

Detection of Parkinson's disease by selection of acoustic variables

Mónica Giuliano¹, Silvia N. Pérez¹, Luis Fernández²

¹ Universidad Nacional del Oeste, Argentina

² Universidad Nacional de La Matanza, Argentina

Abbreviated abstract:

We present the progress made in the comparative speech processing of people with and without Parkinson's disease (PD) of the same age range. The experiments have been performed with phonation samples of the vowel /a/ for which 297 parameters have been extracted. Were selected significant variables using ANOVA, PCA and logistic regression models with different penalty options (LASSO), obtaining a model with high predictive capacity. This allows the identification of speech disorders that correlate with the disease.

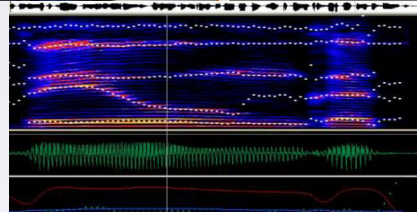
Related publications:

- M. Giuliano; L. Fernández.; S. Pérez. Selección de Medidas de Disfonía para la Identificación de Enfermos de Parkinson. (2020) IEEE Congreso Bienal de Argentina (ARGENCON).
- Giuliano, D. Adamec, M.I. Debas. Construcción de una base de voz de personas con y sin enfermedad de Parkinson. REDDI, 6 (1) (2021).



Problem, Data, Previous Works

- Problem: Differentiating the speech of patients with and without Parkinson's disease
- Database: own construction with Hospital Rivadavia and Hospital Posadas
- Donation vowel /a/
- Previous analysis: Initial parameters using VAT (Voice Analysis Toolbox):
339 → 297 after cleaning.
- The challenge is to **select few parameters for PD prediction.**

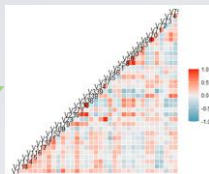


Methods

297 variables classified in 4 groups

- G1: F0 variations- Jitter
- G2: Shimmer
- G3: Noise due to incomplete vocal cord closure
- G4: Articulation problems- MFCC coefficients

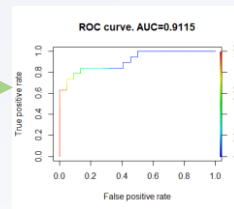
35 variables selected with PCA+ANOVA



Predictive power
Average(AUC)>0,8

LASSO y ELASTIC-NET in Logistic Regression (LR):
iterative method by recording the most selected variables (11 var)

Stepwise on LR with
11var+age+sex:
5 variables selected!



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

We obtained a model with high predictive capacity and only 5 significant variables. This allows the identification of speech disorders that correlate with the disease.

What's next?

- ▶ Increase our database.
- ▶ Speech Evaluation System for Acoustic Diagnosis (SEHDA): PD and NPD.

Challenge: Web system that allows the physician to record the patient's voice, upload it to the cloud and then analyze it and give a feedback.

- ▶ Find ranges for selected variables to enable case diagnosis.
- ▶ Explore other analysis methodologies.



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022