

선형-비선형 특징추출에 의한 비정상 심전도 신호의 랜덤포레스트 기반 분류

김혜진 · 김병남 · 장원석 · 유선국

연세대학교 의과대학 의학공학교실

Random Forest based Abnormal ECG Dichotomization using Linear and Nonlinear Feature Extraction

Hye-Jin Kim, Byeong-Nam Kim, Won-Seuk Jang and Sun K. Yoo

Department of Medical Engineering, Yonsei University College of Medicine

(Manuscript received 15 October 2015; revised 22 March 2016; accepted 23 March 2016)

Abstract: This paper presented a method for random forest based the arrhythmia classification using both heart rate (HR) and heart rate variability (HRV) features. We analyzed the MIT-BIH arrhythmia database which contains half-hour ECG recorded from 48 subjects. This study included not only the linear features but also non-linear features for the improvement of classification performance. We classified abnormal ECG using mean_NN (mean of heart rate), SD1/SD2 (geometrical feature of Poincaré HRV plot), SE (spectral entropy), pNN100 (percentage of a heart rate longer than 100 ms) affecting accurate classification among combined of linear and nonlinear features. We compared our proposed method with Neural Networks to evaluate the accuracy of the algorithm. When we used the features extracted from the HRV as an input variable for classifier, random forest used only the most contributed variable for classification unlike the neural networks. The characteristics of random forest enable the dimensionality reduction of the input variables, increase a efficiency of classifier and can be obtained faster, 11.1% higher accuracy than the neural networks.

Key words: Electrocardiogram (ECG), Linear feature, Nonlinear feature, Spectral Entropy, Dichotomization

1. 서 론

우리나라의 3대 주요 사망 원인중의 하나는 심장 질환이다. 심장 질환에 의한 사망률이 증가함에 따라 심장질환에 대한 연구를 비롯하여 진단의 정확성을 높일 수 있는 알고리즘 개발이 필요하다. 심장질환 중에서도 두드러지게 나타나는 원인은 부정맥으로 심장 박동이 정상보다 빠르거나 느려 규칙적이지 않은 상태이다. 이러한 심장질환을 진단하는 방법으로 심전도가 사용된다. 심전도는 심장의 전기적 활동

을 측정하는 것으로 비정상적인 리듬을 측정하고 진단하는 가장 좋은 방법이다. 심전도 신호에서 획득한 심박수변이도는 심장질환을 예측하는데 좋은 지표가 되기 때문에 부정맥을 진단하는 방법으로 이용되고 있다[1].

심장관련 질환을 자동적으로 진단하기 위한 방법으로 퍼지이론이나 신경망, SVM 등의 분류기가 주로 사용되었다 [2,3]. 이전에는 시간 영역 분석 방법과 주파수 영역 분석 방법을 사용하여 선형분석을 이용하여 연구가 진행되었다[4-6]. 하지만 심혈관계는 선형적으로 표현하기에는 너무 복잡하므로 비선형 분석 방법이 유용할 것으로 제안되었고, Owis가 심전도 신호를 이용한 부정맥 검출을 위해 비선형 분석 방법이 의미가 있음을 밝혀냈다[7,8].

본 논문에서는 부정맥을 검출하기 위하여 심박수변이도를 사용하였고 기존의 선형 분석 방법과 비선형 분석 방법을

Corresponding Author : Sun K. Yoo
Department of Medical Engineering, Yonsei University
College of Medicine
TEL: +82-2228-1919, E-mail : sunkyoo@yuhs.ac
본 연구는 연세대학교 의과대학 교내연구비(6-2014-0010)의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

이용하여 특징을 추출하였다. 선형 연구에서 사용된 신경망이나 SVM을 이용한 분류기는 변수 선정 과정을 거치지 않아 최적화된 분류 모델이라 하기 어려우며, 데이터의 수가 많아지면 학습 시간이 길어진다는 단점이 있다[9]. 특히나 신경망의 경우 트레이닝데이터에 과대적합(Overfitting) 되거나 분류 결과 지역 최소점에 빠질 수 있으며, 분류 과정을 알 수 없어 결과 해석이 불가하다[10]. 반면, 랜덤포레스트는 기존의 분류기와 달리 정확도에 영향을 미치는 변수를 평가하여 중요도가 높은 입력 변수를 선택적으로 사용함으로써, 모든 변수들을 사용하는 분류기보다 학습 속도가 빨라 분류 과정이 효율적이다[11]. 분류 정확도에 부정적인 영향을 미치는 입력변수는 분류과정에서 제외되므로 상대적으로 높은 분류 정확도를 가지게 된다. 따라서 본 논문에서는 랜덤포레스트를 사용하여 부정맥 분류기의 정확도를 향상시키는 것을 목적으로 한다.

II. 연구 방법

1. 신호 전처리

부정맥 분류기의 성능평가를 위해 MIT-BIH 부정맥 데이터베이스를 사용하였다. MIT-BIH 부정맥 데이터베이스는 Beth Israel 병원의 부정맥 연구소에서 1975년과 1979년 사이에 수행한 연구로부터 얻어진 4,000개 이상의 장시간 홀터 기록이다[12]. 이 중 임의로 선택된 23명의 기록과 임상적으로 중요한 증상을 가진 25명의 기록을 포함하여 총 48명 피험자의 데이터베이스를 구성하였다.

심박수변이도를 이용한 비선형 특징을 추출하기 위해서 R 피크(peak) 값을 정확히 추출하는 것이 중요하다. Pan과 Tompkins가 개발한 실시간 QRS 검출 알고리즘을 이용하여 R 피크를 검출한 후, R-R 간격이 32개 포함된 심박수변이도 구간을 만들었다[13]. 한 사람당 적게는 40개, 많게는 100개 이상의 구간을 가지고 있으며 모든 구간의 수는 총 3,420개이다. 정상 심전도의 R-peak 지점은 모두 검출하였으나 전체 심전도 중에서 15% 에 해당하는 부정맥 심전도는 정확한 검출이 되지 않아 MIT-BIH에서 제공하는 Annotation을 참고하여 R-peak 지점을 검출하였다.

2. 선형 분석 방법을 이용한 특징 추출

32개의 R-R간격으로 이루어진 심박수변이도 구간으로부터 선형 분석 방법을 이용하여 부정맥을 분류하기 위한 특징을 추출하였다. mean_NN은 R-R 간격의 평균값으로 심박동의 정도를 나타낸다. SDNN은 전체 R-R 간격의 표준편차를 나타낸다. SDNN의 감소는 좌심실의 이상과 관련이 있고, 좌심실 빈맥이나 갑작스러운 심정지 등을 예측할 수 있다[14]. RMSSD는 단기간의 심박변이를 나타내며 스트

레스와 같은 질환과 연관되어 값이 커지면 부교감신경의 활동이 활성화되었다는 것을 의미한다[15]. SDD는 R-R 간격 차이의 표준편차를 나타낸다. NN5, NN10, NN50, NN100은 연속적인 R-R 간격의 쌍이 각각 5 ms, 10 ms, 50 ms, 100 ms 이상 차이가 나는 것을 말한다[16]. pNN5, pNN10, pNN50, pNN100은 전체의 R-R 간격에 대해 NN5, NN10, NN50, NN100 값들의 비율(%)이며 값이 작을수록 건강한 상태를 의미한다.

3. 비선형 분석 방법을 이용한 특징 추출

심박수변이도의 비선형 분석 방법으로 스펙트럼 엔트로피와 푸앵카레 그래프를 이용하였다. 스펙트럼 엔트로피는 심박수변이도와 같이 예측하기 어려운 시계열 신호를 정량화하기 위해 사용되는 분석 방법이다[17]. 스펙트럼 엔트로피 값은 먼저 푸리에 변환 후 스펙트럼 파워를 계산한다. 모든 주파수에 대해 정규화 한 뒤, 식 1을 이용하여 스펙트럼에 대한 확률 밀도 함수 값을 추정한다[18].

$$p_i = S(f_i) \sum_{k=1}^n S(f_k) \text{ for } 1 \leq i \leq n \quad (1)$$

여기서 p_i 는 주파수 f_i 에 상응하는 확률 밀도함수이며, $S(f_k)$ 는 같은 주파수 요소에 대한 스펙트럼 에너지이다. 그리고 n 은 FFT의 주파수 성분 총 개수이다. 스펙트럼 엔트로피의 정의는 식 2와 같다.

$$H = -\sum_{i=1}^n p_i \log_2(p_i) \text{ for } 1 \leq i \leq n \quad (2)$$

다음으로, 푸앵카레 그래프를 통하여 심박수변이도의 비선형 특징을 추출하였다. 푸앵카레 그래프는 연속적인 R-R 간격의 상관관계를 그래프로 표현하는 방법이다. 푸앵카레 그래프에서 x축은 n번째의 R-R간격인 RR_n 값이며, y축은 n+1번째의 R-R간격인 RR_{n+1} 값, SD1과 SD2는 푸앵카레 타원에서 얻을 수 있는 파라미터이다[19]. SD1은 $y = -x$ 로 투영시킨 직선의 표준편차이고 SD2는 $y = x$ 로 투영시킨 직선의 표준편차이다. SD1은 식 3으로 인접한 R-R 간격 차이의 표준편차인 SDD 값을 이용해서 계산하였고, SD2는 R-R간격의 표준편차인 SDNN과 SDD 값으로 계산하여 식 4와 같다[20]. 이때 p_1 과 p_2 는 식 5로부터 구할 수 있다.

$$SD1^2 = \text{Variation}(p_1) = \text{Var}\left(\frac{1}{\sqrt{2}}RR_n - \frac{1}{\sqrt{2}}RR_{n+1}\right) \quad (3)$$

$$= \frac{1}{\sqrt{2}}\text{Var}(RR_n - RR_{n+1}) = \frac{1}{2}SDD^2$$

$$SD2^2 = 2SDNN^2 - \frac{1}{2}SDD^2 \quad (4)$$

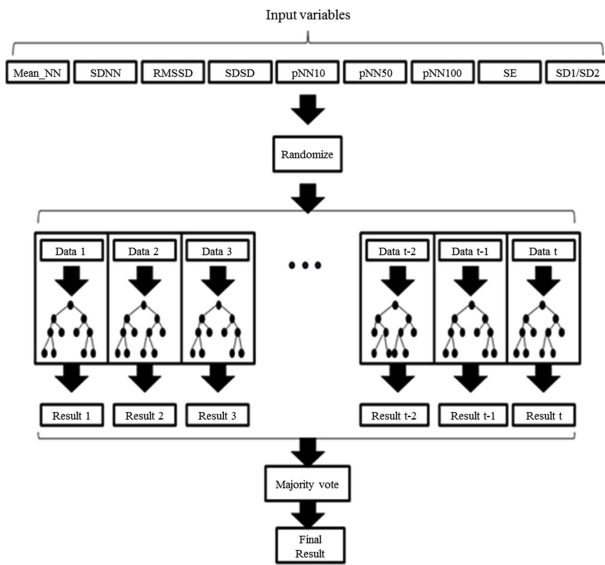


그림 1. 랜덤포레스트 알고리즘 다이어그램.
Fig. 1. Flow diagram of Random forest.

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} RR_n \\ RR_{n+1} \end{bmatrix} \quad (5)$$

4. 부정맥 분류

심전도 신호에서 R-R간격을 추출한 후 32개의 R-R간격을 포함하는 한 구간이 만들어지면 선형·비선형 분석방법을 이용하여 특징을 추출하였다. 추출한 특징들의 조합에 따라 트리의 개수를 여러 경우로 바꾸어가면서 랜덤포레스트를 이용하여 10-fold cross validation으로 정상구간과 부정맥 구간을 분류하였다. 랜덤포레스트는 그림 1과 같이 다수의 의사결정나무(Decision tree)로 구성된 앙상블 학습 방법이다[11]. 결정 트리는 복잡한 문제를 계층 구조 형태로 간단히 나눔으로써 간편하고 빠른 학습과 테스트가 가능하며, 임의의 의사결정 트리를 구성하여 앙상블 학습을 수행함으로써 높은 정확성과 일반화의 성능을 나타낸다. 앙상블 학습 방법은 여러 개의 분류기를 생성하고 그것들의 예측을 결합함으로써 새로운 가설을 학습하는 방법이다. 앙상블 학습의 목표는 여러 분류기의 예측 결과들을 종합함으로써 단일 분류기와 비교하여 신뢰성이 높은 예측 값을 얻는 것이

표 1. MIT-BIH 부정맥 데이터베이스의 비트 주석.
Table 1. Beat annotation of MIT-BIH arrhythmia database.

Code	Description
B	Bundle branch block beat (unspecified)
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature or ectopic beat
V	Premature ventricular contraction
R	R-on-T premature ventricular contraction
F	Fusion of ventricular and normal beat
e	Atrial escape beat
j	Nodal (junctional) escape beat
N	Supraventricular escape beat (atrial or nodal)
E	Ventricular escape beat
?	Beat not classified during learning
!	Ventricular flutter wave
x	Non-conducted P-wave (blocked APC)

다. 랜덤포레스트는 배깅을 이용하여 앙상블 학습을 수행한다. 배깅은 분석에 사용되는 데이터 세트에 대하여 중복을 허용하는 부트스트래핑 표본 추출 방법을 적용하여 다수의 샘플 집합을 구성하고, 각 샘플 집합을 학습용 데이터 세트로 하여 분류기 집합을 만든 후 복합 분류기를 생성하는 방식이다.

랜덤포레스트의 분류 결과를 성능평가하기 위해 신경망 이론을 적용하여 부정맥을 분류한 M.G. Tsipouras의 선형 연구와 결과를 비교하였다[21]. 이때 심전도 데이터의 좌각 블록, 우각블록, 심박조율기(pace maker)에 의한 비트는 선형 연구를 참고하여 정상으로 가정하였다[22]. 부정맥 분류 기준은 부정맥의 경우 표 1과 같이 MIT-BIH에서 제공하는 비트 주석(beat annotation)을 참고하였으며, 그림 2와 같이 한 구간에서 정상 주석(*)이 30개 이상이면 정상, 부정맥과 관련한 주석이 3개 이상이면 부정맥이라 분류한 결과와 비교하여 각 알고리즘의 분류 정확도를 계산하였다.

III. 결 과

심박수변이도에 대한 선형적 분석 방법으로 추출된

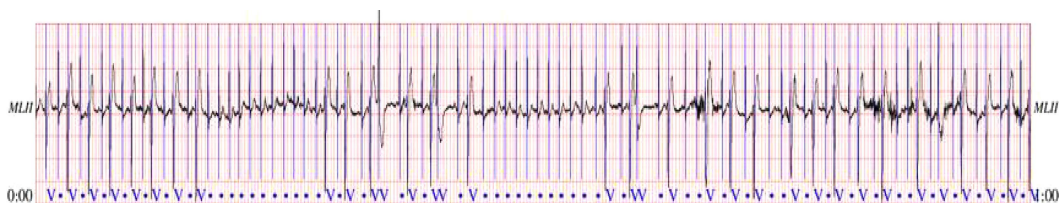


그림 2. MIT-BIH 부정맥 데이터베이스 주석 (*: 정상, V: 조기심실수축).
Fig. 2. Annotation of MIT-BIH arrhythmia database (*: Normal, V: Premature ventricular contraction).

mean_NN, SDNN, RMSSD, SDDSD, pNN5, pNN10, pNN50, pNN100 특징들에 대해 정상구간과 부정맥 구간에 대해 비모수 검정(mann-whitney test)을 실시하여 통계적으로 유의미한 차이가 있음을 확인하였다. pNN5은 유의미한 차이를 확인 할 수 없었다. 정상과 부정맥의 각 특징별 평균(표준편차)값은 mean_NN의 경우 0.81(0.17)와 0.74(0.15)이며, SDNN은 0.05(0.05)와 0.16(0.14), RMSSD는 0.07(0.06)과 0.25(0.21), SDDSD는 0.07(0.06)과 0.25(0.21), pNN10은 75.20%(13.79)와 86.81%(11.76), pNN50은 28.96%(24.54)와 56.23%(28.89), pNN100은 21.86%(19.51)과 47.11%(28.16) 이었다. 검정 결과 유의 수준 95%를 만족하면서 0.001 이하의 유의 확률로 그룹간의 통계적 유의미한 차이가 있음을 확인하였다.

비선형적 분석 방법으로 심박수변이도로 부터 추출된 Spectral Entropy와 SD1/SD2 특징들을 정상구간과 부정맥 구간에서도 비모수 검정(mann-whitney test)을 실시하여 통계적으로 유의미한 차이가 있음을 확인하였다. Spectral Entropy는 정상과 부정맥의 평균(표준편차)이 0.11(0.01)과 0.15(0.05)였으며, SD1/SD2의 정상과 부정맥의 평균값은 0.85(0.37)과 1.58(1.55) 였다. 검정 결과 모든 특징들에서 유의 수준 95% 에서 유의 확률이 0.001 이하로 그룹간의 유의미한 차이가 있음을 확인하였다.

랜덤포레스트를 이용한 부정맥 분류 성능을 평가하기 위하여 신경망 이론을 적용하여 부정맥을 분류한 M.G. Tsipouras의 선형 연구와 분류 결과를 비교하였다[21]. M.G. Tsipouras는 심박수변이도에서 선형적 분석 방법을 통해 추출한 총 6가지의 특징으로 조합 가능한 63가지의 경우에 따라서 신경망을 이용하여 정상 구간과 부정맥 구간을 분류하였고 민감도와 특이도 값을 계산하였다. 같은 특징 조합을 이용하여 랜덤포레스트로 부정맥 구간을 분류한 결과와 비교하였다. 그림 3는 신경망과 랜덤포레스트 분류 방법에 따른 민감도와 특이도의 차이를 분석하였다. 정규성 검정을 통해 정규분포를 띄지 않는 것을 확인하였으며, 이에 각 분석 방법간의 비모수 검정(mann-whitney test)을 실시하였다. 신경

망과 랜덤포레스트의 민감도 평균(표준편차)값은 74.14%(4.86)와 84.72%(3.89)이며, 특이도의 평균(표준편차)값은 71.71%(10.30)과 66.00%(7.66) 이었다. 검정 결과 유의 수준 95% 에서 유의 확률이 0.001 이하로 그룹간의 유의미한 차이가 있음을 확인하였다.

심박수변이도의 특성을 표현하는데 있어 선형적 방법으로 추출한 특징만으로는 한계가 있어 비선형적 방법으로 추출한 특징을 포함하여 랜덤포레스트 방법으로 부정맥을 분류하였다. 심박수변이도의 선형 특징에 mean_NN과 비선형 특징 SE(Spectral Entropy), SD1/SD2을 추가적으로 사용하였으며, pNN5 특징 대신에 pNN100을 사용하였다. pNN5의 경우는 정상 데이터와 부정맥 데이터의 값이 유사하게 나타나 특징으로 적합하지 않다고 판단하였다[22]. 총 9가지의 특징을 입력으로 하여 랜덤포레스트 분류기를 사용하였다. 표 2와 같이 트리의 수를 변경하여 분류기의 민감도, 특이도, 정확도를 측정하였다. 트리의 수가 많아질수록 정확도가 높아졌으며 100개 이상일 때는 거의 일정한 값을 가지며 정확도 변화가 없었다.

랜덤포레스트의 장점은 분류기의 입력 특징으로 사용된 변수의 중요도를 평가할 수 있는 것이다. 여기서 입력 변수는 총 9개의 심박수변이도 특징을 말한다. 변수 중요도를 평가하기 위해 통계프로그램 R을 이용하였고, 비교를 위해 신경망과 랜덤포레스트를 함께 분석하였다. 총 9개의 입력 변수를 모두 사용하였을 때 두 분류기 모두 높은 정확도를 나타내었다. 신경망의 경우 입력 변수로 이용되는 특징들을 패턴 분류하는데 모두 사용하는 반면 랜덤포레스트는 무작위로 선택된 특징들의 조합들 중에서 중요한 변수로 영향을 미치는 특징만을 패턴 분류에 사용하므로 최소 1개 최대 4개의 변수를 사용하였다. 입력변수의 수를 늘려가며 선택된 조합들의 부정맥 분류 결과의 평균값은 표 3과 같다.

패턴분류에 실질적으로 사용되는 특징의 수가 신경망은 1개에서 9개까지 증가함에 따라 랜덤포레스트는 신경망과 같은 수이거나 적은 수의 특징만을 이용함에도 더 나은 분류 정확도를 나타내었다.

64

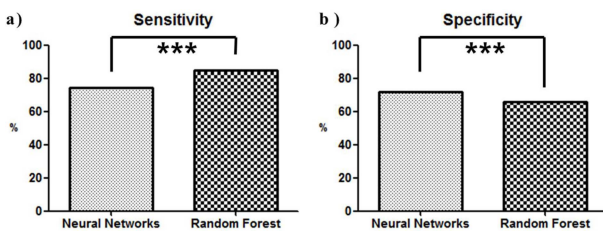


그림 3. 신경망과 랜덤포레스트 분류 방법에 따른 비모수 검정 a)민감도 b)특이도.

Fig. 3. Nonparametric test graph by classification method a) Sensitivity, b) Specificity.

표 2. 랜덤포레스트의 트리 수에 따른 평균 분류 정확도.

Table 2. The average classification accuracy by the number of tree.

number of trees	Sensitivity(%)	Specificity(%)	Accuracy(%)
10	94.6	84.7	90.8
20	94.6	86.6	91.6
30	94.9	87.1	91.9
40	94.6	87.6	92.0
50	94.6	88.1	92.1
100	94.5	88.8	92.3

표 3. 분류에 사용된 특징 수에 따른 평균 정확도.

Table 3. The average accuracy by the number of features used for classification.

number of features		Accuracy (%)	
Neural Network	Random Forest	Neural Network	Random Forest
1	1	62.0	62.0
2	2	80.5	86.1
3	3	82.1	88.5
4	3	82.1	88.6
5	3	80.9	89.4
6	3	83.7	90.9
7	4	83.8	91.7
8	4	83.9	92.2
9	4	84.4	92.3

표 4. 변수 중요도에 따른 4개의 특징을 사용한 평균 분류 정확도.

Table 4. The average classification accuracy using four features in accordance with variable importance.

	Neural Network	Random Forest
Accuracy (%)	83.0	92.2

군 92.2% 로 확인되었다. 두 분류 방법의 정확도 평균을 비교한 결과 유의 확률 0.01 미만으로 알고리즘에 따른 분류 정확도에 유의미한 차이가 있음을 확인하였다.

IV. 고 찰

본 논문에서는 심전도 신호를 이용하여 부정맥을 진단하기 위해 심박수변이도와 랜덤포레스트를 이용하는 방법을 제시하였다. 심박수변이도로부터 선형적 방법으로 추출한 특징만으로는 한계가 있어 비선형적 방법을 추가하여 특징들을 획득하였고 이를 이용하여 부정맥 분류 정확도를 향상할 수 있었다. 신경망을 이용한 부정맥 분류 결과와 비교하였을 때 민감도는 랜덤포레스트 방법을 사용하였을 때 유의미한 차이로 더 높게 나왔고, 특이도는 신경망 방법이 유의미한 차이로 높게 확인되었다. 이는 부정맥이 때 순간 나타나는 것이 아니라 잠시 나타났다 사라지기 때문에 부정맥 구간의 데이터 수보다 정상 구간의 데이터 개수가 훨씬 많아 바이어스가 생기기 때문이다. 이러한 데이터 불균형의 문제를 해결하기 위해서 랜덤포레스트의 적은 수를 가지는 클래스에 가중치 주는 weighted 랜덤포레스트나 데이터의 수를 균형을 맞추는 balanced 랜덤포레스트를 이용하는 방법을 적용하면 정확도가 더 높아질 것으로 생각된다. 같은 조건에서 신경망을 이용한 부정맥 분류 결과 보다 높은 정확도를 얻을 수 있었던 것은 랜덤포레스트가 가지는 장점 때문이라 생각된다. 신경망을 이용한 부정맥 분류의 경우 모든 특징을 사용하기 때문에 부정맥을 분류하는데 불필요한 특징까지 포함하고 있다. 반면 랜덤포레스트의 경우는 결정 트리가 분기할 때 모든 특징을 평가하여 선택하는 것이 아니라 특정 개수를 무작위로 선택하여 평가함으로써 랜덤포레스트의 결정 트리상의 상관관계를 줄여주고 분류기의 정확도를 향상하는 결과를 가져왔다. 이는 변수 중요도를 평가함으로써 불필요한 특징에 인한 정확도가 저하되는 것을 막고 효율적인 분류기를 만들 수 있을 것으로 생각된다. 신경망의 경우 학습률(learning rate), 모멘텀(momentum), 반복 횟수(iteration), 은닉층(hidden node)의 개수 등 고려해야 할 요인이 많아 최적의 모델을 찾는 것이 어려워 정확도가 높은 분류기 조건을 찾는 것이 어려운 반면, 랜덤포레스트의 경우 트리의 개수만을 바꿔 주는 것만으로도 상대적으로 나은 결과를 얻을 수 있다.

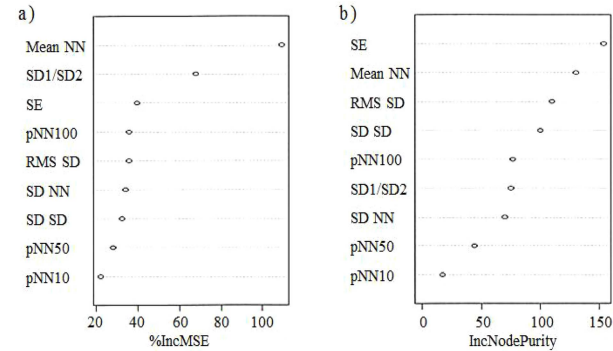


그림 4. 랜덤포레스트 변수 중요도.
Fig. 4. Random forest variable importance.

그림 4는 랜덤포레스트의 결정 트리를 결정하는데 상위 노드에 선정되는 특징들을 확인하기 위해 R을 이용하여 변수 중요도 평가를 수행한 결과이다. 그림 4(a)의 X축은 각각의 변수를 입력변수로 포함되는 유무에 따라 분류 결과값의 평균제곱오차(Mean Square Error)를 나타내며 이는 분류 결과에 미치는 영향을 의미한다. 그림 4(b)의 X축은 각각의 변수를 대신하여 무작위 값을 입력변수로 하였을 때 분류 결과가 하락할 확률로 값이 클수록 분류 결과에 미치는 영향이 큰 것을 의미한다. 그림 4를 통해 분류 결과에 영향을 미치는 중요한 특징들이 mean_NN, SD1/SD2, SE, pNN100 순으로 확인됨을 알 수 있었다. 표 4는 중요도에 따른 4개의 특징들만을 이용하여 분류기를 구성한 신경망과 랜덤포레스트의 분류 정확도에 대해 통계분석을 하였다. 유의수준 95%의 정규성 검정을 실시한 결과 두 그룹 모두 정규성을 띄는 것으로 확인되었다. 정규성 검정 결과를 바탕으로 분류 알고리즘에 따른 유의수준 95%의 t-test를 실시하였다. 검정 결과 신경망을 이용하여 분류한 정확도 평균은 83.0%이며 랜덤포레스트를 적용한 분류 정확도는 평

심박수변이도에서 예측하기 어려운 시계열 특징을 추출하기 위하여 비교 논문과는 다르게 비선형 분석 방법을 추가로 사용하였다. 그 결과 선형 분석만 사용했을 때보다 정확도가 향상되었음을 알 수 있었지만, 탈경향변동분석(detrended fluctuation analysis)이나 최대 리아프노프 지수(largest lyapunov exponent) 등 비선형 특징들은 긴 시간의 데이터가 요구되므로 짧은 시계열 구간에서 부정맥 특징을 추출하기에는 한계점이 있었다. 짧은 시계열 구간에도 심박수변이도의 비선형 분석을 통하여 부정맥 분류할 수 있는 연구에 대한 조사가 필요할 것으로 생각된다.

V. 결 론

심전도로부터 심박수변이도 구간을 추출한 후에 선형, 비선형 분석을 통하여 특징들을 획득한 후 랜덤포레스트를 통해 부정맥을 분류하는 방법을 제시하였다. 분석 결과 선형, 비선형 특징 추출 방법과 랜덤포레스트를 이용했을 때 정확도가 92.3%로 신경망을 적용하였을 때의 85% 분류 정확도보다 높음을 확인하였다. 부정맥 분류기의 분류 정확도를 향상시킴으로써 보다 정확한 부정맥 분류를 할 수 있었다. 더 나아가 부정맥의 조기 진단과 치료를 통해 심장질환으로 인한 사망률을 낮출 수 있을 것으로 생각된다. 추후 연구로 시간주파수 영역 분석방법이나 상관차원(correlation dimension) 등으로부터 추출된 특징변수들을 부정맥 분류에 추가한다면 보다 더 정확하며 세분화된 부정맥 분류가 가능할 것으로 생각된다. 랜덤포레스트를 이용하여 변수 중요도 평가를 하면 단순히 부정맥을 잘 분류하는 것뿐만 아니라 나이, 성별 등의 명목형 변수도 고려하여 어떤 경우에 부정맥이 생기는지에 대한 연구도 진행할 수 있을 것으로 생각된다.

References

- [1] U.R. ACHARYA, K.P. JOSEPH, N. KANNATHAL, C.M. LIM and J.S. SURI, "Heart rate variability: a review", *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031-1051, 2006.
- [2] E. ZELLMER, F. SHANG and H. ZHANG, "Highly accurate ECG beat classification based on continuous wavelet transformation and multiple support vector machine classifiers", *In Proceeding of the 2009 2th International Conference on Biomedical Engineering and Informatics (BMEI 2009)*, 2009, pp. 1-5.
- [3] M. ENGIN, "ECG beat classification using neuro-fuzzy network", *Pattern Recognition Letters*, vol. 25, no. 15, pp. 1715-1722, 2004.
- [4] Y.-C. YEH, W.-J. WANG and C.W. CHIOU, "Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals", *Measurement*, vol. 42, no. 5, pp. 778-789, 2009.
- [5] I. CHRISTOV, G. G. MEZ-HERRERO, V. KRASTEVA, I. JEKOVA, A. GOTCHEV and K. EGI AZARIAN, "Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification", *Medical Engineering & Physics*, vol. 28, no. 9, pp. 876-887, 2006.
- [6] B. ANURADHA, K.S. KUMAR and V.V. REDDY, "Classification of cardiac signals using time domain methods", *ARPN Journal of Engineering and Applied Sciences*, vol. 3, no. 3, pp. 7-12, 2008.
- [7] R. ACHARYA, A. KUMAR, P. BHAT, C. LIM, S. LYENGAR, N. KANNATHAL and S. KRISHNAN, "Classification of cardiac abnormalities using heart rate signals", *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 288-293, 2004.
- [8] M. OWIS, A.H. ABOU-ZIED, A.-B.M. YOUSSEF and Y.M. KADAH, "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification", *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 7, pp. 733-736, 2002.
- [9] E. YILMAZ, "An expert system based on Fisher score and LS-SVM for cardiac arrhythmia diagnosis", *Computational and Mathematical Methods in Medicine*, vol. 2013, 2013.
- [10] J.V. TU, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes", *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225-1231, 1996.
- [11] L. BREIMAN, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [12] G.B. MOODY and R.G. MARK, "The impact of the MIT-BIH arrhythmia database", *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001.
- [13] J. PAN and W.J. TOMPKINS, "A real-time QRS detection algorithm", *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 230-236, 1985.
- [14] G.C. CASOLO, P. STRODER, C. SIGNORINI, F. CALZOLARI, M. ZUCCHINI, E. BALLI, A. SULLA and S. LAZZERINI, "Heart rate variability during the acute phase of myocardial infarction", *Circulation*, vol. 85, no. 6, pp. 2073-2079, 1992.
- [15] T.F.O.T.E.S.O. CARDIOLOGY, "Heart rate variability standards of measurement, physiological interpretation, and clinical use", *European Heart Journal*, vol. 17, pp. 354-381, 1996.
- [16] J. MIETUS, C. PENG, I. HENRY, R. GOLDSMITH and A. GOLDBERGER, "The pNNx files: re-examining a widely used heart rate variability measure", *Heart*, vol. 88, no. 4, pp. 378-380, 2002.
- [17] R. ACHARYA, S.M. KRISHNAN, J.A. SPAAN and J.S. SURI, *Advances in cardiac signal processing*, Springer, 2007.
- [18] C. KAMATH, "Quantification of electrocardiogram rhythmicity to detect life threatening cardiac arrhythmias using spectral entropy", *Journal of Engineering Science and Technology*, vol. 8, pp. 588-602, 2013.
- [19] J. PISKORSKI and P. GUZIK, "Filtering poincare plots", *Computational Methods in Science and Technology*, vol. 11, no. 1, pp. 39-48, 2005.
- [20] M. BRENNAN, M. PALANISWAMI and P. KAMEN, "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?", *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1342-1347,

- 2001.
- [21] M.G. TSIPOURAS, D.I. FOTIADIS and D. SIDERIS, "An arrhythmia classification system based on the RR-interval signal", *Artificial Intelligence in Medicine*, vol. 33, no. 3, pp. 237-250, 2005.
- [22] M.G. TSIPOURAS and D.I. FOTIADIS, "Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability", *Computer Methods and Programs in Biomedicine*, vol. 74, no. 2, pp. 95-108, 2004.