U-healthcare system for pre-diagnosis of Parkinson's disease from voice signal

Sylvio Barbon Junior
Department of Computer Science
State University of Londrina (UEL)
Londrina, Brazil
barbon@uel.br

Victor G. Turrisi da Costa
Department of Computer Science
State University of Londrina (UEL)
Londrina, Brazil
victorturrisi@uel.br

Shi-Huang Chen
Dep. of Computer Science & Information Engineering
Shu-Te University
Kaohsiung City, Taiwan (R.O.C.)
shchen@ieee.org

Rodrigo Capobianco Guido
Instituto de Biociências, Letras e Ciências Exatas, Unesp
São Paulo State University
São José do Rio Preto - SP, Brazil
guido@ieee.org

***Abstract*—With the ageing and growth of the population, some chronic diseases, such as Parkinson's disease (PD), urge the society to a health-conscious looking for better health system designs. Some recent research endeavour has been supported by solutions grounded in ubiquitous healthcare (u-Health) coupling telemedicine, context awareness and decision support capabilities. In this work, we propose a u-healthcare system to pre-diagnose PD based on the speech signal of people under voice call. The speech stream is sampled as well as processed to support the pre-diagnose using machine learning (ML). Experiments were conducted over a PD voice dataset composed of 40 individuals by using five different ML algorithms. Based on a linear Support Vector Machine (SVM) model, a false negative rate of 10% was obtained when classifying the locution of number "three".**

***Index Terms*—Signal Processing, Ubiquitous Computing, Health, Machine Learning, Speech**

I. Introduction

Early diagnosis of Parkinson disease (PD) is an important goal not only for clinical, pharmacology, and epidemiological studies, but also for further prognostic and therapeutic reasons. Rizzo et al. [1] reported a pooled diagnostic accuracy of 73.8% for clinical diagnosis of PD performed mainly by nonexperts. Part of this misdiagnose is related to the overlap of pathological characteristics with different diseases. Sakar et al. in [2] described voice signals as a rich descriptor for identifying PD. Moreover, from speech features, telediagnosis and telemonitoring systems based on speech signals leverage solutions with low cost and easy to self-use. In the same direction, Tsanas et al. in [3] highlighted the need for the development of accurate and objective tools in assessing PD, since the current diagnosis is poor. The authors proposed a solution of telemonitoring of PD using noninvasive speech tests.

In this scenario, the Ubiquitous healthcare (u-Healthcare) is an emerging option to support medical care, lightening the burden of frequent and often inconvenient visits to the clinic. Furthermore, this reliable technology tackles the excessive additional workload, reducing the costs with similar accuracy to clinical evaluation. U-Healthcare systems also aid service providers and practitioners to remotely track the patient's physiological data in real-time and provide feedback [4].

In a typical u-Healthcare system, sensors, such as Electrocardiography, Electroencephalography and Electromyography forward their data via a wireless network interface to a base station. Kim and Lee [5] emphasised that in a u-Healthcare, physical life objects and space are coupled with numerous sensors and devices to give them context awareness and decision support capabilities. Robust implementations of decision support models addressing voice signal classification in health solutions are grounded in machine learning (ML) algorithms [6].

Considering the voice signal as a reliable descriptor for PD diagnosis [2], [3], [7], we propose a general u-Healthcare system for pre-diagnosing of PD based on ML to classify speech signals gathered from voice calling. The model induction is based on supervised ML over voice signal features with labels related to the Unified Parkinson's Disease Rating Scale (UPDRS). After the voice gathering step, the speech is converted to known signals, that is 26 sustained phonations and running speech, used to create the pre-diagnosis model. Base on features extracted from the voice signal, the model classifies the speaker as PD pre-diagnosed or without PD manifestation. The result is stored in a repository and returned to the user.

Experiments were conducted on a real-life dataset composed of 20 PD diagnosed patients and 20 healthy individuals. Five different ML algorithms - Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost) - were compared seeking for low false negative rates (FNR) in PD diagnoses.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

II. Proposed Approach

Our u-Healthcare system comprehends to several capabilities, but two key-points are imperative: the sensoring and capacity to provide insights. Addressing these points, our proposal presents a Web Service [8] for voice gathering and pre-diagnosis modelling. More specifically, as shown in Fig. 1, we propose a six-step solution, composed of Voice Gathering (Step 1), Speech Parsing (Step 2), Speech Recognition (Step 3), Feature Extraction (Step 4), Classification (Step 5) and Pre-diagnosis (Step 6). Afterwards, the classification report is stored into a repository and returned to the user whose started the voice call.

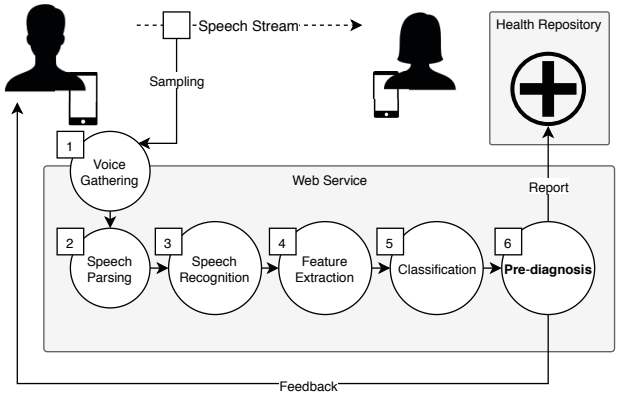

Fig. 1. Proposed approach overview.

A. Voice Gathering

The proposed approach starts when a user makes a voice call with an application allowed to monitor PD. During the conversation, voice signals are sampled from the speech stream by the Voice Gathering step. Several approaches to detect speech have been developed in last years, including solutions for challenging scenarios, such as hands-free communication in noisy environments and dialog in background music [9]. In this step, Voice Activity Detection (VAD) algorithms are employed to avoid forwarding silence or noisy streams to the Speech Parsing step.

B. Speech Parsing

Once the voice data have been acquired, the components of speech can be straightforwardly analysed and parsed [10]. There are several toolkits and frameworks widely employed for speech parsing tasks, e.g., Stanford CoreNLP toolkit [11]. Considering that parser performance tends to deteriorate with increasing length of string, we propose the usage of a restrict list of short sentences, small words and sustained vowels to support PD pre-diagnosis.

C. Speech Recognition

In this step, each parsed data is verified in a set of voiced signals (locutions) for checking its capacity to describe PD. The set of voiced signals that match the trained dataset are used to induce the classification model. It is important to highlight that additional filtering [12], normalisation [13] and scaling [14] task are required to perform a suitable feature extraction in the next step.

D. Feature Extraction

Some speech characteristics of PD subjects, e.g., reduced loudness, breathiness, roughness and exaggerated vocal tremor are extracted from voice data signal [2]. These time-frequency features are grounded on parameters from frequency, pulse, amplitude, voicing, harmonic and pitch [15]. All of them have been wide explored in several studies on digital signal processing [2], [3], [15]. The extracted features are forwarded to the classification model for predicting task.

E. Classification

Supervised ML algorithms are able to induce classification models during the training phase based on previously labelled samples. There are several algorithms that lead to accurate models, e.g., DT [16], RF [17], GBM [18], SVM [19] and XGBoost [20]. More specifically, classification, a subarea of problems from ML, consists of inducing a model capable of predicting a target class y given a feature vector X. After obtaining an accurate model built during the training phase, a given new unlabelled sample is predicted. It is important to highlight the focus on reducing the FNR since the disease should not be undetected [21].

F. Pre-diagnosis

The proposed approach can offer a PD pre-diagnoses report, considering the obtained outcome from a low FNR model. Moreover, the result can be stored in a Health Repository and triggered to a specialist for further medical actions. For a more reliable pre-diagnose, the Web Service could gather more speech data to reinforce the final decision.

III. Materials and Methods

A. Parkinson's Disease Dataset

To evaluate the proposed approach, a voice dataset with health and diseased individuals was employed through our Web Service. The PD database [2] consists of data extracted from sound recordings from 20 individuals with Parkinson's disease and 20 healthy individuals whom appealed at the Department of Neurology in Cerrahpasa Faculty of Medicine of Istanbul University. Each individual has 26 voiced samples which are divided into sustained vowels ("a", "o" and "u"); numbers from 1 to 10; short sentences; and words. We extracted traditional signal features from each voiced sample as in [22]. The features are: Jitter (local), Jitter (local, absolute), Jitter (rap), Jitter (ppq5), Jitter (ddp), Shimmer (local), Shimmer (local, dB), Shimmer (apq3), Shimmer (apq5), Shimmer (apq11), Shimmer (dda), Autocorrelation (AC), noise-to-harmonic ratio (NTH), Harmonic-to-noise ratio (HTN), median, mean, standard deviation, minimum and maximum pitch, number of pulses, number of periods, mean and standard deviation of periods, fraction of locally unvoiced frames, number of voice breaks, and degree of voice breaks.

B. Machine Learning

The ML algorithms chosen to be experimented in our proposed approach are the most recent and accurate ones, as described. The algorithm implementations from the scikit-learn library [23] were used, except for the XGBoost. For that, the library developed by [20] was used. All algorithms were performed with default hyperparameters.

1) Decision Tree: DTs are a class of ML algorithms which employ a divide-and-conquer procedure to group similar instances [16]. It works by greedily selecting the best split point considering current instances and dividing them into subgroups. This process is repeated until instances from a subgroup are from the same class or according to a stop criteria. At test time, instances are sorted into their respective leaves and assigned the majority class of those leaves.

2) Random Forest: RF [17] is a ML algorithm which combines multiple DTs. In this sense, it is categorised as an ensemble algorithm. Given a dataset of size n, this algorithm builds t datasets of size n by sampling with repetition from the original dataset. Then, each dataset is used to induce a different DT. To further add statistical differences between DTs, each tree can only consider m features at each split attempt. At test time, the RFs use all trees in conjunction and output the model of their predictions.

3) Gradient Boosting Machine: The GBM [18] is a framework for function approximation. When used as a ML algorithm, it consecutively fits new models with the goal of providing a more accurate estimation. This is done by making each model address the error of the previous models. It is highly flexible since any function can be used as the loss function, giving freedom to the researchers to choose, or even create, a function that best fits any data-driven task. At test time, trees are employed sequentially and their predictions are combined.

4) Support Vector Machine: SVM [19] is a ML algorithm which creates an optimal separation hyperplane that splits the data into two groups. The boundaries of this hyperplane are defined by points referred to as support vectors. At test time, instances are assigned the class values of instances in the same part of the divided hyperplane.

5) Extreme Gradient Boosting: XGBoost [20] is a variation of the GBM algorithm. The two main differences between them rely on modelling details and the speed-focus of the XGBoost. When inducing trees, the XGBoost algorithm uses a regularisation parameter, which makes it outperform the traditional GBM. Testing procedures work in the same way as GBM.

IV. Results and Evaluation

In Table I the accuracy values per locution for each algorithm are presented and the best accuracy values are highlighted in bold. Due to differences in the nature of each algorithm, there was no clear best locution. For the RF, the best results were for locutions 7 and 15. SVM presented high accuracy values for locutions 6 and 17. GBM had a peak accuracy for locution 11. DT obtained the highest accuracy values for locutions 5, 22 and 25. Lastly, XGBoost performed better for locutions 6 and 17.

TABLE I
Accuracy per locution for each algorithm.

Locution	RF	SVM-linear	GBM	DT	XGBoost
1	0.475	0.100	0.450	0.500	0.500
2	0.475	0.200	0.675	0.525	0.525
3	0.375	0.250	0.425	0.450	0.425
4	0.650	0.600	0.675	0.525	0.600
5	0.650	0.625	0.575	**0.675**	0.575
6	0.650	**0.725**	0.425	0.425	**0.700**
7	**0.700**	0.525	0.550	0.650	0.650
8	0.600	0.475	0.550	0.600	0.475
9	0.550	0.275	0.425	0.375	0.475
10	0.600	0.575	0.625	0.650	0.575
11	0.675	0.450	**0.700**	0.600	0.500
12	0.475	0.500	0.575	0.475	0.500
13	0.400	0.525	0.350	0.400	0.350
14	0.525	0.500	0.525	0.450	0.625
15	**0.700**	0.525	0.625	0.500	0.600
16	0.575	0.575	0.550	0.450	0.650
17	0.650	**0.725**	0.525	0.375	**0.700**
18	0.550	0.400	0.500	0.575	0.475
19	0.375	0.475	0.575	0.600	0.525
20	0.500	0.575	0.600	0.425	0.450
21	0.500	0.200	0.525	0.475	0.475
22	0.350	0.350	0.425	**0.675**	0.550
23	0.500	0.450	0.600	0.575	0.650
24	0.450	0.200	0.350	0.375	0.450
25	0.650	0.650	0.625	**0.675**	0.650
26	0.375	0.325	0.300	0.275	0.475

In Figure 2 a heatmap of the FNR per algorithm for each locution is presented. This metric measures the percentage of subjects with PD that the system falsely identifies as healthy, which corresponds to the worst scenario. Smaller values (denoted by colours close to white) represent better performance, whereas larger values (denoted by red colours) comprehend to worse results. First, it is possible to see that the linear SVM presented the lowest FNR of 0.10 for Locution 6 (the subject speaking the number three) and also other low FNR values for locutions 13, 15, 19, 20 and 25. GBM and XGBoost also presented low FNR values for locution 2 and 6 for the first and latter, respectively.

To emphasise a possible superiority of an algorithm, Friedman's statistical test and the Nemenyi post hoc test have been applied to the averaged FNR values. Figure 3 shows the Critical Difference (CD) diagram obtained from the statistical test results. It was possible to observe that they are not significantly different, since each performs better given a different locution.

As stated in [1], recent results indicate that PD detection is not highly precise when compared to tests as pathological evidence at autopsies, e.g., vascular lesions, striatal plaques and diffuse Lewy body disease. However, it is comparable to other recent literature [2], [3]. In this way, the outcome found is important mainly concerning a u-Healthcare system and further advantages as strategic insights for promoting and accelerating the realisation of PD pre-diagnoses in early stages.

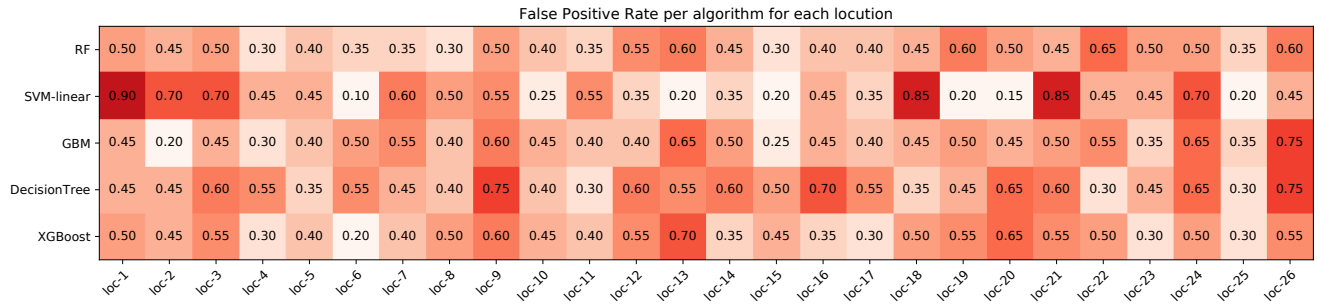

Fig. 2. Heatmap of FNR per locution for each algorithm.

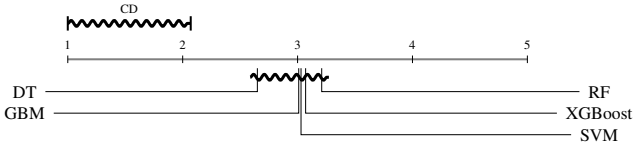

Fig. 3. Comparison of the averaged FNR values from algorithm according to the Nemenyi test, all of them are not significantly different (at $\alpha = 0.05$ and CD $= 1.07$)

V. Conclusion

In this paper, we proposed a u-Healthcare system that combined voice data features and machine learning classification for supporting Parkinson's disease pre-diagnosis embedded in a Web Service. Based on voice gathering, the user whose started a voice call receives feedback about his health and, then further information is stored for additional analysis. Our results exposed an FNR of 10% using SVM over features extracted from the word "three". As future work, we will compute the final classification using an ensemble of signals processed, instead of a single pre-diagnosis per voice gathered.

References

[1] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of parkinson disease," *Neurology*, vol. 86, no. 6, pp. 566–576, 2016.

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

[3] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[4] F. Touati and R. Tabish, "U-healthcare system: State-of-the-art review and challenges," *Journal of medical systems*, vol. 37, no. 3, p. 9949, 2013.

[5] S.-y. Kim and Y.-T. Lee, "Accelerating the growth of u-healthcare in korea: Conceptualizing and classifying from service perspective," *Journal of Service Science*, vol. 1, no. 1, pp. 57–81, 2009.

[6] R. A. Greenes, *Clinical decision support: the road ahead*. Elsevier, 2011.

[7] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[8] D. Roman, U. Keller, H. Lausen, J. De Bruijn, R. Lara, M. Stollberg, A. Polleres, C. Feier, C. Bussler, and D. Fensel, "Web service modeling ontology," *Applied ontology*, vol. 1, no. 1, pp. 77–106, 2005.

[9] C. Uhle and T. Bäckström, "Voice activity detection," in *Speech Coding*, pp. 185–203, Springer, 2017.

[10] A. Caines, M. McCarthy, and P. Buttery, "Parsing transcripts of speech," in *Proceedings of the Workshop on Speech-Centric Natural Language Processing*, pp. 27–36, 2017.

[11] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60, 2014.

[12] P. C. Fantinato, R. C. Guido, S.-H. Chen, B. L. S. Santos, L. S. Vieira, S. B. Júnior, L. C. Rodrigues, F. L. Sanchez, L. M. Souza, C. D. Maciel, *et al.*, "A fractal-based approach for speech segmentation," in *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, pp. 551–555, IEEE, 2008.

[13] L. Polka, O.-S. Bohn, and D. J. Weiss, "Commentary: Revisiting vocal perception in non-human animals: a review of vowel discrimination, speaker voice recognition, and speaker normalization," *Frontiers in psychology*, vol. 6, p. 941, 2015.

[14] K. Pisanski, V. Cartei, C. McGettigan, J. Raine, and D. Reby, "Voice modulation: a window into the origins of human vocal control?," *Trends in cognitive sciences*, vol. 20, no. 4, pp. 304–318, 2016.

[15] S. Aich, K. Younga, K. L. Hui, A. A. Al-Absi, and M. Sain, "A nonlinear decision tree based classification approach to predict the parkinson's disease using different feature sets of voice data," in *Advanced Communication Technology (ICACT), 2018 20th International Conference on*, pp. 638–642, IEEE, 2018.

[16] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, Oct 2001.

[18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sep 1995.

[20] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2016.

[21] A. Ukil, S. Bandyoapdhyay, C. Puri, and A. Pal, "Iot healthcare analytics: The importance of anomaly detection," in *Advanced Information Networking and Applications (AINA), 2016 IEEE 30th International Conference on*, pp. 994–997, IEEE, 2016.

[22] P. Boersma *et al.*, "Praat, a system for doing phonetics by computer," *Glot international*, vol. 5, 2002.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.