



[Center for Machine Learning and Intelligent Systems](#)

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Search

☒ Repository ☐ Web



[View ALL Data Sets](#)

Parkinson Speech Dataset with Multiple Types of Sound Recordings Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken.

Data Set Characteristics:	Multivariate	Number of Instances:	1040	Area:	Life
Attribute Characteristics:	Integer, Real	Number of Attributes:	26	Date Donated	2014-06-12
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	104129

Source:

1. Olcay KURSUN, PhD.,
Istanbul University,
Department of Computer Engineering,
34320, Istanbul, Turkey
Phone: +90 (212) 473 7070 - 17827
Email: okursun '@' istanbul.edu.tr

2. Betul ERDOGDU SAKAR, PhD.,
Bahcesehir University,
Department of Software Engineering,
34381, Istanbul, Turkey
Phone: +90 (212) 381 0589
Email: betul.erdogdu '@' eng.bahcesehir.edu.tr

3. M. Erdem ISENKUL, M.S.,
Istanbul University,
Department of Computer Engineering,
34320, Istanbul, Turkey
Email: eisenkul '@' istanbul.edu.tr

4. C. Okan SAKAR, PhD.,
Bahcesehir University,
Department of Computer Engineering,
34381, Istanbul, Turkey
Phone: +90 (212) 381 0571
Email: okan.sakar '@' eng.bahcesehir.edu.tr

5. Ahmet SERTBAS, PhD,
Istanbul University,
Department of Computer Engineering,
34320, Istanbul, Turkey
Email: asertbas '@' istanbul.edu.tr

6. Fikret GURGEN, PhD.,

Bogazici University,
Department of Computer Engineering,
34342, Istanbul, Turkey
Email: gurgen '@' boun.edu.tr

7. Sakir DELIL, M.D., PhD.,
Istanbul University,
CerrahpaÅYa Faculty of Medicine,
Department of Neurology,
34098, Istanbul, Turkey
Email: sakir.delil '@' ctf.edu.tr

8. Hulya APAYDIN, M.D., PhD.,
Istanbul University,
CerrahpaÅYa Faculty of Medicine,
Department of Neurology,
34098, Istanbul, Turkey
Email: hulya.apaydin '@' ctf.edu.tr

Donor:
C. Okan SAKAR, PhD.,
Bahcesehir University,
Department of Computer Engineering,
34381, Istanbul, Turkey
Phone: +90 (212) 381 0571
Email: okan.sakar '@' eng.bahcesehir.edu.tr

Data Set Information:

The PD database consists of training and test files. The training data belongs to 20 PWP (6 female, 14 male) and 20 healthy individuals (10 female, 10 male) who appealed at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University. From all subjects, multiple types of sound recordings (26 voice samples including sustained vowels, numbers, words and short sentences) are taken. A group of 26 linear and timeâ€“frequency based features are extracted from each voice sample. UPDRS ((Unified Parkinsonâ€™s Disease Rating Scale) score of each patient which is determined by expert physician is also available in this dataset. Therefore, this dataset can also be used for regression.

After collecting the training dataset which consists of multiple types of sound recordings and performing our experiments, in line with the obtained findings we continued collecting an independent test set from PWP via the same physicianâ€™s examination process under the same conditions. During the collection of this dataset, 28 PD patients are asked to say only the sustained vowels 'a' and 'o' three times respectively which makes a total of 168 recordings. The same 26 features are extracted from voice samples of this dataset. This dataset can be used as an independent test set to validate the results obtained on training set.

Further details are contained in the following reference -- if you use this dataset, please cite:
Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013

Training Data File:

Each subject has 26 voice samples including sustained vowels, numbers, words and short sentences. The voice samples in the training data file are given in the following order:

sample# - corresponding voice samples
1: sustained vowel (aaaâ€™|â€™|)
2: sustained vowel (oooâ€™|...)
3: sustained vowel (uuuâ€™|...)
4-13: numbers from 1 to 10
14-17: short sentences
18-26: words

Test Data File:

28 PD patients are asked to say only the sustained vowels 'a' and 'o' three times respectively which makes a total of 168 recordings (each subject has 6 voice samples) The voice samples in the test data file are given in the following order:

sample# - corresponding voice samples

1-3: sustained vowel (aaaâ€¦â€¦)

4-6: sustained vowel (oooâ€¦â€¦)

Attribute Information:

Training Data File:

column 1: Subject id

column 2-27: features

features 1-5: Jitter (local),Jitter (local, absolute),Jitter (rap),Jitter (ppq5),Jitter (ddp),

features 6-11: Shimmer (local),Shimmer (local, dB),Shimmer (apq3),Shimmer (apq5), Shimmer (apq11),Shimmer (dda),

features 12-14: AC,NTH,HTN,

features 15-19: Median pitch,Mean pitch,Standard deviation,Minimum pitch,Maximum pitch,

features 20-23: Number of pulses,Number of periods,Mean period,Standard deviation of period, features 24-26:

Fraction of locally unvoiced frames,Number of voice breaks,Degree of voice breaks

column 28: UPDRS

column 29: class information

Test Data File:

column 1: Subject id

column 2-27: features

features 1-5: Jitter (local),Jitter (local, absolute),Jitter (rap),Jitter (ppq5),Jitter (ddp),

features 6-11: Shimmer (local),Shimmer (local, dB),Shimmer (apq3),Shimmer (apq5), Shimmer (apq11),Shimmer (dda),

features 12-14: AC,NTH,HTN,

features 15-19: Median pitch,Mean pitch,Standard deviation,Minimum pitch,Maximum pitch,

features 20-23: Number of pulses,Number of periods,Mean period,Standard deviation of period,

features 24-26: Fraction of locally unvoiced frames,Number of voice breaks,Degree of voice breaks

column 28: class information

Relevant Papers:

Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013.

Isenkul, M.E., ErdoÄŸdu, B., Sakar, C.O., GÃ¼rmen, E., Delil, M.S., GÃ¼rmen, F., Sertbas, A., Kursun, O., Parkinson HastalÄŸÄ±n Ses Disfonilerinden TeÄŸhisi iÃŸin bir Ses VeritabanÄ± OlusturulmasÄ± ve Ã–lÃ¼mlerinin KullanÄ±lmasÄ±, 16. Biyomedikal MÃ¼hendisliÄŸi Ulusal ToplantÄ±sÄ± (BÃ–YOMUT 2011), Antalya, Turkey, October, 2011.

Citation Request:

Please cite the following paper if you use this dataset:

Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013.

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)