Parkinson's Diagnosis through Compressed Speech Signals at remote Locations

Adil Pervaiz[1], Ammar Tahir[2]

[1]School of Sciences and Technology, University of management and technology Lahore, Pakistan
[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
*Correspondence: ammar_tahir796@yahoo.com

Rapid increase in Parkinson's disease all over the world before the age of 60's is alarming. Parkinson's affects the neurological controls of human body. In this paper we intend to adopt machine based learning to automate the diagnosis of Parkinson's disease with help of compressed speech signals sent via network. An advanced Solution is proposed to automate the diagnosis process more rapid than that of ordinary process. In this study Speech samples from 55 subjects have been collected, 43 with Parkinson Disease and 12 Healthy subjects. Speech signals from clients captured from various sensors and devices transmitted to the cloud for processing. In the cloud Speech samples are compressed using compression codecs MP3, MP4, G.722, G.226, GSM-EFR, AMR-WB, SVOPC/SILK, and OPUS, and then same are diagnosed without compression impact and deliver accurate results. Three classifiers SVM (Support Vector Machine), NN (Neural Network), and GA (Genetic Algorithm) are applied and calculated accurate results as compared with compressed and un-compressed voice samples. Successful experiments enabled us to achieve 90.7% accuracy. Same process can be used for online and social media applications which are available in portable devices as well.
Keywords: Parkinson Disease (PD), compression codecs, neurological disorder, voice sampling

Introduction

Parkinson disease effects speech impairment and disorder in more than 90% of patients, it affects PD patients' daily life actions especially speech, swallowing, dressing, walking, writing, and movement in bed, are common symptoms. Speech impairment is one of the most common problems of Parkinson's patient. Tele monitoring and diagnosis of Parkinson's through speech is now being researched widely, sustained phonation and running speech tests are commonly used methods for this purpose, sustained phonation deals with fixed letters specially vowel characters, whereas running speech tests contain full sentence. In case of voice impairment Patients have decreased loudness, pause, less stress on sentences, breathy sound and noised speech. Movement of PD patient is also another major issue which is exhaustive for such patients to move for regular clinical tests, diagnosis of developing stage and ranking of disease .Although the same work has been performed by community in traditional way by using compression codecs some work has been performed by J. Rafael for compressed speech signal processing,[1] Voice samples are compressed using different compression codecs and converting them to uniform bit rate. Samples are compressed using, G.722, G.226, GSM-EFR, AMR-WB, /SILK, and OPUS. All these compression methods are being used in telecommunication and VOIP solutions to transfer the voice signals over internet or data links, [1, 2, 6]. A software

defined a cloud-based telemedicine system was proposed in [7] Perform our tests to discriminate speakers with PD and without PD. In this solution voiced and un-voiced durations are segmented and further diagnosis is performed by classification of voice frames on the basis of segment noise, perturbation and cepstral features. Further addition in this method is measurement of frequency / pitch of vocal oscillation, sound pressure or loudness, jitter (variation in speech) shimmer (extent of variation) in speech is calculated. After performing pre-processing we would perform classification of voiced samples on basis of Linear Discriminant analysis, following classifiers are selected for our work, Neural Network (NN), SVM (Support Vector Machine), another new approach for GA (Genetic Algorithm) has also been applied to accommodate our results. We are going to discriminate voice samples with PD or without PD on basis of Voice signals boundaries identified on the basis of our classified features. In our solution we prior applied filtration to remove redundant measures, followed by an meta heuristic approach, to test possible combinations and further extension of the filtered measures with a Genetic Algorithm (GA) classifier In this paper we aim to develop such an algorithm which can diagnose the voice samples even after compression with accuracy. This solution will help all those patients who cannot travel for diagnoses through this solution they will get ranked severity of PD, and will support in

allocating treatment and medicine quantity accordingly. In our solution we have covered possible compression methods used in online and mobile applications. our work is an addition by adopting further compression codecs, using two parallel datasets and performing dual pre-processing before and after compression of voice samples. After getting these results we make comparison of all the results and upload them to cloud for later comparisons, these three additions and further classification techniques increase the accuracy of our results and help in effects of medicines in different patients.
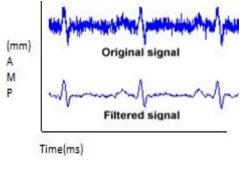
Methodology

Our proposed solution consists of following sections, 1- Pre-processing of voice samples, 2- Encoding of voice samples, 3-Feature extraction and selection, 4- Applying Classifiers.5- Updating the cloud storage and compare the results.

Pre-Processing

Voice samples are recorded in controlled environments in both datasets, and some are picked randomly recorded from different devices using [6] Filtration of signals is performed in Mat lab by using Normalized frequency functions, further using Butterworth design technique to adjust magnitude spectrum of signal, six or seven samples from each subject are selected, all samples are based on selection criteria as suitability of samples for classification, and best voice samples are selected for our study. Voice segments of 3 seconds are selected from each sample due to limitations of applications available for our experiments; available voice samples have 705.6 kbps size.

Feature Collection

In our solution we are using 50% data set from online repository as well as 50% voice samples from Medical lab of Capital City Islamabad Pakistan. We have collected voice samples of PD patients along with detailed discussion session to judge our experiments. Our data set contains sound proof voice recorded samples in order to minimize noise factor. We have focused over collection of dataset within controlled environment to increase the accuracy of our results. All voice samples have the sampling frequency of 44.1 kHz with 16 bits of resolution. As we aim to transmit our samples remotely and over the network a uniform bit rate of these recordings would be $44.1 \times 16 = 705.5$ kbps. And then compressed rate will vary as according to codecs used. Dataset contains male and female patients as well as healthy people voice samples. University of California, (UCI) machine learning dataset repository, in which People have asked to read some vowels and sentences in a series of sessions. Suitable samples are collected in our experiment and the rest are discarded. Selection of voice samples is based on analysis of strong voice signals containing noticeable dysphonia and voice

deviations. Local dataset has been collected with help of National Hospital Lab by conducting voice recording sessions with PD patients, our dataset contains 42 PD patients in which 21 collected from online repository whereas 21 selected from local hospital of Capital City. This ratio enabled us to verify or results more accurately, on the basis of two different datasets which were recorded in similar manner. Adoption of this technique provides diverse mechanisms for verification of results. All voice recordings are selected on basis of minimized noise ratio, to enhance the accuracy and to meet desire results, in our experimental design we have proposed noise removal before compression and another review of noise removal after compression. This technique enhances the pre-processing results and support to build state-of-art design. Voice samples are filtered segmented on the basis of frame size.

Figure 1

Total samples of 412 have been selected from both data sets, balancing ratio of both data sets. It consists of 35 male, 7 female subjects with PD and 9 male, 4 female with HC. The age of all subjects is in 60 to 78 years, voice samples consists of reading of vowels "O", "U", "I" and few words containing vowels "go", "do", "you".

Table 1 Comprehensive Details of Selected Subjects

Subject Code	S/P Gender	Age	UPDRS/H-Y
P-14	M	60.9	2.5
P-12	F	61.1	2
P-02	F	63	3
P-04	M	63	1.5
P-13	M	63	2
P-17	F	66.1	1.5
P-10	M	66.3	1.5
P-16	M	67.4	2
P-18	M	67.9	1.5
P-01	M	68	3
S-24	M	73	2.5
S-26	F	53	2.0
S-08	F	48	2.0

S-39	M	64	2.0
S-33	M	68	2.0
S-32	M	50	1.0
S-01	M	78	3.0
S-22	M	60	1.5
S-37	M	76	1.0
S-21	F	81	1.5
S-04	M	70	2.5
S-19	M	73	1.0
S-35	F	85	4.0
S-05	F	72	3.0
S-18	M	61	2.5
S-16	M	62	2.5
S-27	M	72	2.5
S-10	F	46	HC
S-07	F	48	HC
S-13	M	61	HC
P-26	M	68	HC
P-27	F	68.6	HC
P-28	M	69	HC
P-29	F	70	HC

UPDRSunified Parkinson's disease rating scale, 0-4 ratings, where 0 = normal, 1 = slight, 2 = mild, 3 = moderate, and 4 = severe, Whereas H-Y defines Hoehn and Yahr staging of severity of Parkinson's disease. Here P-subjects represents local dataset and the S-subjects represent online repository dataset

Feature Selection

Voice data is compressed using different codecs e.g. MP3, MP4, G.722, G.226, GSM-EFR, AMR-WB, SVOPC/SILK and OPUS. All codecs are selected because these being used widely in social media applications. Evaluation process is repeated for selection of acoustic features with help of Praat, MDVP, Audacity and Speech analyzer. Selection is made for compared values mainly for Jitter local (Jlocal), Shimmer dB, pitch perturbation (pp), Noise to Harmonic ratio (NHR) are calculated for each sample, values calculated from open source applications adopted for our experiment, after performing statistical analysis and correlation of values, the mean is assigned against each labelled feature.

Table 2: Labelled Feature

Feature Description	Mean $\pm$ Deviation
Jitter percent- (MDVP)	1.512 ± 1.134
Jitter (local)- (PRAAT)	0.861 ± 0.811
Absolute jitter (MDVP)	78.113 ± 81.880

Feature	Value
Jitter local, absolute (PRAAT)	27.445 ± 31.879
Jitter Relative average perturbation (MDVP)	0.916 ± 0.865
Jitter (rap) (PRAAT)	0.313 ± 0.301
Jitter Pitch perturbation quotient (MDVP)	0.962 ± 0.921
Jitter (ppq5) (PRAAT)	0.333 ± 0.301
Shimmer in dB (MDVP)	0.467 ± 0.287
Shimmer (local, dB) (PRAAT)	0.503 ± 0.334
Shimmer percent (MDVP)	5.325 ± 3.070
Shimmer (local) (PRAAT)	5.325 ± 3.570
Amplitude perturbation quotient (MDVP)	3.875 ± 2.312
Shimmer (apq11) (PRAAT)	4.008 ± 2.344
Noise-to-harmonic ratio (MDVP)	0.143 ± 0.049
Mean noise-to-harmonics ratio (PRAAT)	0.029 ± 0.051

Table 3: Compression Code

Codec	Sample Rate	Bit Rate
MP3	8	16
MP4	8	16
GSM-EFR	8	12.2
AMR-WB	16	6.6
G.726	8	16
G.722	16	64
Opus	16	64
SILK	24	25

All voice samples are compressed using series of codecs as follow GSM-EFR, AMR-WB, G.722, G.226, SVOPC/SILK, MP3, MP4 and OPUS, description of is given in detail. The codecs used in this solution to compress the speech signals in order to reduce the bit-rate and thus to make a more easy use of the network resources e.g., G.722 and G.726 are based on the adaptive differential pulse code modulation (ADPCM) method. While GSM-FR, AMR-WB, SILK, and Opus are based on the analysis-by-synthesis concept. A brief description of each codec is provided below.

ADPCM

In this method the difference between the original signal x(n) and the predicted signal ˜x(n) is quantized. The prediction

process is based on a linear prediction (LP) filter, thus the parameters of the LP filter correspond to the
Model of the vocal tract. The difference between the predicted signal and the original (d(n)) corresponds to the excitation. The parameters of the LP filter and the excitation signal are encoded. Similarly encoding used for MP3 compression is CBR 128 Kbit/s resulting in file size 1/11 (=9% or 91% compression) of the original size. MP4 file format (ISO/IEC 14496-14:2003), commonly named as MPEG-4 file format version 2. The MP4 file format was generalized into the ISO Base Media File format ISO/IEC 14496-12:2004, audio file is extracted for necessary processing.

Training of Dataset
After collection of data set, percentage of Severity is calculated in below equation for each dataset
$(S) = 100 * \sum_{i=1}^{Tds} xi/Tds$
In this equation, "S" represents percentage of severity, Tds is total sum of data used for training, and categorization of obtained percentage is scaled as follows. If Result is >0 and <=25 = Low, 25 > and <=65 =Medium, 65> and <=100 =High, whereas No =0.
Evaluation using MCC
(Matthews's correlation coefficient)
TP (True Positive) variable is the total number of PD subjects who are positively classified as PD patients TN (True Negative) is the variable for healthy subjects, FN (False Negative) false labeling PD subject as

$N = TN + TP + FN + FP$
$S = \dfrac{TP + FN}{N}$
$P = \dfrac{TP + FP}{N}$
$MCC = \dfrac{TP/N - S \times P}{\sqrt{PS(1 - S)(1 - P)}}$
$MCC = \dfrac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

healthy, FP (False Positive) false labeling for Healthy Subject as PD, whereas S, and P represent Above Mathew correlation coefficient returns 3 value(−1, 0, +1) +1 relevance shows positive prediction,0represents no random guess,−1 describes a disagreement b/w actual values vs predicted ones

Results
Table 4: Recorded Results

Subject	Available-Record	Result	Level
S-04	2.5	58.20%	Medium

S-19	1	56%	Medium
S-35	4	91%	High
S-05	3	76%	High
S-18	2.5	52%	Medium
S-16	2.5	48%	Medium
S-27	2.5	46%	Medium
S-10	HC	0%	No
S-07	HC	0%	No
P-12	2	39%	Medium
P-02	3	75%	High
P-04	1.5	23%	Low
P-13	2	35%	Medium
P-17	1.5	22%	Low
P-10	1.5	21%	Low
P-16	2	41%	Medium
P-27	HC	0%	No
P-28	HC	0%	No
P-29	HC	0%	No

Results are measured in 10 folds cross validation process, 3 classifiers are applied in this experiment, results are indicated in table-3 showing the three classifiers before compression and after compression, In our experiments we observed that MP4 codec results is reduced, whereas G.722, MP3 and SILK has no significant impact after compression. Significant results by application of three classifiers is also indicated in table-4 showing high accuracy results from NN (Neural Networks), whereas slightly reduced results of SVM (Support Vector Machine) and very close to this is GA (Genetic Algorithm), by applying GA it has been observed that GA is time consuming but with passage of time its chromosome develop improvements, GA provides us future work sight.

Table 6: Code Results

Codecs	Bit Rate	NN	MCC
Un-Compressed	705.6	91+4	97+5
MP3	16	98+5	98+5
MP4	16	97+5	97+5
GSM-EFR	12.2	91+5	91+5
AMR-WB	6.6	94+7	94+7
G.726	16	92+3	92+3
G.722	64	96+6	96+6
Opus	64	91+5	91+5

SILK	25	97+5	97+5

Table 7: Code Results

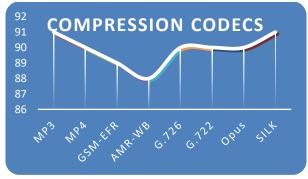| Codecs	Bark Band	Mean	GA	SVM
Un-Compressed	95+5	97+5	90+2	90+3
MP3	93+5	95+5	98+5	98+5
MP4	94+5	95+5	91+5	94+5
GSM-EFR	91+5	91+5	91+5	91+5
AMR-WB	92+7	93+7	90+7	91+7
G.726	92+3	92+3	92+3	91+3
G.722	95+6	96+6	92+6	95+6
Opus	91+5	91+5	91+5	90+5
SILK	95+5	96+5	97+5	95+5

Figure 2

Conclusion

In this solution we are presenting our work by implementing dual preprocessing mechanism to increase the accuracy of the Parkinson's decease diagnoses over network. Moreover number of compression codecs discriminate this design from previous work. Other solutions are solved with Support Vector Machine (SVM) but we are applying Neural Network (NN), SVM and further GA (Genetic Algorithm) for our solution. Our results present more accuracy as we have performed study analysis for three classifiers moreover our focus for enhanced pre-processing for jitter and shimmer calculations in detailed in voice frame classification makes our results more reliable. Our results are more acceptable with combination and enhanced techniques for featuring and classification of voice frames.

References

[1]. C. OkanSakar&OlcayKursun Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia

[2]. Juan Rafael Orozoc-Arroyave, Nicanor Garcia, Jesus Francisco Vargas-Bonilla, ElmarNoth Automatic Detection of Parkinson Disease from Compressed speech Signals

[3]. McSharry, P.E.,Roberts, S.J.,Moroz, I. M xploiting nonlinear recurrence and fractal scaling properties for voice disorder detection).Biomed.Eng.

[4]. McSharry, P.E.,Roberts, S.J.,Moroz, I. M xploiting nonlinear recurrence and fractal scaling properties for voice disorder detection).Biomed.Eng.

[5]. Max A. Little,PatrickE.McSharry, Eric J. Hunter, Jennifer Spielman, and Lorraine O. Ramig. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease

[6]. M. Alhussein, "Monitoring Parkinson's Disease in Smart Cities," in IEEE Access, vol. 5, pp. 19835-19841, 2017.doi: 10.1109/ACCESS.2017.2748561

[7]. L. Hu et al., "Software Defined Healthcare Networks," IEEE Wireless Communication magazine, vol. 22, no. 6, pp.67-75, Dec. 2015.