Original Article

PTRS Physical Therapy Rehabilitation Science

https://doi.org/10.14474/ptrs.2023.12.2.80
eISSN 2287-7584
pISSN 2287-7576

Phys Ther Rehabil Sci
2023, 12(2), 80-91
www.jptrs.org

A Comparative Study of Predictive Factors for Hypertension using Logistic Regression Analysis and Decision Tree Analysis

SoHyun Kim[a] and SungHyoun Cho[b*]

[a]Department of Integrative Medicine, The Graduate School, Nambu University, Gwangju, Republic of Korea
[b]Department of Physical Therapy, Nambu University, Gwangju, Republic of Korea

Objective: The purpose of this study is to identify factors that affect the incidence of hypertension using logistic regression and decision tree analysis, and to build and compare predictive models.

Design: Secondary data analysis study

Methods: We analyzed 9,859 subjects from the Korean health panel annual 2019 data provided by the Korea Institute for Health and Social Affairs and National Health Insurance Service. Frequency analysis, chi-square test, binary logistic regression, and decision tree analysis were performed on the data.

Results: In logistic regression analysis, those who were 60 years of age or older (Odds ratio, OR = 68.801, $p < 0.001$), those who were divorced/widowhood/separated (OR = 1.377, $p < 0.001$), those who graduated from middle school or younger (OR = 1, reference), those who did not walk at all (OR = 1, reference), those who were obese (OR = 5.109, $p < 0.001$), and those who had poor subjective health status (OR = 2.163, $p < 0.001$) were more likely to develop hypertension. In the decision tree, those over 60 years of age, overweight or obese, and those who graduated from middle school or younger had the highest probability of developing hypertension at 83.3%. Logistic regression analysis showed a specificity of 85.3% and sensitivity of 47.9%; while decision tree analysis showed a specificity of 81.9% and sensitivity of 52.9%. In classification accuracy, logistic regression and decision tree analysis showed 73.6% and 72.6% prediction, respectively.

Conclusions: Both logistic regression and decision tree analysis were adequate to explain the predictive model. It is thought that both analysis methods can be used as useful data for constructing a predictive model for hypertension.

Key Words: Data mining, Decision tree, Logistic regression analysis, Machine learning, Hypertension

서론

최근 건강의 패러다임은 질병이 없는 것만이 최적의 건강한 삶이 아니라 가지고 있는 만성질환을 잘 관리하고 건강상태를 유지하는 것에 가치가 부여되고 있다. 만성질환 중 고혈압은 그 자체가 질병으로 인식됨과 동시에 높은 유병률 및 심혈관질환, 심근경색, 뇌졸중 등과 같은 다양한 합병증을 유발하여 사망률을 증가시킨다[1, 2]. 그러나 이러한 심각성에도 불구하고 중증질환이 발생하기 전까지는 특별한 자각 증상을 느끼지 못하고 방치하는 경우가 빈번하기 때문에 고혈압의 예방과 관리는 필수적이라 할 수 있다.

다수의 연구에서 고혈압의 위험요인으로 흡연[3], 음주[4], 스트레스[5] 등을 제시하고 있으나 이는 주로 소규모 데이터를 활용한 고혈압과 개별 요인들의 관련성 또는 영향에 초점을 맞추어 수행되어 각 요인이 고혈압의 유병 확률을 예측하는 것에 한계점이 있다. 따라서 대상자들의 혈압에 영향을 주는 특성을 다각적인 면에서 살펴볼 필요가 있으며 고혈압이 발병된 대상자들과 발병되지 않은 대상자들의 특성을 파악할 필요가 있다.

이에 선행연구에서는 고혈압의 유병 예측을 위해 기계 학습을 활용한 인공신경망분석과 결정트리결합[6]과 의사결정나무[7, 8], 랜덤 포레스트와 CatBoost, 다층 신경망 및 로지스틱 회귀분석 비교연구[9] 등이 이루어

Received: Feb 7, 2023 **Revised:** Apr 9, 2023 **Accepted:** Apr 11, 2023
Corresponding author: SungHyoun Cho (**ORCID** https://orcid.org/0000-0002-5108-4342)
Department of Physical Therapy, Nambu University
23, Cheomdanjungang-ro, Gwangsan-gu, Gwangju, Republic of Korea [62271]
Tel: +82-62-970-0232 **Fax:** +82-62-970-0492 **E-mail:** shcho@nambu.ac.kr

지고 있다. 그러나 기존의 선행 연구는 국외 대상자로 한 연구이거나 비교적 작은 집단의 데이터를 활용하였으며, 최근 조사된 대규모 데이터가 아닌 2016년 이전의 데이터를 근거로 하였다.

다른 분야에서는 각 연구 주제에 대한 예측 요인을 파악하기 위하여 기계 학습의 다양한 알고리즘의 성능을 비교하고 분류 및 예측하는 연구가 활발히 진행되고 있다[10-13]. 또한, 기계 학습의 다양한 알고리즘은 서로 다른 결과를 제시할 수 있기 때문에 관련 요인을 분류하고 예측하기 위해서는 기계 학습의 다양한 분석방법을 적용해야 하지만 단편적으로 하나의 알고리즘에 대한 고혈압 예측 결과의 부분만 고찰되어 각 알고리즘의 성능 비교 연구는 미비한 실정이다.

기계 학습의 알고리즘 중 하나인 로지스틱 회귀분석은 대상이 두 집단 이상으로 나누어진 경우에 개별 관측된 값들이 어느 집단으로 분류되는지를 분석하거나 예측하는 모델을 구축할 때 사용된다[14]. 또한 의사결정나무는 도표화 된 나무 구조를 통하여 분석결과를 쉽게 이해하고 설명하는 장점을 가지고 있으며, 하향식 노드에 따라 결정된 경로를 통과함으로써 집단을 분류하고 예측하는데 유용하다[15, 16]. 즉, 고혈압의 위험요인과 개인의 특성을 고려한 맞춤형 예측 모형 개발과 모형의 성능 비교는 성공적인 고혈압 예방 시스템 구축에 기여할 수 있는 가치 있는 연구라고 판단된다.

본 연구는 이러한 측면에서 로지스틱 회귀분석과 의사결정 나무 분석을 활용하여 고혈압 발병에 대한 영향요인을 파악하고 각 분석 방법에 따라 제시되는 분류정확도를 비교함으로써 가장 타당한 고혈압 예측 모형을 제시하여 고혈압 관련 연구에 기초자료를 제공하고자 한다.

연구 방법

연구 대상

본 연구는 한국보건사회연구원과 국민건강보험공단이 공동으로 진행한 2019년 한국의료패널 2기 연간데이터(Version 2.0)를 분석에 활용하였다. 한국의료패널의 홈페이지에서 연간데이터에 대한 자료활용동의서를 작성한 후 사용승인을 받아 본 연구의 원시데이터로 사용하였다[17]. 연간데이터의 조사 시기는 2019년 08월부터 2019년 11월까지 조사되었으며, 조사원이 대상가구를 직접 방문하여 컴퓨터를 활용하여 설문지를 보며 조사하는 대인면접방법인 CAPI(Computer Assisted Personal Interviewing) 방식으로 진행되었다. 표본 추출은 확률비례 2단 층화집락추출 방식으로 표본조사구와 표본가구를 선정하였다. 조사 대상자는 약 700개 조사구에 거주하고 있는 8,500가구와 14,741명의 가구원으로 이루어졌다. 그러나 조사된 14,741명 중, 변수에 결측치가 있는 경우는 제외되어 본 연구의 최종 대상자로 9,859명이 선정되었다.

데이터 가공

본 연구는 고혈압 예측 모형에 대한 선행연구를 기반으로 다음과 같이 데이터를 가공하였다[6, 8]. 준거 변수(Criterion variable)인 고혈압 유병여부는 고혈압이 아닌 경우 0, 고혈압인 경우는 1로 구분하여 코드화 하였다. 예측 변수(Predictor variable)인 성별은 남자는 0, 여자는 1로 코드화 하였으며, 나이는 39세 이하는 0, 40세～59세 사이는 1, 60세 이상은 2로 코드화 하였다. 결혼여부는 기혼 0, 미혼 1, 이혼/사별/별거 2로 코드화 하였으며, 교육수준은 중학교 졸업 이하는 0, 고등학교 졸업은 1, 대학 졸업 이상은 2로 코드화 하였다. 가구총소득은 2,000만원 이하인 경우 0, 2,000～5,000만원 사이인 경우는 1, 5,000만원 이상인 경우는 2로 코드화 하였으며, 규칙적 운동 여부는 예 0, 아니오 1로 코드화 하였다. 걷기 횟수(일/주)는 전혀 걷지 않으면 0, 3일 이하는 1, 4～6일 사이는 2, 매일 걷는 경우는 3으로 코드화 하였다. 체질량 지수(Body mass index, BMI)는 대한비만학회의 비만 진료지침의 기준[18]에 따라 대상자의 체중과 신장을 계산하여 산출하였다. 저 체중은 0, 정상 체중은 1, 과 체중은 2, 비만은 3으로 코드화 하였으며, 주관적 건강상태는 좋음 0, 보통 1, 나쁨 2로 코드화 하였다. 흡연여부는 과거 흡연과 현재 흡연을 통합하여 0으로 범주화 하였으며, 비 흡연은 1로 코드화 하였다. 음주여부는 비 음주자와 최근 1년간 한 잔도 마시지 않은 자를 0으로 범주화 하였으며, 한 달 이내 음주한 자는 1로 범주화 하였다.

자료 분석

본 연구의 자료 분석은 IBM SPSS 소프트웨어(version 26.0, IBM Corp., USA) 프로그램을 사용하였다. 대상자의 일반적 특성은 빈도분석과 일반적 특성에 따른 차이를 알아보기 위해 카이제곱(Chi-squared) 검정으로 분석하였다. 또한, 예측요인의 파악과 모형 구축을 위해 로지스틱 회귀분석(Logistic regression)과 의사결정나무(Decision tree) 분석을 실시하였다. 의사결정나무의 알고리즘 선정은 3번의 평균 값을 비교하여 CHAID (Chi-squared automatic interaction detection) 알고리즘을 선택하였다. 적용 비율은 선행연구에 따라 훈련표본(Training data)과 검정표본(Testing data)을 80:20으로 설정하였다[19]. 통계적 유의수준은 양측검정에서 0.05로 설정하였다.

연구 결과

연구대상자의 일반적 특성

연구 대상의 일반적 특성은 Table 1과 같다. 분석 대상자 총 9,859명 중, 고혈압 유병자는 31.3%, 고혈압 비유병자는 68.7%이었다. 성별은 남자 46.7%, 여자 53.3%이었고, 나이는 39세 이하 11.7%, 40～59세 사이 31.0%, 60세 이상이 57.3%였다. 결혼여부는 기혼 72.1%, 미혼 11.2%, 이혼/사별/별거 16.7%였으며, 교육수준은 중학교 졸업 이하 36.5%, 고등학교 30.4% 대학교 졸업 이상 33.0%였다. 총 가구 소득은 2,000만원 이하 25.8%, 2,000～5,000만원 사이 37.4%, 5,000만원 이

Table 1. General characteristics of subjects (n＝9,859)

Variable	Categories	Total N(%)	Yes N(%)	No N(%)	x^2 (Pvalue)
Hypertension		9,859(100)	3,086(31.3)	6,773(68.7)	
Sex	Male	4,600(46.7)	1,427(31.0)	3,173(69.0)	0.314 (0.576)
	Female	5,259(53.3)	1,659(31.5)	3,600(68.5)	
Age	≤39 years old	1,158(11.7)	9(0.8)	1,149(99.2)	1861.714*** (P＜0.001)
	40～59	3,052(31.0)	337(11.0)	2,715(89.0)	
	≥60 years old	5,649(57.3)	2,740(48.5)	2,909(51.5)	
Marital status	Married	7,111(72.1)	2,217(31.2)	4,894(68.8)	592.813*** (P＜0.001)
	Single	1,104(11.2)	59(5.3)	1,045(94.7)	
	Divorce/Widowhood/Separation	1,644(16.7)	810(49.3)	834(50.7)	
Level of education	≤Middle School	3,602(36.5)	1,854(51.5)	1,748(48.5)	1238.374*** (P＜0.001)
	High school	3,001(30.4)	826(27.5)	2,175(72.5)	
	≥College	3,256(33.0)	406(12.5)	2,850(87.5)	
Gross household income	≤2,000	2,545(25.8)	797(31.3)	1,748(68,7)	3.433 (0.180)
	2,000～5,000	3,685(37.4)	1,190(32.3)	2,495(67.7)	
	≥5,000	3,629(36.8)	1,099(30.3)	2,530(69.7)	
Regular exercise	Yes	5,113(51.9)	1,668(32.6)	3,445(67.4)	8.624** (0.003)
	No	4,746(48.1)	1,418(29.9)	3,328(70.1)	
Number of walking days/week	Not walking	1,695(17.2)	658(38.8)	1,037(61.2)	56.334*** (P＜0.001)
	3 days	2,223(22.5)	634(28.5)	1,589(71.5)	
	4~6 days	2,720(27.6)	810(29.8)	1,910(70.2)	
	Every day	3,221(32.7)	984(30.5)	2,237(69.5)	
Body mass index	Underweight	4,091(41.5)	1,041(25.4)	3,050(74.6)	125.610*** (P＜0.001)
	Normal weight	4,734(48.0)	1,664(35.1)	3,070(64.9)	
	Overweight	629(6.4)	258(41.0)	371(59.0)	
	Obesity	405(4.1)	123(30.4)	282(69.6)	
Subjective health status	Good	3,605(36.6)	830(23.0)	2,775(77.0)	429.056*** (P＜0.001)
	Moderate	4,488(45.5)	1,359(30.3)	3,129(69.7)	
	Poor	1,766(17.9)	897(50.8)	869(49.2)	
Smoking	Yes	6,233(63.2)	1,929(30.9)	4,304(69.1)	0.983 (0.321)
	No	3,626(36.8)	1,157(31.9)	2,469(68.1)	
Drinking	Yes	4,625(46.9)	1,802(39.0)	2,823(61.0)	237.766*** (P＜0.001)
	No	5,234(53.1)	1,284(24.5)	3,950(75.5)	

**p＜0.01, ***p＜0.001

상이 36.8%였다. 규칙적 운동 여부는 운동을 하는 자가 51.9%, 하지 않는 자가 48.1%이었으며, 걷기 횟수(일/주)는 전혀 걷지 않는 자 17.2%, 3일 22.5%, 4～6일 27.6%, 매일 걷는 자 32.7%였다. BMI는 저 체중 41.5%, 정상 체중 48.0%, 과 체중 6.4%, 비만 4.1%였으며, 주관적 건강 상태는 좋음 36.6%, 중간 45.5%, 나쁨 17.9%였다. 흡연여부는 흡연을 하는 자가 63.2%, 하지 않는 자가 36.8%였으며, 음주여부는 음주를 하는 자가 46.9%, 하지 않는 자가 53.1%였다.

카이제곱검정 결과, 고혈압 유병자와의 비교 집단은 나이, 결혼여부, 교육수준, 규칙적 운동여부, 걷기 횟수(일/주), BMI, 주관적 건강상태, 음주여부에서 통계적으로 유의한 차이가 있었다($p<0.05$).

고혈압 예측요인에 대한 로지스틱 회귀분석

본 연구대상자의 고혈압 예측요인에 대한 로지스틱 회귀분석 결과는 다음 Table 2와 같다. 유의미한 차이를 보인 나이, 결혼여부, 교육수준, 규칙적 운동여부, 걷기

Table 2. Logistic regression analysis results (n=9,859)

Variable	Categories	B	S.E	Odds Ratio	P-value	95% C.I	
						Lower	Upper
Age	≤39 years old	−	−	1	Ref.	−	
	40～59	2.437	0.350	11.443	<0.001***	5.762	22.725
	≥60 years old	4.231	0.353	68.801	<0.001***	34.472	137.317
Marital status	Married	−	−	1	Ref.	−	
	Single	−0.335	0.162	0.715	0.039*	0.521	0.983
	Divorce/Widowhood/Separation	0.320	0.063	1.377	<0.001***	1.217	1.557
Level of education	≤Middle School	−	−	1	Ref.	−	
	High school	−0.332	0.061	0.717	<0.001***	0.636	0.809
	≥College	−0.623	0.079	0.536	<0.001***	0.459	0.627
Regular exercise	Yes	−	−	1	Ref.	−	
	No	−0.089	0.056	0.915	0.110	0.821	1.020
Number of walking days /week	Not walking	−	−	1	Ref.	−	
	3 days	−0.062	0.082	0.940	0.454	0.800	1.105
	4～6 days	−0.066	0.081	0.936	0.415	0.798	1.098
	Every day	−0.164	0.079	0.849	0.038*	0.727	0.991
Body mass index	Underweight	−	−	1	Ref.	−	
	Normal weight	0.637	0.054	1.891	<0.001***	1.700	2.104
	Overweight	1.489	0.111	4.431	<0.001***	3.566	5.506
	Obesity	1.631	0.145	5.109	<0.001***	3.847	6.786
Subjective health status	Good	−	−	1	Ref.	−	
	Moderate	0.322	0.058	1.379	<0.001***	1.231	1.546
	Poor	0.772	0.073	2.163	<0.001***	1.875	2.496
Drinking	Yes	−	−	1	Ref.	−	
	No	−0.098	0.053	0.906	0.065	0.817	1.006
Constant		−4.785	0.367	0.008	<0.001***	−	

−2 Log likelihood	9477.326	Specificity (%)	85.3
Cox & Snell R^2	0.245	Sensitivity (%)	47.9
Nagelkerke R^2	0.345	Classification accuracy (%)	73.6
Chi-square	18.700		

*p<0.05, ***p<0.001

횟수(일/주), BMI, 주관적 건강상태, 음주여부 변수를 로지스틱 회귀분석의 예측요인으로 포함시켰다. 예측요인에서 나이, 결혼여부, 교육수준, 걷기 횟수(일/주), BMI, 주관적 건강상태에서 유의한 것으로 나타났다($p<0.05$).

나이에서 고혈압 유병률은 39세 이하인 대상자에 비해 40~59세 사이($B=2.437$, $p<0.001$)인 대상자는 약 11.443배, 60세 이상($B=4.231$, $p<0.001$)은 68.801배 높은 것으로 나타났다. 결혼여부에서는 기혼인 자에 비해 미혼인 자($B=-0.335$, $p=0.039$)가 유병률이 0.175배 낮았으며, 반면에 이혼/사별/별거($B=0.320$, $p<0.001$)를 경험한 자는 1.377배 높은 것으로 나타났다. 교육수준에서는 중학교 졸업 이하인 자에 비해 고등학교($B=-0.332$, $p<0.001$), 대학 이상($B=-0.623$, $p<0.001$) 졸업자가 각각 0.717배, 0.536배 유병률이 낮아지는 것으로 나타났다. 걷기 횟수(일/주)에서는 전혀 걷지 않은 자에 비해 매일 걷는 자($B=-0.164$, $p=0.038$)가 0.849배 유병률이 낮아지는 것으로 나타났다. BMI에서는 저 체중인 자에 비해 정상체중($B=0.637$, $p<0.001$)은 1.891배, 과체중($B=1.489$, $p<0.001$) 4.431배, 비만($B=1.631$, $p<0.001$) 5.109배 순으로 유병률이 높아지는 것으로 나타났다. 주관적 건강상태에서는 좋음에 비해 중간($B=0.322$, $p<0.001$) 1.379배, 나쁨($B=0.772$, $p<0.001$) 2.163배 순으로 유병률이 높아지는 것으로 나타났다.

또한, 로지스틱 회귀분석의 모형에 대한 특이도는 85.3%, 민감도 47.9%, 전체 분류 정확도 73.6%로 나타났다.

고혈압 예측요인에 대한 의사결정나무 분석

본 연구의 고혈압 예측 요인에 대한 의사결정나무의 결과는 Figure 1, 2에 제시하였다. 전체 나무구조의 최상위 뿌리 마디인 고혈압 여부는 유병률이 32.2%, 비유병률 67.8%로 나타났다. 고혈압 모델에서 최우선으로 관여하는 요인은 나이($x^2=1513.932$, $p<0.001$)였으며, 60세 이상인 경우가 유병률이 48.7%로 증가되었으며,

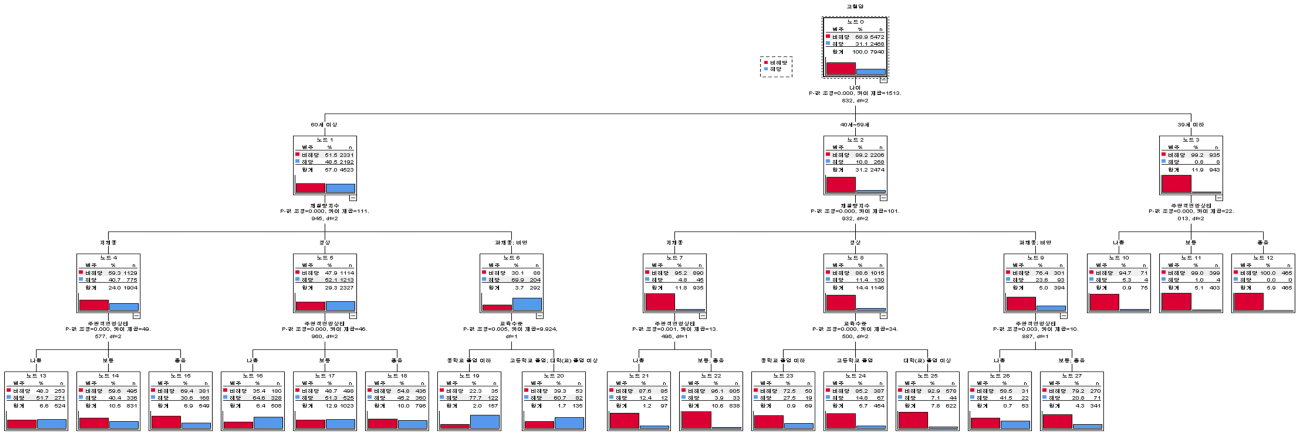

Figure 1. Training-data

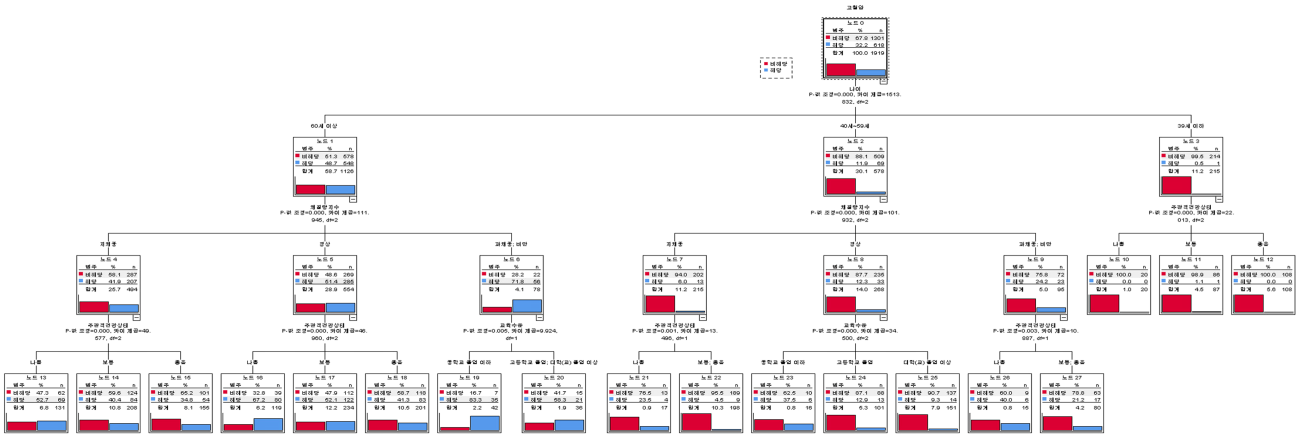

Figure 2. Testing-data

두 번째 분류 변수인 BMI(x^2=111.945, p<0.001)가 과체중이나 비만인 경우 71.8%로 크게 증가되었으며, 세 번째 분류 변수인 교육수준(x^2=9.924, p=0.005)에서 중학교 졸업 이하일 경우 83.3%로 유병률이 더욱 증가하였다. 반면, 비 유병률은 나이가 39세 이하인 경우 99.5%로 크게 증가하였으며, 주관적 건강상태(x^2=22.013, p<0.001)가 좋은 경우 100%로 증가된 것으로 나타났다.

따라서 Table 3의 이익 도표와 마찬가지로 고혈압 유병률에 가장 영향력이 높은 경우는 60세 이상이면서, BMI가 과체중 또는 비만에 해당하며, 교육 수준이 중학교 졸업 이하인 경우로 나타났다. 즉, 19번 노드의 지수(Index)가 258.8%로 해당 노드의 조건을 가진 대상자의 고혈압에 걸릴 확률이 약 2.58 배로 나타났다. 반면에, 비 유병률에 가장 영향력이 높은 경우는 나이가 39세 이하이면서, 주관적 건강상태가 좋은 경우로 나타났다. 즉, 12번 노드의 지수가 147.5%로 해당 노드의 조건을 가진 대상자가 고혈압에 걸리지 않을 확률이 약 1.47 배로 나타났다(Table 4).

의사결정나무의 타당성 평가를 위해 훈련표본과 검정표본을 비교한 결과, 훈련표본의 위험추정값은 0.272, 표준오차 0.005이며, 검정표본의 위험추정값은 0.274, 표준오차 0.010로 모형에 대한 일반화는 무리가 없는 것으로 나타났다. 또한, 의사결정나무의 특이도는 81.9%, 민감도 52.9%, 전체 분류 정확도는 훈련표본에서 72.8%, 검정표본에서 72.6%로 나타났다.

고찰

본 연구는 고혈압에 영향을 주는 요인들을 로지스틱 회귀분석과 의사결정나무 분석을 활용하여 분류하고 예측함으로써 가장 적합한 예측 모형을 선정하여 고혈압 예방과 관리를 위한 기초자료를 제공하고자 시행되었다. 본 연구 주요 연구 결과에 대한 논의는 다음과 같다.

첫째, 로지스틱 회귀분석과 의사결정나무 분석의 고혈압에 대한 주된 공통 요인은 나이로 나타났다. 로지스틱 회귀분석에서 나이는 39세 이하인 대상자에 비해 40~59세 사이인 대상자는 약 11.443배, 60세 이상인 대상자는 68.801배로 크게 고혈압 유병률이 높아지는 것으로 나타났다. 의사결정나무 분석에서도 유병률에 대하여 32.2%이었던 것이 60세 이상인 경우에 48.7%로 증가된 것을 알 수 있었다. 이는 고혈압의 위험요인을

Table 3. Profit index of decision tree analysis related to prevalence of hypertension

Category	Node	Profit index						Category	Node	Profit index					
		Node		Gain		Response	Index			Node		Gain		Response	Index
		N	Percent	N	Percent					N	Percent	N	Percent		
Training data	19	157	2.0%	122	4.9%	77.7%	250.0%	Testing data	19	42	2.2%	35	5.7%	83.3%	258.8%
	16	508	6.4%	328	13.3%	64.6%	207.7%		16	119	6.2%	80	12.9%	67.2%	208.8%
	20	135	1.7%	82	3.3%	60.7%	195.4%		20	36	1.9%	21	3.4%	58.3%	181.1%
	13	524	6.6%	271	11.0%	51.7%	166.4%		13	131	6.8%	69	11.2%	52.7%	163.6%
	17	1023	12.9%	525	21.3%	51.3%	165.1%		17	234	12.2%	122	19.7%	52.1%	161.9%
	18	796	10.0%	360	14.6%	45.2%	145.5%		18	201	10.5%	83	13.4%	41.3%	128.2%
	26	53	0.7%	22	0.9%	41.5%	133.5%		26	15	0.8%	6	1.0%	40.0%	124.2%
	14	831	10.5%	336	13.6%	40.4%	130.1%		14	208	10.8%	84	13.6%	40.4%	125.4%
	15	549	6.9%	168	6.8%	30.6%	98.4%		15	155	8.1%	54	8.7%	34.8%	108.2%
	23	69	0.9%	19	0.8%	27.5%	88.6%		23	16	0.8%	6	1.0%	37.5%	116.4%
	27	341	4.3%	71	2.9%	20.8%	67.0%		27	80	4.2%	17	2.8%	21.3%	66.0%
	24	454	5.7%	67	2.7%	14.8%	47.5%		24	101	5.3%	13	2.1%	12.9%	40.0%
	21	97	1.2%	12	0.5%	12.4%	39.8%		21	17	0.9%	4	0.6%	23.5%	73.1%
	25	622	7.8%	44	1.8%	7.1%	22.8%		25	151	7.9%	14	2.3%	9.3%	28.8%
	10	75	0.9%	4	0.2%	5.3%	17.2%		10	20	1.0%	0	0.0%	0.0%	0.0%
	22	838	10.6%	33	1.3%	3.9%	12.7%		22	198	10.3%	9	1.5%	4.5%	14.1%
	11	403	5.1%	4	0.2%	1.0%	3.2%		11	87	4.5%	1	0.2%	1.1%	3.6%
	12	465	5.9%	0	0.0%	0.0%	0.0%		12	108	5.6%	0	0.0%	0.0%	0.0%

Table 4. Profit index of decision tree analysis related to non-prevalence of hypertension

Category	Node	Profit index						Category	Node	Profit index					
		Node		Gain		Response	Index			Node		Gain		Response	Index
		N	Percent	N	Percent					N	Percent	N	Percent		
	12	465	5.9%	465	8.5%	100.0%	145.1%		12	108	5.6%	108	8.3%	100.0%	147.5%
	11	403	5.1%	399	7.3%	99.0%	143.7%		11	87	4.5%	86	6.6%	98.9%	145.8%
	22	838	10.6%	805	14.7%	96.1%	139.4%		22	198	10.3%	189	14.5%	95.5%	140.8%
	10	75	0.9%	71	1.3%	94.7%	137.4%		10	20	1.0%	20	1.5%	100.0%	147.5%
	25	622	7.8%	578	10.6%	92.9%	134.8%		25	151	7.9%	137	10.5%	90.7%	133.8%
	21	97	1.2%	85	1.6%	87.6%	127.2%		21	17	0.9%	13	1.0%	76.5%	112.8%
	24	454	5.7%	387	7.1%	85.2%	123.7%		24	101	5.3%	88	6.8%	87.1%	128.5%
	27	341	4.3%	270	4.9%	79.2%	114.9%		27	80	4.2%	63	4.8%	78.8%	116.2%
Training	23	69	0.9%	50	0.9%	72.5%	105.1%	Testing	23	16	0.8%	10	0.8%	62.5%	92.2%
data	15	549	6.9%	381	7.0%	69.4%	100.7%	data	15	155	8.1%	101	7.8%	65.2%	96.1%
	14	831	10.5%	495	9.0%	59.6%	86.4%		14	208	10.8%	124	9.5%	59.6%	87.9%
	26	53	0.7%	31	0.6%	58.5%	84.9%		26	15	0.8%	9	0.7%	60.0%	88.5%
	18	796	10.0%	436	8.0%	54.8%	79.5%		18	201	10.5%	118	9.1%	58.7%	86.6%
	17	1023	12.9%	498	9.1%	48.7%	70.6%		17	234	12.2%	112	8.6%	47.9%	70.6%
	13	524	6.6%	253	4.6%	48.3%	70.1%		13	131	6.8%	62	4.8%	47.3%	69.8%
	20	135	1.7%	53	1.0%	39.3%	57.0%		20	36	1.9%	15	1.2%	41.7%	61.5%
	16	508	6.4%	180	3.3%	35.4%	51.4%		16	119	6.2%	39	3.0%	32.8%	48.3%
	19	157	2.0%	35	0.6%	22.3%	32.3%		19	42	2.2%	7	0.5%	16.7%	24.6%

빅데이터 기반의 기계학습 분류 중 하나인 Permutation feature importance(PFI)의 방법으로 분석한 선행연구에서도 나이가 고혈압 발병의 가장 주요한 위험인자인 것으로 나타났다[20]. 나이가 들어감에 따라 혈관의 탄력성이 저하되어 혈압이 서서히 높아지게 되며, 고혈압으로 인해 잔여로 남는 평생 위험은 55세에서 65세에서 약 90%까지 높아지는 것으로 추정되기 때문에[21], 나이가 증가할수록 혈압에 대한 예방과 관리가 필수적이라 할 수 있다. 또한, 고혈압은 하나의 원인에서만 유발되는 것이 아니라 다양한 원인이 모여 고혈압을 일으키게 되므로 혈압의 예방이나 관리 차원에서 나이만 고려하는 것이 아닌 타 위험 요인을 줄이는 것도 주요한 과제라 볼 수 있다.

둘째, 로지스틱 회귀분석에 따르면 의사결정나무 분석 수준과 다르게 결혼 여부와 걷기 횟수(일/주)에서도 유의한 차이를 보였다. 기혼에 비하여 미혼이 고혈압의 발병 확률이 0.175배 낮고, 이혼/사별/별거의 경험을 한 대상자는 기혼에 비하여 1.377배 높았다. 고혈압 환자의 생활 패턴을 분석한 연구에 따르면 기혼자가 이혼 및 사별한 자에 비하여 고 위험군에 속할 확률이 7배 낮아진다고 보고 하여 본 연구와 일치하였다[22]. 이와 유사한 결과로 고혈압 환자의 삶의 질은 배우자 유무에 영향을 받는 것으로 나타났다[23]. 또한, 재가 고혈압 환자를 대상으로 한 연구에서는 미혼인 자와 기혼인 자가 배우자가 없는 경우보다 고혈압 질환 관련 지식이 높음에 있어서 가족 구조가 안정적인 상태에 있는 경우와 심리적으로 불안한 것이 하나의 원인이 될 수 있다고 하였다[24]. 또한 같은 1인 가구임에도 미혼인 자가 이혼/사별/별거한 자보다 고혈압 확률이 낮은 이유는 아마도 미혼인 1인 가구보다 이혼이나 별거를 경험한 1인 가구가 고혈압과 관련성이 높은 우울을 더 심하게 경험하여[25, 26], 혈압에 대한 악영향이 발생할 가능성이 높다고 판단된다.

또한, 본 연구는 미혼인 자가 기혼인 자보다 고혈압이 발생할 확률이 낮은 것으로 나타났다. 본 연구와 일치하는 선행연구에서는 미혼 여성은 기혼 여성에 비해 고혈압 위험이 낮았으며[27, 28], 반면에 남성은 미혼 남성이 혈압 위험이 높았다[29]. 따라서 결혼여부 요인은 국가와 성별에 따라 각 결과의 차이가 있을 수 있을 수 있으므로 추가적인 연구가 이루어져야 할 것이다. 더

욱이 본 연구의 데이터가 단편적인 2차 조사 자료이기 때문에 사회적 관계와 같은 사회문화적 변수를 고려할 수 없었으며, 결혼 상태의 변화에 따라서는 분석이 이루어지지 않아 일반화하기 어려운 경향이 있었다. 그러므로 향후에 결혼 상태의 변화, 즉, 기혼에서 이혼과 같이 변화하는 경우에 대하여 고혈압 유병률이 달라지는가를 파악하는 코호트 연구를 진행하여야 할 것이다.

또한, 걷기 횟수(일/주)에서는 전혀 걷지 않은 자에 비해 매일 걷는 자가 고혈압 확률이 0.849배 낮은 것으로 나타났다. 반면에 규칙적인 운동은 고혈압 환자에게 다양한 이점을 시사함에도 불구하고[30], 걷기를 포함하는 규칙적인 운동 요인에서는 유의한 차이를 보이지 않았다. 이는 아마도 최근 1년간의 규칙적 운동 시행 여부에 대한 설문으로 진행되어 이미 고혈압이 발병된 대상자를 포함하여 분석이 이루어졌기 때문에 고혈압 진단 이후에 건강 관리를 위해 더욱 규칙적인 운동을 시행한 것일 수 있다. 또한, 운동에 대한 내용, 근력 훈련 및 유산소 훈련에 대한 구체적인 언급이 없었다. 즉, 최근 1년 동안 걷기를 포함한 운동에 대한 포괄적인 응답으로써 실제적인 고혈압 환자의 혈압 증가 감소를 파악하기는 어렵다. 그럼에도 구체적인 걷기 횟수(일/주)는 고혈압의 발병 확률에 영향을 미치는 것으로 나타났다. 본 연구와 마찬가지로 앞선 연구에서는 고혈압과 관련하여 유산소 운동은 고혈압을 예방하거나 고혈압이 발병된 자의 혈압을 낮추는 데 효과적인 것으로 입증되었으며, 고혈압인 자에게 빈도 높은 유산소 운동이 권장되어야 한다고 하였다[31]. 또한, 10분 이상 중증도에서 격렬한 신체활동에 참여한 대상자가 고혈압이 있을 가능성이 낮았다고 보고되었다[7]. 따라서 매일 10분 이상의 유산소 운동을 포함한 다양한 운동의 적용은 고혈압 관리와 예방에 긍정적 영향을 미칠 것이라고 생각한다.

셋째, 흡연과 음주여부는 심혈관계의 강력한 위험인자로 뇌졸중, 심근경색 등을 포함한 다양한 질환 예방을 위해 생활습관 개선 방안으로 금연과 금주를 권고한다[32, 33]. 그러나 본 연구의 결과, 로지스틱 회귀분석과 의사결정나무 분석 모두 흡연과 음주는 유의한 결과를 보이지 않았다. 선행 연구에서는 고위험 음주는 고위험 음주를 하지 않는 경우 보다 고혈압 유병률이 8% 증가하는 것으로 보고 되었으며[34], 흡연은 심혈관계 위험을 증가시키는 것으로 알려졌다[33]. 그러나 음주와 흡연 여부는 고혈압 발생과 관련성은 없으나[35, 36], 음주, 흡연의 빈도와 그 양에 따라 고혈압 유병률이 높다는 연구[37, 38]가 주를 이루고 있어 음주와 흡연 여부를 분석한 본 연구와 일치하는 경향을 보였다. 그럼에도 불구하고 음주와 흡연은 기타 심혈관계와 건강에 악영향을 미친다는 연구[39, 40]가 일관성 있게 보고되고 있으므로 금주와 금연은 심혈관계 질환을 유발하는 고혈압 예방을 비롯한 건강관리 측면에서 중요하다 할 수 있겠다.

넷째, 교육수준에서는 로지스틱 회귀분석의 결과, 중학교 졸업 이하인 자에 비해 고등학교 졸업자가 0.717배, 대학 이상 졸업자가 0.536배 유병률이 낮은 것으로 나타났다. 또한, 의사결정나무 분석에서는 60세 이상이면서 과체중이나 비만의 조건을 가진 대상자의 유병률인 71.8%에서 중학교 졸업 이하의 조건을 추가로 가진 대상자의 유병률이 83.3%로 증가하여 가장 고혈압의 유병률에 영향이 높은 경우로 나타났다. 교육수준이 낮은 고혈압 환자는 대사증후군과 심혈관 위험의 증가 및 항고혈압제 사용량이 많은 것으로 보고되고 있으며[41], 교육수준과 고혈압의 상당한 연관성과 함께 교육 수준이 낮을수록 고혈압 유병률이 증가한다고 하였다[34]. 또한, 중학교 졸업 이하에서 고혈압 유병률을 비롯하여 나이, 비만 등이 함께 유의하게 증가되는 것으로 나타나 본 연구 결과와 매우 유사하였다[34]. 이는 교육을 많이 받을수록 질병에 대한 관심도나 건강 정보 이해 능력이 증가되며[42], 이에 관련 정보 습득이 쉬워지거나 질병에 대한 높은 경각심을 갖게 되기 때문인 것으로 생각된다. 따라서 고혈압의 예방 및 관리 방면에서 교육수준을 적절히 고려하고 적용해야 할 필요성이 있다.

다섯째, BMI의 결과로 로지스틱 회귀분석은 저 체중인 대상자에 비해서 정상 체중은 1.891배, 과체중은 4.431배, 비만은 5.190배 순으로 체중이 증가할수록 고혈압 유병률이 증가되는 것으로 나타났다. 의사결정나무 분석에서는 60세 이상인 조건의 대상자에 비해 과체중이나 비만이 추가적인 조건으로 붙는 대상자의 경우 48.7%에서 71.8%로 크게 유병률이 증가되었다. 선행연구에서는 고혈압 환자의 74.4%가 BMI의 기준에 따라 과체중 이상으로 나타났으며[43], 나이가 들고 BMI가 높을수록 고혈압 유병률이 증가하는 결과와 높은 BMI를 주요 고혈압 위험 예측 인자 중 하나로 제시하고 있어 본 연구와 일치하였다[44]. 또한, 국내의 전국 규모 데이터를 활용한 비만 관련 고혈압 연구에서는 비만 매개 변수와 관계없이 비만 자체가 고혈압과 명확하게 연관될 수 있다고 하였다[45]. 즉, 과체중 및 비만은 고혈압의 현저한 유병률과 관련이 있기 때문에[46], 적절한 체중을 유지하기 위한 노력이 필요하며 앞서 설명한 유산소운동을 포함한 신체활동 및 식이요법을 적용해야 한다고 생각된다.

여섯째, 로지스틱 회귀분석의 결과에서 주관적 건강상태의 경우, 좋음에 비해 중간 1.379배, 나쁨 2.163배 순으로 고혈압 유병률이 높아지는 것을 확인할 수 있었다. 의사결정나무 분석에서는 고혈압 비유병률이 나이

가 39세 이하이면서 주관적 건강상태가 좋음인 경우 100%인 것으로 나타났다. 주관적 건강상태는 실제 건강상태의 변량을 상당한 정도로 반영하는 변수로써[47], 고혈압 관리에 영향 요인을 분석한 연구에서는 자신의 건강 상태를 중간이나 나쁨으로 평가할수록 고혈압 인지 가능성이 높아진다고 하였다[48]. 또한, 성인의 고혈압 위험요인을 파악한 연구에서 주관적 건강상태가 좋음이라고 답한 경우에 비해 나쁨이라고 답한 경우에서 고혈압 위험이 2배 이상 증가하였다고 보고 하여 본 연구와 일치하는 것을 볼 수 있었다[49]. 주관적 건강상태는 개인 스스로의 통합적인 건강상태를 포괄하는 개념으로[50], 고혈압 예방과 관리 차원에서 주관적 건강상태를 향상시킬 수 있는 다양한 방안이 고려되어야 할 것이다.

일곱째, 고혈압 유병률에 대한 예측력을 로지스틱 회귀분석과 의사결정나무 분석을 통해 비교 평가한다면, 유병 확률을 예측하는 민감도는 로지스틱 회귀분석 47.9%, 의사결정나무 분석 52.9%로 의사결정나무 분석이 더 높게 나타났으며, 특이도에서는 로지스틱 회귀분석 85.3%, 의사결정나무 분석 81.9%로 로지스틱 회귀분석이 더 높게 나타났다. 분류정확도에서는 로지스틱 회귀분석 73.6%, 의사결정나무 분석이 72.6%로 로지스틱 회귀분석이 더 높게 나타났다. 즉, 고혈압 유병 확률이 있다고 분류한 대상자를 유병될 것이라고 예측하는 민감도에서는 의사결정나무 분석이 더 높았지만, 특이도와 분류 정확도는 로지스틱 회귀분석이 더 높은 것을 확인할 수 있었다. 따라서 두 분석은 정확도에서 다소 큰 차이를 보이지 않아, 로지스틱 회귀분석과 의사결정나무 분석 모두 고혈압 유병률에 대한 예측모형을 구축하는데 유용한 자료로 사용될 것으로 생각된다.

기계학습의 알고리즘은 다양한 통계적, 확률적 및 최적화 방법을 활용하여 대규모의 데이터에서 유용한 패턴을 감지한다[51]. 본 연구의 로지스틱 회귀분석,의사결정나무를 비롯하여 랜덤 포레스트, Extreme Gradient Boosting Decision Tree (XGBoost), Support Vector Machine (SVM) 등의 다양한 알고리즘 방법이 공학,의학분야에서 활용되고 있다[52]. 기계학습의 원활한 적용력을 위해서는 데이터의 크기와 적합한 알고리즘 선택에 따른 최적의 방법을 찾는 것을 우선으로 하기 때문에, 여러 알고리즘을 비교하여 해당 연구에 대한 가장 적합한 알고리즘을 선택해야 한다[52]. 이러한 기계학습의 알고리즘을 비교하고 분석하는 연구는 고혈압 관련 연구에서도 다양하게 시행되고 있다. 나이와 BMI가 고혈압 유병률에 주요 위험요인이라는 본 연구의 결과와 일치한 고혈압 위험 예측 모델 비교 연구[9]에서는 여러 알고리즘을 활용하여 랜덤 포레스트가 성능이 가장 우수한 것이라고 보고하였다. 또한, 고혈압 성인을 대상으로 한 기계 학습 기반의 뇌졸중 예측 모델에서 또한 다른 알고리즘보다는 랜덤 포레스트가 효과적인 예측 모델을 구축했다고 하였다[53]. 심혈관 관련 연구에서는 의사결정나무의 분류 성능이 우수하다고 하였으며[54], 당뇨병과 고혈압 입원 환자의 입원 기간과 사망률을 예측하는 기계 학습 모델 비교 연구[19]에서 입원 기간 예측은 나무 기반의 XGBoost, 사망률 예측은 로지스틱 회귀분석이 가장 좋은 성능을 보였다고 하였다. 다른 맞춤형 고혈압 사후관리 모형 개발 연구[55]에서는 진료 예측에서 로지스틱 회귀분석이 가장 우수한 모형으로 채택되었으며, 고혈압 진료 순응도 세분화 모형은 의사결정나무 모형을 통해 개발되었다.

이와 같이 고혈압 관련 기계학습 연구가 다양하게 이루어진 것을 알 수 있었으며, 최적의 예측 모형 채택에서 서로 다른 결과를 도출한 것을 알 수 있었다. 예를 들어본 연구에서 사용된 의사결정나무는 의학분야에서 가장 많이 사용되는 알고리즘으로써[52], 준거변수에 대한 설명력이 높은 예측변수가 가지를 뻗어 나가도록 하는 기법이다[56]. 의사결정나무는 연속형, 범주형 변수 모두 예측변수로 선택이 가능하며, 결과를 그래프로 제시하여 모형을 직관적으로 쉽게 이해할 수 있다는 장점이 있다. 또 다른 알고리즘인 랜덤 포레스트는 질병 예측에 타 알고리즘에 비해 비교적 우수한 정확도를 보였으며[57], 다수의 표본 생성을 기반으로 의사결정나무에 적용하여 그 결과를 종합하는 방법으로 여러 의사결정나무를 결합시킨 형태이다[58]. 랜덤 포레스트의 장점은 비교적 의사결정나무보다 예측력이 뛰어나고 모형 변경에 대한 위험을 개선한다고 보고 되었다[59]. 즉, 본 연구의 데이터를 랜덤포레스트 알고리즘으로 분석한다면, 또 다른 결과와 적합도를 도출할 수도 있을 것이다. 따라서 향후 관련 연구에서는 로지스틱 회귀분석과 의사결정나무 분석과 함께 더욱 다양한 기계학습 알고리즘을 활용한 연구가 진행되어야 할 것이라고 생각된다. 또한 물리치료 분야의 주요 재활 대상자이면서 본 연구의 고혈압과 함께 주요 만성 질환인 뇌졸중 위험도 예측부터[60] 당뇨병 위험인자 예측까지[61] 기계 학습을 바탕으로 하는 질병에 대한 위험인자 파악은 물리치료 분야에서도 예방과 관리 목적에서 관심을 두고 다양한 질환에 접근하여 다양하게 적용이 가능할 것으로 사료된다. 관련 질환에 대한 예방을 비롯하여 이미 질환이 진행된 상황에서 환자에 대한 교육과 운동 종류 및 기간, 건강상태와 관련하여 맞춤형 치료를 제공하기 위해 다른 보건분야 뿐만 아니라 물리치료 분야에서도 기계학습을 활용한 예측변수 파악 연구가 고려되어야 할 것이다. 아울러 국외의 활발한 기계학습 알고리즘 기반의 고혈압

을 포함한 다양한 만성 질환관련 연구와 달리, 국내의 연구는 다소 미비한 실정이므로 본 연구가 향후 대규모 데이터를 활용한 국내 관련 연구의 기초자료로 제공될 수 있을 것이라고 기대한다.

본 연구는 다음과 같은 제한점을 가진다. 첫째, 대상자의 데이터가 각 변수 별로 차이가 크기 때문에 이러한 영향을 온전히 배제할 수 없었으며 본 연구의 결과를 일반화하기에 무리가 있다. 둘째, 다른 알고리즘의 예측 모형과의 성능 평가와 비교는 이루어지지 않았기 때문에 향후 추가적 외부 검증 및 다른 기계 학습 모델을 기반으로 한 연구가 이루어져야 할 것이다. 셋째, 대상자의 특성과 관련하여 기관에서 제공되는 패널 데이터 자료를 활용함으로써 고혈압 위험인자의 모든 영향요인을 충분히 활용하지 못하였다. 그러나 이러한 제한점에도 불구하고, 본 연구는 우리나라 인구를 대상으로 하는 한국 의료 패널의 대규모 데이터를 분석하여 고혈압에 영향을 미칠 수 있는 다양한 요인들을 분석했다는 점에서 의의가 있다고 판단된다.

결론

본 연구는 주요 관리 대상인 고혈압 발병 확률에 대한 모형을 구축함으로써 향후 고혈압 예방 및 관리 체계에 대한 기초자료를 제공할 수 있다는 점에서 의의를 가진다. 또한, 로지스틱 회귀분석과 의사결정나무를 활용하여 전국 규모의 데이터를 분석한 연구로써 고혈압의 유병률과 비 유병률을 개인적 특성에 따라 파악할 수 있었다. 이를 바탕으로 로지스틱 회귀분석과 의사결정나무에서 구축된 모형은 고혈압 발병 확률 예측에 유용할 것이며, 다양한 분야에서 고혈압 관련 연구에 대한 기반 정보로 활용될 수 있을 것이다. 따라서 추후 연구에서도 더욱 다양한 기계학습 알고리즘을 활용하여 지속적인 고혈압 예방 및 관리 시스템 구축에 기여하여야 할 것으로 생각된다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1F1A1067604).

이해 충돌

본 연구의 저자들은 연구, 저작권 및 출판과 관련하여 잠재적인 이해충돌이 없음을 선언합니다.

참고문헌

1. Doyle AE. Hypertension and vascular disease. Am J Hypertens. 1991;4(2):103-106.
2. Samadian F, Dalili N, Jamalian A. Lifestyle Modifications to Prevent and Control Hypertension. Iran J Kidney Dis. 2016;10(5):237-63.
3. Wu L, Yang S, He Y, Liu M, Wang Y, Wang J, et al. Association between passive smoking and hypertension in Chinese non-smoking elderly women. Hypertens Res. 2017;40(4):399-404.
4. Wang Y, Yao Y, Chen Y, Zhou J, Wu Y, Fu C, et al. Association between Drinking Patterns and Incident Hypertension in Southwest China. Int J Res Public Health. 2022;19(7):3801
5. Kulkarni S, O'Farrell I, Erasi M, Kochar MS. Stress and hypertension. WMJ. 1998;97(11):34-8.
6. Byeon HW, Cho SH. The Predictive Modeling of Middle-aged Hypertension using Integrated Method of Decision Tree and Neural Network. AJMAHS. 2015;5(2):9-18.
7. ou Y, Teng W, Wang J, Ma G, Ma A, Wang J, et al. Hypertension and physical activity in middle-aged and older adults in China. Sci Rep. 2018;8(1):16098.
8. Kim HS, Jung SH, Park SK. Decision-Tree Analysis to Predict Blood Pressure Control Status among Hypertension Patients Taking Antihypertensive Medications. J Korean Biol Nurs Sci. 2019;21(1):85-97.
9. Zhao H, Zhang X, Xu Y, Gao L, Ma Z, Sun Y, et al. Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. Front Public Health. 2021;9:619429.
10. Anh DT, Takakura H, Asai M, Ueda N, Shojaku H. Application of machine learning in the diagnosis of vestibular disease. Sci Rep. 2022;12(1):20805.
11. Dritsas E, Trigka M. Machine Learning Methods for Hypercholesterolemia Long-Term Risk Prediction. Sensors (Basel). 2022;22(14):5365.
12. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. Stroke. 2019;50(5):1263-5.
13. Lee YW, Choi JW, Shin EH. Machine learning

model for predicting malaria using clinical information. Comput Biol Med. 2021;129:104151.
14. Kwon TW, Koo YH. Comparative Analysis of Prediction Taekwondo Trainee's Defection using Decision Tree and Logistic Regression. KSSS. 2008;17(2):71-83.
15. Park MH, Choi SR, Shin AM, Koo CH. Analysis of the Characteristics of the Older Adults with Depression Using Data Mining Decision Tree Analysis. J Korean Acad Nurs. 2013;43(1):1-10.
16. Choi JH, Seo DS. Decision Trees and Its Applications. JKOS. 1999;4(1):61-83.
17. Open Data. Korea Health Panel Annual Data 2019 [Internet]. Sejong and Won ju: Korea Institute for Health and Social Affiairs and National Health Insurance Service; 2020 [cited 2023 Feburary 4]. Available from: https://www.khp.re.kr:444/web/notice/board/view.do?bbsid=53&seq=2832
18. Rhee EJ. Current status of obesity treatment in Korea: based on the 2020 Korean Society for the Study of Obesity guidelines for obesity management. J Korean Med Assoc. 2022;65(7):388-92.
19. Barsasella D, Bah K, Mishra P, Uddin M, Dhar E, Suryani DL, et al. A Machine Learning Model to Predict Length of Stay and Mortality among Diabetes and Hypertension Inpatients. Medicina (Kaunas). 2022;58(11):1568.
20. Oh TS, Kim DK, Won CW, Kim SY, Jeong EJ, Yang JS, et al. A Machine-Learning-Based Risk Factor Analysis for Hypertension: Korea National Health and Nutrition Examination Survey 2016–2019. KJFP. 2022;12(3):173-8.
21. Vasan RS, Beiser A, Seshadri S, Larson MG, Kannel WB, D'Agostino RB, et al. Residual lifetime risk for developing hypertension in middle-aged women and men: The Framingham Heart Study. JAMA. 2002;287(8):1003-10.
22. Ghanbari J, Mohammadpoorasl A, Jahangiry L, Farhangi MA, Amirzadeh J, Ponnet K. Subgroups of lifestyle patterns among hypertension patients: a latent-class analysis. BMC Med Res methodol. 2018;18(1):127.
23. Kim SI, Woo SJ, Jung YH. Factors Related to Hypertension Patients' Quality of Life: The 7th Korean National Health and Nutrition Examination (1st Year, 2016). Leisure Activity Types and Depressive Symptoms among Middle-Aged People Living Alone JKSSCHE. 2020;21(1):61-74.
24. Kwon MS, Noh GY, Jang JH. A study on relationships between health literacy, disease-related knowledge and compliance to medical recommendations in patients with hypertension. J Korean Pubilc Health Nurs. 2013;27(1):190-202.
25. Kang EN, Kim HJ, Kim YS. Leisure Activity Types and Depressive Symptoms among Middle-Aged People Living Alone. HSWR. 2017;37(2):184-215.
26. Jeon DJ, Kim SH, Park SH, Yoon HJ, Kim SG, Kim JH. The Prevalence and Psychosocial Correlates of Depressive Symptoms in Patients with Hypertension. J Korean Soc Biol Ther Psychiatry. 2019;25(3):213-21.
27. Ramezankhani A, Azizi F, Hadaegh F. Associations of marital status with diabetes, hypertension, cardiovascular disease and all-cause mortality: a long term follow-up study. PLoS One. 2019;14(4): e0215593.
28. Tuoyire DA, Ayetey H. Gender differences in the association between marital status and hypertension in Ghana. J Biosoc Sci. 2019;51(3):313-34.
29. Lipowicz A, Lopuszanska M. Marital differences in blood pressure and the risk of hypertension among Polish men. Eur J epidemiol. 2005;20(5):421-7.
30. Boutcher YN, Boutcher SH. Exercise intensity and hypertension: what's new? Journal of human hypertension. J Hum Hypertens. 2017;31(3):157-64.
31. Ruivo JA, Alcântara P. Hypertension and exercise. Rev Port Cardiol. 2012;31(2):151-8.
32. Day E, Rudd JHF. Alcohol use disorders and the heart. Addiction. 2019;114(9):1670-8.
33. Sleight P. Smoking and hypertension. Clin Exp Hypertens. 1993;15(6):1181-92.
34. Choi JH, Park JH, Choi BG. Association between Education Level and Hypertension in Korean Adults Over 30 Years Old: Korea National Health and Nutrition Examination Survey 2019. KJFP. 2022;12(4):247-53.
35. Sohn K. Relationship of smoking to hypertension in a developing country. Glob Heart. 2018;13(4):285-92.
36. Lee HS, Kwun IS, Kwon CS. Prevalence of Hypertension and Related Risk Factors of the Older Residents in Andong Rural Area. JKSFSN. 2009; 38(7):852-61.

37. Eom JS, Lee TR, Park SJ, Ahn YJ, Chung YJ. The risk factors of the pre-hypertension and hypertension of rural inhabitants in Chungnam-do. J Nutr Health. 2008;41(8):742-53.
38. Lee HJ, Lee HS, Lee YN, Jang YA, Moon JJ, Kim CI. Nutritional environment influences hypertension in the middle-aged Korean adults-Based on 1998 & 2001 National Health and Nutrition Survey. Korean J Community Nutr. 2007;12(3):272-83.
39. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. Journal of the American College of Cardiology. J Am Coll Cardiol. 2004;43(10):1731-7.
40. Biddinger KJ, Emdin CA, Haas ME, Wang M, Hindy G, Ellinor PT, et al. Association of Habitual Alcohol Intake With Risk of Cardiovascular Disease. JAMA netw open. 2022;5(3):e223849.
41. Di Chiara T, Scaglione A, Corrao S, Argano C, Pinto A, Scaglione R. Education and hypertension: impact on global cardiovascular risk. Acta cardiol. 2017;72(5):507-13.
42. Kim JE. Measuring the Level of Health Literacy and Influence Factors: Targeting the Visitors of a University Hospital's Outpatient Clinic. J Korean Clin Nurs Res. 2011;17(1):40-7.
43. Diaz ME. Hypertension and obesity. J Hum Hypertens. 2002;16(1):18-22.
44. Rahman M, Zaman MM, Islam JY, Chowdhury J, Ahsan HN, Rahman R, et al. Prevalence, treatment patterns, and risk factors of hypertension and pre-hypertension among Bangladeshi adults. Journal of human hypertension. J Hum Hypertens. 2018; 32(5):334-48.
45. Lee SH, Park YM, Han KD, Yang JH, Lee SW, Lee SS, et al. Obesity-related hypertension: Findings from The Korea National Health and Nutrition Examination Survey 2008-2010. PLoS one. 2020;15(4):e0230616.
46. Leenen FH, McInnis NH, Fodor G. Obesity and the prevalence and management of hypertension in Ontario, Canada. Am J Hypertens. 2010;23(9):1000-6.
47. Choi YH. Is self-rated Health a Sufficient Proxy for True Health? KJGSW. 2018;73(4):7-28.
48. Chang DM, Park IS, Yang JH. Related Factors of Awareness, Treatment, and Control of Hypertension in Korea : Using the Fourth Korea National Health & Nutrition Examination Survey. J Digit Converge. 2013;3(5):6-7.
49. Kim KY. Risk factors for hypertension in elderly people aged 65 and over, and adults under age 65. JKAIS. 2019;20(1):162-9.
50. Yi YM, Park YH. Factors Related to Subjective Health Status in Community-Dwelling Older Adults Living Alone on Low Income. The J of Mus and Joint Health. 2022;29(3):205-17.
51. Mitchell TM. Machine learning. 1st ed. New York: McGraw Hill; 2007.
52. Lee HH, Chung SH, Choi EJ. A case study on machine learning applications and performance improvement in learning algorithm. J Digit Converge. 2016;14(2): 245-258.
53. Huang X, Cao T, Chen L, Li J, Tan Z, Xu B, et al. Novel Insights on Establishing Machine Learning-Based Stroke Prediction Models Among Hypertensive Adults. Front Cardiovasc Med. 2022;9:901240.
54. Shah W, Aleem M, Iqbal MA, Islam MA, Ahmed U, Srivastava G, et al. A Machine-Learning-Based System for Prediction of Cardiovascular and Chronic Respiratory Diseases. J Healthc Eng. 2021;2021: 2621655.
55. Park IS, Yong WS, Kim YM, Kang SH, Han JT. A development of a tailored follow up management model using the data mining technique on hypertension. KSS. 2008;21(4):639-47.
56. Yoo JE. Random forests, an alternative data mining technique to decision tree. J Educ Evaluation. 2015;28(2):427-448.
57. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC. 2019;19(1):281.
58. Seo JD. Foreign Exchange Rate Forecasting Using the GARCH extended Random Forest Model. JIEB. 2016;29(5):1607-1628.
59. Breiman L. Random forests. Machine learning. 2001;45:5-32.
60. Jeong SW, Lee MJ, Yoo SY. Machine Learning-based Stroke Risk Prediction using Public Big Data. J AdvNaving Technol. 2021;25(1):96-101.
61. Hyun JK. Prediction of Diabetic Neuropathy Using Machine Learning Techniques. JKD. 2022;23(4): 338-244