



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

ANALYSIS OF HEART RATE VARIABILITY VIA DATA MODELLING OF POINCARÉ PLOTS

ausgeführt am Institut für
Analysis and Scientific Computing
der Technischen Universität Wien

unter Anleitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.tech. Felix Breitenecker

durch

Matthias Hörtenhuber

Staudach 6
4072 Alkoven

Ort, Datum

Unterschrift

ABSTRACT

Heart diseases are amongst the most common causes of death in the industrialized world. Since the cardiological system is very complex and hard to capture in its entirety, researchers are looking for indicators of its health. One of these is the heart rate variability (HRV), i.e. the variation of the time interval between two heart beats. It reflects many physiological processes which influence the rhythm of the heart. Since these influences of the generation of heart beats are non-linear, researchers use a visualization tool, the Poincaré plot, which has its origins in chaos theory, to analyze HRV. This method, also called Lorenz plot, became popular in the last 20 years. It gives a simple visualization of the heart's beat-to-beat behavior and can be used for various applications, e.g., to predict the mortality of patients with myocardial infarction. Numerous data models exist in order to automatically quantify Poincaré plots.

The main objective of this master thesis is to implement 14 of the most common models in MATLAB[®] and improve them where possible. Afterwards these models are tested with respect to their ability to differentiate between pathological and non-pathological heart beat recordings, compare them to statistical HRV-measures and examine the models' dependences on each other.

The data are filtered via clustering algorithms and used in four different test cases. The first case is to test for data length sensitivity. Therefore, each model is applied to non-pathological and pathological data sets, with a stepwise reduction of their data length. The second case is an application of the models on pathological and non-pathological data sets, at a fixed data length, in order to do a deeper examination of them and their ability to differentiate between these data sets. For the third test, the models are used on data sets of subjects before and after arrhythmia treatment. The final test case is a comparison of younger and older healthy subjects via the data models.

Although not all implemented data models showed significant differences between the tested data sets, some passed all tests, including one, which was improved for this thesis. With the calculated correlations, the number of models which should be considered for further research can also be reduced.

KURZFASSUNG

Herzerkrankungen sind eine der häufigsten Todesursachen in den industrialisierten Ländern der Welt. Aufgrund der Komplexität des Herzkreislaufsystems suchen Forscher*innen nach Indikatoren für den gesundheitlichen Zustand des Systems. Einer von diesen Indikatoren ist die Herzratenvariabilität (HRV), d.h. die Veränderung der Zeitintervalle zwischen zwei Herzschlägen. Die HRV liefert Anhaltspunkte über den Zustand verschiedenster physiologischer Prozesse, die Einfluss auf den Herzrhythmus haben. Nachdem die Entstehung von Herzschlägen ein nichtlinearer Prozess ist, wurde in den letzten 20 Jahren vermehrt auf den Poincaré Plot zurückgegriffen, eine Visualisierungsmethode, die ihren Ursprung in der Chaostheorie hat und auch als Lorenz Plot bekannt ist. Es ist eine einfache Darstellungsweise der Änderung der Herzrate von einem Herzschlag zum nächsten und findet zahlreiche Anwendungen in der Medizin, u.a. zur Vorhersage des Mortalitätsrisikos bei Patient*innen nach einem Myokardinfarkt. Es existieren verschiedenste Datenmodelle zur automatischen Quantifizierung des Poincaré Plots.

Ziel dieser Diplomarbeit war es 14 der gebräuchlichsten Modelle in MATLAB[®] zu implementieren und falls möglich zu verbessern, sowie die Untersuchung ihrer Fähigkeit, verschiedene Arten von pathologischen und nicht pathologischen Herzraten zu unterscheiden, der Vergleich zu statistischen HRV-Größen und die Betrachtung der Abhängigkeit der Datenmodelle zueinander.

Die Daten des Poincaré Plots werden mittels Clusteralgorithmen gefiltert und anschließend in vier Testfällen verwendet. Der erste Test untersucht die Empfindlichkeit der Modelle auf die Datenlänge. Dazu wird die Unterscheidungsfähigkeit zwischen pathologischen und nicht pathologischen Daten bei sich schrittweise verkürzenden Datenlängen betrachtet. Im zweiten Testfall werden die Ergebnisse der Modelle für pathologische und nicht pathologische Daten bei einer fixen Datenlänge genauer untersucht. Für den dritten Test werden Herzratendaten von Arrhythmiepatienten vor und nach einer Antiarrhythmikabehandlung verwendet und erneut die Fähigkeit der Modelle zwischen den Datensätzen zu unterscheiden getestet. Im letzten Test werden die Ergebnisse der Modelle zwischen älteren und jüngeren gesunden Patient*innen untersucht.

Manche, jedoch nicht jedes der implementierten Datenmodelle zeigten signifikant unterschiedliche Ergebnisse zwischen den betrachteten Datensätzen. Darunter ist auch ein Modell, das im Zuge dieser Diplomarbeit verbessert wurde. Anhand der berechneten Korrelationen lässt sich die Anzahl an zu berücksichtigenden Modellen für künftige Untersuchungen reduzieren.

ACKNOWLEDGMENTS

The thesis in front of you is not the result of my work alone, but of the support by numerous people during my five years of studies.

First of all I would like to thank my parents whose support I could always rely on.

Secondly, I would like to thank my two advisors at work, Christopher Mayer and Martin Bachler, whom I could bring all of my questions and would always come back with answers and suggestions.

I am also grateful to Felix Breitenacker for the supervision and for the good times with the ARGESIM group.

I also have to thank my university colleagues without whom I would have not been able to bear the content and stress of my studies.

Out of the last group of people I want to say a special thanks to Antonia Lichtenegger. Without her as a continuous source of motivation this thesis and my studies would never have come to a conclusion in the time it did.

Matthias Hörtenhuber

CONTENTS

1	PREAMBLE	3	
1.1	Scope of Work	3	
1.2	Methodical Approach	4	
1.3	Structure of the Thesis	4	
2	INTRODUCTION AND BACKGROUND	7	
2.1	Physiological Background	8	
3	METHODS	11	
3.1	The Poincaré Plot	11	
3.2	Poincaré Plot Indices	12	
3.2.1	Cluster indices	13	
3.2.2	Asymmetry Indices	15	
3.2.3	3D Histogram Indices	17	
3.2.4	Dispersion Index	19	
3.2.5	Mechanical Indices	20	
3.2.6	Extension Index	21	
3.2.7	LT Indices	22	
3.2.8	Ellipse Fitting Indices	22	
3.2.9	Range Index	23	
3.2.10	Complex Correlation Measure	24	
3.2.11	Contour Index	25	
3.2.12	Statistical Indices	26	
3.3	Data	27	
3.3.1	Non-Pathological Data	27	
3.3.2	Pathological Data	27	
3.3.3	CRIS-Data	28	
3.4	Signal Filtering	29	
3.5	Statistical Tests	31	
3.6	Tests Cases	32	
4	RESULTS	35	
4.1	Test Case I: Data Length	35	
4.2	Overview of Test Cases II-IV	39	
4.3	Parameters of Distributions	41	
4.4	Test Case II: Non-Pathological vs. Pathological Data	48	
4.5	Test Case III: Pre- vs. Post-Antiarrhythmic Treatment Data	49	

4.6	Test Case IV: Non-Pathological Data from Younger vs. Older Subjects	49
4.7	Correlation	50
5	DISCUSSION	55
5.1	Cluster Indices	55
5.2	Asymmetry Indices	55
5.3	3D Histogram Indices	56
5.4	Dispersion Indices	57
5.5	Mechanical Indices	57
5.6	LT-Indices	58
5.7	Ellipse Fitting Indices	59
5.8	Statistical Indices	60
5.9	Ungrouped Indices	60
5.10	Limitations	61
6	CONCLUSION	63
7	APPENDIX	65
7.1	Correlation	65

PREAMBLE

According to a report by the The European Society of Cardiology, heart failure is a leading cause of death in the EU and is on the rise due to an increasing age of the population [37]. Since an early diagnosis of heart conditions leads to more successful treatments, researchers look for markers of heart diseases [10].

More than 30 years ago heart rate variability (HRV) was introduced as such a measure [1]. HRV is the variation of the time interval between consecutive heartbeats. It highly depends on the extrinsic regulation of the heart rate, i.e. the time interval between two beats and reflects changes in the balance of the different regulatory systems, including the autonomous nervous system [41].

Research on HRV has attracted considerable attention in the fields of psychology and behavioral medicine. It has its origin in the search for non-invasive correlates of injury severity which can be extracted from available signals in order to discover new cardiac biomarkers [5]. These signals are usually ones that are routinely measured and include sources like a photoplethysmogram or the electrocardiogram (ECG) [3]. Studies also show a connection between the balance of the autonomic nervous system measured with HRV and cardiovascular diseases [2].

In studies of HRV, both time- and frequency-domain measures are typically used by practitioners and researchers [1, 2]. Since the influences of the beat generation are also non-linear [33], a visualization tool originating in chaos theory, the Poincaré plot, became a popular tool to analyze HRV in the last 20 years [46]. In order to automatically capture and compare the characteristics of a Poincaré plot, numerous data models or Poincaré plot indices have been proposed.

1.1 SCOPE OF WORK

In most cases in the literature a Poincaré plot index is only compared to a standard set of indices, if it is compared at all. Smith et al. gathered in [43] many different indices and tested for correlations if applied on very short signals, but they did not compare their individual differentiation ability. The scope of this thesis is to implement

the existing Poincaré plot quantification methods in MATLAB[®], to improve them (if possible) and to statistically compare their ability to differentiate heart rate data based on their Poincaré plots and the statistical connection of the indices with each other.

Therefore, test cases were chosen in order to answer the following research questions:

- How sensitive are the Poincaré plot indices regarding the length of the data?
- Are there Poincaré plot indices, which can differentiate between non-pathological and not further specified pathological heart rate data?
- Are there Poincaré plot indices, which can differentiate between heart rate data with and without arrhythmia?
- Are there Poincaré plot indices, which show a difference between the heart rate of younger and older healthy subjects?
- Which Poincaré plot indices are dependent on each other?

1.2 METHODOLOGICAL APPROACH

First, a literature research was carried out to get an overview of the existing Poincaré plot indices and their results. Next, the found generally applicable Poincaré plot indices were implemented using MATLAB[®]. During this process their definitions were checked for errors and certain plausible improvements were applied. For the next step data were obtained from the PhysioNet database [16] to establish test cases for the research questions (see section 3.3) and the implemented methods were applied on them. Thereafter, statistical tests were applied to examine the calculated Poincaré plot indices with respect to the research questions.

1.3 STRUCTURE OF THE THESIS

Chapter 2 gives a short introduction into Poincaré plots, HRV and the physiological and the technological background. In chapter 3, the Poincaré plot and its variation are defined, as well as all implemented Poincaré plot indices. Furthermore, this chapter contains a description of the used data, of the applied data filtering and the

constructed test cases. Chapter 4 contains a detailed description of the test case results. These are discussed with respect to each other and to the literature in chapter 5, which also contains remarks on the limitations of the presented work. Chapter 6 contains summarized answers to the research questions, suggestions for the application of the results, as well as for further research in this area.

INTRODUCTION AND BACKGROUND

*Chaos is order in a mask [...] dancing
in the heart of everything.*

Terry Pratchett - Thief of Time

In 1887 Oscar II, King of Sweden and Norway, called out a prize for the solution of the n-body problem, which is posed as follows: *Given n mass points, which attract each other according to Newton's laws, how can you calculate their positions over time.* In the case no one could solve it, they prize would go to the best contribution. At the end Henri Poincaré, a french mathematician, was awarded for proofing that the stated problem could not be solved, since a fluctuation below measurement accuracy would lead to widely diverging end-states. Karl Weierstrass, who was part of the jury, remarked: *"This work cannot indeed be considered as furnishing the complete solution of the question proposed, but that it is nevertheless of such importance that its publication will inaugurate a new era in the history of celestial mechanics"* [22]. This prediction should only tell half the truth, since Poincaré's submission was the starting point of chaos theory, which nowadays finds appliciations in almost all scientific fields.

It is not clear if the Poincaré plot, a visualization method for chaotic process, goes actually back to Henri Poincaré or if it got its name to honor him. One of the first found occurrences is in [28], a paper by Edward Lorenz about non-periodic flow published in 1963. Edward Lorenz, who worked on weather prediction and coined the term "butterfly-effect", made very important contributions to the chaos theory, therefore the Poincaré plot is also sometimes called Lorenz plot. In this work it will always be called Poincaré plot, which is the more established name in medicine. The importance of the Poincaré plot lays in its ability to extract patterns from many chaotic processes with a very simple transformation [24].

In the 1980s, researchers found that the beat generation in human hearts is a chaotic process [15, 33]. Therefore Woo et al. proposed in [51] to use Poincaré plots for the analysis of HRV.

The interest for HRV goes back more than thirty years, when a connection between the autonomic nervous system and the cardiovascular mortality was found [2]. Heart rate variability can be analyzed via time-domain and frequency-domain measures [1, 2], but due to the chaotic features of HRV, non-linear measures capture them in a more effective way [1].

2.1 PHYSIOLOGICAL BACKGROUND

The following section is based on [6], if not further specified.

In a healthy heart the electrical impulse for one beat originates in one center in the right atrium, called the sinoatrial node. Without any influences by regulatory systems, e.g. the nervous or hormonal system, the cells in the node will discharge upwards of 60 times per minute. It is strongly innervated by the parasympathetic nervous system as well as the sympathetic nervous system fibers. This signal moves via both atria to the atrioventricular node, exciting the atria on its way. This node acts as a delay to separate the contraction of the atria from ventricular contraction. After the atrioventricular node the signal goes into the bundle of His, which splits up for the right and the left ventricle. Thereafter, the signal is further split up via the Purkinje fibers, which spread across both ventricles.

These steps can be indirectly observed in an electrocardiogram (ECG). This method measures the polarization and depolarization of the heart muscle via electrodes applied to the skin either at the thorax or at the extremities. The spreading of the depolarization and therefore the spreading of the electrical signal throughout the atria is visible in the P-wave (see figure 1). The sharp QRS-complex corresponds with the depolarization of the ventricles. The following T-wave shows the repolarization of the ventricular muscle tissue.

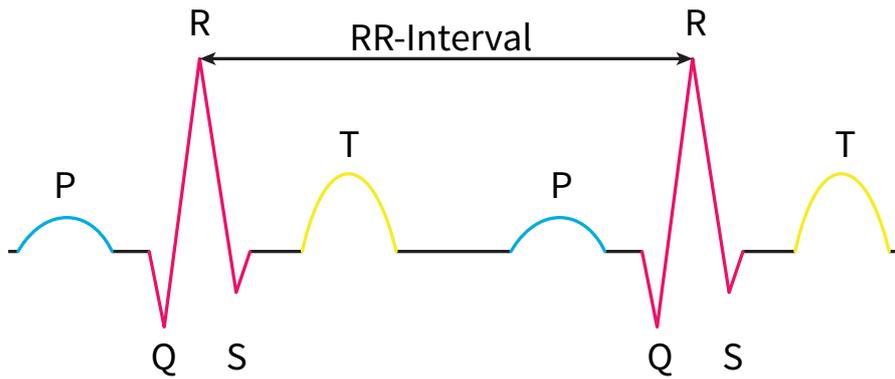


Figure 1: Schematic representation of an ECG-signal with labeled P- and T-waves (blue and yellow) and QRS-complexes (magenta).

The time between two R-peaks, which correspond to two heart beats, is used as the basis of heart rate variability and is called RR-interval (see figure 1). The time between two heart beats can also be measured via the blood pressure, since blood flow is pulsatile, corresponding to each contraction of the heart.

Since the sinoatrial node is not the only part of the heart, which is able to generate excitation signals, beats with different origin can occur. Beats originating in the sinoatrial node are called sinus or sinusoidal beats, whereas beats generated from other sources are called extrasystoles or ectopic beats. A small number of extrasystoles occurs also in healthy hearts [36], e.g., in [8, 38] up to 10 extrasystoles per hour are reported for healthy subjects.

But due to different reasons (e.g., an ectopic focus, i.e. a group of heart muscle cells creating a non-sinusoidal signal), extrasystoles can occur more frequently, which are then diagnosed as arrhythmias. The different forms of extrasystoles are named based on the origin of the electronic signal and its site of effect (e.g., ventricular = signal originates in the ventricles and leads to their abnormal contraction; supraventricular = abnormal contraction of the ventricles from a signal not originating within them) and their duration and frequency (e.g., tachycardia = short sequence of ectopic peaks; fibrillation = long running sequence of ectopic peaks).

HRV reflects the different influences on the sinus beat frequency, besides arrhythmias. Figure 2 (adapted from [50]) shows the main regulatory mechanisms and influences, e.g., the heart rate typically increases during inhalation and decreases during exhalation [18].

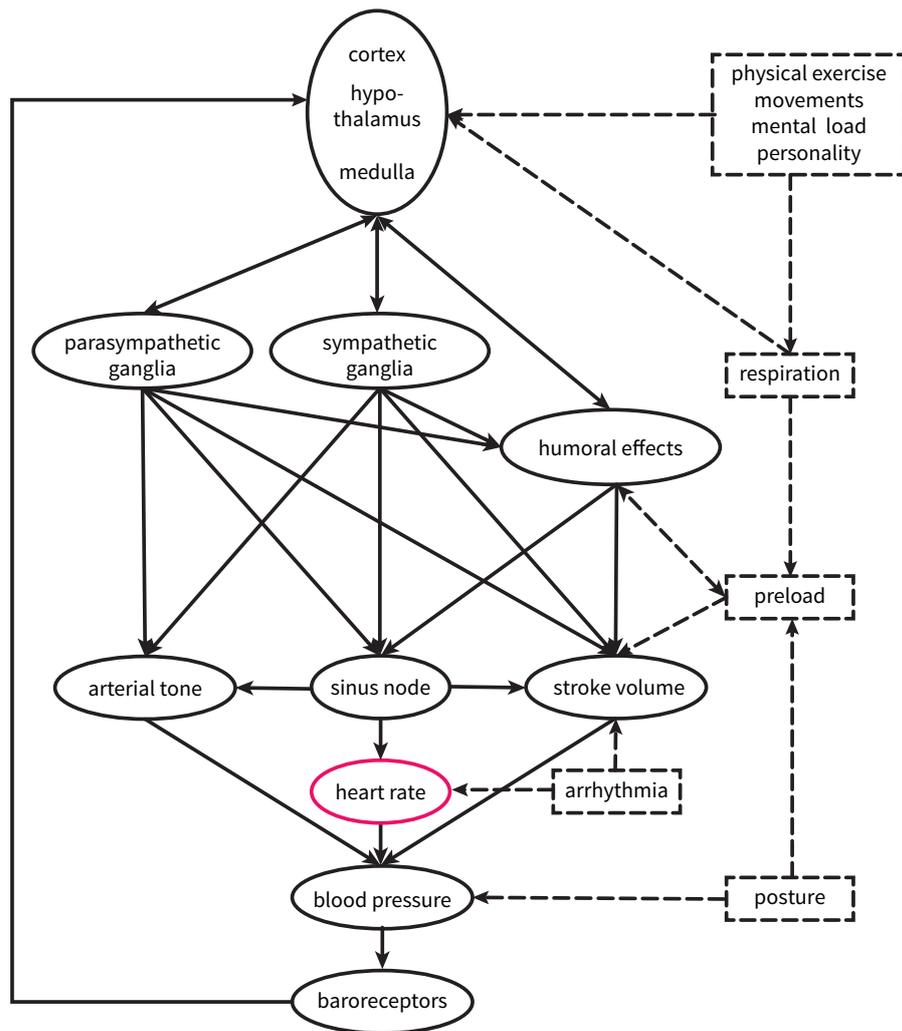


Figure 2: Simplified representation of the regulatory systems and influences of heart beat generation (adapted from [50]).

These various regulators lead to the variable length between two heart beats, i.e. HRV. A failure of one or more regulatory systems, e.g., due to a pathology, thus leads to less variability. Therefore, the general rule of thumb for HRV is: Healthy subjects have a higher HRV compared to unhealthy ones [45].

METHODS

Every plot in this section, except for figure 10 and 11, was created with the recording *f2o10* of the Fantasia Database. For figure 10 the recording 881 of the MIT-BIH Supraventricular Arrhythmia Database was used and figure 11 is based on the recording 116 of the MIT-BIH Arrhythmia Database. For a detailed description of these databases see section 3.3.

3.1 THE POINCARÉ PLOT

Given a data set of N RR-intervals $\{RR_1, \dots, RR_N\}$ a Poincaré plot is obtained by plotting $\mathbf{RR}_1 := \{RR_1, \dots, RR_{N-1}\}$ against $\mathbf{RR}_2 := \{RR_2, \dots, RR_N\}$. A typical Poincaré plot of a non-pathological 2 hour long, unfiltered heart rate recordings is shown in figure 3.

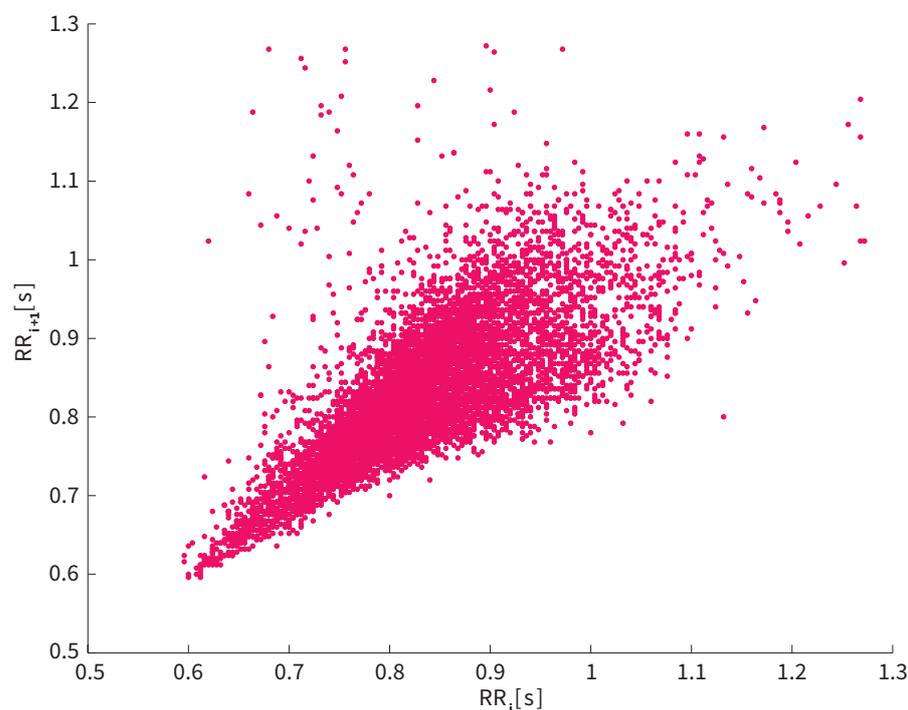


Figure 3: A typical unfiltered Poincaré plot of a non-pathological heart rate data set.

Poincaré plots of heart rate data from healthy patients are typically shaped like a comet, as in figure 3, or like a torpedo or cigar [13]. Plots associated with pathologies do not have a shape as wide spread due to a loss in irregularity or they are multi-centered because of multiple rhythms [13, 46].

There exist some extensions of this definition of the plot, e.g., the lagged Poincaré plot, which is a generalization of the original version. A Poincaré plot with lag m is created by comparing one data point to the m -th after it, i.e., by mapping $\mathbf{RR}_1 := \{RR_1, \dots, RR_{N-m}\}$ against $\mathbf{RR}_{m+1} := \{RR_{m+1}, \dots, RR_N\}$. So the standard Poincaré plot can be interpreted as a Poincaré plot with lag 1.

A further approach is the three dimensional Poincaré plot, where the third axis comes from data points with an additional lag, as shown in figure 4.

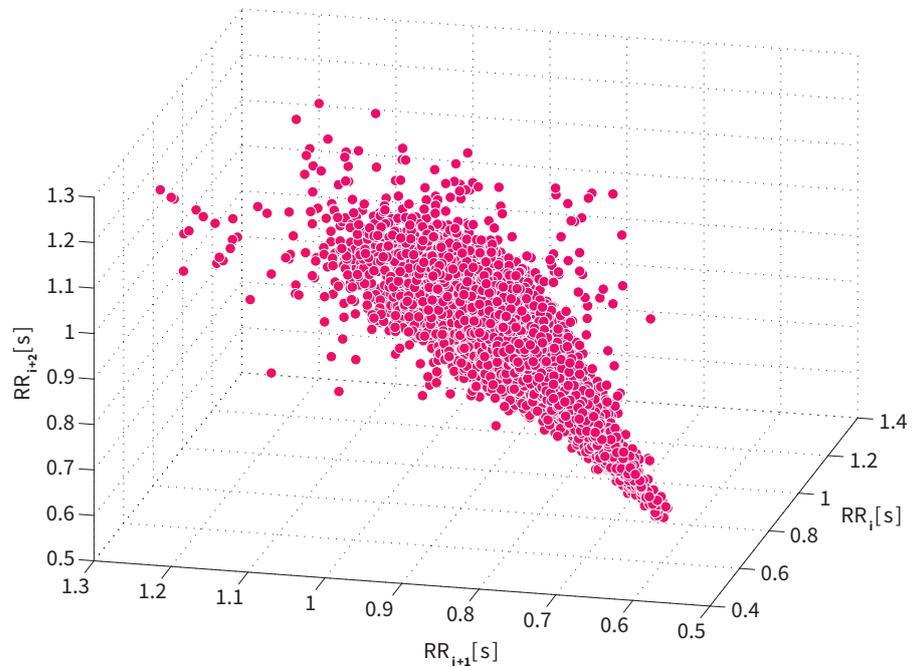


Figure 4: An unfiltered 3D Poincaré plot of a non-pathological heart rate data set with lag 1 and lag 2.

3.2 POINCARÉ PLOT INDICES

In the following section different methods to quantify Poincaré plots are presented.

3.2.1 Cluster indices

The search for similar objects and gathering them into groups is defined as clustering. In the case of Poincaré plots, the aim is to group the data points of different heart rhythms, which appear connected to each other but separated from other data points in the plot. After trying different algorithms (k-means, single linkage and mean-shift), the DBSCAN-Algorithm was chosen, because it does not require an a-priori number of clusters and shows a high robustness against noise.

DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise and was proposed by Ester et al. in [14]. The algorithm needs the parameters ε and *MinPts* as inputs, where ε is a neighborhood threshold and *MinPts* is the minimum number of points in one cluster. DBSCAN distinguishes three types of data points:

- *Core points*: These have *MinPts* or more different points in their ε -environment.
- *Density reachable points*: These have at least one other data point in their ε -environment, but less than *MinPts*.
- *Noise*: These are neither core points nor density reachable points.

The following pseudocode describes the algorithm:

```
function=DBSCAN(D,  $\varepsilon$ , MinPts)
for (all unvisited points P in dataset D)
    mark P as visited
    N=getNeighboringPoints(P,  $\varepsilon$ )
    if(sizeof(N) < MinPts)
        mark P as NOISE
    else
        C = next cluster
        add P to cluster C
        for (P' in N)
            if(P' is not yet member of any cluster)
                recursiveExpandCluster(P', C,  $\varepsilon$ , MinPts)
            end
        end
    end
end
end
end
```

```

function=recursiveExpandCluster(P, C, ε, MinPts)
add P to cluster C
if(P is not visited)
    mark P as visited
    N = getNeighbors(P, ε)
    if(sizeof(N) >= MinPts)
        for{ P' in N}
            if{ P' is not yet member of any cluster}
                recursiveExpandCluster(P', C, ε, MinPts)
            end
        end
    end
end
end
end

```

One of the difficulties lies in the choice of ϵ . If it is too small, the algorithm overclusters, i.e., it separates visibly connected clusters, or it underclusters if ϵ is too large, i.e., it merges visibly unconnected clusters.

Therefore, a refinement of DBSCAN, the Ensemble-DBSCAN (EDB-SCAN) proposed by Xia et al. in [52] was applied.

This algorithm runs DBSCAN r -times iterating ϵ equidistantly from ϵ_{\min} to ϵ_{\max} , where

$$\epsilon_{\min} := D_4^{\text{mean}} - \frac{D_4^{\text{mean}} - D_4^{\text{min}}}{8}, \quad \epsilon_{\max} := D_4^{\text{mean}} + \frac{D_4^{\text{max}} - D_4^{\text{mean}}}{8}.$$

The variable D_4 stands for the set of distances between the data points and their fourth nearest neighbor, D_4^{mean} is its mean value and D_4^{min} and D_4^{max} are the minimal and maximal value of the set. The result of each iteration is saved in the co-association matrix A , by adding 1 to each entry $A_{i,j}$, if the i -th and the j -th data point are in the same cluster and 0 otherwise. After all iterations the co-association matrix is normalized via element-wise division by r . The final clusters are then constructed by using a voting method, described in the following pseudocode:

```

assign the first data point to the first cluster
for (all other points of D)
     $A_{\max} = \max_{j=1, \dots, i-1} A_{i,j}$ 
    if( $A_{\max} < 0.5$ )
        assign current point to a new cluster
    else
        assign current point to cluster of  $D(k)$ ,

```

```

        where  $A_{i,k} = A_{max}$ 
    end
end

```

Afterwards clusters with less data points than a given threshold are considered as noise. The threshold for this categorization is set so that clusters consisting of presumably non-pathological extrasystoles are ignored. Therefore a number of 10 extrasystoles per hour is used as a threshold, based on the findings in [8, 38].

The number of clusters as well as the ratio between the second largest and the largest cluster were used as indices *NumOfClusters* and *ClusterSizeRatio*. If only one cluster was detected *ClusterSizeRatio* was set to 0.

3.2.2 Asymmetry Indices

One of the fundamental laws governing all systems in the physical world is the thrive towards the minimum energy state, i.e., towards its equilibrium. Therefore, a biological system has to function far away from the equilibrium in order to utilize energy for itself, with a larger energy gain the further away they move from the equilibrium [11], usually corresponding to the complexity of the system. This is also true for the heart beat generation [11]. This goes along with a unidirectionality or time irreversibility. Failures of regulatory systems, e.g., due to pathologies, therefore show themselves in a loss of irreversible behavior [11]. The irreversibility is also visible as a geometrical asymmetry in the Poincaré plots [24] with the line of identity as the symmetry axis. There exist different approaches to measure it. Each of the following asymmetry indices is a number between 0 and 100, with 50 denoting perfect symmetry.

Porta's Index

In [40], Porta et al. counted the points above the line of identity and compared them to the number of points below. Therefore, Porta's index, *PI*, is defined as follows. Let U be defined as $U := \{(RR_i, RR_{i+1}) : RR_i < RR_{i+1}\}$, i.e the points above the line of identity and L the set of points below it, i.e., $L := \{(RR_i, RR_{i+1}) : RR_i > RR_{i+1}\}$, see figure 5. Then:

$$PI := \frac{\#U}{\#U + \#L} \cdot 100,$$

where $\#X$ is the cardinality of a set X .

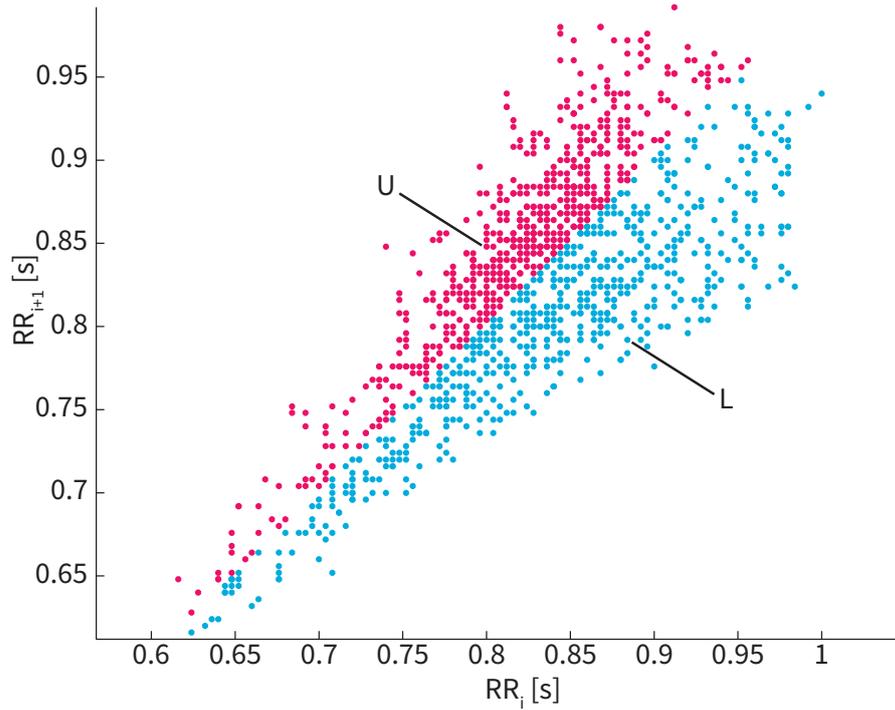


Figure 5: Visualization of the separation of a filtered Poincaré plot along the line of identity to measure asymmetry, with the set U in red and the set L in blue, as used for Porta's and for Guzik's index.

Guzik's Index

Guzik et al. proposed in [19] to measure asymmetry by comparing the distances to the line of identity for the points above and below it. The distance to the line of identity is defined as:

$$d(RR_i, RR_{i+1}) := \frac{|RR_i - RR_{i+1}|}{\sqrt{2}}.$$

Let U and L be defined as in Porta's Index. Then Guzik's Index, GI , is defined as:

$$GI := \frac{\sum_{i \in U} d(RR_i, RR_{i+1})}{\sum_{i \in U \cup L} d(RR_i, RR_{i+1})} \cdot 100.$$

Karmakar-Porta's and Karmakar-Guzik's Indices

Karmakar et al. redefined those two indices in [24] by not only looking at two but at three consecutive points and separating them into

three groups, one for increasing points (I), one for decreasing points (D) and one for constant points (C). They were defined as follows:

$$\begin{aligned}
I &:= \{(RR_i, RR_{i+1}) : RR_i < RR_{i+1} < RR_{i+2} \vee \\
&\quad RR_i \geq RR_{i+1} < RR_{i+2} \vee \\
&\quad RR_i > RR_{i+1} \leq RR_{i+2}\}, \\
D &:= \{(RR_i, RR_{i+1}) : RR_i > RR_{i+1} > RR_{i+2} \vee \\
&\quad RR_i \leq RR_{i+1} > RR_{i+2} \vee \\
&\quad RR_i < RR_{i+1} \geq RR_{i+2}\}, \\
C &:= \{(RR_i, RR_{i+1}) : RR_i = RR_{i+1} = RR_{i+2}\}.
\end{aligned}$$

Based on this, Porta's index was redefined to:

$$PI_p := \frac{\#I}{\#I + \#D} \cdot 100.$$

Guzik's index was redefined to:

$$GI_p := \frac{\sum_{(RR_i, RR_{i+1}) \in I} d(RR_i, RR_{i+1})}{\sum_{(RR_i, RR_{i+1}) \in I \cup D} d(RR_i, RR_{i+1})}. \quad (1)$$

Karmakar's redefinitions can be interpreted as an asymmetry measure for 3D-Poincaré plots. Then, I , D and C can be interpreted as three dimensional sets I_{3D} , D_{3D} and N_{3D} with elements from $\{\mathbf{RR}_1, \mathbf{RR}_2, \mathbf{RR}_3\}$. Therefore, an altered GI_p , called GI_{p3D} , was also examined, which is defined as in (1), but with the three dimensional distance to the line of identity:

$$d_{3D}(RR_i, RR_{i+1}, RR_{i+2}) := \left(\frac{(RR_i - RR_{i+1})^2}{3} + \frac{(RR_{i+1} - RR_{i+2})^2}{3} + \frac{(RR_i - RR_{i+2})^2}{3} \right)^{\frac{1}{2}}.$$

Then GI_{p3D} is defined as:

$$GI_{p3D} := \frac{\sum_{(RR_i, RR_{i+1}) \in I_{3D}} d_{3D}(RR_i, RR_{i+1}, RR_{i+2})}{\sum_{(RR_i, RR_{i+1}) \in I_{3D} \cup D_{3D}} d_{3D}(RR_i, RR_{i+1}, RR_{i+2})}.$$

3.2.3 3D Histogram Indices

Hnatkova et al. showed in [21] that the commonly used 2D Poincaré plot can hide some information, since only its shape is considered as the main quantified feature in many cases. This is the

case if two plots can have the same shape with different beat distributions inside the shape, e.g., one plot has one accumulating area, while the other has two disconnected ones. They proposed in [21] to add its density as an additional dimension, i.e., to count the pairs of $(\mathbf{RR}_1, \mathbf{RR}_2)$ with the same values and therefore appear as one point in the 2D plot. This new kind of Poincaré plot can also be interpreted as a 3D histogram of a Poincaré plot.

Density Index

In order to derive an index from this plot, Hnatkova et al. proposed in [21] to approximate the density function of the plot. For this all points inside a rectangular area around the highest peak of the plot, i.e., the point of highest density, are counted, while iterating through different sizes of the rectangular environment, as shown in figure 6. Afterwards the area under this function is calculated and used as index *area*. According to [21] density functions of Poincaré plots with a sharp peak have a smaller area than those of plots with the same shape but more equally distributed data. To extract further information about the density function, the ratio of the maximal slope of the density function and its mean as was implemented as the additional index *slope*.

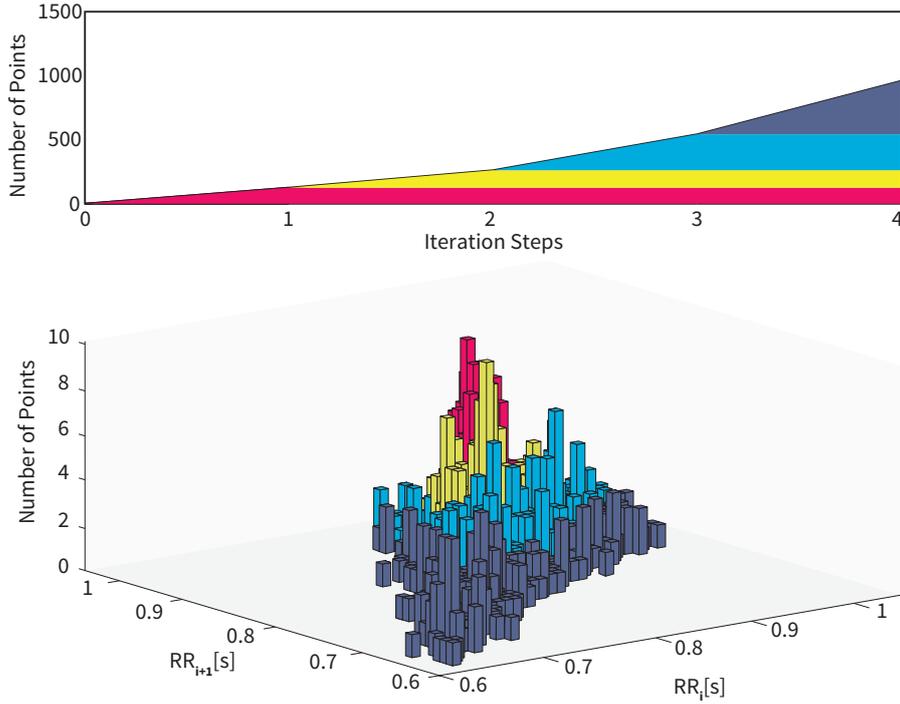


Figure 6: Approximated density function (top) with the corresponding 3D histogram (bottom) of a filtered non-pathologic Poincaré plot. Starting from the highest peak, the number of points in a certain neighborhood is summed up. The diameter of this neighborhood is enlarged with each iteration, as shown color coded in the histogram and the density function.

Peak Indices

In [32], Marciano et al. proposed besides the extension index also an index for the 3D histogram of the Poincaré plot (see figure 6). But instead of a density function as used for the density indices (see section 3.2.3), they used the number of peaks above a certain threshold, NP , as well as the sum of the distances of these peak points to the line of identity in the 2D Poincaré plot, DP , as indices. Their choice for the threshold was a peak height of 8.

3.2.4 *Dispersion Index*

Schechtman et al. quantified in [42] the shape of the Poincaré plot by its longitudinal dispersion at two positions. In order to extract a measurement, they rotated the plot by $\phi = -\frac{\pi}{4}$ and measured its width in vertical direction at the 10th and the 90th percentile of the Poincaré plots horizontal length. Furthermore, they excluded 10% of

the top and bottom points at this positions. The two widths, $width_{10}$ and $width_{90}$, shown in figure 7, as well as their difference, $\Delta width$, were calculated as indices. Similar to the extension index (see section 3.2.6), the same dispersion indices of the unrotated plot were also calculated, i.e., its 80% width in direction of the y-axis at the 10th and the 90th percentile, $rotwidth_{10}$ and $rotwidth_{90}$, as well as their difference, $\Delta rotwidth$.

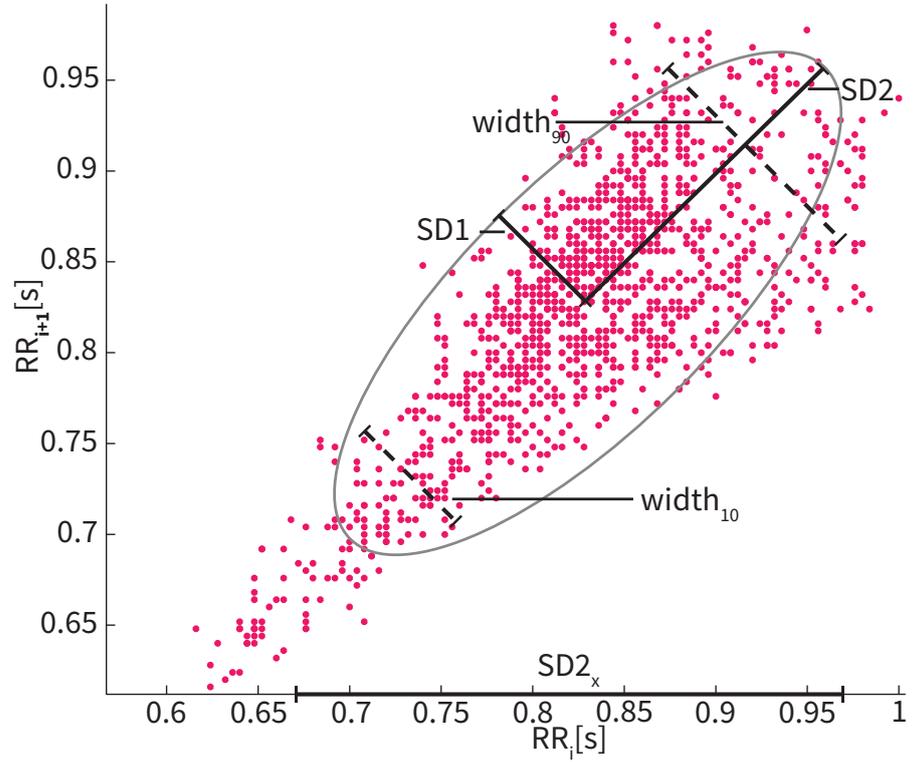


Figure 7: A filtered non-pathological Poincaré plot with the ellipse fitting indices $SD1$ and $SD2$ as the axis lengths, as well as $SD2_x$ from $HiSD$ and the dispersion indices $width_{10}$ and $width_{90}$.

3.2.5 Mechanical Indices

Inertia Index

Marciano et al. interpreted in [32] the Poincaré plot as a mass distribution and used the axis lengths of its ellipse of inertia as indices.

More exactly they calculated the two axes lengths, A and B , using the formulae:

$$A := \frac{\text{Var}(\mathbf{RR}_1) + \text{Var}(\mathbf{RR}_2)}{2} + \sqrt{\frac{\text{Var}(\mathbf{RR}_1) - \text{Var}(\mathbf{RR}_2)}{2} + \text{Cov}(\mathbf{RR}_1, \mathbf{RR}_2)},$$

and

$$B := \frac{\text{Var}(\mathbf{RR}_1) + \text{Var}(\mathbf{RR}_2)}{2} - \sqrt{\frac{\text{Var}(\mathbf{RR}_1) - \text{Var}(\mathbf{RR}_2)}{2} + \text{Cov}(\mathbf{RR}_1, \mathbf{RR}_2)}.$$

Multipole Index

Similar as for the inertia index in section 3.2.5, Lewkowicz et al. looked in [27] at the Poincaré plot as a two dimensional body, where each point has unit mass. Therefore, their distribution can be expressed by moments in a gravitational potential field. They used the entries of the quadrupole and the octupole tensors as indices, as well as the ratio of the kurtosis of the y and x-coordinates. The quadrupoles and the octupoles can be calculated with:

$$\begin{aligned} Q_{xx} &= \sum_{i=1}^{N-1} (2RR_i^2 - RR_{i+1}^2), & T_{xxx} &= \sum_{i=1}^{N-1} (6RR_i^3 - 9RR_i RR_{i+1}^2), \\ Q_{yy} &= \sum_{i=1}^{N-1} (2RR_{i+1}^2 - RR_i^2), & T_{yyy} &= \sum_{i=1}^{N-1} (6RR_{i+1}^3 - 9RR_{i+1} RR_i^2), \\ Q_{zz} &= -Q_{xx} - Q_{yy}, & T_{xxy} &= \sum_{i=1}^{N-1} (36RR_i^2 RR_{i+1} - 9RR_{i+1}^3), \\ & & T_{xyy} &= \sum_{i=1}^{N-1} (36RR_{i+1}^2 RR_i - 9RR_i^3), \\ & & T_{xzz} &= \sum_{i=1}^{N-1} (-9RR_{i+1}^2 RR_i - 9RR_i^3), \\ & & T_{yzz} &= \sum_{i=1}^{N-1} (-9RR_i^2 RR_{i+1} - 9RR_{i+1}^3). \end{aligned}$$

3.2.6 *Extension Index*

Marciano et al. proposed an index which uses the Poincaré plot's extension in the direction of the first and of the second axis. In [32] they

measured the plot extension by projecting it onto the x-axis and calculating the distance, d_x , between the leftmost and the rightmost value. Afterwards the x-position, \tilde{x} , where the plot has the maximal extension in y-direction. Those two measures are shown in figure 8. Their proposed index is defined as:

$$ext := \frac{\tilde{x}}{d_x} \cdot 100.$$

3.2.7 *LT Indices*

The LT Indices presented by Toichi et al. in [47] is similar to ext . They rotated the Poincaré plot by $\phi = -\frac{\pi}{4}$ and measured its elongation in direction of the first (L) and the second axis (T), as shown in figure 8. The ratio of these two lengths, $ratioLT$, as well as the logarithm of their product, LT , were used as indices in [47].

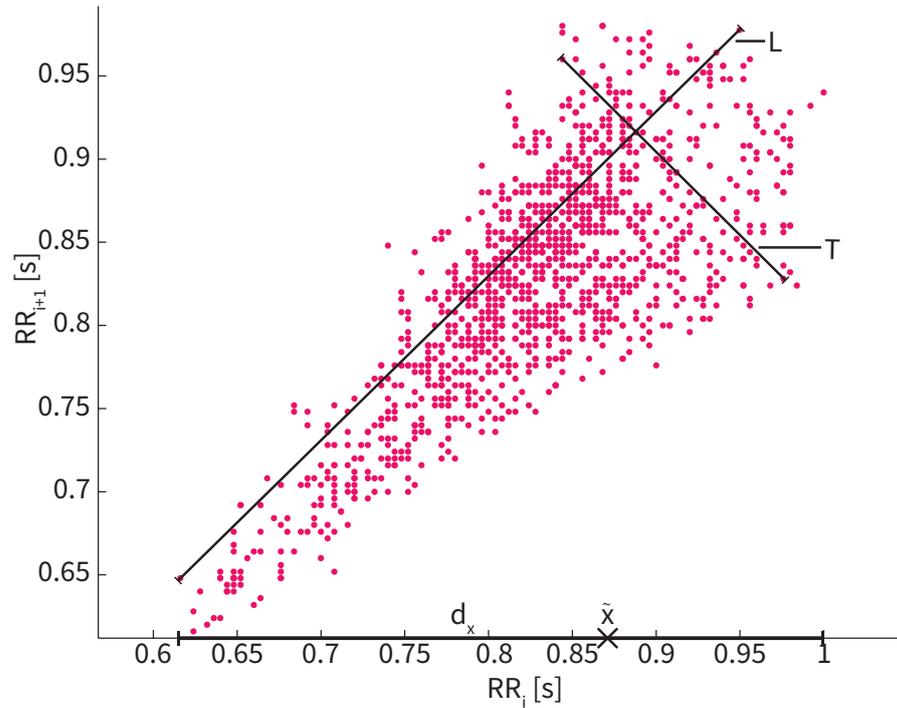


Figure 8: The measurements for ext , \tilde{x} and d_x and of the LT indices L and T of a filtered non-pathological Poincaré plot.

3.2.8 *Ellipse Fitting Indices*

One of the most common methods to quantify a Poincaré plot is to fit an ellipse to it, by rotating its axes by $\phi = \frac{\pi}{4}$ with respect to the axes

of the plot. The center of the ellipse is the mean value of the plotted data. The length of the major axis, $SD2$, and the length of the minor axis of the ellipse, $SD1$, are the standard deviations in each direction as shown in figure 7.

These two lengths are widely used indices to quantify a Poincaré plot. According to Tulppo et al. [48] $SD1$ is related to short-term and $SD2$ to long-term heart rate variability. Brennan et al. showed in [9] that these values can be derived from time domain measures and therefore, they do not make any use of the plot's non-linear features. The length of the the main ellipse axis $SD2$ and the length of the minor axis $SD1$ can be calculated in the following way, according to [9], with $SDNN$ as the the standard deviation of all RR -intervals and $SDSD$ as the standard deviation of $(RR_i - RR_{i+1})$:

$$SD1 = \frac{SDSD}{\sqrt{2}} \quad \text{and} \quad SD2 = \sqrt{2 \cdot SDNN^2 - \frac{1}{2} SDSD^2}. \quad (2)$$

Tulppo et al. showed in [48], that $SD1$ and $SD2$ are strongly correlated with frequency and time domain measures, which isn't the case for the ratio $SD1/SD2$. Therefore this ratio was also considered as an additional index in the following tests.

According to Lerma et al. equation (2) holds only true for stationary processes, which is not the case for HRV [26]. Therefore $TSD1$, $TSD2$ and $ratioTSD$ were also implemented for this thesis according to the original definition of $SD1$ and $SD2$.

Hirose et al. [20] proposed an index, which is very similar to $ratioSD$ and $ratioTSD$. Their proposed index can be written as:

$$HiSD := \frac{SD2_x}{SD1}, \quad (3)$$

where $SD2_x$ is the standard deviation along the x-axis as shown in figure 7.

3.2.9 Range Index

Moraes et al. described in [35] an alternative Poincaré plot. Instead of comparing RR_n to RR_{n+1} they used RR_n vs. $RR_{n+1} - RR_n$. Similar to the density index they also counted the number of overlapping points and showed these counts in z-direction. From this plot they constructed an index by multiplying the longest elongation in x-direction (P_2) with the longest elongation in y-direction (P_3) and the mean slope at the point of highest density (P_1), as shown in figure 9. In [35], the mean slope was calculated as the slope between

10 and 90% of the maximum density in the intersecting plane which contains the maximum density point and is orthogonal to the x-axis. This was implemented by calculating the slope between the peak and the first point with a density of 2 points or lower to the left and to the right of it, in the same plane as in the original definition. Afterwards the mean of the absolute value of these two slopes is used as the mean slope. Their final formula for the range index is:

$$range := (100 - 2P_1) \cdot P_2 \cdot P_3.$$

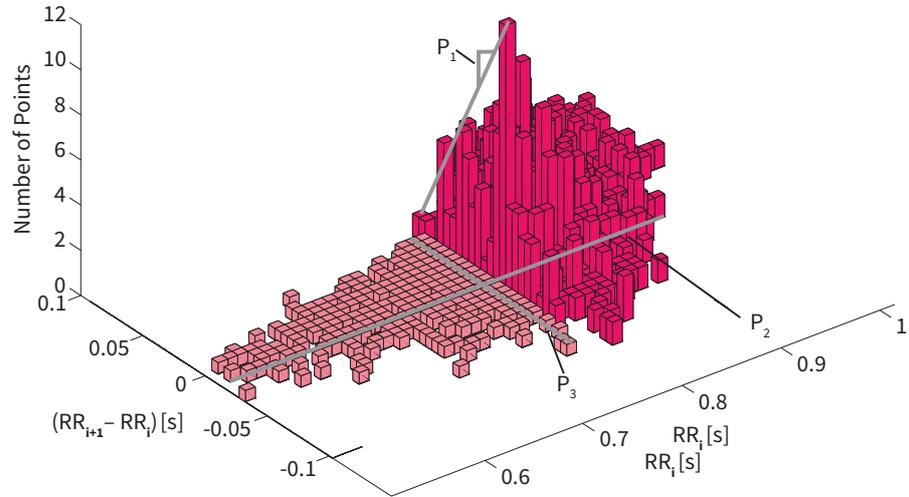


Figure 9: The alternative Poincaré plot according to Moraes et al., with marked mean slope (P_1), maximal elongation in x-direction (P_2) and in y-direction (P_3). The histogram in front of the plane with the highest density point is depleted for visualization reasons.

3.2.10 Complex Correlation Measure

In order to take the temporal beat-to-beat behavior into deeper consideration Karmakar et al. proposed the Complex Correlation Measure (CCM) in [25]. To include this behavior, they used a moving window with a length of three points. Let (x_i, y_i) , (x_{i+1}, y_{i+1}) and (x_{i+2}, y_{i+2}) be the three points of a Poincaré plot in the i -th window. The area A_i of the triangle formed by these three points, is calculated by:

$$A_i = \frac{1}{2} \begin{vmatrix} x_i & y_i & 1 \\ x_{i+1} & y_{i+1} & 1 \\ x_{i+2} & y_{i+2} & 1 \end{vmatrix}. \quad (4)$$

If all three points are on one line, it follows from formula (4) that $A_i = 0$. If they are arranged counter clock-wise $A_i > 0$, and $A_i < 0$, if they are orientated clock-wise. Then, the *CCM* is defined as:

$$CCM := \frac{1}{\pi \cdot SD1 \cdot SD2} \sum_{i=1}^{N-2} A_i.$$

The normalization $\pi \cdot SD1 \cdot SD2$ is the area of the enveloping ellipse.

3.2.11 *Contour Index*

None of the indices found in literature tries to capture the exact profile of the Poincaré plot. Therefore a novel index was additionally implemented. To get information about the plot's shape, the distances between the mean value of (RR_i, RR_{i+1}) and the points farthest away in every direction is calculated. This is implemented by catching all points, where the connecting line to the mean value has the same slope. Afterwards, the distances between all points on this line and the mean value are calculated. The area under this maximal distance function is used as the contour index. The correspondence with the area of the Poincaré plot is shown in figure 5.9.

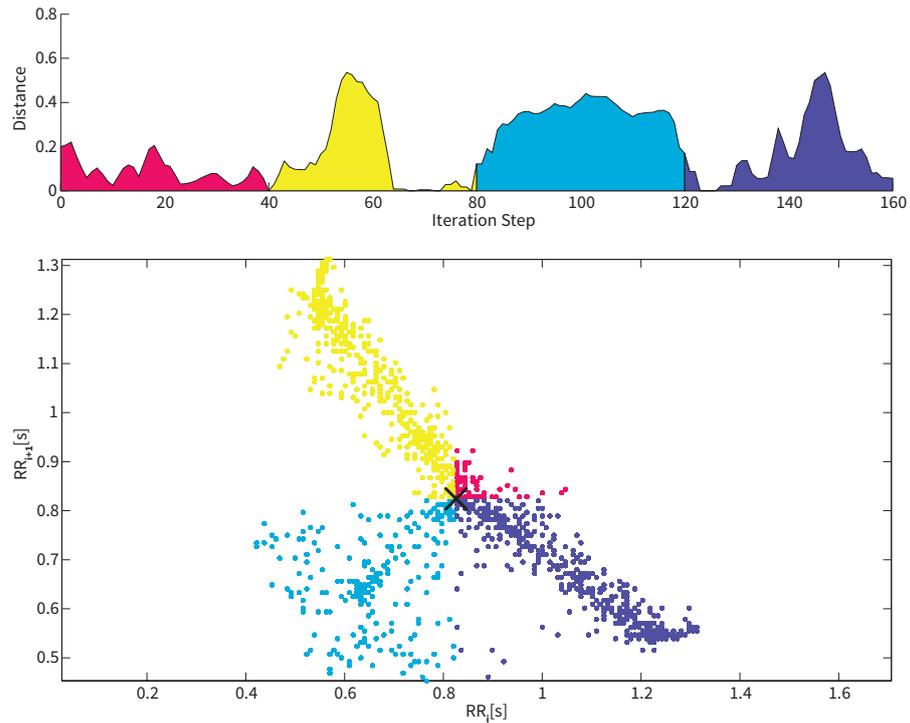


Figure 10: The distance function (top) and the corresponding Poincaré plot (bottom). The plot is scanned counter-clockwise and the mean value is marked with a black cross. The colors show the parts of the Poincaré plot with their equivalent area under the curve.

3.2.12 Statistical Indices

In order to compare the Poincaré indices to time domain indices, the following commonly used statistical measures were used as defined in [9] and [47]:

- *pNN50*: The percentage of sinusoidal RR-Intervals longer than 50ms.
- *SDSD*: The standard deviation of the successive differences, i.e., the standard deviation of $RR_i - RR_{i+1}$.
- *SDNN*: The standard deviation of sinusoidal RR-Intervals.
- *CV*: The coefficient of variation, which is defined as the standard deviation of the data divided by the data's mean value.

3.3 DATA

All data used to test the indices were taken from Physionet.org [16], a free-access, on-line archive of physiological signals. They are described in detail in this section.

To create a control group, databases specified as non-pathological were combined into one database.

3.3.1 *Non-Pathological Data*

- This database includes the *Fantasia Database* consisting of 120 minute long recordings of twenty young (10 men and 10 women; 21 - 34 years old) and twenty elderly (10 men and 10 women; 68 - 85 years old) healthy subjects with ECG digitized at 250 Hz [23].
- *The Normal Sinus Rhythm RR Interval Database*, which consists of 54 ECG recordings, each one approximately 24 hours long, is also part of this database. It contains heart rate data of subjects with normal sinus rhythm (30 men, aged 28.5 to 76, and 24 women, aged 58 to 73) digitized at a sample frequency of 128 Hz [16].
- Furthermore, it includes the *Massachusetts Institute of Technology (MIT) - Boston's Beth Israel Hospital (BIH) Normal Sinus Rhythm Database*, which consists of 18 long-term recordings (5 men, aged 26 to 45, and 13 women, aged 20 to 50) sampled at 128 Hz [16].

This resulted in a total database size of 112 recordings.

3.3.2 *Pathological Data*

To test if pathological heart rate data has an effect on Poincaré plot indices, the following databases were combined to one.

- *The Congestive Heart Failure RR Interval Database*, consisting of 29 ECG recordings, each approximately 24 hours long and

with a sampling frequency of 128 Hz, of subjects aged 34 to 79 (8 men and 2 women; gender unknown for the remaining subjects) with congestive heart failure (NYHA classes I, II, and III) [4].

- Furthermore it consists of the *MIT-BIH Arrhythmia Database*, which contains 48 half-hour recordings, sampled with a frequency of 360 Hz, from 47 subjects (25 men aged 32 to 89 years and 22 women aged 23 to 89 years) [34]. It contains a set of randomly chosen recordings and 25 recordings especially chosen to include examples of uncommon but clinically important arrhythmias recorded at the BIH Arrhythmia Laboratory [34].
- Finally the *MIT-BIH Supraventricular Arrhythmia Database* was also included. It consists of 78 not further specified half-hour ECG recordings of patients with supraventricular arrhythmia, digitized at a 128 Hz [17].

In total 151 recordings of pathological heart rates were used.

3.3.3 *CRIS-Data*

Additionally to the pathological and the non-pathological databases, the indices were tested with data obtained from the *CAST RR Interval Sub-Study Database* [44]. This database was created for the Cardiac Arrhythmia Suppression Trial (CAST), which was originally started to analyze the effect of suppressing ventricular arrhythmias by antiarrhythmic drugs after myocardial infarction on the survival rate [12].

This database consists of 1543 24-hour RR-interval recordings from 809 subjects. For almost all subjects, heart rate data before and during the antiarrhythmia treatment is available. In total, 1462 records for 731 subjects (599 male and 132 female, 20-79 years old) have been used. To avoid daytime-dependent variations, all the data samples were taken from a time window starting at 6 p.m. 75 subjects have been excluded because they had just baseline or just on-therapy data available, three subject were excluded additionally, because there were no recordings for the chosen time window.

3.4 SIGNAL FILTERING

Abnormal data points (e.g., movement artifacts or ectopics beats) are excluded in order to only measure sympathovagal activity. In almost all cases in literature (e.g. [19–21, 24, 25, 29, 32, 35, 40, 42, 47, 48]) data is filtered before usage, to apply the methods only on the sinus beat cluster, if it is distinguishable. This filtering was usually done by hand.

For the following tests, the data was filtered automatically via clustering. All beats longer than 2.5 seconds and their corresponding counterparts were removed, before applying the EDBSCAN-Algorithm (see 3.2.1) to detect the clusters, since they can reasonably be interpreted as artifacts.

The sinusoidal beat cluster was then chosen as the one closest to the mean value of (RR_i, RR_{i+1}) . To reduce assignment errors a correction by adding 0.01s to both coordinates of the mean was implemented. This can be justified by two reasons. First, most of the errors occur because of arrhythmias with a shorter RR-interval length, therefore these beats move the mean closer to zero, away from the actual sinus beats. Second, no case was observed where the mean value was above the sinusoidal cluster, which would be the case for very atypical heart rate of a high amount of single slow beats in connection with a very fast sinus beat.

Figure 11 shows the result of the clustering for a sample data set with arrhythmia.

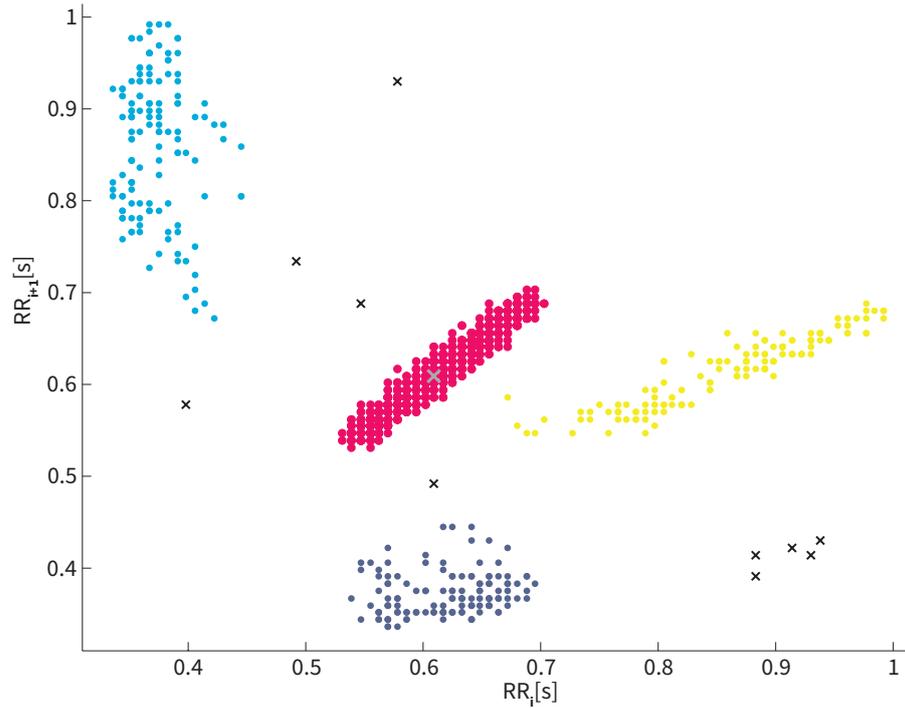


Figure 11: Clustered Poincaré plot of a subject with arrhythmia. The filter algorithm selected the cluster in red as the sinus beat cluster with the mean value marked as the gray cross. The other clusters and the noise (black crosses) are not considered for the calculation of the Poincaré plot indices, except for *NumberOfClusters* and *ClusterSizeRatio*.

Since Poincaré plots represent the relation between two consecutive beats, the filtering can not be done by deleting one beat interval in \mathbf{RR}_1 and the same in \mathbf{RR}_2 , but the preceding one in \mathbf{RR}_1 and the following one in \mathbf{RR}_2 has to be deleted as well, as shown in figure 12 [39].

If a 3D Poincaré plot is used, as it is for GI_{p3D} (see description in section 3.2.2)), for one erroneous RR-interval the preceding two have to be deleted in \mathbf{RR}_1 , the one preceding and the one following in \mathbf{RR}_2 and the two subsequent intervals in \mathbf{RR}_3 , as shown in figure 12 as well.

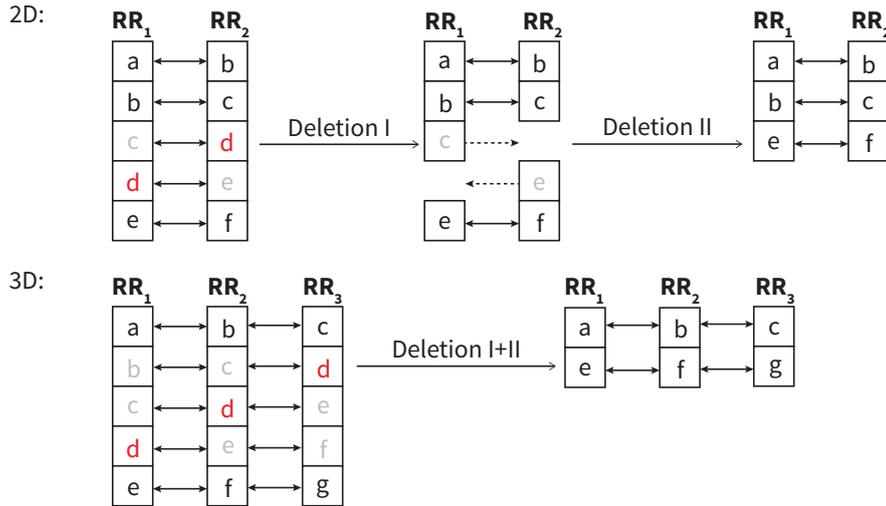


Figure 12: Deletion of the incorrect RR-interval "d" (Deletion I) and the corresponding counterparts "c" in RR_1 and "e" in RR_1 (Deletion II) for a 2D Poincaré plot. In the case of a 3D Poincaré plot for an erroneous interval "d", the preceding intervals "b" and "c" have to be deleted in RR_1 , "c" and "e" in RR_2 and "e" and "f" in RR_3 .

3.5 STATISTICAL TESTS

The following procedure was used to test for a difference between indices in the tests according to [49].

Because most of the results for indices were not normally distributed a Wilcoxon rank sum test was applied to calculate the p-value.

Since the pre- and post-treatment recordings of one subject in the CRIS-Database are dependent and most of their calculated indices did not have a normal distribution the Wilcoxon signed rank sum test was used to examine the differences between the data sets.

A test outcome was declared significant for $p < 0.05$ and very significant for $p < 0.01$.

To test the correlation between the individual indices Spearman's correlation coefficient r was used, since almost all of the indices were not normally distributed. A correlation was declared strong, if $|r| > 0.85$. At the same time the significance of the correlation was calculated.

All tests were done with built-in functions in MATLAB.

3.6 TESTS CASES

The following tests were done to answer the research questions in section 1:

Test Case I: Data Length Sensitivity

In order to see how Poincaré indices behave with respect to the data length, recordings of the pathological and the non-pathological database are compared with different lengths. Starting from 1100 filtered data points, the data sets were reduced by steps of 50 data points and the Poincaré indices of the remaining points are calculated and statistically tested for differences.

Both cluster indices were excluded from this test, because the signal was shortened after filtering and therefore no new clusters can be found most of the time. If they are, this is due to overclustering.

The starting data length of 1100 data points was chosen because it is the upper length limit for short-term recordings [7] and does not exclude too many recordings, since the shortest unfiltered signal consists of 1431 data points.

Test Case II: Non-Pathological vs. Pathological Heart Beats

For the second test 1429 data points are taken from the middle of each recording of the pathological database. These are filtered and the Poincaré plot indices for the first 1000 points are calculated. This length was chosen for the same reasons as in test case I. The same was done for the recordings in the non-pathological database. Afterwards statistical tests were applied, as described in section 3.5, in order to find indices which can differentiate between the two databases. Finally, the correlation of these indices was calculated as well.

Test Case III: Pre- vs. Post-Antiarrhythmic Treatment

To examine if the indices can differentiate between data before and after arrhythmia suppressing medication, 2000 data points recorded around 6 p.m. from the CRIS-Database were filtered. More data points were included in this step than in the first two test cases, because due to the arrhythmias more points were expected to be filtered away. Afterwards the first 1000 data points of each recording were taken to calculate the Poincaré plot indices. The indices of the recordings before and after treatment for each subject were tested for significant differences and correlations.

Test Case IV: Younger vs. Older Non-Pathological Subjects

To test if Poincaré indices are age-dependent, the non-pathological database was split into people of age 63 and less, and people older than this. The threshold of 63 years was chosen, since it is the mean age of the subjects in the non-pathological database. Afterwards the same filtering and statistical tests were applied as in test case II.

RESULTS

4.1 TEST CASE I: DATA LENGTH

Figures 13-16 show the results of the test for the sensitivity of the indices with respect to the data length.

It should be noted that for this case the p-values are comparable, since only the Wilcoxon rank sum test was used for all indices, as not all tested data sets were distributed normally.

As shown in figure 13-A, Guzik's Index GI shows significant differences for signals with more than 1000 data points and gets worse the shorter the signal is. For signals shorter than 300 points the differences gain significance without ever surpassing the significance threshold of $p < 0.05$. Karmakar-Porta's Index PI_p differentiates significantly for data lengths of 950 until 300 points. The slope index shows only significant differences for signals with 250 data points. The contour index has more significant differences the shorter the signal is and surpassing the significance threshold for signal lengths shorter than 400 data points.

The index $width_{10}$ shows no clear trend with respect to the data length, but differentiates significantly for lengths of 100 and 150 points.

Figure 13-B also shows no clear trend for the behavior of the dispersion indices $rotwidth_{10}$ and $rotwidth_{90}$ with changing data lengths, as both pass the significance threshold only once at 550 and at 400 data points, respectively. The same holds true for $width_{90}$ and $\Delta width$, but with lower variability and they have significant differences between non-pathological and pathological data sets for most of the signal lengths, which is completely lost by $\Delta width$ for data shorter than 200 points.

The two inertia indices A and B show no variation over the data length, but can not be computed for signals shorter than 250.

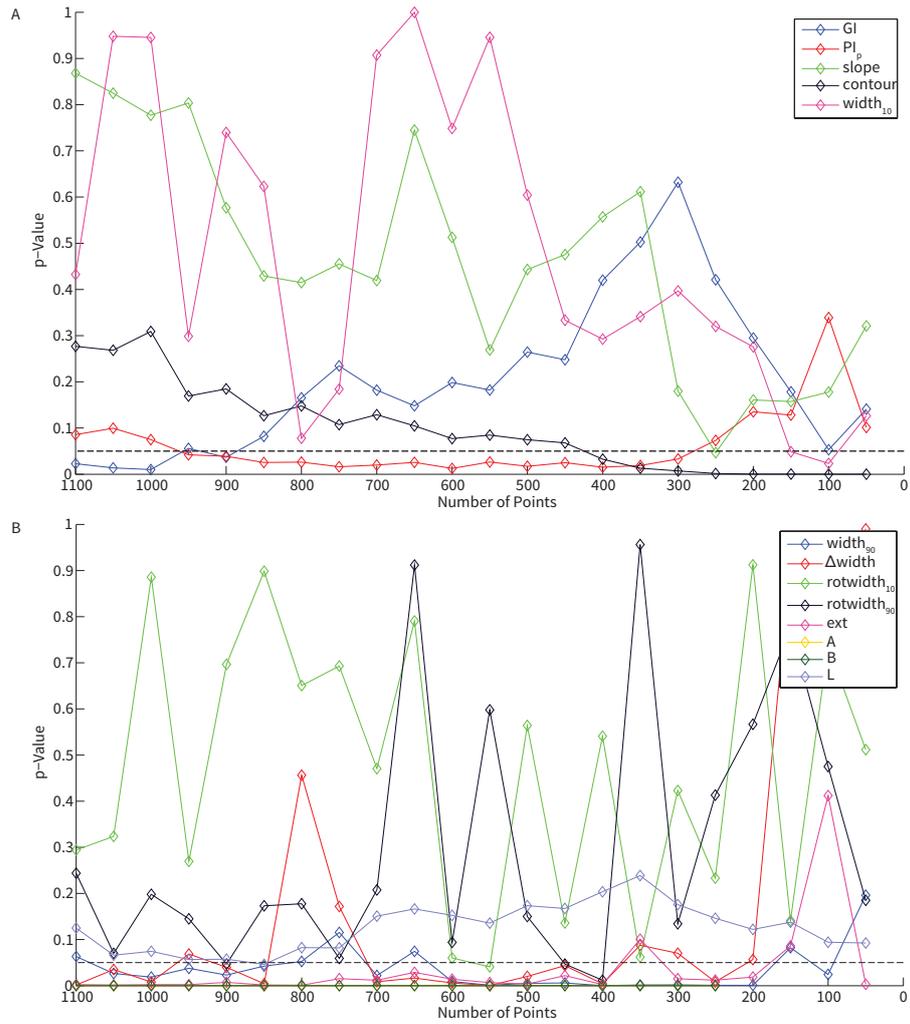


Figure 13: Indices with at least one but not all p-values below the significance threshold of 0.05 (dashed line).

In figure 14-A the octupoles T_{yyy} and T_{xxy} show no trend and are above the significance threshold for most of the signal lengths, while the differences of T_{yzz} are almost always highly significant. This characteristic is lost abruptly for signals of length 500 and 450. The peak index NP shows more significant differences for shorter signals and surpasses the significance threshold at 500 data points. The same holds true for the other peak index DP as shown in figure 14-B. The differences for the range index are not significant most of the time, but tend to be more significant for shorter signals, passing the significance threshold for 100 data points and less. The ellipse fitting index SDI is always differentiating significantly, except for very long and very short signals. The statistical index $SDNN$ shows significant behavior only for sig-

nal lengths of 1050 or longer, which gets worse with shorter data. *SDSD* has almost always significant differences between pathological and non-pathological databases, except for signals of length 1100.

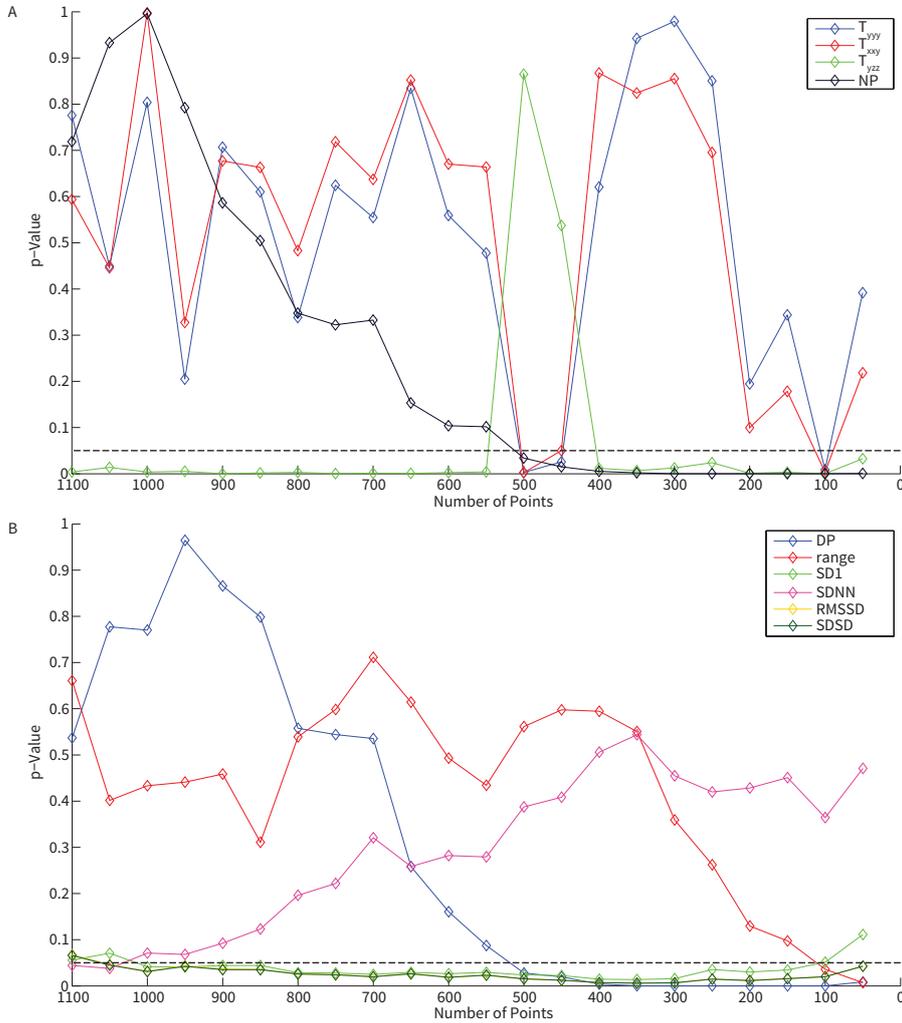


Figure 14: Additional indices with at least one, but not all p-values below the significance threshold of 0.05 (dashed line).

Figure 15 shows that *PI* has a very significant differentiation behavior ($p < 0.01$) for almost all signal lengths shorter than 1000 data points except for signals of length 100. GI_{p3D} on the other hand loses its very significant differences rather fast for signals with less than 200 data points.

The differences of the index *T* have a trend of getting more significant for shorter signals.

The ellipse fitting index *TSD2* has very significant differences for signals of less than 1100 data points.

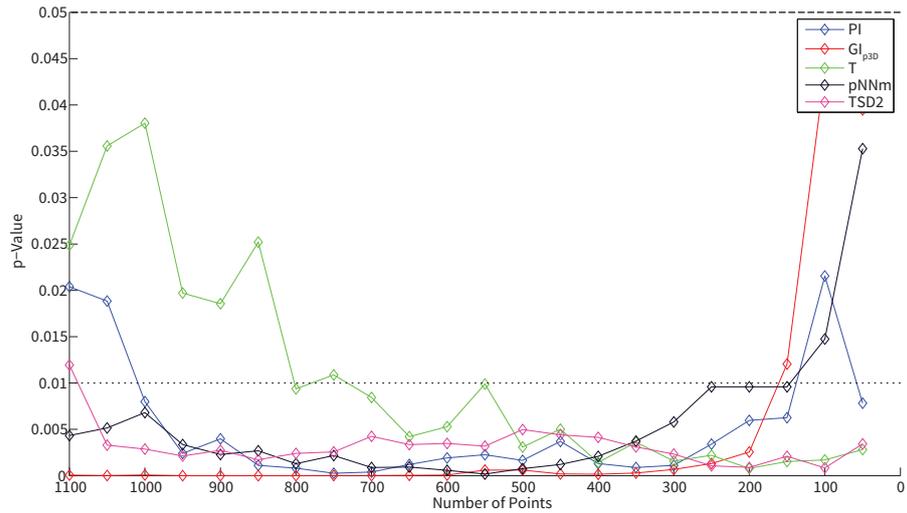


Figure 15: Indices with significant, i.e., $p < 0.05$ (dashed line), and for some data lengths, very significant differences, i.e., $p < 0.01$ (dotted line), between pathological and non-pathological data sets.

In figure 16, one can see that the differentiation behavior of all indices, which differentiate very significantly, hardly change with respect to length of the signal. The only exceptions are minor fluctuations of *SD2*, which still is very significant throughout all lengths.

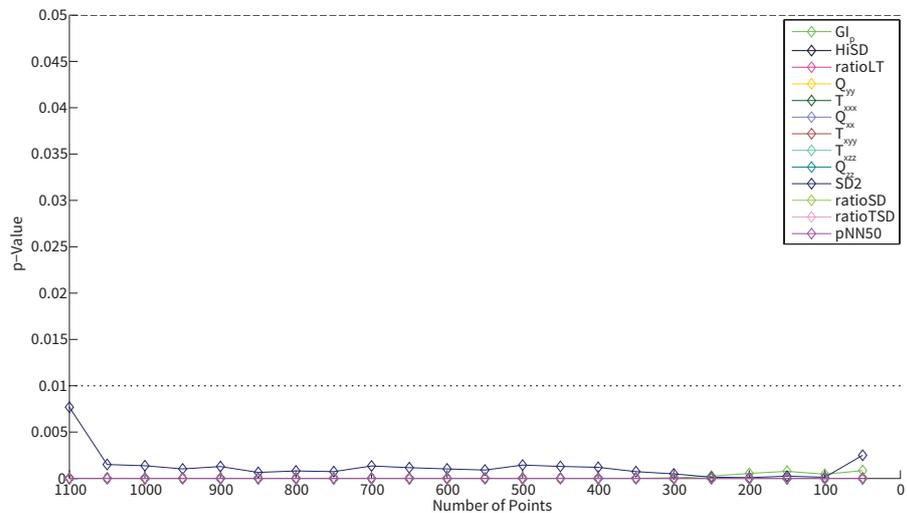


Figure 16: Indices with very significant differences, i.e., $p < 0.01$ (dotted line), between pathological and non-pathological heart rate data for all data lengths.

Figure 17 shows the reaction of different indices to varying signal lengths, without ever reaching significant differences between non-pathological and pathological data.

The index *LT* tends to differentiate more significant for shorter signals.

The indices *CCM* and *kurtosis* show a sudden increase in differences for signals shorter than 200 points.

The opposite is true for *area*, which has an abrupt loss of significance for signals shorter than 150 points.

The index *ratioAB* loses significance in differences for data lengths between 1100 and 800 data points, which is slowly gained back for the following data lengths. Note that it has no values for data lengths below 250 points, because *A* and *B* could not be calculated at this length.

The statistical index *CV* shows less significant differences with some variations throughout the whole test.

The differences of the dispersion index $\Delta rotwidth$ varies without any visible trend for all data lengths.

The index *TSD1* is stable with respect to the data length, with a small loss of significance for signals with less than 300 points.

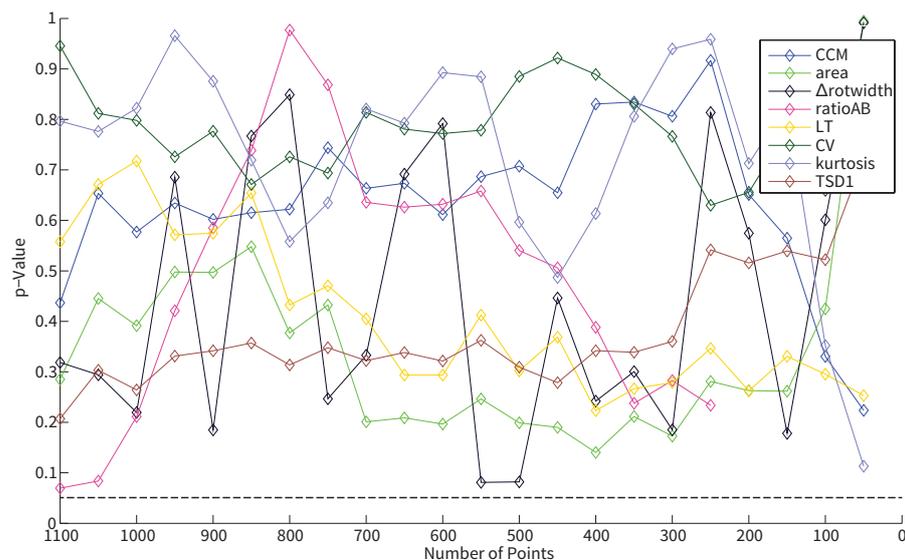


Figure 17: Indices, which have no p-value below the significance threshold of 0.05 (dashed line) for any data length.

4.2 OVERVIEW OF TEST CASES II-IV

Tab 1 shows the p-values for the other three test cases.

Table 1: Results of tests cases II-IV with results of $p < 0.05$ emphasized and marked with * and ** for $p < 0.01$.

Indexgroup/ <i>Index</i>	Test Case II:	Test Case III:	Test Case IV:
	Non-Path/Path.	Pre./Post.	Young/Old
	p-Value	p-Value	p-Value
Cluster Indices			
<i>NumOfClusters</i>	< 0.01 **	< 0.01 **	0.143
<i>ClusterSizeRatio</i>	< 0.01 **	< 0.01 **	0.247
Asymmetry Indices			
<i>PI</i>	< 0.01 **	< 0.01 **	0.045 *
<i>GI</i>	< 0.01 **	0.193	0.995
<i>PI_p</i>	0.014 *	0.177	< 0.01 **
<i>GI_p</i>	< 0.01 **	0.111	< 0.01 **
<i>GI_{p3D}</i>	< 0.01 **	< 0.01 **	< 0.01 **
3D Histogram Indices			
<i>area</i>	0.391	< 0.01 **	0.919
<i>slope</i>	0.777	< 0.01 **	0.329
<i>NP</i>	0.996	< 0.01 **	0.204
<i>DP</i>	0.770	< 0.01 **	0.053
Dispersion Indices			
<i>width₁₀</i>	0.946	< 0.01 **	0.175
<i>width₉₀</i>	0.019 *	< 0.01 **	0.492
<i>Δwidth</i>	< 0.01 **	< 0.01 **	0.061
<i>rotwidth₁₀</i>	0.885	< 0.01 **	0.575
<i>rotwidth₉₀</i>	0.198	< 0.01 **	0.500
<i>Δrotwidth</i>	0.220	0.587	0.324
Mechanical Indices			
<i>A</i>	< 0.01 **	< 0.01 **	< 0.01 **
<i>B</i>	< 0.01 **	< 0.01 **	< 0.01 **
<i>ratioAB</i>	0.238	< 0.01 **	< 0.01 **
<i>Q_{xx}</i>	< 0.01 **	< 0.01 **	< 0.01 **
<i>Q_{yy}</i>	< 0.01 **	< 0.01 **	< 0.01 **
<i>Q_{zz}</i>	< 0.01 **	< 0.01 **	< 0.01 **

T_{xxx}	< 0.01 **	< 0.01 **	< 0.01 **
T_{yyy}	0.804	< 0.01 **	0.746
T_{xxy}	0.998	< 0.01 **	0.636
T_{xyy}	< 0.01 **	< 0.01 **	< 0.01 **
T_{xzz}	< 0.01 **	< 0.01 **	< 0.01 **
T_{yzz}	< 0.01 **	0.529	0.191
<i>kurtosis</i>	0.7498	0.021 *	0.076
LT Indices			
<i>L</i>	0.075	< 0.01 **	0.663
<i>T</i>	0.038 *	< 0.01 **	0.864
<i>LT</i>	0.655	< 0.01 **	0.965
<i>ratioLT</i>	< 0.01 **	< 0.01 **	0.815
Ellipse Fitting Indices			
<i>SD1</i>	0.042 *	< 0.01 **	0.914
<i>SD2</i>	< 0.01 **	< 0.01 **	0.666
<i>ratioSD</i>	< 0.01 **	0.037 *	0.152
<i>TSD1</i>	0.264	< 0.01 **	0.981
<i>TSD2</i>	< 0.01 **	< 0.01 **	0.679
<i>ratioTSD</i>	< 0.01 **	< 0.01 **	0.099
<i>HiSD</i>	< 0.01 **	0.358	0.149
Statistical Indices			
<i>SDNN</i>	0.395	< 0.01 **	0.425
<i>pNN50</i>	< 0.01 **	< 0.01 **	< 0.01 **
<i>SDSD</i>	0.028 *	< 0.01 **	0.900
<i>CV</i>	0.798	< 0.01 **	0.081
Ungrouped Indices			
<i>CCM</i>	0.577	0.413	0.610
<i>range</i>	0.433	< 0.01 **	0.829
<i>contour</i>	0.309	< 0.01 **	1.000
<i>ext</i>	< 0.01 **	< 0.01 **	0.747

4.3 PARAMETERS OF DISTRIBUTIONS

Table 2 shows parameters of distribution of each index for the second test case. Each numerical column contains the median and the

95% central range, i.e., the 97.5th and the 2.5th percentile, of the non-normally distributed indices and the mean value and standard deviation for normally distributed indices for one of the two data bases.

Table 2: Parameters of distribution for the indices in test case II.

Index	median or mean, (95% central range or SD)	
	of Non-Pathological Data	of Pathological Data
Cluster Indices		
<i>NumOfClusters</i>	1, (1, 6)	3, (1, 10)
<i>ClusterSizeRatio</i>	0, (0, 0.04)	0.0084, (0, 0.40)
Asymmetry Indices		
<i>PI</i>	50.7, (44.9, 59.5)	51.5, (47.4, 61.6)
<i>GI</i>	49.2, (37.3, 61.3)	50.6, (21.6, 67.7)
<i>PI_p</i>	50.73, (46.37, 57.43)	50.85, (47.77, 59.69)
<i>GI_p</i>	51.5, (30.0, 67.7)	45.9, (24.2, 60.7)
<i>GI_{p3D}</i>	50.4, (26.8, 55.8)	47.8, (19.6, 68.3)
3D Histogram Indices		
<i>NP</i>	20.5, (0, 55.5)	19, (0, 57)
<i>DP</i>	0.13, (0, 0.51)	0.15, (0, 0.67)
<i>area</i>	18.0, (6.8, 48.7)	19.6, (4.0, 117.7)
<i>slope</i>	7.5875, (2.2482, 27.6666)	7.5853, (1.4086, 34.5761)
Dispersion Indices		
<i>width₁₀</i>	0.046, (0.018, 0.180)	0.050, (0.016, 0.420)
<i>width₉₀</i>	0.063, (0.023, 0.211)	0.055, (0.015, 0.336)
<i>Δwidth</i>	0.008, (-0.088, 0.085)	0.000, (-0.250, 0.165)
<i>rotwidth₁₀</i>	0.0442, (0, 0.2076)	0.0442, (0, 0.3345)
<i>rotwidth₉₀</i>	0.055, (0.017, 0.284)	0.047, (0.006, 0.351)
<i>Δrotwidth</i>	0.008, (-0.146, 0.139)	0.002, (-0.209, 0.194)
Mechanical Indices		
<i>A</i>	0.70, (SD 0.14)	0.61, (SD 0.13)
<i>B</i>	0.704, (SD 0.136)	0.605, (SD 0.127)
<i>ratioAB</i>	1.0036, (1.0012, 1.0103)	1.0041, (1.0012, 1.0168)
<i>Q_{xx}</i>	3493, (SD 1343)	2588, (SD 104)

Q_{yy}	-1744, (SD 670)	-1286, (SD 554)
Q_{zz}	-1748, (SD 673)	-1301, (SD 549)
T_{xxx}	12943, (3678, 38270)	7761, (2175, 26533)
T_{yyy}	0.12, (-2.20, 4.58)	0.05, (-1.76, 35.42)
T_{xxy}	-0.44, (-16.63, 8.82)	-0.17, (-78.55, 7.40)
T_{xyy}	-19399, (-56793, -5515)	-11513, (-39773, -3201)
T_{xzz}	-19432, (-58018, -5520)	-11687, (-39826, -3326)
T_{yzz}	-0.002, (-2.747, 2.381)	-0.216, (-27.165, 2.452)
<i>kurtosis</i>	1.44, (0.49, 5.88)	1.36, (0.54, 13.24)
LT Indices		
<i>L</i>	0.43, (0.16, 0.98)	0.38, (0.08, 1.04)
<i>T</i>	0.21, (0.07, 0.66)	0.25, (0.05, 1.08)
<i>LT</i>	-2.473, (SD 1.016)	-2.414, (SD 1.431)
<i>ratioLT</i>	1.95, (1.23, 4.86)	1.28, (0.66, 3.40)
Ellipse Fitting Indices		
<i>SD1</i>	0.02, (0.008, 0.091)	0.022, (0.006, 0.196)
<i>SD2</i>	0.076, (0.024, 0.174)	0.056, (0.012, 0.190)
<i>ratioSD</i>	0.247, (0.118, 0.65)	0.454, (0.105, 2.206)
<i>TSD1</i>	0.075, (0.024, 0.173)	0.058, (0.013, 0.194)
<i>TSD2</i>	0.020, (0.008, 0.095)	0.022, (0.005, 0.195)
<i>ratioTSD</i>	3.7, (1.5, 8.5)	2.3, (0.6, 9.7)
<i>HiSD</i>	3.0, (SD 1.2)	2.2, (SD 1.6)
Statistical Indices		
<i>SDNN</i>	0.0551, (0.0196, 0.1540)	0.0513, (0.0124, 0.2091)
<i>pNN50</i>	100, (87.7482, 100)	100, (34.7, 100)
<i>SDSD</i>	0.0282, (0.011, 0.1451)	0.0356, (0.0090, 0.2864)
<i>CV</i>	0.0612, (0.020, 0.151)	0.057, (0.017, 0.258)
Ungrouped Indices		
<i>CCM</i>	-0.0231, (-0.0808, -0.0069)	-0.0238, (-0.0983, -0.0072)
<i>contour</i>	8.0, (3.0, 27.2)	6.7, (0.9, 32.6)
<i>ext</i>	53.6, (SD 15.3)	46.4, (SD 19.2)
<i>range</i>	3.89, (0.86, 27.04)	3.47, (0.20, 26.34)

Table 3 shows the parameters of distributions of each index for the third test case.

Table 3: Parameters of distributions for the indices in test case III.

Index	median or mean, (95% central range or SD)	
	Pre-Treatment	Post-Treatment
Cluster Indices		
<i>NumOfClusters</i>	1, (1, 10)	1, (1, 7)
<i>ClusterSizeRatio</i>	0, (0, 0.1118)	0, (0, 0.0363)
Asymmetry Indices		
<i>PI</i>	50.4235, (46.1501, 59.7371)	50.246, (45.9656, 55.8282)
<i>GI</i>	50.1231, (11.9063, 63.1727)	50.2637, (37.4709, 57.7654)
<i>PI_p</i>	50.0688, (47.2444, 55.3693)	50.1941, (47.3538, 54.4184)
<i>GI_p</i>	50.5346, (28.5170, 64.0732)	50.3131, (37.4521, 62.9297)
<i>GI_{p3D}</i>	50.0264, (23.3527, 75.1189)	50.7674, (41.4261, 84.3454)
3D Histogram Indices		
<i>NP</i>	42, (2, 60)	43, (1, 60)
<i>DP</i>	0.2762, (0.0110, 0.7043)	0.2907, (0.0166, 0.7237)
<i>area</i>	10.8708, (2.8443, 32.2084)	9.6067, (2.6841, 25.3397)
<i>slope</i>	4.5852, (1.545, 12.7831)	4.0997, (1.2675, 10.6949)
Dispersion Indices		
<i>width₁₀</i>	0.0390, (0.0150, 0.1535)	0.0390, (0.0150, 0.1140)
<i>width₉₀</i>	0.0390, (0.0155, 0.1520)	0.0390, (0.0150, 0.133)
Δ <i>width</i>	0.00100, (-0.0665, 0.0855)	0, (-0.0390, 0.0705)
<i>rotwidth₁₀</i>	0.0276, (0.0057, 0.1407)	0.0276, (0.0049, 0.0919)
<i>rotwidth₉₀</i>	0.0325, (0.0049, 0.1658)	0.0283, (0.0057, 0.1103)
Δ <i>rotwidth</i>	0.00070, (-0.0774, 0.1025)	0.00420, (-0.0499, 0.0718)
Mechanical Indices		
<i>A</i>	0.5805, (0.4097, 0.8601)	0.5957, (0.4327, 0.8749)
<i>B</i>	0.5782, (0.4087, 0.8573)	0.5934, (0.4306, 0.8713)
<i>ratioAB</i>	1.0034, (1.0012, 1.0098)	1.0029, (1.0011, 1.0075)
<i>Q_{xx}</i>	2276, (1134, 5008)	2401, (1266, 5157)
<i>Q_{yy}</i>	-1138, (-2503, -566)	-1200, (-2575, -633)

Q_{zz}	-1139, (-2504, -570)	-1200, (-2582, -633)
T_{xxx}	7337, (2565, 23937)	7945, (3037, 25099)
T_{yyy}	0.0424, (-2.0338, 2.6925)	-0.0337, (-1.9415, 1.9971)
T_{xxy}	-0.1546, (-10.0115, 8.0948)	0.1402, (-7.4504, 7.3475)
T_{xyy}	-10999, (-35534, -3839)	-11899, (-37564, -4554)
T_{xzz}	-11023, (-35921, -3880)	-11932, (-37734.3803, -4557)
T_{yzz}	-0.00120, (-3.2065, 1.8528)	-0.0366, (-1.7859, 2.0494)
<i>kurtosis</i>	1.3705, (0.5036, 8.3646)	1.3118, (0.5495, 5.2591)
LT Indices		
L	0.2927, (0.0877, 0.7071)	0.2546, (0.0721, 0.6466)
T	0.1439, (0.0445, 0.5989)	0.1209, (0.0445, 0.4087)
LT	-3.1443, (SD 1.1257)	-3.4888, (SD 1.0306)
<i>ratioLT</i>	1.9151, (0.7184, 3.8939)	2.0795, (0.8148, 3.9300)
Ellipse Fitting Indices		
$SD1$	0.0142, (0.00610, 0.0733)	0.0123, (0.0060, 0.0516)
$SD2$	0.0490, (0.0120, 0.1489)	0.0454, (0.0116, 0.1422)
<i>ratioSD</i>	0.2936, (0.1071, 1.6703)	0.2814, (0.1057, 1.2158)
$TSD1$	0.0494, (0.0124, 0.1489)	0.0454, (0.0117, 0.1422)
$TSD2$	0.0140, (0.00610, 0.0717)	0.0123, (0.0060, 0.0478)
<i>ratioTSD</i>	3.4478, (0.7208, 8.7498)	3.5467, (0.8377, 8.9738)
$HiSD$	2.53, (0.8706, 6.6328)	2.6224, (0.9243, 6.7508)
Statistical Indices		
$SDNN$	0.0443, (0.0123, 0.1251)	0.0396, (0.0127, 0.1100)
$pNN50$	100, (87.6, 100)	100, (98.2, 100)
$SDSD$	0.0215, (0.0091, 0.124)	0.0181, (0.0086, 0.0782)
CV	0.0495, (0.0179, 0.142)	0.0439, (0.0154, 0.1147)
Ungrouped Indices		
CCM	-0.0156, (-0.0561, -0.0070)	-0.0153, (-0.0428, -0.0073)
<i>contour</i>	5.479, (1.0835, 17.9417)	4.8982, (0.9865, 15.8618)
<i>ext</i>	48.1733, (SD 16.6949)	50.9067, (SD 16.4308)
<i>range</i>	2.1672, (0.2585, 15.1382)	1.8052, (0.2583, 9.8763)

Table 4 shows parameters of distributions of each index in the forth test case.

Table 4: Parameters of distributions for the indices in test case IV.

Index	median or mean, (95% central range or SD)	
	of Younger Subjects	of Older Subjects
Cluster Indices		
<i>NumOfClusters</i>	1, (1, 6.525)	1, (1, 6)
<i>ClusterSizeRatio</i>	0, (0, 0.0727)	0, (0, 0.043)
Asymmetry Indices		
<i>PI</i>	50.2167, (SD 2.983)	51.6714, (SD 3.6113)
<i>GI</i>	49.0513, (SD 5.424)	48.1692, (SD 7.1248)
<i>PI_p</i>	50.2168, (SD 2.2393)	51.2244, (SD 2.992)
<i>GI_p</i>	53.1708, (SD 9.9182)	48.8854, (SD 7.6011)
<i>GI_{p3D}</i>	51.073, (SD 3.371)	47.905, (SD 7.0538)
3D Histogram Indices		
<i>NP</i>	11, (0, 51.225)	27, (0, 58.125)
<i>DP</i>	0.0849, (0, 0.4667)	0.2154, (0, 0.5085)
<i>area</i>	20.8545, (SD 9.3208)	22.2697, (SD 15.1141)
<i>slope</i>	7.0639, (2.2052, 28.0896)	7.6765, (2.3213, 28.1152)
Dispersion Indices		
<i>width₁₀</i>	0.0469, (0.0156, 0.1449)	0.052, (0.0232, 0.2789)
<i>width₉₀</i>	0.068, (0.0159, 0.2175)	0.0625, (0.0279, 0.2126)
$\Delta width$	0.0156, (-0.0756, 0.1182)	0.0040, (-0.1131, 0.0872)
<i>rotwidth₁₀</i>	0.0442, (0.0043, 0.1714)	0.0442, (0, 0.2786)
<i>rotwidth₉₀</i>	0.0566, (0, 0.234)	0.0537, (0.017, 0.3008)
$\Delta rotwidth$	0.0113, (-0.145, 0.1813)	0.0055, (-0.15, 0.1481)
Mechanical Indices		
<i>A</i>	0.6602, (SD 0.1407)	0.7462, (SD 0.1190)
<i>B</i>	0.6571, (SD 0.1410)	0.7435, (SD 0.1188)
<i>ratioAB</i>	1.005, (SD 0.0026)	1.0036, (SD 0.0018)
<i>Q_{xx}</i>	3072, (SD 1311)	3855, (SD 1272)
<i>Q_{yy}</i>	-1535, (SD 655)	-1925, (SD 632)
<i>Q_{zz}</i>	-1405, (-2947, -539)	-1785, (-3522, -1028)
<i>T_{xxx}</i>	10023, (2392, 30621)	14388, (6350, 39806)

T_{yyy}	0.1215, (-1.6362, 5.3352)	0.1099, (-2.6208, 5.535)
T_{xxy}	-0.4852, (-18.6555, 6.5145)	-0.4315, (-14.9609, 10.4873)
T_{xyy}	-15032, (-45914, -3587)	-21570, (-59487, -9490)
T_{xzz}	-15038, (-45949, -3589)	-21595, (-59931, -9560)
T_{yzz}	0.0785, (SD 1.0553)	-1.0003, (SD 6.1645)
<i>kurtosis</i>	1.5065, (0.6845, 6.5337)	1.2146, (0.4708, 5.5053)
LT Indices		
<i>L</i>	0.446, (SD 0.1775)	0.4634, (SD 0.2324)
<i>T</i>	0.2065, (0.0693, 0.5448)	0.2099, (0.0661, 0.8565)
<i>LT</i>	-2.4678, (SD 0.88)	-2.4765, (SD 1.1278)
<i>ratioLT</i>	1.9452, (1.3985, 3.9515)	1.9592, (1.1497, 5.4893)
Ellipse Fitting Indices		
<i>SD1</i>	0.020, (0.0073, 0.0802)	0.0182, (0.0080, 0.1167)
<i>SD2</i>	0.0794, (0.0357, 0.1834)	0.0754, (0.0231, 0.1789)
<i>ratioSD</i>	0.2292, (0.1135, 0.5815)	0.2733, (0.1388, 0.7711)
<i>TSD1</i>	0.0789, (0.0357, 0.1833)	0.0753, (0.023, 0.1777)
<i>TSD2</i>	0.0218, (0.0073, 0.0806)	0.0187, (0.0080, 0.1152)
<i>ratioTSD</i>	4.1441, (SD 1.7347)	3.6246, (SD 1.5415)
<i>HiSD</i>	3.21, (SD 1.2955)	2.876, (SD 1.1171)
Statistical Indices		
<i>SDNN</i>	0.0554, (0.0268, 0.1504)	0.0535, (0.0193, 0.2057)
<i>pNN50</i>	100, (75.9, 100)	100, (100, 100)
<i>SDSD</i>	0.0303, (0.0102, 0.1201)	0.0270, (0.0112, 0.1575)
<i>CV</i>	0.0649, (0.0272, 0.1615)	0.0547, (0.0186, 0.1512)
Ungrouped Indices		
<i>CCM</i>	-0.0229, (-0.0737, -0.0074)	-0.0243, (-0.0861, -0.0066)
<i>contour</i>	8.2086, (3.1832, 21.8194)	7.9485, (2.8658, 31.5036)
<i>ext</i>	53.11, (SD 16.0164)	54.0562, (SD 14.7121)
<i>range</i>	3.903, (0.9877, 18.4658)	3.8706, (0.8443, 42.0517)

4.4 TEST CASE II: NON-PATHOLOGICAL VS. PATHOLOGICAL DATA

The first column of p-values in table 1 shows that the difference for both cluster indices between pathological and non-pathological heart rate data is very significant.

The same is true for all asymmetry indices, except for the PI_p , which is still significant but not as strongly as the other ones.

The dispersion index $width_{90}$ is also significantly different, while $\Delta width$ is even very significant.

So are the mechanical indices A , B , Q_{xx} , Q_{yy} , Q_{zz} , T_{xxx} , T_{xyy} , T_{xzz} and T_{yzz} .

In the case of the ellipse fitting indices, only $TSD2$ does not have significant differences between the two databases. SDI has significant differences and the other ones show even a very significant behavior. Of the statistical indices, only $pNN50$ is very significant and $SDSD$ shows a smaller, but still significant difference.

Table 2 shows the distribution parameters of all indices for this test case.

Figure 18 shows two representative box plots of the two indices GI_p and CCM . Karmakar-Guzik's index GI_p visibly differs between non-pathological and pathological heart beat data, which is not the case for the index CCM .

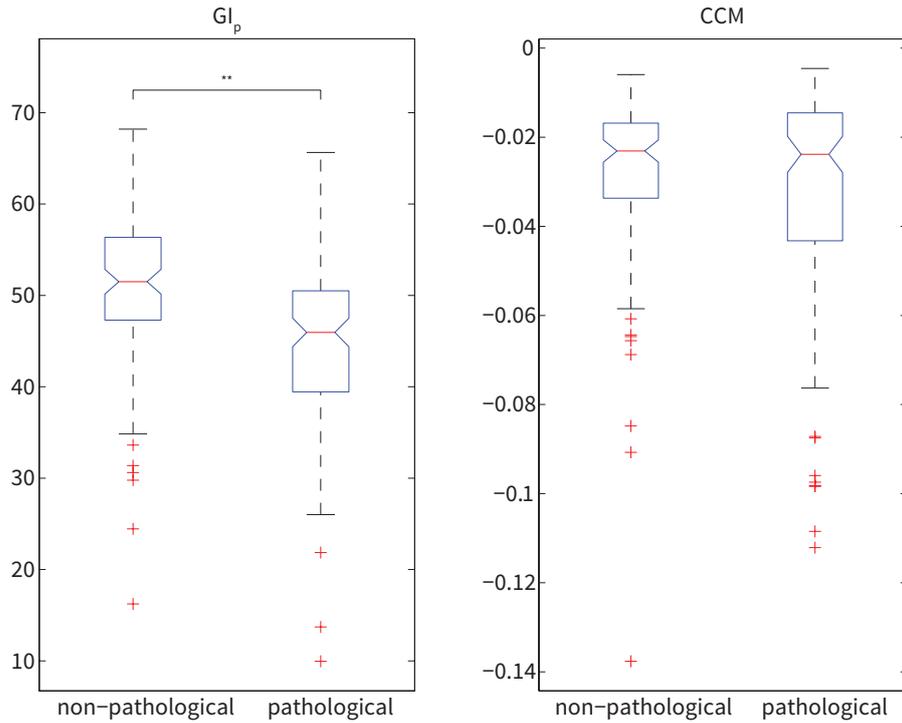


Figure 18: Boxplots for Karmakar-Guzik asymmetry index GI_p and for the peak index NP .

4.5 TEST CASE III: PRE- VS. POST-ANTIARRHYTHMIC TREATMENT DATA

For the third test case, many indices show a very significant difference for data recorded before and after an arrhythmia suppression therapy. The indices *kurtosis* and *ratioSD* show significant differences, while the asymmetry indices GI , PI_p and GI_p , as well as the indices CCM , Δ *rotwidth*, *HiSD* and T_{yzz} show no significance. The distribution parameters of all indices for this test case are shown in table 3.

4.6 TEST CASE IV: NON-PATHOLOGICAL DATA FROM YOUNGER VS. OLDER SUBJECTS

Table 1 shows the results for the test concerning the differences in the Poincaré plot indices due to the age of the subjects. None of the cluster indices shows significant differences in this case. While GI does not differentiate significantly, all the other asymmetry indices do, in the case of PI , GI_p and GI_p even very significantly. None of

the 3D histogram, of the dispersion and of LT-indices, as well as no ellipse fitting index, show any significant differentiation in this test. Most of the mechanical indices have very significant differences between heart rate data of younger and older subjects, except for *kurtosis*, T_{yyy} , T_{xxy} and T_{yzz} . The only statistical index with a p-value below the significance threshold is *pNN50*. Table 4 shows the distribution parameters of all indices for this test case.

4.7 CORRELATION

Figures 19-22 show the results for the test of correlations. Only the correlations between those indices, which had significant differences in either test case II or III were calculated. All the strong correlations, i.e., $|r| > 0.85$, had a p-value below 0.01. For the exact r-values see the appendix (section 7.1).

In figure 19, one can see the correlations in the case of the pathological data set. Almost all mechanical indices ($A, B, Q_{xx}, Q_{yy}, Q_{zz}, T_{xzz}, T_{xyy}, T_{xxx}$) are strongly correlated with each other, but not with any other index. The index T is correlated with $SDSD$, and $SD1$, the later ones are also strongly correlated with each other. $SD2$ is only correlated with $TSD2$. Both ratios of the ellipse fitting indices are correlated with each other, as well as with $HiSD$. Both cluster indices are correlated with each other, as it is the case between PI and PI_p .

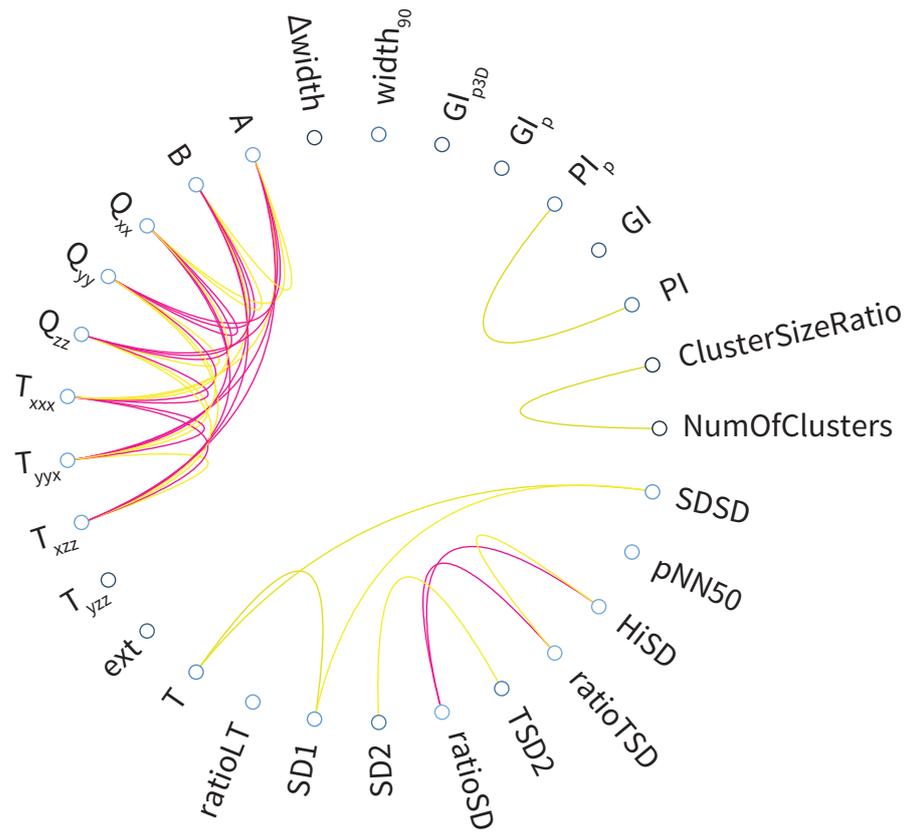


Figure 19: Strong correlations, i.e., $|r| > 0.85$, between Poincaré plot indices of the pathological data set, which showed significant results in test case II. Yellow lines mark positive, magenta lines negative correlation.

Figure 20 shows that the correlations of the indices for non-pathological data are very similar. The only difference is that PI and PI_p are not strongly correlated anymore.

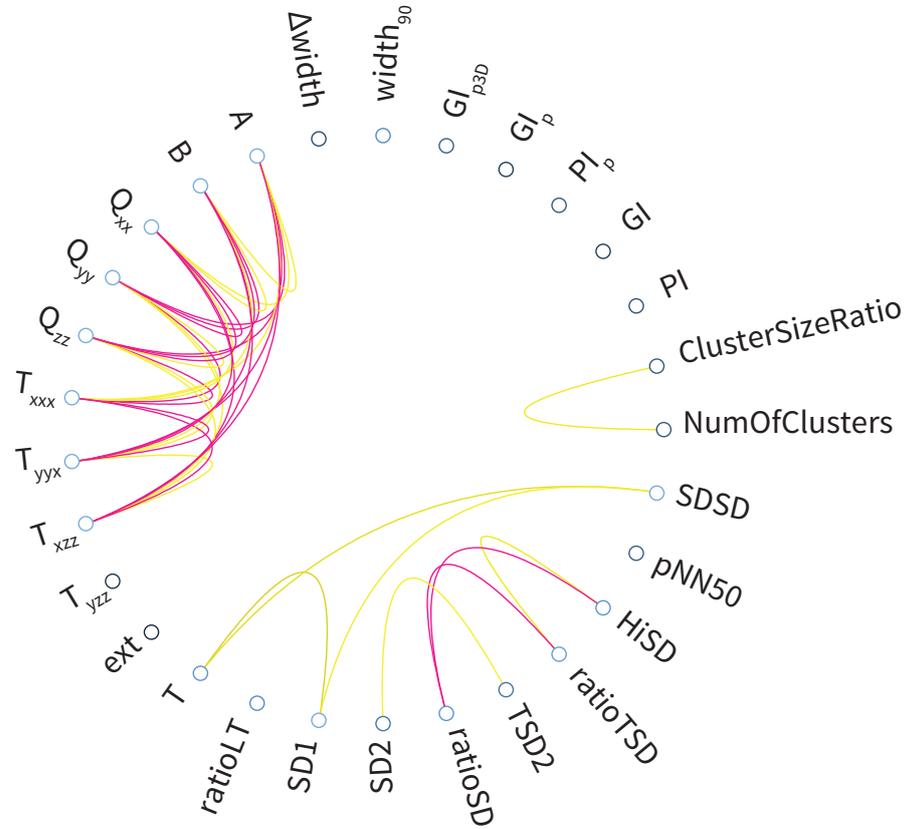


Figure 20: Strong correlations, i.e., $|r| > 0.85$, between Poincaré plot indices of the non-pathological data set, which showed significant results in test case II. Yellow lines mark positive, magenta lines negative correlation.

In figure 21, one can see the same correlations occurring as in figure 19. Additional correlations are found between the *contour* and the *range*, *TSD2*, *L* and *LT*, as well as between *CV* and *ratioAB*, *TSD1* and *SD2*. Additionally, the index *range* correlates strongly with *LT*, *T* and *area*. The later three are also strongly correlated with each other. The ellipse fitting index *SD2* is also correlated with *T*, and *TSD2*. *TSD2* is as well correlated with *CV*, *L* and *SDNN*. The index *TSD1* is correlated with *contour*, *SDSD* and *SD1*.

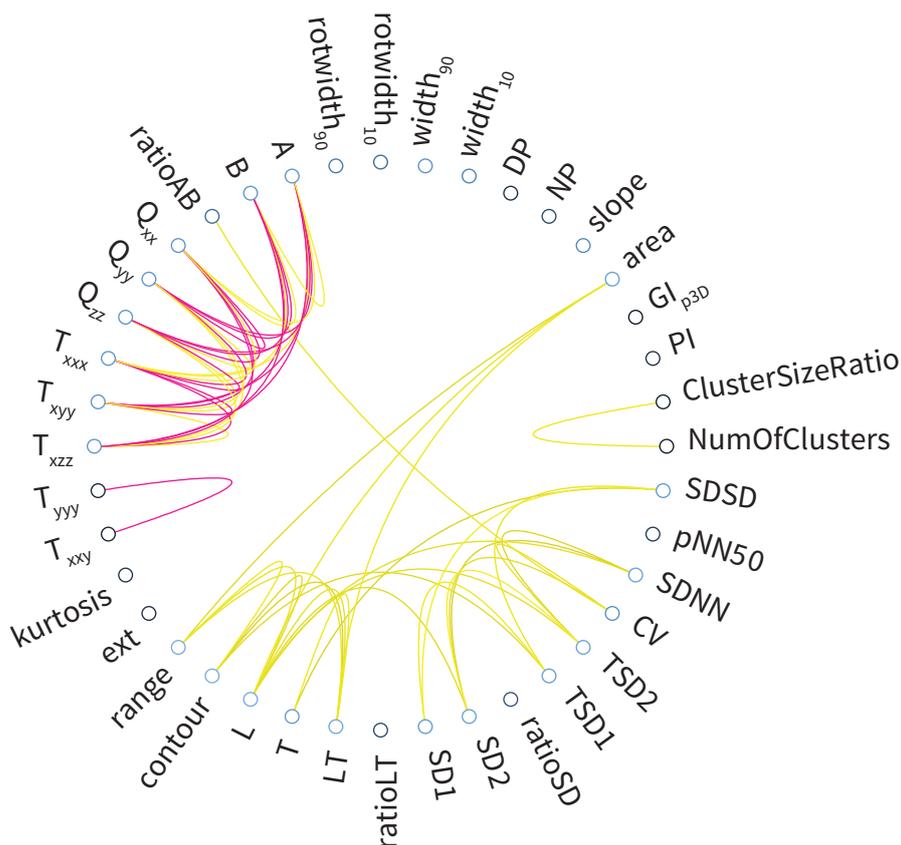


Figure 21: Strong correlations, i.e., $|r| > 0.85$, between Poincaré plot indices of the pre-arrhythmia-treatment data set, which showed significant results in test case III. Yellow lines mark positive, magenta lines negative correlation.

Figure 22 shows that the indices for the post-arrhythmic treatment database has similar correlations as the pre-treatment indices. The differences are that $width_{90}$ is in this case correlated with $TSD1$, the indices $SD2$, $TSD2$, $contour$ and $area$ are also correlated with each other, whereas the index T is no longer correlated with $SDSD$.

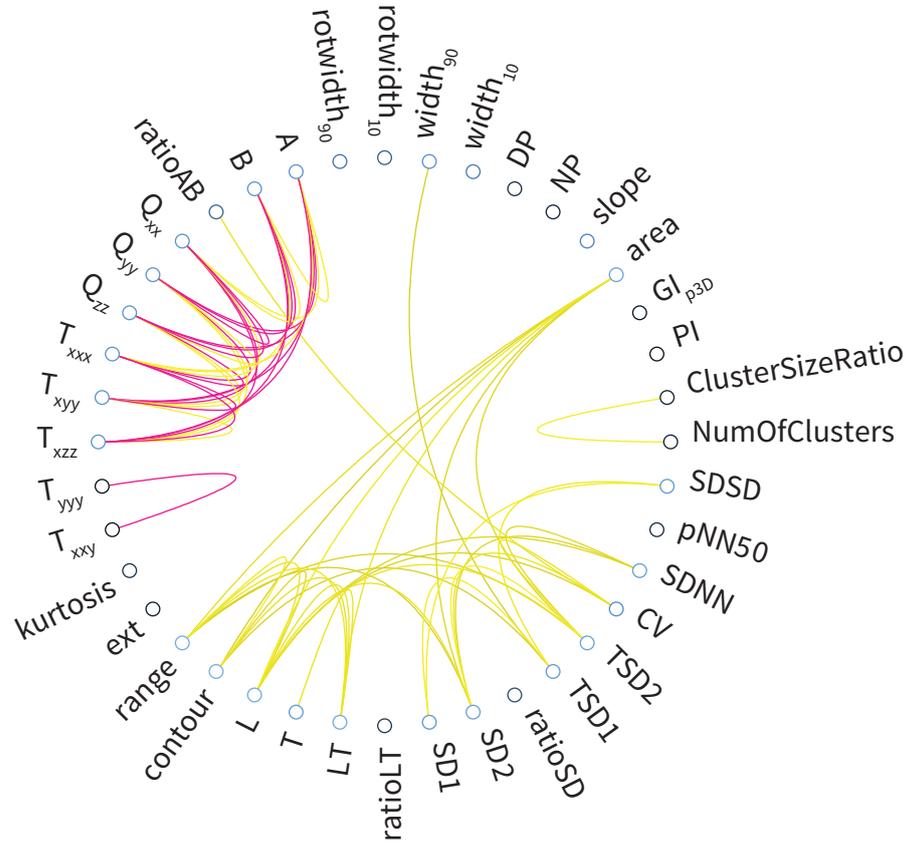


Figure 22: Strong correlations, i.e., $|r| > 0.85$, between Poincaré plot indices of the pre-arrhythmia-treatment data set, which showed significant results in test case III. Yellow lines mark positive, magenta lines negative correlation.

DISCUSSION

5.1 CLUSTER INDICES

Both cluster indices show similar behavior in the last three test cases (see table 1) and are also strongly correlated with each other in all cases (see figure 19-22). One possible explanation for their interdependence is that additional clusters consist of points which otherwise are more often part of the sinus beat cluster and not of clusters of other extrasystoles. This is because the sinus beat cluster is usually the biggest one and therefore the ratio between the number of points in the second cluster and in the first one gets larger with more clusters.

Both indices had, compared to all other indices, the most significant differences for test case III. This can be explained by the fact that for this test data from patients before and after arrhythmia suppression therapy was used and arrhythmia often appears as a multiclustered Poincaré plot (see figure 11).

Both show no difference between younger and older healthy subjects, because most of the data in the non-pathological database only has one cluster, as should be expected from noise filtered Poincaré plots of healthy subjects [13].

5.2 ASYMMETRY INDICES

As shown in table 2 and 18, all asymmetry indices show a stronger asymmetry for pathological heart rate data, i.e., the indices deviate more from the value 50, which represents a symmetrical Poincaré plot. Only PI and GI_{p3D} have significant differences for pre- and post-arrhythmia-treatment subjects, but table 3 shows that only GI_{p3D} indicates a larger asymmetry pre-treatment and it is the opposite for PI .

Although most of them also differ significantly in the fourth test case, all of them, except for GI_{p3D} , showed a higher asymmetry for older subjects. According to [11] time irreversibility and therefore the Poincaré plot's asymmetry degrades with aging. It is unclear why this is

not the case for the other asymmetry indices.

All of them show better results for longer data lengths (see figure 13 and 15), except for GI_p which does not seem to be strongly influenced by data length (see figure 16). This could be the case, because the asymmetry only manifests itself after a certain amount of time, but no research in this direction could be found.

The strong positive correlation between PI and PI_p in the case of pathological data and almost strong positive correlation for non-pathological data (see table 6) is probably due to the reason that both count the number of points on each side of the line of identity and most of the points counted in PI_p are a subset of the points counted in PI .

5.3 3D HISTOGRAM INDICES

Both index groups based on the 3D Histogram of a Poincaré plot, the peak and the density indices, only show significant differences for arrhythmia patients before and after treatment. Hnatkova et al. found similar results for heart rate data of arrhythmia in [21].

The differences between the peak indices NP and DP for pathological and non-pathological data get more significant with shorter data lengths. One explanation for this could be that with a fixed threshold for considered peak heights and a decreasing number of points there is less chance of these points overlapping and only peaks, which are very high at the beginning, are still considered. Therefore the compactness of the Poincaré plot becomes more evident, which corresponds with a decreased HRV [31]. In fact, if one looks at the actual values of NP for shorter data lengths it is generally higher for non-pathological than for pathological data.

Marciano et al. found in [32] significant differences of NP and DP between healthy subjects and subjects with advanced heart failure. A possible reason why this is not reproduced in our test cases is, that they used unfiltered Poincaré plots of 24h recordings, consisting of 90,000 points in general - a data length much longer than the one used in our tests.

The area under the approximated density function seems to have a higher amount of information than its slope, but both density indices, *area* and *slope*, only show significant differences, if arrhythmia patients before and after treatment are compared. In the length test just *slope* shows a sensitivity, with a slow trend towards significance

for shorter signals. The differences of the length sensitivity of NP and DP possibly arise due to the fact that $area$ and $slope$ are defined without any fixed restrictions on the peak heights.

The strong correlation of $area$ to L and $range$ is probably due to the fact that a longer Poincaré plot leads to less overlapping points and therefore to a larger area under density function.

5.4 DISPERSION INDICES

Interestingly the rotated version of the dispersion indices are worse at differentiating in the first test case, compared to the unrotated ones, but are even more significant in the second test case. The most promising is $width_{90}$, which is significant in test case II and III, with no clear trend with respect to its data length. The indices $\Delta width$, $rotwidth_{10}$ and $rotwidth_{90}$ are very sensitive to the number of points in the Poincaré plot.

Considering the statistical position and spread parameters in table 2, all six indices seem to indicate that a comet-shaped Poincaré plot, i.e., $width_{90}$ and $rotwidth_{90}$, respectively, is larger than $width_{10}$ and $rotwidth_{10}$, respectively, is more common in the non-pathological data set. This corresponds with the traditional interpretation of Poincaré plots [13]. For test case III, they indicate larger shape before arrhythmia suppression, probably because of arrhythmia clusters, which merged with the sinus cluster and therefore widened it in general. Considering table 4, the results of the dispersion indices could denote a greater proportion of cigar-shaped Poincaré plots for older subjects compared to younger ones.

If one looks at the visual definitions of $width_{90}$ and $TSDI$ in figure 7, their strong correlation for pre-treatment data can be explained by the fact that depending on the shape of the Poincaré plot, both measures can coincide. Smith et al. found in [43] no correlation of $width_{90}$ and $width_{10}$ with any other Poincaré plot index, but they used very short heart rate signals of patients during anesthesia.

5.5 MECHANICAL INDICES

The radii of inertia showed similarly good results by showing very significant differences in the last three test cases, without strong sensitivity to the data length, but they are not computable for too short signals. Their ratio showed no significant differences for pathologi-

cal and non-pathological data, as well as for heart rates of younger and older subjects. All of them did not react strongly with respect to data length, except for signals shorter than 100 points. Marciano et al. reported in [32] that only A has significant differences between long signals of pathological and non-pathological data. They also found a strong correlation between A and ext , which could not be reproduced. In our test the radii correlated strongly only with each other and the quadro- and octupoles. The strong correlations of A to $SDNN$ and between B and $RMSD$ reported by Smith et al. in [43], could not be reproduced, probably due to the very short signals used by Smith et al.

All of the quadrupoles and some of the octupoles showed very significant differences for all test cases. Lewkowicz et al. reported in [27], that Q_{yy} , T_{xxx} and the kurtosis showed the most promising results. In the case of *kurtosis* this could only be reproduced for the pre- vs. post-treatment test, presumably this is the case, because of shorter signals for the other test cases, compared to the 24h recordings used by Lewkowicz.

Both other quadrupoles Q_{xx} and Q_{zz} showed equally good results compared to Q_{yy} . There are also some octupoles, T_{xyy} and T_{xzz} , which are strongly correlated to T_{xxx} . All of these quadrupoles and octupoles do not show any sensitivity towards the data length in test case I. In the same test case, T_{yzz} shows a sudden loss of significance for a data length of 500 and 550 points. In contrast, T_{yyy} and T_{xxy} show suddenly more significant results for these lengths (see figure 14). Even after closer inspection of the data no reason for this behavior can be found.

The strong correlation between Q_{zz} , Q_{xx} and Q_{yy} is not surprising since one is calculated as a linear combination of the other two. The same explanation holds true for all observed strong correlations between octupoles and between octupoles and quadrupoles.

5.6 LT-INDICES

The index *ratioLT* showed best results of all LT-indices over all tests, as was reported by Toichi et al. in [47]. It was very significant for test case II and III and did not show any sensitivity with respect to data length in test case I. It only did not show significant differences between the heart rate of older and younger subjects, as did none of the other LT-indices. It is not strongly correlated to any other index.

Toichi et al. reported that besides *ratioLT* the index *LT* showed also good results for the assessment of the sympathetic and parasympathetic activities. Although it has a trend towards higher significance for shorter data lengths, this index never surpasses the significance threshold in test case I and in test case II. It had very significant differences between pre- and post-treatment heart rate data of subjects with arrhythmia.

The index *L* also showed just for test case III a very significant differentiation ability and seems to have a weak trend towards significance for longer signals in test case I. The index *T* on the other hand has significant differences between pathological and non-pathological heart rate data of all lengths, which get more significant for shorter signals. This sensitivity to the data length could be explained by the differences in shape stability of Poincaré plots from pathological and non-pathological data with respect to their signal length, as mentioned in 5.9.

Both, *L* and *T*, seem to have no capability to capture the non-linear features of a Poincaré plot since they are strongly correlated to linear, statistical measures.

5.7 ELLIPSE FITTING INDICES

None of the ellipse fitting indices showed significant differences due to the subject's age. The best results had *SD2* with very significant differences for most of the test cases and showing no sensitivity with respect to the data length. The index *ratioSD* has very significant differences for all lengths of pathological and non-pathological data without any visible trend for changing signal lengths. Index *SD1* has more significant differences for the third test case, compared to *ratioSD*, while on the the opposite is true for the second test case, although in both cases both indices show significant differences. Similarly Karmakar et al. also reported in [25] that *SD2* shows a more significant difference between patients with cognitive heart failure and with normal sinus rhythm compared to *SD1*.

The other triplet of ellipse fitting indices *TSD1*, *TSD2* and *ratioTSD*, *TSD2* shows similar results, but with a higher sensitivity to the data length.

It is not surprising that *HiSD* is correlated with *ratioTSD* and *ratioSD*, since it can be calculated as a combination of both of them.

The strong correlation between *SD1* and *TSD1* and between *SD2* and

TSD2 supports the theory by Brennan et al. [9] that the very commonly used indices *SD1* and *SD2* can be derived from statistical ones and therefore do not measure the non-linear characteristics of the Poincaré plot.

5.8 STATISTICAL INDICES

The very simple statistical index *pNN50* shows very good results in all four test cases and is not correlated to any other index. The index *SDSD* only shows no significant differences for age-dependent non-pathological data. This age independence was also found by Malpas et al. in [30] for subjects with diabetes.

The index *SDNN* only differentiates significantly for the third test case. Since *SD1* is only a scaled version of *SDNN*, their strong correlation is not surprising, but the correlation with *TSD1*, again underlines the findings by Brennan et al. in [9].

The index *CV* has significant differences only for data from arrhythmia patients before and after treatment and shows some sensitivity to the length of the data. Similar results were found by Toichi et al., who reported in [47] that *CV* also did not differentiate significantly in all of their test cases.

5.9 UNGROUPED INDICES

The index *CCM* does not show significant differences between two data sets in any test case, contrary to the findings by Karmakar et al. in [25], although they used subsets of the databases used in test case I and II. A reason for this could be different data lengths and unfiltered data in their tests, but since they do not specify them, this can not be verified. It could also be the case that the additional data sets in the combined database dampen the differentiation ability by *CCM*.

The index *contour*, which was introduced in this thesis, only shows significant differences between non-pathological and pathological indices for data lengths below 400 points and for the pre- and post-treatment datasets. After visual examination of the Poincaré plots for the first test case, a possible reason for the higher significance for shorter signals, can be that the shape of non-pathological Poincaré plots tends to not change substantially with less data points, which is not the case for many pathological ones. The strong correlation

to *area*, *range* and *LT* is not surprising, considering that all of them approximate the area of the Poincaré plots. The fact that these other indices do not show better results for shorter data lengths is probably due to their rougher approximations.

The index *ext* differentiates very significantly between data from before and after arrhythmia-treatment, as well as between pathological and non pathological data. In the later case, the index is sensitive to the data length, if it is below 400 points. This is also reported in Marciona et al., who found in [32] significant differences in *ext* for very long signals of healthy subjects and subjects with severe heart failure.

The index *range* tends to have more significant differences for shorter signals in test case I, but reaches significance only for a signal length of 100 and below. It has very significant differences for pre- and post-arrhythmia-treatment data, which is not the case for the age separated non-pathological data in test case IV. This is not the case in the findings by Moraes et al. in [35]. There are many possible reasons that the results could not be reproduced. They used for example data from special disorders of the autonomous nervous system (e.g., drug induced blockade of the sympathetic nervous system). They also applied the index on data sets from 24h recordings in contrast to the 15min long recordings in this work.

5.10 LIMITATIONS

The findings in this thesis are bound by several limitations. First, the merging of data from subjects with different pathologies, including subjects with and without arrhythmias, in the data set for the first and second test case, could have created a too large variety of Poincaré plots and therefore damped the differentiation ability of some Poincaré indices.

Furthermore, the data length used in literature is often almost 90 times larger than the one used in our case. Therefore, some parameters were possibly not chosen optimally, since the same as in the original papers were used.

The fact that an automatic filtering via clustering was used could further distort the results, since the cluster algorithm can not always find all clusters correctly in every case.

CONCLUSION

In section 1 the research questions for this master thesis were declared. The tests showed the following:

- Some indices showed a high sensitivity to the data length, while others, e.g., both ratios of the ellipse fitting indices, showed little to none.
- Many methods have significant differences between non-pathological and pathological data, e.g., the radii of inertia, A and B .
- Almost all Poincaré plot indices are able to differentiate between data before and after antiarrhythmia therapy. Indices, without significant differences between the data sets include all asymmetry indices except for PI and GI_{p3D} .
- Only a few of the indices had significant differences between older and younger healthy subjects. This was the case for almost all mechanical indices.
- There exist some strong correlation between indices inside an index group, e.g., almost all mechanical indices are strongly correlated with each other, but also between different index groups, e.g., SDI and T .

All in all only the quadrupoles Q_{xx} , Q_{yy} , Q_{zz} and the octupoles T_{xxx} , T_{xyy} , T_{xzz} showed excellent results for all tests cases.

The radii of inertia A and B , as well as the redefined asymmetry index GI_{p3D} came very close, but had difficulties with shorter data lengths. Since all of the above multipoles and the radii of inertia are strongly correlated, not all of them have to be considered. But this is not the case for GI_{p3D} , therefore it seems to capture different features of the Poincaré plot and should be considered for further research as well. Further tests should be conducted with Poincaré plots of different pathologies, e.g. diabetes, to see if the indices can capture these as well.

Since NP showed very good results for short signals, it should be

tested, if a higher threshold could lead to a similar result also for longer signals.

Although the newly defined index *contour* had not always significant differences for the test cases, it should be developed further, since in the current version only all its information is cumulated via its area. Therefore, it could for example not differentiate between varying shapes with the same area, although it is capable of finding those in principle.

To conclude, I recommend to use one of the quadrupoles or T_{xxx} , T_{xyy} , T_{xzz} , and additionally GI_{p3D} for future research.

APPENDIX

7.1 CORRELATION

Tables 19-20 show the correlation matrices of the indices which were significant in test case II. Tables 21-22 contain the correlation matrices of the indices which were significant in test case III.

Correlation was declared strong for $|r| > 0.85$.

Table 5: The correlation coefficients for the pathological database.

NumOfClusters	ClusterSizeRatio	PI	GI	PIp	GIp	Clp	Clp	width	width	A	B	Q _{xx}	Q _{yy}	Q _{zz}	T _{xx}	T _{yy}	T _{zz}	ext	T	ratioLT	SDI	SD2	ratioSD	TSDI	ratioTSD	HSD	pNNS0	SDSD
1	0.874	0.138	-0.0910	0.105	-0.0730	-0.299	0.0650	0.100	-0.00900	-0.0110	-0.0240	0.0290	0.0150	-0.0220	0.0300	0.0140	-0.0740	0.0470	0.159	-0.135	0.0840	0.0430	0.0790	0.0580	-0.0410	-0.0380	0.171	0.116
0.874	1	0.207	-0.0630	0.178	-0.140	-0.313	0.0780	0.0630	-0.0260	-0.0270	-0.0360	0.0400	0.0280	-0.0150	-0.00400	0.0280	-0.0150	-0.00400	0.167	-0.171	0.113	0.00100	0.124	0.0280	-0.0660	-0.0570	0.140	0.143
0.138	0.207	1	0.266	0.877	-0.444	-0.394	0.212	-0.149	-0.0680	-0.0720	-0.0910	0.0980	0.0760	-0.0910	0.100	0.0750	-0.204	-0.255	0.393	-0.390	0.429	0.109	0.444	0.129	-0.427	-0.426	0.261	0.460
-0.0910	-0.0630	0.266	1	0.434	-0.522	-0.176	0.167	-0.172	-0.116	-0.118	-0.126	0.133	0.119	-0.126	0.136	0.119	-0.368	0.0370	0.149	-0.231	0.287	-0.0380	0.392	-0.0130	-0.420	-0.431	0.113	0.254
0.105	0.178	0.877	0.434	1	-0.516	-0.458	0.270	-0.199	-0.0160	-0.0190	-0.0380	0.0450	0.0220	-0.0380	0.0480	0.0220	-0.233	-0.197	0.402	-0.415	0.445	0.0950	0.478	0.132	-0.473	-0.469	0.239	0.469
-0.0730	-0.140	-0.444	-0.522	-0.516	1	0.587	-0.138	0.120	0.0800	0.0840	0.0940	-0.100	-0.0870	0.0940	-0.101	-0.0850	0.201	0.0820	-0.234	0.286	-0.270	0.0470	-0.354	0.06000	0.344	0.349	-0.118	-0.274
-0.299	-0.313	-0.394	-0.176	-0.458	0.587	1	-0.109	-0.0390	-0.144	-0.142	-0.131	0.129	0.140	-0.130	0.125	0.141	0.247	0.0510	-0.244	0.216	-0.162	-0.100	-0.161	-0.133	0.149	0.139	-0.0460	-0.199
0.0650	0.0780	0.212	0.167	0.270	-0.138	-0.109	1	0.156	0.201	0.190	0.170	-0.163	-0.187	0.177	-0.164	-0.194	-0.246	0.0150	0.557	0.272	0.692	0.581	0.289	0.608	-0.314	-0.300	0.260	0.664
0.100	0.0630	-0.149	-0.272	-0.199	0.120	-0.0390	0.156	1	-0.0130	-0.0100	0.00200	-0.0130	0.00700	0.00500	-0.0160	0.00600	0.203	0.343	-0.239	0.275	-0.253	0.0200	-0.337	0.06900	0.334	0.346	-0.222	-0.264
-0.0260	-0.0260	-0.0680	-0.116	-0.0160	0.0800	-0.144	0.201	-0.0130	1	1	0.997	-0.994	-0.999	0.997	-0.992	-0.999	-0.00700	0.293	0.0530	0.295	0.0830	0.374	-0.306	0.371	0.299	0.295	-0.518	0.9860
-0.0110	-0.0270	-0.0720	-0.118	-0.0190	0.0840	-0.142	0.190	-0.0100	1	1	0.998	-0.996	-1	0.998	-0.994	-1	-0.00100	0.299	0.0380	0.301	0.0680	0.360	-0.312	0.358	0.305	0.302	-0.531	0.9820
-0.0240	-0.0360	-0.0910	-0.126	-0.0380	0.0940	-0.131	0.170	0.00200	0.997	0.998	1	-0.999	-0.999	1	-0.998	-0.999	0.0210	0.314	0.00900	0.318	0.0350	0.329	-0.333	0.327	0.327	0.324	-0.564	0.0490
0.0290	0.0400	0.0980	0.133	0.0450	-0.100	0.129	-0.163	-0.0130	-0.994	-0.996	-0.989	1	0.997	-0.999	0.999	0.997	-0.0360	-0.323	0.0130	0.328	-0.0150	-0.317	0.349	-0.314	-0.344	-0.341	0.578	-0.0290
0.0150	0.0290	0.0760	0.119	0.0220	-0.0870	0.140	-0.187	0.00700	-0.999	-1	-0.999	0.997	1	-0.998	0.994	1	-0.00200	-0.304	-0.0290	-0.305	-0.0610	-0.350	0.312	-0.348	-0.307	-0.303	0.541	-0.0740
-0.0350	-0.0350	-0.0910	-0.126	-0.0380	0.0940	-0.130	0.177	0.00500	0.997	0.998	1	-0.999	-0.998	1	-0.998	-0.998	0.0270	0.315	0.00800	0.320	0.0370	0.337	-0.341	0.334	0.335	0.332	-0.559	0.0500
0.0420	0.0420	0.100	0.136	0.0480	-0.101	0.125	-0.164	-0.0180	-0.992	-0.994	-0.988	0.999	0.994	-0.998	1	0.985	-0.0470	0.323	0.0170	-0.332	-0.09000	-0.316	0.362	-0.313	-0.357	-0.354	0.583	-0.0220
0.0140	0.0280	0.0750	0.119	0.0220	-0.0850	0.141	-0.194	0.00600	-0.999	-1	-0.999	0.997	1	-0.999	0.995	1	-0.00500	-0.302	-0.0350	-0.306	-0.0640	-0.360	0.320	-0.358	-0.314	-0.310	0.534	-0.0780
-0.0740	-0.0150	-0.204	-0.368	-0.233	0.201	0.247	-0.246	0.203	-0.00700	-0.0100	0.0210	-0.0360	-0.00200	0.0270	-0.0470	-0.00500	1	0.0430	-0.365	0.203	-0.353	-0.256	-0.299	-0.273	0.316	0.303	-0.259	-0.351
0.0470	-0.00400	-0.255	0.0370	-0.197	0.0820	0.0510	0.0150	0.343	0.293	0.299	0.314	-0.323	-0.304	0.315	-0.323	-0.302	0.0430	1	-0.152	0.339	-0.177	0.101	-0.365	0.0980	0.367	0.356	-0.330	-0.184
0.159	0.167	0.393	0.149	0.402	-0.234	-0.244	0.357	-0.239	0.0530	0.0390	0.00500	0.0130	-0.0290	0.00900	0.0170	-0.0350	-0.365	-0.152	1	-0.659	0.900	0.630	0.520	0.659	-0.508	-0.507	0.527	0.921
-0.135	-0.171	-0.390	-0.231	-0.415	0.286	0.216	-0.172	0.275	0.295	0.301	0.318	-0.328	-0.305	0.320	-0.332	-0.306	0.203	0.339	-0.659	1	-0.591	0.0730	-0.801	0.0490	0.768	0.781	-0.416	-0.607
0.0840	0.113	0.429	0.287	0.445	-0.270	-0.162	0.692	-0.253	0.0830	0.0680	0.0350	-0.0150	-0.0610	0.0370	-0.09000	-0.0640	-0.353	-0.177	0.900	-0.591	1	0.562	0.664	0.569	-0.671	-0.668	0.523	0.989
0.0430	0.00100	0.109	-0.0380	0.0950	0.0470	-0.100	0.381	0.0200	0.374	0.360	0.329	-0.317	-0.350	0.337	-0.316	-0.360	-0.256	0.101	0.630	0.0730	0.562	1	-0.196	0.989	0.125	0.130	0.239	0.574
0.0790	0.124	0.444	0.392	0.478	-0.354	-0.161	0.289	-0.337	-0.306	-0.312	-0.333	0.349	0.312	-0.341	0.362	0.320	-0.299	-0.365	0.509	-0.801	0.664	-0.166	1	-0.128	-0.980	-0.986	0.491	0.648
0.0560	0.0280	0.129	-0.0130	0.132	0.00600	-0.133	0.0600	0.00900	0.371	0.58	0.327	-0.314	-0.348	0.334	-0.313	-0.358	-0.273	0.0980	0.659	0.0490	0.589	0.989	-0.128	1	1.03	1.03	0.257	0.603
-0.0410	-0.0660	-0.427	-0.420	-0.473	0.344	0.149	-0.314	0.334	0.299	0.305	0.327	-0.344	-0.307	0.335	-0.357	-0.314	0.316	0.367	-0.508	0.788	-0.671	0.125	-0.980	0.103	1	0.993	-0.486	-0.649
-0.0680	-0.0570	-0.426	-0.431	-0.469	0.349	0.139	-0.300	0.346	0.295	0.302	0.324	-0.341	-0.303	0.332	-0.354	-0.310	0.303	0.356	-0.507	0.781	-0.668	0.130	-0.986	0.106	0.993	1	-0.482	-0.646
0.171	0.140	0.261	0.113	0.239	-0.118	-0.0460	0.260	-0.222	-0.518	-0.531	-0.564	0.578	0.584	-0.531	-0.559	0.583	0.534	-0.330	0.527	-0.416	0.523	0.239	0.491	0.257	-0.486	-0.482	1	0.517
SDSD	0.116	0.460	0.254	0.469	-0.274	-0.199	0.064	-0.264	0.0660	0.0820	0.0490	-0.0290	-0.0740	0.0500	-0.0220	-0.0780	-0.351	-0.184	0.921	-0.607	0.989	0.574	0.648	0.603	-0.649	-0.646	0.517	1

LIST OF TABLES

Table 1	Results of tests cases II-IV.	40
Table 2	Parameters of distribution for the indices in test case II.	42
Table 3	Parameters of distributions for the indices in test case III.	44
Table 4	Parameters of distributions for the indices in test case IV.	46
Table 5	The correlation coefficients for the pathological database.	66
Table 6	The correlation coefficients for the non-pathological database.	67
Table 7	The correlation coefficients for the pre-arrhythmia-treatment database.	68
Table 8	The correlation coefficients for the post-arrhythmia-treatment database.	69

LIST OF FIGURES

Figure 1	Schematic representation of an ECG-signal. 9
Figure 2	Simplified representation of the regulatory systems and influences of heart beat generation. 10
Figure 3	A typical unfiltered Poincaré plot of a non-pathological heart rate data set. 11
Figure 4	An unfiltered 3D Poincaré plot of non-pathological heart rate data set. 12
Figure 5	Visualization of the separation of a filtered Poincaré plot along the line of identity to measure asymmetry. 16
Figure 6	Approximated density function with the corresponding 3D histogram. 19
Figure 7	A filtered non-pathological Poincaré plot with the ellipse fitting and dispersion indices. 20
Figure 8	The measurements for <i>ext</i> and for the LT indices. 22
Figure 9	The alternative Poincaré plot according to Moraes et al., with marked measurements for the index <i>range</i> . 24
Figure 10	The distance function and the corresponding Poincaré plot. 26
Figure 11	Clustered Poincaré plot of a subject with arrhythmia. 30
Figure 12	Deletion of the incorrect RR-interval for a 2D and a 3D Poincaré plot. 31
Figure 13	Indices with at least one but not all p-values below the significance threshold of 0.05. 36
Figure 14	Additional indices with at least one, but not all p-values below the significance threshold of 0.05. 37
Figure 15	Indices with significant, and for some data lengths, very significant differences. 38

Figure 16	Indices with a very significant differences for all data lengths. 38
Figure 17	Indices, which have no p-value below the significance threshold of 0.05 for any data length. 39
Figure 18	Boxplots for Karmakar-Guzik asymmetry index GI_p and for the peak index NP . 49
Figure 19	Strong correlations between Poincaré plot indices of the pathological data set, which showed significant results in test case II. 51
Figure 20	Strong correlations between Poincaré plot indices of the non-pathological data set, which showed significant results in test case II. 52
Figure 21	Strong correlations between Poincaré plot indices of the pre-arrhythmia-treatment data set, which showed significant results in test case III. 53
Figure 22	Strong correlations between Poincaré plot indices of the post-arrhythmia-treatment data set, which showed significant results in test case III. 54

BIBLIOGRAPHY

- [1] U. R. Acharya et al. "Heart rate variability: a review". In: *Medical and Biological Engineering and Computing* 44.12 (2006), pp. 1031–1051.
- [2] American Heart Association Inc.; European Society of Cardiology. "Guidelines – Heart rate variability". In: *Eur. Heart J.* 17 (1996), pp. 354–381.
- [3] M. Bachler et al. "Online and Offline Determination of QT and PR Interval and QRS Duration in Electrocardiography". In: *Pervasive Computing and the Networked World*. Ed. by Q. Zu, B. Hu, and A. Elçi. Vol. 7719. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, 2013, pp. 1–15.
- [4] D. S. Baim et al. "Survival of patients with severe congestive heart failure treated with oral milrinone". In: *J. Am. Coll. Cardiol.* 7.3 (Mar. 1986), pp. 661–670.
- [5] A. I. Batchinsky, J. Salinas, and L. C. Cancio. "Assessment of the Need to Perform Life-Saving Interventions using Comprehensive Analysis of the Electrocardiogram and Artificial Neural Networks". In: *RTO-MP-HFM-182: Use of Advanced Technologies and New Procedures in Medical Field Operations*. 2010, 39:1–16.
- [6] D. Bennett. *Bennett's Cardiac Arrhythmias: Practical Notes on Interpretation and Treatment*. 8th ed. Wiley, 2012.
- [7] J. Bigger et al. "The ability of several short-term measures of RR variability to predict mortality after myocardial infarction." In: *Circulation* 88.3 (1993), pp. 927–934.
- [8] P. Bjerregaard. "Premature beats in healthy subjects 40-79 years of age." In: *European heart journal* 3.6 (Dec. 1982), pp. 493–503.
- [9] M. Brennan and M. Palaniswami. "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?" In: *Biomedical Engineering, IEEE Transactions on* 48 (2001), pp. 1342–1347.

- [10] A. Bundkirchen and R. H. Schwinger. “Epidemiology and economic burden of chronic heart failure”. In: *European Heart Journal Supplements* 6.suppl D (2004), pp. D57–D60.
- [11] M. Costa, A. Goldberger, and C. K. Peng. “Broken Asymmetry of the Human Heartbeat: Loss of Time Irreversibility in Aging and Disease”. In: *Physical Review Letters* 95.19 (Nov. 2005), p. 198102.
- [12] A. E. Epstein et al. “Events in the cardiac arrhythmia suppression trial (CAST): Mortality in the entire population enrolled”. In: *Journal of the American College of Cardiology* 18.1 (July 1991), pp. 14–19.
- [13] H. D. Esperer, C. Esperer, and R. J. Cohen. “Cardiac arrhythmias imprint specific signatures on Lorenz plots.” In: *Annals of noninvasive electrocardiology : the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc* 13.1 (Jan. 2008), pp. 44–60.
- [14] M. Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *Kdd*. Vol. 96. 1996, pp. 226–231.
- [15] L. Glass et al. “Bifurcation and chaos in a periodically stimulated cardiac oscillator”. In: *Physica D: Nonlinear Phenomena* 7.1 (1983), pp. 89–101.
- [16] A. L. Goldberger et al. “PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals”. In: *Circulation* 101.23 (June 2000), e215–e220.
- [17] S. D. Greenwald, R. S. Patil, and R. G. Mark. “Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information”. In: *Computers in Cardiology 1990, Proceedings*. IEEE. 1990, pp. 461–464.
- [18] P. Grossman and E. W. Taylor. “Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions”. In: *Biological psychology* 74.2 (2007), pp. 263–285.
- [19] P. Guzik et al. “Heart rate asymmetry by Poincaré plots of RR intervals”. In: *Biomedizinische Technik/Biomedical Engineering* 51.4 (Oct. 2006), pp. 272–275.

- [20] M. Hirose et al. "Heart Rate Variability during Chemical Thoracic Sympathectomy". In: *Anesthesiology* 89.3 (Sept. 1998), p. 666.
- [21] K. Hnatkova et al. "Numeric processing of Lorenz plots of R-R intervals from long-term ECGs". In: *Journal of Electrocardiology* 28 (Jan. 1995), pp. 74–80.
- [22] V. Ivancevic and T. Ivancevic. *Complex Nonlinearity: Chaos, Phase Transitions, Topology Change, and Path Integrals*. Springer complexity. Springer, 2008.
- [23] N. Iyengar et al. "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics". In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 271.4 (1996), R1078–R1084.
- [24] C. K. Karmakar et al. "Defining asymmetry in heart rate variability signals using a Poincaré plot". In: *Physiological measurement* 30.11 (Oct. 2009), pp. 1227–1240.
- [25] C. K. Karmakar et al. "Complex correlation measure: a novel descriptor for Poincaré plot." In: *BioMedical Engineering On-Line* 8.1 (2009), p. 17.
- [26] C. Lerma et al. "Poincaré plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients". In: *Clinical physiology and functional imaging* 23.2 (2003), pp. 72–80.
- [27] M. Lewkowicz et al. "Description of complex time series by multipoles". In: *Physica A: Statistical Mechanics and its Applications* 311.1 (2002), pp. 260–274.
- [28] E. N. Lorenz. "Deterministic nonperiodic flow". In: *Journal of the atmospheric sciences* 20.2 (1963), pp. 130–141.
- [29] T. H. Makikallio et al. "Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction." In: *Journal of the American College of Cardiology* 28.4 (Oct. 1996), pp. 1005–1011.
- [30] S. C. Malpas and T. J. Maling. "Heart-rate variability and cardiac autonomic function in diabetes". In: *Diabetes* 39.10 (1990), pp. 1177–1181.
- [31] P. Mansier et al. "Linear and non-linear analyses of heart rate variability: a minireview". In: *Cardiovascular research* 31.3 (1996), pp. 371–379.

- [32] F. Marciano et al. “Quantification of Poincare’maps for the evaluation of heart rate variability”. In: *Computers in Cardiology 1994*. IEEE. 1994, pp. 577–580.
- [33] D. Michaels et al. “Chaotic activity in a mathematical model of the vagally driven sinoatrial node.” In: *Circulation research* 65.5 (1989), pp. 1350–1360.
- [34] G. B. Moody and R. G. Mark. “The impact of the MIT-BIH arrhythmia database”. In: *IEEE Eng Med Biol Mag* 20.3 (2001), pp. 45–50.
- [35] R. S. Moraes et al. “Three-dimensional return map: a new tool for quantification of heart rate variability”. In: *Autonomic Neuroscience* 83.1-2 (Sept. 2000), pp. 90–99.
- [36] G. A. Ng. “Treating patients with ventricular ectopic beats”. In: *Heart* 92.11 (2006), pp. 1707–1712.
- [37] M. Nichols et al. “European cardiovascular disease statistics 2012”. In: *European Heart Network, Brussels, European Society of Cardiology, Sophia Antipolis* (2012), P104.
- [38] K. Orth-Gomer et al. “Frequency of extrasystoles in healthy male employees.” In: *British heart journal* 55.3 (Mar. 1986), pp. 259–264.
- [39] J. Piskorski and P. Guzik. “Filtering poincare plots”. In: *Computational Methods in Science and Technology* (2005), pp. 39–48.
- [40] A. Porta et al. “Time reversibility in short-term heart period variability”. In: *Computers in Cardiology, 2006*. IEEE. 2006, pp. 77–80.
- [41] U. Rajendra Acharya et al. “Heart rate variability: a review”. In: *Medical and Biological Engineering and Computing* 44.12 (Nov. 2006), pp. 1031–1051.
- [42] V. L. Schechtman, R. K. Harper, and R. M. Harper. “Development of heart rate dynamics during sleep-waking states in normal infants.” In: *Pediatric research* 34.5 (Nov. 1993), pp. 618–623.

- [43] A.-L. Smith, K. J. Reynolds, and H. Owen. “Correlated Poincaré indices for measuring heart rate variability.” In: *Australasian physical & engineering sciences in medicine / supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine* 30.4 (Dec. 2007), pp. 336–341.
- [44] P. K. Stein et al. “Clinical and demographic determinants of heart rate variability in patients post myocardial infarction: insights from the cardiac arrhythmia suppression trial (CAST)”. In: *Clinical cardiology* 23.3 (2000), pp. 187–194.
- [45] P. K. Stein et al. “Sometimes higher heart rate variability is not better heart rate variability: results of graphical and nonlinear analyses”. In: *Journal of cardiovascular electrophysiology* 16.9 (2005), pp. 954–959.
- [46] T. Thong. “Geometric Measures of Poincare Plots for the Detection of Small Sympathovagal Shifts”. In: *IEEE engineering in medicine and biology magazine* 1 (2007), p. 4641.
- [47] M. Toichi et al. “A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R-R interval.” In: *Journal of the autonomic nervous system* 62.1-2 (Jan. 1997), pp. 79–84.
- [48] M. P. Tulppo et al. “Quantitative beat-to-beat analysis of heart rate dynamics during exercise.” In: *The American journal of physiology* 271.1 Pt 2 (July 1996), H244–52.
- [49] D. Vorberg and S. Blankenberger. “Die Auswahl statistischer Tests und Maße”. In: *Psychologische Rundschau* 50.3 (1999), pp. 157–164.
- [50] A. Voss et al. “Methods derived from nonlinear dynamics for analysing heart rate variability”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367.1887 (2009), pp. 277–296.
- [51] M. A. Woo et al. “Patterns of beat-to-beat heart rate variability in advanced heart failure.” In: *American heart journal* 123.3 (Mar. 1992), pp. 704–710.
- [52] L. Xia and J. Jing. “An Ensemble Density-based Clustering Method”. In: *International Conference on Intelligent Systems and Knowledge Engineering 2007* (Oct. 2007).