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.

**Information**

**Additional 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

**Additional Information**

Training Data File: column 1: Subject id colum 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 colum 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

**Машинный перевод без редактирования**

Обучающие данные принадлежат 20 пациентам с болезнью Паркинсона (БП) и 20 здоровым испытуемым. От всех испытуемых берется несколько типов звукозаписей (26).

Информация

дополнительная информация

База данных PD состоит из обучающих и тестовых файлов. Данные тренинга принадлежат 20 PWP (6 женщин, 14 мужчин)[ **PWP – люди с болезнью паркинсона**] и 20 здоровым людям (10 женщин, 10 мужчин), которые обратились в отделение неврологии медицинского факультета Серрахпаша Стамбульского университета. У всех испытуемых были взяты различные типы звукозаписей (26 образцов голоса, включая устойчивые гласные, цифры, слова и короткие предложения). Из каждого голосового сэмпла извлекается группа из 26 линейных и частотно-временных характеристик. Оценка UPDRS ((Единая шкала оценки болезни Паркинсона) для каждого пациента, которая определяется врачом-экспертом, также доступна в этом наборе данных. Следовательно, этот набор данных также может быть использован для регрессии. После сбора обучающего набора данных, состоящего из нескольких типов звукозаписей, и проведения наших экспериментов, в соответствии с полученными результатами, мы продолжили сбор независимого набора тестов от PWP с помощью того же процесса обследования у врача в тех же условиях. Во время сбора этого набора данных 28 пациентов с БП попросили произнести только устойчивые гласные "а" и "о" три раза соответственно, что составляет в общей сложности 168 записей. Те же 26 признаков извлечены из голосовых выборок этого набора данных. Этот набор данных можно использовать в качестве независимого тестового набора для проверки результатов, полученных на обучающем наборе. Более подробная информация содержится в следующей ссылке - если вы используете этот набор данных, пожалуйста, процитируйте: Эрдогду Сакар, Б., Исенкул, М., Сакар, К.О., Сертбас, А., Гурген, Ф., Делиль, С., Апайдин, Х., Курсун, О., "Сбор и анализ речи Паркинсона". Набор данных с несколькими типами звукозаписей", IEEE Journal of Biomedical and Health Informatics, том 17 (4), стр. 828-834, 2013 Учебный файл данных: У каждого испытуемого есть 26 образцов голоса, включая устойчивые гласные, цифры, слова и короткие предложения. Образцы голоса в файле обучающих данных приведены в следующем порядке: образец # - соответствующие образцы голоса 1: устойчивый гласный (aaa...) 2: устойчивый гласный (oooo...) 3: устойчивый гласный (uuu...) 4-13: числа от 1 до 10 14-17: короткие предложения 18-26: слова Файл тестовых данных: 28 пациентов с БП просят произнести только устойчивые гласные "а" и "о" три раза соответственно, что составляет в общей сложности 168 записей (у каждого испытуемого по 6 образцов голоса). Образцы голоса в файле тестовых данных приведены в следующем порядке: образец# - соответствующие образцы голоса 1-3: устойчивый гласный (аааа...) 4-6: устойчивый гласный (ооооо...)

Информация об атрибуте

дополнительная информация

Файл обучающих данных: столбец 1: Идентификатор объекта столбец 2-27: характеристики характеристики 1-5: Дрожание (локальное), Дрожание (локальное, абсолютное), дрожание (rap), дрожание (ppq5), дрожание (ddp), характеристики 6-11: Мерцание (локальное), Мерцание (локальное, dB), мерцание (apq3), Мерцание (apq5), мерцание (apq11), мерцание (dda), характеристики 12-14: Переменный ток,NTH, HTN, характеристики 15-19: Медианный шаг, средний шаг, стандартное отклонение, Минимальный шаг, максимальный шаг, характеристики 20-23: Количество импульсов, количество периодов, средний период, стандартный отклонение периода, признаки 24-26: Доля локально не озвученных кадров,количество речевых разрывов, степень речевых разрывов столбец 28: UPDRS столбец 29: информация о классе Файл тестовых данных: столбец 1: Идентификатор субъекта столбец 2-27: особенности особенности 1-5: Дрожание (локальное), Дрожание (локальное, абсолютное), дрожание (rap), дрожание (ppq5), дрожание (ddp), особенности 6-11: Мерцание (локальное), Мерцание (локальное, дБ), Мерцание (apq3), мерцание (apq5), мерцание (apq11), мерцание (dda), характеристики 12-14: Переменный ток,NTH, HTN, характеристики 15-19: Медианный шаг, средний шаг, стандартное отклонение, Минимальный шаг, максимальный шаг, характеристики 20-23: Количество импульсов, количество периоды, средний период, стандартное отклонение периода, характеристики 24-26: Доля локально не озвученных кадров, количество речевых разрывов, степень речевых разрывов столбец 28: информация о классе