

# 기계학습을 통한 당뇨병 분석 변수의 확장: 추가 요인의 통합 연구

이소정<sup>1</sup>, 최지웅<sup>2</sup>, 이채린<sup>3</sup>, 김남재<sup>4</sup>, 김종완<sup>5</sup>

<sup>1,3</sup>삼육대학교 보건관리학과 학부생

<sup>2</sup>삼육대학교 인공지능융합학부 학부생

<sup>4</sup>(주)글로벌비즈텍 대표

<sup>5</sup>삼육대학교 SW융합교육원 교수

leesojung0624@naver.com, iostream1204@gmail.com, rampion2713@gmail.com,

knj7360@gbt2019.com, kimj@syu.ac.kr

## Expanding Diabetes Analysis Variables Through Machine Learning: A Study on the Integration of Additional Factors

So-Jung Lee<sup>1</sup>, Ji-Woong Choi<sup>2</sup>, Chae-Rin Lee<sup>3</sup>, Nam Jae Kim<sup>4</sup>, Jongwan Kim<sup>5</sup>

<sup>1,3</sup>Dept. of Department of Public Health, Sahmyook University

<sup>2</sup>Dept. of Division of Artificial Intelligence Convergence, Sahmyook University

<sup>4</sup>Global Biz Tech Co., Ltd.

<sup>5</sup>Software Convergence Education Center, Sahmyook University

### 요 약

본 연구는 당뇨병 예측의 정확도를 높이기 위해 AdaBoostClassifier 기법을 제안한다. 기존 연구에서는 당뇨병 초기 단계를 개인이 인지하기 어렵고, 예측 모델의 정보 부족으로 인해 정확도가 낮았다. 이를 개선하기 위해, 질병관리청의 2022 국민건강영양조사 데이터를 활용하여 당뇨병과 추가 속성들 간의 상관관계를 분석하였다. 본 연구는 기존에 알려진 당뇨병 관련 지표 외에도 만나이, 체질량지수, 요당, A형간염항체 양성여부, 결혼여부를 포함해 다양한 속성을 통합적으로 분석했다. 제안된 모델은 새로운 요인을 효과적으로 통합하여 정교한 예측을 가능하게 하고, 개인 맞춤형 건강관리와 조기 치료에 기여할 수 있다. 이를 통해 당뇨병 관리에 대한 새로운 인사이트를 제공하고 임상적 의사결정에도 도움이 될 것으로 기대된다.

### 1. 서론

최근 고령화 사회로의 진입과 함께 만성질환, 특히 당뇨병의 증가가 두드러지고 있으며, 당뇨병 관리와 예방은 중요한 공중보건 과제가 되었다. 대한당뇨병학회에 따르면, 2020년 기준 우리나라 30세 이상 당뇨병 환자는 약 600만 명으로, 이는 2050년에 도달할 것으로 예측한 환자 수를 30년 앞서 추월한 수치다[1, 2]. 당뇨병은 초기 단계에서 증상이 미미해 자가 진단이 어렵고, 적절한 치료가 없을 경우 심각한 합병증을 초래할 수 있어 조기 진단과 관리가 필수적이다. 기존 연구는 기술역학에 치우쳐 있어, 다양한 속성을 포함한 예측 모델이 부족했다. 한편, 기계학습을 이용한 연구 중 캐글(Kaggle)을 고찰한 결과 기존에 잘 알려진 속성(연령, BMI, 혈압, 혈당 등)을 이용한 당뇨병 예측 및 분류 모델이 대부분이었다. 또한, A형간염으로 인한 당뇨병의 연관

성은 이전 연구에서 발견된 바 있으나 본 연구를 통해 A형간염 항체 여부 자체가 19%의 연관관계를 가지는지는 알려진 바가 없었다[3]. 본 연구에서는 기존 모델의 이러한 한계를 극복하기 위해, AdaBoostClassifier를 활용하여 당뇨병 예측을 보다 정교하게 수행하고자 한다. 본 연구는 기존의 공복혈당, 경구당부하검사, 당화혈색소 외에도 추가적인 5가지 변수: 만나이, 체질량지수, 요당, A형간염항체 양성 여부, 결혼 여부를 포함하여 당뇨병과의 상관관계를 분석했다. 추가 변수는 당뇨병 예측의 정확성을 높이는 데 기여할 수 있으며, 개인 맞춤형 건강 관리와 조기 치료에 도움을 줄 수 있을 것이다.

### 2. 본론

#### 2-1. 데이터 준비

당뇨병 진단 기준 외의 입력 데이터는 5개 항목은 다음과 같다. 목표 라벨은 개인이 속한 이용자군

\* 교신저자(Corresponding Author)

을 나타내고, 세 가지 가능한 범주: 0, 1, 2이다. 총 점수는 다음과 같이 라벨로 매핑된다.

$$label = \begin{cases} 0 & \text{if } sumval < 1 \\ 1 & \text{if } 1 \leq sumval \leq 2 \\ 2 & \text{if } 2 < sumval \end{cases} \quad (1)$$

### 2-2. 알고리즘

AdaBoostClassifier 모델을 학습하고 평가하는 알고리즘은 (그림 1)과 같다.

---

#### Algorithm 1 Training and Evaluating

##### AdaBoostClassifier Model

---

- 1: **Input:** Number of samples  $N = 7380$ , Number of questions  $Q = 5$
  - 2: **Output:** Accuracy score
  - 3:  $N$  : Number of samples
  - 4:  $Q$ : Number of questions
  - 5:  $X$ : Random integer matrix of size (7380, 5)
  - 6:  $y$ : Label vector
  - 7: Initialize random seed to 42
  - 8: Generate random integer matrix  $X$  of size (7380, 5) with float64 values
  - 9: Compute total scores  $sum\_val = np.sum(X, axis=1)$
  - 10: Convert total scores to labels  $y$  using a threshold
  - 11: Split  $X$  and  $y$  into training and testing sets (80 % train, 20 % test)
  - 12: Train AdaBoostClassifier model on training set
  - 13: Predict on test set
  - 14: Compute accuracy of predictions
  - 15: **return** accuracy score
- 

(그림 1) 훈련 모델 평가 알고리즘

### 2-3. 모델의 성능

본 연구에서 사용된 모델의 성능은 <표 1>과 같다. 실험에 사용된 데이터는 총 7,380개이며 정확도는 70%를 보여주었다. 이어서 예측도는 70%, 재현율은 70%, F1-score은 69%의 수치를 나타냈다.

<표 1> 모델 성능 요약

Metric	Accuracy
Test Accuracy	70 %
Precision	70 %
Recall	70 %
F1 Score	69 %

본 연구는 전통적인 당뇨병 판단 기준인 공복 혈당, 당화혈색소, 경구 당부하검사 외에도 추가적인 독립 변수를 활용하여 당뇨병 예측 모델을 구축했다. 이 예측 모델은 약 70%의 정확도를 보였으며, 이는 다소 낮은 수치일 수 있지만, 연구의 핵심 목표는 새로운 변수들과 당뇨병 간의 잠재적 연관성을 탐구하는 데 있다. 기존 연구에서 다루지 않았던 다양한 변수들이 당뇨병 발병에 미치는 영향을 분석함으로써, 당뇨병 연구의 새로운 방향성을 제시하고, 더 나아가 당뇨병 예방과 관리에 유용한 새로운 인사이트를 제공하고자 한다. 결과적으로, 본 연구는 당뇨병 예측 모델의 개선과 발전을 위한 기초 자료를 제공하며, 향후 연구에서 예측 정확도를 향상시키기 위한 방법론 개발에 기여할 것이다.

### 3. 결론

당뇨병은 노인에게만 발생한다고 알려져 있지만, 청소년과 중장년층에서도 발병할 수 있는 복합적인 질환이다. 발병 요인에는 유전적 요인, 생활 습관, 환경적 요인 등이 있으며, 비만, 식습관 변화, 신체 활동 부족, 스트레스가 주요 위험 요인으로 작용한다. 당뇨병은 다양한 합병증을 초래할 수 있어 조기 발견과 관리가 중요하다. 본 연구는 당뇨병 예방과 관리에 대한 새로운 접근법을 제시하고, 당뇨병 발병률을 낮추기 위한 전략을 탐구했다. 이를 위해 공공 캠페인, 교육 프로그램, 정책적 지원의 중요성을 강조하며, 당뇨병의 조기 발견과 예방이 필요하다고 주장한다. 연구 결과는 당뇨병 예방과 관리를 위한 새로운 통찰을 제공하고, 향후 연구와 정책 개발에 기여할 수 있다.

### 감사의 글

본 연구는 2021년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (2021-0-01440).

### 참고문헌

- [1] Diabetes Korea. (n.d.). Diabetes factsheet. Retrieved August 25, 2024
- [2] 권혁상. (2023, 7월). 한국인 당뇨병의 유병률 및 치료 현황. JKMA, 66(7), 404-407.
- [3] 권지연 외. (2009). 급성 A형 간염 후 발생한 전격성 제1형 당뇨병 1예. Korean Clinical Diabetes, 10, 118-122.