

## PPG 신호를 이용한 기계학습 기반 혈압 단계 예측

지예빈, 오현영, 여채운, 최정운, 강대성  
인제대학교

librenoblejyb@oasis.inje.ac.kr, hy1013@oasis.inje.ac.kr, ducodbs0516@oasis.inje.ac.kr,  
choiss5886@oasis.inje.ac.kr, mpkds@inje.ac.kr

## Machine Learning Based Blood Pressure Stage Prediction using PPG Signals

Yebin Ji, HyeonYeong Oh, Chaeyun Yeo, JeongYun Choi, Daesung Kang  
Inje University

## 요 약

본 논문은 피험자의 임상적 정보와 세 가지 특징값 추출 기법을 활용하여 PPG 신호에서 혈압의 단계를 예측하는 기계학습 알고리즘을 제안한다. 연구에서 활용한 특징값 추출 기법은 PPG 신호의 피크 기반 특징값, tsfresh 패키지 기반 특징값, PPG 신호의 면적 기반 특징값이다. 각 특징값 추출 기법의 유효성을 확인하기 위해서 다양한 분야에서 사용되는 기계학습 알고리즘을 활용하였다. 기계학습 알고리즘과 특징값 추출 기법 중, XGBoost와 면적 기반 특징값 추출의 조합이 93%의 정확도와 0.99의 AUC를 보임으로써 혈압의 단계를 예측하는데 효과적인 조합임을 보였다.

## I. 서 론

심혈관 질환의 조기 식별을 위해 비침습적 조기 진단 및 선별 기법에 관한 연구가 많이 진행되고 있다. 이러한 방법 중 PPG(photoplethysmography, 광용적맥파)는 저가이면서 비침습적으로 심혈관 질환을 검출하는 기술로 널리 인식되고 있다. 본 논문에서는 피험자의 임상적 정보와 PPG 신호에서 추출한 여러 특징 정보를 기계학습 알고리즘에 적용하여 피험자의 혈압 상태, 즉 정상 혈압, 경계치 혈압(pre-hypertension), 고혈압을 예측하려고 한다.

## II. 본론

본문에서 사용한 PPG와 혈압 데이터는 Scientific Data 저널에 공개된 데이터로, 중국 Guilin People's Hospital에서 수집한 데이터이다.[1] 그림 1과 같이, 219명의 피험자는 10분간 휴식 후, 30초간 혈압을 측정하고, PPG를 2.1초간 3번 측정하였다.

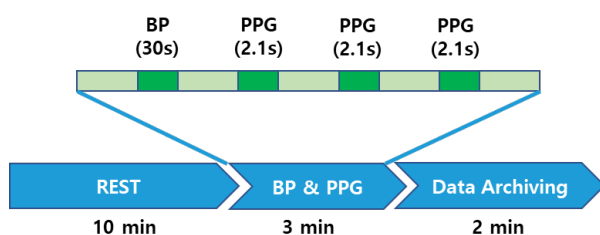


그림 1. 혈압과 PPG 신호 측정 프로토콜

원본 데이터는 혈압을 4 단계(정상, 경계치 혈압, 고혈압 1 단계, 고혈압 2 단계)로 분류하였지만, 각 혈압

단계의 데이터 수가 불균형하여 고혈압 1 단계와 2 단계를 병합하여 고혈압 단계로 분류하였다. PPG 신호는 1kHz로 샘플링 되었지만, 특징값 추출을 위하여 안티-앨리어싱(Anti-Aliasing) 필터가 적용된 데시메이션(decimation) 기법을 사용하여 200Hz로 다운 샘플링하였다. 그리고 PPG 신호에 버터워스 저역 통과 필터(Butterworth low-pass filter)를 적용하여 정상적인 PPG 주파수 범위(0.5Hz~5Hz) 이상의 주파수 신호는 차단하였고, NLM(non-local means)을 이용하여 잔여 노이즈를 제거하였다.

PPG 신호에서 유효한 특징값을 추출하기 위해서 아래와 같은 3가지 기법을 사용하였다. 첫 번째 방법은 그림 2와 같이 피크 기반 통계적 특징값 추출이다.[2] 이는 PPG 신호에서 추출된 피크에 기반하여 사분위수, 중앙값, 최대값, 최소값, 평균값, 표준편차, 분산, 평균제곱오차 등과 같은 통계적 특징값을 구한다.

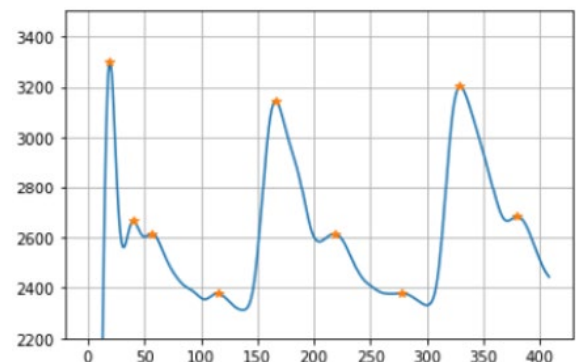


그림 2. PPG 피크 기반 특징값

두 번째 방법은 시계열 데이터의 특징값을 추출하는데 사용되는 tsfresh package를 활용하여 그림 3과 같은

특징값을 787 개 추출하는 방법이다.[3] 실험에서는 오버피팅 문제를 해결하기 위해서 787 개의 특징값 중 중요도가 높은 상위 10 개를 특징값을 사용하였다.

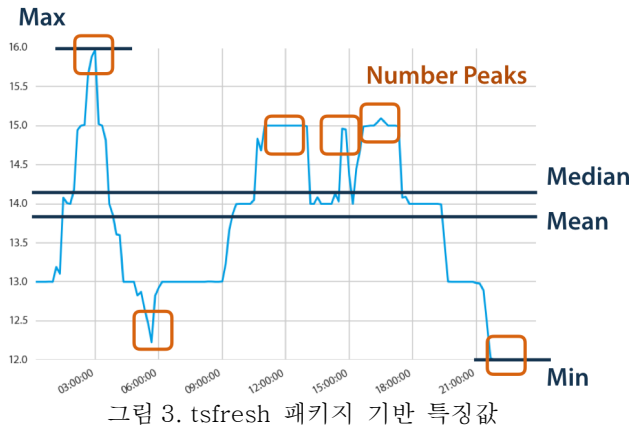


그림 3. tsfresh 패키지 기반 특징값

세 번째 방법은 그림 4 와 같이 PPG 신호의 면적을 기반으로 한 특징값 추출 방법이다.[4] 이 방법은 전체 PPG 신호에서 하나의 PPG 신호가 시작되는 지점에서 수축 피크(systolic peak)까지의 면적인 수축 영역(systolic area), 수축 피크에서 PPG 신호가 끝나는 지점까지의 면적인 이완 영역(diastolic area), 수축 영역과 이완 영역 면적의 합을 특징값으로 사용하였다.

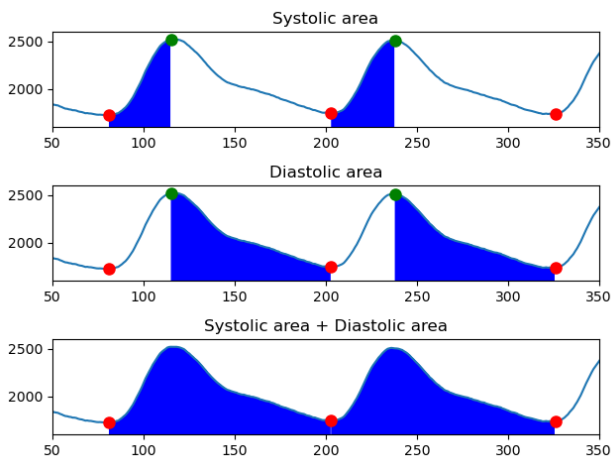


그림 4. PPG 신호의 면적 기반 특징값

위에서 소개한 세 종류의 특징값에 피험자의 임상적 정보(성별, 나이, 키, 체중, 심박수, BMI)를 각각 병합하여 혈압 단계를 예측하는 기계학습 알고리즘을 학습하였다. 이 때, 사용한 기계학습 알고리즘은 결정 트리(DT), 최근접이웃(kNN), 로지스틱 회귀(LR), 서포트 벡터 머신(SVM), 배깅(BAG), 랜덤 포레스트(RF), XGBoost(XGB) 등이다. 학습 모델의 하이퍼파라미터를 최적화하기 위해서 학습 데이터, 검증 데이터, 테스트 데이터를 각각 64:16:20 으로 나누어서 실험하였다.

### III. 결론

본 논문에서는 혈압 단계를 예측하기 위해서 피험자의 임상적 정보 뿐만 아니라 세가지 특징값 추출 기법을 사용하였다. 이렇게 추출된 특징값을 기계학습 알고리즘에 적용하여, 각 특징값 추출 방법에 대한 정확도와 AUC 를 구하였다. 표 1 은 각 특징값에 임상적 정보를 병합하여 기계학습 알고리즘에 적용하여 획득한

정확도(ACC)이며, 표 2 는 AUC 결과이다. 표 1 과 2 에서 나와있듯이 PPG 피크 기반 특징값에 기반한 진단력은 XGBoost 가 87.1%의 정확도와 0.95 의 AUC 를 보였다. 반면, tsfresh 패키지에서 추출된 특징값 중 중요도가 높은 상위 10 개에 기반한 진단력은 XGBoost 가 81.1% 정도의 정확도와 0.94 의 AUC 를 보였다. 마지막으로 PPG 신호의 면적을 기반으로 한 특징값에 기반한 결과는 XGBoost 에서 93.2%의 정확도와 0.99 의 AUC 를 보였다.

ACC	LR	RF	KNN	SVM	DT	XGB	BAG
피크	65.2	72.0	62.9	65.2	60.6	<b>87.1</b>	83.3
TS	80.3	75.8	72.0	80.3	52.3	<b>81.1</b>	78.8
면적	84.1	87.1	81.8	84.1	60.6	<b>93.2</b>	89.4

표 1. 특징값에 따른 머신러닝 알고리즘의 정확도 비교 (피크는 피크 기반 특징값, TS는 tsfresh 패키지 기반 특징값, 면적은 PPG 면적 기반 특징값을 의미)

AUC	LR	RF	KNN	SVM	DT	XGB	BAG
피크	0.82	0.88	0.71	0.82	0.78	<b>0.95</b>	<b>0.95</b>
TS	0.90	0.91	0.79	0.90	0.70	<b>0.94</b>	<b>0.94</b>
면적	0.94	0.98	0.86	0.94	0.77	<b>0.99</b>	<b>0.99</b>

표 2. 특징값에 따른 머신러닝 알고리즘의 AUC 비교

본 연구는 혈압 단계를 예측하기 위해서 PPG 신호의 형태학적 정보를 활용하여 특징값을 추출하여 각 기법의 유효성을 머신러닝 알고리즘을 통하여 보였다. 보다 면밀한 비교를 위해서 웨이블릿과 같은 주파수 영역에서 특징값을 추출하는 방법과 RNN, LSTM 과 같은 순환신경망을 활용하는 방법 등 다양한 방법에 대한 논의가 필요하다.

### ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로

한국연구재단의 지원을 받아 수행된 연구임

(No.2020R1G1A1102881).

### 참 고 문 헌

- [1] Liang, Y., Chen, Z., Liu, G., Elgendi, M. "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China," Sci. Data, 5, 180020, 2018.
- [2] Chowdhury, M.H., Shuzan, M.N.I., Chowdhury, M.E.H., Mahub, Z.B., Uddin M.M., Khandakar, A., Reaz, M.B.I." Estimating Blood Pressure from Photoplethysmogram Signal and Demographic Features using Machine Learning Techniques," Sensors, 20(11), 3127, Jun, 2020.
- [3] Christ, M., Braun, N., Neuffer, J., Kempa-Liehr A.W. "Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh - A Python package)," Neurocomputing, 307, pp. 72-77, 2018.
- [4] Park, J., Seok, H.S., Kim, S.S., Shin, H. "Photoplethysmogram Analysis and Applications: An Integrative Review," Front. Physiol., 12, 808451, 2020.