

# 인공지능 기반의 심혈관질환 발현 가능성 예측 모델 개발

임소영, 배지원, 김재준, 신방호, 박종욱\*

건양대학교 의료인공지능학과

## A Prediction Model of Cardiovascular Disease based on Artificial Intelligence

Soyoung Im, Jiwon Bae, Jaejun Kim, Bangho Shin, Jong-Uk Park\*

Dept. Medical Artificial Intelligence, Konyang University, Korea

\*jupark@konyang.ac.kr

### Abstract

This study proposes a method of prediction of cardiovascular diseases (CVDs) that can develop within ten years in patients with sleep-disordered breathing (SDB). From the data during a baseline period when patients did not have any CVD, we extracted 18 features from electrography (ECG) based on signal processing methods, 30 ECG features based on artificial intelligence (AI), ten clinical risk factors for CVD. We trained the model and evaluated it by using CVD outcomes result, monitored in follow-ups. The optimal feature vectors were selected through statistical analysis and support vector machine recursive feature elimination (SVM-RFE) of the extracted feature vectors. Features based on AI, a novel proposal from this study, showed excellent performance out of all selected feature vectors. Also, new parameters based on AI were possibly meaningful predictors for CVD, when used in addition to the predictors for CVD that are already known. The selected features were used as inputs to the prediction model based on SVM for CVD, determining the development of CVD-free, coronary heart disease (CHD), HF, or stroke within ten years. As a result, the respective recall and precision values were 82.9% and 87.5% for CVD-free; 71.9% and 63.8% for CVD; 57.2% and 55.4% for CHD; 52.6% and 40.8% for HF; 52.4% and 44.6% for stroke. The F1-score between CVD and CVD-free was 76.5%, and it was 59.1% in class four. In conclusion, our results confirm the excellence of the prediction model for CVD in patients with SDB and verify the possibility of prediction within ten years of the CVDs that may occur in patients with SDB.

### 1. 연구 배경

수면호흡장애(sleep-disordered breathing, SDB)는 수면 중 상기도의 부분적 또는 완전한 폐쇄로 무호흡 혹은 저호흡의 환기 장애를 일으킨다[1]. 수면호흡장애는 전 인구의 2~4% 정도로 알려져 있으며[1], 우리나라에서도 3.2~4.5%로 알려져 있는 매우 흔한 질병이다[1].

수면 중 반복적으로 발생하는 수면 무호흡-저호흡은 반복적인 저산소혈증, 재산소화, 흉강압의 급격한 변화, 중추신경계의 각성을 일으키며, 이는 심혈관계의 급성 스트레스 인자로 작용한다[2]. 따라서 수면호흡장애를 장기적으로 치료하지 않고 방치하면 고혈압, 심부전, 심근경색, 부정맥, 뇌졸중 등의 발생이 높아져 심혈관질환으로 인한 사망위험을 증가시킨다[2].

최근에는 개인맞춤의학(personalized medicine)에 대한 관심이 높아지고 있으며, 여러분야에서 인공지능 및 빅데이터와 같은 기술을 적용하여 개인 맞춤형 질환 진단 및 예측을 할 수 있다는 가능성을 제시하고 있다[3].

따라서 본 연구에서는 수면호흡장애 환자의 향후 10년 내의 심혈관 질환을 예측하기 위한 심전도와 CVD 위험인자를 이용한 인공지능 기반의 알고리즘을 제안하였다.

### 2. 연구 방법

The Sleep Heart Health Study (SHHS)는 국립심장폐혈연구소가 심혈관 및 기타 수면장애 결과를 확인하기 위해 시행한 코호트 연구이다. 수면호흡장애가 관상동맥질환, 심부전, 뇌졸중 위험성과 관련성을 확인하기 위한 연구였으며, 미국 내 주요병원에서 참여하였다. 본 코호트에서는 각 참여센터의 임상시험윤리위원회(the Institutional Review Board, IRB)의 승인을 받았으며, 동의서

서명한 사람만을 연구 대상에 포함하였다. 6,441명의 SHHS 데이터셋 중 본 연구에 맞는 대상으로 3,367명(women: 1,521, men: 1,846)을 선정하고 분석하였으며, 알고리즘 설계와 평가를 위해 연구대상을 학습군과 평가군으로 각각 8:2의 비율로 나누었다. 이때 학습군과 평가군 간의 임상적 특징이 통계적으로 차이가 없었다( $p > 0.05$ ).

심혈관질환의 예측모델을 개발하기 위해 18개의 신호처리 기반의 심전도 특징, 30개의 인공지능 기반의 심전도 특징, 10개의 CVD 위험인자, 총 58개의 특징 벡터를 추출하였다.

심전도 신호는 baseline study의 수면다원검사 데이터를 사용하였으며, 125 samples/second로 리샘플링하였다

1.5~20 Hz의 대역통과 필터를 사용하여 기저선 변동과 전원 잡음을 제거한 후, 적응 문턱치 알고리즘과 형태학적 방법을 이용하여 QRS complex와 T wave를 검출하였다. ECG의 특징점을 검출한 이후 QTc와 STTc segment를 계산하였다.

심박변이도를 계산하기 위해 RR의 이소성 비트를 제거하였고, 이 신호를 NN(normal-to normal RR)으로 정의하였다. 심박변이도 분석은 계산된 NN을 등간격으로 보간한 뒤, 4 Hz로 재샘플링(resampling) 하였다. 재샘플링된 신호는 30초 단위로 고속 푸리에 변환(fast Fourier transform, FFT)을 수행 후, FFT에 제곱을 취해 전력 스펙트럼밀도(power spectrum density, PSD)를 계산하였다.

주파수영역 특징들을 계산하기 위해 사용된 각 주파수 대역은 초저주파(very low frequency, VLF: 0~0.04 Hz), 저주파(low frequency, LF: 0.04~0.15 Hz), 고주파(high frequency, HF: 0.15~0.4 Hz)이다. 최종적으로 신호처리 기반의 심전도 특징을 추출했으며, 각 특징별로 전체 수면 동안의 평균과 표준편차를 구하였다.

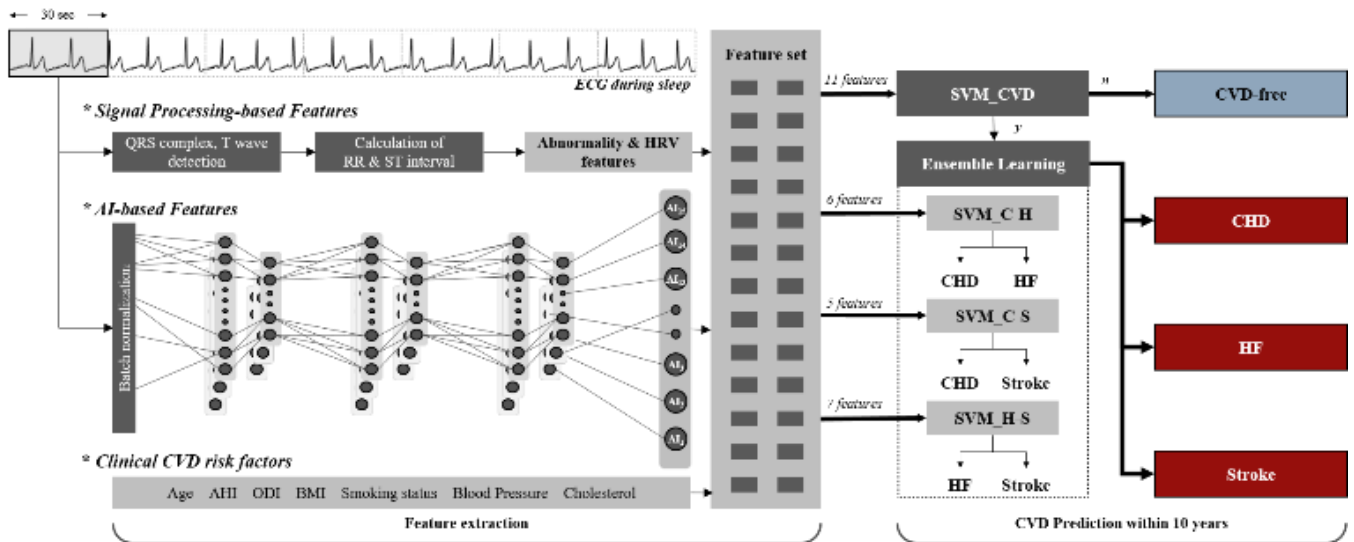


그림 1. 심전도를 이용한 인공지능 기반의 심혈관질환 예측 모델

인공지능 기반의 심전도 특징을 추출하기 위해 CNN (convolutional neural network) 기반의 인공지능 모델을 설계하였다. 모델은 30초 동안의 심전도 신호를 입력으로 받고, batch normalization을 거친다. 이후 3층 구조로 이루어진 convolution layer와 pooling layer를 통해 특징을 추출한다. 이렇게 학습한 CNN 모델의 flatten layer의 노드에서 추출한 AI 기반의 특징들을 각 전체 하룻밤 수면시간 동안 평균과 표준편차를 계산하여 30개의 특징을 추출하였다

심전도 특징 외에 추가적으로 대표적인 CVD 위험인자 10개를 선정하여 CVD 예측 모델의 특징으로 사용하였다.

SVM-RFE는 최적의 특징을 선택하는 방법 중 하나로, 각 특징벡터가 갖는 부류 간 여백을 이용하여 분별력을 측정한다. 부류 간 여백이 작을수록 분별력이 작다고 가정하고, 분별력이 작은 특징 순으로 순차 후방 제거 과정을 수행한다. 즉, 처음에는 모든 특징벡터를 이용하여 분별력을 측정하고, 가장 분별력이 작은 특징을 제거한다. 이 후, 제거되지 않은 특징들을 이용하여 특징 제거 과정을 다시 수행하며, 이 과정은 하나의 특징이 남을 때까지 반복된다. 가장 마지막에 남은 특징이 가장 분별력이 큰 특징이며, 분별력이 큰 순으로 특징들을 조합하면서 반복적인 학습 및 평가를 통해 최소의 특징 개수로 최적의 성능을 보이는 특징을 선택한다. 그 결과 최소의 특징 개수로 최적의 성능을 보이는 11개의 입력 특징이 선택되었다. 입력 특징으로 선택된 것 중 인공지능 기반의 심전도 특징은 3개로 높은 성능을 보이는 특징이다. 이는 기존의 심혈관질환 예측인자 외에 인공지능 기반의 새로운 파라미터가 중요한 인자로 활용될 수 있다는 가능성을 보여준다.

그림 1은 심혈관질환의 발현 가능성을 예측하기 위한 모델이다. 최적 특징을 추출하기 위한 과정을 거친 후 SVM 모델을 통해 10년 안에 CHD, HF, Stroke의 발생 여부를 예측한다. 먼저 CVD와 CVD-free를 예측하는 SVM (SVM\_CVD)을 거쳐 CVD로 분류된 경우, CHD, HF, stroke를 분류하기 위한 OvO (One-Vs-One) 다중 클래스 분류기를 거친다. 본 연구에서는 3개 타겟 클래스(CHD, HF, Stroke)를 분류하기 위해 3개의 SVM 분류기(SVM\_C-H, SVM\_C-S, SVM\_H-S)를 통해 incident CVD outcomes를 예측하였다.

### 3. 연구 결과

전체 학습군의 CVD-free에 대한 recall과 precision 값은 각각 84.5%와 88.9%, CVD는 73.8%와 65.7%, CHD는 63.3%와 57.6%, HF는 55.2%와 45.0%, Stroke는 52.2%와 47.8%의 성능을 보였으며, CVD와 CVD-free 간의 F1-score는 78.2%, 4 class에서는 61.7%의 성능을 보였다. 전체 평가군에서는 CVD-free의 recall과 precision 값은 각각 82.9%와 87.5%, CVD는 71.9%와 63.8%, CHD는 57.2%와 55.4%, HF는 52.6%와 40.8%, Stroke는 52.4%와 44.6%의 성능을 보였으며, CVD와 CVD-free 간의 F1-score는 76.5%, 4 class에서는 59.1%의 성능을 보였다.

본 연구에서 제안한 방법과 기존의 방법 간의 차이를 알아보기 위해 추가 실험을 수행하였다. 입력특징과 인공지능 모델을 다양하게 적용하여 성능평가를 하였으며, 인공지능 모델은 ANN, CNN, SVM, LDA, k-NN을 활용하였고, 입력 특징은 특징추출 과정을 포함한 것과 포함하지 않은 것으로 나누어 평가하였다. 그 결과, 본 연구에서 제안한 모델이 다른 모델에 비해 더 우수한 성능을 보였다. 따라서 본 연구에서 제시한 수면호흡장애환자의 심혈관질환 예측 모델의 우수성을 확인하였으며, 수면호흡장애 환자의 향후 10년 안에 발생하는 개인맞춤형 심혈관 예측의 가능성을 제시하였다. 따라서 본 연구에서 제시한 알고리즘을 통해 수면호흡장애의 위험성을 인지하여 적극적인 치료를 돕고, 수면호흡장애의 동반질환인 심혈관질환을 예방하는데 활용될 수 있을 것으로 사료된다.

### 4. 참고 문헌

- [1] N. Collop, "The effect of obstructive sleep apnea on chronic medical disorders." *Cleve. Clin. J. Med.* Vol 74, No. 1, p72-78, 2007
- [2] J. Kim, et al. "Prevalence of sleep-disordered breathing in middle-aged Korean men and women." *Am. J. Respir. Crit. Care Med.* Vol. 170, No. 10, p1108-1113, 2004
- [3] A. Shah, et al. "Building a rapid learning health care system for oncology: why CancerLinQ collects identifiable health information to achieve its vision." *J. Clin. Oncol.* Vol. 34, No. 7, p756-763, 2016