

♡. °° . 파킨스병의 예측 . °° . ♡ DScan 감소 측정보다 저렴한 Logistic linear Regression

파킨스병은 뇌의 도파민 전달체(DScan)가 감소하면서 움직임에 장애가 나타나는 질환이다. DScan의 감소 측정은 파킨스병을 진단하는 데 정확하다. 하지만 한 번 검사 받는데 약 430만 원 이상의 비용이 들기 때문에 파킨스병의 발병 유무를 더 저렴하게 찾고자 한다.

데이터 세트에는 몇 년 동안 추적 관찰된 파킨스병이 있거나 없는(PD vs. HC) 수백 명의 환자가 포함된다. 이 데이터 세트에 머신러닝을 적용해 파킨스병의 진단에 사용하고자 한다.

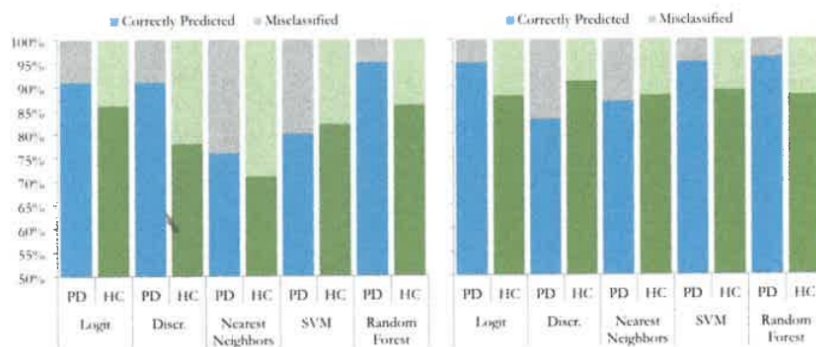
DScan을 측정했을 때와 비슷한 성능을 보이면서 더 저렴한 방법을 찾기 위해 :

- DScan 감소와 상관관계가 있는 임상적 feature는 무엇인가?
- DScan 감소의 예측 모델을 제시할 수 있는가?
- 예측 모델을 제시할 수 있다면, 가장 중요한 feature는 무엇인가?

HC-PD 분류 학습 알고리즘에 의한 DScan 감소를 예측하는 데 중요한 임상적 특징이 세 가지로 발견되었다.

- Hoehn and Yahr motor score
- 의사가 측정한 unified Parkinson Disease Rating score
- University of Pennsylvania Smell Identification score

이 세 가지 특징을 사용해 파킨스병의 예측을 95% 이상 성공할 수 있다.



위의 그림에서 왼쪽에 있는 그림은 20개의 feature를 사용했을 때 머신러닝 알고리즘에 대한 성공률을 나타내고 오른쪽 그림은 세 가지 feature만을 사용해 머신러닝 알고리즘에 대한 성공률을 나타낸다.

Hoehn and Yahr motor score와 의사가 진단하는 Smell Identification score는 파킨스 병과 관련해 DScan 감소의 가장 강력한 예측 변수이다. 따라서 이 두 가지 특징이 있는 linear regression 모델은 DScan 감소의 예측 모델이다.

파킨스병에 걸렸는지 여부를 예측하기 위해 다양한 알고리즘에서 검증을 실시했다. SVM과 Random Forest의 성능이 가장 우수했지만 다른 알고리즘들과 성능 차이가 크게 차이 나지 않는다. 따라서 해석하기 가장 쉬운 logistic linear regression 모델을 사용한다.

"Written by Choi, Ji-Hyeon"

∠(↯_∠)_