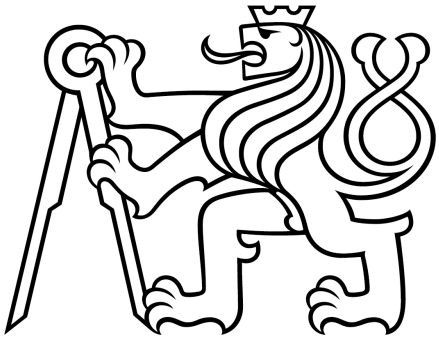

České vysoké učení technické v Praze

Experimental Data Analysis

Task I: Detection of early Parkinson's disease from speech

Valeriia Klimova, Tigran Oganesian, Rudolf Shymon
klimoval@fel.cvut.cz, oganetig@fel.cvut.cz, shymorud@fel.cvut.cz

January 2024

Contents

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder with motor and non-motor symptoms, including speech abnormalities. Traditionally assessed through perceptual tests, recent studies emphasize the significance of speech in early PD diagnosis [1]. Studies have demonstrated that a substantial percentage of untreated individuals in the early stages of PD exhibit some form of speech impairment, with variations in phonation, articulation, and prosody [2]. Speech alterations, linked to basal ganglia disruptions, impact motor planning and execution [3]. Technological advancements, such as machine learning, show promise in automatic PD detection based on vocal features, with notable accuracy (75.76%) [4]. Articulation features are particularly indicative [4]. Optimizing early diagnosis through machine learning aligns with the evolving landscape of PD research [5]. This paper explores speech symptoms in PD patients, their correlation with clinical parameters and demographics, addressing key questions in the context of improving PD understanding and management, addressing the following key questions:

- *Gender-Based Differences in Speech Features*

  Emerging evidence suggests that gender may influence speech characteristics in individuals with PD. This research aims to delineate any significant differences in speech features based on gender among individuals with PD.

- *Distinguishing Speech Features between Patients and Healthy Individuals*

  Numerous studies have suggested that individuals with PD exhibit distinct speech features compared to their healthy counterparts. This investigation seeks to identify and characterize the specific speech features that distinguish individuals with PD from healthy individuals.

- *Differences in Speech Features Based on Age at Diagnosis*

  The age at which PD is diagnosed may have implications for the manifestation of speech features. This research aims to delineate how speech features differ based on the age of onset of PD.

- *Effect of Disease Duration on Speech Features*

  The impact of disease duration on speech characteristics is an area that requires further exploration. This study seeks to investigate how the duration of PD influences speech features.

- *Correlation between MoCA and UPDRS Test Results and Relationship to Speech Features*

  This study aims to explore the correlation between Montreal Cognitive Assessment (MoCA) and Unified Parkinson's Disease Rating Scale (UPDRS) outcomes. Additionally, it seeks to examine how performance on clinical tests relates to speech features in Parkinson's disease.

- *Impact of Speech Features and Patient's Age on Disease Prediction*

  The potential diagnostic value of speech features and patient's age in identifying PD has been a subject of interest. This study aims to assess whether speech features and patient's age can serve as predictive factors for the presence of PD.

2 Methods

The following subsections provide a comprehensive overview of the dataset utilized in this study and outline the methodologies employed to address the specified research objectives.

2.1 Subjects and Measured Parameters

The dataset used in the study consists of the 100 patients with PD, of which there are 60 men and 40 women, 100 people control group of 60 men and 40 women. Demographic data included age and sex. The severity of PD was assessed using the UPDRS III survey, the MoCA assessment and the Purdue pegboard test. The loss of dopamine producing neurons was measured by DAT SPECT. See Table 2.1 for a comprehensive overview of the data set used in this study. The speech of all subjects was analysed during syllable repetition, sustained phonation, monologue and reading, using a total of 45 speech parameters.

	Subject											
	PD						HC					
	all		M		F		all		M		F	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	60.92	12.32	60.77	11.71	61.15	13.32	61.05	12.09	61.07	11.41	61.03	13.20
Symptom duration	1.87	1.52	1.85	1.49	1.89	1.60	n/a	n/a	n/a	n/a	n/a	n/a
MDS-UPDRS III	29.69	12.61	29.58	11.90	29.85	13.76	n/a	n/a	n/a	n/a	n/a	n/a
Bradykinesia	15.95	7.61	15.72	7.32	16.30	8.11	n/a	n/a	n/a	n/a	n/a	n/a
Rigidity	3.75	2.79	3.52	2.70	4.10	2.92	n/a	n/a	n/a	n/a	n/a	n/a
Tremor	6.28	3.85	6.52	3.85	5.93	3.87	n/a	n/a	n/a	n/a	n/a	n/a
PIGD	1.76	1.66	1.60	1.45	2.00	1.92	n/a	n/a	n/a	n/a	n/a	n/a
MoCA	25.08	2.88	24.88	3.25	25.38	2.24	n/a	n/a	n/a	n/a	n/a	n/a
DAT-SPECT:												
Caudate binding ratio	2.97	0.56	2.87	0.56	3.12	0.54	n/a	n/a	n/a	n/a	n/a	n/a
Putamen binding ratio	1.52	0.37	1.50	0.38	1.57	0.36	n/a	n/a	n/a	n/a	n/a	n/a
Sustained phonation:												
VOT	22.34	5.26	23.02	4.66	21.32	5.97	21.16	4.98	23.06	4.85	18.31	3.68
DDKR	6.41	0.94	6.58	0.96	6.15	0.86	6.53	0.76	6.58	0.85	6.44	0.62
DDKI	25.18	16.62	26.04	19.19	23.89	11.88	19.60	9.58	21.68	10.51	16.49	7.02
VD	51.44	12.92	51.07	14.92	51.98	9.31	49.51	10.18	49.91	11.83	48.91	7.14
stdPWR	2.15	0.67	2.29	0.76	1.94	0.44	2.17	0.57	2.34	0.57	1.93	0.46
Monologue:												
MPT	18.07	5.59	18.62	5.95	17.26	4.97	17.14	5.85	18.33	6.38	15.35	4.47
stdF$_0$	0.35	0.23	0.37	0.23	0.32	0.23	0.30	0.19	0.30	0.13	0.31	0.25
jitter	0.52	0.30	0.58	0.37	0.44	0.10	0.49	0.25	0.54	0.31	0.43	0.08
shimmer	2.49	1.08	2.90	1.13	1.88	0.62	2.63	1.08	3.11	1.04	1.93	0.67
HNR	19.37	3.31	17.88	2.74	21.61	2.80	19.61	3.22	18.13	2.76	21.83	2.50
PSI	7.44	16.00	6.06	13.99	9.52	18.62	4.40	9.25	4.21	8.09	4.67	10.87
CPP	23.57	3.56	22.63	3.45	24.98	3.28	23.60	2.93	22.32	2.61	25.52	2.28
FF0T	2.73	1.18	2.57	1.28	2.99	0.96	2.63	1.13	2.37	0.80	3.01	1.42
PF0T	0.17	0.11	0.19	0.11	0.15	0.10	0.15	0.08	0.17	0.08	0.13	0.07
FAT	2.45	0.94	2.31	0.95	2.67	0.89	2.15	0.77	2.11	0.80	2.22	0.72
PAT	6.90	3.31	7.38	3.67	6.17	2.58	6.83	2.57	7.14	2.86	6.37	2.00
Syllable repetition:												
EST	1.54	0.01	1.54	0.01	1.55	0.01	1.55	0.01	1.55	0.01	1.55	0.01
RST	345.28	78.62	329.15	85.22	369.47	60.86	371.46	74.73	346.70	68.44	408.60	68.79
DPI	215.06	80.92	228.58	94.78	194.78	48.27	189.46	75.45	209.57	87.65	159.29	35.68
DVI	261.46	71.14	276.47	85.13	238.95	31.87	258.38	66.08	275.13	75.40	233.26	37.47
GVI	39.49	12.60	34.91	12.06	46.37	10.10	44.18	12.11	38.78	11.42	52.29	7.94
DUS	28.79	13.12	32.91	14.75	22.60	6.52	24.31	9.62	27.14	11.29	20.07	3.36
RFA	8.56	1.27	8.93	1.20	8.01	1.18	9.13	1.37	9.37	1.41	8.77	1.24
RLR	-26.20	4.07	-26.69	3.94	-25.47	4.21	-24.06	3.61	-24.16	4.08	-23.91	2.80
PIR	5.11	1.96	4.62	1.65	5.85	2.17	5.77	2.32	4.96	1.83	6.99	2.47
RSR	16.24	4.08	16.56	4.17	15.75	3.94	16.02	4.34	16.87	4.41	14.75	3.96
LRE	218.10	117.61	236.42	130.09	190.62	90.68	165.41	75.55	172.88	83.19	154.21	61.67
stdPWR	3.48	0.82	3.61	0.94	3.29	0.53	3.91	0.91	4.06	1.05	3.69	0.60
stdF$_0$	1.51	0.46	1.55	0.54	1.47	0.32	1.92	0.65	1.86	0.71	1.99	0.55
Reading:												
EST	1.56	0.01	1.56	0.01	1.56	0.01	1.55	0.02	1.55	0.02	1.55	0.01
RST	429.65	62.81	415.96	63.37	450.19	56.72	438.69	62.09	425.00	63.38	459.23	54.65
AST	17.76	15.42	16.23	14.24	20.06	16.98	17.95	14.73	13.78	13.56	24.20	14.34
DPI	149.85	27.26	155.19	31.02	141.84	17.93	146.57	26.04	153.40	27.68	136.32	19.56
DVI	201.63	36.26	207.31	39.47	193.10	29.27	202.96	32.24	209.95	33.91	192.47	26.66
GVI	56.87	11.81	52.82	10.71	62.95	10.84	55.34	11.91	50.50	10.67	62.60	9.87
DUS	23.96	10.33	27.08	11.76	19.28	4.90	22.38	8.04	23.82	9.24	20.21	5.19
DUF	-0.24	2.29	-0.47	2.18	0.11	2.42	-1.08	2.66	-1.35	2.57	-0.67	2.77
RFA	9.70	1.36	9.96	1.22	9.31	1.48	10.87	1.53	10.97	1.64	10.72	1.36
RLR	-26.54	3.02	-26.80	3.37	-26.15	2.39	-26.30	3.69	-25.96	3.77	-26.81	3.55
PIR	6.94	2.16	6.94	2.44	6.94	1.69	7.00	2.42	7.00	2.61	7.00	2.13
RSR	17.37	7.82	17.85	9.41	16.66	4.51	16.00	4.17	15.54	4.36	16.69	3.82
LRE	138.85	62.58	145.34	69.69	129.13	49.31	147.91	64.49	157.35	71.76	133.75	49.24
stdPWR	3.35	0.63	3.47	0.64	3.15	0.58	3.86	0.67	4.03	0.67	3.59	0.59
stdF$_0$	1.66	0.50	1.59	0.51	1.76	0.48	2.45	0.72	2.35	0.79	2.61	0.60
NSR	6.11	0.82	6.12	0.90	6.09	0.69	6.29	0.83	6.37	0.91	6.18	0.70

Table 1: Overview of data used in the study of measured speech parameters and assessments of PD severity

The box plots in the Figure 1 visually indicate that both the Health Control (HC) and Parkinson's Disease (PD) groups have a uniform distribution of age across gender. This suggests that age-related factors affecting speech parameters will not have a major impact on the further study. The balanced age distribution will result in less biased results.

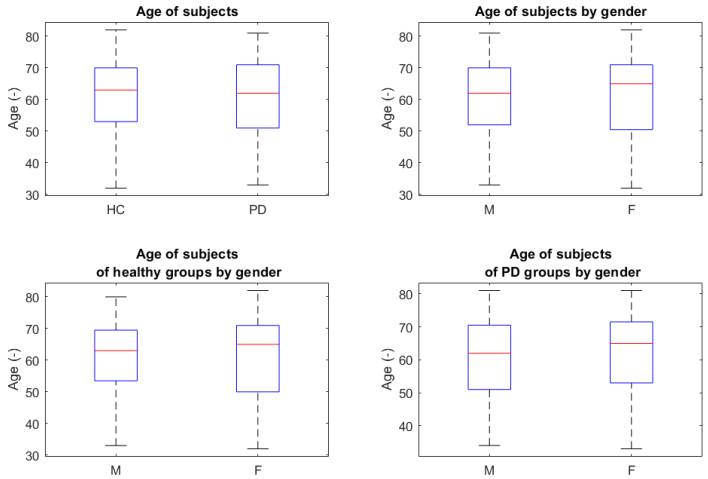

Figure 1: Distribution of subjects' ages across HC and PD group in respect to the gender

2.2 Approach

In statistical analysis, data preprocessing involves removing outliers and testing for normality using the Shapiro-Wilk test. Subsequently, independent t-tests and Mann-Whitney U tests assess the significance of speech features between HC and PD groups, considering normal and non-normal distributions.

Exploring gender differences, additional tests analyze the impact of gender on speech features within both HC and PD groups. Correlation analyses reveal insights into age at diagnosis, disease duration effects on speech features, and relationships between speech characteristics and MoCA/UPDRS scores.

For a deeper understanding, machine learning models (logistic regression, SVM) are developed, integrating speech features and patient demographics to distinguish between HC and PD groups. Results are evaluated using metrics like accuracy, precision, recall, and F1.

Stringent statistical considerations are maintained throughout, adjusting for multiple comparisons. Findings cover group comparisons, gender differences, correlations, and machine learning model evaluations, offering a comprehensive exploration of speech features in the context of PD.

3 Results

3.1 Feature Selection

In the process of feature selection, we meticulously curated the dataset to enhance the robustness and relevance of the speech features under consideration, with a primary focus on achieving meaningful and statistically significant results. The following steps outline the outcomes obtained:

- *Outlier Removal and Bonferroni Adjustment*

  Outliers were diligently identified and removed. The chosen statistical significance level was adjusted using Bonferroni approach due to large amount of tested data.

- *Shapiro-Wilk Normality Test and Data Stratification*

  Out of the original 45 normal speech features, 18 were identified as adhering to a normal distribution.

- *Independent t-Test for Normal Data*

  Post the independent t-test analysis, 4 out of the initial 18 normal speech features emerged as statistically significant contributors: in Reading `RFA` and `stdPWR`, in Monologue `RLR` and `RFA` ($p < 0.001$).

- *Mann-Whitney U Test for Non-Normal Data*

  Following the Mann-Whitney U test, 6 out of the initial 27 non-normal speech features were identified as statistically significant: in Syllable repetition `DDKI` in Monologue `DPI`, `LRE`, `stdPWR`, `stdF0`, and in Reading `stdF0` ($p < 0.001$).

3.2 Gender-Based Differences

Independent t-tests for normally distributed data and Mann-Whitney U tests for non-normally distributed data were used to analyse gender differences in speech characteristics. Tests were performed on those parameters that were found to be significantly different between HC and PD subjects. Tests for difference between HC and PD subjects were repeated separately for each gender.

- *Gender dependent speech parameters*

  On the basis of an independent t-test it was found that 2 out of 4 parametric parameters showed significant differences according to gender `RFA` in monologue and `stdPWR` in reading. The differences between the speech parameters were exaggerated by PD. In particular, the `RFA` parameter in Monologue shifted from $p = 0.03$ to $p = 0.00026$.

  After performing the Mann-Whitney U test for non-parametric data, the `Gender` difference was most evident in the `DPI` parameter in Monologue.

- *Difference in affect produced on speech parameters between genders*

  On the basis of an independent t-test it was found that `RLR` in Monologue has a higher difference between PD and HC groups for male subjects ($p_{\mathrm{M}} < 0.001$, $p_{\mathrm{F}} = 0.055$). While `stdPWR` in reading and `RFA` in reading maintain the difference regardless of gender.

  After performing the Mann-Whitney U test, it was taken into account that for male subjects the greatest difference in non-parametric parameters can be observed on `stdF0` in Reading ($\mathrm{p} < 0.0001$), while at the same time other parameters lost their power. Females maintained the parameters previously determined.

3.3 Correlation Analysis

The correlation analysis between disease-related variables and speech features yielded the following results:

- *Disease Duration vs. Speech Features*

  One speech feature (`stdPWR` from reading test) exhibited a weak correlation value falling within the interval of 0.20-0.39. Additionally, nine speech features demonstrated a very weak correlation ranging from 0.00 to 0.19.

- *Age of Diagnosis vs. Speech Features*

  Two speech features (`stdPWR` from reading test and monologue `DPI`) displayed weak correlation values in the range of 0.20-0.39. Furthermore, eight speech features showed very weak correlations in the interval of 0.00-0.19.

- *UPDRS vs. Speech Features*

  Four speech features (`stdPWR`, `stdF0` and `RLR` from monologue test, `stdF0` from reading test) exhibited weak correlation values falling within the 0.20-0.39 range. Additionally, six speech features demonstrated very weak correlations in the interval of 0.00-0.19.

- *MoCA vs. Speech Features*

  One speech feature (`RLR:monologue`) displayed a weak correlation value in the range of 0.20-0.39. Moreover, nine speech features showed very weak correlations in the interval of 0.00-0.19.

- *UPDRS vs. MoCA*

  The correlation between general score of `UPDRS` and `MoCA` revealed a weak association, with the correlation coefficient falling within the interval of 0.20-0.39.

3.4 Parkinson's disease prediction

3.4.1 Logistic Regression

A logistic regression analysis was conducted to investigate the relationship between the presence of PD and a set of predictor variables, including `Age`, Sustained repetition parameter (`DDKI`), Reading parameters (`RFA`, `stdPWR`, `stdF0`) and Monologue parameters (`RLR`, `RFA`, `DPI`, `LRE`, `stdPWR`, `stdF0`).

The cross-validated results indicated a mean accuracy of M = 0.77 across the 5 folds. Precision, recall, and F1-score were also computed, resulting in 0.77551, 0.76, and 0.76768, respectively. The confusion matrix summarizing the model's classification performance is presented in the Figure 2 on the left.

3.4.2 Support Vector Machine

A Support Vector Machine (SVM) analysis with a linear kernel was employed to investigate the predictive utility of `Age`, Sustained repetition parameter (`DDKI`), Reading parameters (`RFA`, `stdPWR`, `stdF0`) and Monologue parameters (`RLR`, `RFA`, `DPI`, `LRE`, `stdPWR`, `stdF0`) in determining the presence or absence of PD.

The cross-validated results indicated a mean accuracy of M = 0.765 across the 5 folds. Precision, recall, and F1-score were also computed, resulting in 0.77895, 0.74, and 0.75897, respectively. The confusion matrix summarizing the model's classification performance is presented in the Figure 2 on the right.

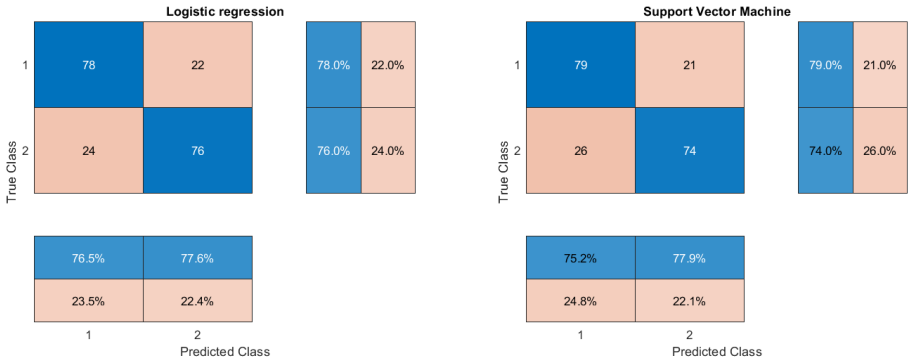

Figure 2: Logistic regression confusion matrix (on the left), Support Vector Machine with a linear kernel confusion matrix (on the right)

4 Discussion

4.1 Gender-Based Differences

Gender differences in the speech features used to detect PD were exaggerated in PD. At the same time, the parameters with the greatest differences between the HC and PD groups were found to be gender dependent. This suggests that models can be improved by making adjustments based on the gender of the subjects.

4.1.1 Limitations and Weaknesses

Sociocultural factors can introduce variations that affect speech characteristics differently in different groups. Lifestyle factors such as occupation and daily habits may also play a role. Emotional states, which are not explicitly addressed, may affect speech characteristics differently in men and women.

4.2 Correlation Analysis

The analysis revealed a scarcity of strong or highly significant correlations between disease-related variables and speech features in Parkinson's disease within our dataset. For instance, when examining the relationship between disease duration and speech features, only one parameter, stdPWR from the reading test, demonstrated a moderate correlation. However, most of the other features showed weak or very weak correlations.

4.3 Parkinson's disease prediction

The use of logistic regression and SVM allowed us to build models that demonstrated promising predictive capabilities, achieving mean accuracies of 0.77 and 0.765, respectively, across five folds.

The results align with our hypothesis that specific speech features, when combined with age, can serve as robust predictors of Parkinson's disease.

4.3.1 Limitations and Weaknesses

Despite promising results, our study has limitations. The dataset used may not represent the entire population of Parkinson's disease patients, and the generalizability of the models needs to be tested on different groups. The representativeness of our models may be also affected by the specific demographic characteristics of our dataset. Additionally, the predictive utility of our models should be validated against other machine learning approaches.

5 Conclusion

Our study identified specific parameters contribute significantly to gender differences. This emphasizes the need for a better understanding of gender variations in speech characteristics in the context of Parkinson's disease.

Although some correlations were identified, the general lack of strong associations emphasizes the complex nature of these relationships.

Based on the correlation results, it can be concluded that speech status is not correlated with age of diagnosis and duration of disease, making it difficult to predict the presence of disease by analysing speech factors at an early stage. Nevertheless, promising accuracy, precision, recall and F1-coefficients of the models shows that it is generally possible to assume a diagnosis based on speech. These models can be improved by also making an assumption based on the gender of the patient.

References

[1] J. Hlavnička, R. Cmejla, T. Tykalová, K. Sonka, E. Růžička, and J. Rusz, "Automated analysis of connected speech reveals early biomarkers of parkinson's disease in patients with rapid eye movement sleep behaviour disorder," *Scientific Reports*, vol. 7, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28144037/

[2] J. Rusz, R. Cmejla, H. Ruzickova, and E. Růžička, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease," *The Journal of the Acoustical Society of America*, vol. 129, 2011. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21303016/

[3] J. Rusz, J. Hlavnička, R. Čmejla, and E. Růžička, "Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction," *Frontiers in Bioengineering and Biotechnology*, vol. 3, 2015. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2015.00104/full

[4] Q. Wang, Y. Fu, B. Shao, L. Chang, K. Ren, Z. Chen, and Y. Ling, "Early detection of parkinson's disease from multiple signal speech: Based on mandarin language dataset," *Frontiers in Aging Neuroscience*, vol. 14, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2022.1036588

[5] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Automatic and early detection of parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method," *Diagnostics (Basel, Switzerland)*, vol. 13, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10253064/