INTERSPEECH 2021: Show & Tell Contribution
30 August – 3 September, 2021, Brno, Czechia

Application for detecting depression, Parkinson's disease and dysphonic speech

Gábor Kiss, Dávid Sztahó, Miklós Gábriel Tulics

Budapest University of Technology and Economics, Budapest, Hungary
{kiss.gabor,sztaho.david,tulics.miklos}@vik.bme.hu

Abstract

In this Show&Tell presentation we demonstrate an application that is able to assess a voice sample according to three different voice disorders: depression, Parkinson's disease and dysphonic speech. Affection probability of each disorder is analyzed along with their severity estimation. Although the acoustic models (support vector machine and regression models) are trained on Hungarian voice samples, English samples can also be utilized for assessment. The results are displayed by as pie chart for probabilities and separate severity scores. The input of the application is a read text with a fixed linguistic content. It is possible to load a pre-recorded voice sample or create a live recording. The developed system could evaluate a speaker's voice sample, assisting medical staff.

Index Terms: depression, SVR, SVM, speech, support vector regression, support vector machines, pathological speech, dysphonia, Parkinson's disease

1. Introduction

Speech as a biomarker has been getting more and more popularity in the past few years. There are numerous affections and diseases that affect speech production and thus can be identified by voice analysis. In this presentation classification of various diseases are demonstrated in practice. The following disorders are considered: depression, dysphonic speech and Parkinson's disease.

Dysphonia is the disturbance of the produced sound: the clear sound of the voice is accompanied by noise. Dysphonic voice is characterized to be hoarse, breathy, harsh or rough [1]. Functional dysphonia is the voice problem in the absence of a physical condition, whereas organic dysphonia caused by physiological change in a speech production subsystem. Depression is a psychiatric disorder caused by an experience of failure with emotional, cognitive and physical symptoms [2]. Parkinson's disease (PD) in considered as a neurological voice disorders [3].

In this work, we put into practice our earlier works [4]–[6]. Namely, the automatic classification of the aforementioned diseases and the automatic estimation of their severity. Various models are built: (1) estimating the probability of each disease in a joint model in one step and (2) estimating the severity of each disease by separately trained model per disease.

By this Show&Tell, presentation we will introduce a decision support tool that aims to detect the affection of the aforementioned disorders: depression, dysphonia and Parkinson's disease (a distinguished type of dysphonia).

2. Overview of the application

The application was developed to analyze a speech sample according to three different voice disorders. This sample can be a prerecorded sound file or it is possible to record a new one using the application. The speech samples must contain a fixed text, a tale titled 'The North Wind and the Sun'. It is a commonly used text in clinical research. The text is displayed on the screen in two languages: English and Hungarian. The output of the algorithm shows multiple evaluations: (1) probabilities of the voice disorders and (2) estimated severity level of each disorder. The severities are shown according to scales specific for each disorder:

- Depression: Beck Depression Inventory II (BDI) scale [7]. The BDI-II scale ranges from 0 to 63, where 0 indicates a healthy state, while 63 is the worst depression condition. The BDI-II scale uses the following rating: 0-13 healthy, 14-19 mild depression, 20-28 moderate depression, 29-63 severe depression.
- Dysphonic speech: hoarseness according to the RBH, ranging from 0 to 3.
- Parkinson's disease: Hoehn & Yahr scale (HY) [8], which ranges from 1 to 5, where 1 indicates minimal PD while 5 is the worst PD condition.

The acoustic models are built by speech samples of Hungarian speakers. The exact number of speakers and their statistics can be found in [9]. Although the models are created on Hungarian samples, the method can process samples of other Languages (such as English) due to the fact that processing algorithms are language independent (for European languages).

The process of decision making is the following. The input sample (live or pre-recorded) is normalized and segmented to phonemes. The feature extractor calculates acoustic features necessary to analyze the sample. The list of acoustic features is shown in Table 1. These features are the fed into four decision units. The joint decision unit (first unit) calculates the probability of each disorder by Support Vector Machine model using linear kernel function (result is displayed as pie chart). Beside the joint probability decision, the severity of each disorder is also estimated (additional three decision units) and

Table 1: *Acoustic features used in decision making.*

Feature	Measurement location	Statistical function
intensity	total sample	mean, standard deviation, percentile (1, 5, 10, 25)
pitch	voiced segments	mean, standard deviation, percentile (1, 5, 10, 25)
formants, HNR, jitter, shimmer	total sample, vowel /e/	mean, standard deviation
12 MFCCs	total sample, vowel /e/	mean
rate of transients	total sample	n.a.
ratio of silence parts	total sample	n.a.
soft phonation index	total sample	n.a.

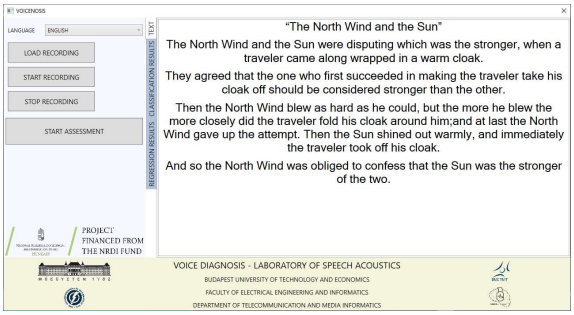

Figure 1: *User interface - recording.*

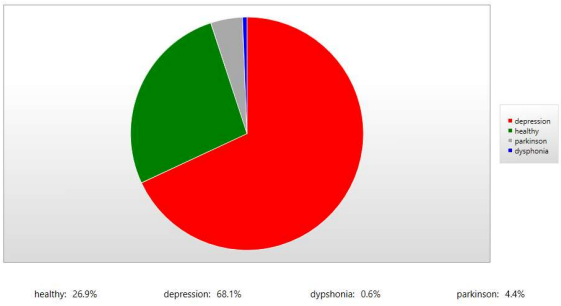

Figure 2: *User interface – joint decision*

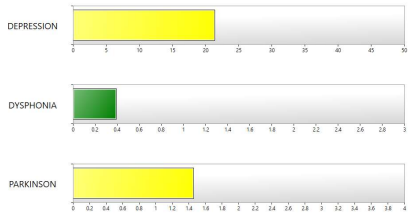

Figure 3: *User interface – severity prediction*

displayed in the application by support vector regression (with linear kernel). The user interface of the application is shown in Figures 1, 2 and 3. Severity is also coded by colors: green-healthy, yellow-mild and red-severe.

3. Evaluation of the application's performance

The models in the application are evaluated in a 10-fold nested cross-validation setup. The confusion matrix of the test samples in shown in Table 2. The cells contain sample number in percentages. The overall balanced accuracy is 81.1%. The estimation of the severity of each disorder is evaluated by RMSE and Pearson correlation metric. These are shown in

Table 2: *Confusion matrix of joint model evaluation. Abbreviations: de-depression, hc-healthy, pd-Parkinson, dy-dysphonia*

		predicted class			
		de	**hc**	**pd**	**dy**
original class	**de**	**73%**	21%	5%	1%
	hc	2%	**93%**	1%	3%
	pd	6%	23%	**70%**	1%
	dy	1%	10%	1%	**88%**

Table 3: *Evaluation of severity level estimations. Abbreviations: de-depression, pd-Parkinson, dy-dysphonia*

	RMSE	**Normalized RMSE**	**Pearson**
de	10.1	0.77	0.59
dy	0.6	0.60	0.82
pa	1.1	0.73	0.67

Table 3. Each severity is estimated by a model trained on the given category.

4. Conclusions

The application described in this short paper is currently capable of estimating the probability of three types of voice disorders: depression, dysphonia and Parkinson's disease in English and Hungarian. Results of the pre-trained models and the first user experiences show that the application can be a promising tool for medicals or GPs.

5. Acknowledgements

Project no. K128568 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_18 funding scheme. The research was partly funded by the CELSA (CELSA/18/027) project titled: "Models of Pathological Speech for Diagnosis and Speech Recognition".

6. References

[1] R. Jani, S. Jaana, L. Laura, and V. Jos, "Systematic review of the treatment of functional dysphonia and prevention of voice disorders," *Otolaryngology—head and Neck Surgery*, vol. 138, no. 5, pp. 557–565, 2008.

[2] "Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies | Family Practice | Oxford Academic." https://academic.oup.com/fampra/article/29/3/255/461869?login=true (accessed May 03, 2021).

[3] R. Balestrino and A. H. V. Schapira, "Parkinson disease," *Eur J Neurol*, vol. 27, no. 1, pp. 27–42, Jan. 2020, doi: 10.1111/ene.14108.

[4] D. Sztahó, I. Valálik, and K. Vicsi, "Parkinson's Disease Severity Estimation on Hungarian Speech Using Various Speech Tasks," in *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2019, pp. 1–6.

[5] G. Kiss and A. Z. Jenei, "Investigation of the Accuracy of Depression Prediction Based on Speech Processing," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, 2020, pp. 129–132.

[6] M. G. Tulics, G. Szaszák, K. Mészáros, and K. Vicsi, "Artificial neural network and svm based voice disorder classification," in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2019, pp. 307–312.

[7] A. T. Beck, R. A. Steer, R. Ball, and W. F. Ranieri, "Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients," *Journal of Personality Assessment*, vol. 67, no. 3, pp. 588–597, 0 1996, doi: 10.1207/s15327752jpa6703_13.

[8] M. M. Hoehn and M. D. Yahr, "Parkinsonism: onset, progression and mortality.," *Neurology*, vol. 17, no. 5, pp. 427–442, May 1967, doi: 10.1212/wnl.17.5.427.

[9] D. Sztahó, G. Kiss, and M. G. Tulics, "Deep Learning Solution for Pathological Voice Detection using LSTM-based Autoencoder Hybrid with Multi-Task Learning," in *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021)*, 2021, vol. 4, pp. 135–141.