Detection of Parkinson's Disease

Hilal Ilgaz, Beyza Akkoyun
hilal.ilgaz@gazi.edu.tr, beyza.akkoyun@gazi.edu.tr

Motivation

Parkinson's Disease (PD) is the most common neurological disease after Alzheimer's disease, mostly among people over 60 years of age. The effects of the disease can be minimized by early diagnosis and good monitoring of the disease[3]. Speech disorders are used to diagnose the disease early and monitor the course of the disease.

Our group aims to predict whether a patient has Parkinson's Disease or healthy, using the voice patterns data of some PD patients and healthy people, and machine learning algorithms.

Data & Features

The dataset was created by extracting 23 features from biomedical voice measurements from 23 Parkinson's disease patients and 8 healthy people. Every column in the table is a biomedical voice measures and every row meets one of 195 voice recording. Out of these 195 rows, 147 rows are Parkinson's disease patients, and 48 rows are healthy people. If data point is labeled as 1, this point belongs to a PD patient. Else, that point belongs to a healthy person[1].

Models

Bagging: Baseline model is Decision Tree with max depth 1 and criterion Gini index.

Stacking: Meta classifier is Logistic Regression. Stack classifiers are Support Vector Machine, Decision Tree, Random Forest Classifier.

Boosting: Base estimator is Decision Tree. Algorithms using base estimator are XGBoost and AdaBoost.

Results

Model	Accuracy	Precision	Recall	F1 Score
SVM	85.64%	84.22%	100%	90.99%
Decision Tree	81.02%	86.94%	92.67%	89.05%
Logistic Regres.	91.05%	87.62%	95.77%	91.06%
Random Forest	88.2%	90.37%	95.69%	93.54%
AdaBoost	92.17%	93.22%	95.27%	94.13%
XGBoost	89.74%	91.30%	94.44%	92.61%

Table 1: Summary of results from different models

Note: These are the best results of different parts.

Discussion

- From our observations of the literature, PD is a subject that is often studied with different variations of similar models.
- By the way, it is a binary classification problem, because of that accuracy score is not the most important metric for the problem.
- Support Vector Machine is most common algorithm for the problem[2], but it is less successful than boosting algorithms.

Future Work

In the future we hope to modify our models to takes into different features like MRI images or EEG signals. Adding these will make the project more comprehensive and provide stronger bases for diagnosis. We also hope to make more advanced systems with various types of neural networks.

References

[1] Little M.A. et al. (2007). Biomedical Engineering Journal 6(23). DOI: https://doi.org/10.1186/1475-925X-6-23

[2] Lahmiri S. et al. (2019). Biomedical Signal Processing and Control vol. 49. DOI:https://doi.org/10.1016/j.bspc.2018.08.029.

[3] Badem H. (2019). Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi 8(2). DOI: https://doi.org/10.28948/ngumuh.524658.