

# 성별에 따른 대사증후군의 위험요인 탐색을 위한 융복합 연구

이소은<sup>1</sup>, 이현실<sup>2\*</sup>

<sup>1</sup>고려대학교 대학원 보건과학과 석사과정, <sup>2</sup>고려대학교 보건과학대학 보건정책관리학부 교수

## Convergence study to detect metabolic syndrome risk factors by gender difference

So-Eun Lee<sup>1</sup>, Hyun-Sill Rhee<sup>2\*</sup>

<sup>1</sup>Master's Course, Department of Health Sciences, Graduate School, Korea University

<sup>2</sup>Professor, School of Health Policy and Management, College of Health Science, Korea University

요 약 본 연구의 목적은 국민건강영양조사 2016-2019년 자료 중 성인을 대상으로 대사증후군의 위험요인 탐색하고, 성별에 따른 위험요인의 차이를 규명하여 대사증후군 예방 및 치료에 기초자료로 제공하기 위함이다. 다양한 선행연구를 통해 대사증후군 위험요인을 수집하고, 4개의 머신러닝(Logistic Regression, Decision Tree, Naïve Bayes, Random Forest)의 방법을 이용하여 분석하였다. 남성과 여성 모두에서 Random Forest의 대사증후군 예측 정확도가 높았다. 대사증후군 유병에 영향을 주는 상위 위험요인으로는 여성과 남성 모두에서 BMI, 식이(지방, 비타민 C, 비타민 A, 단백질, 에너지 섭취), 기저질환의 개수, 연령으로 나타났다. 여성의 경우 교육수준과 초경 연령, 폐경 여부가 추가적으로 주요 위험요인으로 나타났고, 남성에 비해 연령과 기저질환의 개수에서 영향력이 큰 것으로 나타났다. 대사증후군을 예방하기 위해선 BMI, 식이, 질환의 이환, 초경 및 폐경여부를 고려하여 접근해야하며 후속 연구를 통해 다양한 중재 전략을 수립하고 검증해야 할 것이다.

주제어 : 대사증후군, 위험요인, 머신러닝, 예측모델, 랜덤포레스트

Abstract This study was conducted to detect metabolic syndrome risk factors and gender difference in adults. 18,616 cases of adults are collected by Korea Health and Nutrition Examination Study from 2016 to 2019. Using 4 types of machine Learning(Logistic Regression, Decision Tree, Naïve Bayes, Random Forest) to predict Metabolic Syndrome. The results showed that the Random Forest was superior to other methods in men and women. In both of participants, BMI, diet(fat, vitamin C, vitamin A, protein, energy intake), number of underlying chronic disease and age were the upper importance. In women, education level, menarche age, menopause was additional upper importance and age, number of underlying chronic disease were more powerful importance than men. Future study have to verify various strategy to prevent metabolic syndrome.

Key Words : Metabolic Syndrome, Risk factors, Machine learning, Prediction model, Random Forest

\*Corresponding Author : Hyun-Sill Rhee(pridehyun@korea.ac.kr)

Received September 24, 2021

Accepted December 20, 2021

Revised November 1, 2021

Published December 28, 2021

## 1. 서론

대사증후군(Metabolic Syndrome)은 고혈압, 복부비만, 저 고밀도지단백콜레스테롤혈증, 고중성지방혈증, 혈당장애 등의 임상적 진단기준으로 정의된다[1]. 대사증후군은 관상동맥질환의 대표적인 원인 질환이며 제 2형 당뇨병의 위험을 증대시키는 원인으로 알려져있다[2-4]. 급격한 생활환경의 변화와 노화로 인해 대사증후군과 만성질환의 유병이 증가하고 있으며[5], 2016-2018년의 국내 대사증후군 유병률은 19세 이상 성인에서 약 23%, 65세 이상 노인에서 대략 50%의 유병률을 보이며[4] 전세계적으로도 계속해서 증가하고 있는 추세이다[6]. 대사증후군의 위험요인을 복합적으로 가진 사람들이 증가하면서 대사증후군 유병률은 계속 증가할 것으로 예측되며 [5] 대사증후군의 조기발견 및 예방적 차원의 관리가 요구된다.

대사증후군 유병률은 여성보다 남성에서 더 높게 나타나는 등[7] 성별에 따라 차이가 존재하며, 영향을 주는 위험요인이 다르다. 선행연구에 따르면, 여성의 경우 폐경, 조기 초경이 대사증후군 발병 위험도를 높이는 것으로 나타났고[8, 9], 대사증후군 발생에 있어 여성이 남성보다 사회경제적인 요인의 영향이 크게 나타났다[10]. 직업별 대사증후군 위험은 전업주부와 사무직에서 높게 나타나 성별에 따라 다르게 나타남을 보였다[11]. 여성은 교육수준과 소득수준이 대사증후군 위험률에 영향을 주는 반면, 남성은 소득수준 최하위그룹에서만 유의하게 나타났다[12]. 성별에 따라 대사증후군 유병의 위험률과 영향요인이 다르게 나타남에 따라 성별을 구분하여 대사증후군 유병 영향요인을 탐색해볼 필요가 있다.

대사증후군의 위험요인은 다양한 연구를 통해 나타났다. 유산소, 저항운동 등 복합운동이 대사증후군 위험인자를 감소시키고[13] 1인 가구 청년의 아침식사 빈도, 스트레스, 우울 등 식이와 주관적 건강상태가 대사증후군 유병에 영향을 미치고, 다인가구 청년은 음주의 영향이 유의하게 나타났다[14]. 과일과 야채의 섭취가 대사증후군 발병을 낮추고[15, 16], 염증과 폐쇄성 수면 무호흡증, 심폐체력 및 근지구력[17], 악력이 대사증후군의 유병에 영향을 미치는 등[18, 19] 다양한 위험요인이 대사증후군에 미치는 영향을 파악한 연구가 존재했다. 하지만 대부분의 연구는 대사증후군과 특정 위험인자와의 연관성을 보는 연구였으며 다양한 인자를 복합적으로 분석한 연구는 미비했다. 특히 대사증후군의 예측을 위해 머신러닝을 사용한 연구에서 임상 및 유전과 관련된 연구만

진행되었다[20]. 이에 따라 본 연구는 인구사회학적, 식이, 임상 등 다양한 위험인자를 바탕으로 머신러닝 기법을 통해 대사증후군 유병에 영향이 큰 위험요인을 탐색하고, 예측모델 개발에 기초자료를 제공하기 위해 융복합적으로 접근하고자 한다.

본 연구의 목적은 성별에 따른 대사증후군에 영향을 미치는 위험인자를 파악하는 것이다. 선행연구를 바탕으로 알려진 위험인자들을 수집하여 머신러닝을 이용해 대사증후군에 가장 영향을 크게 주는 인자를 파악하고 대사증후군 예측모델 개발 및 예방을 위한 융복합 정책 개발에 기초자료로 제공하고자 한다.

## 2. 연구방법

### 2.1 자료원 및 연구 대상

본 연구에 사용한 자료는 국민건강영양조사 제7기(2016 - 2018년)와 8기(2019년)로, 국민건강영양조사(Korea National Health and Nutrition Examination Survey; KNHANES)는 통계법 제 17조에 근거하여 국민의 건강행태와 만성질환 및 영양섭취에 대한 법정실태조사로 1998년부터 현재까지 시행되고 있다[21].

2016년부터 2019년까지 총 4개연도의 자료를 통합한 뒤 19세 이상의 성인을 대상으로 분석하였다. 데이터는 총 32,379건이었고, 19세 미만, 대사증후군 관련 변수의 결측, 암의 유병, 임신중인 경우, 하루 섭취 칼로리가 500 kcal 이하, 5000kcal 이상, 여성의 여성력 변수가 결측인 경우를 제외한 총 18,616건의 자료를 바탕으로 분석하였다.

### 2.2 대사증후군 진단 변수

대사증후군의 정의는 Modified NCEP - ATP III을 기준으로 사용하였으며[22], 허리둘레의 경우 대한비만학회에서 제시한 한국인 기준을 적용하였다[23].

대사증후군의 진단 기준은 다음과 같다. 1) 복부비만 : 남성 허리둘레  $\geq 90$  cm, 여성 허리둘레  $\geq 85$ cm, 2) 고중성지방혈증 : 중성지방  $\geq 150$  mg/dL 혹은 약물치료중인 경우, 3) 저 HDL 콜레스테롤혈증 : 남성 hdl-콜레스테롤  $< 40$  mg/dL, 여성 hdl-콜레스테롤  $< 50$ mg/dL, 4) 고혈압 : 수축기 혈압  $\geq 130$ mmHg, 이완기 혈압  $\geq 85$  mmHg 혹은 약물치료중인 경우, 5) 고혈당 : 공복혈당  $\geq 110$ mg/dL 혹은 약물치료중인 경우로,

5가지 진단기준 중 3가지 이상에 해당되는 경우에 대사증후군으로 선정하였다.

### 2.3 투입변수

본 연구에서 대사증후군 진단을 위해 투입한 변수는 <Table 1>과 같다. 투입한 변수는 인구사회학적특성, 생활양식, 건강행태, 식이, 여성력으로 구성되어있다. 선행 연구를 바탕으로 수집한 변수들을 바탕으로 '모름, 무응답'을 포함한 결측치가 전체 데이터의 30% 이하인 변수로 선정하였다.

음주는 최근 1년간 월 1잔 이상 음주한 경우를 음주로 나타냈고, 흡연은 매일 피움, 가끔피움을 현재 흡연으로 나타냈다. 신체활동여부는 일주일에 중,고강도 신체활동을 실천한 경우를 실천으로 나타냈다. 스트레스 인지율은

스트레스를 많이 느끼는 경우 스트레스를 인지한다고 정의하였다. 기저질환의 개수는 건강설문조사의 의사진단을 받은 질병의 개수로 나타낸 것으로 0개부터 5개까지로 구성하였으며 5개 이상의 경우엔 5개로 나타냈다. 체질량지수는 신체계측을 통해 수집한 몸무게(kg)를 키의 제곱(m<sup>2</sup>)으로 나눈 값으로 나타냈다. 악력의 경우, 양 손의 3번 측정된 악력을 평균치리한 값을 절대악력(Absolute hand grip strength)으로 정의하고, 절대악력을 체중으로 나눈 값인 상대악력(Relative hand grip strength)을 분석에 투입하였다. 식이의 경우 24시간 회상조사를 통해 하루 동안 섭취한 모든 음식 및 식품의 섭취량의 합으로 나타냈다. 폐경 여부는 자연폐경과 인공폐경 모두 폐경으로 나타냈고, 폐경 연령과 초경 연령은 각 조사 참여자가 기입한 연령을 사용하였다.

Table 1. Input variable descriptions

Category	Variables	Definition	Type
Demographic	Age	19-29, 30-39, 40-49, 50-59, 60-69, 70-	Categorical
	Income level	low, middle-low, middle-high, high	Categorical
	Education level	Under elementary, middle school, high school, Over university	Categorical
	Occupation	Managers, Clerks, Service/sales workers, Agricultural workers, Craft workers, elementary workers, Unemployed	Categorical
	Marital status	Married, Not married	Categorical
	Insurance type	National Health Insurance (Local-subscriber, Employee)	Categorical
Life style	Smoking status	No, Yes	Categorical
	Drinking status	No, Yes	Categorical
	Physical activity	No, Yes	Categorical
	Weekly sleep time		Continuous
	Weekend sleep time		Continuous
Health status	Perceived health	Very good, Good, Normal, Bad, Very bad	Categorical
	Perceived stress	No, Yes	Categorical
	Perceived body shape	Very lean, Little lean, Normal, Little obese, Very obese	Categorical
	Number of underlying diseases	0, 1, 2, 3, 4, ≥ 5	Categorical
	Family history of chronic disease	No, Yes	Categorical
	Body Mass Index		Continuous
	Relative hand grip strength		Continuous
Diet	Energy intake	Kcal	Continuous
	Carbohydrate intake	g	Continuous
	Protein intake	g	Continuous
	Fat intake	g	Continuous
	Vit A intake		Continuous
	Vit C intake		Continuous
Gynecology	Menopause	No, Yes	Categorical
	Menopause age		Continuous
	Menarche age		Continuous

## 2.4 분석 방법

본 연구에서 사용한 머신러닝 방법은 Logistic Regression, Decision Tree, Naïve Bayes, Random Forest를 사용하였고, 자료의 처리와 분석은 SAS 9.4 (SAS Institute, Cary, NC, USA)와 Python (3.9.1)을 사용하였다.

머신러닝(Machine Learning)은 경험(experience)을 통해 데이터를 학습하는 프로그램으로 의료 분야를 포함한 다양한 분야에서 사용되고 있다[24, 25]. 데이터를 기반으로 스스로 자료를 학습하고, 평가하며 모델을 형성하는 방법으로 빅데이터 분석에서 사용되고 있다[26]. 머신러닝 방법은 질병을 예측하고, 분류하는 연구 등 다양하게 활용되고 있다[27]. 국내에서도 머신러닝을 사용하여 알츠하이머, 심혈관질환, 골다공증 등 다양하게 사용되고 있다[28-30].

로지스틱 회귀분석(Logistic Regression)은 선형 회귀분석과 달리 종속변수를 범주형 자료로 사용하여 데이터를 분류하는 분석방법으로 종속변수와 독립변수의 관계를 통해 예측모형을 만들어 사건의 발생 확률을 예측하는 방법이다[31]. 의사결정나무(Decision Tree)는 의사결정규칙에 따라 순차적으로 적용하여 독립변수를 분류하는 분석방법으로 입력 자료에 따라 출력 자료가 어떻게 나오는지 예측하며, 나무 형태의 구조로 시각화되어 종속변수에 미치는 영향의 크기를 이해하기 쉬운 특징을 가진다[24, 32]. 나이브 베이즈(Naïve Bayes)는 베이즈 이론을 기반으로 관측된 특징이 한 클래스에 속할 확률을 예측하는 머신러닝의 지도학습의 한 분석으로 분석하는 시간과 저장 공간에서 매우 효율적인 장점이 존재한다[33, 34]. 랜덤포레스트(Random Forest)는 의사결정나무의 분류방법의 단점을 개선한 데이터 분류 기법으로 학습 및 평가데이터에서 무작위로 선택된 데이터를 통해 의사결정나무가 형성되고, 여러 의사결정나무를 종합하여 숲을 이루며 예측 및 분류의 과정을 거치기 때문에 모형의 성능이 향상되며 분류 정확도가 높은 방법이다[24, 35].

4가지 머신러닝을 통해 예측 모델을 구축한 뒤 AUC(Area Under Curve)와 Accuracy, Precision, Recall, F1-score를 이용하여 모델의 성능 분석을 실시하였다[27]. 모형 검증을 위해 훈련자료(Training set)과 검증 자료(Test set)를 70 : 30 비율로 나누어 모형을 검증하였다. 기초통계분석의 경우 범주형 자료는 빈도(%)로, 연속형 변수는 평균  $\pm$  표준편차로 나타내었다.

## 3. 연구결과

### 3.1 연구대상자의 대사증후군에 따른 특성

Table 2. General characteristics by sex

Variables	Male (N = 8,203)	Female (N = 10,413)
Metabolic Syndrome		
No	5052 (61.59)	6964 (66.88)
Yes	3151 (38.41)	3449 (33.12)
Age		
19 ~ 29	1034 (12.61)	1167 (11.21)
30 ~ 39	1302 (15.87)	1685 (16.18)
40 ~ 49	1468 (17.9)	2064 (19.82)
50 ~ 59	1507 (18.37)	2002 (19.23)
60 ~ 69	1491 (18.18)	1806 (17.34)
70 ~	1401 (17.08)	1689 (16.22)
Income level		
low	1414 (17.24)	2020 (19.40)
middle - low	1991 (24.27)	2592 (24.89)
middle - high	2285 (27.86)	2801 (26.9)
high	2513 (30.64)	3000 (28.81)
Education level		
elementary school and below	1110 (13.53)	2442 (23.45)
Middle school	778 (9.48)	1008 (9.68)
High school	3086 (37.62)	3183 (30.57)
University	3229 (39.36)	3780 (36.3)
Occupation		
Managers, professional, and related workers	1272 (15.51)	1346 (12.93)
Clerks	950 (11.58)	1046 (10.05)
Service/sales workers	770 (9.39)	1555 (14.93)
Agricultural, forestry, and fishery workers	456 (5.56)	283 (2.72)
Craft workers, equipment and machine operation and assembling workers	1959 (23.88)	314 (3.02)
Elementary workers	639 (7.79)	1000 (9.60)
Unemployed	2157 (26.30)	4869 (46.76)
Marital status		
Married	6490 (79.12)	8985 (86.29)
Not Married	1713 (20.88)	1428 (13.71)
Insurance type		
National Health Insurance(Local-subscriber)	2381 (29.03)	2858 (27.45)
National Health Insurance(Employee)	5541 (67.55)	7192 (69.07)
Assistance	281 (3.43)	363 (3.49)

Perceived health		
Very good	464 (5.66)	409 (3.93)
Good	2117 (25.81)	2403 (23.08)
Normal	4392 (53.54)	5514 (52.95)
Bad	1040 (12.68)	1636 (15.71)
Very bad	190 (2.32)	451 (4.33)
Perceived body shape		
Very lean	369 (4.50)	312 (3.00)
Little lean	1152 (14.04)	1007 (9.67)
Normal	3334 (40.64)	4292 (41.22)
Little obese	2806 (34.21)	3669 (35.23)
Very obese	542 (6.61)	1133 (10.88)
Drinking status		
No	2402 (29.28)	6000 (57.62)
Yes	5801 (70.72)	4413 (42.38)
Smoking status		
No	5485 (66.87)	9898 (95.05)
Yes	2718 (33.13)	515 (4.95)
Perceived stress		
No	6279 (76.55)	7473 (71.77)
Yes	1924 (23.45)	2940 (28.23)
Physical activity		
No	4570 (55.71)	6167 (59.22)
Yes	3633 (44.29)	4246 (40.78)
History of chronic disease		
No	3384 (41.25)	3507 (33.68)
Yes	4819 (58.75)	6906 (66.32)
Number of underlying diseases		
0	3527 (43)	3757 (36.08)
1	2099 (25.59)	2422 (23.26)
2	1305 (15.91)	1605 (15.41)
3	709 (8.64)	1063 (10.21)
4	317 (3.86)	699 (6.71)
≥ 5	246 (3.00)	867 (8.33)
Weekly sleep time	298.85 ± 203.11	315.09 ± 196.97
Weekend sleep time	324.01 ± 221.5	342.27 ± 214.35
Relative hand grip strength	1.48 ± 0.34	0.93 ± 0.24
BMI	24.51 ± 3.34	23.48 ± 3.63
Energy intake (Kcal)	2242.24 ± 836.5	1656.91 ± 651.97
Protein intake(g)	80.56 ± 38.58	59.35 ± 29.49
Fat intake(g)	48.84 ± 34.52	37.11 ± 26.81
Vitamin A intake	398.85 ± 420.72	347.45 ± 371.85
Vitamin C intake	65.28 ± 79.71	62.64 ± 81.4
Menopause		
No		5295 (50.85)
Yes		5118 (49.15)
Menopause age		49.72 ± 3.41
Menarche age		13.92 ± 2.03

연구 대상자의 대사증후군 여부에 따른 특성은 <Table 2.>와 같다. 총 18,616건의 자료 중 남성은 8,203명, 여성은 10,413명으로 여성의 빈도가 더 많았고, 대사증후군을 유병하는 자는 총 6,600명으로 나타났다. 성별에 따른 대사증후군 발병률은 남성은 38%, 여성은 33%로 기존 선행연구[11]와 유사하게 남성의 비율이 더 높게 나타났다.

여성의 초졸 이하의 학력을 가진 사람이 23%로 남성의 14%보다 많은 비율을 가지고 있었다. 직업의 경우 남성(26%)과 여성(47%) 모두 무직(주부, 학생)의 비율이 가장 높았다. 주관적 체형인식의 경우 여성에서 약간 비만 혹은 매우 비만으로 인식하는 비율이 46%로 남성의 40%보다 높았다. 70%의 남성이 음주를 하며, 여성은 58%로 남성보다 낮게 나타났다. 흡연에선 남성이 33%, 여성이 5%로 흡연자의 비율이 큰 차이를 보였다.

만성질환 과거력의 경우 여성이 66%, 남성이 59%로 여성에서 크게 나타났으며, 기저질환의 개수는 3개 이상에서 여성의 비율이 높게 나타나 여성에서 질환의 노출에 취약하다고 할 수 있다. 식이의 경우 남성보다 여성에서의 수치가 전반적으로 낮게 나타났다. 여성력의 경우, 폐경의 비율이 49%로 나타났다.

### 3.2 모형의 성능 평가

다양한 연구분야에서 모형의 성능을 평가하는 지표인 AUC(Area Under Curve)를 사용하여 모형의 성능을 평가하였고, 분석 결과는 <Table 3.>와 같다.

Table 3. Confusion matrix of Models

	Model	AUC	Accuracy	Precision	Recall	F1-score
Male	LR	0.68	0.66	0.58	0.42	0.49
	DT	0.624	0.64	0.53	0.54	0.54
	NB	0.785	0.72	0.63	0.65	0.64
	RF	0.807	0.75	0.70	0.59	0.64
Female	LR	0.836	0.77	0.70	0.54	0.61
	DT	0.677	0.71	0.57	0.58	0.57
	NB	0.804	0.74	0.59	0.74	0.65
	RF	0.859	0.79	0.72	0.63	0.67

LR : Logistic Regression, DT : Decision Tree, NB : Naive Bayes, RF : Random Forest

남성의 경우, Random Forest의 모형의 정확도가 80%로 가장 높은 정확도를 보였고, 그 다음으로 Naive bayes, Logistic Regression, Decision Tree 순이었다. 여성의 경우도 Random Forest가 86%로 가장 높은

정확도를 보였고 다음으로 Logistic Regression이 84%로 높았다. 남성과 여성에서 Random Forest가 공통적으로 가장 높은 정확도를 보이는 것으로 나타났다.

정밀도(Precision)은 모형을 통해 양성이라고 출력된 자료 중 실제 데이터가 양성인 경우를 의미하며, 재현율(Recall)은 실제 양성인 자료를 모형에서 양성이라고 출력한 경우를 의미한다. 정밀도의 경우 남성과 여성 모두에서 Random Forest가 가장 높게 나타났고, 재현율은 남성에선 Random Forest, 여성에선 Naive Bayes, Random Forest 순이었다. 정밀도와 재현율의 조화평균으로 나타내는 F1-score을 살펴보면 남성에선 Naive Bayes, Random Forest가 0.64로 가장 높았고, 여성의 경우 Random Forest(0.67), Naive Bayes(0.65), Logistic Regression(0.61) 순으로 나타났다.

### 3.3 대사증후군 유병의 영향요인

대사증후군 유병의 영향요인을 탐색하기 위해 Random Forest의 변수중요도를 추출하였다. 변수중요도를 통해 대사증후군 유병에 미치는 상위 요인을 파악하고 성별에 따라 어떠한 차이가 있는지 탐색하고자 하였다. 남성의 변수중요도는 <Figure 1.>, 여성의 변수중요도는 <Figure 2.>와 같다.

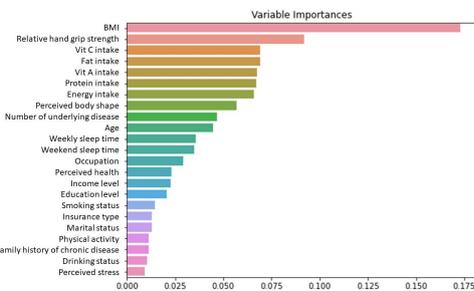


Fig. 1. Variable Importance of Male

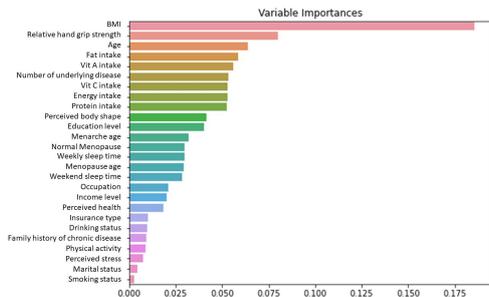


Fig. 2. Variable Importance of Female

<Figure 1.>의 남성의 변수중요도를 살펴보면 BMI, 상대악력, 식이 변수, 주관적 체형인식, 기저질환의 개수, 연령 순으로 나타났고, <Figure 2.>의 여성의 변수중요도를 살펴보면 BMI, 상대악력, 연령, 식이 변수, 기저질환의 개수, 주관적 체형인식, 교육수준, 초경 연령, 폐경 여부 등으로 나타났다.

BMI, 상대악력, 식이, 주관적 체형인식 등의 변수는 성별에 따라 큰 차이가 없이 대사증후군 유병의 상위 위험요인으로 나타났다. 하지만, 여성의 경우 연령의 중요도가 남성과 비교하여 매우 높았고, 기저질환의 개수의 변수도 남성보다 높게 나타나고 있었다. 교육수준의 경우도 남성보다 여성에서 높았으며, 여성력의 변수에선 초경 연령, 폐경 여부 순으로 중요도가 나타났다.

선행연구에서 신체활동이 주요한 대사증후군의 대표적인 위험인자로 나타난 결과에 비해 본 연구에선 남성과 여성 모두에서 신체활동의 중요도가 낮게 나타났다. 이는 신체활동여부보다 식이, 연령 등의 영향이 더 크기 때문에 나타난 결과로, 신체활동여부의 수준을 활동별로 구분하는 등 세부적인 분석이 필요할 것으로 보인다.

흡연과 음주의 경우도 변수중요도에서 중요도가 낮게 나타났다. 이는 남성의 대부분이 음주를 하며, 여성과 남성 모두에서 비흡연자의 비율이 매우 높게 나온 것과 관련하여 나타났다.

## 4. 고찰

본 연구는 국민건강영양조사 자료를 이용하여 성별에 따른 대사증후군 유병에 미치는 위험요인을 탐색하고자 하였다. 4가지 머신러닝 방법을 이용하여 모형의 예측력이 가장 높은 모형을 찾고, 그 모형을 통해 대사증후군 위험요인을 탐색하였다.

성별에 따른 대사증후군 위험인자의 특성을 비교해보았을 때, 남성의 대사증후군 유병 비율이 여성보다 높게 나타났다. 주관적 체형인식에서 여성이 남성보다 약간 비만, 매우 비만의 비율이 46%로 남성의 40%보다 약간 높았고, 기저질환의 개수에서도 여성은 3개(10%), 4개(7%), 5개 이상(8%), 남성은 3개(9%), 4개(4%), 5개 이상(3%)로 여성에서 기저질환의 개수가 더 많았다.

4가지의 머신러닝 방법 중 가장 정확도가 높은 모형은 남성과 여성 모두에서 Random Forest로 나타났다. 남성의 Random Forest AUC는 80%, 여성은 86%로 데이터를 분류하는 모형의 성능이 매우 좋다고 할 수 있다. 이는

Random Forest가 무작위 표본추출을 사용한 앙상블 기법을 사용하여 이상치의 영향을 적게 받고, 과적합 등의 오류가 적게 나타나 예측력이 비교적 높게 나타났다[36,37].

변수중요도 결과 남성과 여성 모두 BMI가 가장 큰 영향을 주는 것으로 나타났다. BMI는 비만의 진단 지표로 사용되며, 허리둘레와 연관이 있음에 따라[20] 변수중요도에서 BMI의 영향력이 매우 높게 나타났다. 이는 비만이 대사증후군의 대표적인 위험인자로 나타난 결과[11]와 함께 높은 BMI가 대사증후군 유병에 큰 영향을 줄 수 있음을 시사한다. 다음으로 영향력이 큰 변수는 상대악력으로 남녀 모두 높게 나타났다. 악력은 신체의 근력을 측정하는 대표적인 지표로, 당뇨, 고혈압 등 대사증후군과 연관성이 높은 것으로 알려져 있다[38]. 한국인유전체역학조사 자료를 사용하여 분석한 선행연구에 따르면 40-64세의 성인에서 상대악력이 높을수록 대사증후군 발생 위험률이 남성에서 0.56배, 여성에서 0.39배 통계적으로 유의하게 낮은 것으로 나타났다[19].

여성과 남성 모두 식이 관련 변수의 중요도가 높게 나타났다. 대사증후군 진단 질병의 수가 증가할수록 식사염증 지표에서 염증의 정도가 높은 것으로 나타나[39] 식이와 대사증후군은 큰 연관성이 존재함을 나타냈다.

여성에서 여성력과 관련하여 초경 연령과 폐경 여부가 가장 큰 영향을 주는 인자로 나타났다. 선행연구에 따르면, 여성은 폐경이 발생하는 연령 즈음에 호르몬 등 다양한 대사성 변화를 보이며, 이러한 변화로 인해 복부 지방의 증가 등 대사증후군의 위험성이 커지는 경향이 있다[40]. 폐경 전 여성과 비교하여 폐경 후 여성에서 대사증후군의 진단 기준에 부합한 개수가 증가하게 되고[40], 대만의 19세 이상 여성을 대상으로 한 연구에서 초경 연령이 14세 이상인 여성보다 12세 이전에 초경을 경험한 여성의 대사증후군 발생 오즈가 1.71배 높고, 12-14세 사이에 초경을 경험한 여성은 1.22배 높게 나타나 초경 연령이 빠를수록 대사증후군의 발생 위험이 커지는 것으로 나타난다[41, 42].

남성과 여성의 변수중요도에서 큰 차이를 보이는 것은 연령, 기저질환의 개수, 교육수준이다. 기저질환의 개수는 기초통계량에서도 남성보다 여성의 기저질환의 개수가 많은 것으로 나타났다. 교육수준의 영향은 남성보다 여성에서 더 크게 나타났다. 한국인유전체역학조사 자료를 사용하여 분석한 연구에서 여성의 경우 교육수준이 낮아질수록 대사증후군의 위험률이 높아졌지만, 남성의 경우 교육수준에 따른 차이가 없었다[12]. 따라서, 본 연구는 선행연구와 유사한 결과를 보인다고 할 수 있다. 대

사증후군의 위험 요인으로 알려진 변수들을 통해 대사증후군 유병에 미치는 영향의 정도를 파악하였으며, 인구통계학적, 건강행태, 식이, 여성력 등 다양한 변수를 이용하여 머신러닝의 방법으로 예측력 및 영향력의 크기를 파악하였다는 점에서 의의가 있다.

본 연구의 제한점은 다음과 같다. 첫 째, 본 연구의 자료는 횡단적 조사자료로서 대사증후군 영향요인 및 대사증후군 유병률 등을 시간적 순서를 반영하는 데에 어려움이 존재했다. 둘째, 24시간 회상 조사를 통해 수집된 식이 자료이기 때문에 해당 식이의 자료가 대사증후군에 정확하게 영향을 미쳤다고 보긴 어렵다. 셋 째, 국민건강영양조사에 존재하지 않는 요인(심폐체력 등)을 제외하여 한정적인 대사증후군 위험요인을 사용으로 모든 위험요인을 대상으로 분석했다고 보긴 어렵다. 넷 째, 모든 모델의 성능이 높다고 할 수 없다. AUC의 측면 뿐만아니라 다양한 측면에서 분석 및 비교해볼 필요가 있다.

## 5. 결론

본 연구는 국민건강영양조사 제 7기와 제 8기 1차년도 자료를 활용하여 다양한 만성질환의 원인으로 알려진 대사증후군 유병에 영향을 주는 요인을 탐색하고, 대사증후군 예측모델 개발 및 예방을 위해 융복합적으로 실시된 연구이다. 대사증후군 유병에 영향을 주는 것으로 알려진 인구통계학적, 건강행태, 식이, 여성력의 변수를 바탕으로 4가지 머신러닝에 투입하였다. 가장 예측력이 높았던 모형은 여성과 남성에서 Random Forest로 나타났다.

대사증후군에 영향을 주는 변수로는 남성과 여성에서 BMI, 식이(지방, 비타민 C, 비타민 A, 단백질, 에너지 섭취), 주관적 체형인식, 기저질환의 개수, 연령으로 나타났다. 여성의 경우에 남성에 비해 연령, 기저질환의 개수, 교육수준, 초경연령, 폐경여부의 영향을 더 크게 받는 것으로 나타났다.

대사증후군 예방 및 치료를 위해 식이 및 체중, 전반적인 질환의 관리가 필수적임을 나타낸다. 본 연구결과와 후속 연구로 대사증후군 예방을 위한 중재 전략을 수립하여 검증이 필요할 것이다.

## REFERENCES

[1] National Cholesterol Education Program (NCEP)

- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*, 106(25), 3143-3421.  
DOI : 10.1161/circ.106.25.3143.
- [2] Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., ... & Costa, F. (2005). Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute scientific statement. *Circulation*, 112(17), 2735-2752.  
DOI : 10.1161/CIRCULATIONAHA.105.169404
- [3] Grundy, S. M. (2007). Metabolic syndrome: a multiplex cardiovascular risk factor. *The Journal of Clinical Endocrinology & Metabolism*, 92(2), 399-404.  
DOI : 10.1210/jc.2006-0513
- [4] Huh, J. H., Kang, D. R., Kim, J. Y., & Koh, K. K. (2021). Metabolic Syndrome Fact Sheet 2021: Executive Report. *CardioMetabolic Syndrome Journal*, 1.  
DOI : 10.51789/cmsj.2021.1.e15
- [5] Shin, S., & Lee, S. (2019). Association between total diet quality and metabolic syndrome incidence risk in a prospective cohort of Korean adults. *Clinical nutrition research*, 8(1), 46-54.  
DOI : 10.7762/cnr.2019.8.1.46
- [6] Grundy, S. M. (2008). Metabolic syndrome pandemic. *Arteriosclerosis, thrombosis, and vascular biology*, 28(4), 629-636.  
DOI : 10.1161/ATVBAHA.107.151092
- [7] Bang, S. Y. (2019). The relations between metabolic syndrome, physical activity, and dietary patterns in Korean adults. *Journal of the Korea Academia-Industrial cooperation Society*, 20(2), 662-672.  
DOI : 10.5762/KAIS.2019.20.2.662
- [8] Han, M. (2011). Metabolic syndrome emerging from menopause. *The Journal of Korean Society of Menopause*, 17(3), 127-135.  
DOI : 10.6118/jksm.2011.17.3.127
- [9] Won JC, Hong JW, Noh JH, Kim DJ(2016) association between age at menarche and risk factors for cardiovascular diseases in Korean women. The 2010 to 2013 Korea National Health and Nutrition Examination Survey. *Med (Baltimore)* 95(18), 3580- 3589.  
DOI : 10.1097/MD.0000000000003580
- [10] Kang, H. M., & Kim, D. J. (2012). Gender differences in the association of socioeconomic status with metabolic syndrome in middle-aged Koreans. *Korean Journal of Medicine*, 82(5), 569-575.  
DOI : 10.3904/kjm.2012.82.5.569
- [11] Kim, E., & Oh, S. W. (2012). Gender differences in the association of occupation with metabolic syndrome in Korean adults. *The Korean Journal of Obesity*, 21(2), 108-114.  
DOI : 10.7570/kjo.2012.21.2.108
- [12] Seo, J. M., Lim, N. K., Lim, J. Y., & Park, H. Y. (2016). Gender difference in association with socioeconomic status and incidence of metabolic syndrome in Korean adults. *The Korean Journal of Obesity*, 25(4), 247-254.  
DOI : 10.7570/kjo.2016.25.4.247
- [13] Oh, S. I., Hwang, Y. S., & Rhyu, M. J. (2013). Effects of a Combined Exercise Program on the Body Composition, Health-related Physical Fitness, and Metabolic Syndrome Risk Factor in Middle-aged Women. *The Official Journal of the Korean Academy of Kinesiology*, 15(3), 91-100.  
DOI : 10.15758/jkak.2013.15.3.91
- [14] Kim, A. (2018). Effect of health behaviors, dietary habits, and psychological health on metabolic syndrome in one-person households among Korean young adults. *Journal of Digital Convergence*, 16(7), 493-509.  
DOI : 10.14400/JDC.2018.16.7.493
- [15] Lim, M., & Kim, J. (2020). Association between fruit and vegetable consumption and risk of metabolic syndrome determined using the Korean Genome and Epidemiology Study (KoGES). *European journal of nutrition*, 59(4), 1667-1678.  
DOI : 10.1007/s00394-019-02021-5
- [16] Yang, H., Kim, H., Kim, J. M., Chung, H. W., & Chang, N. (2016). Associations of dietary intake and metabolic syndrome risk parameters in Vietnamese female marriage immigrants in South Korea: The KoGES follow-up study. *Nutrition research and practice*, 10(3), 313-320.  
DOI : 10.4162/nrp.2016.10.3.313
- [17] Kim, D. I., Kim, J. Y., Lee, M. K., Lee, H. D., Lee, J. W., & Jeon, J. Y. (2012). The relationship between fitness, BMI and risk factors of metabolic syndrome among university students in Korea. *The Korean Journal of Obesity*, 21(2), 99-107.  
DOI : 10.7570/kjo.2012.21.2.99
- [18] Kim, J., Yoon, D. W., Lee, S. K., Lee, S., Choi, K. M., Robert, T. J., & Shin, C. (2017). Concurrent presence of inflammation and obstructive sleep apnea exacerbates the risk of metabolic syndrome: a KoGES 6-year follow-up study. *Medicine*, 96(7).  
DOI : 10.1097/MD.0000000000004488
- [19] Cho, J., Yoon, E., & Park, S. H. (2019). Association of relative handgrip strength with the incidence of metabolic syndrome in Korean adults: a community based cohort study. *Exercise Science*, 28(3), 303-310.  
DOI : 10.15857/ksep.2019.28.3.303
- [20] Choe, E. K., Rhee, H., Lee, S., Shin, E., Oh, S. W., Lee, J. E., & Choi, S. H. (2018). Metabolic syndrome prediction using machine learning models with genetic and clinical information from a nonobese healthy population. *Genomics & informatics*, 16(4).  
DOI : 10.5808/GI.2018.16.4.e31

- [21] Korea Centers for Disease Control and Prevention. (2020). Guideline for raw data use of The Seventh Korea National Health and Nutrition Examination Survey (KNHANES VII), 2016–2018.
- [22] Expert Panel on Detection, E. (2001). Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). *Jama*, 285(19), 2486–2497. DOI : 10.1001/jama.285.19.2486
- [23] Yoon, Y. S., & Oh, S. W. (2014). Optimal waist circumference cutoff values for the diagnosis of abdominal obesity in Korean adults. *Endocrinology and Metabolism*, 29(4), 418–426. DOI : 10.3803/EnM.2014.29.4.418
- [24] Yang, Y. H., Kim, J. S., & Jeong, S. H. (2020). Prediction of dental caries in 12-year-old children using machine-learning algorithms. *Journal of Korean Academy of Oral Health*, 44(1), 55–63. DOI : 10.11149/jkaoh.2020.44.1.55
- [25] Seul, M. S. (2016). Current Status and Future Developments of Machine Learning Artificial Intelligence in Law Focusing the Cusp of Machine Learning in US and Discourses over Legal Profession and Law School Education. *The Justice*, 156, 269–302.
- [26] Cho, S. Y., Kim, S. H., Kang, S. H., Lee, K. J., Choi, D., Kang, S., ... & Chae, I. H. (2021). Pre-existing and machine learning-based models for cardiovascular risk prediction. *Scientific reports*, 11(1), 1–10. DOI : 10.1038/s41598-021-88257-w
- [27] Lee, B. J. (2019). Prediction model of hypercholesterolemia using body fat mass based on machine learning. *The Journal of the Convergence on Culture Technology*, 5(4), 413–420. DOI : 10.17703/JCCT.2019.5.4.413
- [28] Park, J. H., Cho, H. E., Kim, J. H., Wall, M. M., Stern, Y., Lim, H., ... & Cha, J. (2020). Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ digital medicine*, 3(1), 1–7. DOI : 10.1038/s41746-020-0256-0
- [29] Kim, H.-S. (2019). Convergence Analysis of Risk factors for Readmission in Cardiovascular Disease: A Machine Learning Approach. *Journal of Convergence for Information Technology*, 9(12), 115–123. DOI : 10.22156/CS4SMB.2019.9.12.115
- [30] In-Ja, L., & Junho, L. (2020). Predictive of Osteoporosis by Tree-based Machine Learning Model in Post-menopause Woman. *Journal of Radiological Science and Technology*, 43(6), 495–502. DOI : 10.17946/JRST.2020.43.6.495
- [31] Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215–232. DOI : 10.1111/j.2517-6161.1958.tb00292.x
- [32] Lee, S. M., Park, K. D., & Kim, I. K. (2020). Comparison of machine learning algorithms for Chl-a prediction in the middle of Nakdong river (focusing on water quality and quantity factors). *Journal of Korean Society of Water and Wastewater*, 34(4), 277–288. DOI : 10.11001/jksww.2020.34.4.277
- [33] Jeong, M. C., Lee, J. H., & Oh, H. Y. (2020). Ensemble Machine Learning Model Based Youtube Spam Comment Detection. *Journal of the Korea Institute of Information and Communication Engineering*, 24(5), 576–583. DOI : 10.3745/KTSDE.2021.10.7.257
- [34] Lim, J. S., & Kim, J. M. (2014). An empirical comparison of machine learning models for classifying emotions in Korean Twitter. *Journal of Korea Multimedia Society*, 17(2), 232–239. DOI : 10.9717/kmms.2014.17.2.232
- [35] Go, W. S., Yoon, C. G., Rhee, H. P., Hwang, S. J., & Lee, S. W. (2019). A Study on the prediction of BMI (Benthic Macroinvertebrate Index) using Machine Learning Based CFS (Correlation-based Feature Selection) and Random Forest Model. *Journal of Korean Society on Water Environment*, 35(5), 425–431. DOI : 10.15681/KSWE.2019.35.5.425
- [36] Kim, S. J., & Ahn, H. (2016). Application of random forests to corporate credit rating prediction. *The Journal of Business and Economics*, 32(1), 187–211. DOI : 10.22793/indinn.2016.32.1.006
- [37] Jung, H., & Kim, J. W. (2017). A machine learning approach for mechanical motor fault diagnosis. *Journal of the Society of Korea Industrial and Systems Engineering*, 40(1), 57–64. DOI : 10.11627/jkise.2017.40.1.057
- [38] Yi, D. W., Khang, A. R., Lee, H. W., Son, S. M., & Kang, Y. H. (2018). Relative handgrip strength as a marker of metabolic syndrome: the Korea National Health and Nutrition Examination Survey (KNHANES) VI (2014–2015). *Diabetes, metabolic syndrome and obesity : targets and therapy*, 11, 227–240. DOI : 10.2147/DMSO.S166875
- [39] Kim, M., & Sohn, C. (2016). Analysis of dietary inflammatory index of metabolic syndrome in Korean: data from the health examinee cohort (2012–2014). *Korean J Hum Ecol*, 25, 823–834. DOI : 10.5934/kjhe.2016.25.6.823
- [40] Oh, G. C., Kang, K. S., Park, C. S., Sung, H. K., Ha, K. H., Kim, H. C., ... & Lee, H. Y. (2018). Metabolic syndrome, not menopause, is a risk factor for hypertension in peri-menopausal women. *Clinical hypertension*, 24(1), 1–8. DOI : 10.1186/s40885-018-0099-z
- [41] Chang, C. J., Lai, M. M., Lin, C. C., Liu, C. S., Li, T. C., Li, C. I., & Lin, W. Y. (2016). Age at menarche and its association with the metabolic syndrome in Taiwan. *Obesity research & clinical practice*, 10 Suppl 1, S26–

S34.

DOI : 10.1016/j.orcp.2015.10.003

- [42] Jeong E. J. & Jung B. M. (2020). Analysis of Anthropometric and Behavioral Factors of Korean Female Adolescents According to Age of Menarche: 2013~2017 Korea National Health and Nutrition Examination Survey. *The Korean Journal of Community Living Science*, 31(3), 393-409.  
DOI : 10.7856/kjcls.2020.31.3.393

이 소 은 (So-Eun Lee)

【장학원】



- 2020년 2월 : 고려대학교 응용통계학과 (통계학사)
- 2020년 3월 ~ 현재 : 고려대학교 일반대학원 보건과학과 보건정책관리 전공 석사과정
- 관심분야 : 보건통계, 머신러닝, 건강증진, 정신건강

· E-Mail : leese9609@korea.ac.kr

이 현 실 (Hyun-Sill Rhee)

【장학원】



- 1990년 2월 : 서울대학교 보건대학원 (보건학 석사)
- 2005년 2월 : 경산대학교 보건대학원 (보건학 박사)
- 1994년 3월 ~ 현재 : 고려대학교 보건과학대학 보건정책관리학부 교수
- 관심분야 : 보건정보관리, 의료질관리,

개인건강기록관리

· E-Mail : pridehyun@korea.ac.kr