



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

STOP-Bang 및 임상 지표 기반  
폐쇄성수면무호흡증(OSA) 진단 예측모델  
연구: 심층신경망(DNN) 제안 및  
로지스틱 회귀분석과의 성능 비교

연세대학교 보건대학원  
보건정보통계학과 보건정보통계전공  
이 슬 기

STOP-Bang 및 임상 지표 기반  
폐쇄성수면무호흡증(OSA) 진단 예측모델  
연구: 심층신경망(DNN) 제안 및  
로지스틱 회귀분석과의 성능 비교

지도 하 민 진 교수

이 논문을 보건학 석사학위 논문으로 제출함

2025 년 06 월

연세대학교 보건대학원  
보건정보통계학과 보건정보통계전공  
이 슬 기

이슬기의 보건학 석사학위 논문을 인준함.

심사위원 하 민 진 인

심사위원 박 소 희 인

심사위원 주 재 홍 인

연세대학교 보건대학원

2025 년 06 월

## 감사의 말씀

신촌, 용산, 송파. 마치 서울의 세 꼭지점을 잇는 여행이라도 하듯, 낮과 밤 동안 서울을 누비며 보내온 지난 2년반 동안의 시간은 제 인생에서 가장 바쁘면서도 가장 값진 시간이었습니다. 직장 and 결혼, 그리고 대학원이라는 세 가지 여정을 동시에 걸어가는 일은 분명 쉽지 않았지만, 그 여정의 매 순간을 소중하게 기억할 수 있는 건, 곁에서 함께해준 고마운 분들이 있었기 때문입니다.

먼저, 본 논문이 완성될 수 있도록 연구 주제를 구체화하는 과정에서 진심 어린 지도로 이끌어주신 하민진 교수님께 깊이 감사의 말씀을 드립니다. 교수님의 말씀은 연구를 끝까지 이어갈 수 있는 큰 원동력이 되어주었습니다. 아울러 탄탄한 이론과 심도 깊은 강의로 제 연구의 기초를 단단히 다지는데 큰 도움을 주신 박소희 교수님과, 유익한 코멘트로 연구에 방향성을 더해주신 주재홍 박사님께도 진심으로 감사드립니다.

기쁠 때든 지칠 때든 언제나 함께 웃고 고민해준 동기들 덕분에 1시간이 넘는 통학 길도 결코 외롭지 않았습니다. 학기 중에도, 방학 동안에도 어김없이 얼굴 보며 함께 보낸 즐거운 추억들을 되돌아보니, 본 학위과정을 통해 얻게 된 값지고 소중한 선물 중 하나가 바로 동기들이라 생각합니다. 서로 바쁜 일상 속에서도 끈끈한 인연으로 짧지 않은 이 시간을 이어올 수 있게 해준 동기들에게도 감사 인사 드립니다.

또한 회사 동료 여러분께도 감사의 인사를 전합니다. 학업과 업무를 병행하며 힘겨울 때마다 따뜻한 배려와 격려를 보내주신 모든 회사 동료분들께 감사드립니다. 특히 보건정보통계라는 학문 분야에 관심을 갖게 해주시고 새로운 여정에 큰 용기를 주신 Rina 님, 늘 따뜻한 말과 세심한 배려로 실무에서 함께 고민하며 진심 어린 응원을

보내준 선정님, 항상 묵묵히 힘을 북돋아 주신 Joon 님, 그리고 새로운 팀에 적응하는 과정에서도 늘 열린 마음으로 배려해주신 지현님께도 진심으로 감사드립니다.

그리고 저의 가장 큰 원동력이 되어준 가족에게 깊은 사랑과 감사의 마음을 전합니다. 몸도 마음도 지쳐 있을 때마다 사랑이 담긴 따뜻한 말 한마디로 늘 저를 일으켜 주신 엄마, 아빠, 그리고 어떤 고민이든 밤낮없이 함께 나누고 조언해준 언니에게 진심으로 감사드립니다. 늘 제가 잘하고 있다고, 지금 가는 길이 틀리지 않았다고 힘을 실어주신 어머님과 아버님의 따뜻한 격려도 잊지 않겠습니다.

무엇보다, 대학원이라는 새로운 도전을 처음부터 흔쾌히 응원해주고, 누구보다 가까이에서 늘 곁을 지켜주며 기꺼이 현명한 조언과 따스한 위로를 아끼지 않았던 사랑하는 남편, 홍재에게 가장 깊은 감사를 전합니다. 수업을 마치고 늦은 밤 귀가할 때마다 든든한 등대처럼 마중 나와주고, 자신의 피로는 뒤로한 채 언제나 저를 먼저 챙겨주는 당신이 있었기에 이 긴 여정을 끝까지 완주할 수 있었습니다. 학위과정동안 함께 고민하고 기뻐했던 모든 순간들은 잊지않고 소중한 추억으로 간직하겠습니다.

이 자리를 빌려 저를 믿고 응원해주신 모든 분들께 다시 한번 깊은 감사를 드립니다.

2025 년 6 월

이슬기 올림

## 차 례

국 문 요 약.....	v
I. 서론 .....	1
1. 연구 배경 .....	1
2. 연구 목적 .....	4
II. 연구 방법 .....	6
1. 연구 모형 .....	6
2. 연구 대상 .....	7
3. 변수 선정 .....	10
4. 분석 방법 .....	13
III. 연구 결과 .....	20
1. 연구 대상자의 일반적 특성.....	20
2. STOP-Bang 점수와 OSA 진단 분포 .....	25
3. STOP-Bang 점수와 OSA 진단의 단변수 로지스틱 회귀 분석 결과.....	27
4. 건강지표를 포함한 다변수 로지스틱 회귀 분석 결과.....	30
5. DNN 모델 기반 OSA 진단 예측모델 성능평가 .....	35
6. 로지스틱 회귀분석 모형과 DNN 모델 간 성능 비교.....	41

IV. 고 찰.....	43
V. 결 론 .....	52
참 고 문 헌.....	53
Abstract.....	57



## 표 차례

표 1. OSA 진단 여부에 따른 연구 대상자의 일반적 특성 .....	22
표 2. OSA 진단 여부에 따른 STOP-Bang 설문 항목 분포 .....	24
표 3. STOP-Bang 점수대별 OSA 진단 분포 .....	26
표 4. STOP-Bang 점수와 OSA 진단의 단변수 로지스틱 회귀분석 결과 .....	28
표 5. 학습용 데이터셋에서의 OSA 진단과 관련된 건강 및 임상 요인의 다변수 로지스틱 회귀분석 결과 .....	32
표 6. 로지스틱 회귀 모형과 DNN 모델의 ROC 곡선 하의 면적(AUC) 비교 .....	42

## 그림 차례

그림 1. 연구 대상자 선정 흐름도 .....	9
그림 2. 심층신경망(Deep Neural Network, DNN) 구조 .....	16
그림 3. STOP-Bang 점수 기반 OSA 진단 예측의 ROC 곡선 .....	29
그림 4. 검증용 데이터셋에서의 건강 및 임상 요인 기반 OSA 진단 예측 ROC 곡선 ..	34
그림 5. OSA 진단 예측 DNN 모델의 학습 손실 변화 추이 .....	36
그림 6. DNN 모델 기반 OSA 진단 예측의 ROC 곡선 .....	37
그림 7. OSA 진단 예측 DNN 모델에 대한 SHAP 요약 시각화 결과 .....	40

## 국 문 요 약

### STOP-Bang 및 임상 지표 기반 폐쇄성수면무호흡증(OSA) 진단 예측모델 연구: 심층신경망(DNN) 제안 및 로지스틱 회귀분석과의 성능 비교

**배경 및 목적:** 폐쇄성수면무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 반복적인 상기도 폐쇄로 인해 호흡이 제한되는 질환으로, 고혈압, 심혈관질환, 당뇨 등 만성질환과 밀접한 연관이 있는 주요 공중보건 문제로 인식되고 있다. OSA 조기 선별을 위한 도구로 STOP-Bang 설문지 널리 사용되고 있으며, 이 설문지는 높은 민감도를 바탕으로 고위험군을 효과적으로 선별할 수 있는 장점이 있다. 그러나 낮은 특이도로 인해 개별 진단의 정밀도에는 한계가 존재하며, 단독 활용 시 진단 예측의 정확도를 확보하는 데 어려움이 있다는 지적이 제기되고 있다. 이에 본 연구는 STOP-Bang 점수의 한계를 보완하고 예측 성능을 향상시키기 위해 추가적인 임상 및 건강행태 지표를 통합한 예측모델을 구축하고자 하였으며, 전통적 통계 기법인 로지스틱 회귀분석과 머신러닝 기반 심층신경망(Deep Neural Network, DNN)의 예측 성능을 비교하여 보다 효과적인 OSA 진단 예측 모델을 제시하는 데 목적이 있다.

**대상 및 방법:** 본 연구는 2019년부터 2023년까지 수행된 제8-9기 국민건강영양조사(KNHANES) 자료 중 40세 이상 성인 14,939명을 대상으로 하였다. 폐쇄성수면무호흡증(OSA) 진단 여부는 자가 보고된 의사 진단을 기준으로 정의하였으며, STOP-Bang 설문 점수 및 구성 항목 외에도 혈압, 혈당, 지질 수치, 체질량지수(BMI), 음주 및 흡연 여부, 수면시간, 앉아있는 시간 등의 임상 및 건강행태 지표를 분석에 포함하였다. 통계 분석은 복합표본설계를 고려하여 로지스틱 회귀분석을 실시하였으며, STOP-Bang 점수 단일 항목의 설명력을 평가하기 위한 단변수 분석과 건강 지표를 추가로 포함한

다변수 분석을 병행하였다. 예측 성능 검증을 위해 전체 데이터를 훈련용과 검증용으로 7:3 비율로 분할하여 적용하였으며, 추가적으로 심층신경망(Deep Neural Network, DNN) 기반의 예측 모델을 구성하고, 예측 정확도 및 변수 기여도를 비교 평가하기 위해 ROC-AUC 및 SHAP 분석을 수행하였다.

**연구결과:** 단변수 로지스틱 회귀분석에서 STOP-Bang 점수는 OSA 진단과 유의한 양의 연관성을 보였으며(OR=2.31, 95% CI: 2.02-2.65,  $p<0.0001$ ), 해당 모형의 AUC는 0.8268로 확인되었다. 다변수 로지스틱 회귀분석 결과, AUC는 0.8660으로 수면 중 무호흡/목격 여부가 가장 유의한 예측 인자로 확인되었으며(OR=20.68, 95% CI: 9.69-44.14,  $p<0.0001$ ), 코골이 여부와(OR=2.29, 95% CI: 1.09-4.79,  $p=0.0287$ ) 일일 총 잤아 있는 시간(OR=1.14, 95% CI: 1.06-1.22,  $p=0.0003$ ), 음주 여부(OR=0.16, 95% CI: 0.04-0.67,  $p=0.0130$ ) 또한 유의한 관련성을 보였다. 한편, 심층신경망(DNN) 기반 예측모델의 경우 AUC는 0.9020으로, 세 모델 중 가장 높은 예측력을 보였다. 또한 SHAP 분석 결과 로지스틱 회귀분석에서 유의했던 변수들과 유사한 경향성을 보이며, 해석 가능성과 임상적 타당성을 동시에 확보할 수 있음을 시사하였다.

**결론:** STOP-Bang 설문은 OSA 고위험군 선별에 효과적인 도구로 활용될 수 있으나, 개별 진단 수준에서는 특이도의 한계가 뚜렷하여 보완이 필요하다. 본 연구는 STOP-Bang 점수에 임상 및 건강행태 정보를 결합함으로써 예측 정밀도를 향상시킬 수 있음을 확인하였으며, DNN 기반의 머신러닝 모델이 기존 통계모형보다 우수한 예측 성능을 보임을 통해 향후 OSA 조기 선별을 위한 고도화된 예측모델 개발의 가능성을 제시하였다.

---

핵심어: 폐쇄성수면무호흡증(OSA), STOP-Bang, 로지스틱 회귀분석, 심층신경망, 예측 모형, ROC-AUC, 국민건강영양조사

# I. 서 론

## 1. 연구 배경

폐쇄성수면무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 상기도의 반복적이고 부분적 혹은 완전한 폐쇄로 인해 정상적인 호흡이 중단되는 만성 질환으로, 심각한 건강 문제를 초래할 수 있는 주요 공중보건 문제로 대두되고 있다(Senaratna et al., 2017). OSA는 수면 중 산소포화도가 급격히 감소하고 수면 구조가 반복적으로 파괴되는 특성을 가지며, 이는 만성 피로, 주간 졸음, 인지 기능 저하, 그리고 삶의 질 저하와 밀접하게 연관되어 있다(Punjabi, 2008).

임상적으로 OSA는 심혈관계 질환, 고혈압, 이상지질혈증, 제2형 당뇨병 등 주요 만성 질환의 위험 요인으로 작용하며(Marin et al., 2005), 특히 조기 진단과 치료가 이루어지지 않을 경우 심장마비나 뇌졸중과 같은 치명적인 합병증을 유발할 가능성이 높다. 이에 따라, OSA를 조기에 진단하고 고위험군을 선별할 수 있는 효과적인 도구의 개발이 요구되고 있다.

한국에서의 연구에 따르면, OSA의 유병률은 약 3.2-4.5%로 추정되며(Kim et al., 2004), 특히 중년 남성에서 높은 발생률을 보인다. 그러나 이 수치는 표준 검사법을 통한 진단 사례에 국한된 것으로, 진단받지 못한 잠재적 환자는 훨씬 더 많을 것으로 예상된다. 이는 OSA의 초기 증상이 경미하여 환자가 이를 인지하지 못하거나, 접근성이 제한된 진단 방법으로 인해 조기 발견이 어렵기 때문이다.

OSA의 진단을 위한 표준 검사법은 수면다원검사(Polysomnography, PSG)로, 이는 수면 중 호흡 패턴, 산소포화도, 뇌파, 안구운동, 심전도, 흉복부 운동 등 다양한 생리학적 신호를 동시 측정하여 수면 구조 및 호흡 장애의 변화를 종합적으로 평가하여 진단한다(Kapur et al., 2017). 그러나 PSG는 고도의 전문 장비와 기술이 필요함에 따라 피검사자가 수면 전문 기관에 1박 2일 이상 입실해야 하며, 검사 일정 조율, 사

전 준비, 사후 절차 등과 관련된 시간적·행정적 부담이 수반된다. 또한 검사 중에는 다양한 생리학적 신호 측정을 위한 복수의 장비를 부착한 상태에서 수면을 진행해야 하므로, 피검사자의 수면 질 저하 및 불편함을 유발할 수 있다. 이와 함께 고가의 검사 비용과 더불어 장비와 인력 비용 부담에 따라 실제 임상에서는 접근성이 낮다는 단점이 있다(Rosen et al., 2012, Chung et al., 2016). 따라서, 표준 검사법을 대체하거나 보완할 수 있는 간단하고 비용 효율적인 선별 도구가 임상 현장에서 중요성을 더해가고 있다.

이러한 배경에서 개발된 STOP-Bang 설문지는 OSA 선별을 위한 도구로, 코골이(Snoring), 피곤함(Tiredness), 목격된 무호흡(Observed Apnea), 고혈압(Pressure), 연령(Age), 체질량지수(BMI), 성별(Gender), 목둘레(Neck Circumference)와 같은 8개의 항목으로 구성되어 있다(Chung et al., 2008). 설문지의 단순성과 높은 민감도로 인해 OSA의 고위험군을 선별하는 데 유용한 도구로 널리 사용되고 있으나, STOP-Bang 설문지는 높은 민감도로 인해 OSA 고위험군을 선별하는 데 유용하지만, 상대적으로 낮은 특이도로 인해 실제 진단 정확도에는 제한이 있다(Chung et al., 2014). 이에 따라 STOP-Bang 점수 기반 선별의 한계를 극복하고, 보다 정밀한 예측을 가능하게 하기 위해 설문 항목 외에도 임상적 건강 지표를 함께 고려함과 동시에 통계적 또는 머신러닝 기반 접근법이 시도되고 있다.

대표적인 통계적 예측 모형으로 해석력이 뛰어나고 구조가 단순하다는 장점이 있는 로지스틱 회귀 분석의 경우 선형 모델이므로 변수 간 복잡한 상호작용과 비선형적 관계를 반영하는 데 한계가 있다. 반면, 심층신경망(Deep Neural Network, DNN)과 같은 머신러닝 기법은 이러한 제약 없이 복잡한 변수 간 상호작용을 비선형적으로 반영할 수 있다는 장점이 있다(Shickel et al., 2023).

최근 머신러닝(machine learning)은 의료 데이터를 활용한 질병 예측 모델 개발에서 중요한 역할을 하고 있다. 머신러닝은 다차원적이고 복잡한 의료 데이터를 기반으로 숨겨진 패턴을 학습하고, 이를 통해 예측 성능을 크게 향상시킬 수 있다(Buch et al., 2018). 특히, OSA와 같은 다원적이고 복합적인 요인을 가진 질환에서는 기존의

통계적 접근보다 머신러닝 알고리즘이 효과적으로 작용할 가능성이 높다. 선행 연구에서는 다양한 머신러닝 기법을 활용하여 OSA 예측 모델을 개발한 사례들이 보고되고 있으며, 데이터 기반 분석을 통해 민감도와 특이도를 동시에 향상시키는 결과를 보여주었다(Holfinger et al., 2022). STOP-Bang 점수에 임상 정보를 결합한 머신러닝 기반 예측 모델은 기존 설문 도구의 한계를 일정 부분 보완할 수 있으며, 복잡한 변수 간 관계를 반영하고 해석 가능성을 확보할 수 있는 보완적 도구로서 OSA 조기 선별에 활용 가능성이 있다.

본 연구는 STOP-Bang 점수를 기반으로 국민건강영양조사(KNHANES) 데이터를 활용하여 STOP-Bang 점수와 다양한 건강 지표를 바탕으로 예측 모델을 구축하고, 전통적 통계 분석 방법인 로지스틱 회귀분석과 머신러닝 기법의 성능을 비교함으로써 폐쇄성수면무호흡증(OSA)의 효율적인 선별 가능성을 탐색하고자 한다. 이를 통해 비용과 접근성의 제약이 있는 수면다원검사의 한계를 보완할 수 있는 현실적인 대안 도구를 제시하고, 향후 국민건강 데이터를 활용한 머신러닝 모델의 개발 및 적용 가능성을 제시하며, 공공보건 및 임상 진단 분야에서의 적용 가능성을 높이는데 기여하고자 한다.

## 2. 연구 목적

본 연구는 2019년부터 2023년까지 수행된 제8-9기 국민건강영양조사 통합 원시자료를 활용하여, 40세 이상의 성인을 대상으로 폐쇄성수면무호흡증(Obstructive Sleep Apnea, OSA)의 진단 예측 모델을 개발하고 그 성능을 평가하는 것을 목적으로 한다. 이를 통해 STOP-Bang 설문 항목이 갖는 폐쇄성수면무호흡증(OSA) 진단 선별 정도를 파악하고, 개별 진단에서의 특이성을 보완하기 위해 추가적인 임상 및 건강 관련 변수들을 결합하고자 한다. 나아가 이러한 변수들을 활용하여 로지스틱 회귀 분석과 심층신경망(Deep Neural Network, DNN) 기반 예측 모델을 각각 구축하고, 변수 간 복합적인 상호작용을 고려한 진단 성능을 비교한다. 궁극적으로는 폐쇄성수면무호흡증과 관련된 주요 요인들의 복합적인 관계를 반영하여 개별 진단에 있어 높은 예측의 정확도를 가지는 최종 진단 모델을 제안한다.

이를 위한 구체적인 연구 목적은 다음과 같다.

첫째, 연구 대상자의 일반적인 특성을 확인하고 STOP-Bang 점수와 폐쇄성수면무호흡증(OSA) 진단의 분포를 파악한다.

둘째, STOP-Bang 점수를 독립 변수로 하여 단변수 로지스틱 회귀분석을 수행함으로써, 해당 점수와 폐쇄성수면무호흡증(OSA) 진단 여부 간의 통계적 연관성을 평가한다.

셋째, STOP-Bang 항목의 설명력을 보완하기 위해 추가적인 임상 및 건강 지표를 반영하여 OSA 진단 여부에 대한 통계적 연관성을 다변수 로지스틱 회귀분석을 통해 평가한다.

넷째, 변수들간의 복잡한 상호관계를 고려한 진단 예측 모델을 구현하기 위해 심층신경망(Deep Neural Network, DNN) 기반 예측 모델을 구축하고 성능을 평가한다.



다섯째, 로지스틱 회귀분석 결과들과 심층신경망(DNN) 모델을 비교하고 민감도와 특이도를 개선한 개별 맞춤형 폐쇄성수면무호흡증(OSA) 진단 예측 모델을 제안한다.

## II. 연구방법

### 1. 연구 모형

본 연구는 만 40세 이상의 성인을 대상으로 폐쇄성수면무호흡증(OSA) 진단 예측 모델을 구축하는 후향적 단면연구이다. 분석을 위해 폐쇄성수면무호흡증 선별도구인 STOP-Bang가 신규 도입된 2019년부터 2023년까지 수행된 제8-9기 국민건강영양조사 통합 원시자료를 활용하였다. 본 연구는 STOP-Bang 설문 조사 대상인 만 40세 이상의 성인 남녀를 대상으로 하였으며, 종속변수는 폐쇄성수면무호흡증에 대한 의사의 진단 경험 여부에 있어 자가 보고한 결과를 기반으로 정의된 OSA 진단 여부이며, 독립변수는 STOP-Bang 설문 점수 및 8개 문항과 더불어 기존 선행연구를 바탕으로 도출된 OSA와 관련된 다양한 임상 및 건강행태 관련 변수를 선별하고 포함하여 분석에 사용하였다. 본 연구에 사용된 국민건강영양조사 자료는 질병관리청이 국민건강증진법에 따라 시행하는 전국 규모 단위의 건강 조사 자료로서, 원시자료는 공공 데이터로 공개되며, 모든 자료는 개인정보 보호법 및 통계법에 따라 비식별화 처리되어 개인 식별이 불가능한 형태로 제공된다.

## 2. 연구 대상

본 연구는 2019년부터 2023년까지 수행된 제8기와 제9기의 국민건강영양조사에 참여한 40대 이상의 성인을 대상으로 한다. 국민의 건강 상태와 영양 섭취 실태를 평가하고, 만성 질환의 유병률 및 위험 요인을 파악하기 위해 시행되는 국민건강영양조사는 질병관리청이 주관하는 전국 단위의 대규모 조사이다. 1998년 제1기를 시작으로, 제3기(2005년)까지는 3년 주기 체계로 운영된 이후, 2007년 제4기부터는 국민건강증진법에 따라 연중조사 체계로 개편되어 매년 조사가 실시되고 있으며, 대한민국 국민의 건강 관련 지표 산출 및 국가 보건정책 수립의 기초자료로 활용된다.

매년 약 1만 명의 국민을 대상으로 진행되는 국민건강영양조사는 건강 설문조사, 영양조사, 검진 조사의 세 가지 항목으로 구성되어 질병과 관련된 다양한 임상 지표와 생활 습관 데이터를 제공한다. 표본은 시도, 동·읍면, 주택유형(일반/아파트) 등을 기준으로 층화하고, 인구주택총조사 기반의 최신 데이터를 바탕으로 2단계 층화집락표본추출 방식을 통해 조사구 및 가구 단위로 표본을 선정한다. 제8기에서는 2016년 인구주택총조사 자료를 사용하고, 내재적 층화기준으로 성별, 연령, 주거면적 비율을 반영하였다. 이에 따라 총 576조사구, 14,400가구, 30,490명의 대상자 중 건강 설문 조사, 검진 조사, 영양조사 중 1개 이상 참여자는 총 22,559명이다. 제9기에서는 2019년 인구주택총조사 자료를 기반으로 표본추출틀을 구성하였으며, 가구의 학력, 1인 가구 비율, 미혼 가구 비율 등 가구 단위의 인구사회학적 특성을 더 정교하게 반영하여 층화기준을 개선하였다. 이에 따라 총 576조사구, 14,400가구, 19,262명의 대상자 중 건강 설문조사, 검진 조사, 영양조사 중 1개 이상 참여한 자는 총 13,194명이다.

본 연구는 폐쇄성수면무호흡증 선별도구(STOP-Bang) 조사가 이루어진 제8기 2019년부터 제9기 2023년까지의 데이터를 이용하였다. 2019년부터 2023년까지의 5개년도 국민건강영양조사자에서 만 40세 이상의 성인 남녀 전체 참여자 수는 총 22,489명이었으며, 참여자 중 본 연구에서의 폐쇄성수면무호흡증 진단 여부에 답변한 대상자는

20,173명이었다. 이 중 선행연구를 참고하여 주요 분석 변수로 폐쇄성수면무호흡증 선별 도구 항목(STOP-Bang: 코골이, 피곤함, 수면무호흡목격자, 고혈압 여부, BMI, 나이, 목둘레, 성별) 및 임상 관련 변수 (총콜레스테롤, 당화혈색소, 공복혈당, 헤모글로빈 등), 허리둘레, 혈압, 흡연, 음주, 그리고 건강 관련 변수 (수면시간, 잤아있는 시간)에 대해 결측이거나 모름 또는 무응답한 총 7,550명을 제외하였다. 따라서, 최종적으로 본 연구에서 분석한 대상자는 총 14,939명이었다(그림 1).

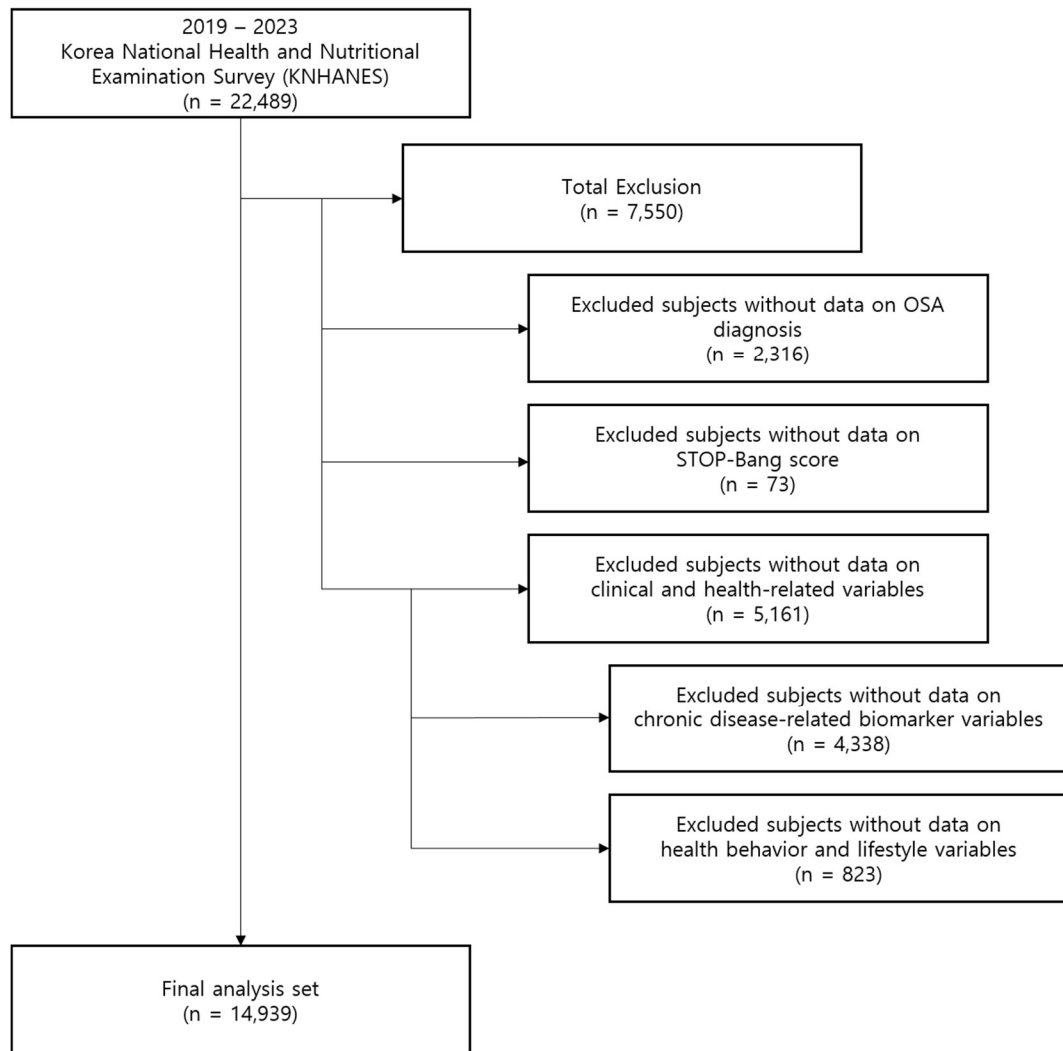


그림 1. 연구 대상자 선정 흐름도

### 3. 변수 선정

#### 가. 종속 변수

본 연구에서는 국민건강영양조사에서 제공하는 자가 보고된 폐쇄성수면무호흡증 진단 여부 항목을 종속변수로 사용하였다. 이 변수는 수면검사(병원 또는 집)를 통해 폐쇄성수면무호흡증을 의사로부터 진단받은 적이 있는지에 대한 질문에 "예"로 응답한 경우를 OSA 진단군으로 분류하였다. 본 변수는 이분형 변수로 처리하였다.

#### 나. 독립 변수

##### 1) STOP-Bang 설문 관련 변수

본 연구의 첫 번째 주요 입력 변수로는 폐쇄성수면무호흡증(OSA) 위험 평가 도구로 널리 사용되는 STOP-Bang 설문지의 점수로 설정하였다. 2008년 Chung et al.에 의해 개발된 STOP-Bang 설문지는 OSA의 고위험군을 선별하기 위한 도구로, OSA의 주요 위험 요인인 코골이, 피로감, 무호흡 목격자, 고혈압, BMI, 연령, 목둘레, 성별을 기반으로 8개의 문항으로 구성된다. STOP-Bang 점수는 총 0점에서 8점까지 범위를 가지며, 총 점수가 높을수록 OSA의 위험이 높다고 평가되어 0-2점인 경우 OSA 저위험군, 3-4점 중등도 위험군, 5-8점은 고위험군으로 분류하였다.

STOP-Bang 설문 항목의 정의 및 처리 기준은 다음과 같다:

- (1) 코골이 (Snoring): 자가 보고된 코골이 여부로, "예" 라고 응답한 경우 1점.
- (2) 피곤함 (Tired): 주간 졸음이나 피곤함을 느끼는지에 대한 응답으로, "예" 라고 응답한 경우 1점.

- (3) 목격자 (Observed): 수면 중 무호흡이 목격되었는지 여부로, "예" 라고 응답한 경우 1점.
- (4) 고혈압 진단 여부 (blood Pressure): 의사로부터 고혈압 진단을 받은 적이 있거나 고혈압 치료 중인 경우 1점.
- (5) 체질량지수 (BMI): 체질량지수(BMI)가  $30(\text{kg}/\text{m}^2)$  보다 높을 경우 1점.
- (6) 나이 (age): 만 50세를 넘을 경우 1점.
- (7) 목둘레 (neck circumstance): 목둘레가 40cm를 넘을 경우 1점.
- (8) 성별 (gender): 남성일 경우 1점.

단, 체질량지수의 경우 기존 STOP-Bang 설문 항목에서는 서양인에서의 고도 비만 (severe obesity)의 기준으로  $35\text{kg}/\text{m}^2$  이상을 사용하였으나, 본 연구에서는 WHO(2000)에서 제안한 아시아인의 고도 비만 기준인  $30\text{kg}/\text{m}^2$  초과 유무로 변경하였다. 또한, 목둘레는 일반인 대상의 경우는 남자는 43cm, 여자는 41cm 이상인지를 확인하도록 되어 있으나, 본 연구에서는 선행 연구에서 많이 사용하고 있는 남녀 공통 40cm 이상을 기준으로 정의하였다.

## 2) 추가 임상 및 건강 관련 변수

추가 독립 변수는 인구학적, 건강행태, 만성질환 관련 변수들로 구성되었으며, 가능한 범위 내에서 연속형으로 수치화하여 다음과 같이 처리하였다:

### (1) 인구학적 변수: 성별, 연령

: 본 연구에서는 인구학적 요인으로 성별과 연령을 분석에 포함하였다. 성별은 국민 건강영양조사 자료에 제공된 남녀 구분 정보를 변경없이 활용하였으며, 연령은 연속형 변수로 사용하였다.

(2) 건강 및 생활 관련 변수: 흡연 상태, 음주 상태, 수면시간, 활동 관련 변수 (앉아있는 시간)

건강 및 생활 관련 변수는 국민건강영양조사 건강면접조사 및 건강행태조사의 설문 문항을 포함하였다. 흡연 상태는 현재 흡연자, 과거 흡연자, 비흡연자 그룹으로 구분하였다. 음주 상태는 최근 1년간 음주 이력에 있어 연속형 변수로의 전환을 위해 각 횟수별 음주 이력에 대해 수치화를 적용하였다.

수면시간의 경우 주중과 주말 하루 평균 수면시간을 통해 연속형 형태의 1주일간 총 수면시간을 계산하였으며, 활동 관련 변수의 경우 평소 하루 앉아서 보내는 시간 (시간, 분)을 최종적으로 시간 형태로 변환하여 산출하여 포함하였다.

(3) 만성질환 관련 생체지표: 만성질환(고혈압, 당뇨병, 이상지질혈증, 빈혈)관련 임상 수치 기반 변수

고혈압 관련 변수의 경우 연속형 형태의 2, 3차 평균 수축기, 이완기 혈압 측정값을 포함하였다. 당뇨병 관련 변수의 경우 연속형 형태의 당화혈색소와 공복혈당 수치를 포함하였으며, 빈혈 관련 변수의 경우 연속형 형태의 헤모글로빈 수치를 포함하였다. 이상지질혈증 관련 변수의 경우 연속형 형태의 HDL-콜레스테롤, 총콜레스테롤, LDL-콜레스테롤, 중성지방 수치를 포함하였으며, LDL-콜레스테롤 수치의 경우 한국지질·동맥경화학회의 이상지질혈증 진료지침(한국지질·동맥경화학회, 2022)에 따라 다음의 Friedewald 공식에 따라 산출된 값을 사용하였다. 모든 변수에 있어 결측·비해당·무응답인 경우 결측으로 처리하였다.

$$\text{LDL 콜레스테롤} = \text{총콜레스테롤} - \text{HDL 콜레스테롤} - \text{중성지방}/5$$



## 4. 분석 방법

본 연구에서는 국민건강영양조사 제8기와 9기의(2019-2023) 5개년도 자료의 건강 설문조사에 대해 SAS version 9.4 통계 프로그램과 Python 프로그램 (version 3.11.7)을 사용하여 분석하였다. 본 연구에서 분석은 총 다섯 단계로 구성되었다.

첫째, 연구 대상자의 일반적인 특성과 분포를 파악하기 위해 인구학적 요인, 건강 행태 요인, 만성질환 관련 변수들을 포함한 기술통계분석을 수행하였다. 범주형 변수는 빈도와 백분율(%)로, 연속형 변수는 평균값과 표준오차(Mean  $\pm$  S.E)로 제시하였다.

둘째, STOP-Bang 점수에 따른 폐쇄성수면무호흡증(Obstructive Sleep Apnea, OSA) 진단 여부의 분포를 확인하고, STOP-Bang 점수를 독립 변수로 하여 단변수 로지스틱 회귀분석(univariable logistic regression)을 수행하였다.

셋째, STOP-Bang 항목 외에 임상 및 건강 관련 변수를 포함하여 OSA 진단 여부에 영향을 미치는 요인을 확인하기 위해 다변수 로지스틱 회귀분석(multivariable logistic regression)을 수행하였다. 로지스틱 회귀 모델의 성능을 객관적으로 평가하기 위해 전체 데이터셋을 학습(train)과 검증(test) 데이터로 7:3 비율로 분할한 후, 각 데이터셋에서 모델을 적합시켰다. 분석 결과는 오즈비(OR)와 95% 신뢰구간(CI)으로 제시하였으며, 모델의 예측 성능은 ROC 곡선(Receiver Operating Characteristic curve)을 기반으로 한 AUC(Area Under the Curve) 값을 통해 정량화하였다.

넷째, 동일한 변수들을 입력 변수로 활용하여 심층신경망(Deep Neural Network, DNN) 기반 OSA 진단 예측모델을 구축하였다. DNN 모델의 성능을 객관적으로 평가하기 위해 전체 데이터셋을 학습(train)과 검증(test) 데이터로 7:3 비율로 분할한 후, 각 데이터셋에서 모델을 적합시켰다. SHAP(SHapley Additive exPlanations) 분석 기법을 적용하여 변수별 예측 기여도를 파악하고 성능 평가는 검증 데이터셋을 기준으로 AUC 및 정확도(Accuracy)를 중심으로 수행하였다.

다섯째, 최종적으로 로지스틱 회귀모델 결과와 DNN 모델의 예측 성능을 AUC를 기준으로 비교하였으며, 모델 간 성능 차이를 정량적으로 평가하였다.

SAS 프로그램을 사용한 분석 과정에서 먼저 국민건강영양조사(KNHANES)의 복합표본 설계조사 데이터의 특성을 고려하여 모든 분석에서 설계 가중치와 층화 및 군집 구조를 반영하였다. 연구 대상자의 일반적 특성은 OSA 진단 여부에 따라 기술통계를 수행하였으며, 연속형 변수는 복합표본설계를 반영한 회귀모형을 기반으로 가중 평균의 차이에 대한 t-통계량을 이용해 비교하였고, 범주형 변수는 Rao-Scott 카이제곱 검정(Rao-Scott chi-square test)을 통해 OSA 진단 여부 간 유의성을 평가하였다.

STOP-Bang 점수와 폐쇄성수면무호흡증(OSA) 진단 간의 관련성을 평가하기 위해 단변량 로지스틱 회귀분석을 실시하였으며, 이를 통해 선별도구로서의 STOP-Bang 점수의 유효성을 확인하고, 점수의 변화에 따른 OSA 진단 여부의 유의성을 통계적으로 검증하였다.

이어, STOP-Bang 항목 외에도 추가적인 임상 및 건강행태 관련 변수들을 함께 고려하여 OSA 진단 여부에 영향을 미치는 독립적인 요인을 파악하기 위해 다변량 로지스틱 회귀분석을 수행하였다. 이 과정에서 모형의 예측 성능을 평가하고 과적합을 방지하기 위해 전체 데이터를 학습용(train)과 검증용(test) 데이터로 7:3의 비율로 분할하여 분석하였으며, 70%의 학습 데이터를 활용해 모델을 훈련시키고, 나머지 30%의 검증 데이터에 모델을 적용하여 예측력을 평가하였다. 이를 통해 추정된 모형의 일반

화 가능성을 확보하고자 하였으며 단순한 상관이 아닌, 다른 변수의 영향을 통제된 상태에서의 독립적인 관련성을 평가하였다. 예측 성능은 ROC curve와 AUC 면적을 통해 산출하였다.

Python 프로그램을 사용한 분석에서는 STOP-Bang 점수로 활용되는 독립변수와 함께 진단 예측 정확도 개선을 위해 임상 변수들을 추가적으로 활용하여 다양한 건강 및 임상 지표들로부터 폐쇄성수면무호흡증(OSA) 진단으로 이어지게 되는 복잡한 상관관계를 심층신경망(DNN, Deep Neural Network)으로 모델링하였다. 모델은 입력층, 은닉층, 출력층으로 구성된 다층 퍼셉트론(MLP) 구조를 기반으로 하였다. 입력층(input layer), 출력층(output layer)과 그 사이 은닉층(hidden layer)들로 구성된 심층신경망 모델의 구성은 아래 (그림 2)에 나타내었다. 입력층에는 STOP-Bang 문항을 비롯하여 임상 및 건강 관련 총 21개의 예측 변수들이 포함되었으며, 은닉층은 총 3개로 구성하였다. DNN 모델에서 또한 모형의 예측 성능을 평가하고 과적합을 방지하기 위해 전체 데이터를 학습용(train)과 검증용(test) 데이터로 7:3의 비율로 분할하여 분석하였으며, 70%의 학습 데이터를 활용해 모델을 훈련시키고, 나머지 30%의 검증 데이터에 모델을 적용하여 예측력을 평가하였다.

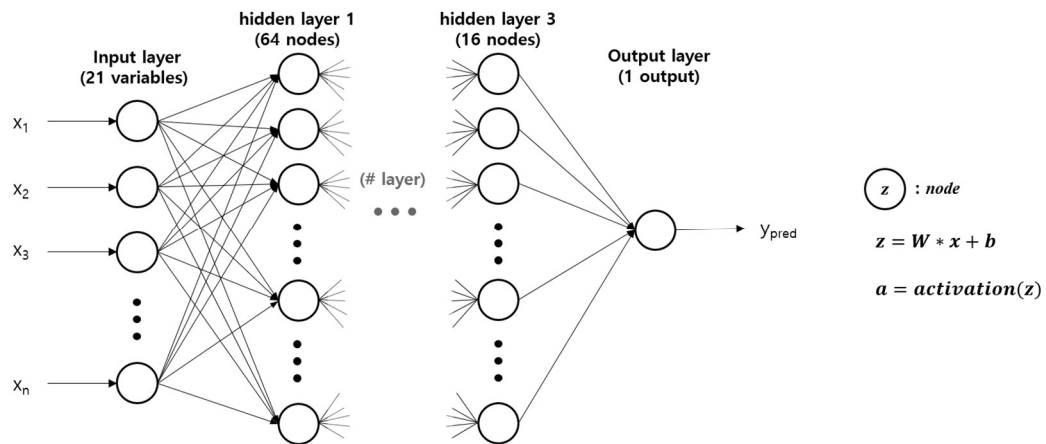


그림 2. 심층신경망(Deep Neural Network, DNN) 구조

## 가. 데이터 전처리

OSA 진단 예측 모델의 학습을 위해서는 데이터 전처리 과정이 필수적으로 선행되어야 한다. 데이터 전처리는 결측치 처리, 변수 유형에 따른 스케일링, 그리고 범주형 변수의 변환 등을 포함하며, 이를 통해 모델 학습의 효율성과 정확도를 높이하고자 하였다.

본 연구에서 사용된 입력 변수는 STOP-Bang 설문 점수를 기반으로 한 변수와 OSA 진단에 영향을 미칠 수 있는 변수로 구성되어 있다. 이러한 변수들은 연속형 변수와 범주형 변수로 구분될 수 있다.

연속형 변수의 경우, 변수마다 값의 범위와 단위가 다르기 때문에 스케일 차이로 인해 학습 과정이 편향될 가능성이 존재한다. 본 연구에서는 경사하강법 기반 학습의 수렴 속도를 개선하고 초기값에 대한 민감도를 줄여 학습 안정성을 높이기 위해 배치 정규화(Batch Normalization) 기법을 적용하였다(Ioffe et al., 2015). 이를 통해 각 변수의 값을 0~1 사이의 범위로 스케일링을 진행하여 편향된 학습을 방지하고 균일하게 학습이 이루어지도록 하였다.

범주형 변수의 경우, 데이터가 “Yes/No”와 같은 이진 분류 또는 다중 범주의 문자형으로 제공되기 때문에, 기계 학습 모델이 이를 처리할 수 있도록 수치형 데이터로 변환하는 과정이 필요하다. 본 연구에서는 범주형 변수의 각 수준을 이진 변수로 변환하여 표현하는 원-핫 인코딩(One-Hot Encoding) 기법을 적용하였다(Hancock et al., 2020). 이를 통해 각 범주를 이진 벡터로 변환하는 방법으로, 데이터의 해석 가능성과 모델의 학습 효율성을 동시에 확보하였다.

이와 같은 데이터 전처리 과정을 통해 연속형 변수와 범주형 변수의 특성을 효과적으로 반영하고, 모델 학습 시 발생할 수 있는 스케일링 불균형과 범주형 데이터의 처리 오류와 같은 잠재적 문제를 최소화하고자 하였으며, 이를 기반으로 최종적으로 OSA 진단 예측모델의 성능을 최적화할 수 있도록 설계하였다.

#### 나. 활성화 함수(Activation Function)과 손실 함수(Loss Function)

각 node에서 출력된 정보를 다음 layer로 전달하기 위해서는 적절한 변환이 필요하다. 이를 위해 본 연구에서는 데이터 간의 비선형적인 관계를 반영하기 위해 각 은닉층에서 ReLU(Rectified Linear Unit) 함수를 활성화 함수로 사용하였다. ReLU는 입력값이 0 이하일 때 0을 출력하고, 0보다 클 경우에는 입력값을 그대로 반환하는 방식으로, 계산이 간단하고 학습 속도를 향상시킬 수 있다는 장점이 있다.

최종 출력층에서는 이진 분류를 위해 Sigmoid 함수를 활성화 함수로 사용하여 0과 1 사이의 확률값을 출력하였다. Sigmoid 함수는 이진 분류 문제에서 널리 사용되며, 모델의 출력값이 예측 확률로 해석될 수 있다. 실제 분석에서는 이진 분류법을 적용하여 출력값이 0.5를 초과하면 OSA 진단(1), 0.5 이하이면 OSA 미진단(0)으로 분류하였다. 과적합 방지를 위해 각 은닉층에 Dropout 기법(비율 0.2~0.4)을 적용하고, 학습 안정성과 일반화를 도모하기 위해 Batch Normalization을 병행하였다.

손실 함수로는 Binary Cross-Entropy를 사용하였고, 최적화 알고리즘으로는 Adam Optimizer를 적용하였다. Binary Cross-Entropy는 이진 분류 문제에서 실제값과 예측값의 차이를 최소화하는 방향으로 학습이 진행되도록 한다. 구체적으로 손실값을 계산하고, 이를 최소화하기 위해 역전파(Backpropagation) 과정을 통해 출력층에서 입력층으로 가중치(Weight)와 편향(Bias)를 조정하는 방식으로 모델을 최적화하였다. 초기 학습률은 0.0001로 설정하였으며, 검증 손실이 개선되지 않을 경우, 학습률을 40% 감소시키는 ReduceLROnPlateau 기법을 활용하였다. 또한, EarlyStopping을 도입하여 검증 손실이 20 Epoch 동안 개선되지 않으면 학습을 조기 종료하고 가장 성능이 좋은 모델 가중치를 복원하였다. 모델은 총 150 Epoch동안 학습하였고 배치 크기는 32로 설정하였다.

그러나 OSA 진단 예측 모델의 경우, 실제 진단 여부 케이스가 전체 데이터셋에서 차지하는 비율이 상대적으로 낮아 데이터 불균형으로 인한 편향된 학습이 발생할 수 있다. 이를 보완하기 위해 가중치 함수(Weight Function)를 적용하여 실제 진단 케이

스에 대한 예측 오류에 더 큰 패널티를 부여하였다. 학습 데이터의 실제 클래스 분포를 반영하여 `class_weight='balanced'` 옵션을 적용하고, `compute_class_weight` 함수를 이용해 클래스별 가중치를 자동 산출하였다. 이를 통해 소수의 진단 케이스의 손실값에 상대적으로 더 큰 가중치를 부여함으로써, 전체 학습 과정에서 예측 편향을 줄이고 민감도(Sensitivity)의 향상을 유도하여 모델의 예측 성능과 정합성을 개선하고자 하였다. 또한, 분석에 포함된 샘플들은 각자의 가중치 변수를 부여받았으며, 모델 학습 시 해당 변수 값을 `sample_weight`로 적용하여 추정의 대표성과 정밀도를 높였다. 이를 통해 각 샘플의 실제 조사 비중을 반영한 학습이 가능해졌으며, 모집단의 특성을 보다 잘 반영하는 예측 모델을 구현하였다.

모델의 성능평가는 이진 분류 결과를 기준으로 실제 진단 여부와 비교하여 OSA 진단 예측률을 정확도(Accuracy)의 지표로 정량화하였다. 또한, 민감도(Sensitivity)와 특이도(Specificity)의 균형을 종합적으로 고려하기 위해 ROC curve(Receiver Operating Characteristic curve)를 작성하고, 그 면적(AUC, Area Under the Curve)을 산출하여 예측 모델의 판별력(discriminative ability)을 정량적으로 평가하였다. 이를 통해 제안된 예측 모델의 진단 성능을 평가하고, 그 유효성을 검증하였다.

### III. 결과

#### 1. 연구 대상자의 일반적 특성

본 연구는 총 연구 대상자 14,939명을 폐쇄성수면무호흡증(OSA) 진단 여부에 따라 일반적인 특성을 비교하였다 (표 1). 전체 대상자 중 OSA 진단군은 110명(0.7%), 비진단군은 14,829명(99.3%)이었다.

연령은 OSA 진단군이  $56.34 \pm 1.03$ 세, 비진단군이  $57.38 \pm 0.16$ 세로, 통계적으로 유의한 차이는 없었다( $p = 0.3076$ ). 성별 분포는 OSA 진단 여부에 따라 통계적으로 유의한 차이를 보였다( $p < 0.0001$ ). OSA 진단군에서는 남성이 80.0%, 여성이 20.0%였으며, 비진단군에서는 남성이 43.04%, 여성이 56.96%로 OSA 진단군에서 남성의 구성 비율이 비진단군에 비해 유의하게 높았다. 수축기 혈압은 OSA 진단군에서  $122.89 \pm 1.46$ mmHg로, 비진단군보다 다소 높게 나타났으나 통계적으로 유의하지 않았으며 ( $p = 0.2143$ ), 이완기 혈압은  $78.32 \pm 0.80$ mmHg로 비진단군보다 유의하게 높게 나타났다 ( $p = 0.0122$ ).

체질량지수(BMI)는 OSA 진단군이  $26.45 \pm 0.60$ kg/m<sup>2</sup>로 비진단군( $24.25 \pm 0.03$ kg/m<sup>2</sup>)보다 유의하게 높았으며( $p = 0.0003$ ), 목둘레( $38.27 \pm 0.41$ cm vs.  $35.21 \pm 0.03$ cm) 및 허리둘레( $93.07 \pm 1.50$ cm vs.  $85.27 \pm 0.10$ cm)도 OSA 진단군에서 유의하게 높았다( $p < 0.0001$ ).

혈당 및 지질 지표에서도 일부 유의한 차이가 확인되었다. 공복혈당은 진단군이  $104.89 \pm 2.16$ mg/dL로 다소 높았으나 유의하지 않았고( $p = 0.5868$ ), HbA1c도 진단 여부에 따른 차이는 유의하지 않았다( $p = 0.6646$ ). 반면, HDL-콜레스테롤 수치는 비진단군( $54.04 \pm 0.16$ mg/dL)에 비해 진단군에서 유의하게 낮았으며( $51.18 \pm 1.64$ mg/dL,  $p = 0.0144$ ), 총콜레스테롤( $179.45 \pm 5.01$  vs.  $191.93 \pm 0.42$ ,  $p =$



0.0132)은 비진단군이 진단군보다 높은 수치를 보이며 이는 통계적으로 유의하였다. LDL-콜레스테롤( $107.61 \pm 4.66$  vs.  $115.83 \pm 0.38$ ,  $p = 0.0784$ )과 중성지방( $137.88 \pm 8.48$  vs.  $138.30 \pm 1.18$ ,  $p=0.9603$ )은 통계적으로 유의한 차이가 없었다. 혈색소 수치는 OSA 진단군이  $14.84 \pm 0.17$ g/dL로 비진단군( $13.90 \pm 0.01$ g/dL)보다 유의하게 높았다( $p < 0.0001$ ).

알코올 섭취량은 진단 여부에 따른 유의한 차이가 없었으며( $0.13 \pm 0.014$  vs.  $0.12 \pm 0.001$ ,  $p = 0.6840$ ), 흡연 상태에서는 통계적으로 유의한 차이가 관찰되었다( $p < 0.0001$ ). OSA 진단군의 현재 흡연자 비율은 18.18%, 과거 흡연자는 52.73%, 비흡연자는 29.09%였으며, 비진단군에서는 각각 15.07%, 25.03%, 59.90%로 나타났다. OSA 진단군은 비진단군에 비해 과거 흡연자 비율이 높고, 비흡연자 비율은 낮은 경향을 보였다.

하루 평균 앉아 있는 시간은 OSA 진단군이  $9.71 \pm 0.34$ 시간, 비진단군이  $8.42 \pm 0.04$ 시간으로 진단군에서 유의하게 높았고( $p < 0.0001$ ), 주당 평균 수면 시간 역시 진단군( $47.73 \pm 0.75$ 시간)이 비진단군( $47.15 \pm 0.08$ 시간)보다 유의하게 길게 나타났다( $p = 0.4471$ ).

표 1. OSA 진단 여부에 따른 연구 대상자의 일반적 특성

Variables	OSA Diagnosis			P-value
	Total (N=14,939)	No (N=14,829)	Yes (N=110)	
	Mean $\pm$ S.E	Mean $\pm$ S.E	Mean $\pm$ S.E	
Age	57.37 $\pm$ 0.17	57.38 $\pm$ 0.16	56.34 $\pm$ 1.03	0.3076
Sex (N, %)				<.0001
Male	6470 (43.31%)	6382 (43.04%)	88 (80.00%)	
Female	8469 (56.69%)	8447 (56.96%)	22 (20.00%)	
Systolic blood pressure	121.09 $\pm$ 0.20	121.07 $\pm$ 0.20	122.89 $\pm$ 1.46	0.2143
Diastolic blood pressure	76.31 $\pm$ 0.11	76.31 $\pm$ 0.11	78.32 $\pm$ 0.80	0.0122
BMI (kg/m <sup>2</sup> )	24.27 $\pm$ 0.04	24.25 $\pm$ 0.03	26.45 $\pm$ 0.60	0.0003
Neck circumference	35.24 $\pm$ 0.03	35.21 $\pm$ 0.03	38.27 $\pm$ 0.41	<.0001
Waist measurement	85.34 $\pm$ 0.10	85.27 $\pm$ 0.10	93.07 $\pm$ 1.50	<.0001
Fasting blood sugar (mg/dL)	103.73 $\pm$ 0.25	103.72 $\pm$ 0.25	104.89 $\pm$ 2.16	0.5868
HbA1c	5.81 $\pm$ 0.01	5.81 $\pm$ 0.009	5.79 $\pm$ 0.06	0.6646
Total cholesterol	191.82 $\pm$ 0.42	191.93 $\pm$ 0.42	179.45 $\pm$ 5.01	0.0132
HDL-cholesterol	54.01 $\pm$ 0.16	54.04 $\pm$ 0.16	51.18 $\pm$ 1.64	0.0144
Triglyceride	138.30 $\pm$ 1.17	138.30 $\pm$ 1.18	137.88 $\pm$ 8.48	0.9603
LDL-cholesterol	115.76 $\pm$ 0.38	115.83 $\pm$ 0.38	107.61 $\pm$ 4.66	0.0784
Hemoglobin	13.91 $\pm$ 0.01	13.90 $\pm$ 0.01	14.84 $\pm$ 0.17	<.0001
Alcohol consumption	0.12 $\pm$ 0.001	0.12 $\pm$ 0.001	0.13 $\pm$ 0.014	0.6840
Smoking (N, %)				<.0001
Current smoker	2254 (15.09%)	2234 (15.07%)	20 (18.18%)	
Previous smoker	3770 (25.24%)	3712 (25.03%)	58 (52.73%)	
Non-smoker	8915 (59.68%)	8883 (59.90%)	32 (29.09%)	
Total sitting time per day (hours)	8.43 $\pm$ 0.04	8.42 $\pm$ 0.04	9.71 $\pm$ 0.34	0.0001
Total sleeping time per week (hours)	47.16 $\pm$ 0.08	47.15 $\pm$ 0.08	47.73 $\pm$ 0.75	0.4471

표 2는 연구대상자의 STOP-Bang 설문 항목에 대한 응답 결과를 OSA 진단 여부에 따라 비교한 결과이다. 코골이(Snoring)는 전체 대상자의 19.10%(2,854명)에서 관찰되었으며, OSA 진단군에서는 64.55%로, 비진단군(18.77%)보다 유의하게 높은 비율을 보였다( $p < 0.0001$ ). 피로감(Tiredness)은 OSA 진단군에서 49.09%로, 비진단군에서는 29.76%로 진단군에서 유의하게 높았다( $p < 0.0001$ ). 수면 중 무호흡 목격(Observer) 여부는 진단군에서 74.55%, 비진단군에서 8.13%로, 진단군에서 매우 높은 비율로 관찰되었으며 통계적으로 유의하였다( $p < 0.0001$ ).

고혈압 병력 여부는 OSA 진단군에서 42.73%, 비진단군에서 33.74%로 유의한 차이를 보였으며( $p = 0.0472$ ), BMI  $\geq 30\text{kg/m}^2$  는 비율은 진단군이 16.36%, 비진단군이 5.36%로 진단군에서 유의하게 높았다( $p < 0.0001$ ). 연령 50세 초과 여부는 전체의 74.05%였으며, OSA 진단군 70.0%, 비진단군 74.08%로 집단 간 유의한 차이는 관찰되지 않았다( $p = 0.3310$ ). 목둘레  $\geq 40\text{cm}$ 는 진단군 23.64%, 비진단군 7.18%로, 진단군에서 유의하게 높았으며( $p < 0.0001$ ), 성별 구성은 OSA 진단군의 80.0%가 남성으로, 비진단군(43.04%)보다 유의하게 높은 비율을 보였다( $p < 0.0001$ ).

STOP-Bang 점수에 따른 위험군 분포에서는, 고위험군(5점 이상)에 속한 대상자는 전체의 6.04%(902명)였으며, OSA 진단군에서는 42.73%가 고위험군에 해당해, 비진단군의 5.77%에 비해 유의하게 높은 분포를 보였다( $p < 0.0001$ ).

표 2. OSA 진단 여부에 따른 STOP-Bang 설문 항목 분포

STOP-Bang variables	OSA Diagnosis			P-value
	Total (N=14,939)	No (N=14,829)	Yes (N=110)	
	N (%)	N (%)	N (%)	
<b>Snoring</b>				<.0001
No	12,085 (80.90%)	12,046 (81.23%)	39 (35.45%)	
Yes	2,854 (19.10%)	2,783 (18.77%)	71 (64.55%)	
<b>Tiredness</b>				<.0001
No	10,472 (70.10%)	10,416 (70.24%)	56(50.91%)	
Yes	4,467 (29.90%)	4,413(29.76%)	54(49.09%)	
<b>Observer</b>				<.0001
No	13,652 (91.38%)	13,624 (91.87%)	28 (25.45%)	
Yes	1,287 (8.62%)	1,205 (8.13%)	82 (74.55%)	
<b>Hypertension</b>				0.0472
No	9,888 (66.19%)	9,825 (66.26%)	63 (57.27%)	
Yes	5,051(33.81%)	5,004 (33.74%)	47 (42.73%)	
<b>BMI &gt; 30kg/m<sup>2</sup></b>				<.0001
No	14,126 (94.56%)	14,034 (94.64%)	92 (83.64%)	
Yes	813 (5.44%)	795 (5.36%)	18 (16.36%)	
<b>Age &gt; 50</b>				0.3310
No	3,877 (25.95%)	3,844 (25.92%)	33 (30.00%)	
Yes	11,062 (74.05%)	10,985 (74.08%)	77 (70.00%)	
<b>Neck circumference &gt; 40cm</b>				<.0001
No	13,848 (92.70%)	13,764 (92.82%)	84 (76.36%)	
Yes	1,091 (7.30%)	1,065 (7.18%)	26 (23.64%)	
<b>Gender - male</b>				<.0001
No	8,469 (56.69%)	8,447 (56.96%)	22 (20.00%)	
Yes	6,470 (43.31%)	6,382 (43.04%)	88 (80.00%)	
<b>STOP-Bang risk</b>				<.0001
Low risk	9,344 (65.55%)	9,237 (62.90%)	17 (15.45%)	
Intermediate risk	4,693 (31.41%)	4,647 (31.33%)	46 (41.82%)	
High risk	902 (6.04%)	855 (5.77%)	47 (42.73%)	

## 2. STOP-Bang 점수와 OSA 진단 분포

본 연구에서는 STOP-Bang 점수에 따른 OSA 진단 분포를 확인하고자, 설문 점수를 기준으로 저위험군(Low risk), 중위험군(Intermediate risk), 고위험군(High risk)으로 구분하여 각 군별 OSA 진단률을 분석하였다(표 3).

저위험군(0-2점)에서는 OSA 진단률이 매우 낮은 수준으로 나타났다. 구체적으로, 0점과 1점의 경우 각각 1명(0.08%)과 3명(0.08%)만이 진단되었으며, 2점에서도 전체 4,510명 중 18명(0.40%)만이 OSA 진단을 받은 것으로 확인되었다. 중위험군(3-4점)의 진단률은 점수 증가에 따라 소폭 상승하는 양상을 보였다. 3점에서는 진단률이 0.59%(19/3,226), 4점에서는 1.84%(27/1,467)로 나타났다. 고위험군(High risk)에서는 상대적으로 높은 OSA 진단률을 보였으며 5점의 경우 601명 중 OSA 진단을 받은 사례가 22명(3.66%), 6점의 경우 228명 중 17명(7.46%), 7점의 경우 63명 중 7명(11.11%), 마지막으로 8점의 경우 10명 중 1명(10.00%)이 확인되었다.

전체적으로 STOP-Bang 점수별 OSA 진단율의 추이를 보았을 때, 점수가 높을수록 진단율이 뚜렷하게 증가하는 경향이 확인되었다. 특히, STOP-Bang 점수가 0 - 1점인 군에서는 0.1% 미만의 진단율을 보이는 반면, 7 - 8점인 군에서는 10% 이상의 진단율이 나타나는 것으로 보아, 이는 해당 점수가 OSA 고위험군을 선별하는 데 있어 유의미한 민감도를 지닌 선별 도구로 작용함을 시사한다.

그러나 진단률이 증가함에도 불구하고 STOP-Bang 점수만으로는 개별 진단에 있어서 특이도가 부족함을 확인할 수 있다. 예를 들어, 고위험군(5점 이상)에서조차 진단률은 3.66%~11.11% 수준에 머물고 있으며, 이는 고위험군 내 대다수가 실제로는 진단되지 않은 경우임을 의미한다. 또한, 저위험군에서도 OSA 진단 사례가 다수 존재하는 점에서 STOP-Bang 점수 기반 분류의 한계를 보이는 결과이다. 이와 같은 점에서, STOP-Bang 설문만으로는 OSA 진단의 민감도와 특이도 모두를 충족하기 어려우며, 보다 정밀한 진단 예측을 위해 추가적인 임상 및 건강 정보를 반영한 예측 모델의 필요성이 강조된다.

표 3. STOP-Bang 점수대별 OSA 진단 분포

STOP-Bang Risk Group	STOP-Bang Score	OSA No	OSA Yes	Total	Diagnosis Rate (%)
Low risk	0	1,193	1	1,194	0.08%
	1	3,642	3	3,645	0.08%
	2	4,492	18	4,510	0.40%
Intermediate risk	3	3,207	19	3,226	0.59%
	4	1,440	27	1,467	1.84%
High risk	5	579	22	601	3.66%
	6	211	17	228	7.46%
	7	56	7	63	11.11%
	8	9	1	10	10.00%
Total		14,829	110	14,939	-

### 3. STOP-Bang 점수와 OSA 진단의 단변수 로지스틱 회귀 분석 결과

STOP-Bang 점수가 폐쇄성수면무호흡증(OSA) 진단 여부에 미치는 영향을 평가하기 위해 단변수 로지스틱 회귀분석(Univariable Logistic Regression)을 수행하였다. 분석에는 총 14,939명의 대상자가 포함되었으며, 이 중 OSA로 진단된 경우는 110명 (0.74%)이었다.

분석 결과, STOP-Bang 점수는 OSA 진단 가능성과 유의한 양의 연관성을 보였으며( $p < 0.0001$ ), 점수가 1점 증가할수록 OSA 진단 가능성이 통계적으로 유의하게 증가하는 것으로 나타났다(표4).

모형의 예측력을 평가한 ROC 곡선의 AUC는 0.8268로, STOP-Bang 점수 단일 변수만으로도 비교적 높은 예측 정확도를 보였다(그림 3).

표 4. STOP-Bang 점수와 OSA 진단의 단변수 로지스틱 회귀분석 결과

Variables	OR	95% CI	P-value
STOP-Bang score	2.31	2.02 - 2.65	<.0001



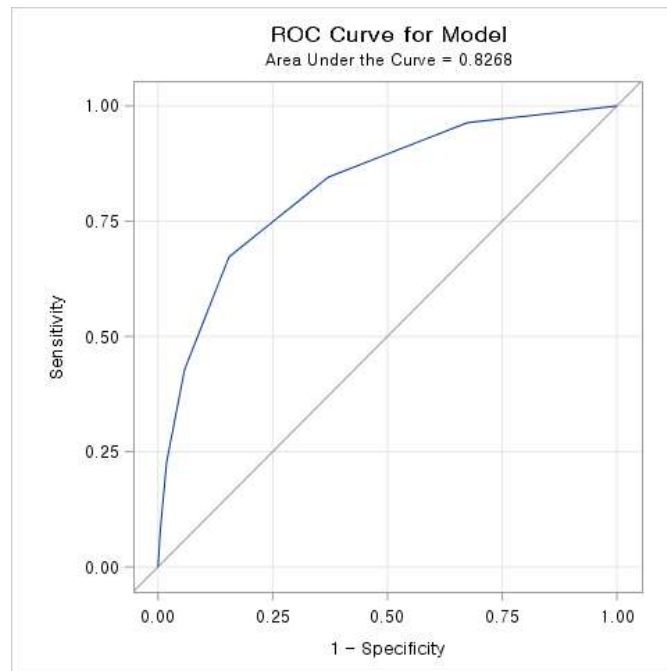


그림 3. STOP-Bang 점수 기반 OSA 진단 예측의 ROC 곡선

#### 4. 건강지표를 포함한 다변수 로지스틱 회귀 분석 결과

STOP-Bang 점수 항목 외에 추가적인 임상 및 건강 관련 변수를 포함하여 OSA 진단 여부에 대한 예측력을 평가하기 위해 다변수 로지스틱 회귀분석(Multivariable Logistic Regression)을 실시하였다. 분석에서의 신뢰도를 높이고 과적합 가능성을 최소화하기 위해 전체 데이터셋을 학습용(70%)과 검증용(30%)으로 나누어 모형을 구축하였다.

학습용 데이터셋(N = 10,565)을 기반으로 수행한 다중 로지스틱 회귀분석 결과 (표 4), 폐쇄성수면무호흡증(OSA) 진단 여부에 통계적으로 유의한 영향을 미치는 변수는 무호흡 목격 여부(Observer), 코골이(Snoring), 음주 여부(Alcohol consumption), 하루 평균 앉아있는 시간(Total sitting time)이었다.

먼저, 수면 중 무호흡을 목격한 대상자는 그렇지 않은 대상자에 비해 OSA 진단 오즈가 20.68배로, 통계적으로 매우 유의하였다( $p < 0.0001$ ). 코골이 항목 또한 오즈비 2.29(95% CI = 1.09-4.79,  $p = 0.0287$ )로 통계적으로 유의하였으나, 반면에 피로감(Tiredness)은 오즈비 1.51로 높은 수치를 보였으나,  $p = 0.0993$ 으로 유의수준에는 도달하지 못하였다. 성별의 경우에도 남성은 여성보다 오즈가 1.79배 높게 나타났으나,  $p = 0.3474$ 로 통계적으로 유의하지 않았다.

건강 및 생활 관련 변수 중 하루 앉아서 보내는 시간은 오즈비 1.14(95% CI = 1.06-1.22,  $p = 0.0003$ )로 통계적으로 유의하였으며, 음주 여부는 오즈비 0.16(95% CI = 0.04-0.67,  $p = 0.0130$ )로 통계적으로 유의한 부적 관련성을 나타냈다. 흡연 변수의 경우, 현재 흡연자는 비흡연자에 비해 오즈가 0.70배로 낮았으나 유의하지 않았고( $p = 0.4276$ ), 과거 흡연자와도 유의한 차이를 보이지 않았다( $p = 0.8934$ ).

혈액 및 생화학 지표 중에서는 헤모글로빈 수치(OR = 1.19,  $p = 0.1511$ )가 비교적 높은 오즈비를 보였으나 통계적 유의성은 확인되지 않았으며, HDL-콜레스테롤 역시 OR = 1.02,  $p = 0.3425$ 로 유의하지 않았다. 당화혈색소(HbA1c), 공복혈당, 총콜레스테롤, 중성지방, LDL-콜레스테롤 등도 통계적으로 유의하지 않았다.

이외에 연령, 수축기/이완기 혈압, 체질량지수, 목둘레, 허리둘레, 수면시간 등은 OSA 진단과 통계적으로 유의한 관련성을 보이지 않았다.

학습용 데이터로 구축한 로지스틱 회귀모델을 검증용(test) 데이터에 적용하여 예측 성능을 평가한 결과, ROC 분석에서 AUC(Area Under the Curve)는 0.8660으로 나타났다. 이는 본 모델이 OSA 진단 여부를 비교적 우수하게 분류함을 알 수 있다 (그림4).

표 5. 학습용 데이터셋에서의 OSA 진단과 관련된 건강 및 임상 요인의 다변수 로지스틱 회귀분석 결과

Variables	OR	95% CI	P-value
<b>Age</b>	0.99	0.96 - 1.03	0.6278
<b>Sex</b>			
Male	1.79	0.58 - 6.03	0.3474
Female	1	Reference	
<b>Snoring</b>			
No	1	Reference	
Yes	2.29	1.09 - 4.79	0.0287
<b>Tiredness</b>			
No	1	Reference	
Yes	1.51	0.93 - 2.46	0.0993
<b>Observer</b>			
No	1	Reference	
Yes	20.68	9.69 - 44.14	<.0001
<b>Systolic blood pressure</b>	1.03	1.00 - 1.06	0.0944
<b>Diastolic blood pressure</b>	0.96	0.94 - 1.03	0.4097
<b>BMI (kg/m<sup>2</sup>)</b>	1.02	0.87 - 1.21	0.7852
<b>Neck circumstance</b>	0.95	0.78 - 1.15	0.6049
<b>Waist measurement</b>	1.03	0.96 - 1.10	0.4398
<b>Fasting blood sugar (mg/dL)</b>	1.00	0.98 - 1.02	0.7053
<b>HbA1c</b>	0.61	0.35 - 1.06	0.0784

(계속)

표 5. 학습용 데이터셋에서의 OSA 진단과 관련된 건강 및 임상 요인의 다변수 로지스틱 회귀분석 결과 (계속)

Variables	OR	95% CI	P-value
Total cholesterol	0.99	0.95 - 1.04	0.6691
HDL-cholesterol	1.02	0.98 - 1.07	0.3425
Triglyceride	1.00	0.99 - 1.01	0.8302
LDL-cholesterol	1.00	0.96 - 1.04	0.9457
Hemoglobin	1.19	0.94 - 1.51	0.1511
Alcohol consumption	0.16	0.04 - 0.67	0.0130
Smoking			
Current smoker	0.70	0.28 - 1.71	0.4276
Previous smoker	1.06	0.45 - 2.50	0.8394
Non-smoker	1	Reference	
Total sitting time per day (hours)	1.14	1.06 - 1.22	0.0003
Total sleeping time per week (hours)	1.02	0.99 - 1.04	0.1518

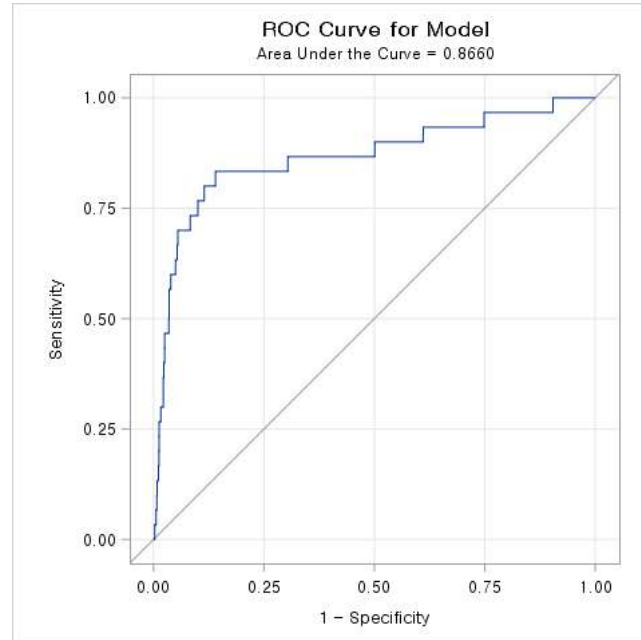


그림 4. 검증용 데이터셋에서의 건강 및 임상 요인 기반 OSA 진단 예측 ROC 곡선

## 5. DNN 모델 기반 OSA 진단 예측모델 성능평가

폐쇄성수면무호흡증(OSA) 진단 예측을 위해 변수 간의 복잡한 상호작용과 비선형성을 효과적으로 반영할 수 있는 심층신경망(DNN) 모델을 구성하고, 그 성능을 평가하였다.

학습 결과는 Training Loss와 Validation Loss를 비교하여 평가하였다. Loss 그래프에서 확인한 바와 같이 초기 학습 단계에서 Validation Loss가 급격히 감소한 후, 약 40 Epoch 이후부터 안정화되는 경향을 보였다. Training Loss는 지속적으로 감소하여 최종적으로 약 0.3 수준으로 수렴하였으며, Validation Loss는 약 0.4 수준에서 수렴하였다 (그림 5). 본 학습모델의 실제 진단 여부와 예측 확률을 이용하여 최종 예측 성능을 평가한 ROC 곡선의 AUC는 0.9020으로 확인되었다 (그림 6). 이진 분류법을 적용하여 출력값이 0.5를 초과하면 OSA 진단(1), 0.5 이하이면 OSA 미진단(0)으로 분류하여 모델의 정확도(Accuracy) 성능을 평가한 결과, 최종 OSA 진단 정확도는 약 94.85%를 기록하며 우수한 예측 성능을 보였다.



그림 5. OSA 진단 예측 DNN 모델의 학습 손실 변화 추이



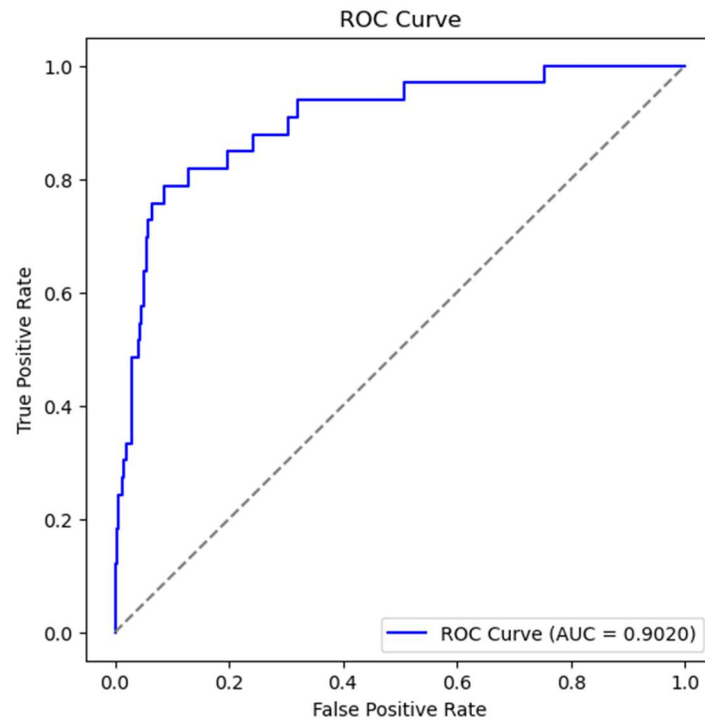


그림 6. DNN 모델 기반 OSA 진단 예측의 ROC 곡선

본 연구에서는 심층신경망(DNN) 기반 OSA 진단 예측모델에 포함된 주요 임상 및 건강 관련 입력 변수들이 예측에 기여하는 정도를 시각화하고 해석하기 위하여 SHAP(SHapley Additive exPlanations) 분석을 이어 수행하였다. SHAP 값은 각 변수의 OSA 진단 확률에 대한 기여도를 정량적으로 제시하며, SHAP summary plot 구현을 통해 전체 변수의 영향력을 시각화하였다(그림 7).

분석 결과, 수면 중 무호흡 목격자 변수가 모델에서 가장 큰 영향을 미치는 변수로 나타났으며, 목격자가 있을수록 해당 변수의 SHAP 값은 전반적으로 양의 방향으로 분포하였다. 이는 무호흡 목격 경험에 있는 경우 OSA 진단 확률을 높이는 방향으로 작용함을 의미하며, 이는 기존의 다변수 로지스틱 회귀 분석에서도 오즈비 20.68 (95% CI: 9.69-44.14,  $p < 0.0001$ )로 확인되어, 가장 강력한 예측 인자로 작용하였다.

다음으로 영향력이 큰 변수로는 앓아있는 시간, 성별, 총콜레스테롤, 수면 시간 등이 확인되었다. 특히 생활 습관과 관련된 건강 변수인 앓아있는 시간은 SHAP 분석에서 비교적 높은 위치를 차지하며 진단 확률에 기여하는 변수로 나타났다. SHAP 값은 양의 방향을 보였고, 이는 앓아있는 시간이 길수록 OSA 진단 확률이 증가함을 시사한다. 이는 다변수 로지스틱 회귀 분석에서 오즈비가 1.14 (95% CI: 1.06-1.22,  $p = 0.0003$ )로 유의한 영향력을 보인 것과 방향성이 동일하다.

성별의 경우 로지스틱 회귀분석에서는 통계적으로 유의하지는 않았으나 (OR = 1.79, 95% CI: 0.58-6.03,  $p = 0.3474$ ), 남성의 SHAP 값이 양의 방향으로 분포하여, 남성이 여성보다 OSA 진단 확률이 높게 나타나는 방향성을 보였다. 이는 SHAP 분석에서도 남성의 SHAP 값이 양의 방향으로 분포하며 예측 확률 증가에 기여하는 경향과는 일치하였다.

수면시간의 경우 SHAP 분석에서 높은 수치일수록 OSA 진단 확률을 높이는데 기여하는 양상을 보이나, 다변수 로지스틱 회귀 분석 결과에서는 통계적 유의수준에 도달하지 않았다 (OR = 1.02, 95% CI: 0.99-1.04,  $p = 0.1518$ ).

한편, 임상적 기대와 상반되는 비직관적인 결과가 관찰되어 로지스틱 회귀 분석 결과와의 비교를 요하는 일부 변수들도 확인되었다. 이 중 총콜레스테롤의 경우 SHAP

분석에서 높은 수치일수록 OSA 진단 확률을 낮추는데 기여하는 양상을 보이나, 다변수 로지스틱 회귀 분석에서도 유사한 방향성을 보이되 통계적으로 유의하지 않았다 ( $OR = 0.99$ , 95% CI: 0.95-1.04,  $p = 0.6691$ ). 목둘레의 경우 SHAP 분석에서는 목둘레가 클수록 진단 확률이 낮아지는 방향성을 보였으나, 이는 다변수 로지스틱 회귀 분석 결과 동일한 방향성을 보이나 통계적으로 유의하지 않았다 ( $OR = 0.95$ , 95% CI: 0.78-1.15,  $p = 0.6049$ ). 당화혈색소의 경우에도 마찬가지로 SHAP 상에서 진단 확률에 음의 영향을 주는 경향을 보였으나, 이는 다변수 로지스틱 분석에서 방향성은 일치하나 통계적으로 유의하지 않았다 ( $OR = 0.61$ , 95% CI: 0.35-1.06,  $p = 0.0784$ ).

반면, 임상적으로 중요한 변수로 알려진 BMI, 나이, 혈압, 지질 수치 등은 SHAP 분석에서 변수 중요도가 낮았으며, 로지스틱 회귀 분석 결과에서도 통계적으로 영향이 없었던 점에서 일관된 결과를 보여준다.

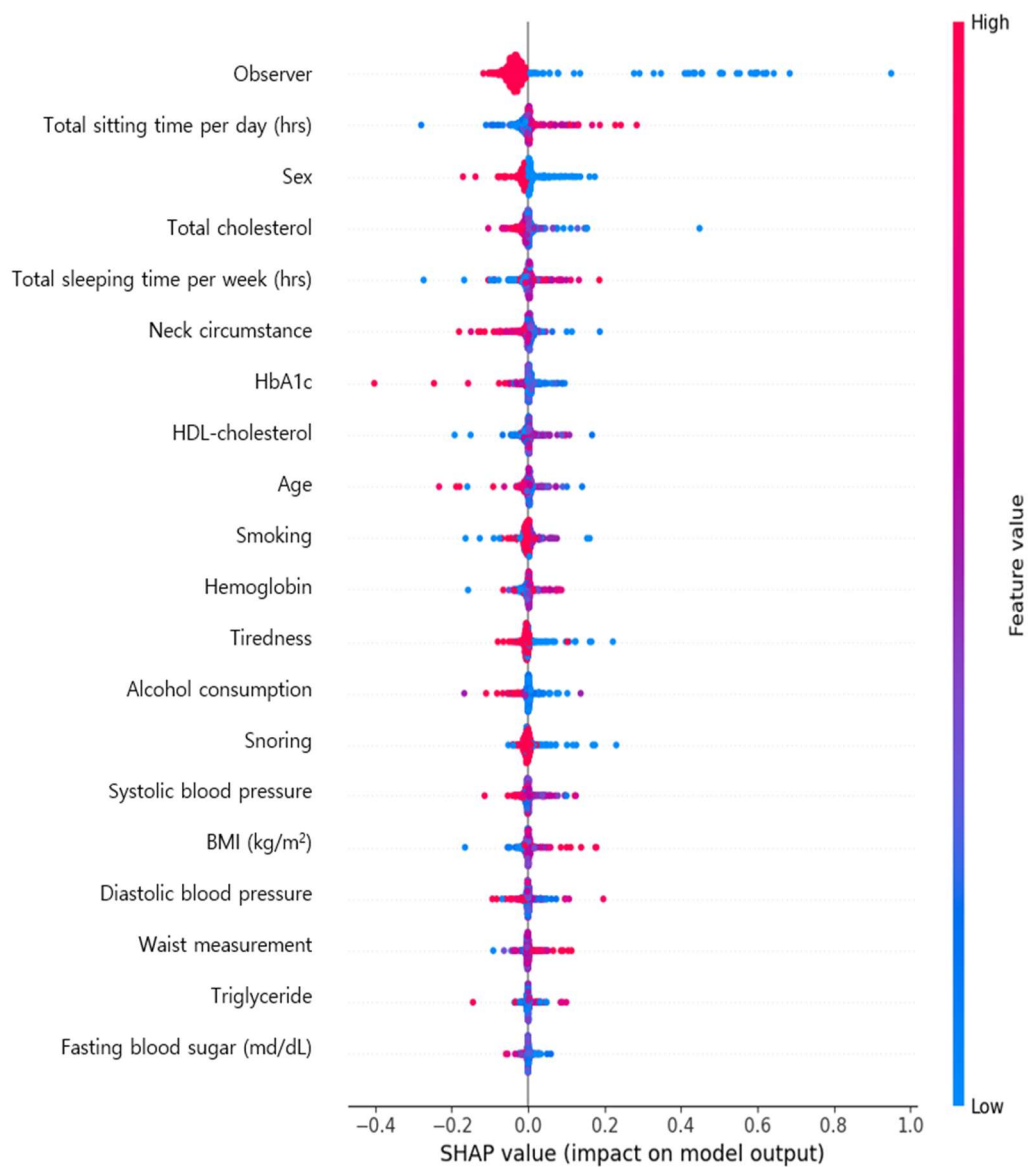


그림 7. OSA 진단 예측 DNN 모델에 대한 SHAP 요약 시각화 결과

## 6. 로지스틱 회귀분석 모형과 DNN 모델 간 성능 비교

본 연구에서는 폐쇄성수면무호흡증(OSA)의 진단 여부를 분류하기 위한 세 가지 분류 기법을 비교하였다. 첫 번째로 STOP-Bang 점수 단일 항목을 독립변수로 한 단변수 로지스틱 회귀 모형, 두 번째로 건강 및 임상 지표를 포함한 다변수 로지스틱 회귀 모형, 그리고 마지막 세 번째로 다변수 로지스틱 회귀 모형과 동일한 변수들을 활용한 심층신경망(DNN) 기반 분류 모형을 구성하였다. 각 모형의 분류 성능은 ROC curve 하의 면적(AUC)을 기준으로 평가하였다 (표 6).

그 결과, 단변수 로지스틱 회귀 분석에서는 AUC가 0.8268로 나타났다. 이후 다양한 임상 및 건강 관련 변수를 포함하여 학습 데이터셋으로 회귀계수를 추정한 다변수 로지스틱 회귀 분석 결과, 테스트 데이터셋에 적용한 AUC는 0.8660으로 확인되었다. 동일한 변수들을 입력으로 사용한 DNN 분류 모형의 경우, 테스트 데이터셋에 적용한 AUC는 0.9020으로 가장 높은 값을 나타냈다.

이처럼 세 분류 모형의 AUC는 각각 0.8268, 0.8660, 0.9020으로, 설문 기반 단변수 분류보다 다변수 접근 방식에서 향상된 분류 성능이 나타났으며, DNN 기반 모형에서 가장 높은 분류 성능을 보였다.

표 6. 로지스틱 회귀 모형과 DNN 모델의 ROC 곡선 하의 면적(AUC) 비교

Classification Model	Set of Variables	ROC AUC
Univariable Logistic Regression	STOP-Bang Score	0.8268
Multivariable Logistic Regression (Test Set)	STOP-Bang Items and Clinical & Health Indicators	0.8660
Deep Neural Network (Test Set)	STOP-Bang Items and Clinical & Health Indicators	0.9020

## IV. 고찰

본 연구는 국민건강영양조사(KNHANES) 제8~9기(2019-2023년) 데이터를 활용하여, STOP-Bang 설문지와 추가적인 임상 변수를 기반으로 폐쇄성수면무호흡증(Obstructive Sleep Apnea, OSA)의 진단 예측모델을 구축하고, 로지스틱 회귀분석과 심층신경망(Deep Neural Network, DNN) 모델의 성능을 비교하였다. 이를 통해 기존 STOP-Bang 설문지의 한계를 보완하고, 보다 정확한 예측모델의 가능성을 탐색하였다.

먼저 STOP-Bang 점수를 기반으로 폐쇄성수면무호흡증(OSA)의 진단 여부를 예측하고자 하였으며, STOP-Bang 점수 단일 항목이 가지는 선별 도구로서의 성능을 평가하였다. STOP-Bang 점수는 기존 연구에서도 입증된 바와 같이 OSA 고위험군을 신속하게 선별하는 데 유용한 도구로 알려져 있으며, 실제로 본 연구에서도 점수가 높아질수록 OSA 진단율이 증가하는 뚜렷한 양상이 확인되었다.

한편, 복수의 위험 요인을 조합하여 OSA 고위험군을 선별하는 도구로 사용되는 STOP-Bang 설문지는 설문 점수를 기반으로 질환 위험을 정량화한다는 점에서, 각 항목 간의 내적 일관성을 확인하는 것이 중요하다. 이에 따라 산출한 Cronbach's  $\alpha$  계수는 원자료 기준 0.39, 표준화된 변수 기준 0.44로 나타났다. 이는 STOP-Bang 설문의 모든 항목이 동일한 개념을 측정하는 척도라기 보다는 상이한 임상적 위험요인을 조합한 체크리스트형 도구임을 고려할 때, 도구의 구조적 특성을 반영한 결과로 해석할 수 있다.

이어서 구체적인 연구 결과에 따르면, STOP-Bang 점수 0점과 1점에서는 각각 0.08%의 낮은 진단율을 보였으며, 2점에서는 0.40%, 4점에서는 1.84%, 7점에서는 11.11%의 진단율로 점수가 증가함에 따라 진단율이 증가하는 양상을 보였다. 특히, 고위험군(5점 이상)에 해당하는 대상자에서 진단율은 최대 11.11%에 도달하였으며, 이는 저위험군(0~2점)의 진단율 대비 약 100배 이상 높은 수준이다. 이는 STOP-Bang 점수가 OSA 고위험군 선별 도구로서 높은 민감도를 가진 도구임을 뒷받침하는 결과로 즉, 점수가

높을수록 OSA에 대한 진단 가능성이 증가함을 보이며, 이는 선별 검사로서 잠재적 고위험군을 식별하는 데 유용하다는 점을 시사한다(Farney et al., 2011, Duarte et al., 2017).

그러나, 특이도 관점에서 살펴보면 STOP-Bang 점수만으로는 개별 진단에 대한 정확도는 제한적임을 알 수 있다. 예를 들어, 고위험군(5점 이상) 대상자 중 실제 OSA로 진단된 비율은 10% 내외에 불과하며, 이는 전체 고위험군의 약 90% 이상이 실제 OSA가 아님에도 불구하고 고위험군으로 분류된다는 점에서 특이도가 낮음을 의미한다. 또한, 저위험군 내에서도 소수이지만 OSA 진단 사례가 존재하는 점은 민감도의 한계 역시 시사한다. 이는 STOP-Bang 점수가 질환 유무를 이분법적으로 판단하기에는 한계가 있으며, 진단을 위한 최종적 지표보다는 선별 목적의 도구로 활용되어야 함을 강조한다.

따라서 STOP-Bang 점수는 전반적인 경향성을 포착하는 데 유용한 지표이나, 개별 환자 수준의 정확한 진단을 위해서는 다양한 임상 지표와 건강 정보를 결합한 복합 예측모델의 보완이 필수적임을 확인할 수 있었다. 본 연구의 후속 분석에서는 이러한 한계를 보완하고자, STOP-Bang 항목 외에 임상 및 건강 정보를 포함한 로지스틱 회귀 분석 및 딥러닝 기반 예측모델을 구성하여 진단 예측 성능을 비교하였다.

먼저, 단변수 로지스틱 회귀분석 결과에서 STOP-Bang 점수는 OSA 진단 여부에 대해 통계적으로 유의한 양의 상관관계를 나타냈으며( $p < 0.0001$ ), 예측 모델의 ROC 곡선(Receiver Operating Characteristic curve)을 통해 계산된 AUC(Area Under the Curve)는 0.8268로 나타났다. 이는 단일 변수 기반의 선별 모델로서 비교적 우수한 예측 정확도를 확보하고 있음을 의미한다.

그러나 이러한 진단력은 OSA의 개별 진단을 위한 정확한 예측 도구로서의 역할에는 다소 한계가 있음을 동시에 보여준다. 고위험군으로 분류된 대상자 중 실제 진단율은 10% 전후에 불과하였으며, 저위험군에서도 다수의 진단 사례가 발생하였다. 특히 2점 이하의 점수를 받은 대상자군에서도 총 22명의 OSA 진단 사례가 관찰되었는데, 이는 STOP-Bang 점수만을 기반으로 위험군을 단순히 분류할 경우 위음성(false negative)



과 위양성(false positive)의 발생 가능성이 모두 존재함을 보여준다.

이러한 경향은 기존 선행연구에서 제기된 STOP-Bang 설문지의 구조적 한계와도 일치한다. Chung 등(2014)의 연구를 포함한 다양한 후속 연구들에서는 STOP-Bang 설문지 높은 민감도에 비해 특이도가 낮아 과도한 고위험군 선별로 이어질 수 있다는 점을 반복적으로 지적하고 있다. 특히 BMI나 목둘레 등 일부 항목은 서구 기준에 맞춰 설정되어 있어, 아시아 인구에서는 민감도 및 예측력이 떨어질 수 있다는 보고도 있다(Waseem et al., 2021). 본 연구에서도  $BMI \geq 30\text{kg/m}^2$  이상에 해당하는 대상자는 전체의 5.4%에 불과하였으며, 해당 항목은 OSA 진단과 유의한 연관성을 보이지 않았다.

더 나아가 본 연구에서는 STOP-Bang 점수 외에도 다양한 임상 및 건강행태 지표를 포함하여 OSA 진단 여부를 예측하는 다변수 로지스틱 회귀 분석을 실시하였다. 이를 통해 단일 설문 점수 기반 선별의 한계를 보완하고, 복합적인 위험 요인들의 영향력을 종합적으로 평가하고자 하였다. 분석 시 학습 데이터에 과도하게 적합되는 과적합(overfitting) 가능성을 방지하고 모델의 일반화 가능성을 확보하기 위해, 전체 데이터를 7:3 비율로 학습용과 검증용 데이터셋으로 분할하여 분석을 진행하였다.

분석 결과, 전체 변수 중 ‘수면 중 무호흡 목격 여부(Observed apnea)’가 가장 높은 예측력을 보이는 변수로 확인되었으며, 이 항목에 ‘예’라고 응답한 대상자는 그렇지 않은 대상자에 비해 OSA 진단 오즈가 약 20.68배(95% CI: 9.69-44.14) 높은 것으로 나타났다. 이는 본 연구에서 단일 점수보다는 항목별 정보가 임상적으로 더욱 중요한 예측 인자를 내포할 수 있음을 보여준다.

무호흡 목격 항목이 갖는 높은 예측력은 기존 선행연구에서도 반복적으로 입증된 바 있다. 무호흡 목격 여부는 외부 관찰자에 의해 수면 중 호흡 정지가 확인된 직접적 증상에 해당하며, 이로 인해 OSA의 주요 진단 기준인 AHI(Apnea-Hypopnea Index)의 객관적 증가와 높은 상관성을 갖는 것으로 알려져 있다. 특히 Shin 등(2021)의 연구에서는 STOP-Bang의 8개 항목 중에서도 무호흡 목격 항목이 가장 높은 특이도와 민감도를 동시에 나타내는 핵심 인자로 확인된 바 있으며, 이는 본 연구의 결과와 동일한 방향성을 보여준다.

반면, 기존에 OSA의 주요 위험 요인으로 제시되어 온 BMI, 목둘레, 수축기 및 이완기 혈압 등의 생체지표들은 본 연구에서는 통계적으로 유의한 예측 인자로 나타나지 않았다. 예를 들어 BMI는 오즈비 1.02(95% CI: 0.87-1.21,  $p = 0.7852$ ), 목둘레는 오즈비 0.95(95% CI: 0.78-1.15,  $p = 0.6049$ )로, 모두 유의수준을 넘지 못하였다. 이는 특히 한국인을 포함한 아시아 인구에서 OSA의 임상적 표현형이 서구와는 상이할 수 있음을 시사한다. 실제로 아시아인은 상대적으로 낮은 BMI에서도 중증 OSA를 보이는 경우가 많으며, 이는 상기도의 구조적 협착 등 인종적 생리적 차이에 기인할 수 있다는 보고가 되어왔다(Li et al., 2000).

Kim 등(2004)은 한국 중년 성인을 대상으로 한 연구에서, BMI가 서구 기준 혹은 아시아 기준으로 비만에 해당하지 않음에도 불구하고 수면 중 무호흡 증상이 발생하는 사례가 관찰된다고 보고한 바 있으며, Lee 등(2010)은 아시아권 전반에서 OSA의 표현형이 서양에 비해 비만도가 낮고 두정부 해부학적 구조의 차이가 더 큰 영향을 미칠 수 있음을 주장하였다. 더불어, Byun 등(2020)은 한국인을 대상으로 STOP-Bang 설문 에 포함된 BMI 및 목둘레 기준을 아시아인에 적합하도록 조정할 경우 선별 정확도가 향상될 수 있음을 제시하였다. 이러한 선행 연구들은 본 연구에서 BMI나 목둘레가 진단 예측에 미치는 영향이 상대적으로 낮게 나타난 결과를 뒷받침하며, 아시아인을 대상으로 한 OSA 예측모델에서는 이러한 특성을 고려한 접근이 필요함을 시사한다.

결과적으로 본 방식으로 도출된 로지스틱 회귀모형은 검증 데이터셋에서 AUC 0.8660을 기록하여, OSA 진단 여부를 구분하는 데 있어 비교적 우수한 판별력을 보였다.

로지스틱 회귀분석이 통계적으로 해석 가능성과 예측 성능을 동시에 확보한 전통적인 접근이라면, 본 연구에서 적용한 DNN(Deep Neural Network) 기반 예측 모델은 그보다 더 높은 판별력을 보여주었다. 실제로 본 연구의 DNN 모델은 검증 데이터셋에서 AUC 0.9020, 정확도 94.85%를 기록하였으며, 이는 로지스틱 회귀모델(AUC 0.8660)보다 우수한 성능을 나타내었다. 이와 같은 결과는 DNN이 변수 간의 입력 변수 간의 비선형적 관계와 복잡한 상호작용을 효과적으로 학습하고 반영할 수 있음을 잘 보여준

다 (Shickel et al., 2023). 특히 OSA와 같이 다양한 생체지표, 건강행태, 설문 기반 항목 등이 복합적으로 얹혀 있는 질환의 경우, 선형 관계만을 가정하는 전통적 회귀 모형보다는, 다층 퍼셉트론 기반의 딥러닝 구조가 실제 질병 발현 패턴을 더 잘 학습할 수 있는 가능성이 있다. 이는 본 연구에서도 확인된 바와 같이, 동일한 예측 대상과 변수 조합을 적용했음에도 DNN 모델이 보다 높은 AUC와 정확도를 나타낸 것으로 설명된다.

그러나 DNN의 가장 큰 단점은 ‘블랙박스’ 구조로 인한 결과 해석의 어려움이다. 예측력은 우수하나, 개별 변수들이 예측에 어떻게 기여했는지를 직접 해석하기 어렵다는 한계가 있으며, 이는 의료 분야에서 딥러닝 모델 도입에 있어 핵심적인 장애 요소로 지적되고 있다. 이를 보완하기 위해 본 연구에서는 복잡한 딥러닝 모델의 내재된 구조를 설명 가능한 형태로 정량화하는 도구로서, 최근 의료 인공지능 연구에서 핵심적으로 활용되고 있는 SHAP(Shapley Additive Explanations) 분석을 활용하여, DNN 모델의 예측 결과에 대한 변수 기여도를 정량화하고 시각화함으로써 설명 가능성을 확보하고자 하였다 (Lundberg et al., 2017).

결과적으로 두 모델에서 도출된 주요 예측 변수의 기여 경향성은 일치하는 양상을 보였다. DNN 모델에서 예측 기여도가 가장 높게 나타난 변수는 ‘수면 중 무호흡 목격 여부’, ‘앓아있는 시간’, ‘성별’ 등으로, 이중 일부는 로지스틱 회귀분석에서도 통계적으로 유의한 변수로 나타났다. 특히 ‘수면 중 무호흡 목격 여부’는 두 모델 모두에서 가장 강력한 예측 인자로 확인되었는데, 이는 수면 중 무호흡을 목격한 경험이 OSA 진단과 밀접하게 연관되어 있음을 시사하며, 해당 항목이 OSA의 선별 및 진단에 있어 핵심적인 임상 지표로 작용할 수 있음을 나타낸다. 이와 같은 결과는 Chen 등(2015)의 다인종 코호트 연구에서도 유사하게 보고된 바 있으며, 다양한 인종 집단에서 수면 중 무호흡 증상의 자각 또는 목격 여부가 높은 예측력을 가진 것으로 나타났다.

또한, 본 연구에서는 앓아있는 시간 역시 두 모델 모두에서 일관되게 높은 중요도를 보였다. 로지스틱 회귀분석에서는 앓아있는 시간이 길수록 OSA 진단 확률이 유의

하게 증가하는 경향을 보였으며, 이는 좌식 생활습관이 OSA 발생에 기여할 수 있음을 시사한다. 기존 문헌에서도 장시간의 앉은 자세는 신체 활동량 저하, 대사기능 저하, 비만 및 상기도 기능 저하 등과 연관되며, 이는 수면 무호흡의 위험 증가로 이어질 수 있음이 보고된 바 있다(Liu et al., 2022).

한편, 성별 변수는 DNN 모델과 SHAP 분석에서 비교적 높은 예측 기여도를 보인 반면, 로지스틱 회귀분석에서는 통계적으로 유의하지 않았다. 이와 같은 결과는 머신러닝 모델이 비선형 상호작용이나 변수 간 복합적인 관계를 포착하는데 더 유리하다는 점에서 기인할 수 있으며, 남성이 여성보다 OSA 유병률이 높다는 기존 역학적 근거와도 일치하는 방향성을 보였다(Young et al., 2002). 즉, 통계적으로 유의하지 않더라도, 일관된 예측 방향성이 여러 분석 방법에서 재현된다는 점은 해당 변수의 임상적 중요성을 시사하는 정성적 근거로 해석될 수 있다.

이와 같은 결과는 기존 문헌에서 제기된 선형 모델(높은 해석적 장점, 제한적인 예측 성능)과 머신러닝 모델(해석의 어려움, 우수한 예측 성능) 간의 트레이드오프(trade-off)를 잘 보여주는 사례라 할 수 있다. 특히, 정확한 진단 예측이 중요한 임상 환경에서는 DNN과 같은 복잡한 알고리즘의 활용이 효과적일 수 있음을 시사한다. 최종적으로, 본 연구에서는 AUC를 기준으로 예측 성능이 가장 우수한 DNN 모델을 폐쇄성수면무호흡증 진단의 대안적 선별 도구로 제안하며, 이는 STOP-Bang 설문 도구의 민감도 위주의 한계를 보완하고 개별 환자의 특성을 고려한 맞춤형 진단 접근이 가능하다는 점에서 실용적 의의가 있다.

최종적으로 본 연구는 OSA 예측을 위한 다양한 통계 및 머신러닝 기법을 비교함으로써, 예측 성능의 향상 가능성과 기법 간 특성을 고찰하였다. 단순 설문 점수 기반의 전통적 로지스틱 회귀모형은 해석력이 뛰어난 반면, 복합적인 임상 및 건강행태 지표를 반영한 다변수 모델은 보다 높은 판별력을 보였다. 특히, 동일한 변수 구성을 기반으로 한 DNN 모델은 가장 높은 AUC(0.9020)를 기록함으로써, 비선형적 상호작용과 복잡한 변수 조합을 효과적으로 반영할 수 있는 머신러닝 기반 예측 기법의 강점을 확인시켜 주었다.

이러한 결과는 로지스틱 회귀분석이 선형 관계에 기반해 변수 간 독립적인 영향을 추정하는 데에 적합한 반면, DNN과 같은 심층신경망은 변수 간 복잡한 비선형 상호작용과 높은 차원의 구조를 효과적으로 학습할 수 있어, 보다 정밀한 예측력을 확보할 수 있음을 시사한다(Shickel et al., 2017, Miotto et al., 2018). 특히 OSA와 같이 다양한 생체 지표, 건강행태, 설문 기반 요인이 복합적으로 작용하는 질환의 경우, DNN과 같은 고차원 알고리즘은 보다 유의미한 패턴을 포착하여 예측 성능을 향상시킬 수 있는 가능성을 제시한다.

하지만 본 연구는 다음과 같은 한계점을 가진다. 첫째, 본 연구에서 종속변수로 사용된 수면무호흡증 진단 변수는 모든 대상자에게 표준화된 진단 수단으로 간주되는 수면다원검사(PSG)를 직접 시행하여 수집된 값이 아니라, 의료진으로부터 진단받은 경험에 대한 자가 보고를 기반으로 정의되었다. 이에 따라, 질환이 없다고 분류된 일부 대상자가 실제로 폐쇄성수면무호흡증을 가지고 있었을 가능성이 있으며, 이는 국민건강영양조사와 같은 건강 조사자료에서 사용되는 진단 변수의 정확성과 포괄성 측면에서 한계로 작용할 수 있다. 또한, STOP-Bang 설문 항목 중 ‘수면 중 무호흡 목격 여부’에 ‘예’로 응답한 대상자의 경우, 본인의 수면 이상에 대한 인식 가능성이 상대적으로 높고, 이는 의료기관의 방문 및 진단으로 이어질 가능성이 있다. 이러한 구조적 특성은 OSA 진단 여부가 특정 항목의 응답에 따라 선택적으로 결정될 수 있는 선택 편향의 원인으로 작용할 수 있다. 따라서 본 연구에서 관찰된 예측력은 선별 도구 자체의 성능뿐 아니라, 진단 변수의 정의 방식 및 편향 가능성에서 기인한 영향 또한 함께 고려되어야 한다.

둘째, 본 연구는 국민건강영양조사(KNHANES)의 복합표본설계에 기반한 대규모 단면 조사 데이터를 활용하였기 때문에, 분석 결과는 전체 모집단을 대표하는 추정에는 유의미하나, 개별 환자 수준에서의 임상적 적용 또는 진단적 도구로서의 직접적인 활용에는 제한이 있을 수 있다. 특히 국민건강영양조사는 특정 질환의 고위험군을 중심으로 표집된 임상 코호트와 달리 건강한 일반인 중심의 인구집단 표본이므로, 모형의 민감도(sensitivity) 측면에서는 낮은 특성을 보일 가능성도 존재한다.

셋째, 연구에 사용된 데이터는 단일 시점에서 수집된 단면적(cross-sectional) 자료로 구성되어 있어, 변수 간의 연관성은 확인할 수 있으나 원인과 결과 간의 인과관계를 도출하기에는 근본적인 한계가 존재한다. 특히 생활습관 요인(예: 음주량 등)과 OSA 진단 간의 관계가 역방향성(reverse causation)을 포함하고 있을 가능성도 존재한다.

넷째, 분석 대상자의 OSA 유병률이 전체 표본 중 극히 일부(약 0.7%)에 해당하는 불균형한 이진 분류 구조를 가지므로, 학습 과정에서 민감도 저하, 모델 수렴 불안정성 등의 문제가 발생할 수 있었다. 본 연구에서는 이를 보완하고자 가중치 조정(class weight)을 적용하였지만, 여전히 민감도와 특이도 간의 균형 유지에는 한계가 존재한다.

다섯째, DNN(Deep Neural Network) 모델의 특성상 내부 구조가 다층의 비선형 연산으로 구성되어 있어 직관적인 변수 간 관계 해석이 어렵다는 문제가 있으며, SHAP 분석을 통해 일부 설명 가능성을 확보하였지만, 이는 개별 예측에 대한 지역적(local) 설명에 국한되며 전역적(global) 추론에는 제한이 따른다. SHAP 값은 변수의 예측 기여에 대한 방향성과 크기를 모두 제공하지만, 실무적으로는 영향력의 크기(magnitude)에 중점을 두는 해석이 주로 이루어지므로 인과적 해석 또는 정책적 해석에는 신중한 접근이 요구된다.

그럼에도 불구하고 본 연구는 다음과 같은 학술적 및 실무적 기여점을 지닌다. 첫째, 본 연구는 국내 대표 보건조사인 국민건강영양조사(KNHANES)의 복합표본자료를 기반으로, 수면다원검사 없이도 OSA 진단을 예측할 수 있는 머신러닝 기반 조기 선별 모형을 제시하였다는 점에서 의의가 있다. 이는 실제 임상 현장에서의 선별검사 접근성이 제한된 상황에서도, 최소한의 정보로 위험군을 효율적으로 추정할 수 있는 기반을 마련하였다는 점에서 보건의료 현장에 실용적 함의를 제공한다.

둘째, 기존의 STOP-Bang 점수 기반 단순 선별 방식을 넘어, 건강행태 변수(예: 음주, 흡연, 앉아있는 시간) 및 임상적 생체지표(혈압, 혈당, 지질수치 등)를 통합적으로 고려함으로써, 현실적 진단 정확도를 제고하고 개별 대상자 수준의 위험도 예측

정밀도를 향상시켰다. 이는 기존 도구의 민감도 한계를 보완할 수 있는 데이터 기반 보완전략의 예시로 볼 수 있다.

셋째, 예측모형의 구축에 있어 전통적 통계기법(로지스틱 회귀)과 딥러닝 기반 모델(DNN)을 병행하여 비교·분석함으로써, 머신러닝 기반 접근의 유효성과 한계 모두를 검토하였다. 특히 SHAP(Shapley Additive Explanations) 분석을 통해 변수별 기여도와 예측 방향성을 시각화함으로써, 복잡한 예측모형의 설명 가능성(interpretability)을 높이고 임상적 수용성을 제고하였다.

넷째, 본 연구는 자가보고형 설문자료 기반의 대규모 보건데이터에서도 머신러닝 모델을 활용한 질환 예측이 가능함을 보여주었으며, 향후 디지털 헬스 환경에서의 위험군 예측 시스템 또는 개인 맞춤형 수면 건강 관리 알고리즘 개발을 위한 기초자료로 활용될 수 있다.

향후 연구에서는 OSA의 보다 정밀한 진단 예측을 위해, 수면다원검사(Polysomnography, PSG) 기반의 객관적인 진단 데이터를 병합한 예측모델 개발이 필요하다. 이를 통해 자가보고형 변수의 정확도 한계를 보완하고, 모델의 진단 타당도와 임상적 활용 가능성을 제고할 수 있을 것이다. 또한, 자가보고 기반이 아닌 충분한 규모의 진단 사례 데이터를 확보하여 분석함으로써 예측 정확도를 향상시키고, 선별도구의 진단 타당성을 정밀하게 검증할 필요가 있다. 더불어, 단면적 분석의 한계를 극복하기 위해 시계열 데이터를 활용한 OSA의 발생 및 예후 예측 연구가 요구되며, 이는 질환의 진행 양상과 장기적 건강 결과를 예측하는 데 기여할 수 있다. 더 나아가 XGBoost, LightGBM, Random Forest 등 다양한 머신러닝 알고리즘과의 성능 비교를 통해 최적의 예측 모형을 도출하고, 각 알고리즘의 해석 가능성과 임상 적용성을 평가하는 과정도 병행되어야함을 제언한다.



## V. 결론

본 연구는 국민건강영양조사(2019-2023년) 데이터를 기반으로, 폐쇄성수면무호흡증 (Obstructive Sleep Apnea, OSA)의 진단 예측을 위한 분류 모형을 구축하고, 전통적인 통계 모형과 머신러닝 기반 모형 간의 성능을 비교하였다. STOP-Bang 설문 점수와 건강행태 및 임상 지표를 활용하여 로지스틱 회귀모형과 심층신경망(Deep Neural Network, DNN) 모형을 각각 구성하였으며, AUC를 기준으로 예측 성능을 평가하고, 변수별 기여도에 대한 해석 가능성을 함께 분석하였다.

분석 결과, STOP-Bang 점수는 OSA 고위험군 선별에 일정 수준의 민감도를 갖추었으나, 단독 사용 시 진단 예측의 정밀도에는 한계가 있었다. 다변수 로지스틱 회귀분석에서는 수면 중 무호흡 목격 여부가 가장 강력한 예측 변수로 확인되었으며, BMI나 목둘레 등 일부 전통적 생체지표는 통계적으로 유의한 영향을 보이지 않았다. 이는 아시아 인구집단에서의 OSA 표현형 차이를 반영하는 결과로 해석된다.

머신러닝 기반 DNN 모델은 AUC 0.9020으로 다변수 로지스틱 회귀모델(AUC 0.8660)보다 우수한 예측 성능을 보였다. 두 모형에서 공통적으로 확인된 변수별 영향력은 SHAP 분석을 통해 시각적으로 검증되었으며, 두 모델에서 주요 예측 변수의 기여 경향성이 일치함을 통해 임상적 타당성과 해석 가능성을 동시에 확보하였다. 이는 OSA 예측에 있어 전통적 선형 모형을 보완할 수 있는 머신러닝 기반 접근의 가능성을 시사하며, STOP-Bang 설문 기반 선별체계를 정교화할 수 있는 방향성을 제시한다.

본 연구는 국가 단위의 대표성 있는 건강조사 자료를 활용하여 OSA 조기 예측 체계를 제시하고, 기존 설문 기반 점수 체계의 한계 보완 가능성을 탐색했다는 점에서 의미가 있다. 다만, OSA 진단이 자가보고 자료에 기반하였고, 변수 간 인과관계를 단면 자료만으로 확정할 수 없다는 한계가 존재한다. 향후 연구에서는 수면다원검사(PSG) 기반의 진단 변수 활용, 변수 조합의 최적화, 다양한 머신러닝 기법 간 성능 비교 등을 통해 보다 정밀한 진단 예측모형 개발이 필요하다.



## 참 고 문 헌

- Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. *Sleep medicine reviews*. 2017;34:70-81.
- Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society*. 2008;5(2):136-43.
- Marin JM, Carrizo SJ, Vicente E, Agusti AG. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *The Lancet*. 2005;365(9464):1046-53.
- Kim J, In K, Kim J, You S, Kang K, Shim J, et al. Prevalence of sleep-disordered breathing in middle-aged Korean men and women. *American journal of respiratory and critical care medicine*. 2004;170(10):1108-13.
- Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *Journal of clinical sleep medicine*. 2017;13(3):479-504.
- Rosen CL, Auckley D, Benca R, Foldvary-Schaefer N, Iber C, Kapur V, et al. A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. *Sleep*. 2012;35(6):757-67.
- Chung F, Abdullah HR, Liao P. STOP-Bang questionnaire: a practical approach to screen for obstructive sleep apnea. *Chest*. 2016;149(3):631-8.

- Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP questionnaire. *Anesthesiology*. 2008;108(5):812-21.
- Chung F, Yang Y, Brown R, Liao P. Alternative scoring models of STOP-bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. *Journal of Clinical Sleep Medicine*. 2014;10(9):951-8.
- Shickel B, Loftus TJ, Ruppert M, Upchurch Jr GR, Ozrazgat-Baslanti T, Rashidi P, et al. Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. *Scientific reports*. 2023;13(1):1224.
- Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *British Journal of General Practice*. 2018;68(668):143-4.
- Holfinger SJ, Lyons MM, Keenan BT, Mazzotti DR, Mindel J, Maislin G, et al. Diagnostic performance of machine learning-derived OSA prediction tools in large clinical and community-based samples. *Chest*. 2022;161(3):807-17.
- 한국지질·동맥경화학회. 2022 Guidelines for the Management of Dyslipidemia, 5th ed. Seoul: Korean Society of Lipid and Atherosclerosis; 2022.
- Ioffe S, Szegedy C, editors. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*; 2015: pmlr.
- Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *Journal of big data*. 2020;7(1):28.
- Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM. The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. *Journal of Clinical Sleep Medicine*. 2011;7(5):459-65.

- Duarte RLdM, Fonseca LBdM, Magalhães-da-Silveira FJ, Silveira EAd, Rabahi MF. Validation of the STOP-Bang questionnaire as a means of screening for obstructive sleep apnea in adults in Brazil. *Jornal brasileiro de Pneumologia*. 2017;43(06):456-63.
- Waseem R, Chan MT, Wang CY, Seet E, Tam S, Loo SY, et al. Diagnostic performance of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea in different ethnic groups. *Journal of Clinical Sleep Medicine*. 2021;17(3):521-32.
- Shin C, Baik I. Evaluation of a modified STOP-BANG questionnaire for sleep apnea in adults from the Korean general population. *Sleep Medicine Research*. 2021;12(1):28-35.
- Li KK, Kushida C, Powell NB, Riley RW, Guilleminault C. Obstructive sleep apnea syndrome: a comparison between Far-East Asian and white men. *The Laryngoscope*. 2000;110(10):1689-93.
- Lee RW, Vasudavan S, Hui DS, Prvan T, Petocz P, Darendeliler MA, et al. Differences in craniofacial structures and obesity in Caucasian and Chinese patients with obstructive sleep apnea. *Sleep*. 2010;33(8):1075-80.
- Byun J-I, Kim D-H, Kim J-S, Shin WC. Usefulness of using alternative body-mass index and neck circumference criteria for STOP-Bang questionnaire in screening South Korean obstructive sleep apnea patients. *Sleep Medicine Research*. 2020;11(1):38-43.
- Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30.
- Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep*. 2015;38(6):877-88.

- Liu Y, Yang L, Stampfer MJ, Redline S, Tworoger SS, Huang T. Physical activity, sedentary behaviour and incidence of obstructive sleep apnoea in three prospective US cohorts. *European Respiratory Journal*. 2022;59(2).
- Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *American journal of respiratory and critical care medicine*. 2002;165(9):1217-39.
- Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics*. 2017;22(5):1589-604.
- Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*. 2018;19(6):1236-46.

## ABSTRACT

### Obstructive Sleep Apnea (OSA) Diagnostic Prediction Model Using STOP-Bang and Clinical Indicators: A Deep Neural Network Approach Compared with Logistic Regression Analysis

SeulKi Lee

Department of Health Informatics & Biostatistics

Graduate School of Public Health

Yonsei University, Seoul, Korea

#### Background and purpose

Obstructive Sleep Apnea (OSA) is a sleep-related breathing disorder characterized by repeated upper airway obstruction during sleep. It is recognized as a major public health concern due to its strong association with chronic diseases such as hypertension, cardiovascular disease, and diabetes. The STOP-Bang questionnaire is widely used as a screening tool for early detection of OSA, offering the advantage of high sensitivity in identifying high-risk individuals. However, its low specificity limits its precision in individual diagnosis, raising concerns about its standalone predictive accuracy.

This study aims to address these limitations by developing an enhanced predictive model that integrates the STOP-Bang score with additional clinical and behavioral health indicators. Furthermore, the study seeks to compare the predictive performance of a traditional statistical method—logistic regression—with that of

a machine learning-based Deep Neural Network (DNN), ultimately proposing a more effective model for OSA risk prediction.

## Methods

This study analyzed data from 14,939 adults aged 40 years and older, derived from the 8th and 9th cycles (2019-2023) of the Korea National Health and Nutrition Examination Survey (KNHANES). Obstructive Sleep Apnea (OSA) diagnosis was defined based on self-reported physician diagnosis. In addition to the STOP-Bang score and its individual components, the analysis incorporated various clinical and behavioral health indicators, including blood pressure, blood glucose, lipid levels, body mass index (BMI), alcohol consumption, smoking status, sleep duration, and sedentary time.

Logistic regression analysis was conducted considering the complex sampling design. Both univariate analysis, using only the STOP-Bang score, and multivariate analysis, incorporating additional health indicators, were performed to assess explanatory power. For model validation, the dataset was randomly split into training and validation sets in a 7:3 ratios. A Deep Neural Network (DNN)-based predictive model was also developed. Predictive performance and variable contribution were evaluated using the area under the receiver operating characteristic curve (ROC-AUC) and SHapley Additive exPlanations (SHAP) analysis.

## Results

In the univariate logistic regression analysis, the STOP-Bang score was significantly and positively associated with OSA diagnosis (OR = 2.31, 95% CI: 2.02-2.65,  $p < 0.0001$ ), with an AUC of 0.8268. In the multivariable logistic regression model, the AUC improved to 0.8660, with witnessed apnea during sleep

emerging as the most significant predictor (OR = 20.68, 95% CI: 9.69-44.14,  $p < 0.0001$ ). Other significant predictors included presence of snoring (OR = 2.29, 95% CI: 1.09-4.79,  $p = 0.0287$ ), total sitting time per day (OR = 1.14, 95% CI: 1.06-1.22,  $p = 0.0003$ ), and alcohol consumption (OR = 0.16, 95% CI: 0.04-0.67,  $p = 0.0130$ ).

The Deep Neural Network (DNN)-based prediction model achieved the highest performance among the three models, with an AUC of 0.9020. SHAP analysis demonstrated variable importance trends similar to those identified in the logistic regression model, indicating both interpretability and clinical relevance of the DNN model.

## Conclusion

While the STOP-Bang questionnaire is an effective tool for identifying individuals at high risk for OSA, its limited specificity at the individual diagnostic level highlights the need for complementary approaches. This study demonstrates that incorporating clinical and health indicators alongside the STOP-Bang score can enhance predictive precision. Moreover, the Deep Neural Network (DNN)-based machine learning model outperformed traditional statistical models, suggesting the potential for developing more advanced and accurate prediction tools for early OSA screening in the future.

---

Keywords: Obstructive Sleep Apnea (OSA), STOP-Bang, Logistic Regression, Deep Neural Network, Predictive Model, ROC-AUC, Korea National Health and Nutrition Examination Survey (KNHANES)