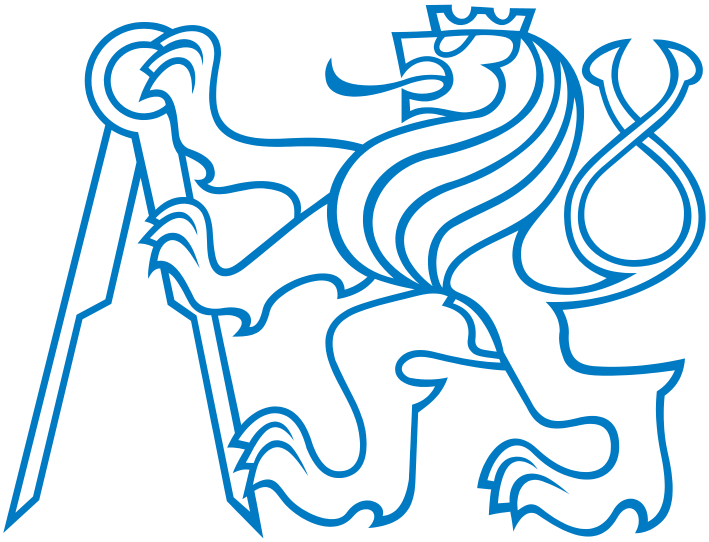
CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering



BACHELOR THESIS

Application of one-class classifiers in differential diagnosis of dysarthria

2018 Tran Duc Minh

ACKNOWLEDGEMENT

I would like to thank my supervisor Ing. Jan Hlavnička for patience, valuable remarks and methodical work management.

DECLARATION

I declare that I have produced the presented work independently and that I have listed all information sources in accordance with Methodical guidelines on maintaining ethical principles during the preparation of higher education theses.

In Prague, 29. December 2018 .....

Czech Technical University in Prague
Faculty of Electrical Engineering

Department of Cybernetics

BACHELOR THESIS ASSIGNMENT

Student: Tran Duc Minh

Study programme: Cybernetics and Robotics

Specialization: System and controls

Title of Bachelor Thesis: Application of one-class classifiers in differential diagnosis of dysarthria

Guidelines:

1. Get familiar with the topic of pattern recognition and dysarthria.
2. Propose the criteria of feature selection for differential diagnosis of dysarthria.
3. Apply the criteria to the real data provided by the supervisor.
4. Evaluate at least three appropriate one-class classifiers learned on selected speech features of healthy controls.
5. Compare the reached results with at least three appropriate multiclass classifiers learned on selected speech features of healthy controls and speakers with dysarthria.
6. Discuss the application of one-class classifiers in clinical practice.

Bibliography/Sources:

[1]. Darley, F. L., Aronson, A. E., and Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. Journal of Speech, Language, and Hearing Research, 12, 246-269.
[2]. Duffy, J. R. (2013). Motor speech disorders: Substrates, Differential Diagnosis and Management, 3rd ed., Mosby, St. Luis.
[3]. Duda, R. O., Hart, P. E., and Stork, D. G. (2001). Pattern Classification, 2nd ed., New York.
[4]. Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., and Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. Scientific Reports, 7: 12.
[5]. Rusz, J., Bonnet, C., Klempíř, J., Tykalová, T., Baborová, E., Novotný, M. Rulseh, A., and Růžička, E. (2015). Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Journal of Neurology, 262, 992-1001.
[6]. Tax, D. (2001). One class classification: Concept-learning in the absence of counter-examples, doctoral thesis, Delft University of Technology.

Bachelor Project Supervisor: Ing. Jan Hlavnička

Valid until: the end of the winter semester of academic year 2018/2019

L.S

prof. Ing. Michal Šebek, DrSc.
Head of Department

prof. Ing. Pavel Ripka, CSc.
Dean

Prague, October 10, 2017

ABSTRACT

Speech analysis of neurodegenerative diseases such as Parkinson's disease (PD) and Huntington's disease (HD) yields tremendous potential for high-throughput screening in the population under the risk of developing neurodegenerative disorders and remote monitoring of progression and treatment efficacy. Big databases of speakers affected by neurodegeneration are necessary for development of predictive models. Unfortunately, difficulties in recruitment of new patients and limitations of the language make the big databases unavailable. This Bachelor's thesis examines the idea that a recognition model can be trained only on speakers with no history of communication or neurological disorder hereby healthy controls (HC) that can be recruited easily. The thesis proposes the criteria of feature selection for differentiation of dysarthria, reviews various one-class classification methods, compares the performance of one-class and multiclass classifiers on this task, and discusses the suitability of one-class classification in the clinical context.

The database used in this thesis consisted of 48 subjects with PD, 43 subjects with HD, and 65 HC subjects. None of the subjects suffered from any additional disease that could negatively influence the speech performance. Each participant performed rhythm task, sustained phonation of vowels /A/ and /I/, monologue, reading passage and diadochokinetic task, of which acoustic signals were recorded using a standardized procedure and then processed by fully automated methods.

The features were selected using series of tests that involved correlation, Bartlett's test of homogeneity of variances, followed with Anova1 test and Kruskal-Wallis test. Selected features of healthy controls were evaluated with one-class classifiers trained on HC and compared with multi-class classifiers trained on both HC and patients with dysarthria. The result of one-class classifiers reached up to 84 percent of accuracy, which was almost comparable outcome with multi-class classifiers for category containing both PD and HD. Unfortunately, one-class classifiers compared to multi-class classifiers performed inconsistently for individual categories of detection PD and HD. Nevertheless, our results suggest that the idea of using the one-class models have potential utilization in clinical practice.

Key words:

Speech pattern recognition, One-class classifier, Multi-class classifier, Parkinson's disease, Huntington's disease, Hypokinetic dysarthria, Hyperkinetic dysarthria.

ABSTRACT

Řečová analýza neurodegenerativních onemocnění jakou je Parkinsonova nemoc (PD) a Hungtingtonova nemoc (HD) přináší obrovský potenciál pro automatizovaný systémy hodnocení u populace pod rizikem vzniku neurodegenerativních onemocnění a následné vzdálené pozorování progrese a účinnosti léčby. Pro vývoj prediktivních modelů je třeba velká databáze lidí postižených neurodegenerací. Bohužel komplikace v získávání nových pacientů a omezení jazyka činí tuto možnost získávání nedostupnou. Tato bakalářská práce zkoumá myšlenku využití rozpoznávání modelu pouze na pacienty bez záznamu komunikačních a neurologických onemocnění, tedy zdravá skupina lidí (HC), kterou lze snadno získat. Práce navrhuje kritéria výběru přízaků vhodných pro diferenciální diagnostiku disartrie, zkoumá různé jednotřídní klasifikační metody, porovnává výkonnost jednotřídních a vícetřídních klasifikátorů, a nakonec diskutuje možnost využití klasifikátorů v klinické praxi.

Databáze v této práci se skládala z 48 pacientů PD, 43 pacientů HD a 65 HC. Nikdo z pacientů netrpěl nemocí, která by mohla negativně ovlivnit průběh jejich testů. Každý z pacientů provedl rytmický test, úlohu prodloužené fonace hlásky /A/ a /I/, čtení textu, monolog a diadochokinetický test, z nichž byly zaznamenány akustické signály pomocí standardizovaného postupu a poté zcela zpracovány automatizovanými metodami.

Příznaky byly vybrány pomocí série testů, které zahrnovaly korelaci, Bartlettův test homogenity odchylek, následovaný Anova1 testem a Kruskal-Wallisovým testem. Vybrané příznaky zdravých pacientů byly hodnoceny jednotřídními klasifikátory trénované na skupinu HC. Dosažené výsledky byly následně porovnávány s výsledky klasifikátorů vícetřídních, které byly naučeny jak na zdravý pacienty, tak na pacienty s dysartrií. Výsledky jednotřídních klasifikátorů dosahovaly až 84 procent přesnosti, což bylo téměř srovnatelné s výsledky vícetřídních klasifikátorů v kategorii obsahující PD i HD. Bohužel jednotřídní klasifikátory ve srovnání s vícetřídními klasifikátory nedosahovaly konzistentních výsledků pro individuální kategorie PD a HD. Nicméně naše výsledky ukázaly, že myšlenka využití modelu rozpoznávání založeného pouze na zdravých pacientech má potenciální využití v klinické praxi.

Keywords:

Rozpoznávání řečových vzorů, Jednotřídní klasifikátor, Vícetřídní klasifikátor, Parkinsonova nemoc, Huntingtonova nemoc, Hypokinetická dysartrie, Hyperkinetická dysartrie.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

NOMENCLATURE

Abbreviation	Meaning
AST	Acceleration of speech timing
CAG	Guanine nucleotide
CV	Constant-vowel
DDK task	Diadochokinetic task
DDKG	Diadochokinetic regularity
DDKR	Diadochokinetic rate
DDKW	Diadochokinetic vowel duration
DPB	Degree of pitch breaks
DPI	Duration of pause intervals
DUF	Decay of unvoiced fricatives
DUS	Duration of unvoiced stops
DVA	Degree of vocal arrest
DVI	Duration of voiced intervals
EFn_M	Degree of hypernasality
EFn_SD	Intermittend hypernasality
EM	Expectation-maximization algorithm
EST	Entropy of the speech timing
F0	Fundamental frequency
FN	False negative
FP	False positive
GVI	Gaping in-between voiced speech
HC	Healthy controls
HD	Huntington's disease
HNR	Harmonics-to-noise ratios
KNN	K-nearest neighbour method for multi-class classifier
KNND	K-nearest neighbour method for one-class classifier
L-DOPA	L-3,4-dihydroxyphenylalanine
LDA	Linear discriminant analysis
LRE	Latency of respiratory exchange
MAP	Maximum a posteriori probability
meanACC	Mean accuracy
ML	Maximum-likelihood estimation
MOG	Mixture of Gaussians
MPT	Maximum phonation time
NAIVE	Naive Bayes classifier
NN	Nearest neighbour algorithm
NSR	Net speech rate
NVB	Number of voice breaks
PARZEN	Parzen density classifer
PCA	Principal component analysis
PD	Parkinson's disease
PDE	Parzen density estimation
PDF	Probability density function
PIR	Pause intervals per respiration
QDA	Quadratic discriminant analysis
RA	Rhythm acceleration
RBF	Radial basis function
RFA	Resonant frequency attenuation
RI	Rhythm instability
RLR	Loudness of respiration
RSR	Rate of speech respiration
RST	Rate of speech timing
SD	Standard deviation

Abbrevation	Meaning
SOM	Self-organizing map
stdF0	Standard deviation of fundamental frequency
stdPWR	Standard deviation of power
SVDD	Support vector data description
SVM	Support vector machines
TN	True negative
TP	True positive
TREE	Tree decision classifier
UHDRS	Unified Huntington's Disease Rating Scale
UPDRS	Unified Parkinson's Disease Rating Scale

Symbols	Meaning
$\boldsymbol{C}$	Regularization parameter of SVM/SVDD which regulates the tradeoff between complexity and errors or the number of classes in a multi-class classification problem
$\boldsymbol{d}$	Dimensionality of input patterns
$\boldsymbol{D_M(\vec{x})}$	Mahalanobis distance of observation
$\boldsymbol{d_{SOM}}$	Euclidean distance in SOM
$\boldsymbol{h}$	Width of kernel
$\boldsymbol{k}$	Number of prototypes per manifold dimensionality in SOM
$\boldsymbol{L}$	Lagrangian, the combination of an error function and constraints
$\boldsymbol{M}$	The number of basis vectors
$\boldsymbol{n}$	Number of free samples
$\boldsymbol{n_{free_p}}$	Number of free parameters
$\boldsymbol{N_{MoG}}$	Number of Gaussians
$\boldsymbol{n_{PCA}}$	Number of free parameters in the PCA
$\boldsymbol{n_{paramMoG}}$	Number of parameters for the MOG
$\boldsymbol{NN^{tr}(z)}$	(First) nearest neighbour of object x
$\boldsymbol{NN_k^{tr}(x)}$	k-th nearest neighbour of object x
$\boldsymbol{p_{\mathcal{N}}(x;\mu_j,\sum_j)}$	Normal or Gaussian distribution, characterized by mean μ and covariance matrix $\sum$
$\boldsymbol{p_{NN}(z)}$	Local density of KNND
$\boldsymbol{p_{MoG}}$	Probability density of Mixture of Gaussians
$\boldsymbol{p_p}$	Parametric density model
$\boldsymbol{p_\varphi}$	Non-parametric density model
$\boldsymbol{S}$	Covariance matrix
$\boldsymbol{\vec{x}}$	Set of observation
$\boldsymbol{\eta(\tau)}$	Learning rate in SOM ranging between 0 and 1
$\boldsymbol{\mu}$	Mean
$\boldsymbol{\sigma}$	Standard deviation

1 Introduction

1.1 MOTIVATION

Parkinson's disease and Huntington's disease patients are the most common after Alzheimer and it is estimated that currently 4 to 6 million people suffer from Parkinson's disease, worldwide. Statistics for the number of affected people with neurodegenerative disorders show an increase in proportion with the overall ageing of the worldwide population as a whole. Thus, considerable attention has been given to progressive neurodegenerative diseases affecting the basal ganglia such as Parkinson's disease and Huntington's disease. Medical treatment mitigates certain symptoms of these diseases but there is no cure available. Still, early diagnosis of the diseases has a very important role in improving patient's live, e.g. appropriate diagnosis and treatment can slow the progress of the disease. Behavioral speech therapy, which includes intensive voice treatment, shows to be one of most effective and objective type of speech intervention at present. For that purposes, many acoustical voice analyses and measurement methods have been tested and reported in previous studies (Rusz et al. **2011**, Postuma et al. **2012**, Harel et al. **2004**). Development of precise acoustic models requires a large database of patients with neurodegenerative diseases because values of speech features can be influenced by several factors such as age and gender. However, collecting a large balanced database of speakers with no history of communication or neurodegenerative disorder hereby healthy controls (HC) can be an easy task. This thesis revolves around the idea to train the model on the HC speakers using one-class classifiers. While a significant amount of research has been devoted to examining the impact of within-class imbalance over multi classifiers, very little attention has been given to their impact on one-class classifiers, which are typically used in cases of extreme between-class imbalance. This work focuses on the application of one-class classifiers in speech pattern recognition, comparison of various types of one-class classifiers with regard to performance; and discussion of the applicability of one-class classifiers in the clinical practice.

1.2 PARKINSON'S DISEASE

Idiopathic Parkinson's disease (PD) is a progressive neurodegenerative disorder. PD was described two centuries ago; still the causes of disease remain unknown. Progressive loss of dopaminergic neurons in the substantia nigra pars compacta results in dopamine deficiency within the basal ganglia (Hornykiewicz **2008**). An imbalance between acetylcholine and dopamine in basal ganglia leads to characteristic parkinsonian motor symptoms represented by rest tremor, rigidity, postural instability and bradykinesia (Jankovic **2008**) (Rodriguez-Oroz **2009**).

PD affects primarily age group older than 50 years. Previous studies show that in average PD effects the most category of people of age 60 (Inzelberg et al. **2002**), however it is predicted that 5-10% of patients get PD in the young age of 20-40 years (Golbe **1991**) depending primarily on genetic dispositions (Inzelberg et al. **2002**). PD is the second most common neurodegenerative disease (after Alzheimer's disease) with prevalence 0,3% in population (Rajput and Birdi **1997**). It is estimated that between 4.1 and 4.6 million patients were diagnosed in the year 2005 and in the year 2030, the number will rapidly increase to

approximately 9 million (Dorsey et al. **2007**). The patient's life expectancy is individually, and it is within the range of 6 to 12 years (Rajput **1992**, Willis **2012**). Despite the fact that medication generally prolongs active life expectancy, the effect of treatment depends upon the stage of the disease during which it is initiated. Furthermore, there is no effective treatment that can cure PD or halt its progression. Patient's therapy is based on relieving the symptoms (Becker **2002**). The main goal is to balance neurotransmitter imbalances by increasing the level of dopamine by its metabolically acceptable form of L-DOPA. Another option is a surgical procedure, primarily used to remove tremors. As it was already mentioned, there are so far no treatments for PD, therefore the early diagnosis of PD plays a vital role in improving the patient's quality of life (National Parkinson Foundation **2013**).

The diagnosis of PD is based upon the presence of primary motor symptoms, which develop after 60-70% of dopaminergic neurons degenerate and dopamine levels are reduced by 80% (Fearnley and Lees **1991**). In addition to the most common motor manifestations, PD is also associated with non-motor symptoms such as autonomic dysfunction, cognitive and neurobehavioral abnormalities, speech impairment, hallucination, and sensory disruptions, and sleep alterations. Non-motor symptoms precede the parkinsonian motor symptoms by more than a decade (Kalia and Lang **2015**).

Both these diseases share features that include delayed onset, abnormal protein processing and aggregation, cellular toxic effects involving both cell autonomous and cell-cell interaction mechanism and selective neuronal vulnerability, despite widespread expression of disease-related proteins during the whole lifetime (Ross et al. **2011**).

1.2.1 Speech impairment

Several studies have found the impaired speech to be one of the earliest manifestations of PD (Postuma et al. **2012**, Harel et al. **2004**). The process of speech impairment can be categorized into interdependent stages, which includes: Respiration problem, phonation problem, articulation problem and prosody.

- *Respiratory problem* - reduction of the ability to control breath for articulation speech (Critchley **1981**, Rusz et al. **2011**). Infliction by the rigidity of the respiratory muscles (Goberman and Coelho **2002**).
- *Phonation problem* - trouble learning the sound system of the language, failing to recognize which sound contrasts, problems in the vocal folds of the larynx, the voice involuntarily sounds breathy, raspy, or strained, or is softer in volume or lower in pitch (Hunker **1982**), impairing the ability to open vocal cords (Weismer **1983**).
- *Articulation problem* - difficulty learning to physically produce the intended phonemes. Infliction of bradykinesia and rigidity of larynx and pharynx (Critchley **1981**, Rusz et al. **2011**)
- *Prosody* - refers to intonation, stress pattern, loudness variations, pausing, and rhythm (Apraxia Kids **2018**). It also includes respiration, phonation, articulation problems and also neurological principal of speech (Hammen **1996**, Goberman and McMillan **2005**). Prosody is expressed mainly by varying pitch, loudness, and duration.

Ho et al. (**1998**) reported that up to 90% of PD patients suffer from vocal impairment, with the most significant impact on phonatory and articulatory features of speech. These vocal deficits can be generally described as hypokinetic dysarthria (Rosen et al. **2006**), (Kent

et al. **2000**). Hypokinetic dysarthria is characterized by reduced vocal loudness, rough and breathy vocal quality, harsh or breathy vocal quality and abnormal speaking rates (slow speaking rates but rushes of fast speech), monopitch, monoloudness and reduced stress (Rudzicz **2011**, Canter **1965**). As already mentioned, PD is associated with hypokinetic dysarthria due to akinesia and bradykinetic-rigid syndromes. The distinctive speech patterns connected with hypokinetic and hyperkinetic dysarthria are generally antagonistic, even though both PD and HD are primarily disorders of the basal ganglia. As an example, it is known that hypokinetic dysarthria in PD typically shows reduced vocal loudness and flattened loudness and pitch inflections, poor voice quality, variable and frequently increased speech rate, inappropriate silences and breathiness (Rusz et al. **2014**).

1.3 HUNTINGTON'S DISEASE

Huntington's disease (HD) is defined as a chronic, degenerative, progressive neuropsychiatric disorder, characterized by progressively increasing of chloreiform movements. HD is caused by an expansion of the number of guanine nucleotide (CAG) repeats located on the short arm of chromosome 4 at 4p16.3 (Kremer et al. **1994**, Hayden **1981**).

Huntington's disease can be regarded as a model neurodegenerative disorder. It is monogenic, fully penetrant, and similar to other neurodegenerative diseases, a disorder of protein misfolding. The gene for HD, huntingtin, was discovered 17 years ago. The condition is typified by progressive degeneration of the medium spiny neurons within basal ganglia, primarily the caudate and putamen (Albin et al. **1992**, Albin **1995**). As the disease progresses, neuronal loss occurs in the white matter, cerebral cortex and thalamus (Vonsattel et al. **1985**). The disease is inherited in an autosomal dominant manner. The penetrance of HD depends on the age of patients. Typically, the onset of symptoms is in middle age after affected individuals have had children, but the disorder can exhibit at any time between infancy and senescence. By using predictive genetic test, it is possible to identify individuals, who are at risk of inheriting the expanded cytosine, adenine and CAG repeats before their clinical onset. Increased CAG repeats predict the earlier onset, accounting for up to 50-70% of the variance in age of onset, with the remainder likely to be due to modifying genes and the environment (Wexler **2004**). On the contrary, expanded CAG repeats contribute less to the rate of progression and there is an opportunity for intervention by understanding the determinants of rate of progression (Rosenblatt et al. **2006**).

Prevalence of Huntington's disease is 4-10 per 100 000 in the western world, with many more people at risk of the disease. The Mean age of onset is 40 years, typically the death occurs 15-20 years from onset. Clinical features of Huntington's disease include progressive motor dysfunction, cognitive decline and psychiatric disturbance, caused by both neuronal dysfunction and neuronal cell death (Walker **2007**, Ross et al. **1997**). The precise pathophysiological mechanism of Huntington's disease is poorly understood. Formal diagnosis of HD is made on the basis of characteristic extrapyramidal motor signs of chorea, dystonia, bradykinesia or incoordination in an individual at risk (Huntington **1996**). Chorea is usually prominent in the early stage of the disease. In the late stage of HD bradykinesia, incoordination and rigidity (motor impairment) are more disabling functionally (Rosenblatt et al. **2003**). Many patients have substantial cognitive or behavioral disturbances before the onset of diagnostic motor signs (Marder et al. **2000**).

1.3.1 Speech impairment

The dysarthrias are differentiated according to perceptual characteristics of speech and verified by the underlying neuropathology. Patients with HD develop a motor speech disorder, which occurs primarily as a consequence of underlying choreatic movements (Hartelius et al. **2003**) (Saldert et al. **2010**). The characteristics of hyperkinetic dysarthria vary considerably across patients. Commonly observed speech patterns in patients with hyperkinetic dysarthria include unexpected variations in pitch or loudness, inappropriate silence, harsh voice, slow speech rate, inappropriate pauses vocal noises, constant or intermittent dysphonia, constant or intermittent hypernasality, intermittent breathy segments and articulatory imprecision (Darley et al. **1969,** Logemann et al. **1978,** Darley et al. **1975**). HD shows hyperkinetic dysarthria resulting from chorea (Duffy **2013**). For HD patients specifically, those symptoms have been described in several studies (Ramig **1986**, Garcia et al. **2011**), Hartelius et al. **2013**, Zwirner et al. **1991**) as harsh, breathy, and strained-strangled voice with occasional pitch fluctuations and vocal arrests. The most notable symptoms of speech deviations in HD include unpredictable breakdowns of articulation, phonatory dysfunction and abnormalities in speech timing and prosody (Duffy **2013**, Darley et al. **1969**). In addition, the gravity of dysarthria is connected to the overall severity of motor symptoms in both diseases (Garcia et al. **2011**, Hartelius et al. **2013**). What is also interested, is the fact that preliminary reports have suggested that speech deficits may precede the onset of the first motor symptoms (Vogel et al. **2012**, Kaploun et al. **2011**).

Considering the potential for early treatment and management strategies in HD and PD due to its genetic predictability (Tabrizi et al. **2011**), speech analysis is potentially an important method for monitoring disease onset and progression, as well as treatment efficacy. It provides subtle and quantitative information. Recent studies considered speech analysis as affordable, objective, and widely available (Postuma et al. **2012**, Harel et al. **2004**, Rusz et al. **2011**). Speech manifestations can be assessed by a wide range of speech tests, such as fast syllable repetition, sustained phonation, various readings and freely spoken monologue. Recorded utterances are subjected to methods as of assessment of sound pressure levels, fundamental frequency, formant frequencies, speech rate and rhythm (Baumgartner et al. **2001**, Chenausky et al. **2011**, Rusz et al. **2011**, Fischer and Goberman **2010**, Goberman and Blomgren **2008**).

1.4 CLASSIFICATION

The classification can be defined as the process of assigning an object represented by a vector of feature values (observation vectors) to a category of objects (class). The training data set is the set of observation vectors along with the corresponding class labels (HC, HD, PD). The test dataset of observations consists of the vector of feature values without class labels. Classifier learns to assign the class labels to unlabeled objects from the test set by using a set of objects from the training set with a purpose to create a new model. New observations are then classified based on the learned model. The classifier's applicability takes into account the type of the features the classifier deals with as well as characteristics of the classification method, such as robustness, computational and storage requirements, and the number of parameters to be estimated/set.

1.4.1 Multiclass classifiers

Naive Bayes classifier

The Naive Bayes (NAIVE) is classifier based on the principle of Maximum a posteriori probability (MAP). Given a problem with C classes $\{C_1, \ldots, C_K\}$ with prior probabilities $P(C_1), \ldots, P(C_K)$. We assign the class label c to the unknown example with features $x = (x_1, \ldots, x_N)$, such that:

$$c = argmax_c P(C = c \| (x_1, \ldots, x_N)\,, \qquad \text{Equation 1}$$

here the class with the maximum a posterior probability is chosen given the observed data. This a posterior probability is defined by the Bayes theorem of probability defined by the following equation:

$$P(C = c \| (x_1, \ldots, x_N) = \frac{p(C_k)p(x|C_k)}{p(x)}\,, \qquad \text{Equation 2}$$

to predict the class of unknown data set. It is a classification technique that assumes the independence among predictors. In other words, a Naive Bayes classifier assumes that the value of a particular feature in a class conditionally independent on the value of any other feature. For example, an observation of fruit can be considered to be a pear when its colour is green, a shape is oval, and diameter is within a certain range. If these features depend on each other or, all of these properties independently contribute to the probability that this fruit is a pear and that is why it is known as 'Naive'. Along with simplicity, Naive Bayes model is easy to build and particularly useful for very large data sets and was shown to perform well despite the simplifying assumption of conditional independence (Aly **2005**, Rish **2001**).

Nearest neighbour classifier

The nearest neighbour algorithm (NN) is among the oldest and simplest of all machine-learning algorithms. The NN represents a non-parametric machine learning since the model is defined by the store training instances. The principle of NN prediction is that the label of classified observation is predicted from the labels of observations nearest to the classified observation. The number of nearest observations is either constant defined by the user or varies with regard to the local density of points called radius-based neighbours learning. Generally, the distance can be any metric measure such as standard Euclidean distance, which is the most common choice. The class is defined by a majority vote of its neighbours, where the class is assigned to the most common among its *k* nearest neighbours (*k* is positive small integer parameter (Pedregosa et al. **2011**).

Quadratic Discriminant analysis classifier

Quadratic discriminant analysis (QDA) and linear discriminant analysis (LDA) are closely related. Both methods assume that the measurements from each class are normally distributed. While LDA can only learn linear boundaries, QDA can learn quadratic boundaries and threfore, it is a more flexible method. Boundaries of QDA are based on the Mahalanobis distance assuming different variance-covariance matrices for each class. Mahalanobis

distance of an observation $\vec{x} = (x_1, \ldots, x_N)^T$ from a set of observation with covariance matrix S and mean $\vec{\mu} = (\mu_1, \ldots, \mu_N)^T$ is defined as:

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}\,. \qquad \text{Equation 3}$$

The boundary produced by QDA is a quadratic, thus, classes can be separated even when their variances are very different. As the variances of the classes become more similar, QDA boundaries will tend toward each other. When classes possess identical variance structures, QDA boundary will be linear, as well as the variance-covariance matrix for each class will be equal to the pooled variance-covariance matrix.

For non-Bayesian form, assuming equal prior, class sizes are equal. For the Bayesian form, the likelihood of each class is determined from posterior distribution (Friedman et al. **2001**). The Gaussian parameters are assessed from training points with maximum likelihood (ML) estimation. The simple Gaussian model assumption is best for cases of lacking information to characterize a class, e.g. little amount of training samples to deduce class distribution. When the number of training samples is small compared to the number of dimensions of each training sample, the maximum-likelihood (ML) covariance estimation can be ill-posed (Dixon, Brereton **2009**, Srivastava et al. **2007**). QDA is recommended for the large training set, so the variance of the classifier is not main concern or if the assumption of the common covariance matrix is unsustainable.

Decision tree classifiers

The Decision trees are support tools that use branching methods to illustrate every possible outcome of a decision with the help of an algorithm that only contains conditional control statements. The classification of a pattern happens through a sequence of questions, where each next question asked depends on the answer to the current question. Non-metric data is best to work with using this approach as all the questions can be answered with a "yes/no" or "true/false". The predictive model-observation about an item can be viewed as the branches and the conclusions about the pattern or item's target value can be viewed as the leaves. The classification of a pattern or item's value begins at the first root node, which is displayed at the top and is connected by directional branches to other nodes. Here is where the first split of data happens. A split is each decision outcome at a node since it corresponds to splitting a subset of the data. The split at each node is based on a feature that gives the maximum information gain of each attribute, which is the expected reduction in entropy caused by the partitioning of the samples according to the attribute. The root node splits the full training data set and so does each and every successive decision. In general, the number of splits depends on the designer and may vary throughout the tree. The different branches correspond to the different possible values. Depending on the answer we choose the appropriate branch, which leads to a descendant node. A descendant node can be considered as the root of a sub-tree, where we again make a decision, which branch we follow next. The algorithm continues this way until it reaches one of the possible conclusions at the terminal nodes or leaf nodes, which have no further connections. A post-pruning process can be carried out to prevent overfitting (Duda et. al. **2012**, Aly **2005**, Li et al. **2004**).

Support vector machines classifier

Support vector machines (SVM) are supervised learning models used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories; an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. The examples are mapped into that same space and predicted to belong to a category according to the side of the gap they fall. The method was developed by Vapnik in early 1960 but was not implemented in current form until 1990 when the concept of nonlinear kernels was introduced.

For our models, we used SVM with nonlinear kernel function called Radial Basis Function (RBF), which is the most popular employed using SVM. SVM can be divided into hard and soft margin SVM. Hard margin presumes that two classes are separable and finds the optimal boundary that separates classes with the maximum possible margin between the two classes. It is possible to find a feature space in which two classes are separable using RBF and forcing the algorithm to search for this feature space may lead to overfitting of the model. To avoid overfitting, soft margin SVM tolerates a degree of misclassification. They are also designed to balance classification error with the complexity of the model. The penalty parameter C trades off misclassification of training data set against the simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims to classify all training data correctly by giving the model freedom to select more samples as support vectors. The parameter Gamma (positive number) determines the width of the RBF. Gamma defines how far the influence of a single training data set reached. The gamma parameter can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. Gamma and C have a big effect on the position and complexity of the SVM decision boundary. The optimal values for these parameters need to be determined from dataset itself. For the case, where gamma is large, the radius of the area of influence of the support vectors only contains the support vector itself. And therefore no amount of regularization with C will be able to prevent overfitting. When gamma is very small, the model is too constrained and cannot capture the complexity of the data. The most common method for optimisation of these parameters is the grid search carrying out repeated random sub-sampling validation. (Cortes, Vapnik **1995**, Dixon and Brereton **2009**, "Support vector machines," n.d.).

1.4.2 One-class classifiers

A one-class classifier is a special type of two-class classifier, where only data from one class are used for modelling the data such as in our case where training data are obtained only from HC group. One class classifiers are exceedingly beneficial in medical studies when the real data obtained from measurements of patients with diseases are difficult and costly or almost impossible to obtain. Visible attention has been given to the one-class classifiers in last years and several approaches to one-class classification have been presented and reviewed (Irigoien et al. **2014**). Possible applications of one class classifiers include mammograms for breast cancer detection (Tarassenko et al. **1995**, Costa and Moura **1995**), prediction of protein-protein interactions (Reyes and Gilbert **2007**), identification of patients with Nosocomial infections using clinical and other data collected during the survey (Cohen et al. **2008**), categorization of patients affected with interstitial lung diseases (Depeursinge et al. **2010**), heart murmur diagnosis (Cabral and Oliveira **2014**), image based tumor

recognition or analysis of electrocardiogram (Irigoien et al. **2014**), or the one-class recognition of cognitive brain functions (Boehm et al. **2011**). In the case of the medical alert system, which notifies staff about unusual measurements, e.g. laboratory results or data from sensors, the model of "normality" could be used for detection of outliers. Another example could be for machine condition monitoring, where faults should be detected (Tax **2001)**. To the present date, there have been few applications of the SVDD to clinical medicine, such as in radiology and MR imaging. Cognitive brain states in fMRI images (Song and Wyrwicz **2009**), reconstruction of brain tissue (Wang et al. **2007**) and the segmentation of brain tumors were analyzed (Zhou et al. **2007**).

There have been a handful of one-class classification models proposed. Most of them focus on outlier detection. One way for outlier detection is to generate outlier data around the target set. Then a classifier is trained to differentiate between outliers and the target data (Roberts et al. **1994**). In this thesis, few models for one-class classification are explored and compared. Three types of models were used: the density estimator, the boundary methods and the reconstruction methods. These methods differ in the definition of the function, the error and in the minimization method. This section includes 2 density methods (mixture of Gaussians and the Parzen density estimation); 2 boundary methods (k-nearest neighbour method and Support vector data description); and 2 reconstruction methods (Principal Component Analysis and Self-organizing map).

Density methods

The most straightforward method to get one-class classifier is to estimate and threshold the probability density (PDF) of the training data (Tarassenko et al., **1995**). Due to unknown information about the second class, the PDF for the second class is assumed uniform. A specific form of the distribution of feature values is often unknown but can be approximated by a mixture of multiple Gaussians or other kernels. The data in the training set are assumed to be representative of the true data distribution. The classification can be obtained by comparing the PDF of the current observation vector to a threshold (Mazhelis **2006**).

Mixture of Gaussians

This method assumes that data can be described as a mixture of several normal distributions. A mixture of Gaussians (MOG) represents a linear combination of normal distributions according to the following equation (Bishop **1995**):

$$p_{MoG}(x) = \frac{1}{N_{MoG}} \sum_{j} p_{\mathcal{N}}\left(x; \mu_j, \sum_j\right) P(i)\,, \qquad \text{Equation 4}$$

where mixing coefficient *P(i)* reflects the prior probability that an observation vector is generated from i-th component of the mixture. It has a smaller bias than the single Gaussian distribution, on the other hand, it requires far more data for training. The number of Gaussians N_{MoG} is defined beforehand by the user, the means μ_j and covariances $\sum_j$ of the individual Gaussian components can be efficiently estimated by an expectation minimization routine (Bishop **1995**). The number of parameters for the MOG is:

$$n_{paramMoG} = N_{MoG}\left(n_{paramMoG} + 1\right). \qquad \text{Equation 5}$$

The parameters of MOG are derived by employing the expectation-maximization algorithm (EM algorithm). The learning process using EM algorithm is more computationally demanding, as a number of interactions should be done before the algorithm converges. The classification process is relatively simple. It is a more flexible density method than Gaussian distributions due to its unimodality and convexity (Tax **2001**, Mazhelis **2006**).

Parzen density estimation

Parzen density estimation (PDE) is an extension of the previous method. The estimated density is a mixture of Gaussian kernels centred on the individual training objects. The parametric density model is defined (Parzen **1962**):

$$p_p(x) = \frac{1}{N}\sum_i p_{\mathcal{N}}(x; x_i, \Sigma_i) \,, \qquad \text{Equation 6}$$

where Σ_i is diagonal covariance matrices of Gaussian kernels.

$$\Sigma_i = hI \,. \qquad \text{Equation 7}$$

The equal width *h* in each feature direction means that the Parzen density estimator assumes equally weighted features and it will be sensitive to the scaling of the feature values of the data, especially for lower sample sizes (Tax **2001**).

Training a Parzen density consists of the determination of one single parameter, the optimal width of the kernel *h*. Thus, the number of free parameters in the model:

$$n_{free_p} = 1 \,. \qquad \text{Equation 8}$$

The *h* is optimized using the maximum likelihood solution (Kraaijveld and Duin **1991**). A good description depends on the representatives of the training set. The computational cost for training a Parzen density estimator is lower than other methods, but the testing is expensive. All training objects have to be stored and, during the learning phase, distances to all training objects have to be calculated and sorted. Considering large datasets in high dimensional feature spaces, this could severely limit the applicability of the method. PDE works very well when the sample size used for training is sufficiently high to overcome the curse of dimensionality (Duda et al. **2012**).

The non-parametric density model is defined as (Gramacki **2018**):

$$p_\varphi(x)) = \frac{1}{nh^d}\sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h}\right), \qquad \text{Equation 9}$$

where *h* is window width, *n* is a total number of samples and *d* is dimensionality of the problem. The method is advantageous due to its ability to approximate the arbitrary distribution, whose parametric form is unknown. Its disadvantageous is need of a large number of samples for accurate estimation (Mazhelis **2006**).

Boundary methods

Boundary methods, as the name implies, estimate the boundary of the trained data. The methods derive the distance between the classified observation vector and the boundary around the trained observation vectors. The distance calculation can take into account the distance between the analyzed observation vectors and the observation vectors in the training dataset as well as the distances between the observation vectors in the training dataset. Boundary methods are specifically targeted at one-class classifiers unlikely to density and reconstruction methods that are used often for multiclass classifiers (Mazhelis **2006**).

K-nearest neighbor method

K-nearest neighbor method (KNND) is derived from local density estimation by the nearest neighbor classifier (Duda et. al., **2012**). The probability density is calculated based on the number of observation vectors in a region of a certain volume. The method avoids the explicit density estimation and only uses distances to the first nearest neighbor. In the nearest neighbor density estimation, a cell (hypersphere in *d* dimensions) is centred around the test object z. The volume of this cell is grown until it captures *k* objects from the training set. The local density is then estimated by:

$$p_{NN}(z) = \frac{\frac{k}{N}}{V_k\left(\left\|z - NN_k^{tr}(z)\right\|\right)}, \qquad \text{Equation 10}$$

where $NN_k^{tr}(z)$ is the k nearest neighbor of z in the training set and $V_k$is the volume of the cell containing this object. In the KNND, a test object z is accepted when its local density is larger or equal to the local density of its (first) nearest neighbor in the training set $NN^{tr}(z)$= $NN_1^{tr}(z)$. All the distances to the k nearest neighbors are averaged, and furthermore, the distance of an object z to its k nearest neighbors is replaced by a more robust distance definition. When objects are very near the target data, the k-th nearest neighbor distance is used, instead of the first nearest neighbor distance. This robust measure makes it hard to distinguish between objects which are near the boundary or which are deep within a tight cluster of objects. This algorithm requires that the user defines in advance the number of neighbors k (Tax **2001,** Mazhelis **2006**).

Support vector data description

Support vector data description (SVDD) defines the hypersphere around the positive class data that encompasses almost all points in the data set with the minimum radius. The SVDD is a special case of support vector classifier (Vapnik **2013**). During the training phase (Tax and Duin **1999**)

$$L = \varepsilon_{SVDD} = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j), \qquad \text{Equation 11}$$

the parameter α_i are estimated by minimizing subject to the constraints $\sum_i \alpha_i = 1$ and $0 \leq \alpha_i \leq C$, where C determines the number of vectors that will not be covered by the description. The minimization is solved as a quadratic programming problem. The kernel

function transforms the vectors to a higher dimensional feature space where a more accurate description can be produced. In the training phase, the distance from the new observation vector to the centre of the hypersphere is calculated and compared against its radius. The SVDD classifier rejects a given test point as an outlier if it falls outside the hypersphere. However, SVDD can reject some fraction of positively labelled data when the volume of the hypersphere decreases.

The method is relatively resistant to noise and mislabelling errors. We reduce the noisy and mislabelled vectors from a produced description by adjusting the regularization parameters. Similarly like in SVM for multiclass classifiers, the user provides the regularization parameter. The training procedure is computationally expensive due to the time complexity of solving a quadratic programming problem. The complexity can be reduced when the simplicity of the employed constraints is taken into account in the design of the optimization routine. Large storage space is needed at the training phase since all the observation vectors from the training dataset are used in the optimization. The classification using this method is computationally simple and does not require a significant allocation of memory (Tax **2001**, Mazhelis **2006**).

Reconstruction methods

The model of the data generation process is assumed in reconstruction methods and the parameters of this model are estimated during the learning phase. The fit of the current observation vector the model is evaluated by the reconstruction error. The smaller reconstruction error, the more likely the data were generated by this model. The discriminant function can be implemented as an inverted reconstruction error (Mazhelis **2006**).

Principal Component Analysis

Principal Component Analysis (PCA) is aimed at explaining the internal variance and covariance structure of n-dimensional data in terms of the set of variables (principal components; Bishop **1995**). Principal components are linear combinations of the original variables. The PCA mapping finds the orthonormal subspace, which captures the variance in the data as best as possible. Neural network approaches exist for the optimization of a PCA. The simplest optimization procedure uses eigenvalue decomposition to compute the eigenvectors of the target covariance matrix. The eigenvectors with the largest eigenvalues are the principal axis of the d-dimensional data and point in the direction of the largest variance. These vectors are used as basis vectors for the mapped data. The number of basis vectors M is optimized to explain a certain, user-defined, fraction of the variance in the data. Because they form an orthonormal basis, the number of free parameters in the PCA becomes (Tax **2001**):

$$n_{PCA} = \binom{d-1}{M}. \qquad \text{Equation 12}$$

The PCA performs well when a clear linear subspace is present. Also for very low sample sizes, the data is automatically located in a subspace, e.g., 10 objects are always distributed in a 9-dimensional subspace. When the intrinsic dimensionality of the data is smaller than the feature size, the PCA can still generalize from the low sample size data. When

the data has variance in all feature directions, it might sometimes be impossible to reduce the dimensionality without reducing the fraction of the explained variance too much. For instance, when the user requests that 90% of the variance of some 2-dimensional data should be explained, it might happen that each of the two PCA features explains about 50% of the variance. In this case, no feature reduction can be applied and the complete feature space is described by two features. Therefore, all data will be accepted. Also when data is distributed in separate subspaces, the PCA will produce an average subspace, which may represent the data in each subspace very poorly. The PCA is relatively sensitive to the scaling of the features, it directly influences the feature variances. Scaling changes the order of the large variance directions and thus the PCA basis. When data directions are enhanced, this improves the PCA description, but when noise is amplified, it harms the characterization. Finally, because the PCA only focuses on the variance of the target set, the PCA is incapable of including negative examples in the training phase. The use of the PCA as the one-class classifier is justified when the dimensionality of the data analyzed by the classifier is high. Then, the computational complexity of classification can be decreased by reduction of the dimension using the PCA. The PCA is sensitive to the noise and outliers in training data, since they may distort the estimation of variances and covariances (Tax **2001**, Mazhelis **2006**).

Self-organizing map

Self-organizing map (SOM) is a clustering method, which assumes that data is clustered and can be described by a set of prototype vectors μ_k. In the SOM, the placing of the prototypes is not only optimized with respect to the data, but also constrained to form a low-dimensional manifold (Kohonen **1990**). The number K of prototype vectors should be selected beforehand. During classification, Euclidean distance is used in the definition of the error and the computation of the distance (Tax **2001**)

$$d_{SOM} = \min_k \|x - \mu_k\|^2 . \qquad \text{Equation 13}$$

The placement of prototype vectors is calculated from the training dataset. The simple competitive learning is employed, i.e. each subsequent observation vector x_i is used to update the position of the nearest prototype μ_k (Kohonen **1990**).

$$\mu_k(\tau + 1) = \mu_k(\tau) + \eta(\tau)(x_i - \mu_k) , \qquad \text{Equation 14}$$

where $\eta(\tau)$ is the learning rate ranging between zero and one. The SOM as well employs competitive learning to define the positions of prototype vectors. Not only the nearest prototypes are updated, but also the prototypes in a neighborhood of the nearest neighbor are also updated. More distant prototypes get a smaller update. This neighborhood is determined by a predefined topology. Often a 2- or 3-dimensional regular square grid is chosen for this manifold such that data mapped on this manifold can be visualized afterwards. Higher dimensional manifolds are possible, but the storage and optimization costs become prohibitive. When the dimensionality of the manifold does not fit the data, this topological constraint on the placing of the prototypes might result in suboptimal placing.
Thus, in the optimization of the SOM $k^{d_{SOM}}$ neurons have to be placed in the d- dimensional feature space. This means that the number of free parameters in the SOM becomes (Tax **2001**):

$$n_{freeSOM} = dk^{d_{SOM}} . \qquad \text{Equation 15}$$

The dimensionality of the manifold d_{SOM} and the learning rate are supplied by the user. Furthermore, the user defines a neighborhood function over the grid, which can even change during training.

The method is sensitive to remote outliers, because they may bias the placement of the prototype vectors. The noise present in training data can be compensated by a high number of observation vectors within each cluster. The method is computationally light; a small memory space is needed in order to store the prototype vectors (Mazhelis **2006**).

2 Material and Methods

2.1 SUBJECTS

Data were obtained from a total of 156 Czech native speakers. Participants were recruited for previous studies, but their speech patterns were never examined comprehensively.

The PD group included 48 participants (21 women, 27 men), all of who fulfilled the diagnostic criteria for PD. The diagnosis criteria of PD were established by the UK Parkinson's Disease Society Bank Criteria (Hughes et al. **1992**). All participants were on stable dopaminergic medication for at least 4 weeks before examinations, which were conducted in the on-medication state. All the PD participants were examined immediately after the diagnosis was made and before symptomatic treatment was initiated. The mean age of PD subjects was 62.0 years ± 11.6 standard deviation (SD) with range (34-82) , mean disease duration 6.6 years ± 5.3 SD, disease stage 2.6 ± 0.5 SD (1-3) according to the Hoehn & Yahr staging scale (ranging from 1 to 5, where 1 indicates mild unilateral motor disorder and 5 indicates confinement to a wheelchair; Hoehn and Yahr **1967**), mean motor score 21.0 ± 11.3 SD (6-43) according to the Unified Parkinson's Disease Rating Scale (UPDRS III; ranging from 0 to 108, with 0 representing a symptom-free state and 108 representing severe motor impairment; Stebbing and Goetz **1998**). None of the PD patients reported previous speech disorders unrelated to the present illness.

The HD group consisted of participants (24 women, 19 men), all of who fulfilled the diagnostic criteria for HD. The mean age of HD subjects was 46.5 years ± 14.0 SD with range (22-69), mean disease duration 6.1 years ± 3.7 SD, a mean number of CAG triplets 44.9 ± 3.6 SD (40-54). Most of the patients were treated with monotherapy or a combination of benzodiazepines, antipsychotics, amantadine and antidepressants. All HD patients were further assessed by a movement disorders specialist and were rated according to the motor score of the Unified Huntington's Disease Rating Scale (UHDRS, ranging from 0 to 124, where 0 indicates no motor disability and 124 indicates severe motor disability; Kieburtz et al. **2001**). In addition, the burden of disease score was calculated for each subject using a formula (Penney et al. **1997**):

$$burden\ disease\ score = age\ x\ (CAG\ repeat\ length - 35.5)\ .$$ Equation 16

The UHDRS motor score was 26.9 ± 11.6 (3-54) and the burden of disease score was 426.8 ± 78.9.

The group of 65 subjects (38 women, 27 men) with no history of neurological or speech disorder diagnostic was included as HC. Mean age of HC subjects was 54.9 years ± 12.5 SD (29-80) with range (29-80).

None of the HC, PD or HD subjects suffered from a chronic obstructive pulmonary disease, allergy, asthma, respiratory tract infection, facial paresis, or another malady that could negatively influence participant speech performance. The study was approved by the Ethics Committee of the General University Hospital in Prague and all participants were provided with written, informed consent.

2.2 PROTOCOL

Recordings were taken in a quiet examination room with a low ambient noise level using a condenser microphone at a distance of approximately 5 cm from the subject's mouth in order to minimize the influence of environmental noise. Data were digitalized with a sampling frequency of 48 kHz and 16-bit quantization. All participants were recorded shortly after the diagnosis was established. Each utterance was recorded during a single session by a speech-language pathologist. All participants were instructed before each task with information to know what specifically to perform in their task. The whole protocol required recording session approximately 30 minutes in duration.

Each speaker performed rhythm task, sustained phonation of vowels /A/ and /I/, monologue, reading passage and diadochokinetic task according to instructions summarized in Table 1. All participants were asked to repeat their production of an attempt that resulted in the flawless production of any task. Participants could repeat their production at any time if they or the speech therapist were not fully satisfied with their performance. All recorded acoustic signals were examined using digital signal processing.

Tasks	Speech data
Sustained phonation of vowel /I/	Sustained phonation of /I/ at a comfortable pitch and loudness as constant and long as possible, at least 5s. The task was performed on one breath.
Diadochokinetic task	Rapid steady /Pa/-/Ta/-/Ka/ syllables repetition as constant and long as possible, repeated at least 5 times. The task was performed on one breath.
Sustained phonation of vowel /A/	Sustained phonation of /A/ at a comfortable pitch and loudness as constant and long as possible, at least 5s. The task was performed on one breath.
Monologue	Monologue, at least 90s. The participants were generally instructed to speak about what they did that current day, their interests, job or family.
Reading passage	Reading the standardized text of 71 words.
Rhythm	Repeat the syllable /Pa/ at least 20 times at a comfortable and steady pace without acceleration or deceleration.

Table 1: List of vocal tasks

Note. Adapted from Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. The Journal of the Acoustical Society of America, 129(1), 350-367.

2.3 METHOD

All speech signals were analysed using specialized digital signal processing methods. Processing of signals consisted of pre-processing using high pass filter to attenuate main hum and other non-speech low-frequency noise and resampling of the sampling frequency, segmented to recognize the individual speech events using machine learning, and calculation of speech features using measurement methods to detect certain properties of speech performance.

The criteria for selection of measurement methods were that feature extraction should be fully automated and compatible with characteristics of the speech disorders analyzed. All data including extracted features were given by supervisor and in the following section, basic information about speech features are explained.

Phonatory features were measured by using intervals of voiced speech. Phonatory features provide information about disabilities to control the closing and opening of vocal folds. Standard methods include measures of the fundamental frequency (F0) mean, F0 range, standard deviation of power (stdPWR) and a standard deviation of the fundamental frequency (stdF0). StdPWR represents an abnormal variation in loudness observed in dysarthria. StdF0 measurements were determined using several vocal tasks such as sustained vowel phonation, reading a text and monologue. StdF0 represents a dysphonic symptom of impaired control of stationary voice pitch. By using sustained vowel phonation, the most popular measurements of voice functions were obtained, which are the perturbation measures jitter (extent of variation of voice range), shimmer (the extent of variation of expiratory flow) and harmonics-to-noise ratios (HNR, the amplitude of noise relative to tonal components in the speech) (Baken and Orlikoff **2000**), (Vokřál and Novák **1995**). Gaping in-between voiced speech (GVI) describes clear pauses between voiced speeches (Hlavnička et al. **2017**). To determine insufficiency of breath support for speech production, maximum phonation time (MPT) was examined. To investigate aperiodicity, a number of voice breaks (NVB), a degree of pitch breaks (DPB) and the degree of vocal arrest (DVA) were evaluated.

Speech rhythm abnormalities are commonly present in patients with the neurodegenerative disorder. A lower ability to reproduce perceived speech rhythm is one of the deficits in PD and HD speech. By measuring speech in task 8, it was possible to determine speech features that represent rhythm instability (RI) and rhythm acceleration (RA). Previous studies (Skodda et al. **2010, 2012**) have revealed that patients with impaired function of basal ganglia showed similar instabilities in speech production, in particular PD and HD manifest difficulties in the steady performance of single syllable repetition without speed alterations. There is a correlation between motor severity scores and RI for HD groups as well (Rusz et al. **2015**). Features involving information about the rhythmic organization of speech were evaluated as the rate of speech timing (RST) including voiced, unvoiced and pause intervals. Acceleration of speech associated with PD was computed by the acceleration of speech timing (AST). Net speech rate (NSR) is standard measurement used by pathologists. NSR is measured when a number of syllables is known, the total number of syllables is divided by the total duration of the speech. Duration of pause intervals (DPI) relates to the quality of speech timing. The heterogeneity of speech, in terms of the matter of voiced, unvoiced, pause and respiratory intervals, is described as the entropy of the speech timing (EST), which is Shannon entropy (Hlavnička et al. **2017**).

Another method is a measurement of articulation rate and pause characteristics that reveal differences of PD subjects in comparison with HC (Skodda and Schlegel **2008**; Forrest et al. **1989**). Articulation rate was calculated for reading the text. PD speakers have been found to have an overall lower intensity level, deficits in intensity range, and intensity variations during speech production (Watson and Munson **2008**). The measurement of intensity variations was determined using the reading text and the monologue. The most common method of evaluating **articulatory skills** is a diadochokinetic (DDK) task. DDK task measures the subject's ability to repeat a rapid and steady consonant-vowel (C-V) combination and usually consists of two measures. The average DDKR is the number of syllable vocalizations per second. The coefficient of DDKG measures the degree of rate variations in the period and assesses the ability to maintain a constant rate of C-V combinations. We can also measure the voice onset time (VOT) duration in DDK task. VOT is typically measured as the duration of time from the articulatory release of a stop consonant to the onset of voicing for the following vowel. The duration of unvoiced stops (DUS)

measured stop consonants, which is one of the most challenging articulatory movements. In addition, the temporal quality of articulation was determined from unvoiced fricatives using the decay of unvoiced fricatives (DUF; Hlavnička et al. **2017**). By reading the passage, articulatory decay (resonant frequency attenuation) was extracted. Resonant frequency attenuation (RFA) represents a decrease of spectral energy as a result of decayed articulatory movements (Rusz et al. 2016).

The **respiratory aspects** were assessed by using data from detected respiratory intervals and expirations represented by voiced and unvoiced intervals. The respiratory aspects were evaluated on a connected speech from monologue and reading passage tasks. The latency of respiratory exchange (LRE) measures the pause between expiration represented by the time speech ends and respective inspiration, the rate of speech respiration (RSR) calculates respiratory rate during speech and was computed from median duration between respiration, pause intervals per respiration (PIR) depicts breath groups. The relative loudness of respiration (RLR) evaluates audibility of respiration relative to the loudness of speech, eliminating dependence on microphone gain (Hlavnička et al. **2017**).

The feature calculation stage involved the application representative selection of traditional and non-standard measurement methods to all the speech signals. Altogether there were used 44 features, which produced a single number for each of the 156 signals (participants). See table 2 for the list of features of the measurement, which also includes standard deviation and the mean of all groups (HC, HD and PD) for each feature.

Many features of signals could be highly correlated with other features. Therefore, the correlation of normalized speech features with clinical scales was carried out by Pearson's correlation coefficient and for non-normally distributed speech features Spearman's correlation coefficient was used. Correlations were done separately for each group of participants, due to the different characteristics of these groups. Out of all 44 features, 4 features were not used. More specifically shimmer phonation (-0.88 with HNR phonation), NSR text (0.84 with RST text), DPI text (-0.86 with RST text) and PIR monologue (-0.84 with DPI monologue). These features showed high correlation amongst each other like shown in figure 2, figure 3, figure 4 and figure 5.

Using statistical tests, the hypothesis that the appropriate symptom may have abnormal values for a given group were verified. Some statistical tests, for example the analysis of variance, assume that variances are equal across groups or sample. For this purpose, Bartlett's test was used to evaluate the normality of the variances. Equal variances across features are called homogeneity of variances. The one-way analysis of variance (ANOVA) test with *post hoc* Bonferroni adjustment was used for homogeneous features and the Kruskal-Wallis test with *post hoc* Bonferroni adjustment was used for non-homogeneous features. With respect to the explorative nature of the current study and the fact that each acoustic variable represents a unique speech aspect, adjustment for multiple comparisons with regard to correlations was not performed. The level of significance was set to $p < 0.05$. In practice, only the most distinguishing features were selected to reduce the computational burden. Figure 5 highlights the feature selection process using a flowchart diagram.

Feature	HC		HD		PD	
	μ	σ	μ	σ	μ	σ
VOT	20,31	4,48	25,23	5,13	23,24	4,89
DDKR	6,72	0,74	4,34	1,48	6,21	0,90
DDKG	22,81	11,91	101,04	68,31	32,72	16,99
DDKW	0,39	0,04	0,32	0,09	0,39	0,05
EFn_M phonationI	-34,96	1,64	-32,63	2,47	-34,61	1,71
EFn_SD phonationI	3,80	0,87	3,86	1,14	3,58	0,99
RA rhythm	2,38	3,65	-1,42	11,74	3,38	2,94
RI rhythm	5,95	2,68	18,71	8,39	7,47	2,66
DVA phonationA	0,00	0,00	3,18	5,23	0,01	0,08
MPT phonationA	14,80	4,80	6,99	4,79	13,14	4,62
stdF0 phonationA	0,54	0,71	1,34	1,25	0,80	1,17
jitter phonationA	0,47	0,20	0,83	1,32	0,62	0,64
shimmer phonationA	2,21	0,84	3,56	2,88	2,56	1,56
HNR phonationA	21,14	3,23	17,09	4,74	19,39	3,86
pitchBreaks phonationA	0,79	2,27	2,44	5,72	2,51	9,19
EST text	1,56	0,01	1,55	0,01	1,56	0,01
RST text	420,48	48,81	318,88	63,89	386,82	60,39
AST text	14,22	13,30	10,91	13,71	11,75	12,57
DPI text	139,51	23,79	231,35	85,59	157,42	55,22
DVI text	215,86	29,92	265,41	47,92	227,88	37,17
GVI text	63,66	10,41	53,43	11,06	58,22	13,33
DUS text	28,16	7,93	43,46	10,99	34,61	13,49
DUF text	-2,81	2,70	-1,28	1,46	-1,42	2,34
RFA text	10,72	1,22	9,95	1,39	9,87	1,33
RLR text	-22,18	3,31	-21,46	4,44	-23,42	3,29
PIR text	7,69	2,99	4,40	1,49	6,28	2,19
RSR text	16,30	4,68	19,97	5,48	18,51	4,20
LRE text	150,33	86,78	185,37	103,29	144,92	63,33
stdPWR text	3,32	0,38	3,49	0,61	3,05	0,49
stdF0 text	2,67	0,80	2,07	0,84	1,65	0,59
NSR text	2,53	0,32	1,74	0,40	2,40	0,32
EST monologue	1,56	0,01	1,55	0,02	1,55	0,02
RST monologue	350,48	62,33	326,40	76,46	338,10	49,43
DPI monologue	185,42	62,53	234,06	83,44	205,91	66,34
DVI monologue	263,40	38,13	274,07	74,36	256,87	41,73
GVI monologue	52,48	10,43	51,68	11,60	47,96	11,11
DUS monologue	36,74	13,85	41,81	12,13	43,38	16,65
RFA monologue	9,01	1,28	8,52	1,52	8,77	1,24
RLR monologue	-20,49	4,36	-19,77	4,99	-22,64	3,83
PIR monologue	6,51	4,17	4,61	2,27	5,82	2,96
RSR monologue	15,04	4,33	18,94	5,58	16,01	4,36
LRE monologue	176,97	92,18	255,42	147,79	204,01	86,46
stdPWR monologue	3,86	0,66	4,18	0,61	3,57	0,71
stdF0 monologue	2,25	0,74	2,37	0,80	1,59	0,48

Table 2: List of features of acoustic measures

Table 2 includes the characteristics of each measurement, with the mean and standard deviation of values in the PD, HD and HC groups. The marked features represent the features that were chosen after eliminating process.

***Symbols**: μ = mean, σ = standard deviation.*

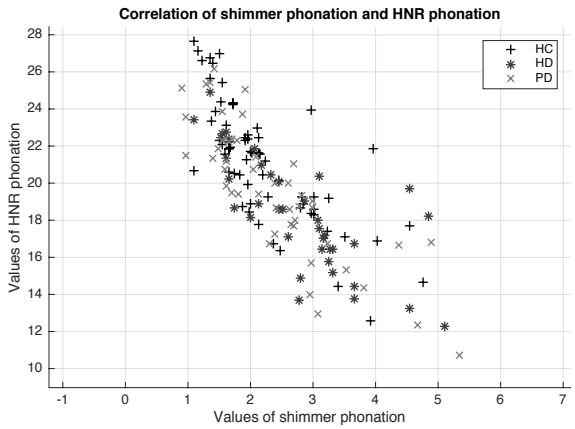


Figure 1: Correlation between shimmer and HNR
***Abbreviations**: HNR = harmonic noise ratio*

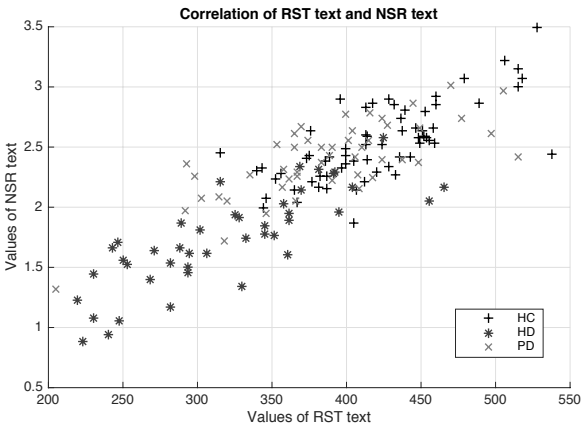


Figure 2: Correlation of RST and NSR
***Abbreviations:** RST = rate of speech timing, NSR = net speech rate*

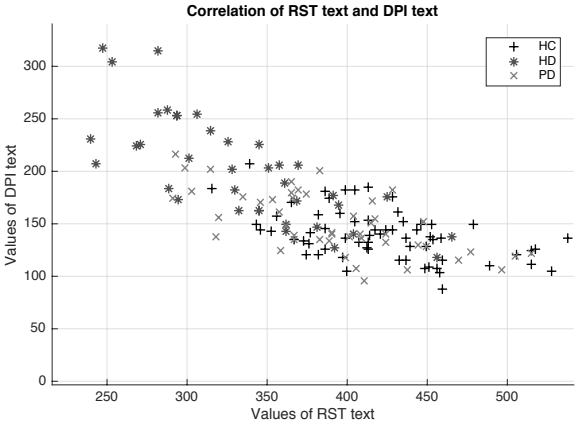


Figure 3: Correlation of RST text and DPI text
***Abbreviations**: RST = acceleration of speech timing, DPI = duration of pause timing*

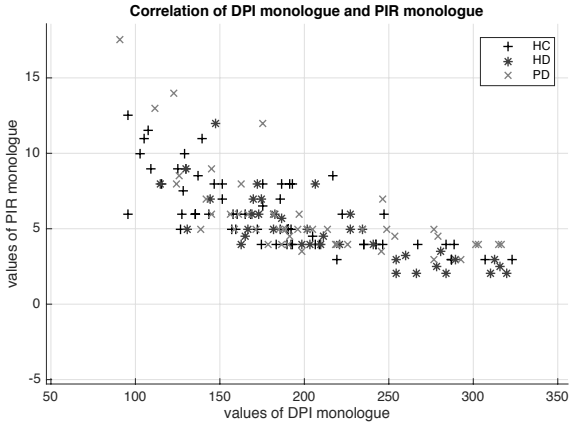
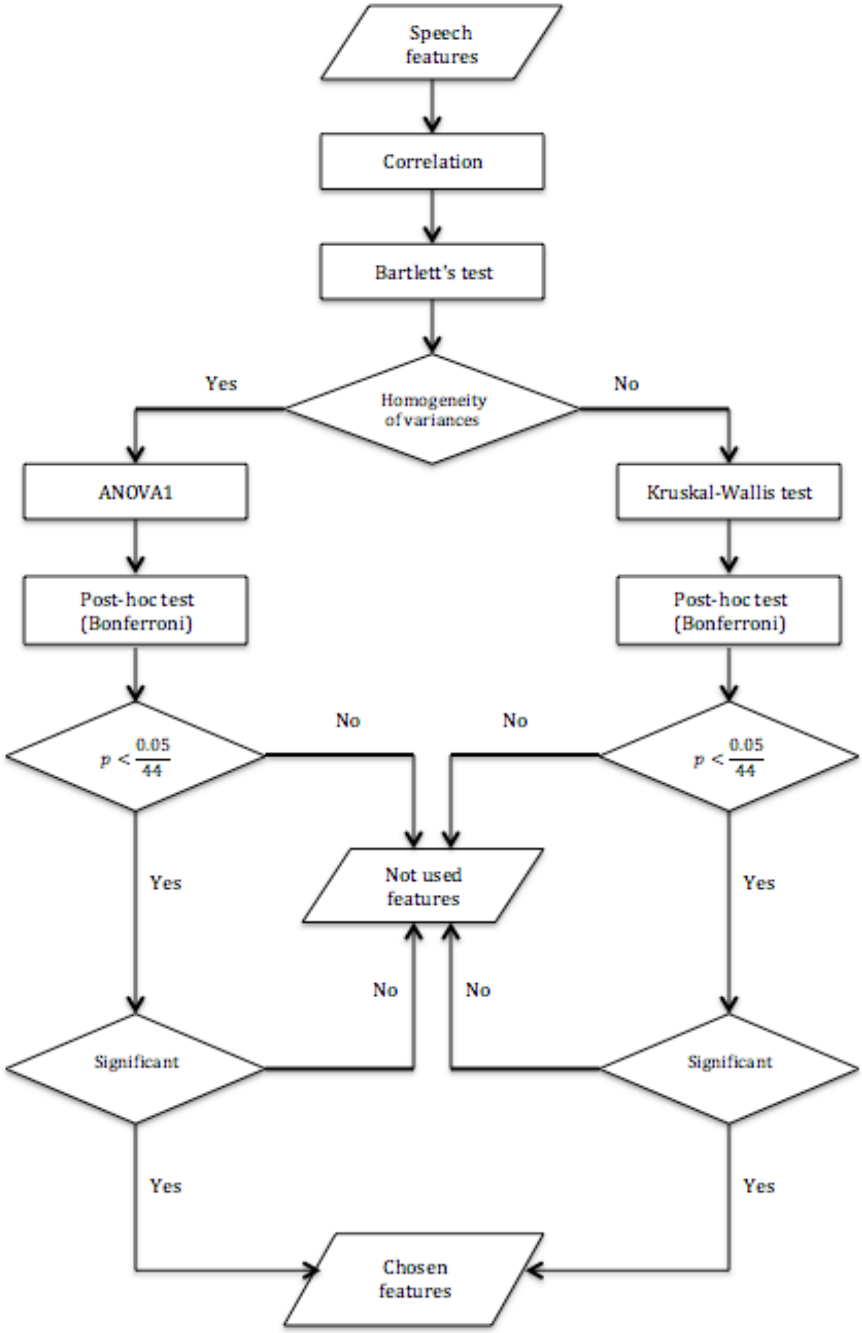


Figure 4: Correlation of DPI monologue and PIR monologue
***Abbreviations**: PIR = pause interval per respiration, DPI = duration of pause timing*

The symptomatology of Parkinson's disease and Huntington's disease is heterogeneous and so it made sense to check for accuracy of prediction models in order to find the best combination of features that contribute the most to the diagnosis of PD and HD. The following processes of cross-validation and classification were therefore done for each combination of features (total number of combinations = $2^{11} - 1$).

Because such a model is excessively complex due to a lot of descriptive features, all combinations of features were tested to select the most efficient ones. A technique called Cross-validation was used for assessing how the results of statistical analysis will generalize to an independent data set and to try to limit the overfitting problem, which may lead to outstanding performance on the trained data, but poor predictive performance in new data. Cross-validation methods split the sample into simulated training samples and testing samples.

Figure 5: Flowchart diagram depicting selections of features



The model was then trained on a training sample and evaluated on the testing sample. More specifically repeated random sub-sampling validation was used for all of the classifiers. The dataset was randomly shuffled into two subsamples where one subsample was retained as a validation data for testing (20% of data) and the other sample was used as training data (80% of data). The two groups of data were put into an algorithm that implements classification. Naive Bayes classifier, k-nearest neighbour classifier, QDA, Decision tree classifier and Support vector machines classifier were used for multi-class classification. For

k-nearest neighbour classifier, parameter k was set to 5. For the Bayesian form of QDA classifier, QDA modelled the likelihood of each class as a Gaussian distribution, then used the posterior distributions to estimate the class for a given test point. RBF kernel was applied in our SVM classifier to search for all feature combinations across acoustic features, which depended on parameters gamma and C. At first, the model was trained based on intuitively picked gamma (scale from 0.5 to 2 with iteration step 0.5) and c (scale from 0.5 to 2 with iteration step 0.5). To improve the accuracy of the prediction model, the function imagesc was used based on parameters c, gamma and meanACC, where new gamma and c were visually chosen based on the area of meanACC for the best combination. Various combination of C and gamma were tested and for value c = 3.7 and gamma = 1.5 we received the best accuracy. The overall classification performance of the SVM-based model was computed as the average percentage of correctly classified subjects into an appropriate group through all 10 iterations. For one-class classification we used Mixture of Gaussian, Parzen density estimation, k-nearest neighbor method, Support vector data description, Principal Component Analysis and Self-organizing map classifiers. For SOM classifier, a neighborhood function is defined by the user. We used the defaults in the SOM-tool Matlab toolbox, i.e. a Gaussian neighborhood, which decreases in size over time.

Each classifier created a prediction model, which was later tested on the test_data set by comparing the predicted labels to the actual test_labels of the test_data set. The outcome of a comparison between predicted labels and test_labels was marked either as true positive (TP), true negative (TN), false negative (FN) or false positive (FP). We had to establish a specific metric for each disease separately and these defined measures (TP, TN, FN and FP) represented different value.

We defined a category of general dysarthria in order to examine characteristics shared between hypokinetic and hyperkinetic dysarthria. For general dysarthria, a variable **true positive** referred to HD and PD correctly identified as dysarthria (HD or PD). **False positive** referred to HC incorrectly identified as PD and HC incorrectly identified as HD. **True negative** referred to HC correctly identified as HC. **False negative** to PD and HD people incorrectly identified as HC.

A category of hypokinetic dysarthria was made in order to distinguish hypokinetic and non-hypokinetic speech pattern. For hypokinetic dysarthria, **true positive** referred to PD identified as hypokinetic speech pattern. **False positive** referred to HC identified as hypokinetic speech pattern and HD identified as hypokinetic speech pattern. **True negative** referred to HC identified as non-hypokinetic speech pattern and HD identified as non-hypokinetic speech pattern. **False negative** referred to PD incorrectly identified as non-hypokinetic speech pattern.

A category of hyperkinetic dysarthria was made in order to distinguish hyperkinetic and non-hyperkinetic speech pattern. For hyperkinetic dysarthria, **true positive** referred to HD identified as hyperkinetic speech pattern. **False positive** referred to HC identified as hyperkinetic speech pattern and PD identified as hyperkinetic speech pattern. **True negative** referred to HC identified as non-hyperkinetic speech pattern and PD identified as non-hyperkinetic speech pattern. **False negative** referred to HD incorrectly identified as non-hyperkinetic speech pattern.

From these new variables, the positive predictive value, accuracy, sensitivity and f-score of the prediction model were calculated.

Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made. Calculation of the accuracy of the prediction model was done by using this formula:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad \text{Equation 17}$$

Precision (also called positive predictive value) is a measure that tells us what proportion of patients that were diagnosed as having disease actually had the disease. The calculation was done using the following formula:

$$precision = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ positives}. \quad \text{Equation 18}$$

Precision shows us how accurate our model is out of the predictive positive. It is a good measure to determine when the costs of false positive are high.

Recall (known as sensitivity) referred to the model's ability to correctly detect patients who do have the dysarthria. The calculation was done using the following formula:

$$recall = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ \ negatives}. \quad \text{Equation 19}$$

Recall is metric to use for selecting our best model when there is a high cost associated with false negative. E.g., during detection of general dysarthria amongst patients, if the patient with dysarthria goes through test and is predicted as not sick, this result will leave consequences on earlier treatment and therefore life expectancy.

F-score is needed to seek a balance between precision and recall. F-score is the harmonic average of the precision and recall, where F-score reaches its best value at 1 and worst at 0. It is very important to use this metric, especially when there is an uneven class distribution (a large number of true negatives).

The calculation was done using this formula:

$$Fscore = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \quad \text{Equation 20}$$

Due to the randomization of data in 2-fold cross-validation process the 2-fold cross-validation, classification and calculation of accuracy of the prediction model were repeated 15 times. In the end, mean of all 10 accuracies was determined for the specific combination of features. Finally, mean accuracies of each combination of features were sorted in order to get the best mean accuracy for the specific classifier. It was also possible to determine which features were the most relevant for the identification of speakers. For each classifier, we received an arithmetic average of mean accuracy, mean precision, mean recall and f-score of 3 best combinations of features.

3 Results

After feature selection process and with the process of repeated random sub-sampling validation cross-validation and classification of the prediction model, we received mean recalls, mean precisions and mean accuracies of all combinations.

3.1 GENERAL DYSARTHRIA

3.1.1 Multi-class and one-class comparison

In table 3, we summarize the results of each classifier including one-class and multi-class classifiers. For a case of multi-class classifiers, the highest average mean accuracy 91,21% was reached with Naive Bayes classifier, closely behind it, was QDA with average mean accuracy 90,16%. Lowest accuracy was reached with SVM classifier (average mean accuracy 87,74%) and KNN (average mean accuracy 88,47%) classifier. The difference of accuracy between SVM classifier and Naive Bayes classifier is not significantly large and overall these multi-class classifiers showed a positive outcome. Furthermore, it is shown that QDA and Naive Bayes classifiers reached the highest percentage of recall, which means that model's ability to correctly detect patients with dysarthria is very good, that all is supported with high precision and f-score over 90%.

Compared to the results of multi-class classifiers, results of one-class classifiers were also quite impressive. SOM classifier reached 82,84 % and PCA had the highest average mean accuracy 83,91%. For both of these classifiers, f-score reached a similarly high value as in case of multi-class classifiers. Similarly to the results of multi-class classifiers, KNND (74,93% average mean accuracy) and SVDD (76,62% average mean accuracy) showed the lowest average mean accuracy for one-class classifiers

3.1.2 Feature selection

Table 4 shows the best combinations of features for each multi-class classifier and table 5 for each one-class classifier. Despite high accuracy not all of these features were represented in all classifiers. Only these features (DDKG, RI rhythm, RFA text, PIR text and stdF0 text) have appeared 9 or more times in best combinations, which is shown in table 5 for multi-class classifier. In comparison with multi-class classifiers, only 4 features (DDKR, DDKG, RI rhythm and DUS text) have appeared 9 or more times for one-class classifiers. Therefore, by observing occurrences of features in both classifiers, we conclude that results showed RI rhythm and DDKG significant features in the process of diagnosis of patients with dysarthria.

3.1.3 Reliability of classifiers

The accuracy of prediction models was calculated for each best combination of features from a certain classifier. The resulting comparison of accuracies of each prediction model is shown in table 6 for multi-class classifiers and table 7 for one-class classifiers. We observed from table 6, that Naive Bayes was one of the most reliable for multi-class classifier and from table 7 that SOM was most reliable for the case of the one-class classifier.

	Classifiers	Accuracy	Recall	Precision	F-score
Multi-class	**SVM**	87,74%	84,89%	98,42%	91,15%
	KNN	88,47%	83,78%	98,41%	90,51%
	QDA	90,16%	88,56%	96,51%	92,36%
	NAIVE	91,21%	88,33%	99,61%	93,63%
	TREE	89,58%	85,49%	94,87%	90,15%
One-class	**KNND**	74,93%	65,30%	90,03%	75,69%
	MOG	80,71%	92,99%	94,36%	93,67%
	PARZEN	81,78%	94,36%	99,40%	96,82%
	PCA	83,91%	100,00%	81,68%	89,91%
	SOM	82,84%	95,73%	83,05%	88,94%
	SVDD	76,62%	84,62%	100,00%	91,67%

Table 3: Accuracy, recall, precision and F-score of general dysarthria

Table 3 includes 5 classifiers from multi-class classifiers and 6 classifiers from one-class classifiers. Each row contains one classifier and average of 3 best results of mean accuracy mean recall, mean-precision and mean f-score of each classifier.

	Features										
Classifiers	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
SVM	1	0	1	1	0	1	1	1	0	1	0
	1	0	0	1	0	0	1	0	0	0	1
	1	0	1	1	0	0	1	1	0	1	1
KNN	1	0	1	1	0	0	0	0	1	0	1
	1	1	0	1	1	0	0	1	1	0	1
	1	0	1	1	0	1	0	1	1	0	1
QDA	1	0	0	1	1	0	1	1	1	0	1
	0	0	0	1	0	0	0	1	1	0	1
	0	0	1	1	0	0	1	1	1	1	1
NAIVE	1	1	1	1	1	0	0	1	0	0	1
	0	1	1	0	0	1	1	1	1	0	1
	0	0	1	1	1	0	0	0	1	1	1
TREE	0	0	1	1	1	1	0	0	1	0	1
	0	0	1	1	0	0	0	0	1	0	1
	0	0	1	1	0	1	0	0	0	1	1
Total	8	3	11	14	5	5	6	9	10	5	14

Table 4: 3 best combinations of features for each multi-class classifier

The best 3 combinations of features are listed in the column for each multi-class classifier. They are represented with specific features in table 4. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen

Classifiers	Features										
	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
	0	0	1	1	0	0	1	0	1	0	1
KNND	0	1	1	1	1	0	1	1	1	0	1
	0	0	1	1	1	0	0	0	0	0	0
	0	1	0	0	0	0	1	0	0	0	0
MOG	0	0	1	1	0	0	0	0	0	0	0
	0	1	1	0	0	0	1	0	0	0	0
	0	0	1	1	0	0	0	0	0	0	0
PARZEN	1	1	0	1	0	0	1	0	0	0	0
	0	0	1	0	0	0	1	0	0	0	0
	0	1	1	1	0	0	1	0	1	1	0
PCA	0	1	1	1	0	0	1	1	0	1	1
	1	0	1	1	0	0	1	1	1	0	0
	1	0	1	0	0	1	1	0	0	1	1
SOM	1	1	0	1	0	0	1	1	0	0	1
	0	0	1	1	0	1	1	0	0	0	0
	0	1	0	1	0	0	0	0	0	0	0
SVDD	0	1	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	1	1	0	0	1
Total	4	9	12	12	2	2	13	5	4	3	7

Table 5: 3 best combinations of features for each one-class classifier

The best 3 combinations of features are listed in the column for each one-class classifier. They are represented with specific features in table 5. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen.

Classifiers	QDA	NAIVE	TREE	KNN	SVM
QDA	86,3%	77,5%	69,4%	72,5%	71,3%
NAIVE	77,5%	86,3%	75,0%	70,6%	78,1%
TREE	64,4%	61,3%	78,1%	73,1%	63,1%
KNN	67,5%	61,3%	61,9%	80,6%	63,1%
SVM	66,5%	67,0%	66,0%	70,5%	84,5%

Table 6: Combination of the multi-class classifier accuracy for the best combination of features

Table 6 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

Classifiers	KNND	MOG	PARZEN	PCA	SOM	SVDD
KNND	75,20%	66,40%	69,33%	71,73%	69,07%	66,93%
MOG	65,87%	81,07%	80,53%	64,53%	62,40%	73,87%
PARZEN	69,07%	77,33%	82,67%	58,93%	63,47%	78,67%
PCA	78,67%	59,73%	77,07%	84,00%	76,53%	56,53%
SOM	76,80%	66,40%	75,20%	77,07%	82,93%	77,07%
SVDD	48,00%	75,47%	61,87%	48,00%	48,00%	77,33%

Table 7: Combination of the one-class classifier accuracy for the best combination of features

Table 7 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

3.2 HYPERKINETIC DYSARTHRIA

In the section for results of hyperkinetic dysarthria, we determined how accurate are models in process of distinguishing HD from PD and HC.

3.2.1 Multi-class and one-class comparison

Table 8 showed similar average mean accuracy amongst multi-class classifiers. Correspondingly to results of multi-class classifiers from general dysarthria section, models for hyperkinetic dysarthria showed in general high average mean accuracy. NAIVE and TREE reached the highest average mean accuracy 88,15%. QDA on the other hand reached the lowest average mean accuracy 84,68%, but still only 3,5%, which is not markedly different. On the contrary, one-class classifiers showed poor performance. The results of one-class classifiers were not consistent. Highest average mean accuracy around 73,07% was reached with SVDD and 71,82% average mean accuracy with PARZEN classifiers. However, the result of mean F-score of these two classifiers showed differences. SVDD had 82,90% mean F-score against only purely 41,11% of mean F-score from PARZEN. PCA and SOM classifiers performed poorly with barely 46% average mean accuracy.

3.2.2 Feature selection

In summary, from presented results in table 9 we can see the best combinations of features for each multi-class classifier and table 10 for each one-class classifier. Features (DDKR, RI rhythm, PIR text and stdF0 text) have appeared 9 or more times in best combinations for multi-class classifier. Features (DDKR, GVI text, RFA text, PIR text, RSR text and stdf0 text) have appeared 9 or more times for one-class classifiers. Thus, features DDKR, PIR and stdF0 text were most significant features for both classifiers in the process of separating HD patients from healthy patients and PD. Even though these features were most significant, not all of them were a representative feature for each classifier, e.g. Feature DDKR was not amongst best 3 results of SVDD classifier.

3.2.3 Reliability of classifiers

Comparison of the average mean accuracies of the prediction models based on each classifier's best combination of features is presented in table 11 for multi-class classifiers, respectively table 12 for one-class classifiers. Despite the best result of NAIVE and TREE classifiers, KNN classifier seemed to be one of the most reliable for the multi-class classifier. SVDD classifier was most reliable for the case of one-class classifier. But as mentioned above, reliability of these classifiers was low due to their inconsistency.

3.3 HYPOKINETIC DYSARTHRIA

Similarly to hyperkinetic dysarthria part, where we determined how accurate are models in process of distinguishing HD from PD and HC. In this part, we evaluated the accuracy and the reliability of models to determine how well these classifiers differentiated PD from HC and HD.

Classifiers		Accuracy	Recall	Precision	F-score
Multi-class	**SVM**	86,14%	61,25%	89,02%	72,57%
	KNN	85,63%	64,75%	78,78%	71,08%
	QDA	84,68%	68,75%	74,99%	71,73%
	NAIVE	88,15%	67,75%	87,04%	76,19%
	TREE	88,15%	67,75%	85,23%	75,49%
One-class	**KNND**	55,73%	53,65%	30,28%	38,71%
	MOG	60,36%	81,90%	27,96%	41,68%
	PARZEN	71,82%	82,54%	27,37%	41,11%
	PCA	39,38%	100,00%	30,40%	46,63%
	SOM	39,20%	98,10%	29,71%	45,61%
	SVDD	73,07%	70,79%	100,00%	82,90%

Table 8: Accuracy, recall, precision and F-score of hyperkinetic dysarthria

Table 8 includes 5 classifiers from multi-class classifiers and 6 classifiers from one-class classifiers. Each row contains one classifier and average of 3 best results of mean accuracy mean recall, mean-precision and mean f-score of each classifier.

	Features										
Classifiers	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
SVM	1	1	0	1	0	0	0	1	1	0	1
	0	1	0	1	0	0	0	1	1	0	1
	1	1	1	1	0	0	0	0	1	1	1
KNN	0	1	0	1	0	0	0	1	0	0	1
	0	0	0	1	0	0	0	0	0	0	1
	1	0	1	1	0	0	0	0	1	0	1
QDA	1	0	0	1	1	0	1	1	1	0	1
	0	1	0	1	1	0	1	1	0	1	1
	0	1	0	1	0	0	0	0	0	1	1
NAIVE	1	1	1	1	1	0	0	1	0	0	1
	1	1	0	1	0	1	0	0	1	0	1
	1	0	1	1	0	0	1	1	1	1	1
TREE	0	0	1	1	0	0	0	0	1	0	1
	0	1	1	1	0	0	1	0	1	1	1
	0	0	1	1	0	1	0	0	0	1	1
Total	7	9	7	15	3	2	4	7	9	6	15

Table 9: 3 best combinations of features for each multi-class classifier

The best 3 combinations of features are listed in the column for each one-class classifier. They are represented with specific features in table 9. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen.

Classifiers	Features										
	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
	0	0	0	0	0	1	0	0	1	1	1
KNND	0	1	0	0	0	1	0	1	1	1	1
	1	1	0	0	0	0	0	1	0	1	0
	1	1	1	1	0	0	0	1	1	1	1
MOG	1	1	1	1	0	1	1	1	1	1	0
	1	1	1	1	0	0	1	0	0	1	1
	1	1	1	1	1	1	1	1	1	1	0
PARZEN	1	1	1	1	0	1	1	1	1	1	1
	1	1	1	1	1	0	1	1	1	1	1
	0	1	0	0	0	1	0	0	0	0	0
PCA	0	1	0	0	0	1	0	0	0	0	1
	0	1	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	1	0	0	1
SOM	0	1	0	0	0	1	0	1	0	1	0
	0	0	0	0	0	0	0	0	0	0	1
	1	0	0	0	0	1	0	1	1	0	0
SVDD	1	0	0	1	0	1	0	0	1	0	0
	0	0	1	0	0	0	0	1	1	1	0
Total	9	12	7	7	2	10	5	11	10	12	9

Table 10: 3 best combinations of features for each one-class classifier

The best 3 combinations of features are listed in the column for each one-class classifier. They are represented with specific features in table 10. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen.

Classifiers	QDA	NAIVE	TREE	KNN	SVM
QDA	86,89%	83,28%	79,76%	78,50%	81,23%
NAIVE	82,00%	86,23%	82,53%	80,65%	80,99%
TREE	81,81%	81,29%	85,03%	82,28%	81,67%
KNN	86,73%	84,76%	84,66%	88,85%	84,99%
SVM	82,70%	83,63%	81,03%	83,47%	89,55%

Table 11: Combinations of the multi-class classifier accuracy for the best combination of features

Table 11 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

Classifiers	KNND	MOG	PARZEN	PCA	SOM	SVDD
KNND	56,00%	38,13%	39,20%	45,87%	51,73%	47,73%
MOG	39,47%	61,07%	35,47%	33,07%	40,27%	40,27%
PARZEN	45,07%	67,47%	72,00%	34,13%	34,13%	38,13%
PCA	33,60%	17,33%	14,93%	39,47%	37,07%	32,53%
SOM	31,73%	21,60%	19,47%	35,47%	40,00%	30,40%
SVDD	69,87%	72,00%	72,00%	35,20%	37,07%	73,60%

Table 12: Combination of the one-class classifier accuracy of the best combination of features

Table 12 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

3.3.1 Multi-class and one-class comparison

Table 13 contains average mean accuracies of multi-class classifiers, which were equally very high. These classifiers had around 95% average mean accuracy. Likewise, the average mean F-score of these classifiers also corresponded with high performance up to 97%. Opposite to that, results of one-class classifiers in table 13 showed notably lower performance. Compared to other one-class classifiers, just a few one-class like SVDD and PARZEN classifiers are able to use labelled outliers in the training and therefore they are more robust against outliers. Thus PARZEN (84,36% average mean accuracy) and SVDD (81,87% average mean accuracy) classifiers showed the highest average mean accuracy. Good performance of PARZEN and SVDD classifier underlies also high average mean f-score of these two classifiers.

3.3.2 Feature selection

Features (DDKG, RI rhythm and RSR text) have appeared 9 or more times in table 14 for best combinations of multi-class classifier. Features (DDKR, DDKG, RI rhythm, PIR text, RSR text and stdf0 text) have also appeared 9 or more times for one-class classifiers in table 14. Thus, we can assert that DDKG, RI rhythm and RSR text are most significant features to separate PD from healthy patients and HD patients.

3.3.3 Reliability of classifiers

Comparison of the average mean accuracies of the prediction models based on each classifier's best combination of features is presented for multi-class classifiers and one-class classifiers in table 16 and table 17. Regardless of the good performance of all multi-class classifiers, comparison of their average mean accuracies for each specific multi-class classifier showed that KNN is a most reliable classifier. SVDD classifier was most reliable for one-class classifier just as it was in the case of hyperkinetic dysarthria. SVDD presents low sensitivity to errors in values of features and outliers and thus is a very robust method. It is also comparatively resistant to noise.

Classifiers		Accuracy	Recall	Precision	F-score
Multi-class	**SVM**	95,92%	94,80%	100,00%	97,33%
	KNN	94,57%	89,20%	100,00%	94,29%
	QDA	95,66%	94,80%	98,26%	96,50%
	NAIVE	95,81%	93,60%	97,61%	95,56%
	TREE	96,14%	96,40%	97,00%	96,70%
One-class	**KNND**	66,84%	72,89%	33,34%	45,75%
	MOG	73,69%	97,78%	37,63%	54,34%
	PARZEN	84,36%	95,56%	100,00%	97,73%
	PCA	51,11%	100,00%	27,84%	43,55%
	SOM	54,13%	100,00%	27,50%	43,14%
	SVDD	81,87%	77,78%	100,00%	87,50%

Table 13: Accuracy, recall, precision and F-score of hypokinetic dysarthria
Table 13 includes 5 classifiers from multi-class classifiers and 6 classifiers from one-class classifiers. Each row contains one classifier and average of 3 best results of mean accuracy, mean recall, mean-precision and mean f-score of each classifier.

Classifiers	Features										
	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
	1	0	0	1	1	0	0	1	0	1	0
SVM	0	1	1	1	0	0	0	1	0	0	0
	1	0	0	1	1	0	0	1	0	0	0
	1	0	1	1	1	1	0	0	1	1	1
KNN	1	0	1	1	1	1	1	0	1	1	0
	0	1	1	1	0	0	0	0	0	0	0
	0	0	1	1	1	0	0	1	0	1	1
QDA	0	0	1	1	0	0	1	1	0	1	1
	0	1	1	1	1	0	0	1	0	0	1
	1	1	1	1	0	0	0	0	0	0	1
NAIVE	0	1	0	1	0	0	1	0	0	1	1
	0	1	1	1	1	0	0	1	1	1	0
	0	0	1	1	1	1	0	0	1	1	1
TREE	0	1	1	1	0	0	0	0	1	0	1
	0	1	1	1	0	1	0	0	1	1	0
Total	5	8	12	15	8	4	3	7	6	9	8

Table 14: 3 best combinations of features for each multi-class classifier

The best 3 combinations of features are listed in the column for each one-class classifier. They are represented with specific features in table 14. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen.

Classifiers	Features										
	VOT ddk	DDKR	DDKG	RI rhythm	RST text	GVI text	DUS text	RFA text	PIR text	RSR text	stdF0 text
	0	0	1	1	0	0	0	0	0	1	1
KNND	0	1	0	1	0	0	0	1	1	0	1
	1	0	1	0	1	1	0	1	1	0	1
	0	1	1	1	0	0	1	1	0	1	1
MOG	1	1	1	1	0	0	0	1	0	1	0
	1	0	1	1	0	0	0	0	1	1	1
	0	1	1	1	1	1	0	0	1	1	1
PARZEN	1	1	1	0	1	1	0	1	1	1	0
	1	1	1	1	1	1	0	0	0	1	1
	0	1	0	1	0	1	0	0	0	0	1
PCA	1	0	0	1	0	0	0	0	0	0	1
	0	0	0	1	0	0	0	1	1	1	0
	1	0	0	1	0	0	0	0	1	0	1
SOM	1	1	1	1	0	0	0	1	0	0	0
	0	0	1	1	0	0	0	0	0	0	0
	0	1	1	0	0	0	0	0	0	1	1
SVDD	0	0	0	0	1	0	0	1	1	0	1
	0	0	0	0	1	0	0	0	1	0	1
Total	8	9	11	13	6	5	1	8	9	9	13

Table 15: 3 best combinations of features for each one-class classifier

The best 3 combinations of features are listed in the column for each one-class classifier. They are represented with specific features in table 15. These features (VOT ddk, DDKR, DDKG, RI rhythm, RST text, GVI text, DUS text, RFA text, PIR text, RSR text and stdF0 text) are the most significant in the process of determining whether the patient has dysarthria or not. Only the features that have appeared in the best combinations 9 or more times were chosen.

Classifiers	QDA	NAIVE	TREE	KNN	SVM
QDA	96,78%	86,15%	92,70%	95,18%	88,57%
NAIVE	86,61%	95,26%	92,03%	88,65%	91,28%
TREE	89,58%	91,58%	95,98%	90,96%	91,48%
KNN	92,82%	93,39%	92,92%	95,97%	91,86%
SVM	86,26%	90,84%	88,97%	91,44%	96,40%

Table 16: Combination of the classifier accuracy for the best combination of features

Table 16 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

Classifiers	KNND	MOG	PARZEN	PCA	SOM	SVDD
KNND	67,20%	58,67%	60,80%	59,73%	57,60%	56,53%
MOG	56,80%	74,13%	57,07%	49,07%	60,27%	49,33%
PARZEN	62,13%	77,87%	84,53%	48,53%	52,80%	60,00%
PCA	46,93%	39,73%	41,07%	51,20%	48,00%	31,73%
SOM	47,20%	48,00%	43,73%	39,73%	55,20%	46,13%
SVDD	79,47%	80,00%	80,00%	76,00%	73,07%	82,13%

Table 17: Combination of the classifier accuracy of the best combination of features

Table 17 is a comparison of accuracies. The best combination of features for each classifier was tested on other classifiers. The best accuracy of the best combination of features of each classifier corresponding to each row is evaluated in each column for specific classifier.

4 Discussion

PD is associated with motor symptoms and non-motor symptoms that precede the motor symptoms by more than a decade (Kalia and Lang **2015**). So, the early diagnosis of PD is largely dependent on various non-motor symptoms, speech impairment being one of the earliest manifestations. Machine learning could rapidly aid this process of detection patients with dysarthria and it can be a key tool. To diagnose the diseases, machine learning systems are constructed from data set of healthy and unhealthy patients. Cases chosen for disease prognosis are representative of different disease states. With the use of common multiclass classifiers such as Naive Bayes (NB), k-Nearest Neighbors (KNN), Support Vector Machine classifiers (SVM), classification trees and Quadratic Discriminant Analysis (QDA) we evaluated speech patterns of dysarthrias. We also evaluated speech patterns using one-class classifiers such as Mixture of Gaussians (MOG), Parzen density estimations (PARZEN), Principal component analysis (PCA), Self-organizing map (SOM), Support vector data descriptions (SVDD) and K-nearest neighbor methods (KNND).

4.1 Performance of classifiers

Results in a category of general dysarthria showed best results of Naive Bayes classifier. It makes sense since the features are independent of each other and the NB does not need many observations to perform well. Moreover, Naive Bayes less likely overfit the training data that suffer from smaller a sample size. The reduction of features at the start and taking into account the correlations amongst the features may have contributed to the accuracy of NB, since the performance of NB greatly improves if the data does not contain highly correlated features. The result of QDA was also very high and compared to NB, QDA learns quadratic boundaries and is therefore more flexible to use. On the other hand, KNN and SVM reached the lowest accuracy. The problem with KNN might be because one category occurred more than another. In our case there were more HC than PD and HD and since we did not apply any weight to the more common category it might have affected the classifier's accuracy in the end. Lower accuracy of SVM can be explained by the fact that the accuracy of SVM is very sensitive to the chosen gamma and c parameter.

PCA and SOM reached comparable results to the results of multi-class classifiers. As mentioned, both these classifiers belong to reconstruction methods, where a model is chosen and fitted to the data based on knowledge of the data and it makes presumptions about the generating process. PCA classifier performed well due to the presence of clear linear subspace of data and this method is not so sensitive to the scaling of the features. For the case of SOM classifier, the learning rule was adapted to also repel the outlier objects and therefore this method indicated robustness to outliers in the training data, which derived in the good performance of the model. KNND and SVDD similarly to KNN and SVM showed the lowest accuracy. Both these classifiers are scale sensitive of the feature values. KNND uses the distance in the evaluation of a test object and SVDD by the use of a Gaussian kernel. KNND is also the least robust to noise. An outlier will cause a portion of the feature space to be acceptable. In the case of KNDD no parameters are present and no model is assumed. And so, this method depends completely on the training set and its distribution of features.

In the category of hyperkinetic and hypokinetic dysarthria, PARZEN and SVDD achieved the best results. PARZEN method has an advantage when there is a small sample of size of features; the width of the parameter is equal for all directions in the feature space. For lower sample sizes, a method such SVDD directly estimates the boundary and is also more preferred. Recent developments from the field of statistical learning theory have shown that kernel-based methods such as SVDD are suited to solve machine learning problems in high dimensions (Platt et al. **2001**). SVDD estimates the support of distribution by identification of a region in input space by nonlinearly projecting the data into a feature space and separating data from the origin as a margin without extra computational costs. In the study of outlier detection with one-class SVMs (Dreiseitl et al. **2010**), where the findings indicated that the classification via outlier detection using one-class SVM offered performance comparable to regular classification algorithms, our results of SVDD showed similar outcome.

4.2 Significant features

Previous studies (D'Alatri et al. **2008**) showed that PD patients have problems with articulation, which is the modification of the position and shape of the speech organs (e.g., tongue) in the creation of sound. Diadochokinetic task is the most common method of evaluating articulatory skills, and our results showed that DDKG in the category of general dysarthria has also been chosen as one of the most significant features. Beside DDKG, RI rhythm was a very significant feature. Feature RI rhythm is associated with irregular pace due to increased speech motor control, impaired timing or discoordination. Our finding on RI rhythm as significant feature for detection of dysarthria is in general agreement with previous researches which demonstrated impairment of vocal pace stability in PD and HD (Skodda et al. **2012, 2014**).

In accordance with the majority of previous studies mentioned in the introduction, the characteristic of hyperkinetic dysarthria commonly indicates in form of decreased rate of speech intervals, unpredictable articulatory breakdown and phonatory dysfunction. Thus, features DDKR, PIR and stdF0 text, which were most dominant for both classifiers showed as well as in our results the most significant features in the process of separating HD patients from HC and PD patients. The standard deviation of F0 (stdF0) is the fundamental frequency or pitch of vocal oscillations, which is one of the traditional features measured when examining phonation. HD patients are generally attributed to disruptions of phonation.

Skodda reported impaired steadiness of syllable repetition in early motor stages of PD (Skodda **2015**), likewise that, our results demonstrated RI feature as one of the most significant features for detection of hypokinetic dysarthria as well as and RSR text and DDKG features. As mentioned in the introduction, speech impairment which includes respiration problems has been found to be one of the earliest manifestations of PD (Postuma et al. **2012**, Harel et al. **2004**), our results also documented that. RSR, which is one of respiration feature, relates to the inefficiency of air-flow management during speech production or decreased control of respiratory movements and is therefore a very important feature to detect patients with PD.

4.3 Clinical practice

In this thesis, we assessed the one-class classification methods and compared their reliability and accuracy with multi-class classification methods. On closer examination, the results

showed the satisfying performance of one-class classifiers, especially results of general dysarthria detection showed that all classifiers performed well in comparison with multi-class classifiers. For a category of the general dysarthria detection, the average accuracy of all one-class classifiers is 80% and multi-class classifiers are 89%. We could present the 9% accuracy difference of these models in the following example. In the general population, the prevalence of Parkinson's disease is 0,3% which means that with the population of 1,28 million in Prague, 384 people have Parkinson's disease. This means that one-class classifiers would detect correctly 307 out of 384 cases and multi-class classifiers would detect 343 out of 384 cases. 9% accuracy between these models represents 36 misdetections. The question however is, if this deviation is considerable. Multi-class models describe data with similar attributes better than one-class models, because they retrieve additional information from the other class and by that they minimalize prediction error. However, it does not necessarily mean that they describe better the hypotheses of disease development. Important is how to understand the interpretation of our results. One class-classifiers indicates whether a combination of features is abnormal and despite the occurrence of an error, results could be still interpreted. With multi-class classifiers, we receive better performance, but results can be elucidated only to that specific disease. It is worth mentioning that hypokinesia occurs in HD with cognitive dysfunction. Hypokinesia is characterized by loss of muscle movement due to disruption in the basal ganglia. This corresponds with PD patients, who experience muscle rigidity and inability to produce movement. Therefore, the overlap of results and error rate is not completely defective.

High average mean recall, precision and f-score in results indicated reliable identification of healthy patients from those with diseases. The results of these metrics would be here more suitable to evaluate our models, especially taking in consideration an uneven class distribution of our data. Average recall of one-class classifiers was approximately 89%, which means amongst 384 PD subjects from the previous example, 342 of them would be detected with PD. Given average precision of 91%, we determined that 311 out of 342 detected subjects truly had PD. In comparison with that, multi-class classifiers with 86% recall would detect 330 patients with PD. With average precision 97,5%, it would mean 322 out of 330 patients that were diagnosed with PD actually had the disease. In summary, one-class model would detect precisely 311 out of 384 subjects and for multi-class model it would be 322 out of 384 subjects.

Chosen significant features could ultimately help to focus on certain tests during the examination of patients. The ANOVA1 and Kruskal-Wallis tests showed the statistical significance of these features lower than 5% making them valid. The credibility of our results could have been improved with more features, more cross-validation repetitions (with better computer performance) and more patients (bigger data scale would provide more data for the prediction models to learn from). Especially in case of SVM classifier, more samples would serve potentially as support vectors. Nevertheless, the limitation of cross-validation could arise. Suppose a model is developed to predict risk for being diagnosed with dysarthria within the next year. If the model is trained using data from a study which contain only a specific population group (e.g. old people or women) but is then applied to the general population, the cross-validation results from the training set could vary considerably from the actual predictive performance.

It is important to remark that in one-class classifiers, the ability to learn the true characteristics of the data set in presence of noise or errors in the feature values is particularly important. Moreover, the number of parameters determined by users should be

minimized. We should take into consideration the possibilities of combining several one-class classifiers. It is well known that combining the results of conventional classifiers can significantly improve performance in conventional classification problems. Due to the different nature of one-class classifiers, it will be investigated how far these characteristics are preserved in the combination of one-class classifiers. Also, the computational and storage requirements must be considered, as there are limiting factors of usage for some methods. The feature selection helped us to simplify the computation and as well the whole system by allowing one-class classifiers to be trained on the represented target class. Without that, the presence of noisy samples could impact classifier performance as well as the problem with overfitting and the curse of dimensionality (Devijver and Kittler **1982,** Pudil et al. **1994**).

For all the reasons stated above, one-class classifiers can be very useful in many biomedical, clinical, pathological, or biological applications. The result of one-class classifiers therefore, showed potential utilization in clinical practice.

Appendix A

Content of CD

/TEXT	directory contents include electronic version of thesis *Bachelor.pdf*
/HELP	includes documentation of disk content, each directories and implemented functions *Documentation.pdf*
/METHOD	contains main scripts
/METHOD/prtools	additional toolbox needed for one-class classifiers
/METHOD/dd_tools	additional toolbox for one-class classifier
/RESULT	contains results of one-class and multi-class classifiers and tables of best features, also includes scripts for tables

References

Kremer, B., Goldberg, P., Andrew, S. E., Theilmann, J., Telenius, H., Zeisler, J., ... & Bird, T. D. (1994). A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. *New England Journal of Medicine*, *330*(20), 1401-1406.

Reyes, J. A., & Gilbert, D. (2007). Prediction of protein-protein interactions using one-class classification methods and integrating diverse biological data. *Journal of Integrative Bioinformatics*, *4*(3), 208-223.

Dreiseitl, S., Osl, M., Scheibböck, C., & Binder, M. (2010). Outlier detection with one-class SVMs: an application to melanoma prognosis. In *AMIA Annual Symposium Proceedings* (Vol. 2010, p. 172). American Medical Informatics Association.

Cabral, G. G., & de Oliveira, A. L. I. (2014, October). One-class Classification for heart disease diagnosis. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on* (pp. 2551-2556). Piscatavay, New Jersey: Institute of Electrical and Electronics Engineers.

Irigoien, I., Sierra, B., & Arenas, C. (2014). Towards application of one-class classification methods to medical data. *The Scientific World Journal*, *2014(1-7)*.

Wang, L., Xu, G., Guo, L., Liu, X., & Yang, S. (2007, August). 3D reconstruction of head mri based on one class support vector machine with immune algorithm. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (pp. 6015-6018). Piscatavay, New Jersey: Institute of Electrical and Electronics Engineers.

Zhou, J., Chan, K. L., Chong, V. F. H., & Krishnan, S. M. (2006, January). Extraction of brain tumor from MR images using one-class support vector machine. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (pp. 6411-6414). Piscatavay, New Jersey: Institute of Electrical and Electronics Engineers.

Song, X., & Wyrwicz, A. M. (2009). Unsupervised spatiotemporal fMRI data analysis using support vector machines. *NeuroImage*, *47*(1), 204-212.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural computation*, *13*(7), 1443-1471.

Depeursinge, A., Iavindrasana, J., Hidki, A., Cohen, G., Geissbuhler, A., Platon, A., ... & Müller, H. (2010). Comparative performance analysis of state-of-the-art classification algorithms applied to lung tissue categorization. *Journal of digital imaging*, *23*(1), 18-30.

Boehm, O., Hardoon, D. R., & Manevitz, L. M. (2011). Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms. *International Journal of Machine Learning and Cybernetics*, *2*(3), 125.

Devijver, P. A., & Kittler, J. (1982). *Pattern recognition: A statistical approach*. Prentice hall International, London.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, *15*(11), 1119-1125.

Cohen, G., Sax, H., & Geissbuhler, A. (2008). Novelty detection using one-class Parzen density estimator. An application to surveillance of nosocomial infections. *Studies in health technology and informatics*, *136*, 21.

Hayden MR (1981) Huntington's chorea. New York, *Springer-Verlag*, pp. 59–92.

Wexler, N. S. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3498-3503.

Albin, R. L., Reiner, A., Anderson, K. D., Dure, L. S., Handelin, B., Balfour, R., ... & Young, A. B. (1992). Preferential loss of striato-external pallidal projection neurons in presymptomatic Huntington's disease. *Annals of neurology*, *31*(4), 425-430.

Albin, R. L. (1995). Selective neurodegeneration in Huntington's disease.

Ramig, L. A. (1986). Acoustic analyses of phonation in patients with Huntington's disease: preliminary report. *Annals of Otology, Rhinology & Laryngology*, *95*(3), 288-293.

García, M. J. V., Cobeta, I., Martín, G., Alonso-Navarro, H., & Jimenez-Jimenez, F. J. (2011). Acoustic analysis of voice in Huntington's disease patients. *Journal of Voice*, *25*(2), 208-217.

Zwirner, P., Murry, T., & Woodson, G. E. (1991). Phonatory function of neurologically impaired patients. *Journal of communication disorders*, *24*(4), 287-300.

Vogel, A. P., Shirbin, C., Churchyard, A. J., & Stout, J. C. (2012). Speech acoustic markers of early stage and prodromal Huntington's disease: a marker of disease onset?. *Neuropsychologia*, *50*(14), 3273-3278.

Kaploun, L. R., Saxman, J. H., Wasserman, P., & Marder, K. (2011). Acoustic analysis of voice and speech characteristics in presymptomatic gene carriers of Huntington's disease: biomarkers for preclinical sign onset?. *Journal of Medical Speech-Language Pathology*, *19*(2), 49-65.

Tabrizi, S. J., Scahill, R. I., Durr, A., Roos, R. A., Leavitt, B. R., Jones, R., ... & Kennard, C. (2011). Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. *The Lancet Neurology*, *10*(1), 31-42.

Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., & Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Scientific reports*, *7*(1), 12.

Vonsattel, J. P., Myers, R. H., Stevens, T. J., Ferrante, R. J., Bird, E. D., & Richardson, E. P. (1985). Neuropathological classification of Huntington's disease. *Journal of Neuropathology & Experimental Neurology*, *44*(6), 559-577.

Rosenblatt, A., Liang, K. Y., Zhou, H., Abbott, M. H., Gourley, L. M., Margolis, R. L., ... & Ross, C. A. (2006). The association of CAG repeat length with clinical progression in Huntington disease. *Neurology*, *66*(7), 1016-1020.

Gramacki, A. (2018). *Nonparametric Kernel Density Estimation and Its Computational Aspects*. New York, USA: Springer International Publishing.

Walker, F. O. (2007). Huntington's disease. *The Lancet*, *369*(9557), 218-228.

Ross CA, Margolis RL, Rosenblatt A, Ranen NG, Becher MW, Aylward E. (1997). Huntington disease and the related disorder, dentatorubral-pallidoluysian atrophy (DRPLA). *Medicine (Baltimore), 76, 305-338.*

Huntington, H. S. G. (1996). Unified Huntington's disease rating scale: reliability and consistency. *Movement Disorders*, *11*, 136-142.

Hartelius, L., Carlstedt, A., Ytterberg, M., Lillvik, M., & Laakso, K. (2003). Speech disorders in mild and moderate Huntington disease: Results of dysarthria assessments of 19 individuals. *Journal of Medical Speech-Language Pathology*, *11*(1), 1-15.

Saldert, C., Fors, A., Ströberg, S., & Hartelius, L. (2010). Comprehension of complex discourse in different stages of Huntington's disease. *International journal of language & communication disorders*, *45*(6), 656-669.

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341-347.

Rosenblatt, A., Abbott, M. H., Gourley, L. M., Troncoso, J. C., Margolis, R. L., Brandt, J., & Ross, C. A. (2003). Predictors of neuropathological severity in 100 patients with Huntington's disease. *Annals of neurology*, *54*(4), 488-493.

Kieburtz, K., Penney, J. B., Corno, P., Ranen, N., Shoulson, I., Feigin, A., ... & Goldstein, J. (2001). Unified Huntington's disease rating scale: reliability and consistency. *Neurology*, *11*(2), 136-142.

Marder, K., Zhao, H., Myers, R. H., Cudkowicz, M., Kayson, E., Kieburtz, K., ... & Shoulson, I. (2000). Rate of functional decline in Huntington's disease. *Neurology*, *54*(2), 452-452.

Duffy, J. R. (2013). *Motor Speech Disorders-E-Book: Substrates, Differential Diagnosis, and Management*. Elsevier Health Sciences.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12*(2), 246-269.

Hornykiewicz, O. (2008). Basic research on dopamine in Parkinson's disease and the discovery of the nigrostriatal dopamine pathway: the view of an eyewitness. *Neurodegenerative Diseases*, *5*(3-4), 114-117.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, *79*(4), 368-376.

Rodriguez-Oroz, M. C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., & Obeso, J. A. (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *The Lancet Neurology*, *8*(12), 1128-1139.

Inzelberg, R., Schechtman, E., & Paleacu, D. (2002). Onset age of Parkinson disease. *American journal of medical genetics*, *111*(4), 459-460.

Golbe, L. I. (1991). Young-onset Parkinson's disease A clinical review. *Neurology*, *41*(2 Part 1), 168-168.

Dorsey, E., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kieburtz, K., ... & Tanner, C. M. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, *68*(5), 384-386.

Rajput, A. H. (1992). Frequency and cause of Parkinson's disease. *Canadian Journal of Neurological Sciences/Journal Canadien des Sciences Neurologiques*, *19*(S1), 103-107.

Willis, A. W., Schootman, M., Kung, N., Evanoff, B. A., Perlmutter, J. S., & Racette, B. A. (2012). Predictors of survival in patients with Parkinson disease. *Archives of neurology*, *69*(5), 601-607.

Becker, G., Müller, A., Braune, S., Büttner, T., Benecke, R., Greulich, W., ... & Thümler, R. (2002). Early diagnosis of Parkinson's disease. *Journal of Neurology*, *249*(3), iii40-iii48.

National Parkinson Foundation, I. N. C. (2013) Parkinson's Disease, www.parkinson.org.

Fearnley, J. M., & Lees, A. J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain*, *114*(5), 2283-2301.

Skodda, S., Flasskamp, A., & Schlegel, U. (2010). Instability of syllable repetition as a model for impaired motor processing: is Parkinson's disease a "rhythm disorder"?. *Journal of neural transmission*, *117*(5), 605-612.

Skodda, S., Grönheit, W., & Schlegel, U. (2012). Instability of syllable repetition in progressive supranuclear palsy. *Journal of Neural Transmission*, *119*(4), 457-462.

Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease. *Movement Disorders*, *23*(7), 985-992.

Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46). New York: IBM.

Skodda, S. (2015). Steadiness of syllable repetition in early motor stages of Parkinson's disease. *Biomedical Signal Processing and Control*, *17*, 55-59.

Aly, M. (2005). Survey on multiclass classification methods. *Neural Netw*, *19*, 1-9.

Postuma, R. B., Lang, A. E., Gagnon, J. F., Pelletier, A., & Montplaisir, J. Y. (2012). How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain*, *135*(6), 1860-1870.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural neurology, 11*(3), 131-137.

Tax, D. M., & Duin, R. P. (1999). Support vector domain description. *Pattern recognition letters*, *20*(11-13), 1191-1199.

Rusz, J., Klempíř, J., Tykalová, T., Baborová, E., Čmejla, R., Růžička, E., & Roth, J. (2014). Characteristics and occurrence of speech impairment in Huntington's disease: possible influence of antipsychotic medication. *Journal of Neural Transmission*, *121*(12), 1529-1539.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The journal of the Acoustical Society of America*, *129*(1), 350-367.

Critchley, E. M. (1981). Speech disorders of Parkinsonism: a review. *Journal of Neurology, Neurosurgery & Psychiatry, 44*(9), 751-758.

Rusz, J., Čmejla, R., Růžičková, H., Klempíř, J., Majerová, V., Picmausová, J., ... & Růžička, E. (2011). Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. *Movement Disorders*, *26*(10), 1951-1952.

Hunker, C. J., Abbs, J. H., & Barlow, S. M. (1982). The relationship between parkinsonian rigidity and hypokinesia in the orofacial system: a quantitative analysis. *Neurology*, *32*(7), 749-749.

Hammen, V. L., & Yorkston, K. M. (1996). Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria. *Journal of communication disorders, 29*(6), 429-445.

Goberman, A. M., & McMillan, J. (2005). Relative speech timing in Parkinson's disease. *Commun Science Disord*, *32*, 22-29.

Rosen, K. M., Kent, R. D., Delaney, A. L., & Duffy, J. R. (2006). Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *Journal of Speech, Language, and Hearing Research*, *49*(2), 395-411.

Kent, R. D., Kent, J. F., Weismer, G., & Duffy, J. R. (2000). What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics, 28*(3), 273-302.

Canter, G. J. (1965). Speech characteristics of patients with Parkinson's disease: III. Articulation, diadochokinesis, and over-all speech adequacy. *Journal of Speech and Hearing Disorders, 30*(3), 217-224.

Rudzicz, F. (2011). Articulatory knowledge in the recognition of dysarthric speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(4), 947-960.

Kalia, L. V., Kalia, S. K., & Lang, A. E. (2015). Disease-modifying strategies for Parkinson's disease. *Movement Disorders*, *30*(11), 1442-1450.

Baumgartner, C. A., Sapir, S., & Ramig, L. O. (2001). Voice quality changes following phonatory-respiratory effort treatment (LSVT®) versus respiratory effort treatment for individuals with Parkinson disease. *Journal of Voice, 15*(1), 105-114.

Chenausky, K., MacAuslan, J., & Goldhor, R. (2011). Acoustic analysis of PD speech. *Parkinson's Disease*, 2011(1), 1-13.

Fischer, E., & Goberman, A. M. (2010). Voice onset time in Parkinson disease. *Journal of Communication Disorders, 43*(1), 21-34.

Goberman, A. M., & Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *Journal of Voice, 22*(2), 178-191.

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism onset, progression, and mortality. *Neurology, 17*(5), 427-427.

Stebbins, G. T., & Goetz, C. G. (1998). Factor structure of the Unified Parkinson's Disease Rating Scale: motor examination section. *Movement Disorders, 13*(4), 633-636.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Clifton Park, NY: Cengage Learning.

Novák, A., & Vokřál, J. (1995). Acoustic parameters for the evaluation of voice of future voice professionals. *Folia phoniatrica et logopaedica*, *47*(5), 279-285.

Rusz, J., Hlavnička, J., Tykalová, T., Bušková, J., Ulmanová, O., Růžička, E., & Šonka, K. (2016). Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. *Sleep medicine*, *19*, 141-147.

Tax, D. (2001). One-class classification. *PhD thesis,* Delft University of Technology.

Apraxia Kids. (2018). Disordered Prosody and Articulation in Children with Childhood Apraxia of Speech. [online] Available at: https://www.apraxia-kids.org/library/disordered-prosody-and-articulation-in-children-with-childhood-apraxia-of-speech/ [Accessed 12 Jul. 2018].

Goberman, A. M., & Coelho, C. (2002). Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy. *NeuroRehabilitation*, *17*(3), 237-246.

Li, T., Zhang, C., & Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, *20*(15), 2429-2437.

Dixon, S. J., & Brereton, R. G. (2009). Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector

machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems*, *95*(1), 1-17.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.

Srivastava, S., Gupta, M. R., & Frigyik, B. A. (2007). Bayesian quadratic discriminant analysis. *Journal of Machine Learning Research*, *8*(Jun), 1277-1305.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York, NY, USA: Springer series in statistics.

Vapnik, V. (2013). *The nature of statistical learning theory*. Berlin, Germany: Springer science & business media.

Tarassenko, L., Hayton, P., Cerneaz, N., & Brady, M. (1995). Novelty detection for the identification of masses in mammograms. IET Conference Proceedings, Cambridge, UK, 1995, p. 442-447: Institution of Engineering and Technology Digital Library.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. New York, USA: John Wiley & Sons.

Mazhelis, O. (2006). One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. *South African Computer Journal*, *2006*(36), 29-48.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464-1480.

Wikipedia contributors. (2018). Support vector machine. [online] Available at: https://en.wikipedia.org/w/index.php?title=Support_vector_machine&oldid=872469419 [Accessed 14 December 2018].

Ross, C. A., & Tabrizi, S. J. (2011). Huntington's disease: from molecular pathogenesis to clinical treatment. *The Lancet Neurology*, *10*(1), 83-98.

Costa, M., & Moura, L. (1995). Automatic assessment of scintmammographic images using a novelty filter. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* (p. 537). American Medical Informatics Association.

Roberts, S., Tarassenko, L., Pardey, J., & Siegwart, D. (1994, August). A validation index for artificial neural networks. In *Proceedings of Int. Conference on Neural Networks and Expert Systems in Medicine and Healthcare* (pp. 23-30). Stevenage, United Kingdom: Institution of Electrical Engineers.

Rajput, A. H., & Birdi, S. (1997). Epidemiology of Parkinson's disease. *Parkinsonism & related disorders*, *3*(4), 175-186.