



Original Article

# 공공데이터 기반 지역사회 건강환경 요인을 활용한 치주질환 진료비 인공지능 예측모형 개발

김규석<sup>1+</sup>, 정나래<sup>2+</sup>, 강정윤<sup>3</sup>

<sup>1</sup>한국폴리텍대학 인공지능소프트웨어과, <sup>2</sup>한서대학교 치위생학과, <sup>3</sup>연세대학교 치위생학과

## An artificial intelligence-based predictive model for periodontal disease expenditures using community-level health environmental factors from public data

Kyu-Seok Kim<sup>1+</sup>, Narae Jung<sup>2+</sup>, Jung Yun Kang<sup>3</sup>

<sup>1</sup>Department of Artificial Intelligence Software, Korea Polytechnics

<sup>2</sup>Department of Dental Hygiene, Hanseo University

<sup>3</sup>Department of Dental Hygiene, College of Health Sciences, Yonsei University

**Corresponding Author:** Jung Yun Kang, Department of Dental Hygiene, College of Health Sciences, Yonsei University, 1 Yonseidae-gil, Wonju-si, Gangwon-do, 26493, Korea. Tel: +82-33-760-5564, Fax: +82-33-760-2919, E-mail: hannahkang@yonsei.ac.kr

<sup>†</sup>These authors contributed equally to this work.

### ABSTRACT

**Objectives:** This study aimed to develop and evaluate a deep neural network (DNN)-based predictive model for periodontal disease expenditures using community-level health environmental factors derived from public data. **Methods:** A total of 1,020 monthly records from 17 regions between January 2020 and December 2024 were analyzed. Independent variables included health behaviors, demographic characteristics, socioeconomic factors, and healthcare resource accessibility. A DNN model was constructed and evaluated using the mean absolute percentage error (MAPE), while permutation feature importance (PFI) was applied to quantify the relative contribution of each variable. **Results:** The DNN model achieved a mean MAPE of 11.01% (range: 9.35-12.51; SD: 0.83) across 10 repeated trials, indicating good predictive performance according to the Lewis (1982) criteria. PFI analysis identified total population, proportion of single-person households, and gender ratio as the most influential predictors of periodontal disease expenditures. **Conclusions:** These findings suggest that periodontal disease expenditures are shaped by complex interactions among demographic, socioeconomic, and behavioral factors, which can be effectively captured by AI-based predictive models. This study provides preliminary evidence that healthcare resource allocation and oral health policy development can benefit from AI-based approaches utilizing publicly available data.

**Keywords:** Artificial intelligence, Health expenditures, Periodontal diseases, Public health data, Socioeconomic factors

**주요어:** 인공지능, 의료비 지출, 치주질환, 공공 보건 데이터, 사회경제적 요인

### 서론

치주질환은 경미한 염증 상태인 치은염부터 치조골 파괴를 동반하는 치주염까지를 포함하는 대표적인 만성 염증성 구강질환으로, 초기 단계에서 증상이 미약하거나 자각하기 어려워 진단과 치료가 지연되기 쉽다. 성인 인구에서 높은 유병률을 보이는 치주질환은 전신 건강 악화,

의료비 부담 증가와도 밀접하게 연관되어 있으며, 치아 상실과 삶의 질 저하를 통해 사회·경제적 부담을 초래한다[1,2]. 따라서 치주질환은 국가 보건의료체계 차원에서 지속적인 감시와 개입이 요구되는 주요 만성질환으로 분류되고 있다.

선행연구에 따르면, 치주질환의 발생과 진행은 치면세균막이나 치석과 같은 국소적 요인에 국한되지 않으며, 개인의 건강행태, 인구학적 특성, 사회경제적 여건, 그리고 의료자원 접근성과 같은 다양한 요인의 복합적 영향을 받는다고 보고하고 있다[3,4].

흡연과 음주는 치주조직 파괴와 질환 악화를 촉진하는 주요 위험행태로 널리 검증되어 왔으며, 이러한 행태 요인은 지역 단위 의료비 지출 변동을 설명하는 데 중요한 요소로 작용한다[5-7]. 흡연은 치은 혈류를 감소시키고 면역반응을 저하시켜 치주조직의 방어력을 약화시키며, 결과적으로 치주미생물 환경을 악화시켜 염증 반응을 촉진한다. 그 결과, 흡연자는 비흡연자에 비해 치주질환의 진행 속도가 빠르고 조직 파괴 정도가 크다는 임상적 특징을 보인다[8]. 과도한 음주 역시 구강 내 환경 변화와 염증 반응 증가를 통해 치주조직의 손상을 초래할 수 있다[9]. 특히, 알코올 섭취는 구강 점막 자극과 미생물 생태 변화와 연관되어 있으며, 이러한 변화는 치주질환의 진행과 악화 가능성을 증가시키는 요인으로 작용한다[10]. 또한, 연령 증가에 따른 면역 기능 저하, 만성질환 동반, 약물 복용 증가 등 인구학적 특성은 치주질환에 대한 취약성을 높이는 요인으로 작용하며, 고령화가 심화되는 사회에서는 치주질환 부담이 더욱 증가할 가능성이 있다[11,12]. 선행연구에서는 고령층 인구 비중이 높은 지역일수록 치주질환 진료비가 증가하는 경향을 보인다고 보고하였다[13,14]. 이처럼 개인의 건강행태와 인구구조 변화는 치주질환의 지역별 의료비 격차를 이해하기 위한 핵심적인 분석 틀을 제공할 수 있다.

사회경제적 요인 역시 치주질환 발생과 밀접한 관련이 있는 것으로 보고되는데, 특히 소득 수준은 구강건강 인식, 예방적 치과의료 이용, 치료 접근성에 영향을 미쳐 치주질환 격차를 심화시키는 요인으로 작용할 수 있다[15,16]. 더불어 치과의료 접근성도 치주질환 관련 의료 이용에 큰 영향을 미치는 요인으로 보고되고 있다. 지역 내 치과 병·의원의 수는 주민의 치료 접근성과 진료 이용 패턴을 반영하는 지표로 활용되고 있으며, 의료기관 밀도는 질환의 조기 발견 가능성과 진료비 규모에도 영향을 미친다는 연구 결과가 제시되어 왔다[17,18]. 즉, 의료기관의 분포는 단순한 시설 규모를 넘어 지역 주민의 치료 기회와 예방적 관리 수준을 결정짓는 구조적 요인으로 기능한다. 특히 인구 1만 명당 치과 병·의원 수는 지역 의료자원 분포를 평가하는 대표적 지표로, 치주질환 진료율과 의료비 지출 규모를 설명하는 데 유용하게 활용되고 있다.

최근에는 인구·사회경제적 자료와 국가 단위 건강보험 청구 데이터를 결합하여 질환 발생이나 의료비 지출을 예측하려는 시도가 증가하고 있다[19,20]. 특히 머신러닝(Machine learning) 기반 예측 기법은 기존의 회귀 기반 접근보다 비선형 구조를 설명하는 데 유리하고, 모델 안정성이 높다는 이유로 보건의료 분야에서 활용이 확산되고 있다[21]. 심층신경망(Deep neural network, DNN)은 다층 신경망 구조를 통해 변수 간의 비선형적 관계와 복잡한 상호작용을 효과적으로 학습할 수 있어, 단순 회귀모형으로는 포착하기 어려운 패턴을 정교하게 반영할 수 있다는 장점이 있다. 또한, 대규모 자료 환경에서도 높은 예측 성능과 안정성을 유지하며, 다양한 은닉층 구성과 활성화 함수의 적용을 통해 모델의 표현력을 유연하게 확장할 수 있다. 아울러 DNN은 지역별 의료비 지출에 내재된 구조적 비대칭성이나 시계열적 변동성을 효과적으로 학습하는 데 유리하여, 정책적 의사결정에 활용 가능한 정밀한 예측 결과를 제공할 수 있다. 또한, DNN의 비선형 특징은 복잡한 데이터 패턴을 모델링할 수 있어 예측 정확도를 높이는 데 도움을 준다. 이와 같은 장점은 치주질환처럼 복합적 요인이 결합된 현상의 예측 정확도를 높이는 데 기여할 수 있다. 이에 본 연구는 공공데이터를 활용하여 지역사회 환경요인과 국가 단위 건강보험 청구 데이터를 통합하여, DNN 기반 예측모형을 통해 치주질환 의료비 지출을 예측하고자 한다.

치주질환 진료비는 개인의 건강행태, 인구학적 특성, 사회경제적 여건, 그리고 의료자원 접근성 등 복합적 요인의 영향을 받는다. 따라서 다양한 설명변수를 동시에 고려하는 예측모형의 구축이 요구된다. 흡연율, 음주율, 평균 연령, 성비, 1인가구 비율, 1인당 지역총소득, 인구 1만 명당 치과 병·의원 수 등은 기존 선행연구에서 관련성이 검증된 변수로 보고되었으며, 의료비 지출 변동을 설명하는 데 유용한 지표로 활용되어 왔다[11,22]. 이러한 변수들은 지역사회의 건강환경과 의료 접근성을 구조적으로 반영하는 지표로서, 예측모형의 성능을 향상시키는 데 핵심적인 역할을 한다. 그럼에도 국내에서는 치주질환 진료비를 인공지능(Artificial intelligence, AI) 기법을 활용하여 예측하고, 변수별 영향력을 체계적으로 분석한 연구가 충분히 이루어지지 않은 실정이다.

이에 본 연구는 지역 단위의 건강행태, 인구학적 특성, 사회경제적 여건, 그리고 의료자원 접근성 변수를 활용하여 치주질환의 요양급여비용 총액(Total medical expenditure, TME)을 예측하는 DNN 기반 모형을 구축하고, 실증 분석을 통해 그 활용 가능성을 제시하고자 한다. 나아가 본 연구는 기존의 통계적 분석을 넘어, 설명변수 간의 복잡한 비선형 관계와 상호작용을 반영할 수 있는 새로운 분석 틀을 제안하는 것을 목적으로 한다.

## 연구방법

### 1. 연구 데이터

본 연구는 2020년 1월부터 2024년 12월까지의 월별 자료를 기반으로 치주질환 요양급여비 예측에 영향을 미치는 요인을 분석하였다. 분석 대상 지역은 강원, 경기, 경남, 경북, 광주, 대구, 대전, 부산, 서울, 세종, 울산, 인천, 전남, 전북, 제주, 충남, 충북의 17개 시·도이며, 60개월의 월별 자료와 결합하여 총 1,020개의 관측값을 구축하였다.

지역사회 건강환경 요인이 치주질환 요양급여비용총액(TME) 예측에 미치는 영향을 분석하기 위해 건강행태 요인, 인구학적 특성, 사회경제적 요인, 의료자원 접근성으로 구분된 지역 수준 변수를 설정하였다<Table 1>. 독립변수는 치주질환 발생 및 의료비 지출에 영향을 미치는 것으로 보고된 선행연구와 이론적 배경을 바탕으로 선정되었다.

치주질환 발생 및 진행과 밀접한 관련이 있는 건강행태 요인으로 지역 주민의 흡연율(Smoking rate, SR)과 음주율(Drinking rate, DR)을 포함하였으며, 이는 치주조직 염증 및 질환 악화와 관련된 주요 위험요인으로 보고되어 왔다. 인구학적 특성으로는 지역별 총인구(Total population, TP), 평균 연령(Average age, AA), 성비(Gender ratio, GR), 1인 가구 비율(Single-person household rate, SPH)을 포함하여 지역의 인구구조와 고령화 수준을 반영하고자 하였다. 성비(GR)는 특정 지역의 남성 인구를 여성 인구로 나눈 값에 100을 곱하여 산출되는 지표로, '여성 100명당 남성 수'를 의미한다. 이 변수는 수식 (1)과 같이 정의된다.

$$GR = \left( \frac{\text{Male Population}}{\text{Female Population}} \right) \times 100 \quad (1)$$

사회경제적 요인으로는 1인당 지역내총소득(Per capita regional income, PCRI)을 사용하여 지역의 경제적 수준을 반영하였다. 이 변수는 구강건강 인식 및 의료서비스 이용 수준에 영향을 미쳐 의료비 격차를 설명하는 변수로 활용되어 왔다. 의료자원 접근성 지표로는 인구 1만 명당 치과병원 비율(Ratio of dental hospitals, RDH)과 치과의원 비율(Ratio of dental clinics, RDC)을 포함하여 치과의료 서비스 공급 수준을 반영하였으며, 이는 지역 내 치료 접근성과 의료이용에 영향을 미치는 요인으로 보고되고 있다. 종속변수는 치주질환으로 인한 지역별 요양급여비용총액(Total medical expenditure, TME)으로 정의하였으며, 이는 국민건강보험 급여 항목에 한정된 비용을 의미한다. 본 변수는 지역별 치주질환 의료비 부담 수준을 나타내는 지표이다. 모든 데이터는 엑셀 또는 CSV 파일을 취합하는 형태로 진행되었다. 인구·사회경제 및 의료자원 관련 자료는 국가통계포털(Korean statistical information service, KOSIS)[23]에서, 종속변수인 치주질환 요양급여비용(치은염 및 치주질환, K05 주상병 기준)은 건강보험심사평가원(Health insurance review and assessment service, HIRA) 빅데이터 개방 포털[24]을 통해 확보하였으며, 수집 시점은 2025년 12월이다.

**Table 1.** Definitions and data sources of variables used in the analysis

Variables	Description	Source
Independent variable		
SR(Smoking rate)	Current smoking rate among local residents	KOSIS
DR(Drinking rate)	Drinking rate among local residents	
TP(Total population)	Population by region	
AA(Average age)	Average age by region	
GR(Gender ratio)	Male-to-female ratio per 100 females	
SPH(Single-person household rate)	Ratio of single-person households	
PCRI(Per capita regional income)	Gross regional income per capita	
RDH(Ratio of dental hospitals)	Ratio of dental hospitals per 10,000 people in the region	
RDC(Ratio of dental clinics)	Ratio of dental clinics per 10,000 people in the region	
Dependent variable		
TME(Total medical expenditure)	Periodontal disease expenditures by region	HIRA

KOSIS: Korea statistical information service; HIRA: Health insurance review and assessment service.

월별 자료의 시계열적 특성과 지역 간 이질성을 고려할 수 있도록 전처리 과정을 수행하였다. 변수 간 단위 불일치를 조정하고 결측값 여부를 점검한 결과, 분석 자료에서는 결측치가 확인되지 않아 별도의 결측치 대체 과정은 수행하지 않았다. 기술통계 분석을 통해 변수 분포를 사전 점검하였으며, 일부 변수에서 분산이 크게 나타났으나 실제 지역 간 특성을 반영한 값으로 판단하여 임의적인 이상치 제거는 수행하지 않았다. 종속변수인 영양급여비용총액(TME)의 경우 우측 치우침 분포를 가질 가능성이 있으나, 본 연구에서는 의료비 규모 자체를 예측 대상으로 설정하였으므로 로그 변환과 같은 추가 변환은 적용하지 않았으며, 입력 변수 간 스케일 차이를 보정하기 위해 StandardScaler를 적용하였다. 지역별 인구 규모 차이를 보정하기 위한 비율 변수 산출도 전처리 과정에 포함하였다. 이러한 변수 구성은 지역사회 환경요인이 치주질환 영양급여비용총액 예측에 미치는 영향을 다차원적으로 분석할 수 있도록 하며, DNN 기반 예측모형의 입력값으로 활용되었다. 본 연구는 개인 식별 정보가 포함되지 않은 공공데이터를 활용한 연구로서, 연세대학교 기관생명윤리위원회의 심의 면제 승인(1041849-202602-SB-026-01)을 받았다.

## 2. DNN 모형 구축 및 분석 방법

본 연구는 건강행태 요인, 인구학적 특성, 사회경제적 요인, 의료자원 접근성이 치주질환 영양급여비용총액에 미치는 영향을 분석하고, 이를 기반으로 DNN 기반의 예측모형을 적용하였다

본 연구에서 적용한 DNN 모형의 변수 설정은 수식 (2)와 같다. 지역별 영양급여비용총액(TME)을 종속변수로 설정하고, 흡연율(SR), 음주율(DR), 인구 수(TP), 평균 연령(AA), 여성 100명 대비 남성의 비율(GR), 1인 가구 비율(SPH), 1인당 지역 총소득(PCRI), 인구 1만 명당 치과 병원 수 비율(RDH), 인구 1만 명당 치과 의원 수 비율(RDC)을 독립변수로 포함하였다. 이는  $TME = RF(SR, DR, TP, AA, GR, SPH, PCRI, RDH, RDC)$ 로 표현된다. 이러한 변수 구성은 건강행태 요인, 인구학적 특성, 사회경제적 요인, 의료자원 접근성을 포괄적으로 고려함으로써 의료비 변동을 설명하는 데 필요한 다차원적 정보를 모형에 통합할 수 있도록 하며, 이를 통해 치주질환 진료비 예측의 정확도와 해석 가능성을 동시에 확보하고자 하였다.

$$TME_i = DNN(SR_i, DR_i, TP_i, AA_i, GR_i, SPH_i, PCRI_i, RDH_i, RDC_i) \tag{2}$$

DNN 모형의 세부 설정은 <Table 2>에 제시하였다. 본 연구에서 적용한 DNN 모형은 두 개의 은닉층(64, 32 뉴런)으로 구성되며, 은닉층에는 ReLU 활성화 함수를, 출력층에는 선형(linear) 활성화 함수를 적용하였다. 모형 학습에는 Adam 옵티마이저(learning rate=0.001)를 사용하였고, 입력 변수의 단위 차이를 보정하기 위해 StandardScaler를 적용하여 표준화를 수행하였다. 전체 자료는 학습 데이터와 검증 데이터로 8:2의 비율로 무작위 분할하였으며, 이는 제한된 표본 수를 고려하여 학습 데이터의 활용도를 높이기 위한 설정이다. 다만, 본 연구는 월별 시계열 자료를 포함하고 있으므로 시간 순서를 반영한 분할 방식이 보다 엄밀한 검증 방법이 될 수 있으며, 무작위 분할의 경우 동일 지역의 인접 시점 데이터가 학습 및 검증 데이터에 동시에 포함될 가능성이 있다는 점은 본 연구의 한계로 인식된다. 모형 학습은 총 300 에포크(epochs), 배치 크기(batch size) 1로 설정하였으며, 예측 성능 평가는 평균절대백분율오차(mean absolute percentage error, MAPE)를 지표로 사용하였으며, 이러한 설정을 통해 모형의 학습 안정성과 예측 성능을 확보하고자 하였다. 또한, 동일한 조건에서 반복 학습이 가능하도록 모형 구조와 하이퍼파라미터(hyperparameter)를 일관되게 유지하여 분석의 재현성을 확보하였다. 이러한 모형 설정은 치주질환 의료비 지출 예측에 내재된 비선형적 패턴을 효과적으로 포착하고, 지역 간 차이를 반영한 예측 성능을 확보하기 위한 목적에 부합한다.

**Table 2.** Setting values in deep neural network model

Classification	Setting values
Model architecture	2 hidden layers (64, 32 neurons)
Activation function	ReLU (hidden layer), linear (output layer)
Optimizer	Adam (learning rate = 0.001)
Data preprocessing	StandardScaler
Data splitting	Training data : validation data = 8 : 2
Evaluation metrics	MAPE
Epochs	300
Batch size	1

MAPE: Mean absolute percentage error.

모형의 예측 성능을 통계적으로 신뢰할 수 있는 값으로 제시하기 위해, 동일한 설정으로 DNN 학습을 총 10회 반복 수행하였다. 반복 학습은 초기 가중치 설정 및 학습 데이터 분할 과정에서 발생하는 무작위성의 영향을 상쇄하고, 평균적 성능과 분산을 동시에 평가할 수 있는 장점이 있다. 각 반복에서 계산된 MAPE 값을 종합하여 최솟값·최댓값·평균·표준편차를 산출함으로써 모형의 안정성을 정량적으로 검증하였으며, 이는 단일 학습 결과에 비해 보다 더 엄밀한 예측 성능 평가를 가능하게 한다.

예측 성능의 해석에는 Lewis(1982)의 기준을 적용하였다[25]. 이 기준에 따르면 MAPE가 10% 미만이면 매우 높은 수준의 예측력, 10-20%는 양호한 수준의 예측 성능, 20-50% 범위는 합리적이지만 오차가 다소 존재하는 일반적 수준, 50% 초과는 예측력이 약하여 모델 개선이 필요한 수준으로 분류된다.

또한, 본 연구에서는 모형이 활용하는 입력 변수들이 예측 성능에 미치는 영향을 정량적으로 파악하기 위해 순열 변수 중요도(Permutation feature importance, PFI) 분석을 실시하였다. 이 방법은 학습된 모델을 변경하지 않은 상태에서 특정 변수를 무작위로 치환했을 때 예측 성능이 얼마나 저하되는지를 측정함으로써, 해당 변수의 상대적 기여도를 평가한다. 특히 DNN과 같은 비선형 구조에서는 전통적 회귀계수와 달리 변수의 영향력을 직접적으로 확인하기 어렵기 때문에, PFI는 각 변수의 예측 기여도를 비교적 직관적이면서도 신뢰성 있게 확인할 수 있는 도구로 기능한다. 이를 통해 모형의 예측 결과에 핵심적으로 작용하는 요인을 식별하고, 정책적 해석과 향후 변수 개선 방향을 제시하기 위한 근거를 확보하였다.

### 3. 통계분석

데이터 전처리 및 기술통계 분석은 Python 기반 환경에서 수행하였으며, 주요 라이브러리는 Pandas와 NumPy를 활용하였다. 변수의 정규화 과정에서는 scikit-learn의 StandardScaler를 사용하였으며, 학습 데이터와 검증 데이터의 분할 및 모델 성능 평가도 동일 라이브러리를 기반으로 수행하였다. DNN 모형은 Keras 기반으로 구현되었으며, TensorFlow를 활용하여 학습을 수행하였다. 예측 성능 평가는 MAPE를 기준으로 산출하였고, 변수 중요도 분석을 위해 PFI를 적용하여 각 변수의 상대적 기여도를 평가하였다.

## 연구결과

### 1. 기술통계

<Table 3>에 제시된 기술통계 결과에 따르면, 흡연율(SR)과 음주율(DR)은 각각 평균 19.0%와 56.9%로 지역 간 변동 폭은 비교적 제한적인 반면, 총 인구 수(TP)와 요양급여비용총액(TME)은 지역별로 큰 차이를 보여 규모의 이질성이 뚜렷하게 나타났다. 평균 연령(AA)은 지역별로 36.9세에서 48.6세 사이에 분포하며, 표준편차가 2.4로 나타나 인구 고령화 수준의 지역 간 차이가 존재하고, 1인가구 비율(SPH)은 27.6-39.9 범위로 나타나 지역 인구구조의 다양성을 반영하였다. 1인당 지역총소득(PCRI)은 29,509천 원에서 65,994천 원 사이에 분포하며 평균은 41,353천 원으로, 지역 간 경제적 격차는 지역별 의료비 지출 수준과 예방적·치료적 의료이용 패턴 차이를 반영하는 중요한 요인으로 작용할 수 있다. 의료자원 접근성 변수의 경우, 인구 1만 명당 치과병원 비율(RDH)과 치과의원 비율(RDC)은 각각 0.0-1.6, 0.1-130.1의 매우 넓은 분포를 보였다. 특히 치과의원 비율(RDC)의 표준편차가 13.9로 나타난 것은 지역별 치과의원 밀집도의 편차가 상당함을 의미하며, 이는 의료자원 접근성 차이가 의료이용과 진료비 지출에 직접적으로 영향을 미칠 수 있음을 보여준다. 이러한 결과는 치주질환 관리 수준, 인구구조, 의료 접근성, 경제력 등 복합적 요인이 지역별로 상이하게 작용하고 있음을 보여준다. 특히, 요양급여비용총액(TME)의 높은 변동성은 비선형 구조를 포착할 수 있는 DNN과 같은 머신러닝 기반 모형의 활용이 적합함을 뒷받침한다.

**Table 3.** Descriptive statistics

(n=1,020)

Variables (unit)	Min-Max	Mean±SD
Smoking rate(%)	13.2-22.3	19.0±1.8
Drinking rate(%)	47.4-62.4	56.9±3.0
Total population(persons)	360,907-14,162,083	3,102,932.7±3,379,289.6
Average age(yrs)	36.9-48.6	44.1±2.4
Gender ratio(males per 100 females)	93.3-106.0	100.2±2.9
Single-person household rate(%)	27.6-39.9	34.7±3.0
Per capita regional income(thousand KRW)	29,509-65,994	41,353.0±8,380.9
Ratio of dental hospitals(per 10,000 population)	0.0-1.6	0.1±0.2
Ratio of dental clinics(per 10,000 population)	0.1-130.1	7.2±13.9
Total medical expenditure(thousand KRW)	548,732-59,578,510	9,720,482.9±12,849,934.6

Min: Minimum; Max: Maximum; SD: Standard deviation; KRW: Korean won.

## 2. DNN 예측 결과

DNN 모형을 10회 반복 수행한 결과, MAPE는 최솟값 9.35에서 최댓값 12.51까지 분포하였다<Table 4>. 전체 평균 MAPE는 11.01로 나타났으며, 표준편차는 0.83으로 나타났다. 이는 Lewis(1982)의 분류 기준에 따라 양호한 예측 성능(Good forecasting) 수준에 해당한다. 표준편차가 과도하게 크지 않다는 점은 반복 실험 간 예측 성능의 변동성이 제한적이며, 모형이 전반적으로 일관된 성능을 유지하고 있음을 의미한다. 즉, 일부 반복 실험에서는 MAPE가 10% 미만에 근접하여, 매우 높은 수준의 예측 성능(Highly accurate forecasting)에 준하는 결과도 확인되었다.

**Table 4.** Mean absolute percentage error descriptive statistics for deep neural network model

	Minimum	Maximum	Mean	Standard deviation
MAPE	9.35	12.51	11.01	0.83

MAPE: Mean absolute percentage error.

## 3. 치주질환 진료비 예측에 대한 변수별 기여도 분석

DNN 모형에서 변수별 예측 기여도를 평가하기 위해 PFI 분석을 수행하였다<Table 5>. PFI 분석 결과, 총인구(TP)가 모델 예측 성능에 가장 큰 영향을 미치는 변수로 나타났다. Importance mean 값은 23.93으로, 총인구 변수를 무작위로 치환했을 때 MAPE가 평균 약 23.93%p 증가함을 의미한다. 이는 지역 단위 의료비 총액이 인구 규모에 구조적으로 강하게 의존하고 있음을 반영한 결과로 해석된다.

다음으로 1인가구 비율(SPH)과 성비(GR)가 비교적 높은 기여도를 보였다. 이는 가구 구조와 성별 분포 등 인구 구성 특성이 의료 이용 패턴 및 비용 규모 형성에 유의미한 영향을 미칠 수 있음을 의미한다. 음주율(DR)과 1인당 지역총소득(PCRI)은 중간 수준의 기여도를 나타내어, 건강행태 요인과 지역의 경제적 여건이 치주질환 요양급여비용총액 예측에 일정 부분 기여함을 보여준다.

반면, 평균 연령(AA)과 흡연율(SR)은 상대적으로 낮은 기여도를 보였으며, 치과병원 비율(RDH)과 치과의원 비율(RDC)은 본 모형에서 예측 성능에 미치는 영향이 제한적인 것으로 나타났다. 이는 의료 공급 인프라의 단순한 수적 분포보다 인구 규모와 가구·인구 구조적 요인이 비용 총액을 설명하는 데 보다 중요한 역할을 수행하고 있음을 의미한다.

**Table 5.** Permutation feature importance of variables in the deep neural network model for predicting periodontal disease expenditures

Rank	Variable	FIM	Importance mean (%p) <sup>*</sup>
1	Total population	161.28	23.93
2	Single-person household rate	21.18	1.60
3	Gender ratio	16.48	1.87
4	Drinking rate	8.21	0.62
5	Per capita regional income	6.54	1.02
6	Average age	2.71	0.55
7	Smoking rate	1.91	0.21
8	Ratio of dental hospitals	0.99	0.16
9	Ratio of dental clinics	0.20	0.12

FIM: Feature importance for mean absolute percentage error (MAPE).

<sup>\*</sup>Importance mean indicates the mean increase in MAPE (%) caused by permuting each feature.

## 고안

본 연구는 치주질환이 성인 인구에서 높은 유병률을 보이고, 전신 건강과 의료비 부담에 중대한 영향을 미치는 만성 염증성 질환이라는 문제 의식에서 출발하였다. 이를 위해 전국 17개 시·도를 대상으로 2020년부터 2024년까지 60개월간의 월별 자료를 결합하여 치주질환 진료비를 예측하고자 하였으며, 건강행태 요인, 인구학적 특성, 사회경제적 요인, 의료자원 접근성의 지역사회 건강환경 요인을 반영하는 다양한 설명 변수를 구성하였다. 특히, 다양한 변수를 통합적으로 고려함으로써, 치주질환 의료비를 둘러싼 구조적 맥락을 정량적으로 분석할 수 있는 기반을 마련하였다.

기술통계 결과는 치주질환 진료비 지출이 지역별로 뚜렷한 이질성을 보인다는 점을 보여주며, 이는 인구 규모와 구조, 사회경제적 여건, 건강행태, 의료자원 분포가 복합적으로 작용한 결과로 해석될 수 있다. 총인구와 요양급여비용총액의 큰 지역 간 편차는 치주질환 의료비 부담이 지역 특성에 따라 상이함을 의미하며, 평균 연령과 1인가구비율의 차이는 고령화 및 가구 구조 변화가 의료이용과 비용에 영향을 미칠 가능성을 보여준다. 또한, 1인당 지역총소득의 넓은 분포는 사회경제적 격차가 치과의료 이용과 치료 양상에 반영될 수 있음을 나타낸다. 특히, 치과 병·의원 비율은 광역자치단체별 도시 규모와 의료 인프라의 이질성을 반영하고 있다. 이러한 기술통계 결과는 치주질환 의료비가 단일 요인이 아닌 다양한 구조적 요인의 결합에 의해 설명됨을 보여주며, 선형모형만으로는 이러한 복잡적 관계를 충분히 반영하기 어렵다는 점을 의미한다.

본 연구에서 DNN 모형 구축하여 치주질환 진료비를 예측하고, Lewis(1982)의 MAPE 기준을 활용해 성능을 평가하였다[24]. 동일한 설정 하에서 다회 반복 학습을 수행한 결과, 전반적으로 우수한 예측 성능(Good forecasting)을 보였으며, 일부 반복 실험에서는 매우 우수한 예측 성능(Highly accurate forecasting)에 근접한 결과도 나타났다. MAPE의 평균과 표준편차가 과도하게 크지 않았다는 점은 초기 가중치 설정이나 데이터 분할 방식의 차이에도 불구하고 DNN 모형이 비교적 안정적인 예측력을 유지하고 있음을 의미한다. 더 나아가 PFI 분석을 통해 총 인구 수가 진료비 예측에 가장 큰 기여를 하는 변수로 확인되었고, 그 다음으로 1인 가구 비율, 남녀 성비, 음주율, 1인당 지역내총소득, 평균 연령, 흡연률, 인구 1만 명 당 치과병원 비율과 치과의원 비율이 뒤따르는 것으로 나타났다. 이는 인구구조, 경제력, 건강행태가 치주질환 의료비 예측에서 핵심적인 설명축을 이룬다는 점을 실증적으로 보여준다. 이러한 결과를 통해 본 예측모형에서는 의료자원 관련 변수보다 인구구조 및 건강행태 관련 변수의 예측 기여도가 상대적으로 높게 나타났다. 본 연구에서 제시된 변수 중요도는 인과적 관계를 의미하지 않으며, 예측 모델 내에서 각 변수의 상대적 기여도를 반영한 결과임을 유의해야 한다. 향후 중단 자료나 인과추론 방법을 활용한 추가 연구를 통해 변수 간 인과적 관계를 규명할 필요가 있다.

한편, 종속변수를 요양급여비용총액으로 설정함에 따라, 총인구 변수가 높은 기여도를 나타내는 것은 일정 부분 구조적으로 반영된 측면이 있다. 실제 비용 구조를 결정하는 치료행위의 구성이 고려되지 않았다는 점도 본 연구의 한계이다. 향후 연구에서는 인구 규모의 영향을 통제 한 1인당 진료비 또는 인구 대비 표준화 지표를 활용함으로써 치주질환 부담의 상대적 수준을 보다 정교하게 분석할 수 있을 것이다.

본 연구는 공공데이터를 활용하여 지역 단위 인구학적 특성, 사회경제적 요인, 건강행태, 의료자원 접근성을 통합적으로 고려한 인공지능 기반 치주질환 진료비 예측모형을 구축하였다는 점에서 의의를 갖는다. 기존 연구들이 치주질환의 위험요인 분석이나 의료비 현황 파악에 초점을 둔 선형 회귀 중심의 분석을 주로 활용한 것과 달리, 본 연구에서는 DNN을 적용하여 설명변수 간의 비선형적 관계와 복잡한 상호작용을 반영함으로써 지역별 의료비 지출 변동 구조를 보다 정교하게 제시하였다. 또한, 본 연구는 PFI 분석을 통해 예측모형의 결과를 해석 가능하게 제

시함으로써, 인공지능 모형의 활용에 대한 투명성과 신뢰성을 제고하였다. 이는 보건의료 분야에서 인공지능 기반 예측모형의 적용 가능성을 실증적으로 확장하는 동시에, 단순 예측을 넘어 정책적 해석이 가능한 분석 틀을 제안한다는 점에서 학문적 기여를 갖는다.

본 연구는 제한된 국내 공공데이터를 활용함에 따라 비교 기준 모형의 부재, 총액 지표 사용에 따른 한계, 성능 평가의 불안정성, 타 국가 및 지역으로의 적용 가능성 제한 등의 방법론적 한계를 가진다. 그럼에도 불구하고, 공공보건 및 치위생학 분야에서 인공지능 기반 데이터 분석의 실질적 활용 가능성을 제시한 선도적 연구로서 의의를 가진다.

저소득 국가에서 중요한 과제로 인식되는 의료비 지출을 예측을 주제로 한 선행연구[26]에서는 요르단의 총 의료비 지출을 분석하여 인구 규모, 의료자원, 물가 수준 등이 주요 결정요인임을 제시하였다. 본 연구에서 개발한 DNN 기반 예측모델 또한 치주질환 진료비 지출이 인구구조와 사회적 및 인구학적 요인에 의해 크게 설명됨을 보여주었다는 점에서, 보건의료 지출이 다양한 인구학적, 경제적, 구조적 요인의 복합적 상호작용에 의해 형성된다는 기존 연구와 일관된 경향을 보인다. 다만, 선행연구가 인과성 검정을 기반으로 변수 간 장기 및 단기 관계를 분석한 후 이를 신경망 모델에 적용한 반면, 본 연구는 예측 중심 접근을 통해 변수의 상대적 기여도를 평가하였다는 점에서 방법론적 차이가 있다. 그럼에도 두 연구 모두 의료비 지출이 다양한 요인의 상호작용 속에서 결정되며, 인공지능 기반 모델이 이러한 복잡한 패턴을 효과적으로 반영할 수 있음을 보여준다. 또한, 본 연구는 기존의 국가 단위 연구를 확장하여 지역사회 수준의 인구 및 사회구조적 요인이 치주질환 진료비 지출과 밀접하게 관련됨을 확인하였으며, 보건의료 자원 배분 및 구강보건 정책 수립 시 이러한 요인을 함께 고려할 필요가 있다.

향후 연구에서는 몇 가지 측면에서 본 연구를 확장하고 보완할 필요가 있다. 첫째, 본 연구가 주로 양적 지표에 기반한 거시적 분석에 초점을 맞추었다면, 이후 연구에서는 개별 환자 수준의 임상 정보나 진료 이력, 예방 서비스 이용 여부 등 보다 미시적 데이터를 결합하여 예측모형의 설명력을 강화할 수 있을 것이다. 둘째, DNN 이외에 Gradient Boosting, 시계열 딥러닝 모형 등 다양한 알고리즘과의 비교 분석을 통해 치주질환 진료비 예측에 가장 적합한 모형 조합과 하이브리드 구조를 모색해 볼 필요가 있다. 셋째, 본 연구는 예측에 초점을 두었으나, 향후 인과추론 기법이나 구조방정식 모형 등을 활용하여 각 요인이 의료비에 미치는 직접·간접 효과를 분해하는 분석도 유의미할 것이다. 넷째, 정책 변화나 보장성 강화, 예방 프로그램 도입과 같은 제도적 요인을 모형에 반영함으로써, 시나리오별 의료비 변화와 정책 효과를 사전에 평가할 수 있는 의사결정 지원 도구로 발전시키는 것도 중요한 과제가 될 것이다. 마지막으로, 다양한 국가 및 인구집단을 포함한 외부 데이터셋을 활용하여 모형의 외적 타당도와 재현성을 검증하고, 다기관 데이터를 기반으로 모형을 재학습함으로써 예측 성능의 일반화 가능성을 높일 필요가 있다.

본 연구에서 적용한 DNN 모형은 비교적 제한된 변수 수를 기반으로 함에도 불구하고, 변수 간의 상호작용 및 비선형 관계를 효과적으로 반영할 수 있다는 점에서 활용되었다. 특히 인구구조, 사회경제적 요인, 건강행태, 의료자원 접근성과 같은 변수들은 독립적으로 작용하기보다 상호의존적이고 복합적인 영향을 미칠 가능성이 높아, 선형 모형만으로는 충분히 설명되기 어렵다. 이에 DNN을 적용하여 이러한 비선형적 특성을 반영하고자 하였으며, 반복 학습을 통해 예측 성능의 안정성도 함께 검증하였다. 다만, 다양한 기계학습 알고리즘과의 비교를 통한 최적 모형 탐색은 본 연구의 범위를 넘어서는 부분으로, 향후 연구에서 추가적으로 검토될 필요가 있다.

## 결론

본 연구는 전국 17개 시·도의 2020년 1월부터 2024년 12월까지 월별 공공데이터를 활용하여 치주질환 진료비를 예측하는 DNN 기반 인공지능 모형을 구축하고, 그 적용 가능성을 실증적으로 검토하였다. 본 연구에서는 DNN 모형이 반복 학습에서도 비교적 일관된 예측 성능을 보였으며, 치주질환 진료비 지출의 지역별 변동 패턴을 일정 수준에서 설명할 수 있음을 확인하였다. 다만, 데이터 분할 방식, 비교 모형 부재, 종속 변수 설정 등 방법론적 한계가 존재하므로, 예측 성능을 일반화하거나 정책적 활용 가능성을 단정적으로 해석하는 데에는 신중할 필요가 있다. 본 연구의 결과는 치주질환 진료비 예측에 대한 탐색적 접근으로서의 의의를 가지며, 향후 다양한 모형 비교와 추가 검증을 통해 보다 정교한 예측 및 정책적 활용 가능성을 검토할 필요가 있다. 주요 결론은 다음과 같다.

1. 인구학적 특성 측면에서 총인구를 비롯한 인구 구조 변수들이 치주질환 진료비 예측에 가장 높은 기여도를 나타내어, 지역 인구 규모와 구성의 차이가 치주질환 의료비 부담 형성에 핵심적으로 작용함을 확인하였다.
2. 사회경제적 요인인 1인당 지역총소득은 치주질환 진료비 예측에 중간 수준의 기여도를 보여, 지역의 사회경제적 요인이 진료비 지출에 영향을 미칠 가능성을 확인하였다.
3. 건강행태 요인인 흡연과 음주는 치주질환 진료비 예측에 일정 수준의 기여도를 나타내었다.
4. 의료자원 접근성 요인인 치과 병·의원 분포는 진료비 예측에 일정 부분 기여하였으나, 그 영향력은 인구학적·사회경제적·건강행태 요인에 비해 상대적으로 제한적인 것으로 나타났다.

## Notes

### Author Contributions

Conceptualization: KS Kim, JY Kang; Data collection: KS Kim, NR Jung; Formal analysis: KS Kim, JY Kang; Writing-original draft: KS Kim, NR Jung; Writing-review&editing: KS Kim, NR Jung, JY Kang

### Conflicts of Interest

The authors declared no conflicts of interest.

### Funding

None.

### Ethical Statement

This study was exempted from review by the Institutional Review Board of Yonsei University, as it used de-identified public data (approval number: 1041849-202602-SB-026-01).

### Data Availability

All data generated or analyzed in this study are freely available at 'Korean statistical information service' (<https://kosis.kr/>) and 'Health insurance review and assessment service' (<https://opendata.hira.or.kr/>).

### Acknowledgements

None.

## References

1. Kim YJ, Kwak JS. Estimation of lifetime dental expenditures for periodontitis. *J Korean Soc Dent Hyg* 2021;21(3):245–53. <http://doi.org/10.13065/jksdh.20210024>
2. Jung YS, Kim YS, Kim KR. Association between multimorbidity and periodontal disease in Korean adults: a nationwide cross-sectional cohort study. *Int J Dent Hyg* 2024;22(2):376–83. <https://doi.org/10.1111/idh.12785>
3. Korean Council of Periodontology Professors. *Periodontology*. 5th ed. Seoul: Koonja; 2010: 109–45.
4. Kim BO, Cho MS, Kim SA, Sim HS, Han YK, Go EK, et al. *Periodontology*. 3rd ed. Seoul: Daehannarae; 2012: 45–68.
5. Kim JH. A convergence study on the association between alcohol consumption and periodontal disease. *J Korea Converg Soc* 2018;9(8):95–100. <http://doi.org/10.15207/JKCS.2018.9.8.095>
6. Won YS, Kim JH. Association between cigarette smoking status and periodontal disease in adults: results from the 2012 Korea national health and nutrition examination survey. *J Korean Acad Oral Health* 2016;40(2):133–9. <https://doi.org/10.11149/jkaoh.2016.40.2.133>
7. Yeo JY, Jung HS. Determinants of dental screening and unmet dental needs: interaction effect between geographical accessibility and economic affordability. *Korean J Health Econ Policy* 2012;18(4):109–26.
8. Kim WJ, Shin YJ. A multi-level analysis of factors affecting the unmet needs of dental care service: focusing on comparison by age group. *J Korean Acad Oral Health* 2021;45(3):126–37. <https://doi.org/10.11149/jkaoh.2021.45.3.126>
9. Tezal M, Grossi SG, Ho AW, Genco RJ. The effect of alcohol consumption on periodontal disease. *J Periodontol* 2001;72(2):183–9. <https://doi.org/10.1902/jop.2001.72.2.183>

10. Sala I, Conti S, Antonazzo IC, Rozza D, Losa L, Ferrara P, et al. Random forest regression for predicting healthcare costs using administrative databases from a health protection agency in Northern Italy [Internet]. *Epidemiology, Biostatistics, and Public Health*; 2025[cited 2025 Dec 15]. Available from: <https://riviste.unimi.it/index.php/ebph/article/view/29504>
11. Lee YJ, Choi YH. Disparities in dental healthcare utilization based on regional characteristics in Korea. *J Korean Acad Oral Health* 2024;48(2):64-70. <https://doi.org/10.11149/jkaoh.2024.48.2.64>
12. Hwang SH, Yoo JS. Relationship between chronic disease, oral health behavior of elderly and periodontal disease. *J Korea Acad-Ind Coop Soc* 2024;25(5):116-22. <https://doi.org/10.5762/kais.2024.25.5.116>
13. Kang EJ. Convergent relationship between drinking and smoking behavior and periodontal disease in elderly Koreans. *J Korea Converg Soc* 2019;10(7):295-301. <https://doi.org/10.15207/JKCS.2019.10.7.295>
14. Won YS, Choi CH, Oh HN. Risk factors of periodontal disease in Korean adults. *J Korean Acad Oral Health* 2014;38(3):176-83. <https://doi.org/10.11149/jkaoh.2014.38.3.176>
15. Kim N, Kim CY, Shin H. Inequality in unmet dental care needs among South Korean adults. *BMC Oral Health* 2017;17:80. <https://doi.org/10.1186/s12903-017-0370-9>
16. Jung EJ. Disparities in oral health according to the socioeconomic status of adults: analysis of data from the 7th Korea National Health and Nutrition Examination Survey. *J Korean Soc Dent Hyg* 2024;24:17-26. <https://doi.org/10.13065/jksdh.20240003>
17. Forrest LN, Ivezaj V, Grilo CM. Machine learning v. traditional regression models predicting treatment outcomes for binge-eating disorder from a randomized controlled trial. *Psychol Med* 2023;53(7):2777-88. <https://doi.org/10.1017/S0033291721004748>
18. Lee SR. Oral health status and oral health behavior according to the periodontal disease in Korean elderly [Master's thesis]. Gwangju: Graduate School of Public Health, Chosun University; 2021.
19. Prismsari S, Kim K, Mun HY, Kang JY. A proposal for a predictive model for the number of patients with periodontitis exposed to particulate matter and atmospheric factors using deep learning. *J Dent Hyg Sci* 2024;24(1):22-8.
20. Lee SH, Kim KS, Mun HY, Kang JY. Prediction model for dental implants utilization in the elderly after the national health insurance coverage of dental implants: focusing on socioeconomic factors. *J Korean Soc Dent Hyg* 2024;24(1):9-16. <https://doi.org/10.13065/jksdh.20240002>
21. Lee SY, Lee YH. A convergence study of adults' oral health behaviors and periodontal disease. *J Korea Converg Soc* 2019;10(5):63-70. <http://doi.org/10.15207/JKCS.2019.10.5.063>
22. Park YD, Kang JO, Kim SJ, Kwon HJ, Hwang JH, Hwang KS. Estimation of the costs of smoking-related oral disease: a representative South Korean study. *Int Dent J* 2012;62(5):256-61. <https://doi.org/10.1111/j.1875-595X.2012.00118.x>
23. Korean Statistical Information Service. Gross regional domestic product per capita [Internet]. Statistics Korea[cited 2025 Dec 15]. Available from: <https://kosis.kr/>
24. Health Insurance Review and Assessment Service. Public data list [Internet]. HIRA Big Data Open Portal[cited 2025 Dec 15]. Available from: <https://opendata.hira.or.kr/>
25. Lewis CD. *Industrial and business forecasting methods*. Oxford: Butterworth-Heinemann; 1982.
26. Saleh MH, Alkhaldeh RS, Jaber JJ. A predictive modeling for health expenditure using neural networks strategies. *J Open Innov Technol Mark Complex* 2023;9(3):100132. <https://doi.org/10.1016/j.joitmc.2023.100132>