

Speech Features for Telemonitoring of Parkinson's Disease Symptoms

Hamideh Ramezani, *Student Member, IEEE*, Hossein Khaki, *Student Member, IEEE*,
Engin Erzin, *Senior Member, IEEE* and Ozgur B. Akan, *Fellow, IEEE*

Abstract—The aim of this paper is tracking Parkinson's disease (PD) progression based on its symptoms on vocal system using Unified Parkinsons Disease Rating Scale (UPDRS). We utilize a standard speech signal feature set, which contains 6373 static features as functionals of low-level descriptor (LLD) contours, and select the most informative ones using the maximal relevance and minimal redundancy based on correlations ($mRMR_C$) criteria. Then, we evaluate performance of Gaussian mixture regression (GMR) and support vector regression (SVR) on estimating the third subscale of UPDRS, i.e., UPDRS: motor subscale (UPDRS-III). Among the most informative features, a list of features are selected after redundancy reduction. The selected features depict that LLDs providing information about spectrum flatness, spectral distribution of energy, and hoarseness of voice are the most important ones for estimating UPDRS-III. Moreover, the most informative statistical functions are related to range, maximum, minimum and standard deviation of LLDs, which is an evidence of the muscle weakness due to the PD. Furthermore, GMR outperforms SVR on compact feature sets while the performance of SVR improves by increasing number of features.

I. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegeneration disorders resulting from death of dopamine-generating cells. Dopamine is information carrier for communication among nerve cells responsible for relaying messages that control body movement [1]. Lower amount of available dopamine for release decreases the achievable rate of this communication [2]. Hence, this disease mainly affects the motor system with reducing range of movements and its symptoms include tremor, rigidity and loss of muscle control. Along with movement disorders in other parts of the body, PD also affects the muscles in the face, mouth and throat that are used in vocal system. Main PD symptoms present in speech contain weak, hoarse, nasal or monotonous voice, imprecise articulation, slow or fast speech, difficulty starting speech, impaired stress or rhythm, stuttering and tremor [3]. Hence, speech signal can be used for PD diagnosis and tracking the progression of this disease.

Unified PD rating scale: motor subscale (UPDRS-III) is a metric of PD progress, which reflects the presence and severity of PD symptoms [4]. Finding a statistical mapping

H. Ramezani and O. B. Akan are with Next-generation and Wireless Communications Laboratory, Department of Electrical and Electronics Engineering, Koç University, Istanbul 34450, Turkey, (e-mails: {hramezani13 and akan}@ku.edu.tr).

H. Khaki and E. Erzin are with Multimedia, Vision and Graphics Laboratory, Department of Electrical and Electronics Engineering, Koç University, Istanbul 34450, Turkey, (e-mails: {hkhaki13 and eerzin}@ku.edu.tr).

This work was supported in part by ERC project MINERVA (ERC-2013-CoG #616922), EU project CIRCLE (EU-H2020-FET-Open #665564), and TÜBİTAK graduate scholarship program (BİDEB-2215).

between speech properties and UPDRS-III is shown to be feasible to some degree. Early studies on this problem used sustained vowel recordings to extract a set of dysphonia measurements [5], [6], where phonation and articulation related features are mainly used. However, as it is shown in [7], [8], a reading task is more discriminative for automatically assessing the severity of PD since it contains prosody variations. Hence, the released dataset by [8], which contains 42 speech tasks with different durations for each PD patient, provides appropriate context for PD telemonitoring task. This dataset is analyzed in different studies such as [9], [10], where heuristic regression methods such as random forest and deep neural network are applied for estimating UPDRS-III. However, the most informative features and discriminative statistical functions used to summarize them in estimating severity of PD are not studied in the literature.

In this paper, we use the standard feature set provided by recent INTERSPEECH Paralinguistic Challenges [11], [12] and analyze the most informative features for estimating UPDRS-III based on the linear correlation metric. Moreover, we explore the relationship between these features and PD symptoms. For these aims, the maximal relevance and minimal redundancy based on correlations ($mRMR_C$) is utilized as the feature selection mechanism [13]. Then, support vector regression (SVR) and Gaussian mixture regression (GMR) methods are used for mapping the selected features to the UPDRS-III. Based on experimental results, we attain more accurate estimation of UPDRS-III with a less complex system by using the proposed feature reduction mechanism.

The remainder of this paper is organized as follows. In Section II, we present the database and standard feature set used in this study. Then, we explain the feature selection approach, regression mechanisms and the accuracy metric in Section III. We explore the most informative features for PD telemonitoring and report the performance of our method in Section IV. Finally, we conclude the paper in Section V.

II. DATABASE AND FEATURE SET

Recordings of the training and test sets are taken from [8], which includes a total of 50 patients with Parkinsons disease (25 female, 25 male), where 35 of the patients are included in the training set, and the remaining 15 comprise the test set. Each speaker performed a total of 42 speech tasks including 24 isolated words, 10 sentences, one reading text, one monologue, and the rapid repetition of the syllables /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/. The neurological state of the patients was evaluated by an expert neurologist according

TABLE I
THE CATEGORIZATION OF LLDs.

Group	Low-Level Descriptors
A	Loudness, modulated loudness, MFCC coefficients 1-14, RASTA filtered auditory band levels (Rfilt) 1-26, Energy of frequency bands 250-650 Hz and 1-4 kHz, root mean square (RMS) energy, zero-crossing rate (ZCR), spectral skewness, spectral variance, spectral sharpness, spectral harmonicity, spectral entropy, spectral centroid, spectral flux, spectral kurtosis, spectral slope, spectral roll-off point (ROP) 0.25, 0.5, 0.75, 0.9.
B	Fundamental frequency (F0), voicing probability, Jitter, delta Jitter, Shimmer, logarithmic harmonics to noise ratio (HNR)
C	Ratio of nonzero values and segment length

to the UPDRS-III metric, whose values span over 5 to 92, where larger values represent severe motor impairment.

To extract the most relevant features of speech to PD symptoms, we use the provided feature set in the INTER-SPEECH'15 Paralinguistic Challenge, which contains 6373 static features as functionals of low-level descriptor (LLD) contours [12]. These features are derived from three groups of LLDs as shown in Table I, where group A contains energy, spectral, and cepstral LLDs, group B LLDs are related to the source/excitation signal and LLDs in group C are defined based on the voiced/unvoiced segments. In addition to these LLDs, their first order delta regression coefficients are also used to represent the dynamic LLDs. The 6373 static features are defined as a variety of statistical functionals, which are applied on the LLDs. Some examples of these functions are mean, standard deviation, minimum, maximum, range, skewness, kurtosis and the lower and upper quartiles [14].

III. METHODOLOGY

In this section, we first explain our feature selection method, i.e., $mRMR_C$. Then, we represent the regression mechanisms used for mapping the selected features to the UPDRS-III label. Finally, we describe metrics that we use for evaluating the performance of our UPDRS-III estimator.

A. Feature Selection Method

Our objective is to select features with the highest relevancy to UPDRS-III and minimal redundancy, where relevancy can be characterized in terms of correlation or mutual information. Estimation of the mutual information requires the distribution of features and label, which in turn needs large data. Hence, we utilize Pearson correlation to measure the linear relevancy. Based on [13], we have two criteria, (i) maximizing the correlation between feature, x_i , and UPDRS-III label, y , and (ii) minimizing the average cross-correlation between selected features. Thus, we select features one by one to maximize following objective function,

$$mRMR_C(i) = |corr(x_i, y)| - \frac{1}{M-1} \sum_{\substack{x_j \in S_m \\ j \neq i}} |corr(x_i, x_j)|,$$

where S_m denotes set of selected features and $|corr(x_i, y)|$ is the absolute value of the Pearson correlation between x_i

and y . This incremental search method leads to the optimal feature set based on the aforementioned criteria [13]. We define S_{m, N_F} as the final set of N_F selected features.

B. Regression Methods

To estimate the UPDRS-III values based on the features in S_{m, N_F} , we apply linear regression as $\hat{y} = \bar{\beta}^T \bar{x} + b$, where \bar{x} is the vector containing features in S_{m, N_F} , $\bar{\beta}$ contains the weight of each feature and b is the offset of regression. We utilize two different regression methods, i.e., SVR and GMR, which are described in the following.

1) *Support Vector Regression*: The objective of SVR is to have the linear regression as flat as possible. Hence, it minimizes the norm value ($\bar{\beta}^T \bar{\beta}$) while all residuals have a value less than a given threshold, ϵ , i.e.,

$$\begin{aligned} \min J(\bar{\beta}) &= 0.5 \bar{\beta}^T \bar{\beta} \\ \text{s.t. } |y_n - \bar{\beta}^T \bar{x}_n - b| &\leq \epsilon \quad \forall n, \end{aligned}$$

where \bar{x}_n and y_n are the n -th observation of features and label in training set, respectively. It is possible that no such function \hat{y} exists to satisfy these constraints for all points, hence, slack variables ξ_n and ξ_n^* are defined for each point and the objective function is changed to

$$\begin{aligned} \min J(\bar{\beta}) &= 0.5 \bar{\beta}^T \bar{\beta} + C \sum_{n=1}^{N_F} (\xi_n + \xi_n^*) \\ \text{s.t. } \bar{\beta}^T \bar{x}_n + b - y_n &\leq \epsilon + \xi_n^* \quad \forall n, \\ y_n - \bar{\beta}^T \bar{x}_n - b &\leq \epsilon + \xi_n \quad \forall n, \\ \xi_n, \xi_n^* &\geq 0 \quad \forall n, \end{aligned}$$

where the constant C is a positive parameter that controls the tolerable penalty by observations that lie outside the margin defined by ϵ [15].

2) *Gaussian Mixture Regression*: Gaussian mixture model (GMM) is a classic parametric model used to represent multivariate probability distributions. It states that any general distribution of $\bar{\mathcal{X}}$ can be approximated by sum of weighted Gaussian distributions as $P(\bar{\mathcal{X}}) = \sum_{l=1}^L w_l \mathcal{N}(\bar{\mathcal{X}}; \bar{\mu}_l, \Sigma_l)$, where L is the total number of Gaussian mixtures, $\mathcal{N}(\bar{\mathcal{X}}; \mu_l, \Sigma_l)$ represents the multivariate Gaussian distribution with mean $\bar{\mu}_l$ and covariance matrix Σ_l , and $w_l \geq 0$ is the weight of l -th mixture, which satisfies $\sum_{l=1}^L w_l = 1$.

GMR is a GMM-based probabilistic mapping approach, used in our work to estimate the UPDRS-III label based on selected features by $mRMR_C$. In GMR, feature set and label are assumed to be jointly Gaussian, then the mean square estimation of label within an isolated Gaussian mixture is defined as, $\hat{y}_l = \mu_{y,l} + \Sigma_{y\bar{x},l} (\Sigma_{\bar{x}\bar{x},l})^{-1} (\bar{x} - \mu_{\bar{x},l})$, where $\Sigma_{y\bar{x},l}$ is the cross-covariance matrix between the features in S_{m, N_F} and UPDRS-III label for the l -th mixture. Then, the GMR based estimation of the label over all Gaussian mixtures can be achieved as,

$$\hat{y} = \sum_{l=1}^L P(\gamma_l | \bar{x}) \hat{y}_l,$$

where γ_l is defined as the l -th Gaussian mixture and $P(\gamma_l | \bar{x})$

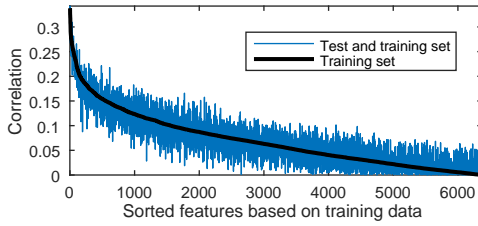


Fig. 1. The correlation among features and UPDRS-III for training and test sets, where features are sorted based on their correlation in training set.

is the probability of the l -th Gaussian mixture given the observation \bar{x} and is derived as follows.

$$P(\gamma_l|\bar{x}) = \frac{\mathcal{N}(\bar{x}; \bar{\mu}_{\bar{x},l}, \Sigma_{\bar{x}\bar{x},l})}{\sum_{i=1}^L \mathcal{N}(\bar{x}; \bar{\mu}_{\bar{x},i}, \Sigma_{\bar{x}\bar{x},i})}$$

C. Accuracy Metric

We define the correlation between estimated values and UPDRS-III labels as the accuracy metric of the regression. For this purpose, we use two different correlation metrics as,

- **Pearson correlation**, i.e., $Corr(\hat{y}, y)$, and
- **Spearman correlation**, which is defined as the Pearson correlation among rank of \hat{y}_n and y_n , where the rank of y_n is its index after sorting all observations of label.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we investigate the importance of LLDs and applied statistical functions for PD telemonitoring. Then, we report the performance of the regression methods.

A. Feature Set Analysis

To investigate the most relevant features to PD, we study the absolute value of the correlation among each feature and level of PD depicted by UPDRS-III metric, where higher correlations show stronger linear dependency. In Fig. 1, all features are sorted based on their correlation with the UPDRS-III in training set. Same pattern is seen for the correlations of these sorted features in test set as depicted in Fig. 1. Hence, the correlation metric is consistent for training and test set. Moreover, the correlation decreases dramatically among more informative features while after knee point the reduction rate of correlation between label and features decreases. We use the knee point, whose position is 743 in sorted features based on the training set, to divide features to two categories as more and less informative.

The summation of the correlation between UPDRS-III and all features defined based on an individual LLD depicts the importance of that LLD in estimating UPDR-III. This summation is calculated over the more informative features defined based on knee point, i.e., 743 features, and shown in Fig. 2(a). The most relevant LLDs to PD can be listed as:

- RASTA filtered auditory band levels 5-11 and 13-18,
- MFCC coefficients 2, 4 and 12, and
- Loudness and modulated loudness.

The auditory bands depict the energy perception in human auditory system for each frequency band. After applying

RASTA filter to these auditory bands, near stationary and high frequency noises are attenuated [14], hence, the impacts of PD symptoms on low and high auditory bands are filtered. Moreover, loudness and modulated loudness LLDs are affected by PD, since patient's voice volume reduces as a result of decreased muscle-movements range.

However the LLDs defined in Group C, i.e., ratio of nonzero values and segment length, can be used as a measure of speed of the speaker, which is affected by PD. These features cannot be used on our dataset to extract meaningful information for estimating level of PD since the duration of majority of recordings is short (82% less than 4 sec), where the patients are asked to repeat a given word or syllabus.

We also analyze the importance of features defined by applying statistical functions on the delta regression of LLDs. The summation of correlations for features defined by directly applying statistical functions on LLDs is 91.7 while this summation for features defined based on the delta regression of LLDs is 40.8. Hence, the temporal variation of LLDs is less correlated to PD symptoms compared to LLDs.

To extract the most important statistical functions of LLDs for estimating level of PD, we calculate sum of correlations for 743 features that have highest correlation with UPDRS-III for different statistical functions. As shown in Fig. 2(b), the most informative statistical functions are related to the range, maximum, minimum and standard deviation of the LLDs since PD makes the muscles weak by reducing their range of movement. On the other hand, the least relevant statistical functions to PD are position of the maximum/minimum value relative to the input length, linear and quadratic regression coefficients, temporal centroid, kurtosis, LP analysis gain and coefficients, range of valley and peak amplitude relative to the arithmetic mean, ratio of nonzero values, and all functions applied on segment length LLD.

Since cross-correlation exists among features, selecting ones that have highest correlation with the UPDRS-III does not lead to the best feature selection mechanism. Cross correlation of 10 features that have highest correlation with UPDRS-III metric for training set is shown in Fig. 3. As it is depicted in Fig. 3, first and second features are highly correlated since they are representing same statistic, i.e., flatness, on same type of LLD, i.e., RASTA filtered auditory band levels, for two adjacent bands that are not uncorrelated, hence, the improvement achieved in estimating UPDRS-III by selecting both of them is not significant compared to selecting one of them. This redundancy between selected features is minimized by using $mRMR_C$ feature selection technique. In Fig. 4, the $mRMR_C$ values achieved by incremental search for feature selection are shown, where the features are sorted based on their $mRMR_C$ value at training set. As it is depicted in Fig. 4, the $mRMR_C$ values achieved by incremental search over both training and test set follows same pattern as the results for training set. Hence, utilizing selected features based on the training set will lead to proper feature set for estimating UPDRS-III in test set.

Finally, we study the structure of selected features by $mRMR_C$. For this objective, we compare the selected

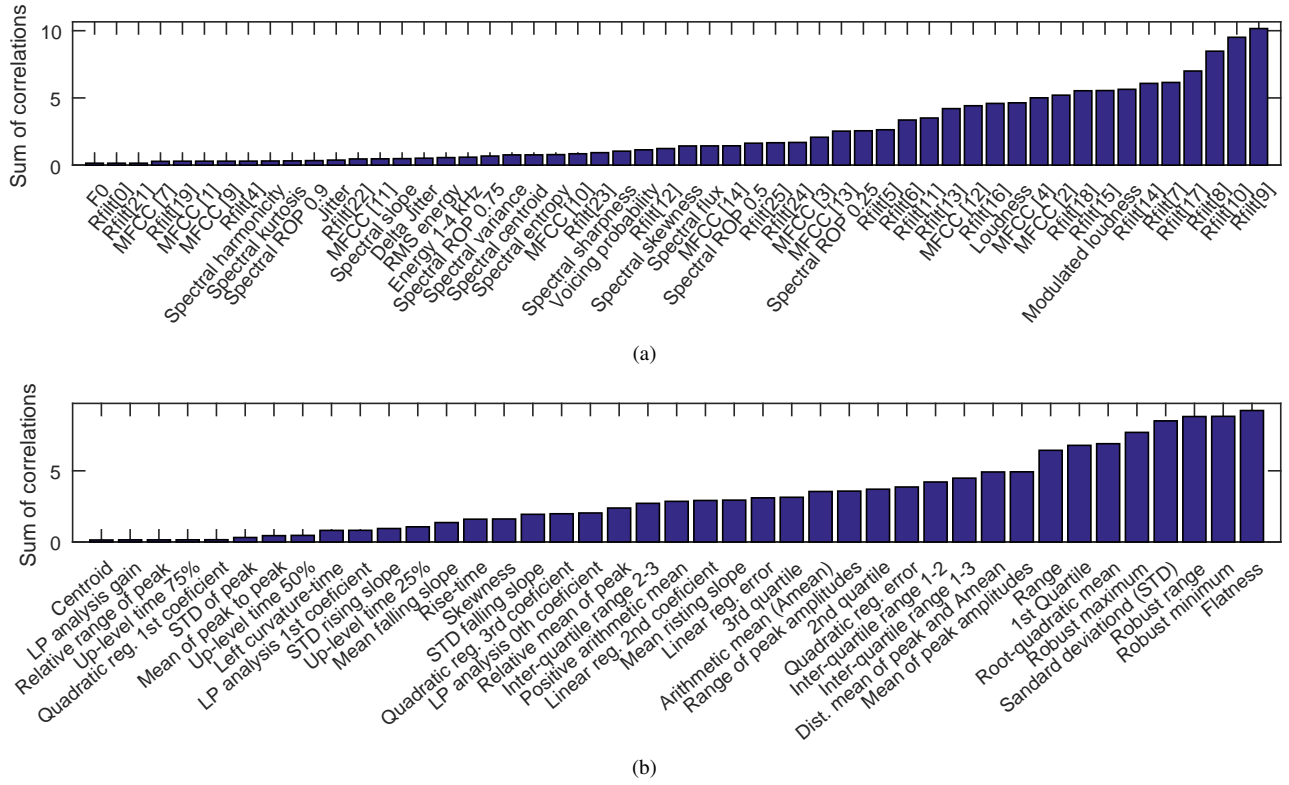


Fig. 2. Summation of correlation between UPDRS-III metric and features in $S_{m,743}$ defined by applying (a) different statistical functions on an individual LLD and (b) a statistical function on different LLDs.

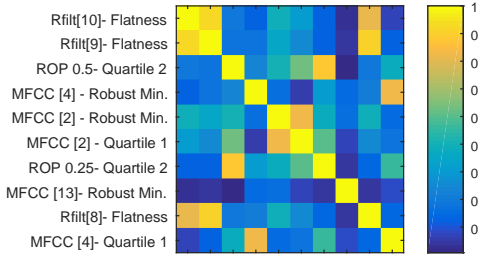


Fig. 3. The cross-correlation among 10 features with highest correlation with UPDRS-III.

features based on $mRMR_C$ by features that have highest correlation with label. The first 50 features selected by this method are defined based on the following LLDs:

- MFCC coefficients 1-4,7, and 12-14,
- RASTA filtered auditory band levels 5,8,9,10,14,
- Spectral ROP 0.25, 0.5 and 0.75,
- Spectral slope, variance and sharpness, and
- Jitter.

Comparing these LLDs and ones that have maximum correlation with UPDRS-III shown in Fig. 2(a), loudness and modulated loudness are not selected here. The main reason is that they are defined based on MFCC coefficients and RASTA filtered auditory spectrum [14], hence they are highly correlated with other selected LLDs. On the other hand, the spectral ROP 0.25, 0.5, 0.75 are selected since the volume of voice decreases as a result of PD.

Moreover, MFCC coefficients are dominant in $S_{m,50}$, while the RASTA filtered auditory bands have higher correlation with UPDRS-III metric as shown in Fig. 2(a). The main reason is existence of lower cross-correlation among MFCC coefficients compared to RASTA filtered auditory bands [16].

Low and high MFCC coefficients selected in $S_{m,50}$ depict the flatness in spectrum and amount of variation, respectively. Hence, the reason of their importance is spectrum frequency changes that occurs to the speech under PD, i.e., symptoms such as hoarse, nasal or monotonous voice, imprecise articulation, impaired stress or rhythm and stuttering.

Although Jitter, spectral slope, variance and sharpness are not among LLDs that have highest correlation with UPDRS-III, they are selected by $mRMR_C$, hence, they have lower correlation with other LLDs in $S_{m,50}$. Another aspect of Jitter is that it can be considered as an indicator of voice hoarseness [17], hence it can be used for PD telemonitoring.

Only 4 LLDs defined based on delta regression coefficients are in $S_{m,50}$. Hence, the information provided by delta regression coefficients of LLDs for estimating UPDRS-III is highly correlated with the information achieved by directly applying statistical function on LLDs.

B. UPDRS-III Estimation

We apply GMR and SVR for mapping S_{m,N_F} to UPDRS-III. We select the parameters of regression methods by maximizing the correlation metric over 4-fold cross validation (CV) of training set and adjust them for test set. We choose $L \in \{1, 2, \dots, 8\}$, $C \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$ and fixed

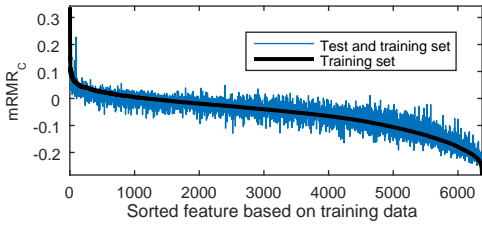


Fig. 4. The $mRMR_C$ values for features selected based on the incremental search method.

TABLE II

CORRELATION OF ESTIMATED UPDRS-III WITH ITS TRUE VALUE.

Method	N_F	Train CV		Train / Test	
		Pearson	Spearman	Pearson	Spearman
$mRMR_C+GMR$	18	0.5117	0.5192	0.4883	0.4418
$mRMR_C+SVR$	50	0.5189	0.5748	0.4793	0.4269
$mRMR_C+SVR$	1000	0.4514	0.5239	0.2719	0.52
Baseline	6373	0.3291	0.4342	0.3407	0.4936

$\epsilon = 1$. Moreover, we normalize the features to reach zero mean and unit variance variables.

The correlation between estimated UPDRS-III and its true value using GMR or SVR is tabulated in Table II. We are not able to apply GMR on large feature sets because of small sample size problem in covariance estimation. Hence, first, we search over 50 best features based on the $mRMR_C$ criteria and find the number of features and regression parameters that give best estimation results in training set. For GMR, 18 features are selected while for SVR the performance improves consistently by increasing number of selected features, i.e., N_F , until 50. Thus, we conclude that GMR leads to higher performances based on both Pearson and Spearman correlations with lower number of features in S_{m,N_F} .

In [12], SVR is applied over the set of all features, which is considered as baseline method in Table II. However, adding another feature to S_{m,N_F} will not improve $mRMR_C$ for number of selected features greater than 1000 as depicted in Fig. 4. By comparing results of baseline and applying SVR over $S_{m,1000}$, we observe that performance of estimation based on the Spearman correlation metric drops by increasing N_F . Moreover, increasing N_F from 50 to 1000 improves the performance based on the Spearman correlation while the Pearson correlation decreases. By looking at the estimated values, we observe very large rare numbers, which can be considered as outliers and Pearson correlation is sensitive to outliers, however, the Spearman correlation is robust.

V. CONCLUSION

In this paper, we studied the impacts of PD symptoms on features extracted from speech signal and utilized regression techniques to estimate level of severity of disease, which is measured by UPDRS-III. For this aim, we used the standard set of features provided in the literature, which contains 6373 static features as functionals of LLDs, and applied $mRMR_C$ feature reduction technique to select ones that have highest

correlation by UPDRS-III and lowest cross-correlation. Then, we utilized GMR and SVR for estimating UPDRS-III based on the selected features. Our feature analysis depicted that LLDs providing information about spectrum flatness, spectral distribution of energy, and hoarseness of voice are the most important ones for estimating UPDRS-III. Moreover, the most informative statistical functions are related to the range, maximum, minimum and standard deviation of LLDs since PD makes the muscles weak by reducing their range of movement. Furthermore, based on the experimental results, we concluded that (i) GMR outperforms SVR for lower number of selected features, and (ii) increasing number of selected features to 1000 and applying SVR improves the performance of estimating UPDRS-III based on the Spearman correlation metric while further increase in number of selected features decreases the accuracy.

REFERENCES

- [1] O. Hornykiewicz, "Biochemical aspects of parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.
- [2] H. Ramezani and O. B. Akan, "Rate region analysis of multi-terminal neuronal nanoscale communication channel," in *17th IEEE NANO Conf.* IEEE, 2017.
- [3] L. O. Ramig *et al.*, "Speech treatment for parkinsons disease," *Expert Review of Neurotherapeutics*, vol. 8, no. 2, pp. 297–309, 2008.
- [4] G. T. Stebbins and C. G. Goetz, "Factor structure of the unified parkinson's disease rating scale: motor examination section," *Movement Disorders*, vol. 13, no. 4, pp. 633–636, 1998.
- [5] A. Tsanas *et al.*, "Enhanced classical dysphonia measures and sparse regression for telemonitoring of parkinson's disease progression." in *ICASSP*, 2010, pp. 594–597.
- [6] —, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE Tran. on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.
- [7] A. Bayestehtashk *et al.*, "Fully automated assessment of the severity of parkinson's disease from speech," *Computer speech & language*, vol. 29, pp. 172–185, 2015.
- [8] J. R. Orozco-Arroyave *et al.*, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease." in *LREC*, 2014, pp. 342–347.
- [9] A. Zlotnik *et al.*, "Random forest-based prediction of parkinson's disease progression using acoustic, asr and intelligibility features," in *INTERSPEECH*, 2015.
- [10] S. Hahm and J. Wang, "Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data," in *INTERSPEECH*, 2015.
- [11] B. Schuller *et al.*, "The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language," in *INTERSPEECH*, 2016.
- [12] —, "The interspeech 2015 computational paralinguistics challenge: Nativeness, parkinsons & eating condition," in *INTERSPEECH*, 2015.
- [13] H. Peng *et al.*, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Tran. on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [14] F. Eyben, *Real-time speech and music classification by large audio feature space extraction*. Springer, 2016.
- [15] A. Smola and V. Vapnik, "Support vector regression machines," *Advances in neural info. processing systems*, vol. 9, pp. 155–161, 1997.
- [16] K. K. Paliwal, "Decorrelated and liftered filter-bank energies for robust speech recognition." in *Eurospeech*, vol. 99, 1999, pp. 85–88.
- [17] T. Jones *et al.*, "Objective assessment of hoarseness by measuring jitter," *Clinical Otolaryngology & Allied Sciences*, vol. 26, no. 1, pp. 29–32, 2001.