Computational diagnosis of Parkinson's Disease from speech based on regularization methods

1st Yolanda Campos-Roca
Departamento de Tecnologías de los Computadores y de las Comunicaciones
Universidad de Extremadura
Cáceres, Spain
ycampos@unex.es

2nd Fernando Calle-Alonso
Departamento de Matemáticas
Universidad de Extremadura
Cáceres, Spain
fcalonso@unex.es

3rd Carlos J. Pérez
Departamento de Matemáticas
Universidad de Extremadura
Cáceres, Spain
carper@unex.es

4th Lizbeth Naranjo
Facultad de Ciencias
Universidad Nacional Autónoma de México
México DF, Mexico
lizbethna@ciencias.unam.mx

***Abstract*—A computational tool to discriminate healthy people from people with Parkinson's Disease (PD) is proposed based on acoustic features extracted from sustained vowel recordings. Several approaches based on different feature sets and regularization methods (LASSO, Ridge, and Elastic net) are experimentally compared. The effectiveness of these methods has been evaluated on a dataset containing acoustic features of 40 healthy people and 40 patients with PD, who have been recruited at the Regional Association for Parkinson's Disease in Extremadura (Spain). The results show relevant differences when varying the initial feature set but high stability when changing the regularization approach. The three considered methods have achieved very promising classification accuracy rates via 10-fold cross-validation analysis, reaching 88.5%.**

***Index Terms*—Acoustic features, Elastic net, Least absolute shrinkage and selection operator, Nonlinear speech signal processing, Parkinson's disease, Regularized regression, Ridge.**

I. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease in the world and its incidence is expected to increase consistently as populations age [1]. PD is still incurable, but an early diagnosis allows to apply different therapies or lifestyle changes aimed at improving the patient's quality of life.

Early diagnosis of PD is a complex process and relies on subjective evaluations by neurologists. PD affects coordination of muscles, including those responsible of speech production, thus voice alterations are present in the majority of people affected by PD. Furthermore, speech impairment may be amongst the first symptoms of PD onset. Consequently, it has been suggested that acoustic analysis of speech may constitute an objective, low-cost and non-invasive technique to diagnose PD in an early stage [2].

This research has been supported by projects MTM2014-56949-C3-3-R and MTM2017-86875-C3-2-R (MINECO), and projects IB16054, GR18108 and GR18055 (Junta de Extremadura/European Regional Development Funds, EU).

In the last years, different machine learning approaches for automatic PD diagnosis from speech have been reported, and a large number of features have been extracted from acoustic recordings [2]–[5]. Some of these features are highly correlated, providing redundant information. Besides, due to the difficulty of recruiting patients, it is common to conduct experiments with a small number of subjects compared to the large number of features considered. This may cause computational problems such as multicollinearity or overfitting, which produce an overestimation and present limited generalization performance.

There are different techniques to deal with high dimensional learning. Filter feature selection methods are computationally fast, but they do not take into account the classifier. Wrapper methods use a predictive model to score feature sets. However, they are computationally expensive. Embedded methods have the advantage that they include the interaction with the classification model, while at the same time being less computationally intensive than wrapper methods [6].

Regularization or penalization methods are the most common type of embedded methods. These techniques use all the variables to create a model and then analyze it to deduce the importance of each variable. In this work three concrete techniques are applied to PD detection and their performance is compared: least absolute shrinkage and selection operator (LASSO) [7], Ridge [8] and Elastic net (Enet) [9].

Several classification experiments, based on different sets of linear and nonlinear features and the three aforementioned regularization approaches, are applied to a dataset containing data of 40 healthy people and 40 patients with PD, who have been recruited at the Regional Association for Parkinson's Disease in Extremadura (Spain). The results are comparatively analyzed in terms of different aspects: impact of the initial feature set on detection performance, influence of the concrete regularization method and complexity of the resulting models.

ISBN 978-90-827970-1-5

The remainder of this paper is organized as follows. Section II presents the main information on participants and speech recordings. Details about feature extraction are given in Section III. Section IV describes the regularization and classification framework. The experiments and obtained results are provided in Section V, together with a discussion. Finally, the paper ends up by some concluding remarks in Section VI.

II. Participants and speech recordings

Speech recordings were collected from 80 Spanish native speakers. Half of them were healthy: 22 men (55%) and 18 women (45%), and the other half were diagnosed with PD: 27 men (67.5%) and 13 women (32.5%). Their mean age ($\pm$standard deviation) was 66.38 ± 8.38 years for the control group and 69.58 ± 7.82 years for the subjects with PD. The people with PD participating in this study were members of the Regional Association for Parkinson's Disease in Extremadura (Spain). All participants provided written informed consent and the research protocol received approval from the Bioethical Committee of the University of Extremadura.

The participants were requested to sustain phonation of /a/ vowel for at least 5 seconds, attempting to maintain steady frequency and amplitude at a comfortable level. Phonation onsets were excluded from the recordings. The task was repeated until three successful recordings per subject were obtained for posterior averaging purpose.

All speech data were recorded in a quiet room with a low ambient noise level. The recording equipment was composed of a portable computer with an external sound card (TASCAM US322) and a headband cardiod microphone (AKG 520). The mouth-to-microphone distance was approximately 4 cm during all recordings. The voice signals were sampled at 44.1 kHz with 16-bit resolution and stored in WAV format by using Audacity software (release 2.0.5).

III. Feature extraction

In the signal pre-processing stage, the initial one-second segments were selected from the sustained vowel recordings. Next, the waveforms were normalized between -1 and 1.

The study is based on linear and nonlinear acoustic features. The considered features may be grouped according to the signal characteristics they are supposed to measure.

The first group of features includes 9 perturbation measures: four of them are pitch-perturbation features (jitter variants) and five of them amplitude-perturbation ones (shimmer variants). The second group is composed of two features related to Signal-to-Noise Ratio (SNR) measures: Harmonic-to-Noise Ratio (HNR) and Glottal-to-Noise Excitation (GNE). The rationale for these measures is that subjects with PD present incomplete vocal fold closure and this may lead to increased acoustic noise [10].

Also, features based on Mel Frequency Cepstral Coefficients (MFCC) are considered. A recent contribution that considers MFCC-based features extracted from sustained vowel recordings to discriminate between patients with PD and healthy controls is [11]. MFCCs are related to articulator placement, which may be affected in PD [12]. Thirteen MFCC parameters (0-12th order) are calculated for each frame. The 0th order one simply represents the average speech energy, and each higher-order MFCC represents increasingly finer spectral detail. Frames had a length of 30 ms, with a 50% overlap. This high number of frame-based parameters is reduced by calculating the mean values (related to average vocal tract configuration) and also the standard deviations of the frame-based parameters (that measure lack of steadiness in the vocal tract configuration during a sustained vowel production).

The last group of features is based on nonlinear speech processing. Previous studies have shown the limitations of the linear source-filter model of speech highlighting the need of including nonlinear features for effective detection of PD [13], [14] . The specific features considered are the following: Recurrent Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), Pitch Period Entropy (PPE) [13], Correlation dimension (D2), and three entropy variants (permutation entropy, fuzzy entropy and sample entropy).

PD seems to have a differential impact on phonation in men and women [15]. Therefore, gender has been used as an additional feature in the classification experiments.

For each acoustic feature, the three replications per individual were averaged, avoiding the usual practice of considering measurements within the same subject as independent. This provided a matrix of 80 rows (one per subject) and 45 columns (one per acoustic feature plus the gender). The complete list of acoustic measures and their notations considered in the present study is shown in Table I.

TABLE I
Acoustic features.

Perturbation and SNR features	**Symbol**
Jitter relative	jitr
Absolute jitter (%)	jita
Relative average perturbation (%)	RAP
Pitch perturbation quotient (%)	PPQ
Shimmer loc (%)	Shimloc
Shimmer in dB (%)	ShimdB
Amplitude perturbation quotient (%)	APQ3
Amplitude perturbation quotient (%)	APQ5
Amplitude perturbation quotient (%)	APQ11
Harmonics-to-noise ratio (%)	HNR
Glottal-to-noise excitation	GNE
Nonlinear features	**Symbol**
Recurrent period density entropy	RPDE
Detrended fluctuation analysis	DFA
Pitch period entropy	PPE
Correlation dimension	D2
Permutation entropy	PermutEn
Fuzzy entropy	FuzzyEn
Sample entropy	SampleEn
MFCC-based features	**Symbol**
Mean of $MFCC^i$	μ^i_{MFCC}
Standard deviation of $MFCC^i$	σ^i_{MFCC}

IV. Feature selection and regularization

Generalized linear models with convex penalties are used in this paper, including LASSO regression [7] and Ridge

ISBN 978-90-827970-1-5 © EURASIP 2018

regression [8]. A mixture of them, Enet regression, is also considered [9]. All of them exploit sparsity in the data matrix.

Although LASSO and Ridge have a common goal, their properties differ substantially. Both penalize the magnitude of features while minimizing the error between predicted and real observations. These regularization techniques vary depending on how they assign penalty to the coefficients. In the case of LASSO regression, it performs a regularization by adding a penalty equivalent to the sum of the absolute values of the coefficients. This leads to some coefficients which are shrunk to zero (or approximately to zero), which effectively means that the features associated with those coefficients are eliminated (or given a low weight). In the case of Ridge regression, the regularization is performed by adding a penalty equivalent to the sum of the squares of the coefficients, but in this case, the coefficients are simply shrunk, not setting any of them to zero. Finally, Enet regression linearly combines the penalties of LASSO and Ridge methods, becoming a mixture of both. The three methods are useful to analyze data affected by multicollinearity and avoid the need of removing features in advance.

The modelling was conducted with R software by using `glmnet` [16] and `caret` [17] packages. The algorithms use cyclical coordinate descent, computed along a regularization path. These algorithms are fast and able to handle large problems, dealing efficiently with sparse features [18]. The shrinkage parameter in each case was valued by using a grid search.

V. Experimental results

Several classification experiments have been carried out to analyze the performance and stability of the three regularization methods (LASSO, Ridge, Enet) with four feature sets extracted from the voice recordings of the 80 people participating in the study. The gender has been included in all the sets. With respect to the acoustic features, the first set consists of perturbation and SNR-based measures, the second set is composed of nonlinear characteristics, and the third one includes MFCC-based features. Finally, all the features from the previous sets are considered.

In order to estimate the model performance on new individuals, 10-fold cross-validation schemes are considered [19]. Specifically, in 10-fold cross-validation, the original sample is randomly partitioned into 10 equal-size subsamples. A single subsample (out of the 10 defined) is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the validation data. For each iteration, the metrics used for analyzing the performance of the detection system are defined in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as follows: accuracy rate = $(TN + TP)/n$, where n is the number of subjects, sensitivity = $TP/(TP + FN)$, and specificity = $TN/(TN + FP)$. The 10 results from the folds are averaged to produce a single estimation for each metric. For diagnostic purposes, the 10-fold cross-validation is performed in a stratified way so that each subsample has the same number of subjects with PD and control subjects.

Table II shows the accuracy rates, sensitivities, and specificities for the four feature sets and the three methods.

TABLE II
Classification performance.

	Accuracy(%)	Sensitivity (%)	Specificity(%)
	Perturbation and SNR features		
LASSO	64.4	64.0	64.8
Ridge	69.4	67.5	71.3
Enet	66.3	65.8	66.8
	Nonlinear features		
LASSO	80.1	79.5	80.8
Ridge	80.0	80.8	79.3
Enet	81.3	81.0	81.5
	MFCC-based features		
LASSO	83.0	82.0	84.0
Ridge	84.5	83.5	85.5
Enet	84.8	84.5	85.0
	All features		
LASSO	85.6	83.8	87.5
Ridge	88.5	89.0	88.0
Enet	87.1	87.8	86.5

The results show the low discrimination capability achieved when using features exclusively based on perturbation and SNR. Accuracy rates are lower than 70% as well as all the sensitivities and 2 out of 3 specificities. When the set of nonlinear features is considered, the accuracies are increased to around 80%, as well as the sensitivities and specificities. This leads to an increase with respect to the results obtained with perturbation and SNR measures from 10% to 15%. This supports the already established and experimentally tested theory that traditional linear perturbation methods of voice signal analysis do not account for the two main biophysical symptoms of voice disorders, which are complex nonlinear aperiodicity and turbulent non-Gaussian randomness [20]. The three metrics are increased when MFCC-based features are exclusively considered. For accuracies, the increase with respect to the results obtained with nonlinear features ranges from 2.9% to 4.5%, whereas for sensitivities and specificities it ranges from 2.5% to 3.5%, and from 3.2% to 6.2%, respectively. Finally, the best results are obtained when all the features are used. The best accuracy, 88.5%, is obtained for Ridge regression, followed by the 87.1% obtained with Enet method, and, finally, LASSO provided an accuracy of 85.6%. All of them improve the accuracy rates obtained with the other feature sets.

ROC (Receiver Operating Characteristic) curves are presented for the three methods when they were applied to all features, see figures 1 to 3. The same order of goodness-of-fit is obtained for the three methods based on AUC (Area Under Curve) ROC values.

The regularization methods applied in this study consider two parameters. The first one is α, the mixing parameter, and it is set to 1 for LASSO, to 0 for Ridge and between 0 and 1 for Enet. The second one is λ that defines the amount

ISBN 978-90-827970-1-5 © EURASIP 2018

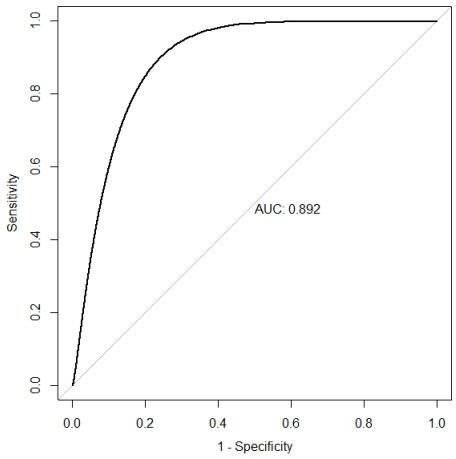

Fig. 1. ROC curves and AUC for LASSO with all features.

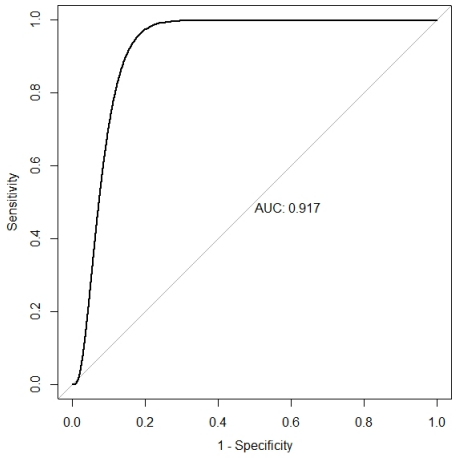

Fig. 2. ROC curves and AUC for Ridge with all features.

of shrinkage in the model, and it is optimized with `caret` [17] by cross-validation to achieve the best possible accuracy results. LASSO and Enet can shrink the regression coefficients to zero, whereas Ridge can shrink them, but not set them to zero. Table III shows the number of regression coefficients remaining in the models after regularization (including the intercept). The coefficient values for Rigde correspond to all the acoustic features, gender and intercept in the four datasets.

TABLE III
NUMBER OF REGRESSION COEFFICIENTS CONSIDERED BY THE THREE APPROACHES.

	LASSO	**Ridge**	**Enet**
Perturbation and SNR features	13	13	12
Nonlinear features	9	9	9
MFCC-based features	6	28	14
All features	7	46	17

When the number of acoustic features is not too large

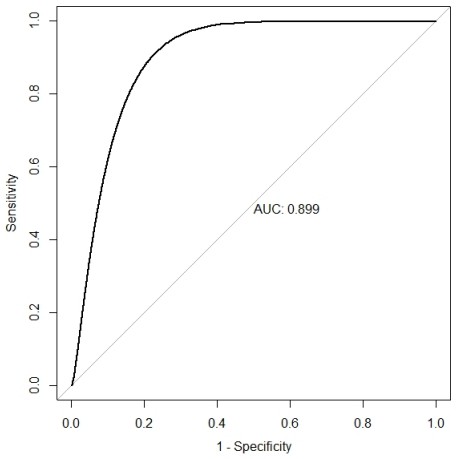

Fig. 3. ROC curves and AUC for Enet with all features.

(perturbation and SNR or nonlinear cases), LASSO is not able to filter features and Enet is able to filter only one perturbation feature. However, when the number of features increases (MFCC-based features and all features), LASSO is able to shrink many regression parameters to zero. In fact, it is able to shrink to zero 22 and 39 parameters out of 28 and 46, respectively. Enet shrinks to zero 14 and 29 parameters, respectively.

The three applied methods have been designed to work with many features and a reduced number of subjects. In the following the results considering all the acoustic features will be discussed. The data matrix has 80 rows (number of subjects) and 45 features (44 acoustic features and gender). In general, this would correspond to a large number of features for a reduced number of subjects. However, in this context, due the difficulty of recruiting volunteers suffering from PD the number of subjects is considered to be moderate.

If the interest is focused on the accuracy, Ridge provides the best one (88.5%), but keeping all the features. LASSO only keeps 7 regression parameters related to intercept, gender, PPE, SampleEn, μ^{10}_{MFCC}, σ^{5}_{MFCC}, and σ^{11}_{MFCC}. This combination of features does not require estimation of the fundamental frequency, which is a hard task in the case of pathological voices. This leads to 85.6% of accuracy. For the Enet method, the accuracy is increased to 87.1% and the number of features also increases. In addition to the ones obtained for LASSO, Enet also considers RPDE, DFA, RAP, μ^{1}_{MFCC}, μ^{3}_{MFCC}, μ^{5}_{MFCC}, σ^{1}_{MFCC}, σ^{3}_{MFCC}, σ^{9}_{MFCC}, and σ^{10}_{MFCC}. The fact that sample entropy has been kept in LASSO and Enet methods instead of other entropy measures allows to avoid the calculation of more refined and computationally heavier features such as fuzzy entropy in further studies. The results also confirm that the gender feature is relevant and should be taken into account in a PD detection system.

ISBN 978-90-827970-1-5 © EURASIP 2018

VI. Conclusion

Based on these findings, the adoption of regularization is recommended as an easy, fast, and effective method to perform PD detection. The prediction results keep rather stable by varying the regularization method, despite the fact that the number of recruited subjects is moderate. Although the best accuracy is obtained with Ridge, it is noticeable that LASSO offers the easiest model interpretability, since this method exclude many features. Enet offers intermediate results both in terms of interpretability and accuracy.

Acknowledgment

Thanks to the anonymous participants and to Carmen Bravo and Rosa María Muñoz for carrying out the voice recordings and providing information from the people with PD.

References

[1] E. R. Dorsey and B. R. Bloem, "The Parkinson pandemic–A call to action," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 75, no. 1, pp. 9–10, 2018.

[2] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[3] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[4] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J.F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, "Automatic detection of Parkinson's disease in running speech spoken in three different languages," *Journal of the Acoustic Society of America*, vol. 139, no. 1, pp. 481–500, 2016.

[5] M. Novotny, J. Rusz, R. Cmejla, and E. Ruzicka, "Automatic evaluation of articulatory disorders in Parkinson's disease," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1366–1378, 2014.

[6] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200–1205.

[7] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[8] A. Hoerl and R. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 2, pp. 301–320, 2005.

[10] J. R. Orozco-Arroyave, *Analysis of speech of people with Parkinson's disease*, Logos-Verlag, Berlin (Germany), 2016.

[11] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 10, pp. 1100–1108, 2016.

[12] V. M. Sauvageau, J. Macoir, M. Langlois, M. Prud'Homme, L. Cantin, and J. P. Roy, "Changes in vowel articulation with subthalamic nucleus deep brain stimulation in dysarthric speakers with Parkinson's disease," *Parkinson's Disease*, no. ID 487035, pp. 1–9, 2014.

[13] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.

[14] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with Parkinson's disease through nonlinear dynamics," in *Advances in Nonlinear Speech Processing*, T. Drugman and T. Dutoit, Eds., vol. LNAI 7911 of *Lecture Notes in Artificial Intelligence*, pp. 112–119. Springer-Verlag, 2013.

[15] I. Hertrich and H. Ackermann, "Gender-specific vocal dysfunctions in Parkinson's disease: electroglottographic and acoustic analyses," *Annals of Otology, Rhinology and Laryngology*, vol. 104, no. 3, pp. 197–202, 1995.

[16] J. Friedman, T. Hastie, and R. Tibshirani, "glmnet: Lasso and elastic-net regularized generalized linear models.," R package version 2.0-13, 2017.

[17] M. Kuhn, "Classification and regression training," R package version 6.0-78, 2017.

[18] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[19] A. Webb, *Statistical Pattern Recognition*, John Wiley and Sons, Chichester, 2002.

[20] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 23, pp. 1–19, 2007.

ISBN 978-90-827970-1-5 © EURASIP 2018