CLASSIFICATION OF NEURO-DEGENERATIVE DISEASES USING SUPPORT VECTOR MACHINES (SVMs)

SYSC 5405 PATTERN CLASSIFICATION & EXPERIMENTAL DESIGN

INSTRUCTOR: ALISTAIR BOYLE

GROUP #5: MO CHEN (101062544), SHIVANGI RAGHAV (101107300)

DATE: NOVEMBER 27TH , 2018

OUTLINE

Pre-processing

- Noise reduction
- Data Normalization – "*Normalize*" Filter
- Class Imbalance – "*SMOTE*" Filter

Process Adjustments

FRAME WORK

Feature Extraction from time series

Feature selection

Parameter adjustments and Training 2 SVM classifiers for the two-class & four-class problems.

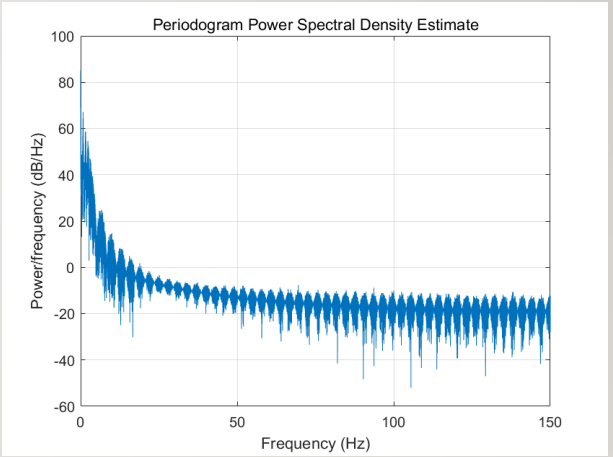
Testing and validation of the classifier.

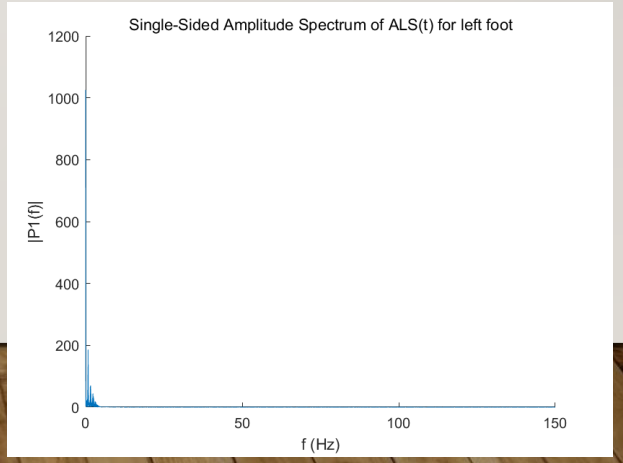
Adoption of Adaboost approach to increase the performance.

Evaluation

EXTRATED FEATURES USING MATLAB

- In time domain
  - Root mean square (RMS)
  - Standard deviation
  - Pearson's Correlation
  - Mean distance & value of local peaks/minimums
- In frequency domain
  - RMS value, Variance
  - Fundamental frequency
  - PSD estimation using FFT





FEATURE SELECTION AND CLASSIFICATION

- "***AttributeSelection***" in Weka Filters
  - Evaluator – "GainRatioAttributeEval"
  - Search method – "Ranker"
- Classifier
  - SVM classification - Sequential Minimal Optimization (SMO) with Poly Kernel
  - 10-fold Cross-Validation
- Meta Learning
  - Bagging, Boosting (AdaBoostM1)
- Synthetic Minority Over-sampling Technique (SMOTE)
  - To handle unrepresented class problem
  - However, it will change the number of sample data, increasing from 44 to 60, in which each of the four types are represented equally (15 instances per class).

PERFORMANCE MEASUREMENTS

- 2-CLASS PROBLEM : H/D

	Actual positive(H)	Actual negative(D)
Predicted positive (H)	9	2
Predicted negative (D)	2	31

- Accuracy: 90.90 %

- 4-CLASS PROBLEM : H/ALS/HD/PD

	Actual H	Actual ALS	Actual HD	Actual PD
Predicted H	10	0	1	0
Predicted ALS	0	4	4	0
Predicted HD	6	0	7	2
Predicted PD	1	1	4	4

- Accuracy: 56.82 %

PERFORMANCE MEASUREMENTS AFTER ADABOOSTING

- 2-CLASS PROBLEM : H/D

	Actual positive(H)	Actual negative(D)
Predicted positive (H)	9	2
Predicted negative (D)	2	31

- Accuracy: 90.90 %

- 4-CLASS PROBLEM : H/ALS/HD/PD

	Actual H	Actual ALS	Actual HD	Actual PD
Predicted H	9	0	2	0
Predicted ALS	0	4	2	2
Predicted HD	3	0	10	2
Predicted PD	0	1	5	4

- Accuracy: 61.36 %
(An increase of 5%)

OVERALL ANALYSIS :

	Accuracy	Standard Deviation
2 Class Problem		
SMO for H/D	0.856	0.155
SMO + Adaboost for H/D	0.861	0.147
SMO + SMOTE for H/D	0.844	0.126
SMO + SMOTE + Adaboost for H/D	0.846	0.127
4 Class Problem		
SMO for H/ALS/HD/PD	0.570	0.241
SMO + Adaboost for H/ALS/HD/PD	0.590	0.180
SMO + SMOTE for H/ALS/HD/PD	0.655	0.176
SMO + SMOTE + Adaboost for H/ALS/HD/PD	0.660	0.176

REFERENCES:

[1] MATLAB, Available Online: https://www.mathworks.com/products/matlab.html.

[2] Weka, Available Online: https://www.cs.waikato.ac.nz/ml/weka/

[3] https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/

[4] H. Zheng, et. al. "Machine Learning and Statistical Approaches to Supportthe Discrimination of Neuro-degenerative Diseases Based on Gait Analysis". Available online : https://link.springer.com/chapter/10.1007/978-3-642-00179-6_4

[5] Course Lecture Notes.

Thank you!