

# Analysing Risk Factors of 5-Year Survival Colorectal Cancer Using the Network Model

Won Jun Park\*, Young Ho Lee\*, Un Gu Kang\*

## Abstract

The purpose of this study is to identify the factors that may affect the 5-year survival of colon cancer through network model and to use it as a clinical decision supporting system for colorectal cancer patients. This study was conducted using data from 2,540 patients who underwent colorectal cancer surgery from 1996 to 2018. Eleven factors related to survival of colorectal cancer were selected by consulting medical experts and previous studies. Analysis was proceeded from the data sorted out into 1,839 patients excluding missing values and outliers. Logistic regression analysis showed that age, BMI, and heart disease were statistically significant in order to identify factors affecting 5-year survival of colorectal cancer. Additionally, a correlation analysis was carried out age, BMI, heart disease, diabetes, and other diseases were correlated with 5-year survival of colorectal cancer. Sex was related with BMI, lung disease, and liver disease. Age was associated with heart disease, heart disease, hypertension, diabetes, and other diseases, and BMI with hypertension, diabetes, and other diseases. Heart disease was associated with hypertension, diabetes, hypertension, diabetes, and other diseases. In addition, diabetes and kidney disease were associated. In the correlation analysis, the network model was constructed with the Network Correlation Coefficient less than  $p < 0.001$  as the weight. The network model showed that factors directly affecting survival were age, BMI levels, heart disease, and indirectly influencing factors were diabetes, high blood pressure, liver disease and other diseases. If the network model is used as an assistant indicator for the treatment of colorectal cancer, it could contribute to increasing the survival rate of patients.

▶Keyword: Colorectal Cancer, Logistic Regression Analysis, Correlation Analysis, Network Model

## I. Introduction

현대인의 불규칙한 식습관과 미세먼지, 황사 등 환경오염물질 영향으로 암 발생률이 꾸준히 증가하고 있다[1]. 세계보건기구에서 2017년 발표한 사망원인 통계자료에 의하면 암으로 인한 사망은 전 세계 사망 원인 중 12%를 차지한다[2]. 특히 대장암은 우리나라에서 폐암, 위암에 이어서 세 번째로 많이 발생하는 암으로 네 번째로 사망률이 높은 암이다[3]. 대장암으로 인한 사망은 2000년대 들어 가장 크게 증가하였으며 매년

대장암으로 인한 사망률은 꾸준히 증가하고 있다[4]. 이에 따라 대장암 생존율을 높이기 위해 생존과 관련이 있는 위험 요인들을 파악하는 많은 연구가 진행되고 있다[5]. [6]은 대장암 진단을 받은 960명의 환자를 대상으로 회귀 분석을 통해 대장암 5년 생존에 영향을 미치는 요인을 파악하였다. [7]은 서울 소재 S 병원에서 대장암 환자 1,878명을 대상으로 로지스틱 회귀분석을 통해 대장암 5년 생존에 영향을 미치는 요인을 파악하였다. [8]은 대장

---

• First Author: Won Jun Park, Corresponding Author: Un Gu Kang  
\*Won Jun Park (dnjswns2280@naver.com), Dept. of Computer Engineering, Gachon University  
\*Young Ho Lee (lyh@gachon.ac.kr), Dept. of Computer Engineering, Gachon University  
\*Un Gu Kang (ugkang@gachon.ac.kr), Dept. of Computer Engineering, Gachon University  
• Received: 2019. 07. 23, Revised: 2019. 09. 18, Accepted: 2019. 09. 19.

암 수술을 받은 환자들을 대상으로 대장암 5년 생존에 영향을 미치는 요인들을 로지스틱 회귀분석을 통해 확인하였다.

위에서 언급한 연구들에서는 대장암 생존에 영향을 줄 수 있는 요인들을 파악했을 뿐, 그 요인들 간의 연관성을 파악하지는 못했다. 본 연구에서는 로지스틱 회귀분석을 통해 대장암 5년 생존에 영향을 미치는 요인을 파악하고, 상관 분석을 통하여 각 요인 사이의 관계를 파악하여 네트워크 모델로 구축함으로써 생존에 직간접적으로 영향을 미치는 요인들을 파악하였다.

## II. Preliminaries

### 1. Network Model Structure Overflow

Fig. 1은 네트워크 모델을 구축하는 전체적인 흐름을 나타낸다. 본 연구는 1996년부터 2018년까지 대장암 수술을 받은 2,540명 환자데이터를 사용하여 연구를 진행하였으며, 대장암 임상과의 의무기록사의 조언과 선행연구를 바탕으로 대장암 5년 생존과 관련이 있는 요인들을 추출하였다. 추가적으로 로지스틱 회귀분석을 통해 생존에 영향을 미치는 요인들을 파악하였고, 상관분석을 통해 요인들 간의 연관성을 확인하였다. 이를 네트워크 모델로 구축하여 대장암 5년 생존에 직간접적으로 영향을 미치는 요인을 파악하였다.

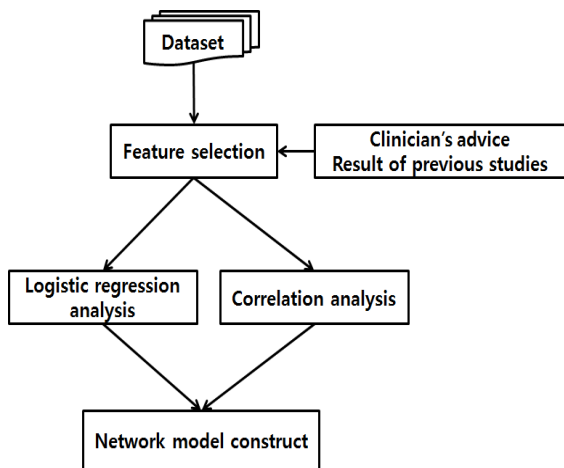


Fig. 1. Network Model Structure Overflow

### 2. Colorectal Cancer Patient Data

대장암 임상과의 의무기록사의 조언과 선행 연구를 바탕으로 대장암 5년 생존과 관련이 있는 요인을 선정한 결과 성별, 나이, BMI 지수, 심장 질환, 폐 질환, 고혈압, 신장 질환, 간 질환, 기타 질환이 대장암 5년 생존과 관련이 있었다. 선정된 요인에서 결측치와 이상치를 제외한 1,839명의 환자 데이터를 추출하여 분석을 진행하였다. Table 1은 추출한 환자의 데이터를 나타낸다. 남자는 1,066명(58.0%), 여자는 773명(42.0%)이었고 50세 이하는 281명(15.3%), 51-60세는 438명(23.8%),

61-70세는 589명(32.0%), 71세 이상은 531명(28.9%)으로 나타났다. 또한 BMI 지수가 20kg/m<sup>2</sup> 이하는 358명(14.6%), 20kg/m<sup>2</sup>-25kg/m<sup>2</sup>는 1037명(61.3%), 25kg/m<sup>2</sup> 초과는 444명(24.1%)으로 나타났다. 각 질환 별 환자의 수는 심장 질환 232명(4.1%), 폐 질환 65명(3.5%), 고혈압 528명(28.7%), 당뇨병 232명(12.6%), 신장 질환 16명(0.9%), 간 질환 26명(1.4%) 기타 질환 253명(13.8%)으로 나타났다.

Table 1. Variables of Factors Affecting 5-Year Survival of Colorectal Cancer (n=1839)

Variables	Frequency	Percent
Survival	Yes	1,329 (72.3)
	No	510 (27.7)
Sex	Female	773 (42.0)
	Male	1,066 (58.0)
Age	≤50	281 (15.3)
	51-60	438 (23.8)
	61-70	589 (32.0)
	≥71	531 (28.9)
BMI (kg/m <sup>2</sup> )	<20	358 (14.6)
	20-25	1037 (61.3)
	>25	444 (24.1)
Heart disease	Yes	232 (4.1)
	No	1764 (95.9)
Lung disease	Yes	65 (3.5)
	No	1774 (96.5)
Hypertension	Yes	528 (28.7)
	No	1311 (71.3)
Diabetes	Yes	232 (12.6)
	No	1607 (87.4)
Kidney disease	Yes	16 (0.9)
	No	1823 (99.1)
Liver disease	Yes	26 (1.4)
	No	1813 (98.6)
Other disease	Yes	253 (13.8)
	No	1586 (86.2)

### 3. Methods

#### 3.1 Logistic Regression Analysis

대장암 5년 생존에 영향을 미치는 요인을 파악하기 위해 로지스틱 회귀분석을 실시하였다. 로지스틱 회귀는 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법이다[9]. 로지스틱 회귀의 목적은 일반적인 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용한다[10]. 이는 독립 변수의 선형 결합으로 종속 변수를 설명한다는 관점에서는 선형 회귀 분석과 유사하지만, 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에 일종의 분류 기법으로 볼 수 있다[11]. 본 논문에서는 대장암 5년 생존과 관련이 있다는 요인을 대상으로 5년 이상 생존 유무를 확인하기 위해 식(1)의 로지스틱 회귀분석을 사용하였다.

$$\ln \frac{p}{1-p} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

### 3.2 Network Correlation Coefficient

대장암 5년 생존과 관련이 있는 요인들 간의 연관성을 확인하기 위해 상관분석(Correlation Analysis)을 실시하였다. 상관분석은 확률론과 통계학에서 하나의 종속 변수와 다른 하나의 독립 변수 사이의 관련성을 분석하는 것이다. 두 변수는 서로 독립적인 관계이거나 상관된 관계일 수 있으며 이때 두 변수 간의 관계의 강도를 상관계수(Correlation Coefficient)라 한다. 상관계수를 구할 때는 보편적으로 피어슨 상관계수를 사용한다[12].

네트워크 모델을 구축하여 대장암 5년 생존에 직간접적으로 영향을 미치는 요인을 파악하기 위해서는 노드와 노드를 연결하는 edge가 필요하다[13]. 하지만 피어슨 상관 계수는 -1에서 1의 값을 나타내기 때문에 피어슨 상관계수로 연관성을 확인할 수는 있지만 edge로 하여 네트워크 모델을 구축할 수는 없다. 따라서 본 연구에서는 피어슨 상관계수의 절댓값을 구하는 NCC(Network Correlation Coefficient)공식을 제안하여 edge를 구하였다. 식(2)는 NCC를 계산하는 식이다.

$$NCC = \left| \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \right| \quad (2)$$

### 3.3 Network Model Construct Tool

대장암 5년 생존에 직간접적으로 영향을 미치는 요인을 파악하는 네트워크를 구축하기 위해 NodeXL을 사용하였다. NodeXL은 MicroSoft에서 개발된 엑셀 add-in 기반의 네트워크 분석 및 시각화 도구이다[14]. NodeXL에 노드와 edge를 설정하여 네트워크 모델을 구축한다[15].

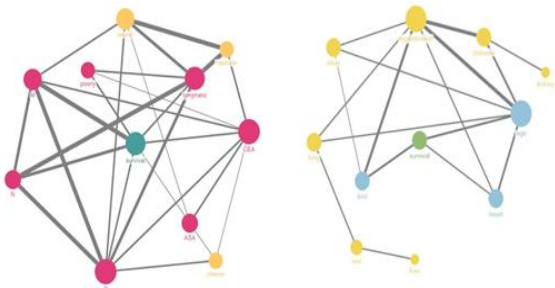


Fig. 2. Example of Network Model Using NodeXL

## III. Result

### 1. Logistic Regression Analysis on the 5-Year Survival of Colorectal Cancer

Table 2는 대장암 5년 생존에 영향을 미치는 요인을 파악하기 위해 로지스틱 회귀분석을 실시한 결과이다. 먼저 선정된 변

수의 다중공선성의 진단결과, 모든 변수의 VIF(Variation Inflation Factor)값이 10미만으로 다중공선성의 문제가 없음을 확인하였으며, 분석결과 나이( $\beta=-0.005$ ,  $p<0.001$ ), BMI 지수( $\beta=0.014$ ,  $p<0.001$ ), 심장 질환( $\beta=-0.178$ ,  $p<0.001$ )은 통계적으로 유의하게 분석되어 대장암 5년 생존에 영향을 미치는 것으로 나타났다. 하지만 성별( $\beta=0.005$ ,  $p=0.808$ ), 폐 질환( $\beta=-0.009$ ,  $p=0.879$ ), 고혈압( $\beta=0.014$ ,  $p=0.563$ ), 당뇨병( $\beta=-0.059$ ,  $p=0.063$ ), 신장 질환( $\beta=-0.085$ ,  $p=0.438$ ), 간 질환( $\beta=-0.138$ ,  $p=0.110$ ), 기타 질환( $\beta=-0.048$ ,  $p=0.105$ )은 통계적으로 유의하지 않게 분석되어 대장암 5년 생존에 영향을 미치지 않았다.

Table 2. Result of Logistic Regression Analysis on the 5-Year Survival of Colorectal Cancer (n=1839)

Variables	$\beta$	t	p	VIF
Sex	0.005	0.242	0.808	1.012
Age	-0.005	-6.108	<0.001	1.082
BMI (kg/m <sup>2</sup> )	0.014	4.532	<0.001	1.044
Heart disease	-0.175	-3.363	<0.001	1.017
Lung disease	-0.009	-0.163	0.879	1.013
Hypertension	0.014	0.579	0.563	1.181
Diabetes	-0.059	-1.861	0.063	1.085
Kidney disease	-0.085	-0.776	0.438	1.006
Liver disease	-0.138	-1.597	0.110	1.006
Other disease	-0.048	-1.623	0.105	1.013

### 2. Correlation Analysis and Network Model Structure on the 5-Year Survival of Colorectal Cancer

Table 3은 대장암 5년 생존과 관련이 있는 요인들을 대상으로 상관분석을 실시한 결과이다. 대장암 5년 생존은 나이가 감소하고( $r=-0.16$ ,  $p<0.001$ ) BMI 지수가 증가할수록( $r=0.11$ ,  $p<0.001$ ) 증가하였다. 또한 심장 질환과( $r=-0.09$ ,  $p<0.001$ ) 당뇨병( $r=-0.06$ ,  $p<0.05$ ), 기타 질환이( $r=-0.06$ ,  $p<0.05$ ) 없을 때 증가하였다. 남성이 여성에 비해 BMI 지수가 낮았고( $r=-0.05$ ,  $p<0.05$ ) 폐 질환과( $r=0.06$ ,  $p<0.001$ ) 간 질환이( $r=0.06$ ,  $p<0.05$ ) 발생할 가능성이 높았다. 나이가 증가할수록 심장 질환( $r=0.11$ ,  $p<0.001$ ), 폐 질환( $r=0.06$ ,  $p<0.001$ ), 고혈압( $r=0.24$ ,  $p<0.001$ ), 당뇨병( $r=0.11$ ,  $p<0.001$ ), 기타 질환( $r=0.08$ ,  $p<0.001$ )이 발생할 가능성이 높았다. BMI 지수가 증가할수록 고혈압( $r=0.18$ ,  $p<0.001$ ), 당뇨병( $r=0.06$ ,  $p<0.05$ ) 발생할 가능성이 높았고 기타 질환( $r=-0.05$ ,  $p<0.05$ )이 발생할 가능성이 낮았다. 폐 질환이 있을 때, 고혈압( $r=0.09$ ,  $p<0.001$ )이 발생할 가능성이 높았고 고혈압이 있을 때, 당뇨병( $r=0.07$ ,  $p<0.001$ )이 발생할 가능성이 높았다.

Table 3. Result of Correlation Analysis on the 5-Year Survival of Colorectal Cancer (n=1839)

Node1	Node2	Correlation coefficient	NCC	p
Survival	Age	-0.16	0.16	<0.001
Survival	BMI	0.11	0.11	<0.001
Survival	heart disease	-0.09	0.09	<0.001
Survival	diabetes	-0.06	0.06	<0.05
Survival	other disease	-0.06	0.05	<0.05
Sex	BMI	-0.05	0.05	<0.05
Sex	lung disease	0.06	0.06	<0.001
Sex	liver disease	0.06	0.06	<0.05
Age	heart disease	0.11	0.11	<0.001
Age	lung disease	0.06	0.06	<0.001
Age	hypertension	0.24	0.24	<0.001
Age	diabetes	0.11	0.11	<0.001
Age	other disease	0.08	0.08	<0.001
BMI	hypertension	0.18	0.18	<0.001
BMI	diabetes	0.06	0.06	<0.05
BMI	other disease	-0.05	0.05	<0.05
Heart disease	hypertension	0.09	0.09	<0.001
Heart disease	diabetes	0.05	0.05	<0.05
Lung disease	hypertension	0.07	0.07	<0.001
Hypertension	Diabetes	0.27	0.27	<0.001
Hypertension	other disease	0.06	0.06	<0.05
Diabetes	kidney disease	0.07	0.07	<0.001

Fig. 3은 대장암 5년 생존과 관련이 있는 요인들을 노드로 하고, p<0.001미만인 NCC를 가중치로 하여 네트워크 모델을 구축하였다. 생존과 직접적으로 연관이 있는 노드는 적색으로 표시하였고 간접적으로 연관이 있는 노드는 황색으로 표시하였다.

네트워크 모델을 확인한 결과 대장암 5년 생존에 직접적으로 영향을 미치는 요인에는 나이, BMI 지수, 심장 질환이었고, 간접적으로 영향을 미치는 요인에는 성별, 폐 질환, 고혈압, 당뇨병, 신장 질환, 간 질환, 기타 질환이 있었다.

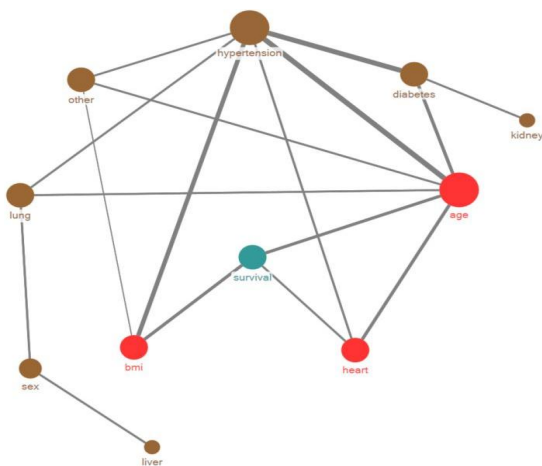


Fig. 3. Network Model Structure on the 5-Year Survival of Colorectal Cancer

#### IV. Consideration

로지스틱 회귀분석을 통해 각 요인들이 대장암 5년 생존에 미치는 영향을 확인한 결과 대장암 5년 생존은 나이, BMI 지

수, 심장 질환에 영향을 받았다. 연령이 높을수록 대장암 5년 생존율이 감소하였는데, 이는 노화가 대장암 생존율을 감소한다는 선행연구 결과와 일치하였다[16-17]. 또한 BMI 지수가 20미만인 저체중 그룹보다 20-25인 정상 그룹과 26이상인 과체중 그룹에서 대장암 5년 생존율이 증가하였는데, 이는 저체중 환자보다 과체중 환자의 암 생존율이 높다는 선행연구 결과와 일치하였다[18-20]. 심장 질환을 가지고 있는 환자의 대장암 5년 생존이 감소하였는데 이는 심장 질환 환자의 대장암 5년 존율이 감소한다는 선행연구 결과와 일치하였다[21].

상관분석을 통해 각 요인들의 연관성을 확인한 결과, 성별과 BMI 지수가 연관성이 높았는데, 이는 성별에 따라 BMI 지수가 달라진다는 결과와 일치하였다[22]. 또한 남성과 폐 질환, 간 질환의 연관성이 높았는데, 이는 남성이 폐 질환, 간 질환이 발생할 가능성이 여성보다 높다는 선행연구 결과와 일치하였다[23-24]. 나이와 다양한 합병증이 연관성이 있었는데, 이는 나이가 증가할수록 합병증이 발생하는 빈도가 청년보다 높았다는 선행연구 결과와 일치하였다[25]. 또한 BMI 지수와 당뇨병, 고혈압이 연관성이 있었는데, 이는 BMI 지수가 당뇨병과 고혈압이 연관성이 있다는 선행연구 결과와 일치하였다[26-27]. 또한 고혈압과 당뇨병이 연관성이 있었는데, 이는 고혈압 합병증으로 당뇨병이 있다는 선행연구 결과와 일치하였다[28].

#### V. Conclusions

본 연구는 1996년부터 2018년까지 대장암 수술을 받은 2,540명 환자데이터를 사용하여 연구를 진행하였으며, 임상적 의의기록사의 조인과 선행연구를 바탕으로 대장암 5년 생존

과 관련이 있는 요인들을 선정하였다. 선정된 요인에서 결측치와 이상치를 제거한 1,839명의 환자 데이터를 추출하여 분석을 하였다. 대장암 5년 생존에 영향을 미치는 요인을 파악하기 위해 로지스틱 회귀분석을 실시한 결과 나이( $\beta=-0.005$ ,  $p<0.001$ ), BMI 지수( $\beta=0.014$ ,  $p<0.001$ ), 심장 질환( $\beta=-0.178$ ,  $p<0.001$ )이 생존에 영향을 미친다고 나타났다.

또한 상관분석을 통해 요인들 간의 연관성을 확인한 결과 대장암 5년 생존은 나이가 감소하고( $r=-0.16$ ,  $p<0.001$ ) BMI 지수가 증가할수록( $r=0.11$ ,  $p<0.001$ ) 증가하였다. 또한 심장 질환과( $r=-0.09$ ,  $p<0.001$ ) 당뇨병( $r=-0.06$ ,  $p<0.05$ ), 기타 질환이( $r=-0.06$ ,  $p<0.05$ ) 없을 때 증가하였다. 남성이 여성에 비해 BMI 지수가 낮았고( $r=-0.05$ ,  $p<0.05$ ) 폐 질환과( $r=0.06$ ,  $p<0.001$ ) 간 질환이( $r=0.06$ ,  $p<0.05$ ) 발생할 가능성이 높았다. 나이가 증가할수록 심장 질환( $r=0.11$ ,  $p<0.001$ ), 폐 질환( $r=0.06$ ,  $p<0.001$ ), 고혈압( $r=0.24$ ,  $p<0.001$ ), 당뇨병( $r=0.11$ ,  $p<0.001$ ), 기타 질환( $r=0.08$ ,  $p<0.001$ )이 발생할 가능성이 높았다. BMI 지수가 증가할수록 고혈압( $r=0.18$ ,  $p<0.001$ ), 당뇨병( $r=0.06$ ,  $p<0.05$ ) 발생할 가능성이 높았고 기타 질환( $r=-0.05$ ,  $p<0.05$ )이 발생할 가능성이 낮았다. 폐 질환이 있을 때, 고혈압( $r=0.09$ ,  $p<0.001$ )이 발생할 가능성이 높았고 고혈압이 있을 때, 당뇨병( $r=0.07$ ,  $p<0.001$ )이 발생할 가능성이 높았다.

대장암 5년 생존과 관련이 있는 요인들을 노드로 하고,  $p<0.001$ 미만인 NCC를 가중치로 하여 네트워크 모델을 구축했을 때, 대장암 5년 생존에 나이, BMI 지수, 심장 질환이 직접적으로 영향을 미치고 성별, 폐 질환, 고혈압, 당뇨병, 신장 질환, 간 질환, 기타 질환이 간접적으로 영향을 미친다는 것을 확인할 수 있었다. 본 연구에서 제안하는 네트워크 모델을 대장암 환자 진단 시에 치료방침을 세우는 보조지표로 활용한다면 환자의 생존율을 높이는데 기여할 수 있을 것이다. 하지만 각 특징들이 대장암 5년 생존에 미치는 영향을 확인했을 뿐, 그 특징들의 그룹별 영향력을 분석하지 못했다는 연구의 한계점이 있다. 향후 연구에서 대장암 5년 생존과 관련이 있는 요인들을 추가하여 네트워크를 확장한다면 더욱 신뢰성 있는 보조지표를 만들 수 있을 것이다.

## REFERENCES

- [1] D. N. Son, S. W. Moon, D. Y. Ha, H. G. Jung, and B. W. Jung, "The Level of Serum CEA Associated with Recurrence after Potentially Curative Surgery for Colorectal Cancer," *Annals of Surgical Treatment and Research*, Vol. 75, No. 4, pp. 250-254, Oct. 2008.
- [2] H. S. Park, H. Cho, and H. S. Kim, "Development of an Evidence-based Nursing Process System to Improve Clinical Decision Making with Colorectal Cancer Nursing Diagnosis," *Journal of Korea Multimedia Society*, Vol. 19, No. 7, pp.1197-1207, Jul. 2016.
- [3] P. J. Oh, and J. H. Kim, "Chemotherapy-related Cognitive Impairment and Quality of Life in People with Colon Cancer: The Mediating Effect of Psychological Distress," *Journal of Korean Academy of Nursing*, Vol. 46, No. 1, pp.19-28, Feb, 2016.
- [4] Y. A. Baek. and M. S. Yi, "Factors Influencing Quality of Life during Chemotherapy for Colorectal Cancer Patients in South Korea," *Journal of Korean Academy of Nursing*, Vol. 45, No. 4, pp.604-612, Aug. 2015.
- [5] B. M. Kang, G. S. Choi, K. H. Lim, I. J. Park, and S. H. Jun, "Risk Factors of Peritoneal Recurrence after Curative Resection of Colorectal Cancer," *Annals of Surgical Treatment and Research*, Vol. 79, No. 5, pp.355-361, Aug. 2010.
- [6] S. J. Joeng, S. Y. Ryu, H. R. Kim, D. Y. Kim, and Y. J. Kim, "Multivariate analysis of prognostic factors in colorectal cancers," *Annals of Surgical Treatment and Research*, Vol. 61, No. 2, pp.173-178, Aug. 2001.
- [7] S. W. Park, D. W. Lee, J. W. Park, S. B. Ryoo, R. Shin, S. Y. Jeong, and K. Y. Park, "Impact of body mass index on overall survival after surgery for colorectal cancer," *Korean Journal of Clinical Oncology*, Vol. 12, No. 2, pp.91-96, Dec. 2016.
- [8] Y. G. Park, B. S. Cho, Y. D. Yoon, Y. J. Kang, and J. S. Park, "Analysis of the survival rate after operations for colorectal cancer," *Annals of Surgical Treatment and Research*, Vol. 55, No. 4, pp.540-548, Oct. 1998.
- [9] C. H. Lee, "Fine-Grain Weighted Logistic Regression Model," *The Institute of Electronics and Information Engineers*, Vol. 53, No. 9, pp.77-81, Jul. 2016.
- [10] H. J. Byung, and D. H. Lim, "Performance Comparison of Logistic Regression Algorithms on RHadoop," *Journal of the Korea Society of Computer and Information*, Vol. 22, No. 4, pp.9-16, Apr. 2017.
- [11] C. H. Lee, and M. N. Jung, "Value Weighted Regularized Logistic Regression Model," *Journal of the Korea Society of Computer and Information*, Vol. 43, No. 11, pp.1270-1274, Nov. 2016.
- [12] M. S. Choi, "An Analysis about Recognition of Indoor Air Quality of Workers at Dental Clinics in Jeollanamdo Area," *Journal of the Korea Society of Computer and Information*, Vol. 23, No. 11, pp.137-142, Nov. 2018.
- [13] Y. T. Kwak, "A Layer-by-Layer Learning Algorithm using Correlation Coefficient for Multilayer Perceptrons," *Journal of the Korea Society of Computer and Information*, Vol. 16, No. 8, pp.39-47, Aug. 2011.
- [14] G. C. Lee, and S. W. Youn, "Smart Speaker Market

- Analysis and Forecast using Google Trends," *KIISE Transactions on Computing Practices*, Vol. 24, No. 11, pp.596-602, Nov. 2018.
- [15] M. N Lee, and J. H. Hong, "Semantic Network Analysis of Government's Crisis Communication Messages during the MERS Outbreak," *The Journal of the Korea Contents Association*, Vol. 16, No. 5, pp.124-136, May. 2016.
- [16] Y. B. Park, and S. E. Moon, "A clinical study of colorectal cancer," *Annals of Surgical Treatment and Research*, Vol. 46, No. 2, pp242-249, Jun. 1994.
- [17] B. P. Kim, K. H. Kwon, and H. S. Min, "A clinical study of colorectal cancer," *Annals of Surgical Treatment and Research*, Vol. 54, No. 3, pp.374-382, Oct. 1998.
- [18] Y. Moriwaki, C. Kunisaki, S. Kobayashi, H. Harada, S. Imai and C. Kasaoka, "Does body mass index (BMI) influence morbidity and long-term survival in gastric cancer patients after gastrectomy?," *Journal of Hepato Gastroenterology*, Vol. 50, No. 49, pp.284-288, Jan. 2003.
- [19] C. Eroglu, O. Orhan, H. Karaca, D. Unal, M. Dikilitas, M. Ozkan, et al., "The effect of being overweight on survival in patients with gastric cancer undergoing adjuvant chemoradiotherapy," *European Journal of Cancer Care*, Vol. 22, No. 1, pp.133-140, Sep. 2012.
- [20] N. M. Tsang, P. C. Pai, C. C. Chuang, W. C. Chuang, C. K. Tseng, K. P. Chang, et al., "Overweight and obesity predict better overall survival rates in cancer patients with distant metastases," *Journal of Cancer Medicine*, Vol. 5, No. 4, pp.665-675, Jan. 2016.
- [21] C. H. Kroenke, R. Neugebauer, J. Meyerhardt, C. M. Prado, E. Weltzien, M. L. Kwan, et al., "Analysis of body mass index and mortality in patients with colorectal cancer using causal diagrams," *JAMA Oncology*, Vol. 2, No. 9, pp.1137-1145, Sep. 2016.
- [22] H. S. Kim, and N. C. Kim, "Difference of the Obesity Index, Blood Pressure and Serum Lipids in Abdominal and Non Abdominal in Men and Women," *Journal of Korean Academy of Nursing*, Vol. 30, No. 4, pp. 948-955, Aug. 2000.
- [23] Y. G. Yoo, Y. R. Park, and E. H. Choi, "A Comparative Study of Life Styles between Lung Cancer Patients and Healthy Adults," *Asian Oncology Nursing*, Vol. 13, No. 3, pp.136-144, Sep. 2013.
- [24] H. J. Lee, and H. S. Park "Experience of Patients' Management of Hepatocellular Carcinoma with Transarterial Chemoembolization," *Journal of Qualitative Research*, Vol 18, No. 12, pp.78-92, May. 2015.
- [25] D. G. Shin, S. H. Seong, W. S. Kim, and D. S. Kim, "The Effect of Old Age on the Development of Postoperative Complication in Surgery for Gastric Carcinoma," *Annals of Surgical Treatment and Research*, Vol. 69, No. 6, pp.455-458, May. 2005.
- [26] S. B. Lee, and S. Y. Park, "Effects of body fat and physical activity habit on cardiovascular risk factors in type 2 diabetic subjects," *KSW*, Vol. 7, No. 3, pp.237-248, May. 2012.
- [27] C. G. Kim, "Effect of Health Behavior and Obesity Indices on Blood Pressure in 20s Man," *The Journal of the Korea Contents Association*, Vol. 11, No. 8, pp.231-238, Aug. 2011.
- [28] S. S. Kim, "The effect of a long term aerobic training on blood lipid profiles and body fat in hypertensive obesity and diabetic obesity patients," *Journal of Sport and Leisure Studies*, Vol. 11, No. pp.183-193, Nov. 1999

## Authors



analysis.

Won Jun Park is a student of computer engineering at Gachon University. He is currently conducting research on Watson, which was first introduced at Gil Hospital in Incheon. His research interests include machine learning, deep learning, and data



Engineering, Gachon University .of medicine and science, Korea. His research interests include AI and mobile healthcare.

Young Ho Lee received a M.S. degrees from Hankuk University of Foreign Studies and a Ph.D. degree from Ajou University, Korea, in 1995 and 2007. He has worked for IBM Korea. He is currently a professor in the School of Computer Information



Software, Healthcare Information, U-healthcare.

Un Gu Kang received Ph.D. degree in Computation Engineering from Inha University in 2001. He is currently a Professor in Department of Computer Engineering at Gachon University. His primary research interests include Mobile