

# Topological indices for heart rate variability

## Description of the dataset

This dataset contains the data of patients that experienced the stroke episode that was a basis for the results discussed in the paper *Persistent homology as a new method of the assessment of heart rate variability* by Grzegorz Graff, Beata Graff, Paweł Pilarczyk, Grzegorz Jabłoński, Dariusz Gąsecki, Krzysztof Narkiewicz, Plos One (2021), DOI: 10.1371/journal.pone.0253851.

### 1 File format

The file is in the csv (comma-separated values) format that is easy to import by most types of spreadsheet software. All the patients are identified by the string “STR” (stroke patients) followed by a 3-digit identifier (such as STR005, for example); note the non-contiguous numbering.

The file `indices_STR.csv` contains all the indices that were calculated for the purpose of their potential use in the paper; in particular, all the standard hrv (heart rate variability) indices and all the topological indices introduced in the paper are contained in this dataset. These indices are listed and described below.

### 2 General information in the file `indices_STR.csv`

There are several columns in the file `indices_STR.csv`; all of them are described in this and the next two sections.

The first few columns contain data regarding general information about the RR sequences, and are described below.

**Patient** – a unique identifier of each patient: “STR” followed by 3 digits identifies a patient suffering from a stroke

**number of original RRs** – the number of RRs read from the input data file; typically 512, unless the data file was shorter

**number of bad RRs** – the number of RRs that were not classified as “normal”; these are typically related to ectopic heartbeats

**number of low outliers** – the number of outliers that were removed, because they were below the lower threshold (described in the paper)

**number of high outliers** – the number of outliers that were removed, because they were above the higher threshold (described in the paper)

**number of normal RRs** – the number of normal RRs taken to the analysis (without the outliers if removed)

**total duration of all RRs** – the total duration of all the RRs, which equals the total duration of the entire sample (in seconds)

**total duration of normal RRs** – the total duration of normal RRs only (in seconds)

### 3 Standard hrv indices in the file `indices.csv`

These indices were computed with the `hrvanalysis` Python module; see <https://pypi.org/project/hrv-analysis/>. The brief descriptions in this section were extracted from the documentation of this module.

#### 3.1 Time domain features

**mean nni** – the mean value of the RR intervals

**sdnn** – the standard deviation of the RR intervals

**sdsd** – the standard deviation of differences between adjacent RR intervals

**pnni 50** – the proportion of the number of interval differences of successive RR intervals greater than 50 ms to the total number of RR intervals

**pnni 20** – the proportion of the number of interval differences of successive RR intervals greater than 50 ms to the total number of RR intervals

**rmssd** – the square root of the mean of the sum of the squares of differences between adjacent RR intervals

**median nni** – the median of the absolute values of the successive differences between the RR intervals

**range nni** – the difference between the maximum and minimum RR intervals

**cvsd** – the coefficient of variation of successive differences (equal to the `rmssd` divided by `mean nni`)

**cvnni** – coefficient of variation of the sequence of RR intervals (equal to the ratio of **sdnn** divided by mean **nni**)

**mean hr** – mean heart rate

**max hr** – maximum heart rate

**min hr** – minimum heart rate

**std hr** – standard deviation of heart rate

### 3.2 Geometrical features

**triangular index** – the hrv triangular index measurement, computed as the integral of the density distribution (the number of all RR intervals) divided by the maximum of the density distribution

### 3.3 Frequency domain features

**lf** – variance (= power) in hrv in the low frequency (0.04 to 0.15 Hz)

**hf** – variance (= power) in hrv in the high frequency (0.15 to 0.40 Hz)

**lf hf ratio** – the ratio lf/hf

**lfnu** – normalized lf power

**hfnu** – normalized hf power

**total power** – total power density spectral

**vlf** – variance (= power) in hrv in the very low frequency (0.003 to 0.04 Hz)

### 3.4 Non-linear domain features

**csi** – the Cardiac Sympathetic Index

**cvi** – the Cardiac Vagal Index

**Modified csi** – Modified CSI is an alternative measure in research of seizure detection, as proposed in a 2014 paper by Jesper Jeppesen et al.

**sd1** – the standard deviation of the projection of the Poincaré plot on the line perpendicular to the line of identity ( $y=x$ )

**sd2** – the standard deviation of the projection of the Poincaré plot on the line of identity ( $y=x$ )

**ratio sd2 sd1** – the ratio  $sd2/sd1$

**sampen** – the sample entropy of the sequence of RR intervals

## 4 Topological indices in the file `indices.csv`

`number of intervals` – the total number of persistence intervals

`longest interval` – the length of the longest persistence interval

`2/1 ratio` – the ratio of the length of the second longest persistence interval to the longest one

`3/1 ratio` – the ratio of the length of the third longest persistence interval to the longest one

`length mean` – the mean length of the persistence intervals

`length median` – the median of the lengths of the persistence intervals

`length stdev` – the standard deviation of the lengths of the persistence intervals

`length sum per rr` – the sum of the lengths of the persistence intervals divided by the number of RRs considered

`pers entropy` – persistent entropy

`normed entropy` – normed persistent entropy

`length threshold` – the actual threshold on the length of a persistence interval that is used to distinguish between short intervals and long intervals; it corresponds to 5% of the length of the longest persistence interval

`frac5%` – the quotient of the number of intervals longer than the threshold (defined above) to the number of all the intervals

`frac100` – the quotient of the number of intervals longer than 100 ms to the number of all the intervals

`frac200` – the quotient of the number of intervals longer than 200 ms to the number of all the intervals

`middle mean` – the mean of the middles of the intervals longer than the threshold (defined above)

`middle stdev` – the standard deviation of the middles of the intervals longer than the threshold (defined above)

`signal to noise` – the quotient of the sum of lengths of the intervals longer than the threshold (defined above) to the sum of lengths of the intervals whose lengths do not exceed the threshold

`birth mean` – the mean of the birth times of the intervals longer than the threshold (defined above)

**birth stdev** – the standard deviation of the birth times of the intervals longer than the threshold (defined above)

**death mean** – the mean of the death times of the intervals longer than the threshold (defined above)

**death stdev** – the standard deviation of the death times of the intervals longer than the threshold (defined above)

**dispersion** – spatial dispersion of the points in the persistence diagram measured by means of the determinant of the covariance matrix of the coordinates of the points; the entries of this matrix were divided by 1000 in order to obtain moderate values of the dispersion

**triangle width** – the width of the topological triangle, as described in the paper

**triangle height** – the height of the topological triangle

**triangle location** – the location of the topological triangle

**triangle proportion** – the proportion of the lengths of the side edges of the topological triangle

**triangle misalignment** – the misalignment indicator of the topological triangle