

# 머신 러닝을 이용한 발작 상태의 근전도 신호 분류

정수빈, 남상훈, 서보민, 최여은, 안혜성, 신태민\*

\*연세대학교 의공학과

integerbin@naver.com, tmshin@yonsei.ac.kr

## Classification of EMG signal in seizure state using machine learning

Jung Su Bin, Nam Sang Hoon, Seo Bo Min, Choi Yeo Eun, An Hye Seong, Shin Tae Min\*

Department of Biomedical Engineering, Yonsei University, Korea

### 요 약

본 논문에서는 각 피험자에게서 추출한 근전도 데이터셋의 상관관계를 분석하고 발작 및 정상 상태를 분류할 특징점들을 선별하였다. 연관도가 높은 파라미터로 TM, DASDV, PSR, VCF 가 선정되었다. 데이터 분류에 사용된 모델은 MLP, ANN, K-NN, Naive Bayes, Random Forest, Decision Tree, SVM, Logistic Regression 으로 총 8 종류가 사용되었고 각 모델에 대한 성능평가를 진행하였다.

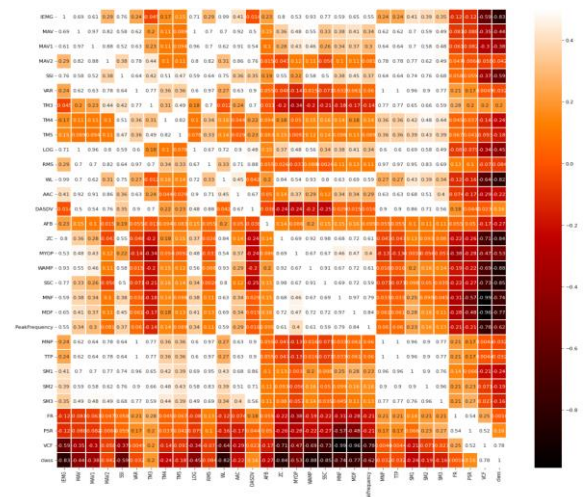
### I. 서 론

발작은 뇌의 정상적인 전기 신호 시스템에 문제가 발생하여 뇌세포에서 통제되지 않은 전기신호가 생성되어 나타나는 증상을 말한다. 발작이 없는 정상 상태일 때의 환자의 신경학적 검진 결과는 보통 정상이다. 대부분의 발작은 불확실성을 가지고 있으므로 의료진이 존재할 때 발작이 일어날 가능성은 희박하다. 또한 발작을 진단하는 데 있어 의료진의 주관이 개입하기 때문에 정량적인 진단 방법이 필요한 실정이다. 본 논문에서는 각 피험자에게서 추출한 근전도 데이터셋의 상관관계를 분석하고 발작 및 정상 상태를 분류할 특징점들을 선별하였다. 연관도가 높은 특징점 데이터셋에 대해 8 종류의 모델을 사용하여 분류성능평가를 진행하였다.

### II. 본론

발작 데이터셋은 Dryad 의 <Ictal quantitative surface electromyography correlates with postictal EEG suppression> 공개 데이터셋을 사용하였다. 발작 상태에서 측정된 EEG 신호와 동시에 측정된 EMG 신호를 추출하여 사용하였고 연구에 사용된 EMG 신호는 삼각근, 이두근, 삼두근에 부착되어 측정되었다. 정상적인 상태의 근전도 데이터는 IEE-IA 의 공개 데이터셋 sEMG 데이터베이스를 사용하였다. 각 피험자들은 비디오를 보며 17 종류의 특정 행동을 반복하였으며 전극은 이두근 및 삼두근에 부착되었다. 발작 상태의 데이터셋과 정상 상태의 데이터셋의 측정 환경이 다르다는 문제점을 해결하기 위해 데이터 전처리를 진행하였고 시계열 데이터를 사용하지 않고

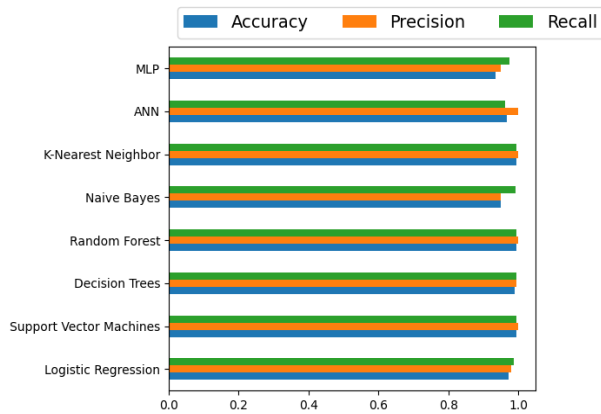
전처리 과정에서 획득한 특징점들의 데이터셋을 새로 구축하였다. 총 18,360 개의 EMG 데이터가 사용되었으며 30 개의 특징점들을 추출하였다. Heatmap 을 통해 데이터의 상관관계를 분석한 결과 연관도가 높은 파라미터는 TM(Temporal Movement of order x of the signal), DASDV(Difference Absolute Standard Deviation Value), PSR(Power Spectrum Ratio of the signal), VCF(Variance of the Central Frequency of the PSD) 였다.



<그림 1. Heatmap 을 통한 데이터 상관관계 분석>

### III. 결론

발작 데이터 분류에 사용된 모델은 MLP, ANN, K-NN, Naive Bayes, Random Forest, Decision Tree, SVM, Logistic Regression 으로 총 8 종류의 모델을 사용하여 분류성능평가를 진행하였다. 정확도, 정밀도, 재현율을 비교한 결과는 다음과 같다.



<그림 2. 사용한 모델의 분류성능평가 결과>

정확도 측면에서 K-NN, SVM, Random Forest 모델이 가장 우수한 성능을 보였고 정밀도 측면에서 ANN, K-NN 및 Random Forest, SVM 모델이 가장 우수한 성능을 보였다. 재현율 측면에서는 K-NN, Random Forest 및 SVM 모델이 가장 우수한 성능을 보였다.

### 참 고 문 헌

- [1] Tanvir Ahmed and Md. Kafiul Islam 2021 J. Phys.: Conf. Ser. 1921 012043
- [2] Raez MB, Hussain MS, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biol Proced Online. 2006;8:11-35. doi: 10.1251/bpo115. Epub 2006 Mar 23. Erratum in: Biol Proced Online. 2006;8:163. PMID: 16799694; PMCID: PMC1455479
- [3] [https://www.ufrgs.br/ieelab/resource\\_IEE\\_sEMG\\_db.php](https://www.ufrgs.br/ieelab/resource_IEE_sEMG_db.php)
- [4] <https://datadryad.org/stash/dataset/doi:10.5061/dryad.14004q5>