

머신러닝 기법을 활용한 고혈압 환자의 건강 관련 삶의 질 요인 예측

정재혁¹ · 조성현^{2*}

¹남부대학교 일반대학원 물리치료학과 박사과정 학생, ^{2*}남부대학교 물리치료학과 교수

Using Machine Learning Techniques to Predict Health-Related Quality of Life Factors in Patients with Hypertension

Jae-Hyeok Jeong, PT, MS¹ · Sung-Hyoun Cho, PT, Ph.D^{2*}

¹*Dept. of Physical Therapy, Graduate School of Nambu University, Ph.D-Student*

^{2*}*Dept. of Physical Therapy, Nambu University, Professor*

Abstract

Purpose : This study aims to identify the factors influencing health-related quality of life through machine learning of the general characteristics of patients with hypertension and to provide a basis for related research on patients, such as intervention strategies and management guidelines in the field of physical therapy for health promotion.

Methods : Annual data from the second Korean Health Panel (Version 2.0) from 2019 to 2020, conducted jointly by the Korea Health and Social Research Institute and the National Health Insurance Service, were analyzed (Korea Health Panel, 2024). The data used in this study was collected from January to July 2020, and the data was collected using computer-assisted face-to-face interviews. Of the 13,530 household members surveyed, 1,368 were selected as the final study participants after removing missing values from 3,448 individuals diagnosed with hypertension by a doctor.

Results : The results showed that walking (P2) was the most significant factor affecting health-related quality of life in random forest, followed by perceived stress (HS1), body mass index (BMIc), total household income (TOTc), subjective health status (SRHc), marital status (Marr), and education level (Edu).

Conclusion : To prevent and manage chronic diseases such as hypertension, as well as to provide customized interventions for patients in advanced stages of the disease, research should be conducted in the field of physical therapy to identify influencing factors using machine learning. Based on the findings of this study, we believe that there is a need for additional content that can be utilized in the field of physical therapy to improve the health-related quality of life of patients with hypertension, such as diagnostic assessment and intervention management guidelines for hypertension, and education on perceived stress and subjective health status.

Key Words : health related quality of life, hypertension, machine learning, physical therapy, predictors

*교신저자 : 조성현, shcho@nambu.ac.kr

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었음(No. 2022R1F1A1067604).

제출일 : 2024년 4월 15일 | 수정일 : 2024년 6월 1일 | 게재승인일 : 2024년 6월 21일

I. 서론

1. 연구의 배경 및 필요성

고혈압은 만성질환 중 높은 유병률과 여러 합병증으로 인한 사망이 매우 높은 질환이다(Samadian 등, 2016). 고혈압은 뇌졸중 및 관상동맥 질환의 주요 위험 인자이며, 고혈압 환자는 신부전, 심부전, 말초혈관질환 등 합병증에 걸리기 쉽다(Mills 등, 2016). 전 세계적으로 14억 명의 성인(18세 이상 여성의 20%, 남성의 25%)이 고혈압을 앓고 있는 것으로 추정된다(Mills 등, 2016). 이와 같은 이유로, 전 세계적으로 사망과 질병 부담 측면에서 커다란 문제가 되고 있다(Clement, 2015).

고혈압 환자들은 고혈압의 합병증, 진단에 대한 의식, 동반질환, 항고혈압제 복용에 대한 부작용으로 인해 삶의 질이 저하된다. 고혈압 환자는 심혈관 문제 및 합병증의 위험에 있어 매우 취약한 집단이며 고혈압은 당뇨병보다 건강 관련 삶의 질이 낮은 질환인 것으로 나타났다(Chang 등, 2017). 그러므로 고혈압 환자의 삶의 질을 향상시키기 위해 건강 관련 삶의 질과 관련된 요인들을 파악하고 중재를 계획하는 것은 주요한 과제이다(Lee & Cho, 2016). 고혈압은 단일 요인에 의해 발생하는 질환이 아니라, 유전적 요인, 식습관 및 영양섭취, 운동 부족, 스트레스 등과 같은 다양하고 복합적인 요인들이 작용하여 발생하기 때문에(Lee, 2018; The Korean Society of Hypertension, 2024), 고혈압의 건강 관련 삶의 질에 영향을 미치는 요인을 분석함에 따라 이러한 다양한 측면의 요인들을 종합적으로 고려해볼 필요가 있다.

4차 산업혁명의 핵심 기술 중 하나인 인공지능(AI)은 전례 없는 속도로 발전하며 다양한 분야에 혁신을 가져오고 있다. 보건의료분야에서의 AI 기술의 도입은 진단, 치료, 예방, 연구 등에 확대되어 최근 인공지능을 활용하고 결과를 임상에 적용하려는 노력들이 점차 많아지고 있다. 앞으로는 여러 형태의 빅데이터를 분석하여 미래를 예측하고 보다 새로운 가치를 만들어냄으로써 그에 따른 결과를 기반으로 보건의료정책 수립에 보탬이 될

수 있다고 보고하였다(Back 등, 2018). 인공지능이란 빅데이터의 시각화(visualization), 머신러닝(machine learning), 딥러닝(deep learning)의 기술을 통해 자료를 분석하고 더 발전된 결과를 찾아내는 것이다(Lee & Kim, 2016). 인공지능은 방대한 양의 데이터 축적과 컴퓨팅 파워의 개선 및 네트워크의 활성화 등으로 인해 기술 수준이 급성장하여 헬스케어시스템을 포함해 다양한 분야에서 사용할 수 있다(Kim, 2018).

또한, 최근 머신러닝, 딥러닝에 대한 알고리즘의 발전함에 따라 이를 활용하여 다양한 질병들을 예측하고 진단하기 위한 연구들이 진행되고 있다(Kang 등, 2022). 머신러닝은 기존의 데이터를 활용하여 환경에 따른 데이터의 특징들을 추출 및 테스트를 거쳐 최적화 또는 자가 발전을 하는 일련의 과정이며, 이러한 과정의 반복은 환경에 따라 변하는 데이터를 주어진 환경과 연관하여 보다 정확한 분류를 가능하게 한다(Lee 등, 2016).

머신러닝을 활용한 고혈압의 진단 및 예측 모델과 관련된 연구들에서는 소득, 가구원의 수, 결혼 여부, 성별 등의 사회경제적 요인, 흡연, 음주, 운동 등과 같은 건강행태, 가족력이나 건강지표 등 다양한 요인들을 사용하고 있다(Kang 등, 2022). 빅데이터 기반의 머신러닝을 활용한 실제 임상현장 적용의 선행연구를 살펴보면, 국민건강영양조사 데이터를 활용하여 심근경색 및 협심증 발생을 머신러닝으로 예측하는 연구가 보고되었다(Lim, 2018). 또한, 국민건강보험공단에서 제공하는 공공데이터를 이용하여 뇌졸중 위험도를 높은 정확도로 예측하는 머신러닝 모델이 제시되었다(Jeong 등, 2021).

2. 연구의 목적

본 연구는 고혈압 환자의 성별, 나이, 교육 수준, 가구 총 소득, 결혼 여부, 체질량 지수, 우울감, 불안감, 자살생각, 스트레스 인지 정도, 주관적 건강 상태, 규칙적 운동 여부, 걷기, 흡연 여부, 음주 여부 정도를 머신러닝의 알고리즘 중 하나인 랜덤포레스트와 로지스틱 회귀분석 알고리즘을 적용하고 물리치료 영역에서 고혈압 환자의 건강 관련 삶의 질에 대한 영향요인을 파악하고자 한다. 각 분석방법에 따라 제시되는 주요 요인들을 비교하고 제시하여 고혈압 환자에게 고혈압 관리지침, 교육, 접근

가능한 내용 등을 살펴봄으로써 물리치료 연구의 기초 자료를 제공하고자 한다.

II. 연구방법

1. 연구자료 및 대상

본 연구는 한국의료패널의 홈페이지에서 연간 데이터에 대한 자료 활용 동의서를 작성한 후 사용승인을 받아 본 연구의 원시데이터로 사용하였다. 한국보건사회연구원과 국민건강보험공단이 공동으로 진행한 2019년부터 2020년까지 한국의료패널 2기 연간 데이터(Version 2.0)를 분석에 활용하였다(Korea Health Panel, 2024). 2기 한국의료패널은 전국적인 규모의 대표성 유지를 위해 2016년 인구주택총조사의 등록 센서스를 표본추출틀로 하고 있다. 17개의 시도, 동부 및 읍면 부를 지역별로 표본 조사구를 선정한 후, 확률비례 2단 층화집락추출법으로 708개의 표본 조사구 내의 가구에서 조사대상을 선정하였다. 본 연구에서 활용한 2기 연간 데이터의 조사 시기는 2020년 01월부터 07월까지 조사되었으며, 컴퓨터를 활용한 대인 면접 방법(computer assisted personal interviewing; CAPI)으로 자료가 수집되었다. 본 연구는 조사된 13,530명의 가구원 중, 의사로부터 고혈압 진단을 받은 3,448명에서 결측값을 제거한 1,368명을 최종 연구대상자로 선정하였다.

2. 변수정의

본 연구는 고혈압 환자의 건강 관련 삶의 질의 예측 모형에 대한 선행연구를 바탕으로 다음과 같이 변수를 정의하였다(Byeon & Cho, 2015; Kim 등, 2019).

1) 일반적 특성

일반적 특성은 성별, 나이, 결혼 여부, 교육 수준, 가구 총 소득, 규칙적 운동 여부, 걷기, 체질량지수(body mass index; BMI), 흡연 여부, 음주 여부, 스트레스 인지 정도, 우울감, 불안감, 자살 생각, 주관적 건강 상태를 확인하였다. 성별은 남성과 여성으로 분류하였으며, 연령은 60

세 미만과 60세 이상으로 분류하였다. 결혼 여부는 기혼, 미혼, 이혼/별거/사별로 분류하였으며, 교육 수준은 중학교 졸업 이하, 고등학교 졸업, 대학 졸업 이상으로 구분하였다. 가구 총 소득은 2,000만 이하, 2,000~5,000만, 5,000만 이상으로 분류하였으며, 규칙적 운동 여부는 예, 아니오로 분류하였다. 걷기는 전혀 걷지 않음, 3일 이하, 4~6일 사이, 매일 걸음으로 분류하였으며, 체질량지수는 대한비만학회의 비만 진료지침의 기준에 따라 대상자의 체중(weight; WT)과 신장(height; HT)을 계산하여 산출하였다(Rhee, 2022). 저체중, 정상, 과체중, 비만으로 분류하였다. 흡연 여부는 과거 흡연 및 현재 흡연, 비흡연으로 분류하였으며, 음주 여부는 비음주자 및 1년간 안마심, 한 달 내 음주로 분류하였다. 우울감과 불안감 및 자살 생각은 예, 아니오로 분류하였으며, 주관적 건강 상태는 좋음, 보통, 나쁨으로 분류하였다.

2) 건강 관련 삶의 질

Euroqol 그룹이 개발한 삶의 질 도구(Euroqol-5 dimension; EQ-5D)를 활용하여 연구대상자의 건강 관련 삶의 질을 나타냈다(Devlin & Brooks, 2017). EQ-5D index는 건강관련 삶의 질을 5가지 항목과 3가지 수준으로 구성되어 있다. 건강 관련 삶의 질의 5가지 항목은 운동 능력(mobility), 자기 관리(self-care), 일상 활동(usual activity), 통증/불편감(pain/discomfort), 불안/우울(anxiety/depression)이다. 각 항목은 ‘전혀 문제없음’(1점), ‘다소 문제있음’(2점), ‘많이 문제 있음’(3점)의 3가지 수준(3 level version)으로 분류된다(Kim & Min, 2020). 각 항목의 점수를 합산하면 최소 5점(모든 문항에서 ‘전혀 문제 없음’을 선택한 경우)에서 최대 15점(모든 문항에서 ‘많이 문제 있음’을 선택한 경우)이 나오게 된다. 설문 응답자가 자신의 상태를 응답하면, 모든 EQ-5D 상태에 대해 일반 인구집단에서 질 가중치를 구한 가치평가 세트(valuation set)를 이용하여 각 EQ-5D 상태에 해당하는 질 가중치 값을 EQ-5D index라고 한다. 이 중 5문항 모두에 대해서 중간 값인 10점(모든 문항에서 ‘다소 문제 있음’을 선택한 경우)은 질 가중치를 반영한 EQ-5D index 값 0.677에 해당된다. 따라서 2개의 그룹으로 분류하기 위해 cut-off point를 EQ-5D index 값 0.678로 정하였다

으며, 이 값을 기준으로 0.678 이상인 그룹을 1, 0.678 미만인 그룹을 0으로 코드화하여 분류하였다. 0.678 미만인 그룹은 0.678 이상인 그룹에 비해 상대적으로 건강 관련 삶의 질이 낮다는 것을 의미한다.

3. 자료 처리 및 분석

본 연구의 빈도분석과 기술통계량은 IBM SPSS software 프로그램(version 22.0, IBM Corp., USA)을 사용하여 분석하였고, 머신러닝과 그에 따른 시각화는 R program version 4.2.3의 패키지 중 random forest와 foreign을 사용하였다. 대상자의 개인적, 환경적 특성, 생리적 요인, 증상 경험, 일반적 건강 지각, 건강 증진 행위 및 건강 관련 삶의 질은 빈도 분석과 기술 통계량을 이용하여 실수와 백분율, 평균과 표준편차를 산출하였다. 연구대상자의 건강 관련 삶의 질에 미치는 영향 요인을 파악하기 위해 랜덤포레스트, 로지스틱 회귀분석 알고리즘을 활용하였다. 통계적 유의수준은 양측검정에서 .05로 설정하였다.

1) 랜덤포레스트

랜덤포레스트는 훈련 데이터 세트에서 n개의 데이터 포인트를 활용하여 부트스트랩 샘플을 구성하고 입력 변수의 하위 집합을 무작위로 선택하여 의사 결정 트리를 생성한 다음 이를 선형 방식으로 집계하여 최종 학습자를 형성한다. R의 랜덤포레스트에서는 변수에 대한 중요한 지수를 제공한다. 특정 변수에 대한 중요도 지수는 그 변수를 포함하지 않을 경우에 대하여 그 변수에 포함할 때에 예측오차가 어느 정도 줄어드는지를 보여주는 것이다(Park, 2015). 머신러닝에서의 모형 평가같은 경우, 모형 수립 시 포함하지 않았던 데이터를 예측 모형에 적합해 봄으로써 모형의 일반화 가능성을 평가하는 방법을 활용하여 예측 성능을 평가한다.

머신러닝 분류 모델의 성능을 평가할 수 있는 지표는 정확도(accuracy), 재현율(recall), 정밀도(precision), 그리고 F1-score이다. 재현율은 실제 긍정인 것들 중에서 긍정으로 예측한 비율(true positive / (true positive + false negative))이고, 정밀도는 긍정으로 예측한 것들 중에서 실제 긍정인 경우의 비율(true positive / (true positive +

false positive))이다. F1-score는 재현율과 정밀도의 조화 평균값(2*(재현율*정밀도)/(재현율 + 정밀도))으로, 재현율과 정밀도를 종합하여 보완한 것이다.

2) 로지스틱 회귀분석

로지스틱 회귀는 대표적인 머신러닝의 지도 학습 방식 중 하나이다. 해당 기술은 독립변수의 선형결합을 사용하여 0부터 1 사이의 범주에 데이터가 속할 확률을 예측하는 수학적 기법이다(Connolly, 2020). 로지스틱 회귀 분석은 직접적인 값을 예측하는 선형 분석과는 다르게 종속변수가 특정 범주에 속할 확률을 이용해 두 개의 클래스 중 한 개의 클래스로 예측할 수 있다. 이 때 해당 확률에 따라 분류 기준 값(cut-off)을 적용하여 더 높은 가능성이 있는 범주에 속하는 것으로 분류하는 방법이다. 이러한 분석 방법은 원시적이고 기본적인 방법이지만 여전히 여러 분야에서 분류 및 예측을 위해 사용되고 있으며 예측하고자 하는 데이터가 두 개의 클래스를 갖는 범주형 데이터일 때 사용할 수 있다(Cho 등, 2023).

III. 결 과

1. 연구 대상자의 일반적 특성

본 연구의 연구대상자의 일반적 특성은 다음과 같다(Table 1). 성별은 남성보다 여성이 더 많았으며(53.3 %), 나이는 60세 이상(91.0 %)의 분포가 가장 많았다. 교육 수준은 중학교 졸업 이하(59.2 %)가 가장 많았으며, 가구 총 소득은 2,000~5,000만원(38.9 %)이 가장 많았다. 결혼 여부는 기혼이 가장 많았으며(71.5 %), 체질량지수는 비만(40.6 %)이 가장 많았다. 우울감은 있음보다 없음(91.9 %)이, 불안감은 있음보다 없음(94.3 %)이, 또한 자살 생각은 있음보다 없음(93.6 %)이 가장 많았다. 스트레스 인지 정도는 많이 느끼는 편이다가 가장 많았으며(48.7 %), 주관적 건강 상태는 보통(46.6 %)이 가장 많은 것으로 나타났다. 규칙적 운동 여부는 아니오보다 예가 가장 많으며(53.1 %), 걷기는 매일 걷는 경우가 가장 많았다(29.9 %). 흡연 여부는 과거 흡연 및 현재 흡연보다 비흡연이 더 많았으며(61.3 %), 음주 여부는 한달 내 음주보

Table 1. Descriptive statistics for observed variables

(n= 1,368)

Characteristics	Categories	n (%)
Gender	Male	639 (46.70)
	Female	729 (53.30)
Age (year)	<60	123 (9.00)
	≥60	1245 (91.00)
Educational level	Middle school ≥	810 (59.20)
	High school	378 (27.60)
	College ≤	180 (13.20)
Household income (ten million won)	2,000 ≥	478 (34.90)
	2,000~5,000	532 (38.90)
	5,000 ≤	358 (26.20)
Marital status	Married	978 (71.50)
	Single	19 (1.40)
	Separation/Bereavement/Divorce	371 (27.10)
Body mass index (kg/m ²)	Underweight	29 (2.10)
	Normal weight	386 (28.20)
	Overweight	398 (29.10)
	Obesity	555 (40.60)
Depression	Yes	111 (8.10)
	No	1257 (91.90)
Anxiety	Yes	78 (5.70)
	No	1290 (94.30)
Suicidal thoughts	Yes	87 (6.40)
	No	1281 (93.60)
Level of stress	More stress	359 (26.20)
	A lot of stress	666 (48.70)
	A little stress	307 (22.40)
	None stress	36 (2.60)
Subjective health status	Poor	430 (31.40)
	Moderate	638 (46.60)
	Good	300 (21.90)
Regular exercise	Yes	727 (53.10)
	No	641 (46.90)
Number of walking (days/week)	Not walking	306 (22.40)
	1~3 days	312 (22.80)
	4~6 days	341 (24.90)
	Every day	409 (29.90)
Smoking	Current smoker + Former smoker	530 (38.70)
	Never	838 (61.30)
Drinking	Non-drinking + Not drinking for 1 year	701 (51.20)
	Drinking within a month	667 (48.80)

다 비 음주자 및 1년간 안 마심이 더 많았다(51.2 %).

2. 연구 대상자의 건강 관련 삶의 질

대상자의 건강 관련 삶의 질은 5가지의 하부 영역 중 운동능력, 자기 관리, 일상활동, 불안/우울에 대한 결과, 응답은 문제없음(68.5~90.1 %), 다소 문제 있음(9.1~31.1 %), 매우 심각한 문제 있음(0.3~0.7 %) 순으로 문제없음이 가장 많은 것으로 나타났다. 통증/불편은 문제없음이 48.7 %, 다소 문제 있음은 49.1 %, 심각한 문제 있음이 2.2 %로, 상대적으로 다른 영역과 비교하여 통증/불편을

겪는 대상자가 상대적으로 많은 것으로 나타났다(Table 2). 운동 능력(EQ1)에서 불안/우울(EQ5)까지의 변수들은 모두 최솟값이 1이고 최댓값이 3인 변수이다. 평균값과 중앙값이 1에 가깝다는 것은 대부분의 응답이 가장 낮은 범주를 선택했다는 것을 의미한다. EQ-5D 종합 점수는 5에서 15 사이의 값을 가지며, 대상자의 전반적인 삶의 질에 대한 평균 및 표준편차는 6.35±1.60점이며, EQ-5D Index의 평균 및 표준편차는 0.94±0.24점으로 나타났다. 이는 응답자들 사이에 건강 상태의 다양성을 나타내며, 비교적 더 넓은 범위의 분포를 보인다.

Table 2. Health-related quality of life in study subjects (n= 1,368)

Characteristics	Categories	n (%)
Mobility (EQ1)	No problems	937 (68.50)
	Some problems	425 (31.10)
	Extreme problems	6 (.40)
Self-care (EQ2)	No problems	1233 (90.10)
	Some problems	125 (9.10)
	Extreme problems	10 (.70)
Usual activity (EQ3)	No problems	1070 (78.20)
	Some problems	289 (21.10)
	Extreme problems	9 (.70)
Pain/Discomfort (EQ4)	No problems	666 (48.70)
	Some problems	672 (49.10)
	Extreme problems	30 (2.20)
Anxiety/Depression (EQ5)	No problems	1140 (83.30)
	Some problems	224 (16.40)
	Extreme problems	4 (.30)
EQ-5D index	< .678	80 (5.80)
	≥ .678	1288 (94.20)

3. 연구 대상자의 건강 관련 삶의 질에 대한 영향요인 분석

1) 랜덤포레스트

본 연구의 랜덤포레스트를 활용하여 고혈압 환자의 건강 관련 삶의 질을 종속변수로 설정한 모형에 대한 검증데이터의 예측 결과는 다음 (Table 3)과 같다. 정확도 (accuracy)는 94.5 %이며, 재현율(recall)은 100 %, 정밀도 (precision)는 94.5 %이다. 재현율(recall)과 정밀도

(precision)의 조화평균값에 해당하는 F1-score는 97.1 %로 나타났다.

머신러닝의 모형은 통계적 유의성 진단을 수행하지는 않지만 랜덤포레스트는 예측에 영향을 미친 설명변수들의 상대적 정확도와 중요도를 분석할 수 있다. Table 4는 상대적으로 정확도 및 중요도가 높은 설명변수부터 순서대로 작성한 표이며 Fig 1은 하나의 범주 및 척도를 구성하는 설명변수를 중심으로 상대적 정확도 및 중요

도를 시각화 한 것이다. 각 설명변수별 정확도 중 가장 높은 정확도는 우울감(HS2)으로 나타났으며, 그 뒤를 이어 불안감(HS3), 결혼 여부(Marr), 주관적 건강 상태(SRHc), 자살 생각(HS4), 성별 (Gender) 등의 순으로 나타났다.

고혈압 환자의 건강 관련 삶의 질에 영향을 주는 주요 요인은 Table 4와 같다. 변수중요도를 측정하는 지표인 평균 지니불순도 감소량(mean decrease gini)의 값이 1 이상인 경우는 해당 설명변수를 기반으로 한 분할이 노드의 불순도를 최대한으로 감소시켜 모델의 예측에 중요

한 역할을 했다는 것을 의미한다. 반대로 값이 1 미만인 경우에는 노드의 불순도를 충분히 감소시키지 못하여 모델의 예측에 1이상인 경우에 비해 상대적으로 중요한 역할을 하지 못했다는 것을 의미한다. 따라서 건강 관련 삶의 질에 영향을 미치는 요인들 중 가장 큰 영향을 미치는 요인은 걷기(P2)로 나타났으며, 그 뒤를 이어 스트레스 인지 정도(HS1), 체질량지수(BMIc), 가구 총 소득(TOTc), 주관적 건강 상태(SRHc), 결혼 여부(Marr), 교육 수준(Edu) 등의 순으로 나타났다(Fig 1).

Table 3. Random forest classification model

(n= 1,368)

Random forest classification model (%)	
Accuracy	94.50
95 %CI	91.30~96.90
Recall	100
Precision	94.50
F1-Score	97.10

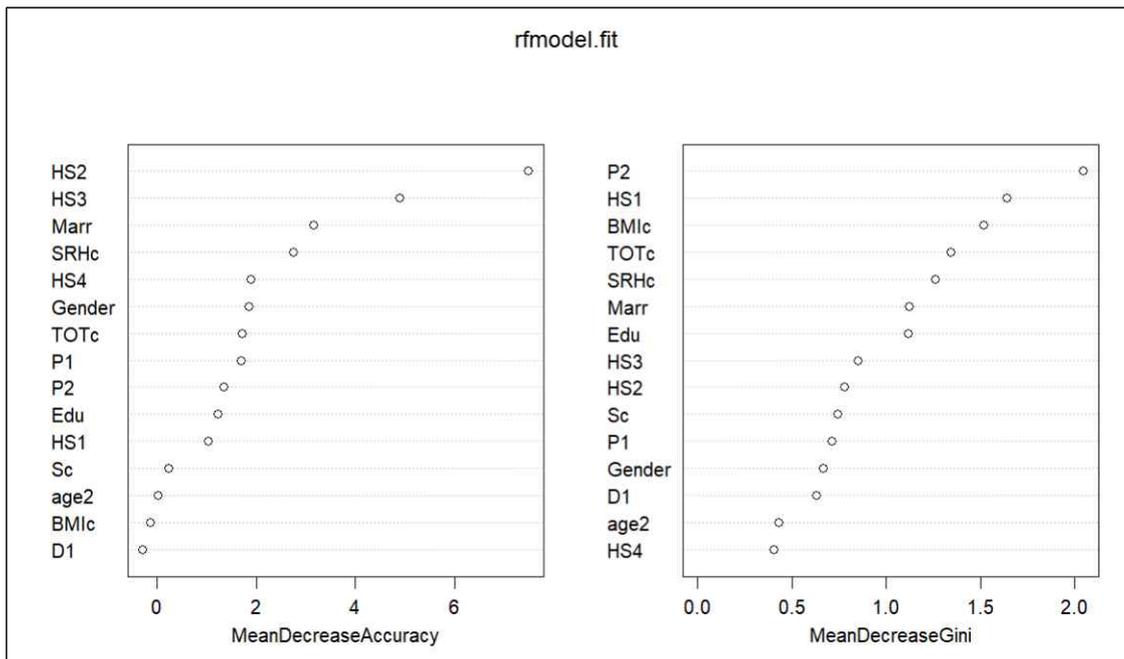


Fig 1. Random forest mean decrease accuracy & gini

Table 4. Random forest mean decrease accuracy & gini (n= 1,368)

Mean decrease accuracy		Mean decrease gini		
1	Depression (HS2)	7.47	Number of walking (P2)	2.04
2	Anxiety (HS3)	4.88	Level of stress (HS1)	1.63
3	Marital status (Marr)	3.27	Body mass index (BMIC)	1.51
4	Subjective health status (SRHc)	3.22	Household income (TOTc)	1.34
5	Suicidal thoughts (HS4)	1.89	Subjective health status (SRHc)	1.26
6	Gender (Gender)	1.84	Marital status (Marr)	1.12
7	Household income (TOTc)	1.70	Educational level (Edu)	1.11
8	Regular exercise (P1)	1.70	Anxiety (HS3)	.85
9	Number of walking (P2)	1.33	Depression (HS2)	.77
10	Educational level (Edu)	1.22	Smoking (Sc)	.73
11	Level of stress (HS1)	1.03	Regular exercise (P1)	.71
12	Smoking (Sc)	.23	Gender (Gender)	.66
13	Age (age2)	.02	Drinking (D1)	.63
14	Body mass index (BMIC)	-.13	Age (age2)	.43
15	Drinking (D1)	-.28	Suicidal thoughts (HS4)	.40

2) 로지스틱 회귀분석

고혈압 환자의 일반적 특성이 건강 관련 삶의 질에 미치는 영향 요인을 검증하기 위해 로지스틱 회귀분석을 실시하였다(Table 5).

로지스틱 회귀분석 결과, 고혈압 환자의 일반적 특성에 해당하는 모든 독립 변수들이 건강 관련 삶의 질에 통계적으로 유의미한 영향을 미치지 않는 것으로 나타났다(p>.05)(Table 6).

Table 5. Logistic regression model (n= 1,368)

Logistic regression model (%)	
Accuracy	94.10
95 %CI	92.70~95.30
Recall	100
Precision	100
F1-score	100

Table 6. The results of logistic regression analysis

Independent variable		B	S.E.	OR	95 %CI		p
					Lower	Upper	
Gender	Male			1			Ref.
	Female	.19	.27	1.214	.71	2.07	.477
Age (year)	60 >			1			Ref.
	60 ≤	-.00	.45	.997	.41	2.42	.995
Marital status	Married			1			Ref.
	Single	.11	1.06	1.125	.14	8.98	.911
	Separation /Bereavement /Divorce	-.45	.26	.637	.37	1.07	.092
Educational level	≤Middle school			1			Ref.
	High school	.39	.30	1.485	.81	2.69	.192
	≥College	.46	.42	1.590	.69	3.64	.273
Household income	≤2,000			1			Ref.
	2,000~5,000	.08	.27	1.084	.63	1.85	.767
	≥5,000	-.03	.29	.967	.54	1.72	.909
Regular exercise	Yes			1			Ref.
	No	-.15	.27	.852	.49	1.46	.562
Number of walking	Not walking			1			Ref.
	1~3 days	.11	.34	1.121	.56	2.22	.742
	4~6 days	.26	.37	1.309	.62	2.74	.474
	Every day	.30	.37	1.352	.64	2.84	.425
Body mass index	Underweight			1			Ref.
	Normal weight	.32	.78	1.380	.29	6.46	.682
	Overweight	-.00	.78	1.005	.21	4.65	.994
	Obesity	-.14	.76	.864	.19	3.88	.849
Smoking	Current smoker + Former smoker			1			Ref.
	Never	.10	.23	1.106	.69	1.76	.673
Drinking	Non-drinking + Not drinking for 1 year			1			Ref.
	Drinking within a month	-.14	.25	.867	.52	1.44	.581
	More stress			1			Ref.
Level of stress	A lot of stress	.23	.65	1.260	.34	4.57	.725
	A little stress	.30	.65	1.353	.37	4.85	.642
	None stress	.66	.69	1.936	.49	7.57	.342
Depression	Yes			1			Ref.
	No	.16	.52	1.176	.41	3.32	.759
Anxiety	Yes			1			Ref.
	No	.49	.59	1.641	.50	5.28	.406
Suicidal thoughts	Yes			1			Ref.
	No	-.44	.54	.638	.22	1.84	.407
Subjective health status	Good			1			Ref.
	Moderate	.18	.30	1.202	.65	2.19	.548
	Poor	.30	.35	1.355	.67	2.72	.394

Nagelkerke R²=.030

IV. 고 찰

본 연구는 고혈압 환자의 삶의 질을 파악하고, 고혈압 관련 질환 연구에 기초자료로 제공하기 위해 분석한 연구로 2019년부터 시행한 한국의료패널 2기 연간 데이터를 활용하여 2차 자료 분석을 통해 진행하였다. 대상자의 일반적 특성은 빈도분석과 기술통계 분석을 실시하였고, 고혈압 환자의 건강 관련 삶의 질에 미치는 영향요인을 확인하기 위해서 머신러닝의 기법 중 하나인 랜덤포레스트와 로지스틱 회귀분석 알고리즘을 활용하였다.

랜덤포레스트 알고리즘의 중요도 결과, 고혈압 환자의 건강 관련 삶의 질에 영향을 주는 가장 큰 주요 요인은 걷기(P2)로 나타났으며, 그 뒤를 이어 스트레스 인지 정도(HS1), 체질량 지수(BMIc), 가구 총 소득(TOTc), 주관적 건강 상태(SRHc), 결혼 여부(Marr), 교육 수준(Edu) 순으로 나타났다.

본 연구에서는 남녀 모두 걷기 실천을 하는 사람이 건강 관련 삶의 질이 유의하게 높게 나타났다는 선행연구의 결과와 일치하였고(Lee, 2014), 65세 이상 여성 노인, 고혈압 노인을 대상으로 한 연구에서도 신체활동을 할수록 건강 관련 삶의 질에 영향을 미쳤다는 선행연구의 결과와도 일치하였다(Kim & Bae, 2020; Kim, 2021). 고혈압 환자에게 걷기 등 운동을 통하여 혈압 저하, 심폐기능 개선, 체중 감소, 이상지질혈증 개선, 스트레스 해소의 효과가 있다고 보고되었다(The Korean Society of Hypertension, 2024). 이러한 이유로 매일 걷는 고혈압 환자가 상대적으로 적게 걷는 환자들에 비해 건강 관련 삶의 질에 대해서 높게 나타난다고 생각된다.

또한 스트레스 인지 정도는 건강 관련 삶의 질에 영향을 미치는 것으로 나타났다. 여러 선행연구들을 살펴보면, 대부분의 연구에서는 스트레스가 고위험군일수록 건강 관련 삶의 질이 낮아지는 것으로 나타났다. 이러한 연구결과는 스트레스라는 요인이 고혈압 환자뿐만 아니라 노인, 암환자, 다른 만성 질환을 가지고 있는 대상자의 삶의 질에도 영향을 미치는 것으로 알 수 있었다(Choi & Lee, 2015; Kim & Kim, 2018; Lee & Cho, 2016; Lee, 2014; Park & Hong, 2017). 보건소를 이용하는 고혈

압 환자의 건강 관련 삶의 질을 연구한 다른 연구에서도 스트레스가 많을수록 건강 관련 삶의 질은 낮아지는 것으로 확인되었다(Choi & Lee, 2015). 이러한 결과는 고혈압과 같은 만성질환을 앓고 있는 대상자들이 장기적인 질환으로 인해 스트레스를 경험하고, 이 스트레스가 지속되기 때문에 삶의 질이 낮아지는 것으로 생각된다. 본 연구 결과를 통해 스트레스는 다른 만성 질환과 마찬가지로 고혈압을 경험하는 대상자의 삶의 질에서도 중요한 요인인 것으로 판단된다.

체질량 지수(BMI)와 관련된 다른 선행연구 결과에 의하면 비만인 경우, 남자는 삶의 질 수준이 높은 반면, 여자는 삶의 질 수준이 낮았다(Lee, 2014). 또 다른 연구에서는 지난 1년 동안 체중이 증가한 대상자는 그렇지 않은 대상자보다 건강 관련 삶의 질이 낮게 나타났으며(Lee & Cho, 2016), 고혈압 노인을 대상으로 한 선행 연구에서는 건강 생활습관과 삶의 질에 관한 연구에서 저체중 노인보다 체질량 지수가 비만 바로 전 단계에서 노인의 삶의 질은 가장 높다고 보고하였다. 이는 노인에게 있어 저체중인 경우나 과체중인 경우는 오히려 영양섭취의 불균형 및 운동 부족 등 자기 관리 부족이 삶의 질 저하의 원인이 될 수 있다고 보고하였다(Bang & Hyeon, 2018). 고혈압의 유병률은 고도 비만군이 가장 높으며 그 다음으로 비만, 과체중, 정상군의 순서였다는 선행연구의 결과를 바탕으로(Youn 등, 2015), 올바른 영양섭취와 규칙적인 운동 등을 통해 체중조절을 실천한다면 건강 관련 삶의 질이 보다 높아질 것으로 생각된다.

가구 총 소득과 건강 관련 삶의 질과 관련된 선행 연구를 살펴보면, 노인 고혈압 환자를 대상으로 한 연구에서는 경제활동을 하는 경우 삶의 질이 높다고 나타났고(Bang & Hyeon, 2018), 암 환자를 대상으로 한 연구에서도 가구소득은 암 생존자들에게 있어서 삶의 질에 영향을 주는 요인으로 나타났으며(Kang, 2016), 여성 노인을 대상으로 한 연구에서도 경제활동을 하는 경우 삶의 질이 높은 것으로 나타나(Kim & Min, 2020), 고혈압을 대상으로 한 본 연구의 결과와 일치했다. 노인들의 경우 교통비, 병원비 등 경제적인 문제로 자신들의 건강관리 또는 질병치료를 보류할 경향이 높아 이러한 결과가 나왔을 것이라 생각된다. 또한, 이와 같은 결과는 고혈압과 같은 경우, 다른 심혈관 질환이나 만성 질환과 동반되어

이환할 수 있기 때문에 지속적으로 약에 대한 비용이나 검사비용 등 의료와 관련된 지출이 많은 것이 영향을 미쳤다고 생각된다.

주관적 건강 상태는 고혈압 환자의 건강 관련 삶의 질에 있어 영향을 미치는 것으로 나타났다. Korean Statistical Information Service(2024)의 조사에 의하면, 주관적인 건강 상태와 건강 관련 삶의 질 점수는 나이가 증가할수록 낮아지는 것으로 나타났다. 다른 연구에서는 주관적 건강 상태는 노년층에 비해 청장년층이 높은 것으로 나타났다(Kim, 2022). 고혈압 환자의 생애주기별 건강행태와 삶의 질을 조사한 연구의 결과에서는 남성, 배우자가 있는 경우, 경제활동을 하는 경우, 교육 수준이 높은 경우, 주관적 건강 상태가 좋을수록 삶의 질이 높은 것으로 나타났다(Kim & Min, 2020). 또한 주관적 건강 상태를 나쁘다고 판단하는 이유에 대해 신체적 증상 혹은 심리적 증상을 파악하는 것이 중요하다고 보고하였다(Kim & Min, 2020). 따라서 주관적 건강 상태는 본인이 고혈압 환자라는 긍정적 인식과 더불어 개인적 건강관리 차원에서 고혈압을 관리 하는 경우도 있으며, 부정적인 경우에도 건강관리가 필요하다는 동기부여가 되는 경우가 있는 것으로 보인다. 그러므로 향후 고혈압 환자의 건강 관련 삶의 질과 관련된 변인들을 조사할 때, 주관적 건강 상태에 대한 사전 정보를 체계적으로 파악한 후 연구를 진행하는 것이 필요하다고 생각된다.

교육 수준은 건강 관련 삶의 질에 높은 영향을 미치는 것으로 나타났다. 이는 노인의 건강 관련 삶의 질에 대한 연구를 한 Lee(2014)의 연구 결과에서도 교육 수준이 높을수록 삶의 질 수준이 높게 나타났다. 또 다른 연구에서도 나이가 낮을수록, 교육 수준이 높을수록, 직업이 있는 경우, 배우자가 있는 경우에서 건강 관련 삶의 질이 높은 것으로 나타났으며(Lee & Cho, 2016), 본 연구의 결과로 미루어 볼 때 선행연구와 마찬가지로 우리나라의 교육 수준이 삶의 질의 전반적인 수준을 높이는데 기여한 것으로 생각된다.

Korea Disease Control and Prevention Agency(2024)에 의하면, 2020년 고혈압 유병률은 남성 28.6%, 여성 16.8%로 여성에 비해 남성이 유병률이 높은 것으로 나타났다. Korea Hypertension Management Association(2024)의 고혈압 건강 상식에 의하면, 고혈압 환자에게 있어 운동

량 증가, 체중감량, 건강식사법, 스트레스 완화 등을 통해 고혈압을 관리하는데 있어 중요하다고 보고하였다. 그러므로 한국의료패널 2기 연간데이터를 활용하여 고혈압 환자의 건강 관련 삶의 질에 미치는 영향 요인을 파악한 본 연구의 결과들을 토대로, 고혈압과 같은 만성 질환에 대한 예방 및 관리를 비롯하여 이미 질환이 진행된 상황에서 환자에 대한 교육과 운동 종류 및 기간, 건강 상태와 관련하여 맞춤형 중재를 제공하기 위해 고혈압 진단평가와 중재 관리지침, 스트레스의 인지 정도와 주관적 건강 상태에 대한 교육 등 활용 가능한 추가적인 내용이 필요할 것으로 생각된다. 이를 위해서는 다른 보건의료분야 뿐만 아니라 물리치료 분야에서도 머신러닝을 활용하여 영향 요인을 파악하는 연구가 시행되어야 할 것이다.

본 연구는 다음과 같은 제한점을 가진다. 첫째, 랜덤포레스트와 로지스틱 회귀분석 간의 성능 평가와 비교는 이루어지지 않았기 때문에 향후 추가적 외부 검증 및 의사결정나무(decision tree), 서포트 벡터 머신(support vector machine; SVM)과 같은 다른 머신러닝 알고리즘을 기반으로 한 심층적인 연구가 이루어져야 할 것이다. 둘째, 랜덤 포레스트와 로지스틱 회귀분석 결과, 재현율(recall)이 100%가 나오는데, 이는 특이도(specificity)가 0이라는 뜻과도 같다. 특히 로지스틱 회귀분석의 결과, 고혈압 환자의 일반적 특성은 건강 관련 삶의 질에 유의미한 영향을 미치지 않는 것으로 나타났는데($p>.05$), 이러한 결과는 변수 중 EQ-5D index의 데이터 분포가 1,368명 중 0.678이상인 그룹이 1,288명, 0.678미만인 그룹이 80명으로 불균형하게 분포되어 로지스틱 회귀분석 알고리즘이 효과적으로 구별하지 못한 것으로 생각된다. 이러한 영향을 완전히 배제할 수 없었으며, 본 연구의 결과를 일반화하기에 어려움이 있다. 셋째, 대상자의 특성과 관련하여 한국의료패널에서 제공되는 자료 중 고혈압과 관련된 위험인자의 모든 요인을 충분히 활용하지 못하였다. 특히 건강 관련 삶의 질은 단일 요인 보다는 여러 요인의 복합적인 영향을 받을 가능성이 있으며, 고혈압 환자 개개인의 특성에 따라 건강 관련 삶의 질에 영향을 미치는 요인들이 다를 수 있다고 생각된다. 이에 따라 단일 변수의 영향보다는 다중 요인의 상호작용을 고려한 맞춤형 접근이 필요하다고 생각된다. 하지만 이

러한 제한점에도 불구하고, 본 연구는 우리나라 인구를 대상으로 하는 한국의료패널의 연간 데이터를 분석하여 고혈압의 건강 관련 삶의 질에 영향을 미칠 수 있는 성별, 나이, 결혼 여부 등을 비롯해 걷기, 규칙적 운동 여부 등 건강관리 요인, 스트레스 인지 정도, 주관적 건강 상태 등 여러 요인들을 분석하고 제시했다는 점에서 의의가 있다고 판단된다.

V. 결론

본 연구는 한국의료패널 2기 연간데이터를 활용하여 고혈압 환자의 건강 관련 삶의 질에 대한 영향요인 분석 모형을 활용하여 연구하였다. 본 연구에서 분석된 요인들을 활용하여 향후 물리치료적 관점에서의 고혈압과 관련된 진단평가와 중재 관리지침, 스트레스와 주관적 건강 상태에 대한 관리 교육 등에 대한 기초자료를 제공할 수 있을 것이라고 생각된다. 또한, 머신러닝의 분석 기법 중 랜덤포레스트와 로지스틱 회귀분석 알고리즘을 활용하여 전국 규모의 데이터를 분석한 연구로써 고혈압 환자의 일반적 특성에 따라 건강 관련 삶의 질 수준을 파악할 수 있었다. 이를 바탕으로 머신러닝을 통한 고혈압 환자의 건강 관련 삶의 질의 영향요인 분석과 관련된 다양한 분야에서 고혈압 관련 연구에 대한 중요한 기반 정보로 활용될 수 있을 것이다. 추후 연구에서도 고혈압 환자와 연관된 요인들을 보다 충분히 고려하고, 다양한 머신러닝 알고리즘을 적용하여 고혈압과 그에 관련된 질환에 대해서 물리치료 연구에 활용되기를 기대한다.

참고문헌

Baek SK, Park JH, Kang SH, et al(2018). A study on the development of severity-adjusted mortality prediction model for discharged patient with acute stroke using machine learning. *Journal of the Korea Academia-Industrial Cooperation Society*, 19(11),

126-136. DOI: 10.5762/KAIS.2018.19.11.126
Bang SY, Hyeon SS(2018). Health behaviors and quality of life in the elderly with high blood pressure. *Journal of Digital Contents Society*, 19(11), 2159-2166. DOI: 10.9728/dcs.2018.19.11.2159
Byeon HW, Cho SH(2015). The predictive modeling of middle-aged hypertension using integrated method of decision tree and neural network. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 23(4), 13-28. DOI: 10.14257/AJMAHS.2015.04.25
Chang SJ, Jang SJ, Lee SH, et al(2017). Factors influencing quality of life and its measurements in patients with hypertension : a systematic review. *Journal of Muscle and Joint Health*, 24(1), 24-36. DOI: 10.5953/JMJH.2017.24.1.24
Cho SU, Kim KL, Lee SE, et al(2023). A correlation analysis between depression and economic activity using logistic regression. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 159, 1086-1087.
Choi MN, Lee EH(2015). Relationships of depression symptom, self-esteem, and stress to health-related quality of life in patients with hypertension registered to a community health center. *Journal of Korean Public Health Nursing*, 29(2), 165-176. DOI: 10.5932/JKPHN.2015.29.2.165
Clement DL(2015). Control of hypertension: is the goal reached?. *65(1)*, 25-26. DOI: 10.1161/HYPERTENSIONAHA.114.04256
Connelly L(2020). Logistic regression. *Medsurg Nursing; Pitman*, 29(5), 353-354.
Devlin NJ, Brooks R(2017). EQ-5D and the euroqol group : past, present and future. *Appl Health Econ Health Policy*, 15(2), 127-137. DOI: 10.1007/s40258-017-0310-5
Jeong SW, Lee MJ, Yoo SY(2021). Machine learning-based stroke risk prediction using public big data. *Journal of Advanced Navigation Technology*, 25(1), 96-101. DOI:

- 10.12673/jant.2021.25.1.96
- Kang SA, Kim SH, Ryu MH(2022). Analysis of hypertension risk factors by life cycle based on machine learning. *Journal of the Korea Industrial Information Systems Research*, 27(5), 73-82. DOI: 10.9723/jksis.2022.27.5.073
- Kang SJ(2016). Factors influencing quality of life among cancer survivors : using KNHANES 2010-2014. *Journal of the Korea Contents Association*, 16(9), 628-637.
- Kim AC, Bae HJ(2020). Impact of physical activity, body mass index and depression on the health related quality of life according to the presence of hypertension in the elderly women. *Journal of the Korea Academia-Industrial Cooperation Society*, 21(11), 543-553. DOI: 10.5762/kais.2020.21.11.543
- Kim ES(2021). Factor influencing health-related quality of life in Korean hypertensive seniors with osteoarthritis. *Journal of the Korea Academia-Industrial cooperation Society*, 22(3), 169-180. DOI: 10.5762/KAIS.2021.22.3.169
- Kim HS, Jeong SH, Park SK(2019). Decision-tree analysis to predict blood pressure control status among hypertension patients taking antihypertensive medications. *Journal of Korean Biological Nursing Science*, 21(1), 85-97. DOI: 10.7586/jkbns.2019.21.1.85
- Kim HJ, Min ES(2020). Health behaviors and quality of life by life cycle of hypertensive patients. *Journal of Convergence for Information Technology*, 10(7), 58-66. DOI: 10.22156/CS4SMB.2020.10.07.058
- Kim KS(2022). A study on health behaviors of the young adults and the elderly with hypertension : a secondary analysis of the 2018 Korea health panel data. *Journal of Industrial Convergence*, 20(11), 141-148. DOI: 10.22678/JIC.2022.20.11.141
- Kim SH, Kim MJ(2018). Factors influencing quality of life in a serious case : the sixth Korea national health and nutrition examination survey. *Journal of Korean Society for School & Community Health Education*, 19(1), 13-25.
- Kim YJ(2018). Issues and responses to advances in AI technology. *Korea Institute of Science & Technology Evaluation and Planning, Issue Weekly*, 34(252), 1-29.
- Lee HH, Chung SH, Choi EJ(2016). A case study on machine learning applications and performance improvement in learning algorithm. *Journal of Digital Convergence*, 14(2), 245-258. DOI: 10.14400/JDC.2016.14.2.245
- Lee HS(2014). The factors influencing health-related quality of life in the elderly - focused on the general characteristics, health habits, mental health, chronic diseases, and nutrient intake status: data from the fifth Korea national health and nutrition examination survey (KNHANES V), 2010~2012. *Korean Journal of Community Nutrition*, 19(5), 479-489. DOI: 10.5720/kjcn.2014.19.5.479
- Lee HY(2018). Evaluation and management of hypertensive patients according to new hypertension guideline. *Korean Journal of Medicine*, 93(5), 447-451. DOI: 10.3904/kjm.2018.93.5.447
- Lee KE, Cho EH(2016). Factors influencing health related quality of life in patients with hypertension : based on the 5th Korean national health and nutrition examination survey. *Journal of the Korea Contents Association*, 16(5), 399-409. DOI: 10.5392/JKCA.2016.16.05.399
- Lee KY, Kim JH(2016). Artificial intelligence technology trends and IBM watson references in the medical field. *Korean Medical Education Review*, 18(2), 51-57. DOI: 10.17496/KMER.2016.18.2.51
- Lim HK(2018). Prediction of myocardial infarction/angina and selection of major risk factors using machine learning. *Journal of The Korean Data Analysis Society*, 20(2), 647-656. DOI: 10.37727/jkdas.2018.20.2.647
- Mills KT, Bundy JD, Kelly TN, et al(2016). Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*, 134(6), 441-50. DOI: 10.1161/CIRCULATIONAHA.115.018912
- Park CY(2015). *Data mining with R*. Rev. 1st, Seoul,

- Kyowoo media, pp.431.
- Park JA, Hong JY(2017). Factors influencing quality of life in adult cancer patients : the sixth Korea national health and nutrition examination survey (KNHANES VI-2), 2014. Journal of the Korea Academia-Industrial cooperation Society, 18(5), 382-390. DOI: 10.5762/KAIS.2017.18.5.382
- Rhee EJ(2022). Current status of obesity treatment in Korea: based on the 2020 Korean society for the study of obesity guidelines for obesity management. Journal of the Korean Medical Association, 65(7), 388-392. DOI: 10.5124/jkma.2022.65.7.388
- Samadian F, Dalili N, Jamalian A(2016). Lifestyle modifications to prevent and control hypertension. Iran J Kidney Dis, 10(5), 237-263.
- Youn JH, Kim JS, Cho KH, et al(2015). Trends in cardiovascular disease risk factors by obesity: Korea national health and nutrition examination survey 2001~2013. Korean Journal of Family Practice, 5(3), 382-388.
- Korea Disease Control and Prevention Agency. Trends in hypertension prevalence 2012-2022, 2024. Available at <https://doi.org/10.56786/PHWR.2024.17.17.3> Accessed May 29, 2024.
- Korea Health Panel. Sampling and panel characteristics(version 2.0). Available at <https://www.khp.re.kr:444/web/survey/sampling.do> Accessed March 10, 2024.
- Korea Hypertension Management Association. Hypertension health facts-cure, 2024. Available at <http://www.khma.or.kr/hypertension/cure.php> Accessed May 29, 2024.
- Korean Statistical Information Service, Subjective Health Status, Available at https://kosis.kr/statHtml/statHtml.do?orgId=143&tblId=DT_11430222_2022&vw_cd=MT_ZTITLE&list_id=114_11437_002_005&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE Accessed March 10, 2024.
- The Korean Society of Hypertension. Korean society of hypertension guidelines, 2024. Available at <https://www.koreanhypertension.org/reference/guide?mode=read&idno=10081> Accessed March 10, 2024.