به نام خدا

Written by Mehdi Emari

Parkinson's Disease Detection Dataset

About Dataset

Parkinson's data set

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient, the name of the patient is identified in the first column. For further information or to pass on comments, please contact Max Little (littlem '@' robots.ox.ac.uk).

Further details are contained in the following reference -- if you use this dataset, please cite:
Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', IEEE Transactions on Biomedical Engineering (to appear).

Matrix column entries (attributes):

name - ASCII subject name and recording number
MDVP:Fo(Hz) - Average vocal fundamental frequency
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
MDVP:Flo(Hz) - Minimum vocal fundamental frequency
Five measures of variation in Frequency
MDVP:Jitter(%) - Percentage of cycle-to-cycle variability of the period duration
MDVP:Jitter(Abs) - Absolute value of cycle-to-cycle variability of the period duration
MDVP:RAP - Relative measure of the pitch disturbance
MDVP:PPQ - Pitch perturbation quotient
Jitter:DDP - Average absolute difference of differences between jitter cycles
Six measures of variation in amplitude
MDVP:Shimmer - Variations in the voice amplitdue
MDVP:Shimmer(dB) - Variations in the voice amplitdue in dB
Shimmer:APQ3 - Three point amplitude perturbation quotient measured against the average of the three amplitude
Shimmer:APQ5 - Five point amplitude perturbation quotient measured against the average of the three amplitude
MDVP:APQ - Amplitude perturbation quotient from MDVP
Shimmer:DDA - Average absolute difference between the amplitudes of consecutive periods
Two measures of ratio of noise to tonal components in the voice
NHR - Noise-to-harmonics Ratio and
HNR - Harmonics-to-noise Ratio
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
Two nonlinear dynamical complexity measures
RPDE - Recurrence period density entropy

D2 - correlation dimension
DFA - Signal fractal scaling exponent
Three nonlinear measures of fundamental frequency variation
spread1 - discrete probability distribution of occurrence of relative semitone variations
spread2 - Three nonlinear measures of fundamental frequency variation
PPE - Entropy of the discrete probability distribution of occurrence of relative semitone variations

About this file

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient, the name of the patient is identified in the first column.