는 좀 더 상세한 증상을 정의하고 일반인들이 충분히 구별할 수 있도록 가이드가 마련되어야 한다.

또한 향후 진단 시스템으로 개발 시에 선택지를 True/False 만으로 판단하는 것이 아니라 증상의 강도를 여러 단계로 나누어 반영할 수 있도록 개발 하고 실제 췌장암 후보 군을 발굴하여 검증하는 작업이 필요하다.

## Reference

- [1] A.K. Sarnecka, M. Zagozda, M. Durlak, "An Overview of Genetic Changes and Risk of Pancreatic Ductal Adenocarcinoma", *Journal of Cancer*, vol. 7, no. 14, pp. 2045-2051, 2016.
- [2] M. Wahutu, K. Sara et al., "Pancreatic Cancer: A Survival Analysis Study in Oklahoma", *Journal of Oklahoma State Medical Association*, vol. 109, no. 7-8, pp. 391-398, 2016.
- [3] S. Ngamruengphong, M. I. Canto, "Screening for Pancreatic Cancer", *Surgical clinics of North America*, vol. 96, pp. 1223-1233, 2016.
- [4] A.Mohamed, Y.Saad et al., "Can Serum ICAM1 distinguish pancreatic cancer from chronic pancreatitis", *Asian Pacific Journal of Cancer Prevention*, vol. 17, no. 10, pp. 4671-4675.
- [5] G. Ge, G.W. Wong, "Classification of premalignant pancreatic cancer mass-spectrometry data using decision tree ensembles", *BMC Bioinformatics*, vol. 9, pp. 275, 2008.
- [6] F. Gorunescu, M. Grunescu et al., "Competitive collaborative neural computing system for medical diagnosis in pancreatic cancer detection", *Biological Journal of the Linnean Society*, vol. 28, no. 1, pp. 33, 2011.
- [7] F.T. Dombal, D.J. Leaper et al., "Computer-aided diagnosis of acute abdominal pain", *British Medical Journal*, vol. 2, pp. 9-13, 1972.
- [8] A. Alvaro-Meca, R. Gil-Prieto, Gil de Miguel A., "In-Hospital Death Caused by Pancreatic Cancer in Spain: Application with Bayesian Network", *International Journal of Biomedical Science*, vol. 7, no. 2, pp. 125-130, 2011.
- [9] L. You, J. Xiao et al., "Analysis of clinical characteristics and treatment of pancreatic cystic tumors", *Chinese Journal of Cancer Research*, vol. 28, no. 5, pp. 519-527, 2016.
- [10] A. Engebretson, L. Matrisian, C. Thompson, "Patient and caregiver awareness of pancreatic cancer treatments and clinical trials", *Journal of Gastrointestinal Oncology*, vol. 7, no. 2, pp. 228-233, 2015.