UNIVERSITÀ DEGLI STUDI DI MILANO

Doctoral School of Computer Science
Department of Computer Science

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science
(XXVII Cycle)

FEATURE EXTRACTION AND CLASSIFICATION THROUGH ENTROPY MEASURES

INF/01

DOCTORAL DISSERTATION OF:
Md AKTARUZZAMAN

ADVISOR:
Prof. Roberto Sassi

DIRECTOR OF DOCTORAL SCHOOL:
Prof. Ernesto Damiani

Academic Year 2013/'14

Dedicated to
—*My beloved wife*

"Research is to see what everybody else has seen,
and to think what nobody else has thought."

— Albert Szent-Gyorgyi

ACKNOWLEDGMENTS

All praise to the Omnipotent, who has created me as a human, and then provided me the ability to pursue higher study. First and foremost, I wish to thank my honorable adviser, Professor Roberto Sassi, head of the Biomedical image and Signal Processing (BiSP) Laboratory, Crema, University of Milan, Italy. I feel proud to be the first international PhD student of him. I appreciate all contributions of his valuable time, ideas and advices those make productive and stimulating my PhD experiences. Professor Sassi has been supportive to me since the day I started working with him both academically and mentally through the rough road to finish the thesis. He helped me to come up with the thesis title and guided me over the years of completion. During the most difficult times of my research, he not only gave me the moral support but also provided me the full freedom to move on. Besides group working, Professor Sassi encouraged me to think and implement independently to be a leading researcher in future. Professor Sassi always inspired me to make things perfect, be honest in representing reports, be self confident, and high ambitious.

Then, I would like to give thanks to the director of the Doctorate School Professor Ernesto Damiani, who has provided all types of academic and administrative supports for participating in different summer schools, seminars, and presentations. Dozens of internal and external professors have taught me and cooperated me directly or indirectly during my PhD thesis. I would like to mention with full respects the names of Prof. Vincenzo Piuri, Prof. Sergio Cerutti, Prof. Anna Maria Bianchi, Prof. Luca T. Mainardi, Prof. Valentina Corino, Prof. Mattia Monga, Prof Walter Cazzola, Prof. Annamaria R. Varkonyi-Koczy, Prof. Simona Ferrante, and Prof. Rafael Accorsi.

I also want to thank the referees Prof. Sergio Cerutti, Prof. Olivier Meste, and Prof. Luca Citi for their time spent in reading my thesis and for giving me the valuable suggestions to improve the quality of my work.

Massimo Walter Rivolta, who is my best and bosom colleague accompanied me at the first day of my PhD school. He cooperated me in managing all official procedures. Besides this, he also helped me in times and out of times to understand some topics those I feel difficult to understand. His contribution during my PhD thesis is not comparable anyway. I am also grateful to Mr. Matteo Migliorini, who has shared his knowledge in working together. I found him cooperative and very friendly in group works. Besides these, special thanks go to the people that volunteered for acceleration signals acquisitions that I needed to evaluate my work on physical activity recognition. Among them I want to mention my new colleague Mr. Ebadollah Kheirati Roonizi, Gerson Antunes Soares, and Mrs. Teresa Rutigliano.

I thank all my senior and junior colleagues, who helped me by their valuable suggestions. Among them I want to mention, with no particular order, Dr. Angelo Genovese, Dr. Paolo Arcaini, Dr. Ruggero Donida Labati, and Dr. Ravi Jhawar, Mrs. Giovanna Janet Lavado, and Mr. Bruno Guillon.

A special thank goes to Mrs. Claudia Piana, an officer of the department of Computer Science, Crema, Milan University. She cordially helped me in some issues like finding residence, opening bank accounts, translating documents,etc. She did never say “no" in any assistance.

Lastly, I wish to give thanks to my family members, especially my wife Mrs Mahbuba Sharmin, who has shared all good and bad experiences with me during our stay in abroad. She has sacrificed a lot for the sake of my PhD. I will never forget the contributions of parents in law Mr. Alhaj Md Mahbubur Rahman and Mrs. Alhaj Firoza Begum Fatema, my parents Mr. Alhaj Md Abdul Mozid Sheikh and Mrs. Sufia Khatun, my elder brothers Mr. Harun-Or-Rashid, Mr. Rezaul Haque, my younger brothers Hamidul Haque, and Nuruzzaman Khokon.

Finally, I am grateful to my younger uncle Mr. Abdul Bari, my respectful teachers, my friends here and abroad, my relatives and well wishers who have inspired me in time and out of time for doing higher study.

ABSTRACT

Entropy is a universal concept that represents the uncertainty of a series of random events. The notion “entropy" is differently understood in different disciplines. In physics, it represents the thermodynamical state variable; in statistics it measures the degree of disorder. On the other hand, in computer science, it is used as a powerful tool for measuring the regularity (or complexity) in signals or time series. In this work, we have studied entropy based features in the context of signal processing.

The purpose of feature extraction is to select the relevant features from an entity. The type of features depends on the signal characteristics and classification purpose. Many real world signals are nonlinear and nonstationary and they contain information that cannot be described by time and frequency domain parameters, instead they might be described well by entropy.

However, in practice, estimation of entropy suffers from some limitations and is highly dependent on series length. To reduce this dependence, we have proposed parametric estimation of various entropy indices and have derived analytical expressions (when possible) as well. Then we have studied the feasibility of parametric estimations of entropy measures on both synthetic and real signals. The entropy based features have been finally employed for classification problems related to clinical applications, activity recognition, and handwritten character recognition. Thus, from a methodological point of view our study deals with feature extraction, machine learning, and classification methods.

The different versions of entropy measures are found in the literature for signals analysis. Among them, approximate entropy (ApEn), sample entropy (SampEn) followed by corrected conditional entropy (CcEn) are mostly used for physiological signals analysis. Recently, entropy features are used also for image segmentation. A related measure of entropy is Lempel-Ziv complexity (LZC), which measures the complexity of a time-series, signal, or sequences. The estimation of LZC also relies on the series length.

In particular, in this study, analytical expressions have been derived for ApEn, SampEn, and CcEn of an auto-regressive (AR) models. It should be mentioned that AR models have been employed for maximum entropy spectral estimation since many years. The feasibility of parametric estimates of these entropy measures have been studied on both synthetic series and real data. In feasibility study, the agreement between numeral estimates of entropy and estimates obtained through a certain number of realizations of the AR model using Montecarlo simulations has been observed. This agreement or disagreement provides information about nonlinearity, nonstationarity, or nonGaussinaity presents in the series. In some classification problems, the probability of agreement or disagreement have been proved as one of the most relevant features.

After feasibility study of the parametric entropy estimates, the entropy and related measures have been applied in heart rate and arterial blood pressure variability analysis. The use of entropy and related features have been proved more relevant in developing sleep classification, handwritten character recognition, and physical activity recognition systems.

The novel methods for feature extraction researched in this thesis give a good classification or recognition accuracy, in many cases superior to the features reported in the literature of concerned application domains, even with less computational costs.

CONTENTS

LIST OF TABLES

1

INTRODUCTION

The focus of this thesis is on the research of innovative methods for features extraction and their application in classification purposes. In particular, the emphasis has been given on the study and implementation of original methods for entropy related features extraction for classifications of signals acquired from the human body (*i.e.* physiological signals) and to recognize handwritten digits. This study is composed of theoretical derivations of entropy and related metrics, the feasibility study on both synthetic and real data, and finally the applications of entropy and entropy related features in various possible application domains.

In order to have a better understanding of this research, first an introduction to entropy and related measures are presented, followed by objectives and contributions of this thesis.

1.1 GENERAL CONCEPT OF ENTROPY

The term “entropy" is used in information theory to quantify the amount of information or uncertainty inherent in a system. *e.g.* “Desk entropy", as shown in figure 1.1, represents the degree of disorder of a PhD student’s desk space and the inability to find something, when she/he really needs it. The concept of entropy was coined first by Rudolf Clausius in thermodynamics in order to explain the energy loss in any irreversible process of a thermodynamics system. An increase in entropy of a system refers to the decrease in order of that system. Disorder in statistical mechanics means unpredictability due to the lack of knowledge.

The concept of entropy has been found applicable in the field of information theory since the mid of 20th century, describing an analogous loss of data in information transmission system. In 1948, Claude Shannon set out to mathematically quantify the statistical nature of “lost information” in phone-line signals. To this aim, he applied the general concept of entropy in information theory, and developed a function for

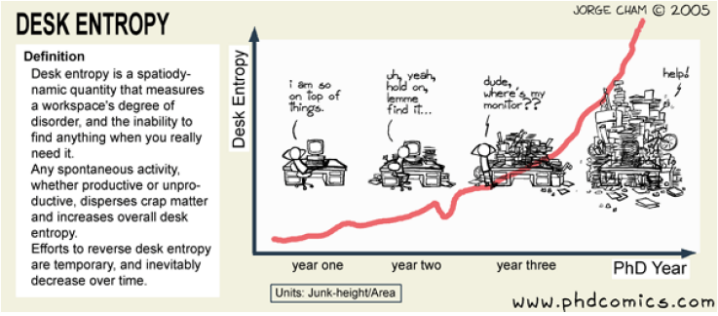


Figure 1.1: Entropy of a PhD students desk at different years.
Source:www.phdcomics.com

estimating the entropy. In Physics, entropy is used to measure system disorder. In Computer Science, it is used to measure the regularity (or complexity) of time-series or signals. The growing rate of entropy of a sequence with increasing the length N of the series is defined as the entropy rate.

Entropy has been represented with many names in many application domains since introduced. The major application areas of entropy are cryptography [1], data compression [2, 3], uncertainty or predictability [4, 5, 6] in information theory [3]. Besides this, some applications of entropy and entropy rate are found in signal processing [7, 8, 9], text classification [10], pattern recognition [11, 12, 13], image and speech signals processing [14, 15].

1.1.1 LEMPEL-ZIV COMPLEXITY: A RELATED METRIC

Lempel-Ziv complexity (LZC), first introduced by Lempel and Ziv [16], measures the rate of generation of new patterns of a sequence. It is closely related to the entropy rate of the sequence. It has become an standard algorithm for file compression on computers [3]. To compute LZC of a time-series, the sequence is first converted to a symbolic sequence, and then the sequence is parsed to get distinct words. Finally, the LZC is estimated from the parsing words. There are different methods of parsing. The one popular method is proposed by the inventor [16], and another attractive one is illustrated by Cover and Thomas [3]. The only difference between them is in parsing technique.

1.2 BACKGROUND CONCEPTS

Some fundamental concepts about the signals and the methods discussed in this thesis are provided in this section.

1.2.1 SIGNAL FUNDAMENTALS

A signal is a single-valued representation of information as a function of an independent variable (*e. g.* time) [17]. The type of information may have real or complex values.

A signal may be a function of another type of variables instead of time, and even two or more independent variables. If the signal is a function of a single independent variable, then it is called one-dimensional signal. On the other hand, if it is a function of two or more independent variables, it is called multidimensional signal (for D-independent variables, it is called D-dimensional signal). An example of a signal of two independent variables is an image. These independent variables represent the spatial coordinates.

Signals can be classified further as continuous-time and discrete-time, depending on the characteristics of the independent variable (time). If they are defined on continuous interval of time, then they are called continuous time signals. On the other hand, signals defined at only certain specific values of time are called discrete-time signals.

The mathematical analysis and processing of signals require the availability of a mathematical representation for the signal itself. This mathematical description, often referred to as the signal model, leads to another important classification of signals [18]. Any signal that can be uniquely expressed by an explicit mathematical expression, or a well-defined rule is called deterministic. This term is used to emphasize the fact that without any uncertainty, all future values can be predicted exactly if past values of the signal are known. The signals for which it is impossible to predict an exact future value, even if its entire past information is available, are referred to as stochastic signals. There is some aspect of the signal that is random, and hence it is often referred to as random signal. Random signals cannot be expressed by any mathematical expression. In fact, there are cases where such a functional relationship is unknown or too complicated for any practical use. *e. g.* speech signals cannot be described by any mathematical expression. Some other physiological signals of this category include electrocardiogram (ECG) , which provides information about the electrical activity of the cardiovascular system, and accelerometry signal [19], which is captured by an accelerometer sensor.

1.2.2 FEATURE EXTRACTION AND CLASSIFICATION

Classification deals with mathematical and technical aspects of grouping different signals through their descriptive information. The objective of signal classification is achieved in a three step procedure as shown in figure 1.2. The input signal may contain noise or artifacts [1]. So, it is preprocessed first. Then features/parameters suitable for classification are extracted. The preprocessed signal is then classified in the final step based on the extracted features.

Feature extraction addresses the problem of finding the most informative and compact set of features for improving the efficiency of classification or recognition. Researchers in machine learning, soft-computing, and statistics who are interested in predictive modeling, are exploiting their efforts together to advance the feature extraction problems.

Machine learning problems arise when a classification or recognition task is defined by a series of cases instead of some specific predefined rules. Such problems are found in a wide variety of application domains, ranging from medical applications in diagnosis, prognosis, drug discovery, engineering applications in robotics and pattern

1 Something observed during scientific investigation, but naturally is not present

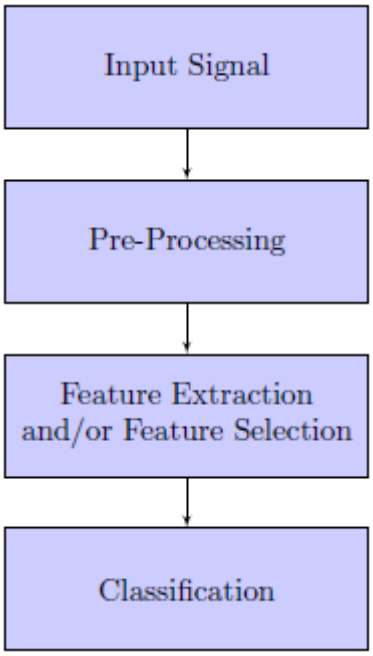


Figure 1.2: A typical classification procedure

recognition (optical character, speech and face recognition), and internet application (text classification) [20]. Given a number of training samples associated with the target outcome, the machine learning process consists of finding the relationship between the training samples and the target outcomes from the training samples. There is a lot of advancements in machine learning since mid 1950s, when it was introduced first by Samuel [21]. Feature extraction lies at the core of theses advances. The type of features depends on the signal characteristics and its application.

Feature extraction may lead to another additional step of feature selection. In the feature extraction step, information relevant to the signal classification is extracted from the input data first, and a feature vector v of m-dimension is formed. The feature vector may contain irrelevant or less relevant information, which just increases the complexity of classification without any significant contribution to the classification task. In the feature selection step, the vector v is transformed into another vector, which has the dimensionality n (n<m). If the feature extractor is properly designed, the feature vector is matched to the pattern classifier with low dimension. Then, there is no need of feature selection. However, machine learning algorithms are highly computationally intensive to the number of features; with increasing the number of features the system requirements increase, as well as the training and classification times of the system grow exponentially with the number of features. The correlated features do not contribute anymore in accuracy of the system than comes from their single one. So, the features should be decorrelated before feeding them into the classifier to reduce requirements ans hence computational cost of the system [22].

1.3 OBJECTIVES OF THIS THESIS

The estimates of entropy and entropy rate have become very popular for signal analysis. However, the use of these tools are limited by the series (signal) length. Some of them are not defined for short series. On the other hand, some are defined by accepting their bias estimates and there is a large difference between the estimates for short and longer series. In fact, real applications often require processing of very short series with sufficient reliability. Thus, the issues of convergence may arise when estimating entropy of short series. Research for developing methods to analyze very short series with enough reliability is still unveiled. A related problem of convergence arises in spectral analysis, where long time stationary series are required to achieve lower variance of the estimates, and parametric spectral analysis of time series is commonly performed since the works of Ulrich et al., [23] and Kay & Marple [24]. The simple stationary stochastic processes *i.e* autoregressive (AR) models have been used as maximum entropy spectral analysis.

The robustness of any estimation depends on the availability of the amount data used for the estimation. The amount of data in any estimation can be used to compensate missing knowledge and vice-versa. That means uninformative priors can be assumed from plenty of data. On the other hand, stronger assumptions (*e.g.* Gaussian distribution, models that characterize the data, etc) about the data should be made for any estimation on a limited amount of data. However, complex models (*i.e.* weaker assumptions) may lead to high variance (poor estimation). On the other hand, stronger assumptions may lead to high bias. So, depending on the amount of data available, it is wise to devise the optimal model complexity in order to limit the overall estimation error which is the sum of these two components.

The direct estimation of entropy, differential entropy or entropy rate from their expressions are not used in practical, even if they are major tools in signal processing. The estimated values are used because entropy depends on the probability density (or probability distribution) of the data, which is unknown; and users do not know the robustness of these estimates. Even if the probability density is known, another major difficulty arises due to the numerical integration in the definition of differential entropy, and hence entropy rate [25]. So, deriving values theoretically from their expressions and, then comparing numerically estimated values might be helpful to assess the measures. In addition some more information may also be obtained.

The objectives of this thesis are:

1. To develop an alternative parametric method for entropy estimation on very short series
2. To derive analytical expressions for different entropy measures (when possible)
3. To study the feasibility of the developed method on synthetic and real data
4. To introduce new entropy related features
5. To use these new features in some real applications

1.4 THESIS NOVELTY AND CONTRIBUTION

The performed research started with the preliminary study on entropy and entropy rates. The entropy metrics were chosen because the literature is rich with various versions of entropy and entropy rate in real applications. However, some dependence of the estimates of this popular metrics generates a headache for its real applications in very short series. So, the reduction of dependency may make the metric more popular for practical applications. To this aim, we studied alternative methods for estimating entropy from very short series and then extracted features based on this new method. New features have been derived based on this new metrics of entropy. Finally, the usage of these features have been found effective in some real world applications.

In this study, we developed a linear parametric approach of entropy estimation for finite short series. This estimation truly comes from a linear AR model, which cannot capture any nonlinear behavior of the series. Our hypothesis is based on the assumption that signals for very short series seems to be stationary. The entropy itself is a nonlinear statistic. But, the estimates come from the model is a linear statistic. The readers may be confused with linear estimates of entropy. To aid the readers, we studied the feasibility of the new parametric approach. If numerical estimation of entropy is truly reflected by the non-linearity of the series, then the numerical estimations should be different from the estimations obtained for the models. On the other hand, if the two estimates agree *i.e.*, numerical estimations are within 95% standard range of the estimations obtained through the model, then it implies entropy does not capture any more information than the models can. Inspired by this, we have introduced new features related to the entropy and parametric estimates of entropy.

We derived analytical expressions for parametric approach of entropy estimations (when possible), so that we can compare how far the real estimates are from the true theoretical one, which is impossible in case of the traditional entropy estimates. The feasibility of the method has been studied on both synthetic and publicly available real data, which comprise a large set of subjects. After feasibility study, the entropy based features have been applied in real applications such as heart rate and arterial blood pressure variability analysis, sleep classification, physical activity recognition. Besides physiological time-series analysis, we have extended the study for extracting features from patterns. Applications of entropy features reported effective for segmentation and image processing applications [26, 27]. We have used entropy features directly in pattern recognition. In particular, we have studied features for recognition of Bengali handwritten numerals.

In addition, the Lempel-Ziv complexity and its estimates have been studied. The parsing methods of LZC estimation and their pros and cons have been explained in detail. The parametric estimates of entropy and LZC for finite series have been compared during sleep classification.

Thus the total efforts payed on this thesis include studying entropy, entropy rates, Lempel-Ziv complexity, parametric entropy estimations, their theoretical expressions derivation, feasibility study of parametric entropy estimation, feature extraction, fea-

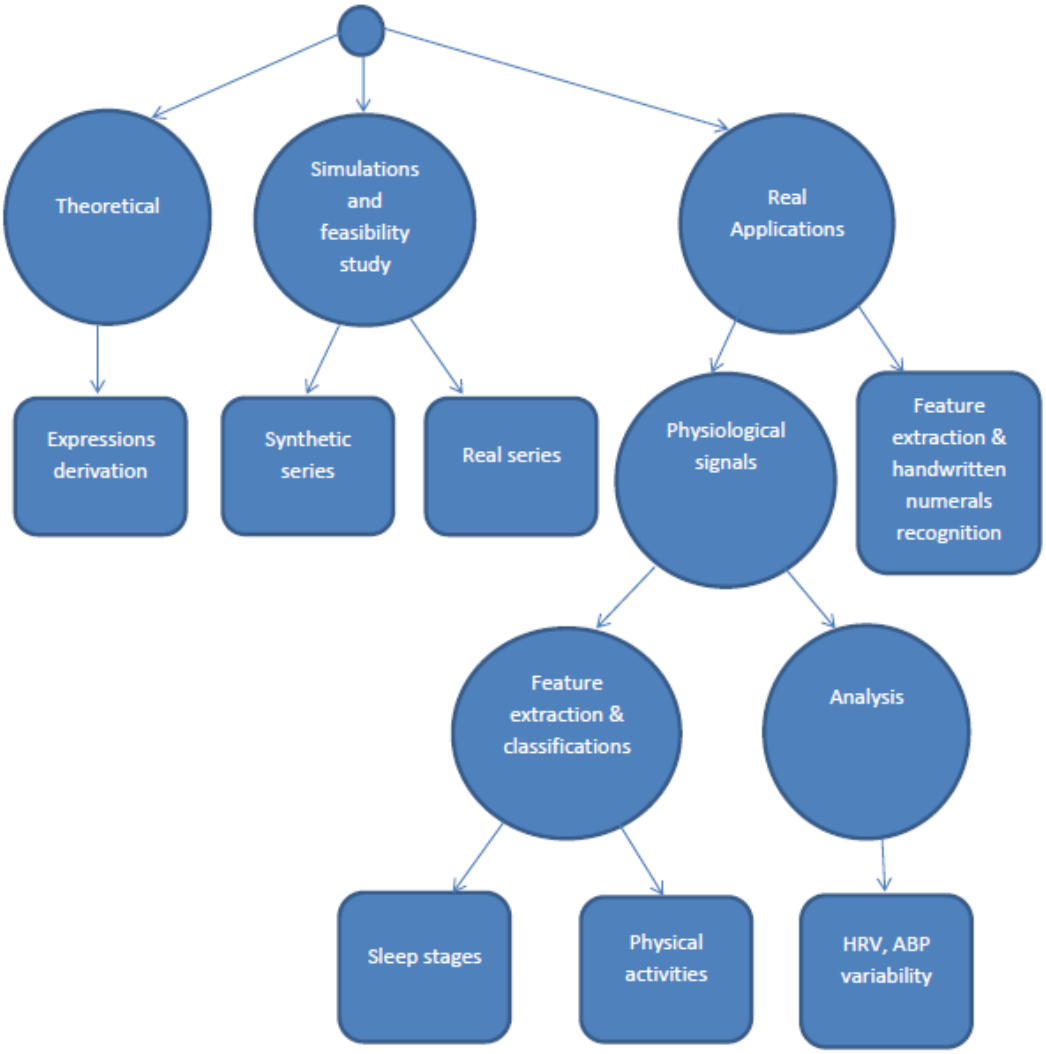


Figure 1.3: General overview of the thesis.

ture dimensionality reduction, and usage of the extracted features in real applications especially in physiological signals processing and handwritten character recognition.

1.5 THESIS STRUCTURE

The general overview of this thesis is illustrated by figure 1.3. The structure of the thesis is divided into 3 major parts: (i) theoretical analysis, (ii) simulations, and (iii) real applications. The derivation of expressions for parametric entropy estimations have been explained in theoretical part. Then, the feasibility of parametric entropy estimations have been justified using simulations on both synthetic and real data. After feasibility study, the proposed method has been used for analysis of physiological signals such as HRV and arterial blood pressure (ABP) variability. Finally, the use of entropy and related measures have been applied for classification of sleep stages, physical activities, as well as handwritten numerals recognition.

This thesis is organized as follows:

- Chapter 2 contains background details about the entropy and entropy related metrics. The numerical estimations of some entropy measures of a time series have been given, including the Lempel-Ziv complexity of a time series. The entropy of continuous and discrete time variables, and their difference has been explained. It also includes mathematical definitions of some common measures of entropy and entropy rates in the literature.

- Chapter 3 contains discussions about the parametric approaches. The parametric estimations of different entropy measures. The theoretical derivations of approximate entropy, sample entropy, and conditional entropy of a stochastic process have been explained. This chapter also includes the simulations on synthetic series, and the convergence of the expected values of entropy of the synthetic signals generated through the models to their corresponding theoretical values. The relationship between entropy rate and Lempel-Ziv complexity of a Gaussian stochastic process has also been discussed in this chapter.

- Chapter 4 is dedicated to the validation checking of the proposed parametric estimations of entropy. The feasibility study of the parametric entropy estimations has been studied on synthetic series and real data. The feasibility of Lempel-Ziv complexity on short series has also been studied in this chapter. Besides this, a comparison on the robustness of Lempel-Ziv complexity and entropy for short series has been also provided in this chapter.

- Chapter 5 presents the background of different physiological signals acquired from the human body, especially the electrocardiogram (ECG), the physiology of sleep and arterial blood pressures to provide the basic knowledge to the computer scientists about the physiological signals and their processing. Methods for extracting entropy features from some physiological signals have been explained in this chapter. Finally, the effective use of these entropy features in some real applications such as sleep and physical activity classifications have been included. A brief discussion about the obtained results have also been added for every method.

- Chapter 6 describes the method for extracting features from Bengali handwritten digits. The basic concept about the Bengali numerals has been provided at the starting of the chapter, followed by a review on existing features for their recognition. Then a set of features including the entropy one (corrected conditional entropy), and their extraction from the Bengali numerals have been illustrated. The complete handwritten numeral recognition method has been explained, including feature dimension reduction (*i.e.*) best feature selection strategy.

- Chapter 7 summarizes the work and obtained results, then dictates a series of possible future works.

- Appendix A contains the list of publications in which some of the ideas and/or significant results from this thesis have been published (or accepted) in refereed journals and international conferences.

2

ENTROPY AND RELATED METRICS

2.1 INTRODUCTION

Entropy refers to the uncertainty of a random variable, and it depends on the probability distribution of the variable. Entropy is usually used to refer to the discrete random variable and the entropy of a continuous random variable is called differential entropy. The entropy of continuous and discrete variables are different, as their probability distribution is different. The rate of generation of information or the average rate of uncertainty added for each variable is called entropy rate.

A. Lempel and J. Ziv [16] in 1976 first introduced a related measure of entropy rate called Lempel-Ziv complexity, which is also a metric of complexity to evaluate the randomness of a finite sequence. Since then it has been widely employed besides entropy to solve information-theoretic problems [28, 29, 30] and applications such as data compression [31, 32] and coding [33]. For a stationary ergodic process, the entropy rate and Lempel-Ziv complexity (LZC) converge to a common value [3]. Despite the popularity of these measures, the interpretation of LZC and the comparison of theses two measures has not been well addressed in the literature.

In this chapter, we will give basic concepts about entropy and entropy rates used for measuring the regularity or complexity of time series. Besides this, a short discussion on LZC and its estimation will be given. Moreover, we will provide a comparison between the estimates of LZC and entropy rate.

2.2 ENTROPY DEFINITION

Entropy, usually called the Shannon entropy (ShEn) of a discrete random variable X with probability mass function $F(x)$, is defined by

$$ShEn(X) = -\sum_{x \in \mathcal{X}} F(x) \log_b F(x) = -E[\log_b F(x)] \tag{2.1}$$

From the probabilistic point of view, entropy of X is the measure of information carried by $F(x)$, with less entropy corresponding to more information. The entropy of a random variable depends on its probability distribution instead of the actual values taken by the variable. If variable X takes discrete values {x[1], x[2],$\cdots$,x[N]}, then $\mathrm{ShEn}(X) \leqslant \log N$ for any probability mass function of X, and the maximum value, logN is achieved only when the variable has uniform distribution [34].

The unit of the measure of entropy depends on the base (b) of the logarithm (log). If b is 2, entropy is expressed in bits. Entropy is measured in nats, if log of base e is used. In the following discussion, we will use log of base e.

2.2.1 JOINT ENTROPY

The definition of entropy can be extended to a process of multiple variables. The joint entropy of a process of n discrete random variables $(X_1, X_2, \cdots, X_n)$ with joint distribution $F(x_1, x_2, \cdots, x_n)$ can be defined as

$$\mathrm{JEn}(X_1, X_2, \cdots, X_n) = -\sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_n \in \mathcal{X}_n} F(x_1, x_2, \cdots, x_n) \log\{F(x_1, \cdots, x_n)\} \quad (2.2)$$

2.2.2 CONDITIONAL ENTROPY

The conditional entropy of a random variable given another random variable is the expected value of the entropies of their conditional distributions. In information theory, given two random variables X_1 and X_2, the conditional entropy $\mathrm{CEn}(X_2|X_1)$ quantifies the amount of information needed to predict the outcome of X_2 for some known value of X_1. Conditional entropy, $\mathrm{CEn}(X_2|X_1)$ is the result of averaging $\mathrm{ShEn}(X_2|X_1 = x)$ for all possible values of X_1. Mathematically, if X_1 and X_2 are discrete random variables, then $\mathrm{CEn}(X_2|X_1)$ can be computed as [3]

$$\begin{aligned}
\mathrm{CEn}(X_2|X_1) &= \sum_{x_1 \in X_1} F(x_1)\mathrm{ShEn}(X_2|X_1 = x_1) \\
&= -\sum_{x_1 \in X_1} F(x_1) \sum_{x_2 \in X_2} F(x_2|x_1)\log F(x_2|x_1) \\
&= -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} F(x_1, x_2)\log F(x_2|x_1) \\
&= -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} F(x_1, x_2)\log\{F(x_1, x_2)/F(x_1)\} \\
&= -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} F(x_1, x_2)\log F(x, y) - \sum_{x \in X_1} \sum_{x_2 \in X_2} F(x_1, x_2)\log F(x_1) \\
&= -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} F(x_1, x_2)\log F(x_1, x_2) - \sum_{x_1 \in X_1} F(x_1)\log F(x_1) \\
&= \mathrm{JEn}(X_1, X_2) - \mathrm{ShEn}(X_1),
\end{aligned}$$

where $F(x_1, x_2)$ is the joint probability distribution of X_1 and X_2. Thus the relationship between the joint entropy, $JEn(X_1, X_2)$ of a pair of random variables and the conditional entropy of one of them can be explained by the fact that the joint entropy of the pair of random variables is the sum of the conditional entropy ($CEn(X_2|X_1)$) of X_2 given X_1 plus the entropy of X_1 (*i.e*, simply $ShEn(X_1)$. This is called the chain rule of entropy for two discrete random variables. Thus the chain rule of entropy for n discrete random variables, $X_1, X_2, \cdots, X_n$ can be expressed as

$$JEn(X_1, X_2, \cdots, X_n) = \sum_{j=1}^{n} CEn(X_j|X_{j-1}, X_{j-2}, \cdots, X_1) \tag{2.3}$$

2.2.3 DIFFERENTIAL ENTROPY

So far, we discussed about the entropy of discrete random variable(s). The random variables can also take continuous values, called continuous random variables. The entropy of a continuous random variable is called differential entropy. The differential entropy of a continuous random variable X with probability density $f(x)$ is given by

$$DEn(X) = -\int_S f(x) \log f(x) dx, \tag{2.4}$$

where S is the set of all possible values of the random variable. An important property of differential entropy is that among all random variables with the same variance, it acquires the maximum value with a normal (Gaussian) distribution [35]. Thus the differential entropy of Gaussian distribution defines the upper bound of entropy.

2.2.4 DIFFERENTIAL ENTROPY WITH GAUSSIAN PROBABILITY DISTRIBUTION

The differential entropy (DEn) of a continuous random variable, X with Gaussian probability density function, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\pi\sigma^2}$ can be written [using 2.4] as

$$\begin{aligned}
DEn(X) &= -\int_{-\infty}^{\infty} \frac{e^{-(x-\mu)^2/2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} \log \left\{ \frac{e^{-(x-\mu)^2/2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} \right\} dx \\
&= \frac{1}{2}\log(2\pi\sigma^2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\pi\sigma^2} dx + \\
&\frac{1}{2\sigma^2} \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\pi\sigma^2} dx \\
&= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sigma^2 \\
&= \frac{1}{2}\log(2\pi\sigma^2 e),
\end{aligned}$$

where μ and σ^2 are the mean and variance of $f(x)$. Thus, differential entropy depends only on the variance of the distribution.

Let $X_1, X_2, \cdots, X_n$ have a multivariate Gaussian distribution, $\mathcal{N}(\boldsymbol{\mu_n}, \boldsymbol{\Sigma_n})$ with mean vector $\boldsymbol{\mu_n}$ and covariance matrix $\boldsymbol{\Sigma_n}$. Then their joint differential entropy

$$DEn(X_1, X_2, \cdots, X_n) = \frac{1}{2}\log\{(2\pi e)^n|\boldsymbol{\Sigma_n}|\}, \tag{2.5}$$

2.2.5 CONDITIONAL DIFFERENTIAL ENTROPY

Given a continuous random varibale X_1, the conditional differential entropy of another continuous random variable X_2 can be defined as

$$DEn(X_2|X_1) = -\int f(x_1, x_2)\log f(x_2|x_1)\,dx_2\,dx_1,$$

where $f(x_1, x_2)$ is the joint probability density of X_1 and X_2. Using the relationship $f(x_1, x_2) = f(x_1) * f(x_2|x_1)$, the conditional entropy can be alternatively expressed as

$$DEn(X_2|X_1) = DEn(X_1, X_2) - DEn(X_1) \tag{2.6}$$

2.3 DIFFERENCES BETWEEN ENTROPY AND DIFFERENTIAL ENTROPY

The definition of differential entropy is simply an extension to continuous variables of the Shannon entropy for discrete variables. Due to this transformation, there are certainly some differences observed in entropy and differential entropy. A fundamental difference arises from the fact that $F(x)$ used in equation 2.1 is a probability, whereas $f(x)$ appearing in equation 2.4 is the probability density and attains the meaning of probability only when it is integrated over a finite interval. Thus, the use of entropy to mean average uncertainty measure is meaningful, but this claim does not hold for differential entropy [36]. Another important difference between differential entropy and the entropy is that entropy is always positive but differential entropy might have negative values. In spite of some conceptual difficulties, the concept of differential entropy has many potential applications in fields beyond statistical mechanics and communication theory [36] including time series analysis, biomedical signal processing, image processing, econometrics, biostatistics, and population research. If any continuous random variable X is discretized with a quantization step of size Δ, then differential entropy (for continuous variable) and the entroy of discretized form (X_d) holds the following relation

$$DEn(X) + \log\Delta = ShEn(X_d), \tag{2.7}$$

as $\Delta \to 0$

2.4 ENTROPY RATE OF A STATIONARY STOCHASTIC PROCESS

A stochastic process $\mathcal{X}_i$, a time indexed sequence of random variables, is said to be stationary if the joint distribution of any subset of the sequence of random variables

is time invariant. A Gaussian stationary process is defined as a family of random variables $X(n)$ such that

- (i). $X(n)$ is normally distributed with mean and variance being independent on the time parameter
- (ii). the joint probability distribution of $X(n_1, n_2, \cdots, n_k)$ is multivariate normal whose parameters depend only on the differences between two time indices n_i and n_j.

The autocorrelation function of the process may decay slowly or exponentially. The processes with slowly decaying autocorrelation functions are called “long-memory", “long-range correlations", or “strongly dependence memory". An example of this type of stationary process is a fractional Gaussian noise (fGn). On the other hand, the processes with quickly decaying autocorrelation functions are called “short memory", “short-range correlations", or “weakly dependence memory". An example of this is AR process, where the autocorrelation function decays exponentially. AR processes are simple Gaussian stationary processes and have been used for maximum entropy spectral estimation. The more details about the AR processes are given in section 3.2.1.

The entropy of a stationary stochastic process increases (asymptotically) with increasing the number of random variables in the process. The entropy rate of a stationary stochastic process is the increment in entropy for adding a new variable in the process. Mathematically, the entropy rate (*i.e.,* the entropy per symbol) of a stationary stochastic process $\{X_i\}$ is defined in [3] by

$$\mathrm{ShEn}(\mathcal{X}) = \lim_{n\to\infty} \frac{\mathrm{ShEn}(X_1, X_2, \cdots, X_n)}{n} \tag{2.8}$$

If $X_1, X_2, \cdots, X_n$ are independent and identically distributed (*i.i.d.*), then

$$\begin{aligned} \mathrm{ShEn}(\mathcal{X}) &= \lim_{n\to\infty} \frac{\mathrm{ShEn}(X_1, X_2, \cdots, X_n)}{n} \\ &= \lim_{n\to\infty} \frac{1}{n} n\mathrm{ShEn}(X_i) \\ &= \mathrm{ShEn}(X_i), \end{aligned} \tag{2.9}$$

for any $1 \leqslant i \leqslant n$. Thus for a stationary stochastic process of *i.i.d.* random variables, the entropy rate *i.e.* the entropy per symbol is constant and is equal to the entropy of any random variable in the process.

Another notion of entropy rate is the conditional entropy of the last random variable given the past ones, which can be expressed as

$$\mathrm{CEn}(\mathcal{X}) = \lim_{n\to\infty} \frac{\mathrm{CEn}(X_n|X_{n-1}, X_{n-2}, \cdots, X_1)}{n} \tag{2.10}$$

For stationary and stochastic processes the limits exist and two quantities in 2.9 and 2.10 are equal [3].

The differential entropy rate of a stochastic process $\{\mathcal{X}_i\}$, where X_i $(1 \leqslant i \leqslant n)$ are continuous variables,

$$dEn(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} ShEn(X_1, X_2, \cdots, X_n) = \lim_{n\to\infty} ShEn(X_n|X_{n-1}, X_{n-2}, \cdots, X_1),$$

when the limit exists. This expression can be simplified exploiting the chain rule of entropy by

$$dEn(\mathcal{X}) = \lim_{n\to\infty} ShEn(X_n, X_{n-1}, \cdots, X_1) - \lim_{n\to\infty} ShEn(X_{n-1}, X_{n-2}, \cdots, X_1) \quad (2.11)$$

In practice, often the entropy rate is estimated by dropping the limit in equation 2.11

2.4.1 ENTROPY RATES FOR TIME SERIES ANALYSIS

In 1957 Kolmogorov led a seminar on dynamical systems, where Yakub Sinai, Alexeev, Arnold, and some other people attended [37]. From the ideas discussed in that seminar, Kolmogorov proposed the entropy notion to distinguish probabilistic dynamical system and deterministic dynamical system. Kolmogorov defined the entropy only for quasi-regular dynamical system and is called Kolomogorov entropy. Later, Sinai thought about generalization of the Kolmogorov entropy which can be applied to all dynamical system, and is known as measure-theoretic entropy or Kolomogorov-Sinai entropy (KSEn).

Consider a discrete time dynamical system (X, Ω, T, μ) with the state space Ω. Suppose the system is equipped with a σ-algebra (the collection of events to which probabilities can be assigned [38]) and a probability measure μ is defined on it. In general ergodic theory, dynamics is given by a measurable transformation T of Ω onto itself preserving the measure μ. If the state space Ω is partitioned into $\mathcal{A} = (\Phi_1, \Phi_2, \cdots, \Phi_p)$, then it generates a stationary random process of probability theory [37] with values 1, 2,$\cdots$,p if $\Phi_k(x) = i$,for $x \in T^{-k}\Phi_i$, -$\infty < k < \infty$. The entropy of the partition $\mathcal{A}$ is simply the Shannon entropy defined as [39, 40]

$$ShEn(\mathcal{A}) = -\sum_{i=1}^{k} \mu(\Phi_i) \log \mu(\Phi_i),$$

where $\mu(\Phi_i)$ is the probability that the system state resides in partition Φ_i. To compute entropy of a dynamical system, it is required to consider its dynamics $T : \Omega \to \Omega$ with respect to the partition, $\mathcal{A}$. Hence the entropy (*i.e.* actually the entropy rate) of the dynamical system is given by:

$$h(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} ShEn(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}) \quad (2.12)$$

This is called measure-theoretic entropy or dynamical entropy [41].

In other words, it is the limit of the Shannon entropy of the product of partitions with increasing dynamical refinement. The Kolomogorov-Sinai entropy (KSEn) or met-

ric entropy [42] is the upper bound of the dynamical entropy and is defined as the supremum of the dynamical entropy over all partitions. Thus the KSEn of a dynamical system (X, Ω, T, μ) is

$$KSEn(T) = \sup_{\mathcal{A}} h(T, \mathcal{A}), \tag{2.13}$$

where *sup* is taken over all finite partitions. The KSEn is generally inappropriate for any statistical applications. Most obviously, it is usually infinite for correlated stochastic processes, rendering it useless as a mean of discriminating general data sets [43].

2.4.1.1 K_2 ENTROPY

In 1983, Grassberger and Procaccia [8] proposed a new estimate, K_2 that estimates KSEn directly from the finite length real time series. Let us illustrate the computation of K_2 entropy now. Let $x(1), x(2), \cdots, x(N)$ constitute a time series of N points. Now a sequence of vectors, $u_m[i] = [x(i), x(i+1), \cdots, x(i+m-1) | 1 \leqslant i \leqslant N-m+1]$ with embedding dimension m in $\mathbb{R}^m$ is constructed from the time series. Let $n_j^m(r)$ denotes the number of vectors $u_m[j]$ that are close to $u_m[i]$. Here closeness of two vectors means that the Euclidean distance between the two vectors is within a tolerance r of mismatch. Now, the probability of closeness of any vector $u_m[i]$ to the vector $u_m[j]$ is given by $C_j^m(r) = \frac{n_j^m}{N-m+1}$ The probability, $C^m(r)$ that any two vectors are within maximum distance r of each other can be obtained by averaging $C_j^m(r)$ for $1 \leqslant j \leqslant N-m+1$. Thus $C^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} C_j^m(r)$. Then K_2 entropy is defined as

$$K_2 = \lim_{m \to \infty} \lim_{r \to 0} \lim_{N \to \infty} \log \left\{ C^m(r) - C^{m+1}(r) \right\}, \tag{2.14}$$

2.4.1.2 ECKMANN-RUELLE ENTROPY

Following the same technique, Eckmann and Ruelle suggested calculating the KSEn by considering the distance between two vectors as the maximum absolute difference of their corresponding elements, *i.e.*, if $u_m[i]$ and $u_m[j]$ are two vectors of length m, then the distance between them is $d(u_m[i], u_m[j]) = max\{|(x(i+k) - x(j+k)|\}$. The estimation of Eckmann and Ruelle entropy (EREn) is illustrated by the following steps:

1. Form templates $u_m[j] = [x(j), \cdots, x(j+m-1)]$ of size m, for $1 \leqslant j \leqslant N - m + 1$;
2. Define the distance between $u_m[j]$ and $u_m[i]$: $d(u_m[j], u_m[i]) = max_{0 \leqslant k \leqslant m-1} |u_m[j+k] - u_m[i+k]|$;
3. Let A_j^m be the number of templates $u_m[i]$ such that $d(u_m[j], u_m[i]) \leqslant r$, where $1 \leqslant i \leqslant N - m + 1$, and $C_j^m(r) = A_j^m/(N-m+1)$;
4. Define $\Phi^m(r) = (N-m+1)^{-1} \sum_{j=1}^{N-m+1} \log C_j^m(r)$;

5. Increase m by 1 and repeat steps 1 to 4;
6. Finally, $EREn = \lim_{m\to\infty} \lim_{r\to 0} \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]$

It is noted that $\Phi^{m+1}(r) - \Phi^m(r)$ represents the average of the natural logarithm of the conditional probability that sequences that are close for m successive data points will remain close also for the next incremental point. Although this measure has been found useful in discriminating low dimensional chaotic systems, but it cannot be applied to experimental data since the estimate is infinity for a process with superimposed noise of any magnitude [44, 45].

2.4.1.3 APPROXIMATE ENTROPY

Pincus [46] introduced a similar way of estimating EREn called approximate entropy, for getting finite estimates with real and noisy experimental data. ApEn measures the likelihood that runs of patterns that are close remain close at the next incremental comparison. Pincus proposed fixed values for parameters m and r in the definition of EREn. So, the definition of ApEn becomes

$$ApEn(m,r) = \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)] \tag{2.15}$$

Estimated values of approximate entropy, ApEn(m,r,N) is obtained by dropping the limit in equation 2.15. ApEn(m,r,N) approximates EREn for very large values of N, m and for very smaller values of r. The novelty of ApEn(m,r) is that they can distinguish a wide variety of system, and that the estimation of ApEn(m,r,N) is possible even with small m and finite short series length N. It can potentially distinguish low dimensional deterministic systems, high dimensional chaotic systems, periodic and multiply periodic systems, stochastic and mixed systems [46]. It has small standard deviation. However, it has some limitations also. The matching of a template with itself *i.e.* selfmatching has been considered to make it define even with very short series. This consideration of selfmatching makes the measure biased [43] for finite short series. Besides this, ApEn for short series is uniformly lower than expected [47]. Another important shortcoming of ApEn is more prone to practical inconsistency. Pincus [48] considered the problem of assessing if any stochastic process A was more regular then process B, by means of computing ApEn. He defined "consistent" those processes for which ApEn of A was always larger (or smaller) than ApEn of B, for any value of the parameters m and r. Here, we defined "practical consistency" by the fact that ApEn of series S_A is larger than ApEn of series S_B for a broad range of m and r values.

2.4.1.4 SAMPLE ENTROPY

To address some manifest limitations of ApEn (Pincus himself [43] reported ApEn to be a biased statistic), Richman and Moorman [47] introduced sample entropy (SampEn), which is similar but improved version of ApEn. The estimation of SampEn is performed by defining the following steps:

1. Let A_j^m be the number of templates $u_m[i]$ such that $d(u_m[j], u_m[j]) \leqslant r$, where $1 \leqslant i \neq j \leqslant N - m$, and $C_j^m(r) = A_j^m/(N - m - 1)$;

2. Let A_j^{m+1} be the number of templates $u_{m+1}[i]$ such that $d(u_{m+1}[j], u_{m+1}[i]) \leqslant r$, where $1 \leqslant i \neq j \leqslant N - m$, and $C_j^{m+1}(r) = A_j^{m+1}/(N - m - 1)$;

3. Define $A_m(r) = \sum_{j=1}^{N-m} C_j^m(r)/(N - m)$ and $A_{m+1}(r) = \sum_{j=1}^{N-m} C_j^{m+1}(r)/(N - m)$, then

4. $\text{SampEn}(m, r) = \lim_{N\to\infty}[\log A^m(r) - \log A^{m+1}(r)]$ and,
for a finite series, $\text{SampEn}(m, r, N) = \log A^m(r) - \log A^{m+1}(r)$.

SampEn does not consider self matches of templates like ApEn, which makes ApEn a biased estimate. Besides this, SampEn is less prone to practical inconsistency, as it requires less lengthy series to converge to the final value.

2.4.1.5 CORRECTED CONDITIONAL ENTROPY

All estimations of entropy rate (ApEn, SampEn, and CEn) are dependent on the selection of two parameters: m (the length of templates) and r of mismatch between corresponding elements of the templates. In CEn estimation, it may happen that there is only a unique pattern matching of lengths m and $m+1$. i.e. the unique appearance of the pattern of length $m+1$ can be completely predicted by the pattern of length m. As a consequence, the conditional probability of 1 leads CEn estimation to zero. Even with completely random series, the estimated CEn value decreases quickly with increasing the pattern length. But, in some cases, it may necessary to select a large enough m, for recognition of deterministic patterns [49]. This limitation has been overcome by adding a corrective term with the CEn, and is referred to as corrected conditional entropy [49]. as:

$$\text{CcEn}(m+1) = \text{CEn}(m+1|m) + \text{Perc}(m+1) * \text{ShEn}(1), \tag{2.16}$$

where $\text{Perc}(m+1)$ is the percentage of single (unique) points in the $m+1$ dimensional phase space and ShEn(1) is the Shannon Entropy for templates of length 1. The CcEn of a process decreased with increasing its regularity as other entropy rate. In addition, when ApEn underestimates the entropy of finite short length series, the CcEn estimates is still high.

2.4.1.6 TRANSFER ENTROPY

The transfer entropy (TEn), an another information theoretic quantity, provides a slightly different definition of statistical dependency using conditional probability to define what it means for one random process to provide information about another [36]. Transfer Entropy measures the amount of information transferred from one process to another. This measure was initially proposed to quantify information transport in dynamical systems [50] and was later extended to continuous random variables [51]. Mutual information (MI) [52] is often used to quantify the statistical dependence between signals. But, it cannot be applied to determine the predominant direction of information flow. The TEn of two discrete processes $\{X\}_i$ and $\{Y\}_j$ has been defined

in [51] as: suppose state x_{i+1} of $\{X\}_i$ depends on the m past states X_i^m, but do not depends on the l past states Y_j^l of $\{Y\}_j$. Then it holds the generalized Markov property

$$F(x_{i+1}|X_i^m, Y_j^l) = F(x_{i+1}|X_i^m).$$

If there exists any such dependence, then the TEn can quantify this, which is obtained [51] by

$$TEn(X_{i+1}|X_i^m, Y_j^l) = \sum F(X_{i+1}|X_i^m, Y_j^l) \log \left\{ \frac{F(X_{i+1}|X_i^m, Y_j^l)}{F(X_{i+1}|X_i^m)} \right\}, \tag{2.17}$$

where $F(X_{i+1}|X_i^m, Y_j^l)$ and $F(X_{i+1}|X_i^m)$ are considered as the underlying transition probability (the probability associated with various state changes of of a process) and a priori transition probability, respectively. The TEn can be expressed as the difference of conditional Shannon entropies as

$$TEn(X_{i+1}|X_i^m, Y_j^l) = ShEn(X_{i+1}|X_i^m) - ShEn(X_{i+1}|X_i^m, Y_j^l)$$

Similary, the transfer entropy (tEn) of two continuous processes $\{\mathcal{X}_i\}$ and $\{\mathcal{Y}_i\}$, can be written as

$$tEn(X_{i+1}|X_i^m, Y_j^l) = \int\int\int f(X_{i+1}|X_i^m, Y_j^l) \log \left\{ \frac{f(X_{i+1}|X_i^m, Y_j^l)}{f(X_{i+1}|X_i^m)} \right\} dx_{i+1}\, dX_i^m\, dY_j^l, \tag{2.18}$$

where f is the joint probability density function of two continuous processes.

2.4.1.7 RENYI ENTROPY

Shannon's original work [53] has been extended in many alternative measures of entropy. For instance, Renyi [6] extended the Shannon entropy (ShEn) to a family of measures that follows

$$ReEn_q(X) = -\frac{1}{q-1} \log \sum_{i=1}^{n} F(x)^q \tag{2.19}$$

As order, $q \to 1$ in equation 2.19, the Renyi entropy (ReEn) tends to ShEn.

The ReEn of order q for a continuous random varible X is

$$reEn_q(X) = -\frac{1}{q-1} \log \left\{ \int_{-\infty}^{\infty} f(x)^q \right\} \tag{2.20}$$

Letting $q \to 1$ and applying L'Hospitals rule, 2.20 results in differential entropy, i.e., $deEn(X) = reEn_1(X)$. ApEn and SampEn are respectively the differential Renyi entropy rates of order 1 and 2 [35].

2.5 LEMPEL-ZIV COMPLEXITY

Lempel-Ziv complexity measures the rate of generation of new patterns along a sequence of symbols. There are some variations of this metric. We will discuss here the original version (LZ76) and the modified version of it (LZ78) introduced in [3]. The concept of both techniques are same. The only difference is in parsing. For a symbolic sequence, $S = \{s_1, s_2, \cdots, s_N\}$ of length N, such that $s_i \in \mathcal{A}$ (the alphabet of α symbols), parsing refers to the procedure of partitioning S into a set of nonoverlapping substring. The yields of the parsing procedure are called phrases (PhrS). In both techniques of Lempel-Ziv complexity estimation, the symbolic sequence is partitioned into a possible set of distinct phrases, which define the LZC of the symbolic sequence.

2.5.1 SYMBOLIC REPRESENTATION

The complexity analysis of a time series is based on coarse-graining of the measurements, i.e. the time series is transformed into a sequence of symbols, called symbolic sequence. In doing so, some amount of detailed information is lost, but some of the invariant, robust properties of the dynamics may be kept [54]. Using a larger number of symbols may be better, since it can keep more information than two symbols [55]. There are dozens of techniques for producing different variants of the symbolic representation [56, 57] of the time series. Some symbolic representations reduce the dimensionality, and is the major concerning issue of data storage, while using some other representations, the intrinsic dimensionality of the symbolic sequence is the same as that of the original. However, this is not a concerning issue for complexity analysis. Instead, we are interested about probability distribution. the symbolic sequence should closely resembles the time series. Here, we will explain two methods for symbolic transformation (quantization) :(i) quantization with source distribution, (ii) quantization with equiprobable distribution.

- “Quantization with source distribution" In this quantization method, the range of the distribution (*i.e.* the difference between maximum and minimum of the series) is partitioned into a fixed number of bins. Each bin is labeled by a distinct symbol (or letter of the alphabet). All values of the series fall within a specific bin, is represented by the symbol assigning to the bin. In this way, a symbolic sequence of same length of the series is obtained. Thus, if we use 2 levels of quantization, then values less than a threshold (mean or median) is represented by the symbol ’0’ and values greater than the threshold is represented by ’1’. A time series and its symbolic representation is depicted in figure 2.1. This method of quantization is affected by the presence of artifacts (being the maximum and minimum value of the series are dependent on the artifacts in the series).

- “Quantization with equiprobability" This quantization technique will produce a symbolic sequence with equiprobability from the given time series. Given a nor-

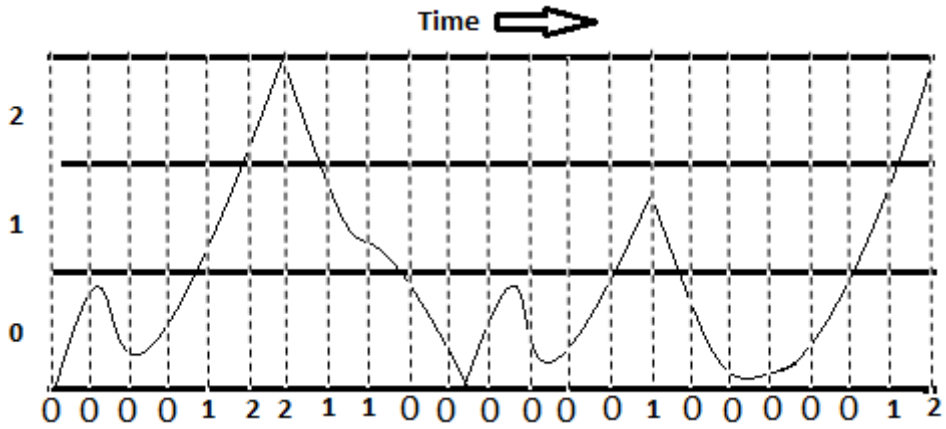


Figure 2.1: A discrete time series and its symbolic representation. The vertical dashed lines denote the value of the signal at time and the discrete series has been represented by 3 symbols.

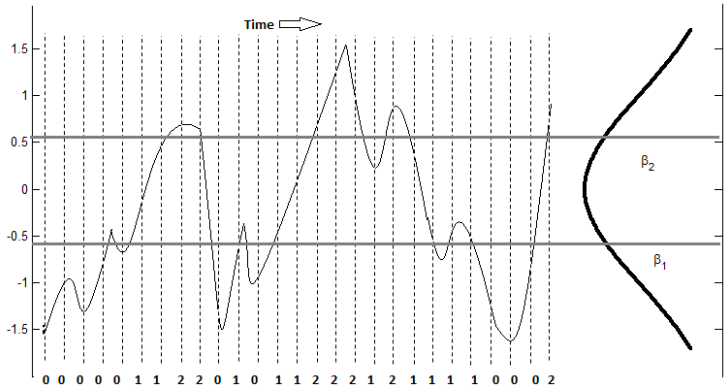


Figure 2.2: A time series and its discrete values denoted by the vertical dashed lines. The discrete time series is then mapped to the symbolic dynamics using equiprobablie quantization technique. The two breakpoints β_1 and β_2 define the 33th and 66th percentile values.

malized time series, we can simply define the breakpoints[1] that will produce a equal sized areas under the curve [58]. Once the breakpoints have been determined, we can symbolize the time series as follows: all values below the smallest breakpoints are mapped to a symbol (suppose 'a'). Then, the values smaller than the second smallest breakpoints but greater than or at least equal the smallest breakpoint are mapped to another symbol (suppose 'b'), and continues. This method is illustrated in figure 2.2.

2.5.2 LZ76 PARSING METHOD

Suppose S denotes a finite length sequence of N symbols. The procedure of partitioning S into a set of non-overlapping distinct substrings are referred to as 'parsing'. A

1 Breakpoints [57] are a sorted list of numbers $B = \beta_1, \beta_2, \cdots, \beta_{b-1}$ such that the area under a curve $\mathcal{N}(0,1)$ from β_j to $\beta_{j+1} = 1/b$, where β_0 and β_b are defined as -∞ and ∞, respectively.

phrase $PS(i,j)$ staring at position i and ending at j is a substring $S(i,j)$, for $i \leqslant j$ of S. The parsing of the phrases involves a left to right scan of the given sequence. A substring $S(i,j)$ is compared to all substrings of S(1,j-1). If $S(i,j)$ is matched with any substring of S constructed from the S(1,j-1), then $S(i,j)$ is updated to $S(i,j+1)$. The matching check and updating are repeated until the no matching of $S(i,j)$ is found. When no match of $S(i,j)$ with any substring of the sequence of S(1,j-1) is found, $S(i,j)$ is defined as a new phrase. Now $S(i,j)$ is updated to $S(j+1,j+1)$ to point to the next $(j+1)^{th}$ symbol and the procedure is repeated for another new phrase. The procedure begins with the first symbol $S(1,1)$ and continues until the end of the sequence. The number of phrases is denoted by $c(S_N)$. It should be mentioned that $c(S_N)$ is updated by 1, even if the last substring is not a new phrase. Consider a sequence of 3 symbols, S=0101202110200120102221001122 01. The parsing of this sequence using LZ76 will generate the following phrases (the phrases are separated by a period):
{0.1.012.02.11.020.0120.1022.210.011.220.1}.
Thus, $c(S_{30})$=12.

2.5.3 LZ78 PARSING METHOD

This parsing procedure is particularly simple and has become popular as an standard algorithm because of its speed and efficiency [3]. In this scheme, the lesftmost symbol S(1) of the sequence is cosidered as the first phrase, and is stored in a vocabulary, VOCAB (PhrS). Then we look immediately for the next shortest substring S(2,j), where $j \leqslant N$, that is not appeared in the vocabulary, and store it in the VOCAB. This searching and storing process will be repeated until all symbols of the sequence are precessed. Thus with the for the same symbolic sequence used to explain LZ76 parsing, LZ78 returns the following phrases:
Symbols:0101202110200120102221001122 01
Phrases:0.1.01.2.02.11.020.012.010.22.21.00.112.20.1
The total number of phrases generated using LZ78, $c(S_{30})$= 15.

2.5.4 LEMPEL-ZIV COMPLEXITY ESTIMATE USING LZ78

In LZ78 parsing, all symbols except the last one of the current phrase, might have appeared earlier. So, each phrase can be broken up into a reference to a previous phrase and a letter of the alphabet. The procedure is explained in table 2.1. In the code words, the symbol ϕ is used to refer to the empty prefix of the phrase. The location of the prefix to the phrase can be referred by at most $\log_2(c_N)$ bits, and $\lceil \log_2(Q) \rceil$ bits to

Position:	1	2	3	4	5	6	7	8	9	10	11	12
Phrases:	0	1	01	2	02	11	020	012	010	22	21	00
Code Words:	ϕ,0	ϕ,1	1,1	ϕ,2	1,2	2,1	5,0	3,2	3,0	4,2	4,1	1,0

Table 2.1: Coding of the sequence {0101202110200120102221001122 01} using LZ78

code the last symbol, where Q is the number of symbols. Therefore, the total length (in number of bits) required to encode the sequence S is

$$\mathrm{LZC78}(S) = c_N \left\{ \log_2(c_N) + \lceil \log_2 Q \rceil \right\} \tag{2.21}$$

Thus, the rate of new patterns appearing in the sequence is given by

$$\mathrm{LZC78}_{norm}(S) = \frac{c_N (\log_2(c_N) + \lceil \log_2 Q \rceil)}{N \lceil \log_2 Q \rceil} \tag{2.22}$$

On the other hand, if the complexity is represented in symbols, instead of bits, then equation 2.21 can be written as

$$\mathrm{LZC78}(S) = c_N \left\{ \log_Q(c_N) + 1 \right\}, \tag{2.23}$$

and normalized LZC78 is:

$$\mathrm{LZC78}_{norm}(S) = \frac{c_N}{N} \left\{ \log_Q(c_N) + 1 \right\}, \tag{2.24}$$

where $\log_Q(c_N)$ symbols are used to encode the prefix of a phrase and 1 for last symbol for the sequences of Q symbols.

2.6 SUMMARY

The main objective of this chapter was to provide a review on entropy, entropy rate, and Lempel-ziv complexity. The estimates of entropy, entropy rate and LZC has been explained in detail.

Entropy and entropy rate are very common measures in physics and information theory. It gains its popularity since Shannon introduced it. The definition given by Shannon has been developed with times in order to extend its applicability with reducing limitations. Due to its vast usefulness, the further improvement of the estimates is still an on going research yet today.

The estimates of LZC using different parsing methods as well as different quantization techniques have been explained. Different parsing techniques provides different number of phrases, and hence also the LZC for a given sequence. The comparison of different parsing methods and the effects of quantization methods in the estimation of LZC are explained in detail in the next chapter.

3

ENTROPY PARAMETRIC ESTIMATION

3.1 INTRODUCTION

Entropy and entropy rates have been used as powerful tools in different applications of signal processing [59, 60, 61, 62, 63, 64]. Some most common measures of entropy used for measuring the regularity (or complexity) of time series, and also for analyzing physiological signals include ApEn, SampEn, and CcEn which are measures of differential entropy rate. Entropy rates are becoming more and more interesting due to the possibility they offer to distinguish between a periodic repetition of the same patterns and aperiodic dynamics [49]. However, their estimations are highly dependent on series length. Hence the estimates of entropy may be far away from what expected on very long series. On the other hand, some metrics such as sample entropy may be quietly undefined or has large variance in their estimates for very short series. Unfortunately, real applications require analysis of very short series. A related problem in spectral analysis has been solved using parametric approach. Parametric estimation means that any given signal $y[n]$ is generated by a known mathematical model and the estimation is made as a function of both model parameters and inputs. The research for parametric estimates of entropy and related metrics have the purpose of designing and implementing parametric methods for different entropy measures to overcome the limitations in the usability of their traditional measures for very short series length. With short series, stronger assumptions (using a Gaussian AR model) can be made which reduces the risk of high variance in estimates. However, this may also introduce some bias in the events when the assumptions are partially violated. This violation of assumption can be predicted by testing the disagreement between the numerical estimates of entropy and the acceptable range of values obtained through several realizations of the model. The agreement (or disagreement) between numerical and the parametric estimates of entropy might provide some more information about the signal behavior such as the presence of nonlinearity, nonstationarity or non-Gaussianity in the series.

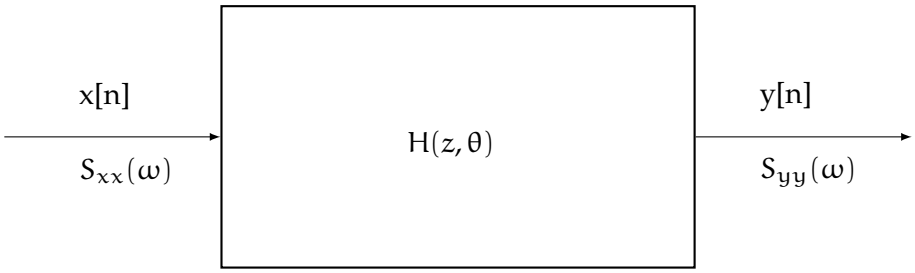


Figure 3.1: The general schematic diagram of parametric approach PSD estimation. The power spectrum $S_{yy}(\omega)$ of a signal $y[n]$ is determined in terms of the power spectrum $S_{xx}(\omega)$ of the input $x[n]$ (i.e. fed into the model) and the function of model parameters. That is why, it is called parametric approach estimation.

A related measure of entropy rate *i.e.* Lempel-Ziv complexity also to be a powerful complexity measure and its variants have been demonstrated. However, it is still not clear how many samples are required for proper estimates of these measures. Although LZC and entropy rate are used in same application domains. The relationship between LZC and entropy rate with respect to the convergence issue is still not well addressed.

In this chapter, an overview of parametric approach will be presented first. Then, we will explain derivation of analytical expressions for some most commonly used measures of entropy rates. After that, the feasibility of parametric estimation of entropy will be studied based on synthetic series. Then, we will demonstrate the number of necessary samples for reliable estimations of LZC through the Gaussian stationary processes. The convergence of LZC and entropy rate with respect to the series length (N) will be compared. Besides this, a relative correlation will be verified on LZC and SampEn using synthetic series generated through the autoregressive (AR) models. In addition, the dependence of LZC estimation on probability distribution will be verified.

3.2 PARAMETRIC ESTIMATIONS

Parametric approaches are based on the use of models, assuming that the data are generated in a certain way and the estimation is made as a function of both model parameters and inputs. *e.g.*, the estimation of power spectral density (PSD) of a signal through parametric approach is explained in figure 3.1.

The first step of a parametric approach is to select the most appropriate family of models. There are many models for parametric approach, but the most common choices are a family of linear time-invariant models, whose transfer function is defined by a set of M parameters $\Theta = [\Theta_1, \Theta_2, \cdots, \Theta_M]$. The set of parameters defines the property of transfer function and characterizes the signal generated by the model.

The most commonly used models for a given signal are autoregressive moving average (ARMA), AR, and moving average (MA). Identification of AR models has been

largely explored in the literature and it requires solving linear (simpler) equations than those required for MA or ARMA models. Also AR models are maximum-entropy models (among all sharing the same autocorrelation function). Here for simplicity, we will describe the parametric entropy measures through AR models.

We will first discuss the conditions under which a parametric estimation of entropy is possible. We will limit our attention to the estimations of ApEn, SampEn, and CcEn with linear AR models. Pincus [43] and then Lake [35] already tackled the problem of deriving analytical formulas of ApEn and SampEn for an AR process. Following the suggestion in [43], our objectives are to extend the analytical expression of ApEn to any m value, and also to derive an analytical expression for SampEn. Then to compare the numerical estimations, the estimations obtained through simulated series, and the theoretical ones.

3.2.1 AR PROCESS

An AR process of order M can be expressed as

$$x[n] = -\sum_{i=1}^{M} a_i x[n-i] + w[n]$$

where a_i are real coefficients and $w(n)$ is a white Gaussian noise (WGN) with mean zero and variance σ_w^2. An AR model of order M is a wide-sense stationary process, if the roots of the polynomial $z^M - \sum_{i=1}^{M} a_i z^{M-i}$ lie within the unit circle, *i.e.* each root must satisfy $|z_i| < 1$.

The parameters of the model and the autocovariance function values γ_k, are linked by the Yule-Walker's equations

$$\begin{pmatrix} 1 & a_1 & a_2 & \cdots & a_M \\ a_1 & 1+a_2 & a_3 & \cdots & 0 \\ a_2 & a_1+a_3 & 1+a_4 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_M & a_{M-1} & a_{M-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \cdots \\ \gamma_M \end{pmatrix} = \begin{pmatrix} \sigma_w^2 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}. \tag{3.1}$$

The m consecutive values, $X_m[n] = \{x[n], \cdots, x[n+m-1]\}$, are multivariate normal on $\mathbb{R}^m$, with Normal joint probability density
$f(X_m) = \mathcal{N}(0, \Sigma_m) = e^{(-X_m^T \Sigma_m^{-1} X_m/2)/[(2\pi)^m \det(\Sigma_m)]^{1/2}}$ and Toeplitz covariance matrix

$$\Sigma_m = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{pmatrix}.$$

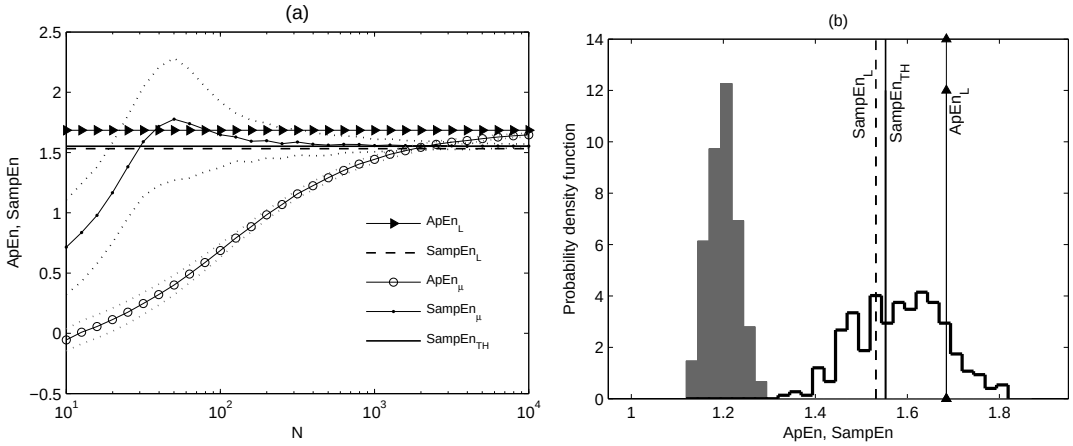


Figure 3.2: ApEn and SampEn of the arbitrary AR model of coefficients $[1, -0.87, 0.02]$, with $m=2$ and $r = 0.2 \times \text{STD}$. Panel (a): Entropies of the model as a function of N. ApEn_μ and SampEn_μ were estimated by taking the average of $K = 10000$ and $K = 300$ Monte Carlo's runs for $N \leqslant 100$ and $N > 100$, respectively. The dotted lines define the boundary of $\text{SampEn}_\mu \pm \text{SampEn}_\sigma$ and $\text{ApEn}_\mu \pm \text{ApEn}_\sigma$. Panel (b): Probability density functions derived from $K = 300$ realizations of ApEn (left) and SampEn (right) for $N = 360$. ApEn_μ does not match ApEn_L yet, as N is too small. Instead, SampEn_L=1.532, SampEn_{TH}=1.553 and SampEn_μ=1.584 approximately coincide. On the other hand, $\text{SampEn}_\sigma = 0.096$ is larger than $\text{ApEn}_\sigma = 0.033$. Lake (2002) derived an expression for estimating SampEn_σ, but in this case it underestimates it (0.016).

The values γ_m, for $m \leqslant M$, are defined by equation (3.1). When $m > M$, further elements in Σ_m are still dictated by the Yule-Walker's equation $\gamma_k = -\sum_{i=1}^{M} a_i \gamma_{k-i}$.

Denoting $\rho_k = \gamma_k/\gamma_0$ the autocorrelation coefficient, the variance $\sigma_y^2 = \gamma_0$ of the series generated by the AR process is

$$\sigma_y^2 = \sigma_w^2 (1 + a_1\rho_1 + \cdots + a_M\rho_M)^{-1} = \sigma_w^2 c, \tag{3.2}$$

where $c = (1 + a_1\rho_1 + \cdots + a_M\rho_M)^{-1}$.

3.2.2 ASYMPTOTIC THEORETICAL VALUES FOR ENTROPY OF A GAUSSIAN AR PROCESS

The analytical expression of $\text{ApEn}(m = 1, r)$ for a stochastic (thus also for an AR) process is given by Pincus in [43]. Let

$$Q_m = \int_{x[m]-r}^{x[m]+r} \cdots \int_{x[1]-r}^{x[1]+r} f(\Xi_m) d\xi_1 \cdots d\xi_m$$

be the probability that the values X_m lie within the hypercube of side $2r$, where $f(X_m)$ is the multivariate probability density of the ergodic stochastic process. Then

$$\text{ApEn}_{TH}(1, r) = \iint_{\mathbb{R}^2} f(X_2) \log\left(\frac{Q_1}{Q_2}\right) dx[1]dx[2].$$

This equation can be extended to derive a general analytical expression of $\mathrm{ApEn}(m, r)$ for any m as

$$\mathrm{ApEn}_{\mathrm{TH}}(m, r) = \int_{\mathbb{R}^{m+1}} \cdots \int f(X_{m+1}) \log\left(\frac{Q_m}{Q_{m+1}}\right) dX_{m+1}. \tag{3.3}$$

where $dX_m = dx[1]dx[2]\cdots dx[m]$,

Following a similar approach, a theoretical value for SampEn of an AR process can be derived from the definition. In fact, the probability of matching two templates of size m within error tolerance r (i.e. the maximum absolute difference between the corresponding elements of any two templates is r) is given by:

$$P_m = \int_{x[m]-r}^{x[m]+r} \cdots \int_{x[1]-r}^{x[1]+r} \frac{e^{-\Xi_m^T \Sigma_m^{-1} \Xi_m}}{(2\pi)^{m/2} \det(2\Sigma_m)^{1/2}} d\xi_1 \cdots d\xi_m.$$

In fact, the difference $X_m[i] - X_m[j]$ is distributed as $\mathcal{N}(0, 2\Sigma_m)$. Hence, the theoretical value of SampEn of an AR model is

$$\mathrm{SampEn}_{\mathrm{TH}}(m, r) = \log(P_m) - \log(P_{m+1}). \tag{3.4}$$

It should be noted that the asymptotic theoretical values in equations (3.3) and (3.4) depend on r and m and are valid in the limit $N \to \infty$. However, when computing P_{m+1} for $m \geqslant M$, the Yule-Walker's equations allow the factorization of the covariance matrix into

$$\Sigma_{m+1} = T \begin{pmatrix} \Sigma_m & \mathbf{0} \\ \mathbf{0} & \sigma_w^2 \end{pmatrix} T', \tag{3.5}$$

with the change of variable $\hat{x}[n+m] = x[n+m] + \sum_{i=1}^{M} a_i x[n+m-i]$, where T is a transformation matrix. Hence, the value of $\mathrm{SampEn}_{\mathrm{TH}}(m, r)$ stabilizes for $m \geqslant M$.

Similarly, we can write an analytical expression for conditional entropy (*i.e.* the entropy rate) of a Gaussian AR process. The conditional entropy (CEn) is the entropy of a conditional distribution of the present m^{th} observation, given the previous $m-1$ observations. For a Gaussian stochastic process (thus for an AR process), the conditional entropy and differential entropy rate should represent the same value. Thus from equation 2.11, we can write the conditional entropy

$$\mathrm{CEn}(X) = \lim_{m\to\infty} \{\mathrm{dEn}(\mathbf{X}_m) - \mathrm{dEn}(\mathbf{X}_{m-1})\}, \tag{3.6}$$

If $X_1, X_2, \cdots, X_m$ is a stationary random sequence, having m^{th} order normal probability density function, then their joint differential entropy from equation 2.5 is,

$$\mathrm{DEn}(X_1, X_2, \cdots, X_m) = \frac{1}{2}\log(2\pi e)^m |\Sigma_m|, \tag{3.7}$$

where Σ_m is the Toeplitz covariance matrix of order m, and $|.|$ denotes the determinant of it. Thus, the expression for CEn of a Gaussian AR process of order M becomes

$$\begin{aligned} CEn(X) &= \lim_{m\to\infty}\{DEn(X_1, X_2, \cdots, X_m) - DEn(X_1, X_2, \cdots, X_{m-1})\} \\ &= \lim_{m\to\infty}\left\{\frac{1}{2}\log(2\pi e)^m|\Sigma_m| - \frac{1}{2}\log(2\pi e)^{m-1}|\Sigma_{m-1}|\right\} \\ &= \frac{1}{2}\left\{\log(2\pi e) + \lim_{m\to\infty}\log\left(\frac{|\Sigma_m|}{|\Sigma_{m-1}|}\right)\right\}, \end{aligned} \tag{3.8}$$

for $m > M$, further elements are dictated by the Yule Walker's equation 3.1. The estimated value of the conditional entropy rate is obtained by discarding the limit in the definition, and thus using factorization of the covariance matrix from equation 3.5, the expression for CEn(X) reduces to

$$CEn(X) = \log\sigma_w + \frac{1}{2}\log(2\pi e) \tag{3.9}$$

This is the theoretical value for continuous random variables. But, in practice, the computation is based on discrete series. So, the expected value of conditional entropy of a Gaussian AR process, from equations 3.9 and 2.7 is

$$CEn(X) = \log\sigma_w + \frac{1}{2}\log(2\pi e) - \log\Delta, \tag{3.10}$$

The quantization step size, Δ depends on the expected range of the distribution of AR model and number of quantization levels (ξ). Expressing Δ as a function of the expected range and ξ. Finally, the expected theoretical value of conditional entropy of a Gaussian AR process is given by

$$CEn(X) = \log\sigma_w + \frac{1}{2}\log(2\pi e) - \log\left(\frac{expectedRange}{\xi}\right), \tag{3.11}$$

Now let us consider the expression for differential entropy rate of a multivariate Gaussian AR process. By definition, the differential entropy rate, dEn(X) of a Gaussian AR process of m random variables is

$$\begin{aligned} dEn(X) &= \lim_{m\to\infty}\frac{1}{m}dEn(X_m) \\ &= \lim_{m\to\infty}\frac{1}{2}\log(2\pi e)^m|\Sigma_m| \\ &= \frac{1}{2}\log(2\pi e) + \lim_{m\to\infty}\frac{1}{m}\log|\Sigma_m| \end{aligned}$$

For a Gaussian AR process, Σ_m is generated by the power spectral density f, determinant of Σ_m is the Toeplitz determinant [65]. If $f(\lambda)$ denotes the Fourier transform of the covariance function, then as stated in [65], the limiting term in the right hand side

of differential entropy rate can be replaced by the $\frac{1}{2\pi}\int_0^{2\pi}\log\left(f(\lambda)\right)d\lambda$, and hence the differential entropy rate for a Gaussian AR becomes

$$dEn(X) = \frac{1}{2}\left\{\log(2\pi e) + \frac{1}{2\pi}\int_0^{2\pi}\log f(\lambda)d\lambda\right\} \tag{3.12}$$

As the number of variables m in the AR process approaches ∞, the ratio of the determinants of Toeplitz covariance matrices [65] in equation 3.8 approaches to $e^{\frac{1}{2\pi}\int_0^{2\pi}\log f(\lambda)d\lambda}$ *i.e*

$$\frac{|\Sigma_m|}{|\Sigma_{m-1}|} \to e^{\frac{1}{2\pi}\int_0^{2\pi}\log f(\lambda)d\lambda},$$

Putting the values of the ratio in equation 3.8, we see that the CEn of a Gaussian AR process becomes

$$CEn(X) = \frac{1}{2}\left\{\log(2\pi e) + \frac{1}{2\pi}\int_0^{2\pi}\log f(\lambda)d\lambda\right\} \tag{3.13}$$

Thus the differential entropy rate and conditional entropy of a multivariate Gaussian AR process is the same. In estimation of CEn with finite series, the relative patterns in m dimensional space becomes single and hence the estimation is negatively biased. To overcome this problem with short series, Porta et al., [49] added a correction factor with the CEn estimation. This correction factor is solely determined by the percentage of the number of unique patterns of dimension m and the Shannon entropy for m=1. The additional term is not a required for determining the theoretical value of the estimation. Thus, the theoretical value of CEn is also the theoretical value for CcEn.

3.2.3 THEORETICAL VALUES OF ENTROPY FOR $m \to \infty$ AND $N \to \infty$

Pincus [48] showed that ApEn is related to differential entropy rate, a central concept of information theory, and later Lake [35] proved that ApEn and SampEn are the differential Renyi entropy rates of order 1 and 2, respectively.

In practice, Lake [35] derived the theoretical expressions of both ApEn and SampEn, from the definition of differential entropy rate, in the limit $m \to \infty$. If r is chosen independently of the the standard deviation (STD) of the sequence (σ_y), the expressions for ApEn and SampEn (according to Lake's derivation) become

$$\begin{aligned} ApEn_L(r) &= \log(\sigma_w) + \frac{1}{2}\left[\log(2\pi) + 1\right] - \log(2r), \\ SampEn_L(r) &= \log(\sigma_w) + \frac{1}{2}\log(4\pi) - \log(2r). \end{aligned}$$

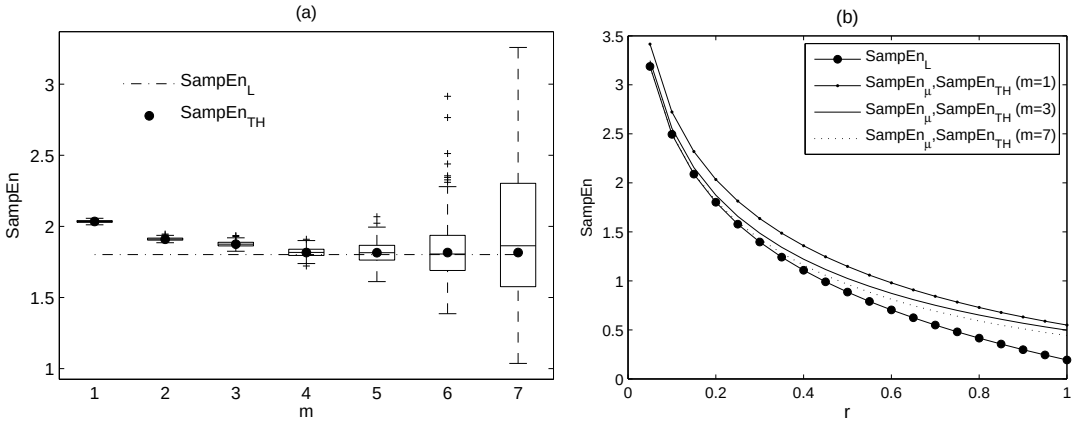


Figure 3.3: SampEn of the arbitrary AR model of coefficients $[1, -0.80, 0.46, 0.02, -0.33]$ for $N = 6000$, r=0.2 $\times$ STD and different values of m. Panel (a): boxplots represent the probability density of SampEn derived for 300 realizations of the model. SampEn_{TH} lies inside the standard range of numerical estimations for every m. On the other hand, SampEn_L is constant due to it's independence on m. Although, both SampEn_{TH} and SampEn_μ differ from SampEn_L for any $m < M = 4$ (the model order), they meet at a common value for any $m \geqslant 4$. Panel (b): SampEn_{TH} approximately overlaps with SampEn_μ for any m, r. They progressively converge to SampEn_L for $m \geqslant 4$.

On the other hand, if r is chosen as a percentage $\hat{r}$ of the STD such that $r = \hat{r} \times \text{STD}$, then the expressions for ApEn_L and SampEn_L become

$$\text{ApEn}_L(\hat{r}) = \log(\sigma_w) + \frac{1}{2}[\log(2\pi) + 1] - \log(2\hat{r}\sigma_y) \tag{3.14}$$

$$= \log\left(\frac{\sigma_w}{\sigma_y}\right) + \frac{1}{2}[\log(2\pi) + 1] - \log(2\hat{r}), \tag{3.15}$$

$$\text{SampEn}_L(\hat{r}) = \log\left(\frac{\sigma_w}{\sigma_y}\right) + \frac{1}{2}\log(4\pi) - \log(2\hat{r}). \tag{3.16}$$

To keep consistency of the notation used in the literature, the letter r will be used in place of $\hat{r}$ in rest of this thesis. Now, if σ_y in equations (3.15) and (3.16) is replaced by $\sigma_w\sqrt{c}$ using equation (3.2), we get

$$\text{ApEn}_L(r) = \frac{1}{2}[\log(2\pi) + 1] - \log(2r\sqrt{c}), \tag{3.17}$$

$$\text{SampEn}_L(r) = \frac{1}{2}\log(4\pi) - \log(2r\sqrt{c}). \tag{3.18}$$

Hence, if r is fixed, Lake's estimates depend on the variance of the prediction error (σ_w^2). On the other hand, if r varies with σ_y (as common in practice), they depend on the coefficients of the model (thus also on the model order M) but not anymore on σ_w^2.

Obviously, for a white Gaussian noise, which is an AR process of order zero with $c = 1$ and $\sigma_w = \sigma_y$, the theoretical values in equations (3.17) and (3.18) still apply.

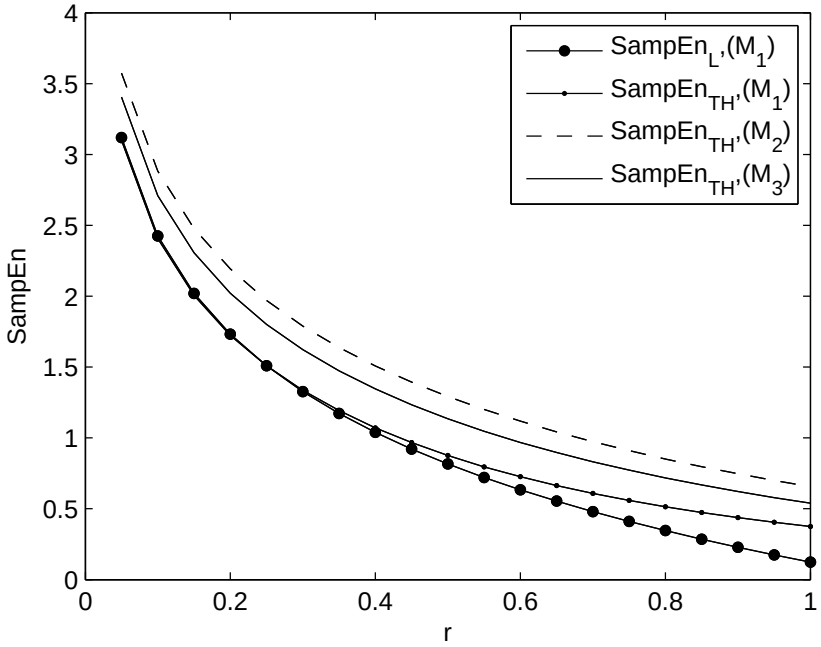


Figure 3.4: The convergence of SampEn for models, M_1: $[1, -0.77]$, M_2: $[1, -0.04, 0.87]$, and M_3: $[1, -0.56, 0.03, 0.4]$ with $N = 10000$, $m = 1$, and values r over the range $(0.05, \cdots, 1) \times$STD. SampEn_{TH} and SampEn_{μ} approximately coincides for every model. SampEn_{TH} and SampEn_{μ} closely converges with SampEn_{L} only for M_1. This does not happen in the other two cases, since the value of m is less than the order of the other two models

3.2.4 COMPARATIVELY RELIABLE ENTROPY ESTIMATIONS FOR FINITE N

The expressions provided in the previous two sections are asymptotic in the limit $N \to \infty$. Numerical estimates for short series obtained using the classical numerical algorithms ([44] for ApEn and [66] for SampEn) might be still far from the expected values. This is illustrated in figure 3.2(a).

A possible operative approach to obtain expected values of these estimates for finite and small values of N is to perform a specific number of Monte Carlo simulations[1], and then measure ensemble statistics. While this is seldom possible with real series, given the lack of stationarity over time, it is fairly easy with synthetic series obtained from AR models, which can be generated at will. Hence, in each realization, a synthetic series of length N is generated. The values of entropies are estimated, using the classical algorithms [44, 66], for specific values of m and r. Finally, an estimate of the probability density function (PDF) of the statistics is obtained (*i.e.* with an histogram), from which mean and STD can be estimated. In the following, the mean values of ApEn and SampEn computed from K realizations of the process will be referred to as ApEn_{μ} and SampEn_{μ}. Correspondingly, ApEn_{σ} and SampEn_{σ} will label their standard deviations.

1 A problem solving technique used to approximate the probability of certain outcomes by running multiple trial runs, called simulations, using random variables [67].

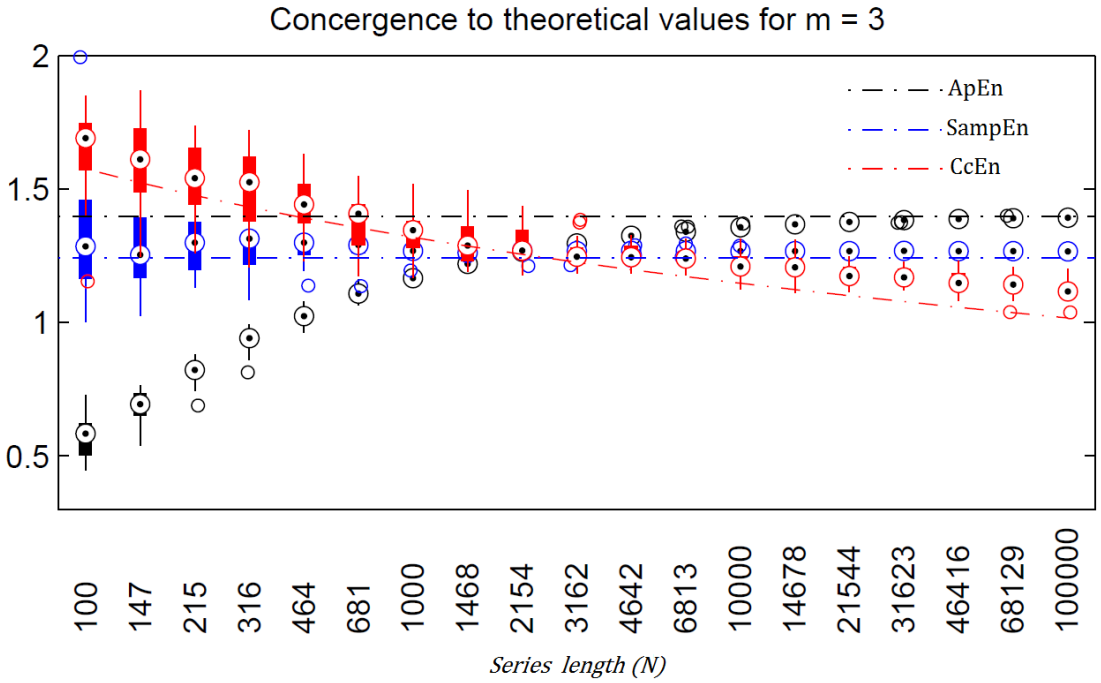


Figure 3.5: Convergence to theoretical values for an AR model with coefficients [1 -0.2 0.1] with σ_w=0.1, number of symbols $\xi = 6$. The boxplots represent the estimated values of ApEn, SampEn, and CcEn obtained through K=300 realizations of the Montecarlo's approach and the lines denote their corresponding theoretical values: $ApEn_L$, $SampEn_L$, and CEn_{TH}. The difference in CcEn and CE_{TH} with large series (N>3162) is due to the fact that the CE_{TH} is obtained using the equation 3.11, which requires $\xi \to \infty$ to get convergence with very large series.

Figure 3.2(b) illustrates the procedure and, together with figures 3.3 and 3.4, shows how the asymptotic values of the previous two sections match the numerical estimates, for various values of m, r and AR model order M.

As a rule of thumb, for an AR model, we can expect $SampEn_\mu$ to converge much earlier ($N \approx 100$) to $SampEn_{TH}$. However $SampEn_\sigma$ is always larger than $ApEn_\sigma$, suggesting that SampEn trades a much smaller bias at the expenses of a larger variance of the estimates. Also $SampEn_\sigma$ grows significantly with m, if N is fixed. Regarding the asymptotic values, $SampEn_\mu$ converges to $SampEn_{TH}$ when N is large enough, and the two values start matching $SampEn_L$ only when m is larger than the AR model order. The convergence of theoretical values for ApEn, SampEn and CcEn is illustrated in figure 3.5.

3.3 ENTROPY RATE VERSUS LEMPEL-ZIV COMPLEXITY

We have explained entropy rate and Lempel-Ziv complexity in the previous chapter. Here, we will discuss some relative properties of them for a stochastic stationary process. Let $\{\mathcal{X}_i\}_{i=1}^n$ be a stationary stochastic process (*e.g.* an AR process) and

$l(X_1, X_2, \cdots, X_n)$ be the Lempel-Ziv codeword length (using LZ78) associated with $X_1, X_2, \cdots, X_N$. Then it is shown in [3] that

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} l(X_1, X_2, \cdots, X_n) &\leqslant ShEn(\mathcal{X}) \\ \lim_{N\to\infty} \frac{1}{N} (LZC78(\mathcal{X})) &\leqslant ShEn(\mathcal{X}) \\ LZC78_{norm}(\mathcal{X}) &\leqslant ShEn(\mathcal{X}), \end{aligned} \tag{3.19}$$

where, $ShEn(\mathcal{X})$ and $LZC_{norm}(\mathcal{X})$ denote respectively, the entropy rate and normalized LZC78 of the process.

Thus the LZC of the encoding process using LZ78 algorithm of a stochastic process asymptotically should not exceed the entropy rate of the source. To verify this, we consider an arbitrary Gaussian AR process of order 2, and we generate synthetic series of some lengths ranging from 100 to 100000. The discrete time series generated through the process is converted to some symbolic sequences using 2, 3, and 4 different symbols. The expected entropy rate and Lempel-Ziv complexity are estimated by the average of K=100 realizations of the synthetic series generated through the process. The expected entropy rate and Lempel-Ziv complexity with different numbers of symbolic representation is shown in figure 3.6.

We see that LZC is less dependent than entropy rate on the number of symbols. The $LZC78_{norm}$ tends to converge to the entropy rate at very long series, when only sequence of two symbols are used. With increasing the number of symbols the entropy rate increased, and it upper bounds the $LZC78_{norm}$ for any series length.

If we keenly observe figure 3.6, it is seen that there is a drift in $LZC78_{norm}$ value in panel (b) than other two panels. This drift is due to the fact that the value of the term $\lceil \log_2(Q) \rceil$ is equal for Q=3 and 4. Similarly, the value of this term for Q=5, 6, 7 are same and is equal to the value for Q=8. Thus Lempel-Ziv complexity estimation for any number of symbols q, such that $\log_2(q) < \log_2(Q)$ differs only in the number of distinct phrases.

So far we discussed the behavior of LZC78 and entropy rate of a Gaussian AR process. Let us see, what happens with LZC76. The estimates of LZC76 and the entropy rate for a sequence of 2 symbols, which are mainly shown in the literature are depicted in figure 3.7.

If we compare, figures 3.7 and panel (a) of 3.6, we see that the $LZC78_{norm}$ is always larger than 1, even at very long 100000 series lengths. On the other hand, the $LZC76_{norm}$ using LZ76 parsing always remains below 1 for any series length and entropy rate upper bounds the $LZC76_{norm}$, as expected theoretically. So, we can conclude that LZC estimate using LZ76 is more reliable than LZ78, even at very short series. The one major disadvantage of LZ76 parsing is that it requires large parsing time.

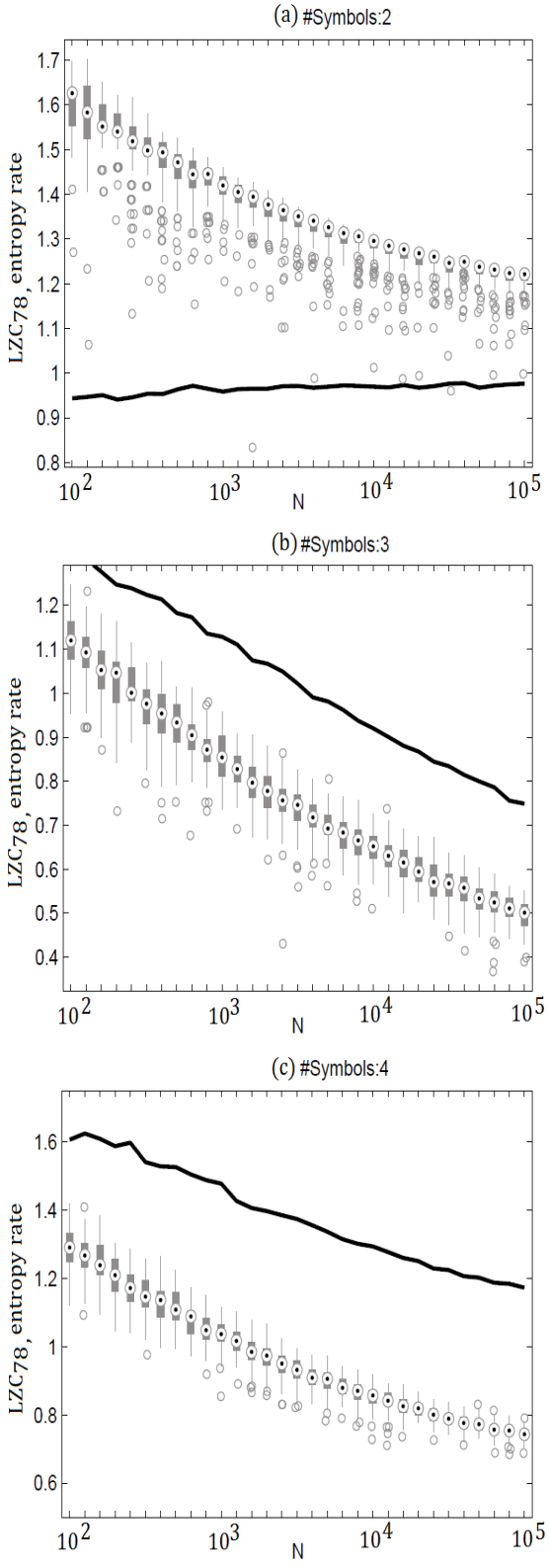


Figure 3.6: $\mathrm{LZC78}_{\mathrm{norm}}$ and the entropy rate of an AR process with coefficients { 1 -0.2 0.1} and variance of prediction error σ_w=0.1. The solid black line and boxplots denote entropy rate and $\mathrm{LZC78}_{\mathrm{norm}}$, respectively. Panel (a), (b), and (c) show the $\mathrm{LZC78}_{\mathrm{norm}}$ and entropy rate for sequences of 2, 3, and 4 symbols, respectively.

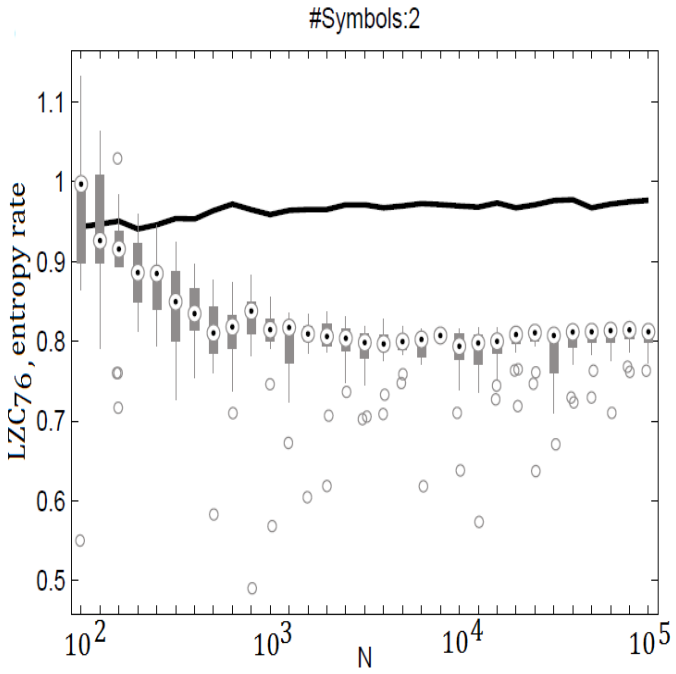


Figure 3.7: $LZC76_{norm}$ and the entropy rate of an AR process with coefficients { 1 -0-2 0.1} and variance of prediction error σ_w=0.1.

3.4 SUMMARY

In this chapter, we have set up a theoretical basis for the parametric estimation of entropy through autoregressive models, which are popularly used for power spectrum analysis of time-series. We started from the results available and developed new theoretical ones. The theoretical results matched the outcomes of the simulations performed.

SampEn and CcEn converges to their theoretical values earlier than ApEn. The standard deviation of ApEn is less than that of other estimates, even at short series. The difference in the expected value and the theoretical one of CcEn for long series ($N \geqslant 3000$) is due to its dependence on the number of quantization levels, ξ. It should converge with $\xi \to \infty$. The value of ξ should be increased with increasing N.

The $LZC78_{norm}$ converges very slowly, even with stationary stochastic (thus for AR) process at very long series. The quick convergence of Lempel-Ziv complexity is found for $LZC76_{norm}$, and its upper bound is defined by the entropy rate of the Gaussian stochastic process. The estimation of LZC expressed in symbols instead of bits is more appropriate, when estimating the complexity of a time series.

4

VALIDATION OF PARAMETRIC ESTIMATIONS ON REAL SERIES

4.1 INTRODUCTION

The common and major problem of entropy estimation is their dependence on series length. In the previous chapter, we have introduced parametric estimation of entropy, for very short series, when the traditional nonlinear measures of entropy may be impossible or suffers from the question of convergence. We have shown better convergence of parametric entropy with the theoretical ones for arbitrary AR models. We have also discussed the estimation of a very similar measure of entropy, the Lempel-Ziv complexity and its convergence with the entropy rate for a Gaussian AR process.

The validity of these parametric estimations should be justified on real series. To this aim, here we have considered RR series *i.e.* the series of intervals between the successive R peaks of an electrocardiogram (ECG), because they seem to be stationary on short period and they are available on some public database. The variation in RR intervals is called heart rate variability (HRV). The more details about RR series and HRV are provided in the next chapter.

In this chapter, at first we will explain the feasibility of parametric entropy estimations on HRV signals, which are available on the Physionet[1]. After studying the feasibility of the parametric estimations of entropy, the effects of the number of samples for reliable estimate of LZC from RR series extracted during sleep will be justified, as well as the LZC will also be compared with another related measure of entropy rate (*i.e.* SampEn). The study is focused on RR series extracted during the sleep, due to its small signal to noise ratio. The HRV of a subject can be significant during her/his different sleep stages. Only three sleep stages: light sleep (LS), *i.e.* stages 1 and 2 of NREM; deep sleep (DS), *i.e.* stages 3 and 4 of NREM; and rapid-eye-movement (REM). During sleep a person goes normally into different sleep stages:rapid-eye movement

1 Physionet offers free access to the large collections of recorded physiologic signals on the web

(REM), light sleep (LS), and deep sleep (DS), which are associated with different brain activity, and hence cardiovascular activity as well. The more details about the sleep physiology have been included in the next chapter.

4.2 DATA AND METHODS

In order to investigate if a parametric approach is feasible in practical sense, Physionet's normal sinus rhythm (nsrdb & nsr2db), and congestive heart failure (chf2db) databases are considered for this study. The "nsrdb" database consists of 18 (5 men, age: 26 to 45, and 13 women, age: 20 to 50) long-term ECG recordings of subjects without any significant arrhythmias. "nsr2db" includes beat annotation files for 54 Holter recordings of subjects in normal sinus rhythm (30 men, age range 28.5 to 76, and 24 women, age range 58 to 73). The third database, "chf2db" includes 29 long-term ECG recordings of subjects (age 34 to 79) suffering from congestive heart failure (CHF). The beat annotations were obtained by automated analysis followed by manual review and correction. The ECG sampling rate was 128 samples/second in all cases.

To study the effects of series length on LZC and entropy (here we consider SampEn), a sub-population (13 out of 16 subjects) of the "Cyclic Alternating Pattern (CAP) of the EEG activity during sleep" of [68, 69] database has been considered. Among 16, three subjects were removed due to the low quality or absence of the ECGRR series are extracted from the ECG signals, collected during sleep period. The sleep stage annotation series provided with the database are used for segmenting sleep stages.

4.3 FEASIBILITY OF PARAMETRIC ENTROPY ESTIMATIONS

The feasibility study verifies if parametric approach of entropy estimation is feasible on real series. Feasibility is justified by the agreement between numerical estimate and the expected average value of entropy for the series generated from the AR model. The coefficients define the model, which has generated the series, and hence the coefficients themselves contain all information of the signal dynamics including its entropy. Therefore, estimating entropy of the synthetic signals generated through the AR models gives an idea of entropy values for a purely linear process. On the other hand, ApEn and SampEn are nonlinear tools for measuring the regularity (or complexity of a time series). The numerical esitmates of the entropy may fall inside or outside the 95% standard range of the values obtained through K realizations of the model. If the numerical estimate lies within 95% of standard range of the values obtained through K realizations of the model, then they are considered in agreement, otherwise disagree. The agreement of numerical estimates with the expected average value of the model implies that the entropy is only due to an aggregated index of linear properties of the series, which is also shown with other traditional temporal or spectral parameters [70].

The disagreement between numerical estimate and the expected average value of the model implies that the series contains nonlinear, nonstationary or nonGaussian components, and its dynamics cannot be described by the purely linear AR process. The agreement between numerically estimated value and the value obtained through

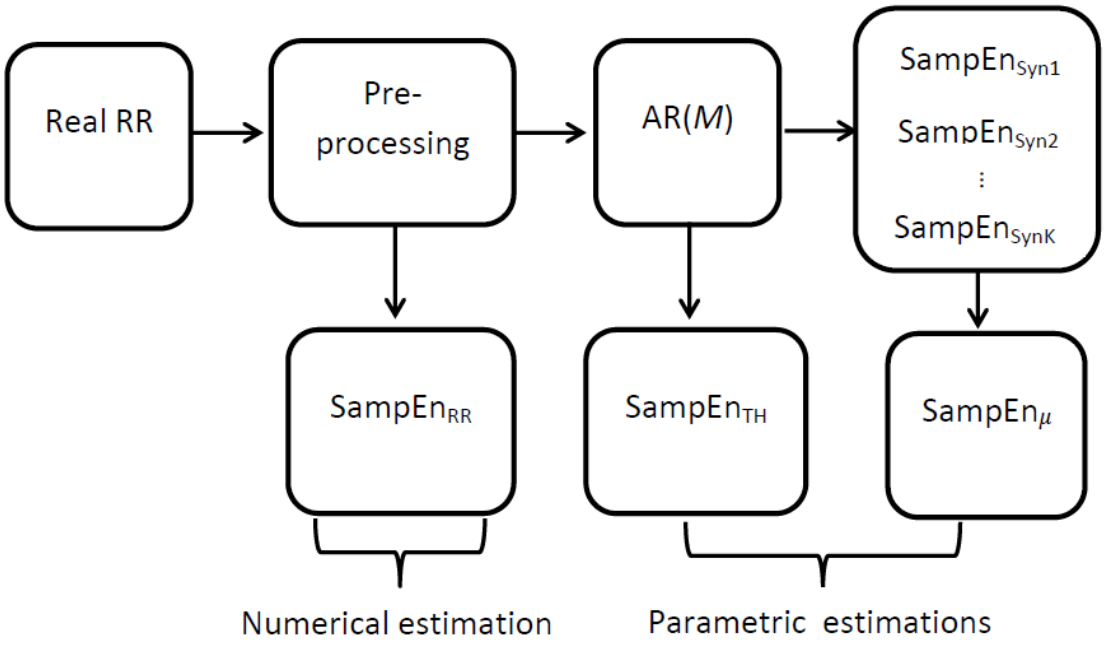


Figure 4.1: Block-diagram of the parametric SampEn estimation on real data. SampEn$_{RR}$ is the SampEn numerical estimations of the RR series after pre-processing

the models may decrease with increasing the series length due to increasing the non-stationarity. To investigate the source of this disagreement, in particular if it might be ascribed to non-Gaussianity, surrogated data can be used.

The methodology for parametric estimation of entropy (*e.g.* SampEn) are listed in the block-diagram (Fig. 4.1). It consists of the following steps:

- "Pre-processing". The RR series may contain ectopic beats and artifacts. To remove them, two levels of pre-processing are performed. In the first stage, only those extreme artifacts which lie outside the range $[Q_1\ Q_3]\pm 3\times$ IR are removed. Here IR=Q_3-Q_1 is the interquartile range, between the third (Q_3) and first (Q_1) quartiles. Ectopic beats are instead excluded in the second step, when only those RR which lie within 20% of the previously accepted RR interval were retained. The first accepted RR value of each series must lie within the IR.

- "AR(M)". For each subject, RR series of some specific lengths (N=75, 150, 225, 375, 750, and 1500) are chosen in overlapping fashion (overlapped by 50%) and fitted to an AR model. The model order, M is determined by satisfying the Akaike information criterion (AIC) [71] and the Anderson's whiteness test [72].

- "SampEn estimation". All three measures of SampEn: SampEn$_{RR}$, SampEn$_{TH}$, and SampEn$_{\mu}$ have been considered. The value of SampEn$_{RR}$ is estimated numerically using the procedure described in section 2.4.1.4. The value of SampEn$_{TH}$ is determined using equation 3.4. Besides these, SampEn$_{\mu}$ is obtained by averaging estimations of $K = 300$ Monte Carlo simulations of the AR models as described in section 3.2.4.

The agreement between numerically estimated value of entropy ($SampEn_{RR}$) and the value obtained for the series generated through $K = 200$ realizations of the AR models for different series lengths N={75,150,225,300,375,750,1500} are observed for the considered datasets.

4.4 ANALYSIS OF SERIES LENGTH FOR ROBUST ESTIMATION OF LZC

The variation in complexity or regularity of physiological signals is sensitive to different pathological conditions. Lempel-Ziv complexity (LZC) has been used to discriminate between WAKE and SLEEP in patients under anesthesia [73] from electroencephalogram (EEG) and to compute the EEG background activity in patients with or without Alzheimer's disease [74], and the complexity computation of autonomic nervous system (ANS) from HRV analysis. There are many common applications of LZC and entropy reported in the literature [75, 76, 77, 78]. In fact, there are some researches those show how changes in complexity can distinguish between ventricular tachycardia [79] and atrial fibrillation [80], or how ANS control is modified by pathological conditions, such as sleep apnea or heart failure.

Despite to their proven capability, what LZC can really measure from biomedical signals is not clear enough. We want to re-mention that the estimation of this metric of a time series depends on the transformation of the time-series into symbolic sequence. Aboy *et al.* [76] has reported that LZC is dependent on the frequency related quantities (using binary symbolic sequence) on running window of 10 sec. The effect of the number of samples and long-term nonstationarity have been totally neglected in their study.

In this study, to move a step forward, we planned a few synthetic simulations and real data analysis to verify which minimum number of samples should be employed for obtaining a robust estimate on HRV signals expressed as RR series. Moreover, we compared LZC with another complexity measure, *i.e.* SampEn, in order to assess at which extend they are related. We focused our efforts on RR series extracted during sleep because of its optimal signal-to-noise ratio. However, the RR series during sleep can vary significantly with the different sleep stages [81]. For this reason, here we have considered three sleep stages (according to standard sleep labeling [82, 83]): Light Sleep (LS), represented by NREM stage 1 and 2; Deep Sleep (DS), represented by NREM stage 3 and 4; and Rapid-Eye Movements (REM).

The proposed system consists of the following methodological steps:

- "Preprocessing and data modeling". The RR series is first high pass filtered using median filter of 200 samples, and then extreme artifacts are removed. Each RR segment of 400 samples from successive windows (with no overlapping) is fitted to the AR models of fixed order 9. Model fitting is performed for individual sleep stages of each subject using the Yule-Walker equations. The model, whose prediction error is more than 0.05 sec^2 is excluded from the analysis. Thus a variable number of models are obtained depending on the subject and the sleep stage (*e.g.* 219 DS, 272 LS, and 169 REM).

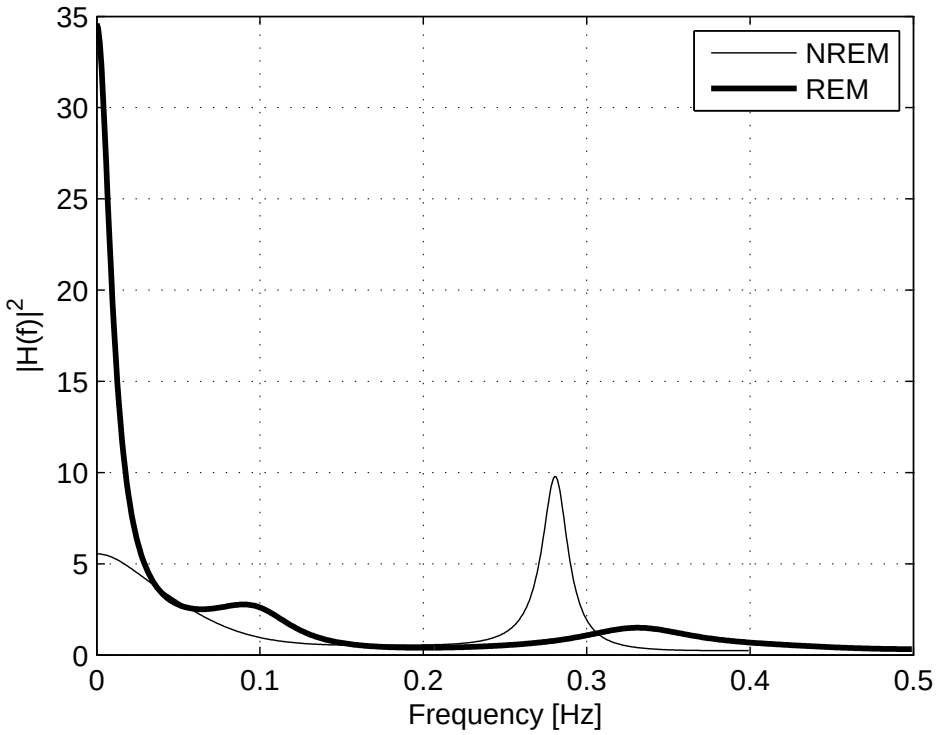


Figure 4.2: Squared magnitude of the frequency response of the two AR models mREM (light) and mNREM (bold line).

- "Symbolic representation". The RR segments have been converted to the symbolic sequence using "quantization with equiprobable distribution", as described in section 2.5.1 of chapter 2. The break points are defined by the percentile values (*i.e.*, each interval has the same probability).

- "Expected value estimation of LZC". The control of the ANS is reflected on a predominant power in the high frequency during NREM, and in the low frequency band during REM sleep. At first, two general AR models: mNREM (during NREM) and mREM (during REM) are estimated from the real data such that they contain the dominant power in the corresponding (NREM: [0.15 to 0.4]Hz and REM: [0.04 to 0.15] Hz) frequency bands. The squared magnitude of the frequency response of both models are shown in figure 4.2. Synthetic series of different lengths $N : \{10, 10^2, \cdots, 10^5\}$ are generated for the AR models through K=30 realizations following Montecarlo's approach. The series is then converted to a symbolic sequence of 2, 3, and 4 symbols using the technique described in section 2.5.1 of chapter 2. The expected value LZC ($LZC76_\mu$) for each series is estimated using LZ76 parsing technique by taking the average estimates of 30 realizations.

- "Evaluation of the series length". The evaluation of the number of samples is carried out by two different approaches. In the first approach, the percentage of variation for each series length (N) is compared with respect to the average value obtained at N maximum (10^5), and the number of samples for having

variation (less than 5%) is determined by the linear interpolation. In the second approach, a statistical (double tail t-test is applied to compare the average values $LZC76_{\mu}$ between successive values of N (*e.g.* $LZC76_{\mu}$ for N=10 vs N=100; N=100 vs N=1000. In addition, the double tail t-test is also applied on the mean estimates of two models to observe if it can distinguish the two population with significant ($p<0.05$) values.

- "Relationship between Lempel-Ziv complexity and sample entropy". To evaluate the relationship between LZC and SampEn, $LZC76_{\mu}$ is estimated for a set of K=30 synthetic series of length N=23000, generated by the same AR models identified in "Inter-subject variability". The value of $SampEn_{TH}$) is determined with m=1 and r=0.2×STD. The linear correlation between $SampEn_{TH}$ and $LZC76_{\mu}$ is computed, considering: (i) no groups at all, and (ii) groups LS, DS, REM separately. Here, we are interested in the average correlation, not on the specific realization.

- "Effects of source distribution on Lempel-Ziv complexity". The effect of source distribution on the estimation of Lempel-Ziv complexity is evaluated by comparing $LZC76_{\mu}$ using both symbolic representations in section 2.5.1. It is clear that the symbolic transformation using equiprobable representation gives uniformity to the source distribution. However, the other representation technique transforms the series into symbolic sequence without any transformation of the distribution of the series. The estimated $LZC76_{\mu}$ is compared for both mREM and mNREM models using both symbolic representation techniques with different series lengths of the synthetic series.

4.5 RESULTS

The results obtained from the studies are presented separately. The results on feasibility of parametric entropy estimations are given first. Then the results on the effects of series length during different sleep stages are provided in this section.

4.5.1 RESULTS ABOUT FEASIBILITY OF PARAMETRIC ENTROPY ESTIMATION

In this section, the experimental results based on the feasibility study of applying parametric approach for entropy estimations from real data are presented. The possible reasons of infeasibility have also been explained from experimental point of view.

- "Agreement between numerical and parametric estimations". The agreement between $SampEn_{RR}$ and $SampEn_{\mu}$ is investigated for RR series of lengths N={75,150, 225,375,750,1500} from all considered databases. The total number of RR series analyzed is large (*i.e.* 2.7×10^5 and 1.4×10^4, respectively for $N = 75$ and $N = 1500$). The average number of times that $SampEn_{RR}$ falls within the standard range of the distributions of $SampEn_{\mu}$, for m=1 and r=0.2×STD, is reported in table 4.1(a). For $N = 75$, $SampEn_{RR}$ is in agreement with $SampEn_{\mu}$ for more than 83% cases

Table 4.1: Average agreement (%) of $SampEn_{RR}$ with $SampEn_{\mu}$

(a) Three database

	Series length N					
Database	75	150	225	375	750	1500
nsrdb	83.63	72.15	69.19	58.93	46.13	37.61
nsr2db	83.20	73.89	66.60	44.64	46.13	35.86
chf2db	83.52	73.90	67.31	56.95	41.64	28.15

(b) "nsr2db" database with N=300

	r						
m	0.1	0.15	0.2	0.25	0.3	0.35	0.4
1	58.26	60.58	60.56	63.16	65.55	67.44	69.28
2	66.84	62.20	57.27	56.97	57.83	58.80	60.64
3	83.02	73.21	62.66	57.88	55.65	55.10	56.14
4	92.67	83.91	72.05	63.85	58.54	55.28	54.51

Table 4.2: Average agreement (%) of $ApEn_{RR}$ with $ApEn_{\mu}$

(a) Three database

	Series length N					
Database	75	150	225	375	750	1500
nsrdb	84.12	74.17	66.96	45.93	46.32	36.16
nsr2db	83.88	74.36	67.76	57.85	44.67	34.65
chf2db	84.39	75.04	68.10	56.61	42.10	30.51

(b) "nsr2db" database with N=300

	r						
m	0.1	0.15	0.2	0.25	0.3	0.35	0.4
1	58.26	60.58	60.56	63.16	65.55	67.44	69.28
2	66.84	62.20	57.27	56.97	57.83	58.80	60.64
3	83.02	73.21	62.66	57.88	55.65	55.10	56.14
4	92.67	83.91	72.05	63.85	58.54	55.28	54.51

in each database. This figure reduces to about 28% for $N = 1500$. The average agreement is shown in table 4.1.

Similar results have been obtained for ApEn and CcEn(as long as the parameters employed were suitable to the situation at hand). The average agreement of $ApEn_{RR}$ with $ApEn_{\mu}$ on "nsr2db" database, for $m = 1$, $r = 0.2 \times STD$ and

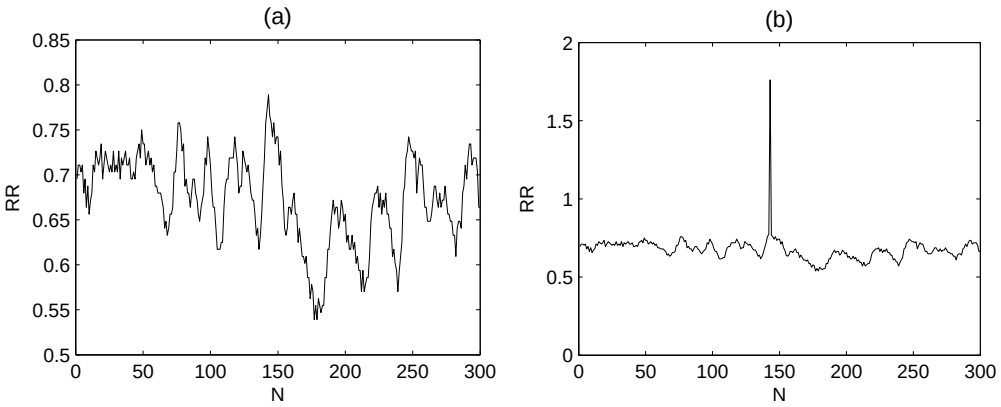


Figure 4.3: Effect of spikes. Panel (a): RR series of N = 300 points with $\text{SampEn}_{RR} = 1.0102$ and $\text{SampEn}_{\mu} = 1.1182$. Panel (b): a spike of amplitude 20×STD of the series has been artificially added and now $\text{SampEn}_{RR} = 0.5983$ and $\text{SampEn}_{\mu} = 2.1376$ respectively. Please notice that in both cases, AR model identification satisfied AIC and whiteness test. Also, the STD of the series increased significantly with the addition of the artifact.

N = 300 is presented in table 4.2, which is very similar to the obtained results for SampEn under the same conditions.

- "Possible reasons for disagreements". To investigate the source of this disagreement, in particular if it might be due to non-Gaussianity, surrogated data of the original RR segments from the "chf2db" database are constructed in such a way that the original distribution is replaced by a Gaussian one having the same STD while preserving the order of ranks. In practice, WGN of the same length, mean and variance of the RR series is generated first. Then the samples in the RR series are replaced by the ones in the WGN, such that the ordering of the samples is preserved, *i.e.* a sample that is at position i, once sorted the original series, is in the same position in the sorted surrogate one. Once repeated the procedure described in the "Preprocessing" part of section 4.3 on the surrogate series, the figure of agreement of SampEn_{RR} with SampEn_{μ} increased for every considered lengths. In specific, it rose to 88.12% (N = 75), 80.54% (N = 150), 75.26% (N = 225), 66.56% (N = 375), 52.93% (N = 750), and 39.88% (N = 1500), respectively. The percentage of agreement decreases (as expected) with increasing the series length which suggests an additional effect due to non-stationarity.

- "Effects of the editing methods". The estimation of SampEn is influenced by the editing method. In fact, if the series contains artifacts due to missing or wrongly detected beats (or ectopic beats, spikes, ...) then a large disagreement is found between SampEn_{RR} and SampEn_{μ}. In [84], the authors have explained the effect of spikes for numerical estimation of SampEn. Here, the effect of spikes is further investigated by considering a RR series of N = 300 points from a subject of the "nsr2db" database (record# nsr001), as shown in figure 4.3. A spike, of amplitude 20×STD of the series, is added at the time-index of maximum amplitude. SampEn_{RR} and SampEn_{μ} are estimated before and after the spike is added.

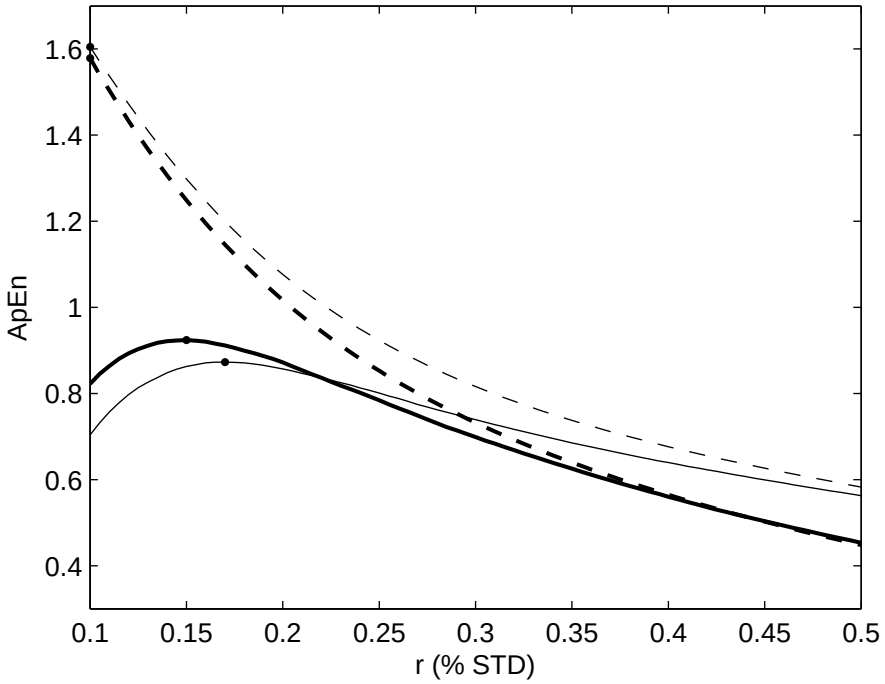


Figure 4.4: Values of $ApEn_\mu$, for $m = 2$, as a function of r. The series were generated from the same AR model of figure 4.5 with $\rho_1 = 0.4$ and $\rho_2 = 0.2$ (case 1, thick lines) or $\rho_1 = 0.9$ and $\rho_2 = 0.8$ (case 2, thin lines). The two continuous lines are for $N = 300$ ($K = 200$) and the two sketched ones for $N = 10000$ ($K = 10$). Dots mark the largest values of approximate entropy obtained varying r. They should be used to characterize the complexity of the series, as suggested by Lu *et al.* [85] as shown in figure 4.4. However, case 1 would appear less regular than case 2 for $N = 300$, but not for $N = 10000$.

Although the model identification satisfied AIC and whiteness test in both cases, with adding the spike the value of $SampEn_{RR}$ decreased, followed by a large increase in $SampEn_\mu$. Also, the SampEn of the original series matched with that of the model, but not after including the spike. In fact, spikes increase the STD of the RR series, as well as the effective value of r (which is proportional to STD), leading to a smaller $SampEn_{RR}$ (analogously to figure. 3.3).

- "Effects of the parameters m and r". The choice of the parameters m and r gained large attention due to the inherent sensitivity of both ApEn and SampEn. For smaller value of r poor estimates of conditional probability are achieved, while for larger value too much information about the system is lost. Also, to avoid remarkable contribution of noise in entropy calculation, the value of r must be larger than most of the noise. Values of r in the range of $[0.1, 0.25] \times$ STD have been shown to be applicable to measure the regularity (or complexity) of a variety of signals [44, 66, 43]. Traditionally, $r = 0.2 \times$ STD is used for measuring the regularity of HRV. The selection of the value m might depend on the series length.

  The effect of modifying the values of parameters m and r, on the average agreement of $SampEn_{RR}$ with $SampEn_\mu$, has been investigated for the subjects of the "nsr2db" database. The results are mentioned in table 4.1(b).

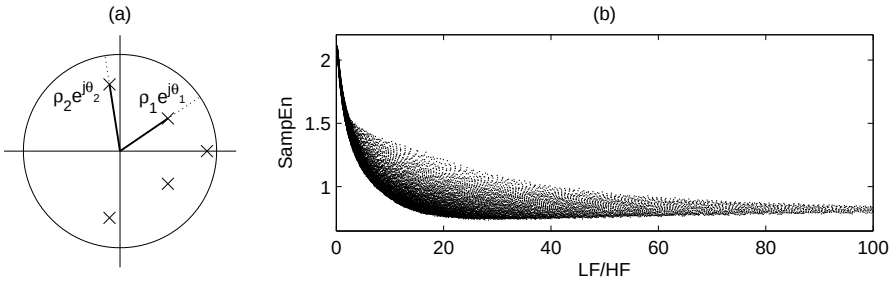


Figure 4.5: Values of $SampEn_\mu$, with $m = 2$, $r = 0.2$ and $N = 300$, for series generated by a fifth order AR model. Panel (a): the poles of the model were located along the real axis ($\rho_0 = 0.9$) and at middle of the LF and HF bands: $\theta_1 = 2\pi(0.04 + 0.15)/2$ and $\theta_2 = 2\pi(0.15 + 0.40)/2$. The magnitudes of the four complex poles, ρ_1 and ρ_2, were varied in the range 0.05–0.95 (with step: 0.005). For each case, $SampEn_\mu$ was obtained from $K = 200$ Monte Carlo realizations. The power's contents in the LF and HF bands were computed by integrating analytically the power spectral density of the AR process. The individual values of $SampEn_\mu$ are plotted in panel (b) as a function of the LF/HF ratio.

In our study, the highest agreement of $SampEn_{RR}$ with $SampEn_\mu$ was found for larger m and smaller r. This was due to the fact that in this range the number of matches in the SampEn computation was very limited, and making any distinction very hard. However, as table 4.1 shows, when m=1, the average agreement slightly increased with r, as expected for short series of $N = 300$ points. So, this study tentatively favors the selection of m=1, r=0.2 for short series.

As a rule of thumb, the analysis performed here confirms that the bias of ApEn and the variance of SampEn decrease with N. However, to the best of our knowledge, there is no recognized consensus in the literature on how to select the values of the parameters r and m, especially when using approximate entropy. Even very popular rules, like the one by Lu *et al.* [85], might fail, as figure 4.4 shows.

On the other hand, AR models are stationary and can provide series of uniform characteristics of any lengths. These, along with the theoretical results available, can offer an insight of the variability of the estimates for the values of the parameters (and the number of points) selected. Given the problem at hand, an AR model which is close enough (discarding possible nonlinearity and non-Gaussianity) to the series under study is a good test-bench for selecting m and r.

4.5.2 RESULTS ABOUT RELIABLE ESTIMATES OF LZC

In this section, the results obtained from the analysis of LZC during sleep will be explained. First, the assessment on the minimum number of samples required by LZC to have no changes in its mean values will be presented. Then, we will discuss the relationship between SampEn and LZC. The study results are summarized as:

- "Evaluation of series length".

The number of necessary samples required by the LZC76, *i.e.* for having no statistical significant changes in its average value ($LZC76_\mu$), is evaluated for two AR models: mNREM and mREM. In particular, considering mNREM, the 1% of variation is reached at N=9300 with number of symbols Q=2 (1000 at 5%), 52000 with Q=3 (5500 at 5%) and 66100 with Q=4 (6800 at 5%). Similar results are obtained considering mREM (1%: at N=53500 with Q=2, 45300 with Q=3 and 64900 with Q=4; at 5%: 3900 with Q=2, 3500 with Q=3 and 6300 with Q=4). However, $LZC76_\mu$ at $N=10^4$ is not different to that at $N=10^5$ only when considering Q=2 for both models (figure 4.6; $p<0.01$).

- "Relationship between LZC and SampEn". The relationship between LZC and SampEn is evaluated by considering (i) LS, DS, and REM groups separately (figure 4.7) and (ii) with no grouping. In the first case, the linear correlation between LZC and SampEn is always more than 0.75 ($p<0.01$). In particular, when considering LS and REM with Q=2, the correlation is more than 0.90. In the second case, the linear correlation is more than 0.90 ($p<0.01$) for both Q=2 and Q=3 (the sample is not balanced). Table 4.3 summarizes the relations and the correlations found.

Table 4.3: The linear relationship between LZC and SampEn (ALL) is shown. Also, relations are reported as function of the sleep stage and the level of quantization Q. Linear correlation is shown in brackets ($*$ refers to $p<0.01$).

Sleep stage	Q=2	Q=3
LS	2.20×LZC + 0.52 (0.92 *)	2.17×LZC + 0.38 (0.93 *)
DS	1.52×LZC + 1.02 (0.75 *)	1.65×LZC + 0.81 (0.81 *)
REM	2.25×LZC + 0.44 (0.97 *)	2.26×LZC + 0.28 (0.98 *)
ALL	2.20×LZC + 0.52 (0.90 *)	2.21×LZC + 0.35 (0.93 *)

- "Effects of quantization methods". The effects of the method applied to transform the RR series into symbolic sequence has been investigated. The values of $LZC76_\mu$) are estimated for both mNREM and mREM models, and using both quantization techniques (described in section 2.5.1) as a function of N with Q=4 . Although, a small difference is shown for finite short series, but the estimates are same for very long (N>10000) series. Thus it is verified that the LZC complexity is independent on the distribution of the source.

4.6 OVERALL EVALUATION OF PARAMETRIC ENTROPY ESTIMATION

The theoretical computations of SampEn (also ApEn and CcEn) matched with the outcomes of the simulations performed. Regarding simulations on real series, the obtained results show that when non-stationarity, nonlinearity or non-Guassianity are minimal, $SampEn_{RR}$ does match with $SampEn_\mu$. However, inherent nonlinearitiy, non-

Gaussianity, non-stationarity or the presence of ectopic beats or artifacts induce the two estimates to differ. This is supported by the experiments done with surrogated data and adding spikes to the original RR series. Given the fact that the chance of non-stationarity in biomedical signals (and RR intervals series in particular) increases with the series length, the percentage of agreement of $SampEn_{RR}$ with $SampEn_{\mu}$ decreases accordingly. Hence, the shorter the series the more likely the effectiveness of the parametric approach on real RR series. Also, parametric estimates might be helpful for very short series (less than 90-100 points), where traditional values of SampEn are often undefined [66]. This is not the case for ApEn, which is defined at any length N due to the inclusion of self-matches. However, they produce a large bias in the estimates, which are far away from the asymptotic value (figure 3.2) for short series.

When SampEn values, obtained from an RR sequence and from the AR model fitted onto it, do coincide, the regularity/complexity measured by both depends only from the autocorrelation function of the series, as equations (3.18) and (3.4) prove. In this circumstances, SampEn is only an aggregated index of linear properties of the series, which were likely available using other traditional temporal or spectral parameters [70]. For instance, the relation between a common standard spectral parameters, as the LF/HF ratio, and SampEn, is illustrated in figure 4.5 for series obtained from an AR process (while varying the position of the poles and, thus, of the spectral content). SampEn varies with the LF/HF ratio, being maximal when the latter is minimal. This is a consequence of the fact that SampEn is larger when the power is spread along the entire frequency axis (the signal is "more similar" to a WGN).

4.7 OVERALL EVALUATION ON THE EFFECTS OF SERIES LENGTH FOR LZC

In this experimental study, the number of samples required for getting a reliable estimation of LZC during sleep has been verified. The minimum number of samples required by LZC for having no change in its average value, during a specific sleep stage is 10^4 (which is practically impossible to collect for a single sleep stage), when employing binary quantization (figure 4.6; $p<0.01$). However, a variation (<5%) is found, when employing $N > 1000$ for both Q=2 and Q=3.

A number of quantization levels Q>2 is not recommended because more than 10^5 samples are required (figure 4.6). It remains to accurately evaluate if LZC can distinguish NREM and REM, even before convergence (partially demonstrated; figure 4.6; $p<0.01$) and, if its value reflects physiological information. Furthermore, the study of which quantization technique should be used, results that there is no effect of the quantization method applied for converting the series into symbolic sequence.

The proper evaluation of the number of samples required by LZC during sleep on real data is not feasible due to the presence of non-stationarity. However, the cyclic behavior of sleep stages during the night leads to a reduction of the variability of the LZC, when increasing the number of samples making possible an empirical evaluation.

Finally, the linear correlation between LZC and SampEn is assessed on a synthetic dataset (table 4.3 and figure 4.7, >0.90; $p<0.01$). Such comparison is meant to evaluate if the two metrics carry different information, even though the quantization proce-

dure employed for LZC (standard parameter were considered for SampEn) is not the same. From our experiments, it seems to be excluded. Therefore, the methodology we proposed [86] can be applied for estimating LZC from very short series for better estimation. This result infers to verify which of these two metric converges more quickly, when long series are available.

In conclusion, the metric LZC is suggested to apply for $N \geqslant 1000$ using binary quantization if a variation smaller than 5% is considered, or at least 10^4 for maximal accuracy. The quantization levels Q more than 2 is not recommended.

4.8 SUMMARY

In this chapter, a detailed study on the possibility and significance of performing a parametric estimation of entropy on real series (available on Physionet) has been provided. The feasibility has been justified by the agreement between numerical estimation and the estimation of entropy obtained through K realizations of the model. The feasibility of parametric estimations has been positively justified on short series, even in case of long series, it gives additional information if there are some nonlinearity, nonstationarity, or nonGaussianity.

The work supports the finding that when numerical and parametric estimates of entropy agree, it is mainly influenced by linear properties of the series. A disagreement, on the contrary, might point those cases where numerical estimation truly offers some new information that is not readily available with traditional temporal and spectral parameters. This disagreement also infers that there are some nonlinearity, nonstationarity, or nonGaussianity in the series. Thus, parametric estimation can be used also for statistical analysis of the series.

The linear correlation between LZC and SampEn was computed on a synthetic dataset. When estimation of LZC from short series is required, our proposed methodology of parametric estimation could be applied.

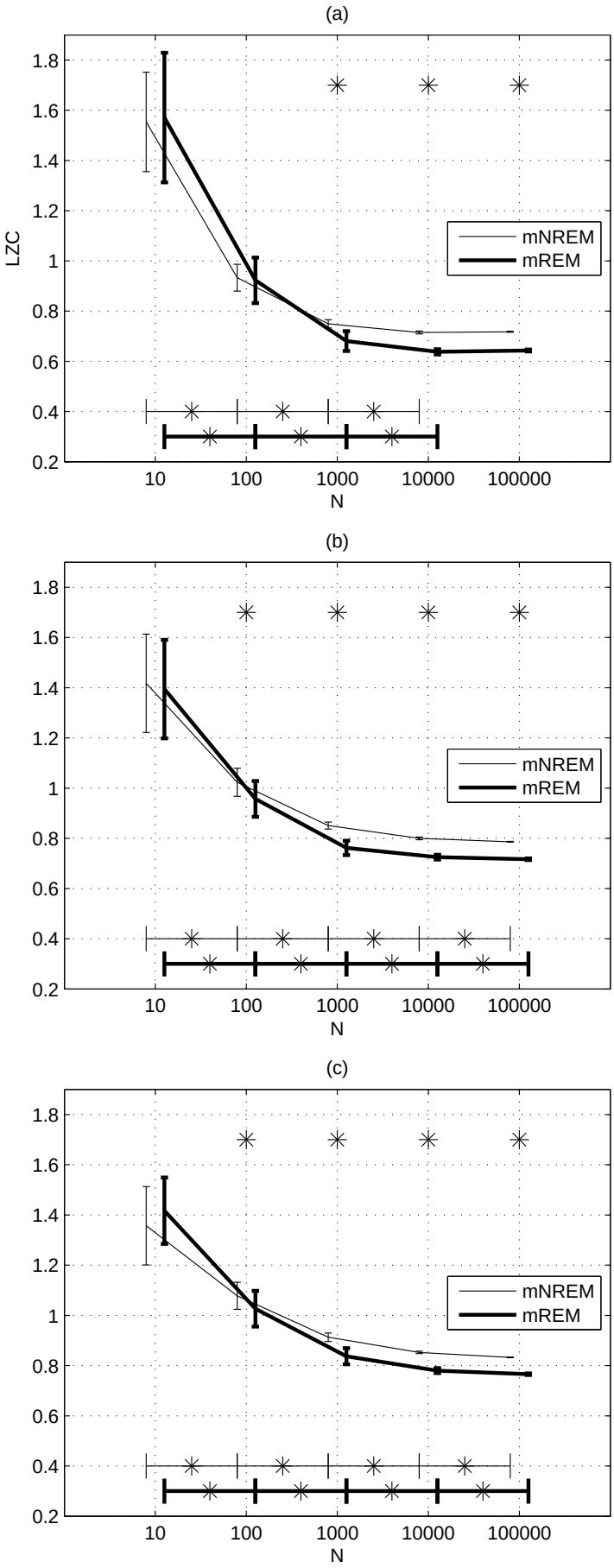


Figure 4.6: Mean and standard deviation of LZC as function of the series length N when considering mNREM (light line) and mREM (bold line) and with levels of quantization $Q = 2$ (a), $Q = 3$ (b) and $Q = 4$ (c). $*$ on the horizontal bars refer to the statistical difference in the average estimation between successive series lengths N, and $*$ on the top are used to denote the statistical difference between mNREM and mREM. $*$ refers to $p<0.01$ of double-tail t-test.

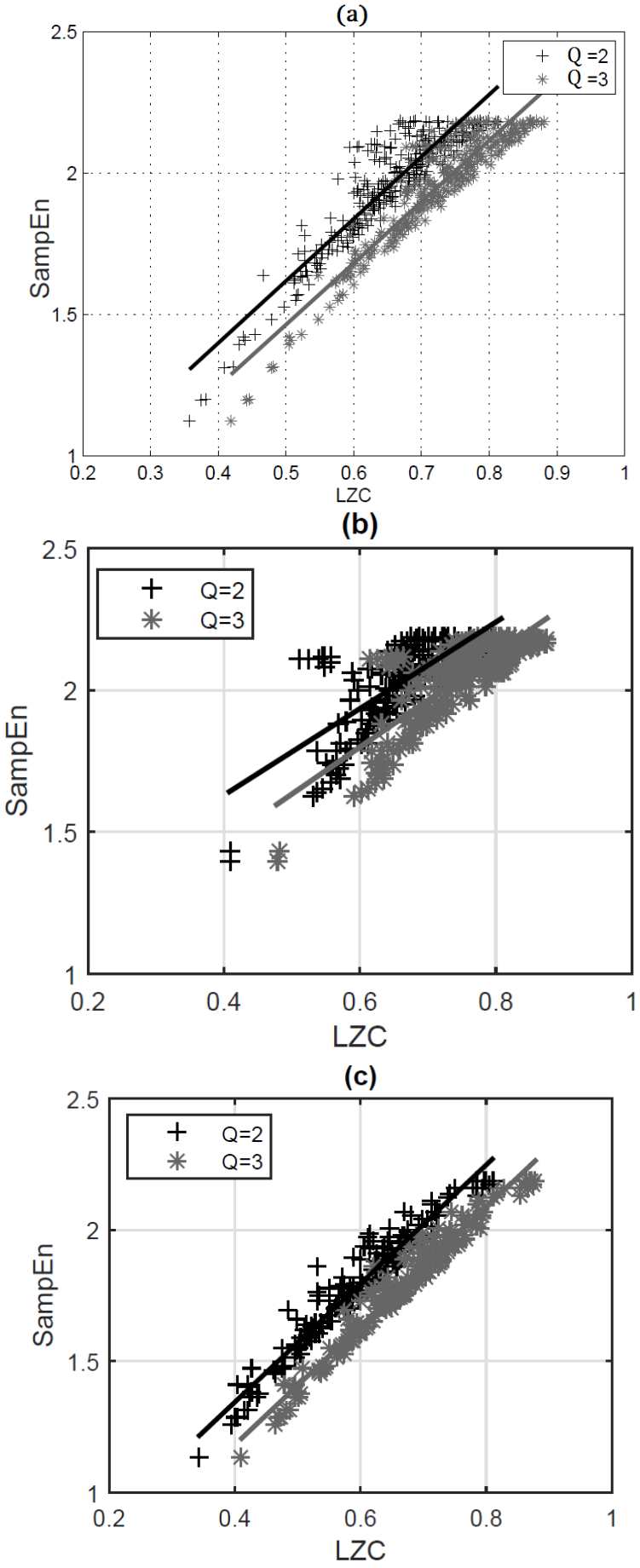


Figure 4.7: Scatter plot and linear regression between LZC and SampEn when considering LS (a), DS (b) and REM (c) with Q=2 (gray) and Q=3 (black).

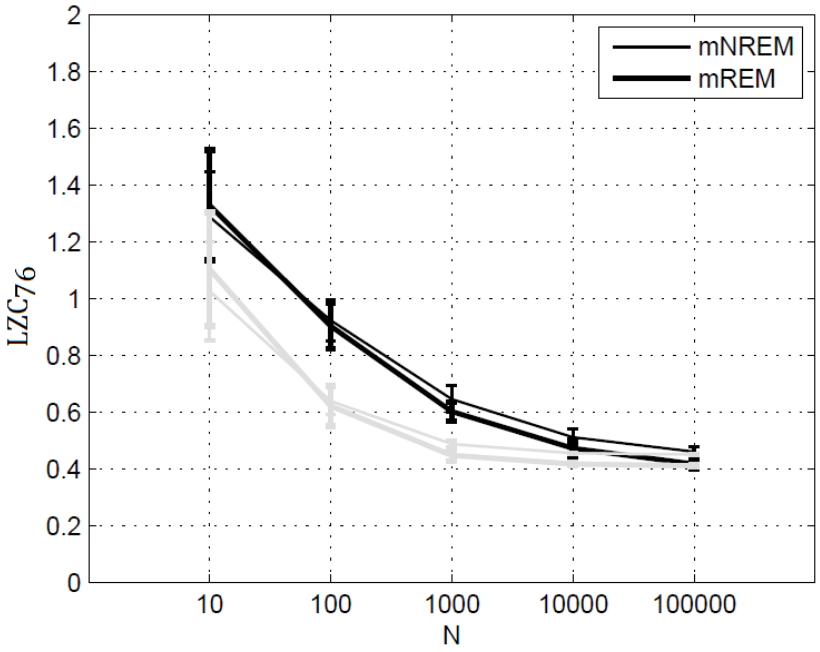


Figure 4.8: Mean and standard deviation of LZC76 as a function of the series length for mNREM (light line) and mREM (bold line), when considering uniform (black lines) and equiprobable (shaded) lines quantization techniques with Q=4.

5

ENTROPY BASED FEATURE EXTRACTION FOR PHYSIOLOGICAL SIGNALS

5.1 INTRODUCTION

Biological Signals such as beat to beat fluctuations in cardiovascular signals contain useful information to detect and characterize the patients with heart diseases, sleep disorders, etc. It is very important to early identify patients with heart dysfunction, because it may cause acute chest pain, breathing problems up to sudden cardiac death (SCD).

In the previous chapter, the feasibility of parametric entropy estimations have been studied. Now, it is supported that parametric estimations of SampEn are feasible for very short series. Moreover, the disagreement of numerical and parametric estimations of entropy might provide more information about the nonlinearity or nonstationarity presents in the series. So, proposed method of parametric estimations can be used for some statistical analysis of a signal.

In this chapter, the usage of entropy features has been shown in some real applications based on physiological signals analysis. To help readers to understand the methods and their applications, a short description of physiological signals (only those considered in this work) has been included at the beginning of the chapter. Moreover, a short state of the art has been given the beginning of each method, so that the readers can understand the importance and objectives of the method.

5.2 PHYSIOLOGICAL BACKGROUND

The human body is a very complex biological system. The heart is one of the most complex and indispensable part of human body. Although, it is usual that older people suffers from heart diseases, but sometimes it may happen to younger, even to the children. Suddenly an unexpected death may happen due to loss of heart functioning.

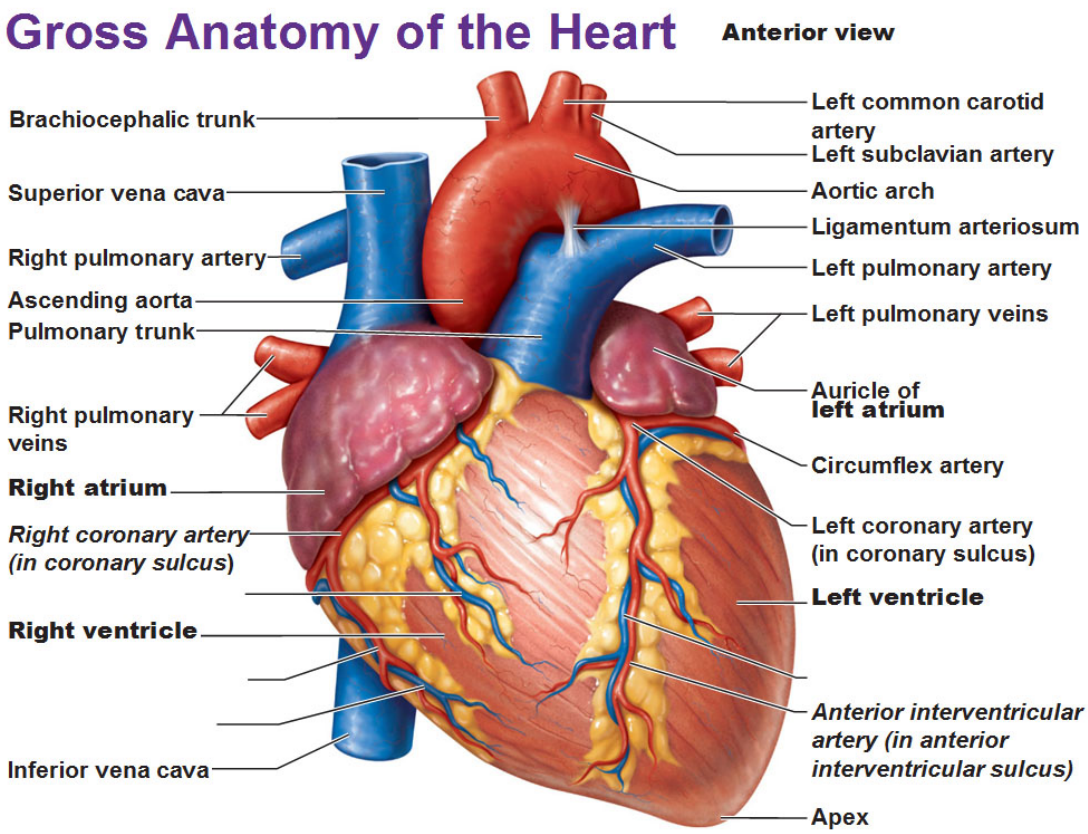


Figure 5.1: Physiology of human heart.Source: http://anatomyandphysiologyi .com/wp-content/uploads/2013/09/gross-anatomy-of-the-heart-anterior-view.jpg

In every year, about 300,000 to 400,000 adult deaths in the USA also due to SCD [87]. Proper treatment applied immediately may save the subject from such a SCD. Besides this, people suffer from many other heart diseases like atrial fibrillation (AF), coronary artery disease, congestive heart failure, heart attack, etc. The heart rate varies with the age, mood, disease, etc. Many heart diseases can be predicted by analyzing the heart rate variability.

In the rest of this section, we will briefly discuss about the physiology of human heart, the common diseases, the methodology used for recording the electrical activity, the ECG signal processing, and sleep physiology .

5.2.1 HEART PHYSIOLOGY

The heart is a muscular organ located just behind and to the left of breastbone. It consists of four chambers (i) the left atrium (LA), (ii) the right atrium (RA), (iii) the left ventricle (LV) and (iv) the right ventricle (RV). This is shown in figure 5.1.

The RA collects blood from the veins and passes it to the RV. The RV pumps the blood into the lungs, where blood is loaded with oxygen. This oxygen enriched blood is collected by the LA which pumps it into the LV. Then the LV supplies the oxygen-rich blood to the rest of the body. The coronary arteries run over the surface of the

heart and provide oxygen-riched blood to the heart muscle. Thus in normal condition, blood is circulated through the body by an ordered contraction and expansion of the chambers.

5.2.2 HEART ELECTRICAL CONDUCTION SYSTEM

Cardiac muscle is composed cardiomyocytes which generate action potentials during heart contraction. Cardiomyocytes are polarized with an electrical membrane potential of about -90 mv, at rest. A sudden increase in electrical potential of myocardial cells due to some external stimulus is called depolarization. This depolariztion is caused due to positively charged sodium ions entering into the cell. After depolarization, the muscle returns to its original electrical state, and the downward swing of the action potential is called repolarization. The repolarization is mainly caused by the movement of potassium ions out of the cell. The electrical activity during depolarization and repolarization produces a heart beat.

The conduction system of the heart described by [88] is shown in figure 5.2. The heart itself has a natural pacemaker that regulates the heart rate. It resides in the upper portion of the RA. It is a collection of electrical cells, commonly known as SINUS or SINO-ATRIAL node (SA). At normal condition, the SA generates a number of impulses which passes through the specialized electrical pathway and stimulates the muscle walls of the chambers to contract in a certain pattern. The rate of generation of impulses depends on the amount of adrenaline released, which is controlled by the autonomic nervous system (ANS). The atria are stimulated first, followed by a slight delay to empty them and then the ventricles are stimulated. The other members of the conduction system are atrioventricular node (AV), bundle of His, anterior and posterior bundles. The final components of conduction system are the Purkinje fibers, whose task is to conduct the wavefronts directly to the two ventricles so that they can contract simultaneously. The AV node acts as the electrical bridge that allows the impulse to go from the atria to the ventricles. The His-Purkinje network carries the impulses throughout the ventricles. The impulse then travels through the walls of the ventricles, causing them to contract. This forces blood out of the heart to the lungs and the body. The pulmonary veins empty oxygenated blood from the lungs to the left atrium.

The complete cycle of depolarization and repolarization of atria and ventricles constitute a heart beat. A normal heart beats in a constant rhythm of about 60 to 100 times per minute at rest. The atrial depolarization results P wave. Atrial repolarization and ventricular depolarization occurs simultaneously, which corresponds to the QRS complex. The ventricular repolarization phase is represented by the T wave. Although rare, but it is possible that a U-wave can be seen after the T-wave, which is may be generated by the mechanical electric feedback. The waveforms of two successive beats and some segments are shown in figure 5.3.

A properly functioning of conduction system guarantees an appropriate heart beat and sequential contractions and expansions of the atrial and ventricles keeps heart rate normal. Cardiac electrical dysfunction can be happened by any damage or improper

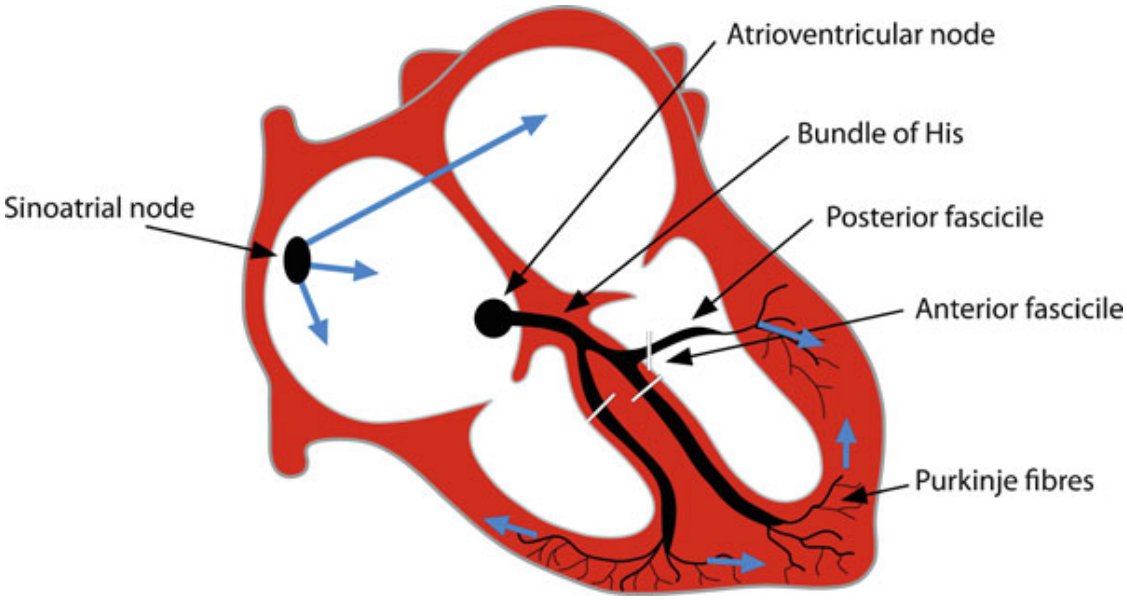


Figure 5.2: The electrical conduction system of heart

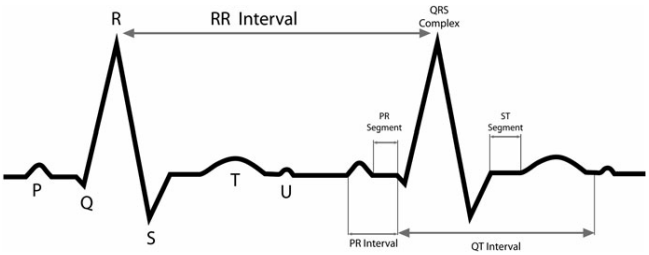


Figure 5.3: Two successive heart beats

functioning of any or combination of components. Other causes of cardiac arrhythmias can be a pathological stimulus generation or pathological conductive loop [88].

The interval between two successive R peaks is called the RR interval. The RR interval represents one cardiac cycle and is used to measure the heart rate (HR).

5.2.3 COMMON HEART DISEASES

The dysfunction of any or more of the components of the heart conduction system might cause different types of heart diseases. The most common heart diseases are:

- **Coronary artery disease**: Over the times, cholesterol plaques can narrow the arteries, supplying blood to the heart. The narrowed arteries are at higher risk for complete blockage from a sudden blood clot (this blockage is known as heart attack). It is the most common cause of death in United States of America (USA) and Europe [89]

- **Congestive heart failure**: The heart is either too stiff or too weak to effectively pump blood through the body. As a result, the heart cannot pump enough oxygen and nutrients to meet the body's needs. The CHF may be subdivided into

systolic and diastolic heart failures. There is a reduced cardiac contractility and an impaired cardiac relaxation during systolic and diastolic heart failures, respectively. The most common cause of heart failure is LV systolic dysfunction, which is symptomized by incresed heart rate, increased aldosterone level, endothelial dysfunction, and organ fibrosis. In diastolic dysfunction, the primary abnormality is impaired LV relaxation, causing high diastolic pressures and poor filling of the ventricle. Patients are often symptomatic with exertion when increased heart rate reduces LV filling time and circulating catecholamines worsen diastolic dysfunction [90].

- **Atrial fibrillation**: During AF, irregular heart beats are generated by abnormal electrical impulses in the atria. It is one of the most common arrhythmia. If someone suffers from AF, the electrical impulse does not travel in an orderly fashion through the atria. Instead, many impulses start simultaneously and spread through the atria and compete for a chance to travel through the AV node. The firing of these impulses causes a very rapid and disorganized heart beat. Atrial fibrillations are classified into (i) paroxysmal atrial fibrillation,(ii) persistent atrial fibrillation, and (iii) permanent atrial fibrillation.
  - **Paroxysmal atrial fibrillation (PAF)**: During paroxysmal atrial fibrillation, the faulty electrical signals and rapid heart rate begin suddenly and then stop on their own. They stop within about a week.
  - **Persistent atrial fibrillation (PeAF)**: In persistent atrial fibrillation, the abnormal heart rhythm lasts more than a week. It may stop on its own or can be stopped with treatment.
  - **Permanent atrial fibrillation**: It refers to the condition in which patient's normal heart rhythm cannot be restored with the usual treatments. Paroxysmal and Persistent Atrial fibrillations can occur more frequently and eventually become Permanent (or long standing persistent) AF.
- **Cardiac arrest**: The sudden loss of heart function is called cardiac arrest and it causes SCD. Certain ECG (we will discus more detail about ECG later) abnormalities can help identify patients at increased risk for sudden cardiac death. These include the presence of AV block or intraventricular conduction defects and QT prolongation, an increase in resting heart rate to >90 bpm, and increased QT dispersion in survivors of out-of-hospital cardiac arrest [87].

5.2.4 HEART RHYTHM DISORDERS

The SA also called natural pacemaker, located in the RA initiates an electrical discharge through AV node, which latter results a heart beat. The repetition of heart beats generates a rhythm. If the SA acts as a pacemaker, which produces heart rate in the normal range of 60-100 beats per minute, then the rhythm is called normal sinus rhythm. On the other hand, if the electrical discharge is made by some other sources instead of SA node, then it causes an extra (or missing) of a beat. This beat is referred to an ectopic

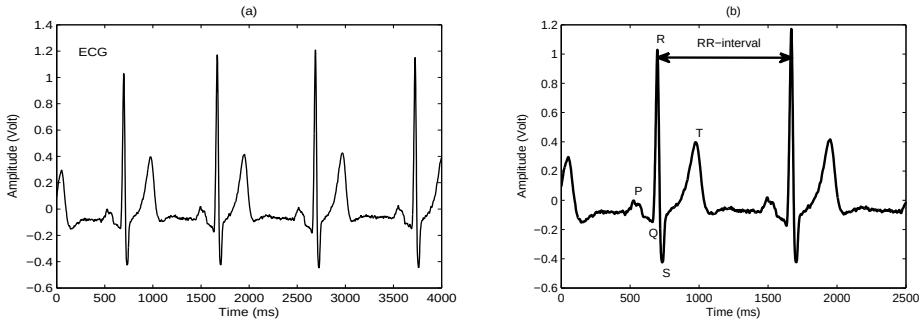


Figure 5.4: An example of sample ECG signal (a) and RR interval of two successive beats (b).

beat. The ectopic beats are irregular and they cause variations in the normally regular pulse.The normal rhythm of heart beats can be interrupted due to abnormalities of action of the stimuli generation system of the heart or with abnormal conduction of these stimuli [91, 88].

5.2.5 ECG SIGNAL PROCESSING

ECG is a method of recording electrical activity generated during depolarization and repolarization from the cells of the heart muscle against time. The ECG waveform is shown either on the computer screen or is printed on a graph paper. The ECG signal is concerned not only about the electro-physiological properties, but also concerned about the anatomical and mechanical properties of the heart. A continuous ECG signal is represented in terms of sequences of symbols. An ECG and RR interval i.e. the time gap between two successive R peaks are shown in figure 5.4. The series of RR intervals is commonly referred to as heart rate variability (HRV) series or RR series. To record the ECG, some small metal electrodes are placed on some specific parts of the body. The electrodes detect electrical impulses coming from different directions within the heart. There are normal patterns for each electrode. Abnormal patterns may produce due to some heart disorders. Careful assessment of the ECG provides lots of useful diagnostic information about heart functioning.

One characteristic feature of ECG signal is the cyclic occurrence of its components consisting of P-QRS-T complex. During ECG signal processing and analysis, an important task is to detect each wave form from the P-QRS-T complex and finding of the so called fiducial points [88, 92]. The most important task in ECG signal processing is the detection of R peaks. In 1985, Pan and Tompkins [93] proposed a real time QRS detection algorithm, which detects correctly about 99.3% of the QRS complexes.

5.2.6 ECG BEAT ANNOTATION

Annotation indicates the time of occurrence and the type of each individual heart beat. For example, many of the recordings that contain ECG signals have annotations that

Table 5.1: Some beat annotations used by Physio bank databases

Code	Description
N	Normal Beat
L	Left bundle branch block beat
R	Right bundle branch block beat
A	Atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature or ectopic beat (atrial or nodal)

indicate the times of occurrence and types of each individual heart beat. The examples of some beat annotations used by Physio bank databases are shown in table 5.1.

5.2.7 HEART RATE VARIABILITY

The heart behavior is not stable with time, instead, there exists a variation in the time interval between consecutive heart beats. The normal heart rhythm is controlled by the SA node, which is modulated by innervation from both the sympathetic and the vagal balances of the ANS. The SA node is the final responsible, for generating heart beats. The ANS controlling functions of the heart is divided into vagal (parasympathetic) and sympathetic systems, and they work in opposite directions. The activity from the sympathetic system increases the heart rate, whereas the vagal activity causes the heart rate to slow down. In rest condition, there is a balance state between these systems, that is responsible for the variability in the consecutive heart beats intervals. At the same time, the ANS is influenced by many other systems (i.e. central nervous system, respiratory system, renin-angiotensin system, vasomotor system), which is also responsible to modulate the heart rate. HRV is simply the variation in the consecutive heart beats intervals, or in other words, the variations between consecutive instantaneous heart rates. Under certain conditions like people taking medications, drug, alcohol, or suffering from some diseases such as AF, infarct of myocardium, and kidney failures, adjustment of heart rhythm is very difficult and HRV is significantly reduced [70, 94, 88]. Heart rate variability analysis [70, 94] plays an important role in cardiac (heart) rhythm disturbance analysis.

5.2.8 SLEEP PHYSIOLOGY AND PHYSIOLOGICAL CHANGES DURING SLEEP

Human spends about one-third of their lives asleep, yet most individuals know little about sleep. In the last years, it became evident, in the scientific community, that sleep has a large effect on many physiological functions and may play a fundamental role in the genesis and insurgence of different pathologies: cardiologic, neurological, metabolic, etc.

During the night, human passes through different sleep stages. Rules and guidelines provided by the American Academy of Sleep Medicine (AASM) allow the evaluation

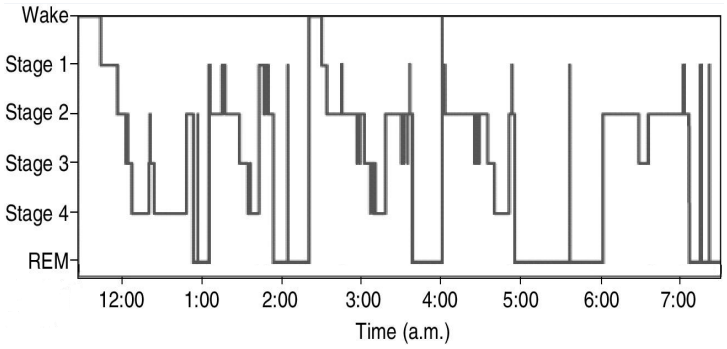


Figure 5.5: Different sleep states that a young adult experienced from 12.00 to 7.00 am

of the macrostructure (alternation of different sleep stages: REM, NREM (LS; DS or slow wave sleep (SWS): stage 3), wakefulness [95, 82, 68]. The different sleep states across a single night in young adult is shown in figure 5.5 (This is a modified version of figure 2-1 in [96]).

Previously NREM sleep was divided into four different stages by [97]. As no physiological or clinical difference exists between NREM stages 3 and 4, they were combined by the American Academy of Sleep Medicine (AASM) commission in 2007 [98] into a single stage (stage 3).

A sleep episode starts with a short period of NREM stage 1, and progresses through stage 2, 3, 4 and finally to REM. The switching of sleep states bteween NREM and REM happens cyclically throughout the night. NREM constitutes about 75 to 80 percent, and REM constitutes the remaining 20 to 25 percent of the total period of sleep [96].

The state of sleep is characterized by some changes in the brain wave activity, heart rate, body temperature, breathing, and other physiological functions. Different physiological functions may be more active and variable (or less active and more stable), depending on the stage of sleep. Many physiological variables are controlled during wakefulness at levels optimal for the bodys functioning. During wakefulness, our blood pressure, body temperature, levels of oxygen, carbon dioxide, and glucose in the blood remain quite constant. However, during sleep, physiological demands are reduced and causes a drop in temperature and blood pressure. In general, many of our physiological functions such as brain wave activity, heart rate, and breathing, are quite variable when we are awake or during REM sleep, but are highly regular when we are in NREM sleep.

5.2.9 BLOOD PRESSURE

Blood pressure also referred to as arterial pressure is the force exerted by circulating blood against the walls of arteries. The blood pressure in the circulation is principally due to the pumping action of the heart [99]. During each heartbeat, blood pressure varies between a maximum (systolic) and a minimum (diastolic) pressure during each heart beat. Systolic is the highest level of pressure, and it happens only when the heart beats. Diastolic is the minimum value of blood pressure observed, when the heart relaxes between the beats. The unit of blood pressure is millimeters of mercury (mm

Hg). Ideally, the normal range of blood pressure for adults is 120 (systolic) and 80 (diastolic) (mm Hg). It is stated as 120 over 80. The beat-to-beat systolic pressure series is termed as blood pressure series, and is commonly obtained by searching for a local maximum in the blood pressure signal following each R-wave.

5.3 HRV REGULARITY ANALYSIS DURING PERSISTENT AF

A number of standard parameters have been recommended in [70] for HRV analysis. They are commonly subdivided into (i) time domain, (ii) frequency domain. Time domain measures consist of mean (Mean_{RR}) and standard deviation (STD) of the RR series, RMSSD (the square root of the mean of sum of square of differences between adjacent normal RR intervals), etc. On the other hand, frequency domain parameters include spectral metrics *i.e* VLF (power in very low frequency : 0.003 to 0.04 Hz), LF (power in low frequency:0.04 to 0.15 Hz), and HF (power in high frequency: 0.15 to 0.4 Hz). Time domain parameters are computationally simple, but are less effective in discriminating between parasympathetic and sympathetic contributions to HRV. On the other hand, spectral parameters can be benefited from robust estimates based on parametric approach. In fact AR models have been employed for maximum entropy spectral estimation since 1975.

The development of nonlinear techniques has paved the way of characterizing biological signals. ApEn and SampEn have been proved effective in discriminating the terminating and nonterminating of PAF. With respect to the PeAF, Ulimoen et al. [100] investigated the effect of four rate control drugs [1] on heart rate (HR) and arrhythmia related symptoms with permanent AF. Controlling the ventricular rate during both paroxysmal and persistent atrial fibrillation (AF), initially with digitalis preparations and subsequently with β-blockers and calcium channel blockers [101], has been a mainstay for the management of this arrhythmia for many years. In the absence of preexcitation, the AV node is the only electric pathway existing for transmission of rapid fibrillatory activity from the atrium to the ventricles. Drugs that prolong AV nodal refractoriness include β-adrenergic receptor blockers (Carvedilol), nondihydropyridine calcium channel blockers, and digitalis glycosides. Ulimoen et al. [100] reported that all four drugs reduced the mean HR. Here, we will investigate if a rate control drug (Carvedilol) can also alter the value of SampEn_{TH}, and hence if SampEn_{TH} can distinguish two groups: the baseline group, B (without drug administered) and the group C (with Carvedilol administered).

5.3.1 DATA

The analysis was performed on a sub-population (a subset of 20 randomly selected subjects) of the rate control in atrial fibrillation (RATAF) database [100]. No subject was suffering in from ischemic heart disease, systolic heart failure, paroxysmal or persistent

1 Drugs used for controlling the ventricular rate during both paroxysmal and persistent atrial fibrillation

atrial fibrillations for less than 3 months. The considered RR series are extracted from 20-min Holter ECG segments starting at 2 PM.

5.3.2 PARAMETERS ESTIMATION

Only the normal RR intervals of each subject are considered. The RR series is preprocessed as described in the "Preprocessing" part of section 4.3. After preprocessing, a set of linear and nonlinear parameters are extracted from each RR series. The set of linear parameters consists of spectral parameters: total power (TP), normalized power in low frequency band LF) (*i.e* normalized by the total power), normalized power in high frequency band HF_{Norm}, time domain parameters: mean of RR series ($Mean_{RR}$), and $SampEn_{TH}$. The only nonlinear parameter $SampEn_{RR}$ is derived also along with the linear parameters. The $SampEn_{TH}$ and spectral parameters are obtained from the same model fitted on the RR series. In case of fitting the AR model, the model order is determined by satisfying the Akaike information criterion (AIC) and Anderson's whiteness test. The estimated value of each parameter is compared without and with the drug administered.

The estimation of $SampEn_{RR}$ is affected by different choices of the embedding dimension, m and tolerance of mismatch r (the maximum difference between the corresponding elements of the templates of length m constructed from the given series of length N). There is no universal indication about the choice of values for m and r. The values of $m = \{1,2\}$ and $r = \{0.1,0.15,0.2,025\} \times STD$ are mostly used in the literature for SampEn estimation. In this experimental study, both $m = \{1,2\}$ and $r = \{0.15,0.2,025\} \times STD$ have been used. The RR series of different lengths $N == \{75,150,300,600,1000\}$ besides the entire series are considered here.

5.3.3 STATISTICAL ANALYSIS

A nonparametric Wilcoxon signed rank [102] has been used to evaluate the differences between the two group "B" and "C". The value of $p < 0.05$ is considered significant.

5.3.4 RESULTS ON HRV REGULARITY ANALYSIS

In this section, the experimental results on parametric assessment of SampEn to characterize the effects of β-blocker ("Carvedilol") on HRV regularity for patients suffering from persistent atrial fibrillation are presented. RR series of different lengths $N=\{75,150,300,600,1000\}$ besides the entire series are considered from both (baseline group:B, and with Carvedilol group: C). SampEn values of each series are estimated with $m=\{1,2\}$ and $r=\{0.15,0.2,0.25\} \times STD$. Each of the time and frequency domain linear parameters ($Mean_{RR}$, TP, LF, HF) is significantly different between the two groups ($p < 0.05$), at any series length. Both $SampEn_{RR}$ and $SampEn_{TH}$ are also different for $N \geqslant 300$, and for nearly any considered values of m and r. A few results (median $\pm$IQR) are summarized for entire series in table 5.2. Box plots for $N = 300$ are reported in Fig. 5.6.

The significant (p<0.05) differences in both numerical and theoretical values of SampEn between the two groups infer that nonlinearities or non-Gaussianities might be altered due to drug administration. Changes in stationarity are less likely given that differences hold even with small N. However, even if it corresponds to the expected value of an entropy rate, SampEn_{TH} is nevertheless based on the same information on which linear spectral parameters are (*i.e.* a few sample autocorrelation coefficients derived from the original series). This is in line with the fact that also spectral parameters are modified by the drug.

The differences between the numerical estimation (SampEn_{RR}) and theoretical computation (SampEn_{TH}) are further investigated by two steps sub-analysis using surrogate data. At first, 2000 synthetic series are generated for each RR sequence of length N=300, fitted on the AR models. The SampEn of the sequence is computed numerically as well as the Monte Carlo probability distribution is considered for the synthetic series. The number of cases numerical estimation is within the 95% standard range of these distributions for both groups are reported in table 5.3.a.

The number of agreements after Carvedilol is always more than that at baseline group. Moreover, the SampEn values of 8-subjects (out of 20) lie outside the standard range in both groups with m=1 and r=0.2. The group of these 8-subjects is referred as SUB-8 in the following. The SampEn values for SUB-8 population are significantly

Table 5.2: Different parameters for entire series (single-tail Wilcoxon test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$)

Parameters	**Groups**	
	B	**C**
Mean_{RR}*** (ms)	538 ± 104	742 ± 181
TP*** (s^2)	0.014 ± 0.011	0.024 ± 0.025
LF** (%)	11.43 ± 5.32	15.12 ± 5.08
HF*** (%)	23.944 ± 6.92	36.14 ± 8.7
$\text{SampEn}_{RR}(1, 0.15)$**	2.201 ± 0.24	2.353 ± 0.15
$\text{SampEn}_{RR}(1, 0.2)$*	1.952 ± 0.22	2.058 ± 0.16
$\text{SampEn}_{RR}(1, 0.25)$**	1.725 ± 0.19	1.846 ± 0.17
$\text{SampEn}_{RR}(2, 0.15)$**	2.183 ± 0.31	2.328 ± 0.15
$\text{SampEn}_{RR}(2, 0.2)$**	1.927 ± 0.21	2.061 ± 0.16
$\text{SampEn}_{RR}(2, 0.25)$**	1.694 ± 0.19	1.851 ± 0.17
$\text{SampEn}_{TH}(1, 0.15)$**	2.461 ± 0.05	2.469 ± 0.01
$\text{SampEn}_{TH}(1, 0.2)$**	2.175 ± 0.05	2.182 ± 0.01
$\text{SampEn}_{TH}(1, 0.25)$**	1.954 ± 0.04	1.961 ± 0.01
$\text{SampEn}_{TH}(2, 0.15)$**	2.456 ± 0.07	2.466 ± 0.01
$\text{SampEn}_{TH}(2, 0.2)$**	2.170 ± 0.07	2.180 ± 0.01
$\text{SampEn}_{TH}(2, 0.25)$**	1.949 ± 0.07	1.958 ± 0.01

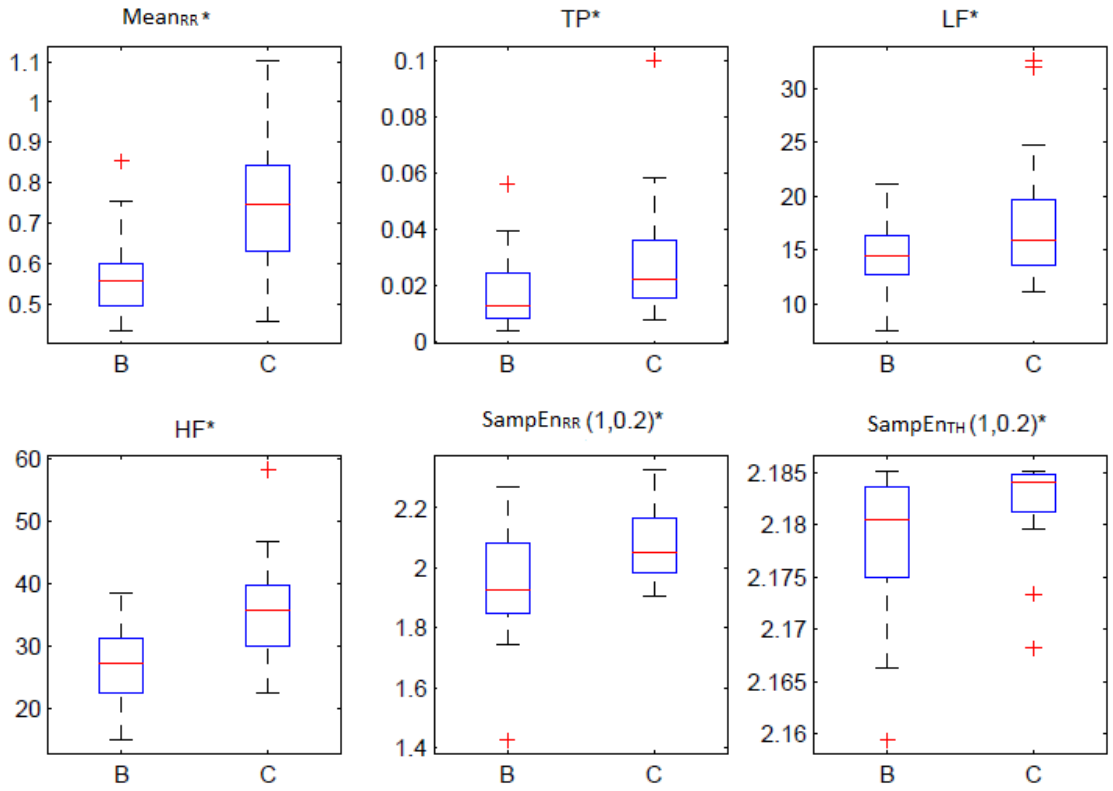


Figure 5.6: Boxplots of the parameters for $N = 300$. SampEn values are computed with $m = 1$ and $r = 0.2 \times STD$.

Table 5.3: # of cases inside the 95% standard ranges

a) AR series	**r**		
m	**0.15**	**0.2**	**0.25**
1	B:7; C:13	B:5; C:10	B:6; C:11
2	B:11; C:17	B8; C18	B:6; C:18

b) IAAFT	**r**		
m	**0.15**	**0.2**	**0.25**
1	B:16; C:19	B:13; C:18	B:19; C:20
2	B:18; C:19	B:16; C:20	B:17; C:20

different (p=0.0391) between the two groups. This procedure is repeated for IAAFT[2] surrogates, instead of the synthetic series generated through the AR process, and the number of agreements is mentioned in table 5.3.b.

2 An amplitude adjusted Fourier transform (AAFT) based surrogate method, where the power spectrum obtained from AAFT is adjusted back to that of the original series by performing a number of iterations. With each iteration the alteration of the power spectrum when rescaling is performed will therefore be smaller than in the previous iteration [103]

5.3.5 EVALUATION ON HRV REGULARITY ANALYSIS

The number of agreements between SE_{RR} and what expected from an AR process (table 5.3.a) is increased after drugs. The same is found when considering IAAFT surrogates (table 5.3.b), which share with the original series the power spectrum and a possible nonlinear static transformation of the samples. This analysis suggests that the drug might reduce, in particular, non-Gaussianity of the RR series. A relatively smaller increase in nonlinear regularity is also supported by the finding that SE_{RR} is different in the SUB-8 population after Carvedilol.

5.4 NONLINEAR REGULARITY ANALYSIS OF ABP VARIABILITY IN PATIENTS WITH AF

It is almost unknown if the irregularity of ventricular response might directly affect arterial blood pressure (ABP) variability during atrial fibrillation. Pitzalis et al. [104] observed a respiratory related high-frequency component of systolic arterial pressure (SAP) variability during AF in absence of a respiratory sinus arrhythmia. More recently, Mainardi et al. [105] have observed a low frequency component of arterial pressure variability during AF, independently from the presence of a corresponding component in RR variability and, very recently, Corino et al. [106] have reported that the low frequency component of SAP variability in patients with AF is increased after tilt test. All of these results are interpreted as indirect evidences for a possible instrumental role of oscillatory components of sympathetic discharge in determining the low frequency oscillations of SAP and diastolic arterial pressure (DAP).

The above-mentioned results are obtained by analyzing ABP variability with traditional linear methods, thus with a limited capability of collecting information on the dynamic patterns used by the cardiovascular regulation systems to adjust heart rate with blood pressure. Nonlinear methods of signal analysis can be useful for characterizing complex dynamics. Nonlinear analysis of heart rate has been widely employed during sinus rhythm [107, 108, 109], and during AF [110, 111] to some extent, providing information related to the irregularity of the series in terms of pattern repetition and their dynamics. On the contrary, a few studies have analyzed irregularity of blood pressure variability in patients during AF [112] or normal sinus rhythm [113, 114].

The purpose of this experimental study is to assess the effects of sympathetic activation induced by tilt on the patterns of blood pressure irregularity in patients with AF: *i.e.* in a physiological model in which the coupling between cardiac cycle duration and pulse pressure is regulated independently of functioning baroreflex control mechanisms for the lack of regularity in RR intervals dynamics. In other words, it is verified that if the effects of sympathetic stimulation acting on blood pressure control can also be observed in patients with AF.

Table 5.4: Demographic characteristics and cardiovascular history in the entire study population and in the two subgroups (group A: patients whose systolic arterial pressure increased during tilt, group B: patients whose systolic arterial pressure did not increase during tilt)

Variable	All patients	GroupA	Group B
Number	20	11	9
Gender (male/female)	15/5	8/3	(7/2)
Age (years)	62± 14	59± 14	65± 14
AF duration (months)	3±4 (2-9)	3±4 (2-9)	3±4 (2-9)
Previous AF	11	5	6
Left atrium diameter (mm)	46±7	47±5	45±8
Ejectin fraction (%)	57±8	54±10	60±5
Diabetes	2	0	2
Hypertension	12	5	7
β-blockers	11	7	4
Flecainide	3	1	2
Cordarone	5	3	2
ACE-inibitori	13	6	7
Ca-antagonist	3	2	1

5.4.1 DATA

In this experimental method, we have considered 20 patients (age: 62±14 years with 75% male) admitted to the hospital for programmed electrical cardioversion of persistent AF according to the international guideline [115] (i.e. an AF episode lasting longer than 7 days and requiring termination by electrical cardioversion). The mean duration of arrhythmia was 3±4 months (2-9 range). Detail characteristics of the patients are mentioned in table 5.4.

Three orthogonal leads, a periodic reference arterial pressure measurement, continuous beat-to-beat non-invasive recordings of arterial pressure, and the respiratory signal were obtained with a Task Force Monitor (CNSystem; Austria) recording system. Surface ECG and blood pressure signals were acquired at rest, and during a passive orthostatic stimulus (head-up tilt test, 75° tilting). Both phases lasted about ten minutes. The sampling frequency was 1 kHz for the ECG signal and 100 Hz for continuous arterial pressure recording. Raw data were exported as ASCII text files for off-line analysis.

5.4.2 BLOOD PRESSURE SERIES EXTRACTION

The beat-to-beat systolic pressure series is usually obtained by searching for a local maximum in the blood pressure signal following each R-wave during normal sinus rhythm. This approach does not work appropriately during AF [105]. In fact, to generate regular pulses in arterial pressure, R waves may not be coupled with an adequate

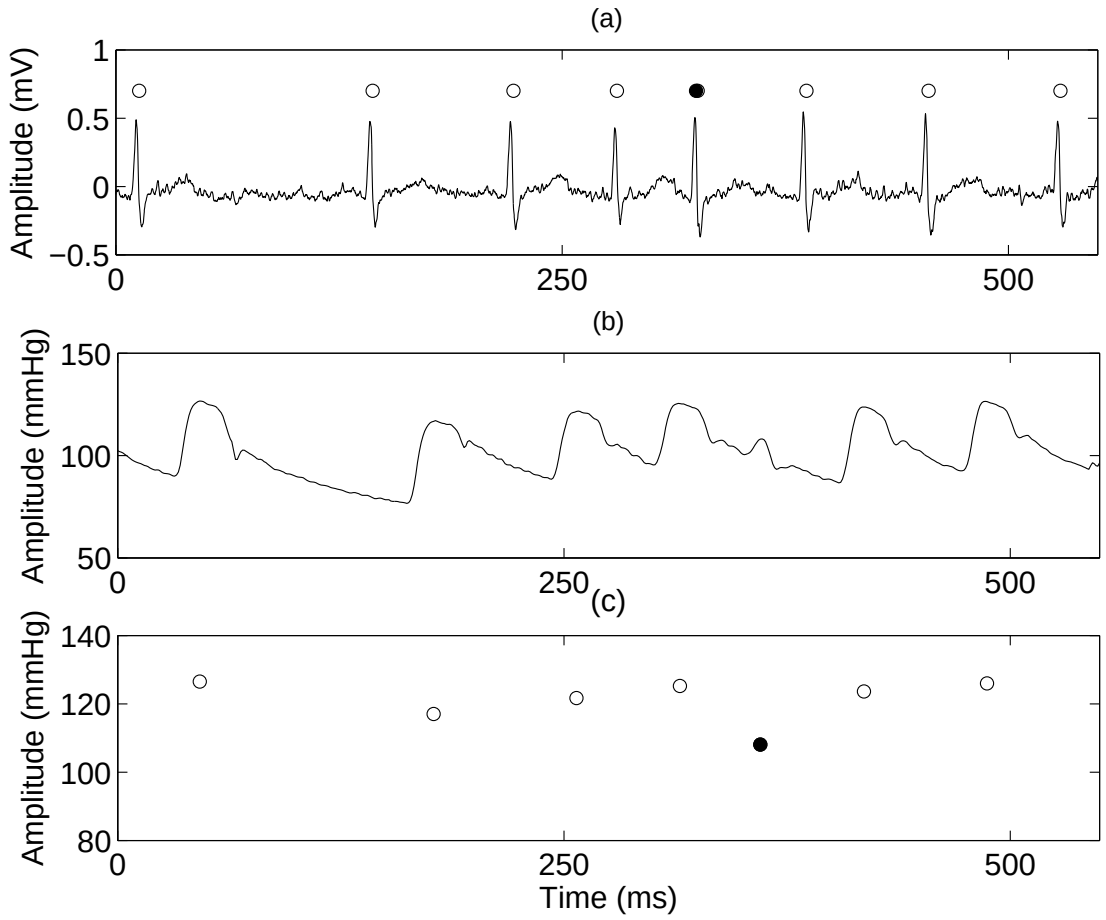


Figure 5.7: Panel (a) ECG signal and (b) blood pressure signal of a patient during AF. The circles in (a) correspond to the detected QRS, being the filled circle a beat which is not followed by a pressure pulse. (c) The systolic arterial pressure series obtained without preprocessing: the filled circle identifies a drop in systolic value due to an insufficient pressure pulse.

left ventricular output (i.e. the left ventricle can be only partially filled when an atrial impulse propagates through the AV node triggering the contraction). Thus the QRS complexes are not necessarily followed by an arterial pressure pulse of regular amplitude as shown in figure 5.7. For this reason, to measure the beat-to-beat systolic pressure values, a method that coarsely localizes arterial pressure systolic peaks has been applied, and then refines their positions, thus obtaining the systolic values not relying on the information about QRS location. An interactive graphic interface allowed the operator to visually identify and correct misdetected arterial pressure pulse events. We have also extracted and analyzed DAP series, whose values are defined as the local minimum preceding all systolic values.

As series length influences the following analysis, we have considered series of 300-samples for all patients and all phases, being 300 the length of the shortest available series of sufficient quality during tilt (which is slightly more than 3 minutes). In particular, the last 300 points of rest are selected, while discarding the first 25 points on tilt, to avoid the initial drift, and then 300 points are selected thereafter.

5.4.3 PARAMETERS ESTIMATION

- "ApEn and SampEn numerical estimations". $ApEn_{RR}$ and $SampEn_{RR}$ have been numerically estimated for each series as described in sections 2.4.1.3 and 2.4.1.4. Since ApEn requires longer series to converge than SampEn, we have used the values of classical parameters r=0.2×STD and m=1 for $ApEn_{RR}$ and 2 for $SampEn_{RR}$ estimations.
- "ApEn and SampEn parametric estimations". Besides numerical estimations, the values of $ApEn_{\mu}$ and $SampEn_{\mu}$ have been estimated for the synthetic signals generated through $K = 1000$ realizations of the AR models, fitted to the series with the same values of parameters m and r, used for numerical estimation.

5.4.4 STATISTICAL ANALYSIS

The data have been presented as mean values±STD. A paired t-test or the Wilcoxon signed rank test [102] has been used to evaluate the differences between parameters obtained during rest and tilt. An unpaired t-test or Wilcoxon-Mann-Whitney [102] test has been used to evaluate the differences between groups A and B. The value of $p < 0.05$ is considered significant.

5.4.5 RESULTS ON NONLINEAR REGULARITY ANALYSIS OF ABP VARIABILITY

In this section, the results obtained from the analysis of nonlinear regularity of arterial blood pressure in patients with AF during tilt-test procedure are explained.

5.4.5.1 ENTIRE POPULATION

Both $ApEn_{RR}$ and $SampEn_{RR}$ are significantly higher during tilt for SAP ($ApEn_{RR}$: 1.73±0.22 vs. 1.81±0.20, $p < 0.05$, rest vs. tilt; $SampEn_{RR}$: 1.68±0.31 vs. 1.84±0.30, $p < 0.05$, rest vs. tilt). On the contrary, no differences are observed in entropy values when comparing rest vs. tilt for DAP series. No significant changes are observed when comparing $ApEn_{\mu}$ and $SampEn_{\mu}$ during rest and tilt phases ($ApEn_{\mu}$: SAP: 1.85±0.21 vs. 1.89±0.18, ns; DAP: 1.99±0.06 vs. 1.98±0.13, ns; rest vs. tilt. $SampEn_{\mu}$: SAP: 1.91±0.26 vs. 1.97±0.24, p-value ns; DAP: 2.07±0.10 vs. 2.04±0.17, p-value ns; rest vs. tilt.). The percentages of agreement between $ApEn_{RR}$ and $SampEn_{RR}$ computed on surrogate data using IAAFT surrogate showed almost no changes when moving from rest to tilt.

5.4.5.2 ARTERIAL PRESSURE RESPONSE TO TILT

Two different patterns of SAP alteration are observed due to tilt. The first group (group A, 11 patients) is composed of patients, whose systolic pressure increased more than 5 mmHg during tilt. In these patients the systolic pressure is increased on average 12±7 mmHg (range 5-26 mmHg). In the remaining 9 patients (group B) the average value of SAP is remained almost unchanged or it even decreased (110±18 vs. 107±19 mmHg,

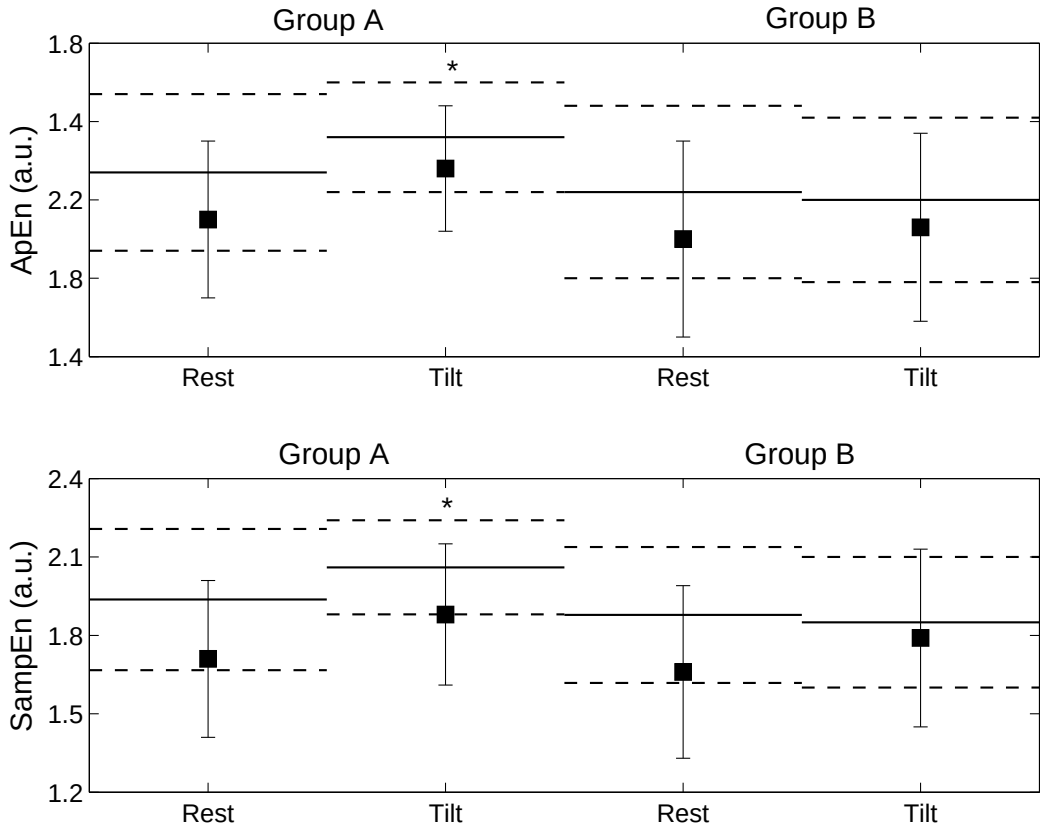


Figure 5.8: Errorbar of $ApEn_{RR}$ (top panel) and $SampEn_{RR}$ (bottom panel) during rest and tilt phases for the two subgroups of patients. Mean (solid line)± standard deviation (dashed lines) of $ApEn_{\mu}$ and $SampEn_{\mu}$ are superimposed. Group A: patients whose systolic arterial pressure increased during tilt, group B: patients whose systolic arterial pressure did not increase during tilt. * $p < 0.05$.

p-value ns). Therefore, the two groups (A and B) are further analyzed, separately. The patients of group A has an increased standard deviation of SAP and DAP series during tilt, whereas no differences are seen in patients of group B.

Figure 5.8 shows the results of entropy values for SAP series in the two subgroups (no differences are found in DAP series, thus data are not shown). In group A, we observed a significant increase in $SampEn_{RR}$ and $ApEn_{RR}$ of SAP series during tilt, together with an augmented $ApEn_{\mu}$ and $SampEn_{\mu}$. On the contrary, no significant differences were found in group B, neither for $ApEn_{RR}$ and $SampEn_{RR}$, nor for $ApEn_{\mu}$ or $SampEn_{\mu}$.

The percentages of agreement between real and synthetic values for ApEn and SampEn are shown in table 5.5. Group A did not display an evident change during tilt and in half of the cases the observed dynamics were consistent with a purely linear process. On the contrary, group B showed a definite increase in agreement, thus suggesting that the series dynamics were well described by a linear process.

Table 5.5: Percentages of agreement between real and synthetic values of ApEn and SampEn for the two subgroups (group A: patients whose systolic arterial pressure increased during tilt, group B: patients whose systolic arterial pressure did not increase during tilt).

	Group A		**Group B**	
	REST	TILT	REST	TILT
ApEn	55%	64%	56%	100%
SampEn	64%	55%	78%	89%

5.4.6 EVALUATION ON NONLINEAR REGULARITY ANALYSIS OF ABP VARIABILITY

The first important finding of this experimental study is that not all patients with AF experienced a similar increase of systolic pressure during tilt: 9 out of 20 patients have blood pressure values that remained almost unchanged or even it sharply decreased. Therefore, the study population has been divided into two groups, depending on the increase (group A)/invariance (group B) of the systolic pressure during tilt. No substantial clinical difference are found between the two groups. However, some consistent tendencies are observed: patients of group A are on average younger (59 ± 14 vs. 65 ± 14, p-value ns; group A vs. group B) than those of group B, they have a slightly lower ejection fraction (54 ± 10 vs. 60 ± 5, p-value ns; group A vs. group B) and smaller mean RR at rest and tilt (rest: 749 ± 140 vs. 800 ± 176 ms, p-value ns; tilt 683 ± 126 vs. 726 ± 149 ms, p-value ns; group A vs. group B).

Frequency domain analysis is the most commonly used technique for assessing the autonomic response of heart rate and blood pressure to head-up tilt test. However, nonlinear dynamics of cardiovascular response can also be assessed in order to evaluate regularity and synchronization among cardiovascular beat-to-beat variability signals during the sympathetic activation induced by head-up tilt.

The second finding of our experimental study is that both measures of irregularity ($ApEn_{RR}$ and $SampEn_{RR}$) are significantly higher during tilt for SAP series, especially when the two subgroups are separately considered. It can be hypothesized that in patients of Group A, the vascular regulatory mechanisms is still efficient (i.e. the response to the autonomic stimulus is similar to what observed in subjects in normal sinus rhythm), in spite of the presence of persistent AF. On the contrary, patients of group B are seemed to have lost their vascular capability of a physiological response to sympathetic stimulation.

5.5 FEATURE EXTRACTION FROM HRV FOR CLASSIFICATION OF SLEEP STAGES

Sleep has a large effect on many physiological functions, and quality of sleep is one of the aspects that mostly influence our everyday life. In fact, it has a strong impact on some important human functioning like memorization, learning and concentration [116]. Poor sleep quality or too short sleep time have been identified among the

main causes of car accidents [117]. Furthermore, disturbances in sound sleep have a strong association with cardiovascular pathologies. Also, a bad quality of sleep has an impact on blood pressure, decreases the immunity defenses and may increase the insurgence probability of metabolic disturbances such as obesity and diabetes [118, 119, 120, 121, 122]. Sleep quality is generally evaluated through polysomnography (PSG), which consists of many physiological signals recorded during one or more nights of sleep: electroencephalogram (EEG), electro-myogram (EMG), and electro-oculogram (EOG), besides respiration activity and ECG.

The different stages of sleep (described in section 5.2.8) are associated with different brain activity and hence cardiovascular activity as well. During stage 1 sleep, people suffer from drowsiness. They drips in and out of the sleep. Stage 1 is considered as a transition between wakefulness (WAKE) and sleep (SLEEP). During this stage, people are still relatively awake and alert. This stage of sleep lasts only for a very few (1 to 5 minutes) of sleep. Stage 2 is the longest lasting period (about 20 minutes) of NREM. During the 2^{nd} stage (stage 2) of sleep, body temperature starts to decrease and heart rate begins to slow. Stage 3 is also referred to as the slow wave sweep because the slow brain waves are seen during this stage.

The standard practice for sleep evaluation is the visual or semi-automatic scoring of polysomnographic traces [123]. This technique requires special instrumentation, and signals which are recorded and scored by trained personnel. In addition, their acquisition may be disturbing to affect the sleep quality itself.

Sleep can strongly affect the peripheral system, particularly the autonomic nervous system, so that the HRV signal presents different patterns during different sleep stages [124, 125, 126] and during sleep phasic events [127, 128]. One of the advantages of using HRV for sleep evaluation is the possibility of employing less intrusive devices, such as a sensorized T-shirt[3] or mattress [129, 130, 131]. For these reasons, many recent studies have focused on the effects of sleep stage transitions on peripheral systems. Most of the works found in the literature have given emphasis on the correspondence of different patterns of heart rate with different sleep stages [132] and more recently a few works described methods to perform sleep staging through HRV analysis.

The quantitative spectral analysis of HRV to evaluate the changes in autonomic influences with sleep stages in adults reveals that the stage differences are most evident using spectral parameters than time domain parameters [133]. In the low frequency band, the REM and WAKE are not significantly different. However, REM and NREM are significantly distinguishable in both low and mid frequency bands. On the other hand, power content in the very high frequency band is insignificant to distinguish any stages of sleep. The modulus and phase of the pole in high frequency band are one of the significant features for distinguishing NREM from REM sleep [134]. To the best of our knowledge, some existing methods [135, 136, 137] with high accuracy reported in

3 An intelligent T-shirt, integrated with sensors and electrodes able to record different physiological signals. The T-shirt is comfortable and made completely in clothes. The patient is asked to wear the T-shirt before going to the bed. The sensorized T-shirt allows for acquisition of both physiological parameters, such as those given by the ECG and breathing frequency, and bio-mechanic parameters such as movement and posture.

the literature used a large set of features. The use of these features make the system computationally expensive .

A set of new features, those reflect the changes in regularity of the RR series among the different sleep stages along with the existing ones are used in this study for automatic classification of sleep stages. During stage 1 of NREM, people drips in and out of sleep. Sleep stage 1 is excluded from the NREM sleep. This exclusion is useful for reducing the ambiguity between WAKE and NREM [130]. The features and selection of relevant features for classifications between and SLEEP, and between REM and NREM stages will be explained in this experimental study. The classification performances are evaluated on the basis of the capability to discriminate between WAKE and SLEEP, and between NREM and REM.

5.5.1 DATA

Full PSG of 20 patients with suspected sleep-disordered breathing, recorded for one night each at the Sleep Center of Tampere University Hospital, Finland. The Ethical Committee of the Pirkanmaa Hospital District has approved the study and all the subjects have given their consent to be included into the study. The age of the subjects are between 49 and 68 years; the BMI varied between 21.8 and 40.6; 13 patients are females. The patients were suffering from a variety of sleep disorders, including either different degrees of nocturnal apnea/hypopnea and/or insomnia.

The inter-beats (RR) series used in this experimental setup, are obtained from the ECG recordings as well as the sleep scoring automatically derived from the complete PSG recordings (mainly using the EEG traces) through the Somnologica software; the scoring is based on 30-second epochs.

The automatic sleep stages classification consists of the following methodological steps:

5.5.2 PREPROCESSING

The artifacts from the RR series are removed by the procedure, explained in the *preprocessing* part of section 4.3. After removing artifacts, only those RR segments of different lengths (2, 6, and 10 epochs pertaining to a common sleep stage, where each epoch corresponds to 30 seconds of recordings) are considered as long as at least 20% of the selected beats were previously marked as normal.

5.5.3 FEATURE EXTRACTION FROM RR SERIES

Feature extraction is a crucial and sensitive step for any classification problem. Nonlinear regularity features, besides those proven effective are considered here. A set of time-domain, frequency-domain, detrended fluctuation analysis, as well as regularity features are extracted from the RR segments. The features are briefly explained by the following steps:

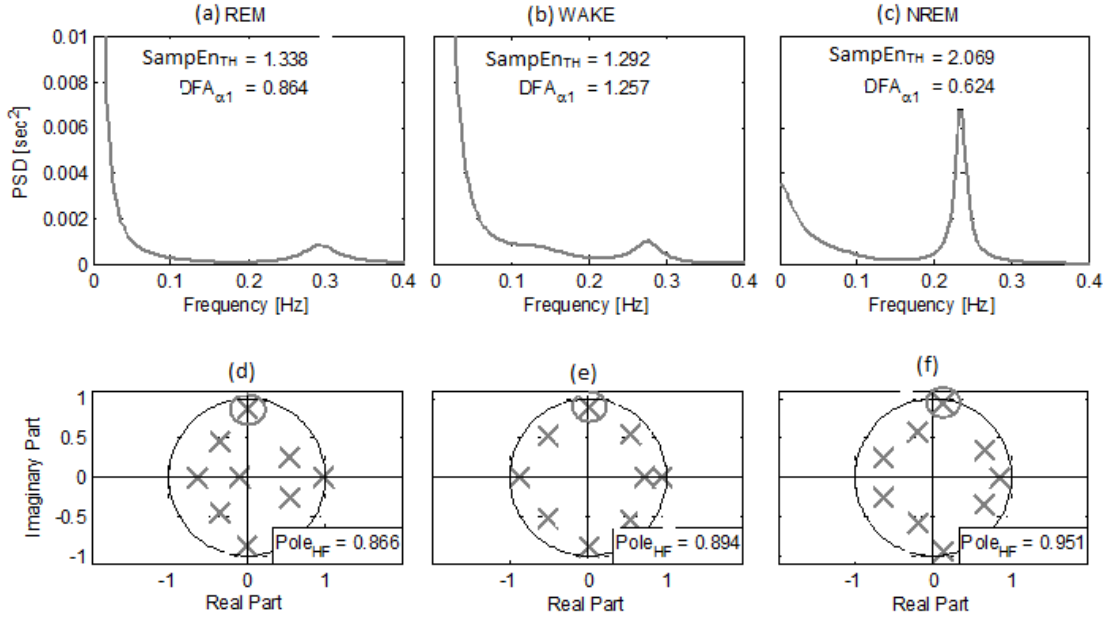


Figure 5.9: Power spectra of the RR series panels (a), (b), (c) and the positions of the poles of an AR model fitted to the series panels (d), (e) and (f), respectively during the three stages of sleep. The parameter Pole_{HF} has been marked with a circle).

- "Time-domain". The normalized mean (Mean_{RR}) and standard deviation (STD) of each RR series are extracted. The mean and standard deviation of each RR segment are normalized, respectively by the mean and standard deviation of the entire series.

- "Frequency-domain". RR segment is fitted to an AR model of fixed order (9). The Andersons test [72], that checks the whiteness of the prediction error is satisfied by more than 95% cases using this fixed order. Then normalized powers (normalized by the total power) in three frequency bands: VLF from 0.003 to 0.04 Hz, LF from 0.04 to 0.15 Hz, and HF from 0.15 to 0.4 Hz are extracted using a spectral decomposition technique described in [138]. Besides these, the ratio of LF to HF and the modulus of pole (Pole_{HF}) in the high frequency band with the largest residual are also considered. The modulus of pole is strongly related to the respiratory frequency [134]. The spectral components of heart rate during different sleep stages are shown in figure 5.9.

- "Regularity features". In the proposed system, we have considered three measures of SampEn *i.e.* SampEn_{RR}, SampEn_{TH}, SampEn_{μ} for each RR series. The value of SampEn_{RR} has been estimated numerically using the procedure described in section 2.4.1.4. The value of SampEn_{TH} has been determined using equation 3.4. Besides these, SampEn_{μ} has been obtained by averaging estimations of $K = 200$ Monte Carlo simulations of the AR models as described in section 3.2.4.

- "Probability of agreement". We have already mentioned in chapter 4 that the parametric estimates SampEn_{TH} and SampEn_{μ} are truly affected by the linear

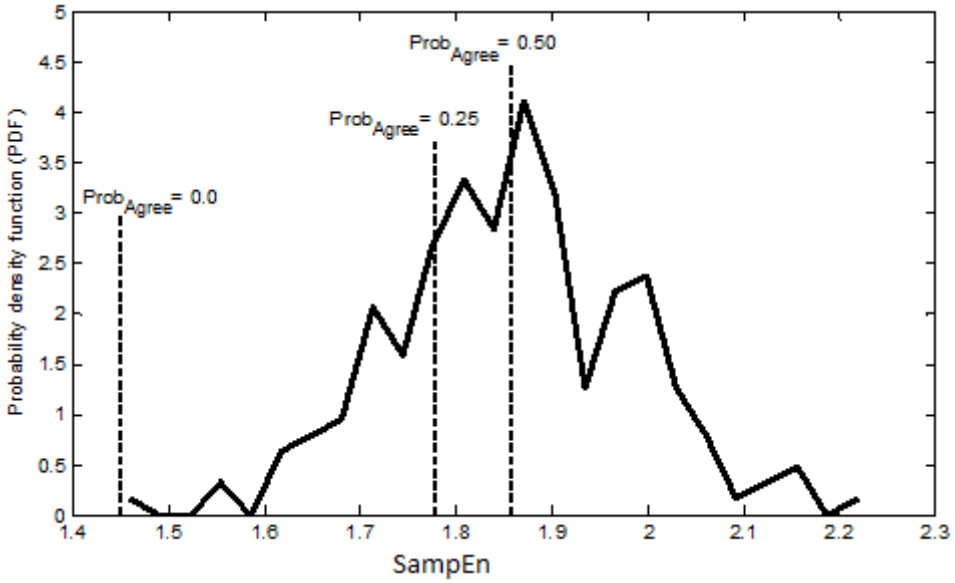


Figure 5.10: The probability density of the values of SampEn computed on 200 synthetic series (thick black line), and the probability of agreement ($Prob_{Agree}$) for three distinct values of $SampEn_{RR}$ (vertical bars). The probability of agreement is indicated for each $SampEn_{RR}$

behavior of the AR model. So, the capability of the AR model to well approximate the series can be tested in terms of $SampEn_{RR}$. To this aim, the distribution of SampEn values estimated from the synthetic series, has been compared with $SampEn_{RR}$. The value of $SampEn_{RR}$ may fall within or outside this distribution (shown in figure 5.10). The probability of agreement ($Prob_{Agree}$), between $SampEn_{RR}$ and the distribution increases from 0 (*i.e* $SampEn_{RR}$ lies outside the distribution) to 0.5 ($SampEn_{RR}$ corresponds to the median of the distribution). The value of $Prob_{Agree}$ has been calculated non-parametrically using the ranks of the distribution of SampEn.

- "Detrended fluctuation analysis feature". Detrended fluctuation analysis (DFA) is a scaling analysis method that provides a simple quantitative parameter to estimate the autocorrelation properties of a non-stationary signal. It has proven useful in characterizing correlations in apparently irregular time series [139]. In DFA, an integrated time series is constructed from the original time series. Then this integrated time series is divided into non-overlapping "time-windows" of increasing window size n and the local or (polynomial) trends are subtracted from the integrated series. The fluctuation of the signal around the trend is determined for increasing the window size. The slope of fluctuation variance versus the window size defines the scaling exponent. The two slopes are termed as "short-range scaling exponent" ($DFA_\alpha 1$) and "long-range scaling exponent" ($DFA_\alpha 2$). In this paper, only $DFA_\alpha 1$ has been considered. The calculation of DFA is summarized in figure 5.11.

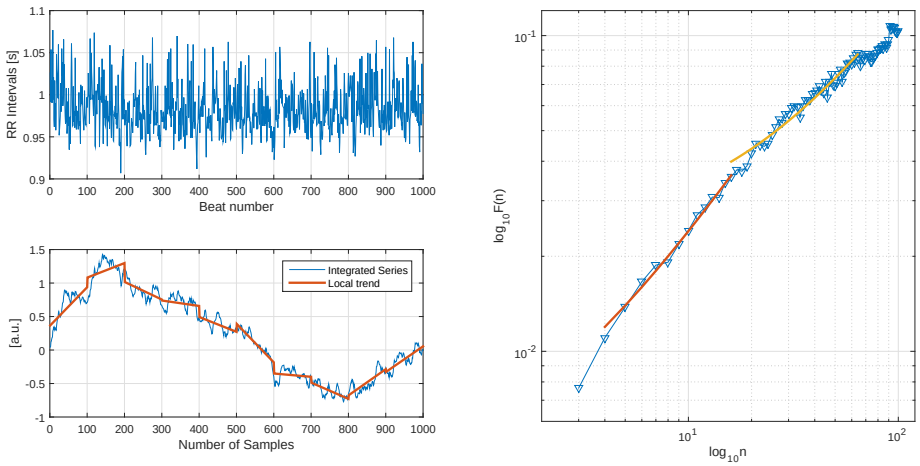


Figure 5.11: The signal is integrated and divided into boxes of equal length (n=100). The local trend (bold line in the plot) is then removed and F[n] is computed.

The RR series of length N is first integrated as

$$y[k] = \sum_{i=1}^{k} (RR[i] - Mean_{RR}), \tag{5.1}$$

where $Mean_{RR}$ is the mean of the RR series. Then, this integrated series is divided into boxes of equal length n. A least-squared line is fitted to the segment in each box for estimating the local trend in that box, as shown in figure 5.11. Finally, the integrated series y is detrended by subtracting the local trends in each box. The root mean square fluctuation of this integrated-detrended series is calculated by

$$F[n] = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y[k] - y_n[k])^2}, \tag{5.2}$$

where $y_n[k]$ is the local trend in each box.

The value of scaling exponent is defined by the slope of the straight line fitted to the log-log graph of n against F[n] using least-squares method. The value of parameter $DFA_{\alpha 1}$ has been defined as the slope in the range $4 \leqslant n \leqslant 11$.

5.5.4 CLASSIFICATION

Classification is the task of categorizing objects into their meaningful groups/classes. The classification performance depends on how robust are the features for describing the objects with respect to the noise and similarities between inter-classes, as well as the proper selection of a classifier. Artificial neural networks (ANN) have been used as classifiers since they were introduced first by McCulloh and Pitts [140] in 1943. The ANN proposed by by McCulloh and Pitts [140] was a basic perceptron, which was

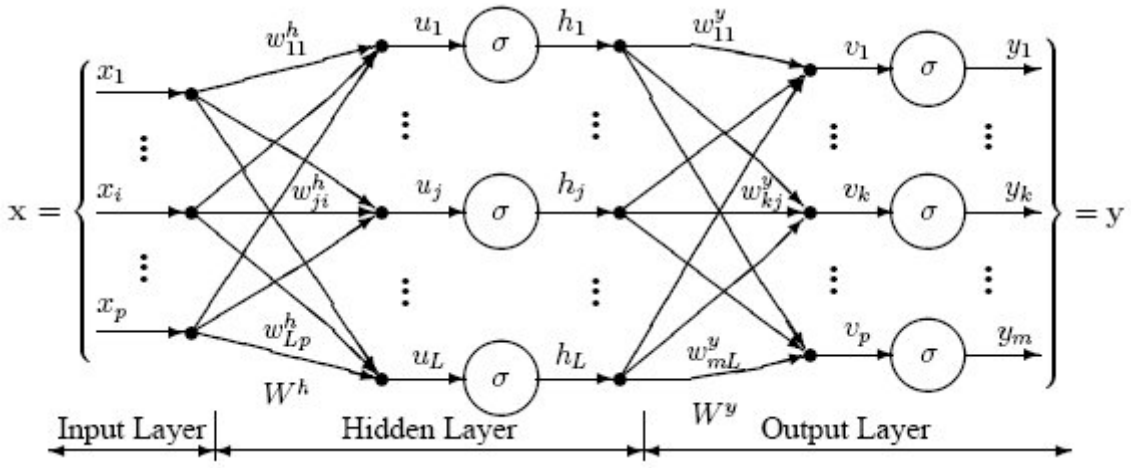


Figure 5.12: Basic structure of a 3 layer ANN. Source: http://www.dtreg.com/mlfn.htm

used as a linear classifier only. They are widely used after the innovation of multilayer perceptron by Hecht-Nielsen in 1990. The multilayer ANN is capable of solving nonlinear classification problems. Although, they require the selection some free parameters: the learning rate and the number of neurons in hidden layer. The learning is process is affected by the values of these parameters settings. If an inadequate number of hidden neurons are used, the network will be unable to model complex data, resulting a poor fitting of the network. On the other hand, using too many hidden neurons, the network may overfit the data and the training time will become too long. The basic structure of a multilayer (3 layer) perceptron ANN is shown in figure 5.12

The number of neurons in the hidden layer has been justified by the average performance of the ANN using different number of neurons. An ANN is first trained using a set of features extracted from the training datset. Then the trained network is used for classification based on the test dataset. In this experimental study, a feedforward back-propagation ANN has been used as a classifier.

The term feedforward is used to mean that neurons are only connected forward, and the term “backpropagation" refers to the mode of training procedure. In backpropagation learning, the neural network is fed with sample inputs and the actual outputs are also provided together. The neural network compares the anticipated output with the actual output for the given input pattern. The backpropagation training function calculates the error based on the anticipated outputs and adjusts the weights at various layers backword from output to the input. In the rest of this study, we will use only ANN to refer to the feedforward backpropagation ANN.

The data collected from 20 recorded subjects are divided according to two different cross validation techniques: Leave One Out (LOO) among recordings, and 10 fold on the total amount of data. In addition, the distribution of the classes in the considered dataset is unbalanced. This may negatively influence the training of the ANN [141]. For this reason, the entire study is repeated using both unbalanced and balanced proportion of classes for training. To balance the populations, equal number of samples of the class with lowest samples are selected randomly from the class with more number

of samples. In addition, the different initializations of the randomly selected weights might lead to (slightly) different classification results. The training and testing of the ANN are repeated 5 times and the average performances are taken into account to minimize this problem.

Finally, the classification performance of the ANN is evaluated by means of four parameters: accuracy (ACC), sensitivity (SENS), specificity (SPEC) and the reliability parameter (Cohens Kappa, k)[142]. The reliability of the classifier is determined by the value of k. The value of k=0, implies that the classification is random, no intelligence is used. On the other hand, the value of k=1, implies the perfect reliability of the classifier.

5.5.5 FEATURE SELECTION

The use of correlated features for training ANN adds noise in training, instead of benefiting the classification accuracy. The full feature set consists of 12 features. The possibility of reducing the features dimension is investigated. The feature selection method is composed by two following approaches:

- **Greedy backward elimination**: The classification reliability, k is checked by leaving one feature out at a time. The feature discarded at each round is the one leading to the highest value of k for the remaining set. This approach is repeated until only the most significant feature is retained.
- **Greedy forward selection** : This procedure starts with a set of features including the single best feature estimated in procedure A such that k value is maximum for the pair of these features. Here, another feature, that maximizes k value from the remaining ones is added to the set at each round. This approach is repeated until all the features are included into the considered feature set.

5.5.6 RESULTS ON CLASSIFICATION OF SLEEP STAGES FROM HRV

In this section, the experiments related to the methods for an automatic classification of sleep stages into WAKE, NREM and REM using RR series analysis are presented. The performance of the system is evaluated by four parameters: accuracy (ACC), sensitivity (SENS), specificity (SPEC), and reliability (k). First, the classification performance of the system is evaluated for full set of features: {$Mean_{RR}$, STD, VLF, LF, HF, LF/HF, $Pole_{HF}$, DFA_{α_1}, $SampEn_{RR}$, $SampEn_{TH}$, $SampEn_{\mu}$ and $Prob_{Agree}$} using ANN, and then the results related to relevant feature selection are described. Finally, the classification performance using relevant features are explained. The distribution of all (12) features {$Mean_{RR}$, STD, VLF, LF, HF, LF/HF, $Pole_{HF}$, $DFA_{\alpha 1}$, $SampEn_{RR}$, $SampEn_{TH}$, $SampEn_{\mu}$ and $Prob_{Agree}$ extracted for each sleep stage is shown in figure 5.13.

5.5.6.1 STATISTICAL ANALYSIS

To get better network training, a logarithmic transformation has been applied to STD, HF and LF/HF in order to get their statistical distributions closer to Gaussian function.

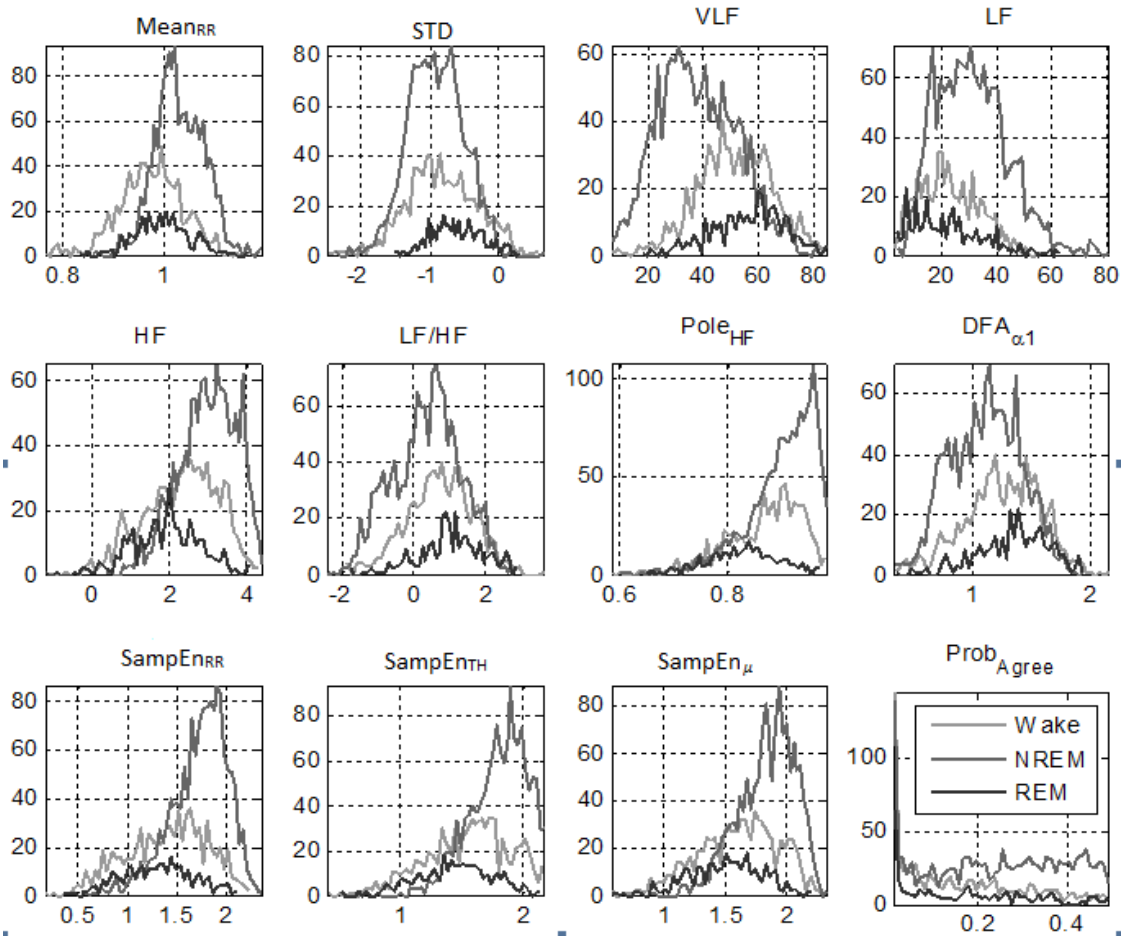


Figure 5.13: The probability distributions of the full set of features during different sleep stages for RR series of 6 epochs long.

The significance of the features has been tested through the analysis of variance. After verifying the non-normality of the features through the Kolmogorov-Smirnov test, a Kruskal-Wallis analysis of variance has been applied. The analysis shows that all the features can significantly ($p < 0.01$) discriminate SLEEP from WAKE and NREM from REM.

5.5.6.2 NUMBER OF HIDDEN NEURONS

The capabilities of FFNN are affected by the number of neurons in its hidden layer. The learning of FFNN can be underestimated using less number of neurons, on the other hand, for more number of hidden neurons, the training of FFNN can be overfitted. To determine the number of hidden neurons in this experimental study, the performances of the classifier are preliminary observed using different number of neurons (8, 12, 15, 20, 25) in a smaller case study (only a subset of the subjects and epochs was used). The results did not improve using more than 12 neurons, which is what we employed in this study.

5.5.6.3 CLASSIFICATION RESULTS USING FULL SET OF FEATURES

There are highly unbalanced proportion of samples of classes. The number of samples of NREM is about 6 times the number of REM class. The training of FFNN can be biased due to this unbalance proportion of samples. So, the performances of the classification system are evaluated for both unbalanced and balanced proportion of samples.

The results for discriminating different sleep stages using the full set of features have been summarized in table 5.6, which reports ACC along with κ, for balanced and unbalanced number of samples and using both 10-fold and LOO validation techniques. The accuracy of WAKE vs SLEEP classification was 77.16% and 71.65%, for 10-fold and LOO techniques respectively, with unbalanced number of samples and 2 epochs. ACC and κ did not change considerably neither increasing the number of epochs nor changing the proportion of samples (balanced or unbalanced). Please notice that SENS and SPEC, respectively, represent here the true recognition of WAKE and SLEEP stages. There was an incremental trend in SENS (from 40.40%, 2 epochs to 52.17%, 10 epochs) and κ (from 0.38 to 0.44) with increasing the window size, when 10 fold validation was used. The same happened in NREM vs REM classification, where ACC increased (from 83.17% to 88.21%), as well as SPEC (29.33% to 57.88%) and κ (0.32 to 0.57). The classifier showed a slightly smaller recognition accuracy (84.62% instead of 88.21%) when LOO was used.

5.5.6.4 FEATURE SELECTION/FEATURE DIMENSION REDUCTION RESULTS

In this subsection, the results of the procedures for selecting the best relevant features in each classification task is illustrated. The results of the feature selection procedure for WAKE vs SLEEP classification is shown in table 5.7. The first row in table 5.7(a) corresponds to the results for the full set of features, as well as the last row of table 5.7(b). The value of reliability parameter, κ is increased from 0.34 (when only Mean_{RR} is considered) to 0.45 (when VLF, DFA_{α_1}, Prob_{Agree} are also added). The addition of the remaining features just increased the features set dimension without any major significant contribution to the value of κ. A similar behavior is observed when reducing the size of the features set. Thus, the subset of feature{Mean_{RR}, VLF, DFA_{α_1}, and Prob_{Agree}} is further considered for WAKE vs SLEEP classification.

In the same way, the results of the feature selection strategy for NREM vs REM classification are illustrated in table 5.8. Also in this case, the value of κ did not improve considerably beyond the addition of 4 features. The four best relevant features for NREM vs REM classification are: {Mean_{RR}, LF, Pole_{HF}, and SampEn_{μ}}.

As a confirmation, if 4 features are sufficient for describing the variability of the data, we further verified using singular value decomposition (SVD) [143] that four transformed variables captured 99% of the variance of the data in both classification problems.

5.5.6.5 CLASSIFICATION WITH RELEVANT FEATURES ONLY

The classification results obtained using only the four most relevant features are reported in table 5.9. When distinguishing WAKE vs SLEEP, there is no significant reduction in ACC, with respect to the full features set (table 5.6), except for unbalanced distributions with 2 epochs, where also SENS as well as k values are highly reduced from 40.0% to 1.98% and from 0.38 to 0.02, respectively. A similar behavior is observed for NREM vs REM classification with no negative effect on the performance parameters (ACC, k) with balanced distributions. Overall, there is no considerable difference between ACC (88.22%) and k (0.56), obtained using only 4 relevant features, and ACC (88.21%) and k (0.57) obtained with the full set of features.

The mean results using 10 fold and the LOO cross validation are comparable for every cases, while the standard deviation is higher using the LOO technique, suggesting that the inter-subject variability is large. As expected, in comparing the results with balanced vs unbalanced number of samples for training and testing, the performances gave privilege to the most represented class (SLEEP for WAKE vs SLEEP and NREM for NREM vs REM classification) when unbalanced samples are used for training. This is typical when the sets are skewed towards one of the classes. This issue supports the statement that the classification is more reliable when training is performed with balanced number of samples.

5.5.7 EVALUATION ON SLEEP STAGES CLASSIFICATION FROM HRV

The features selection strategy has reduced the features set dimension from 12 to 4, by removing redundant ones which did not carry extra information. The overall classification performances (as measured by ACC, k) are not changed significantly when a subset of four features instead of the full set was considered. In addition to time and frequency-domain parameters (already reported in the literature), 3 additional features are included in these sets.

Mean_{RR} and Pole_{HF}, were reported [134] as significant features for sleep staging into NREM vs REM and here also proved so. It is possible that the vagal control reflects its activity in different ways during the three stages considered, being its influence higher during NREM. Similarly, VLF and LF proved here valuable features in discriminating between WAKE vs SLEEP and NREM vs REM, confirming previous studies [134, 135]. VLF has been normalized as a percentage of the total power: its increase during WAKE indicates that the total variance is less influenced by the sympatho-vagal system during WAKE than during sleep. Moreover, while sympathetic activity should increase during REM, in this study, LF is increased during NREM. This result can be explained with the fact that LF is also normalized with respect to the total power. Thus the higher values of LF during NREM are influenced by the lower values of VLF during REM. Finally, the modulus of the strongest pole in the HF band (Pole_{HF}) are proved highly informative in this study for discriminating between NREM and REM, also reported previously as significant [134]. This feature captures the regularity of the respiration rhythm which decreases significantly during REM.

The three methods used to compute entropy showed really similar behavior (figure 5.13). $SampEn_{RR}$, $SampEn_{\mu}$, and $SampEn_{TH}$ in practice carried the same information, even if the latter two are linear indexes while the former is a nonlinear metric. The entropy increases significantly during NREM suggesting a higher regularity during REM sleep.

The relevance of DFA_{α_1} and $Prob_{Agree}$ in the classification process suggests that there are evident changes in short-term correlations and nonlinear regularity of HR during different sleep stages. Apart from increasing the overall classification performances, it also provides information about the physiology of sleep. In fact, during WAKE, the augmented DFA_{α_1} might reflect the increased variability, while a smaller $Prob_{Agree}$ suggests a possible rise of nonlinearity or non-stationarity during such periods.

In both classification problems, the performance parameters ACC and κ increase with increasing number of epochs. This may be due to the fact that short RR series cannot be properly characterized by the selected features.

The accuracy of NREM and REM classification obtained in this study is always larger than 82%, which is an improvement over the results reported in Mendez et al. [134]. Although, Redmond et al. [135] reported a classification accuracy for WAKE vs SLEEP of 89%, which is more than what we achieved, instead they used a set of 30 features, collected not only from ECG but also from respiratory signals.

5.6 PHYSICAL ACTIVITY CLASSIFICATION THROUGH ENTROPY FEATURES

Physical activity refers to any bodily movement produced by skeletal muscles that causes energy expenditure above an underlying level. Human activity recognition (HAR) refers to identify the actions carried out by a person. HAR is an emerging field of research, originating from the major fields of ubiquitous computing, multimedia, and context-aware computing. Recognizing the daily activities is becoming an important application in pervasive computing, with a lot of interesting developments in the healthcare and eldercare. Automatic activity recognition reduces the necessity for humans to oversee difficulties individuals (especially for older adults) might have performing activities, such as falling, when they try to get out of bed [144].

The automation of working process and comfortable travel options in modern society have lead to various mental and physical diseases, such as depression, obesity, cardiovascular diseases, and diabetes, which requires enormous medical costs. According to the World Health Organization (WHO), at least 1.9 million people die every year due to physical inactivity [145]. The Physical activity guidelines advisory committee [146], reports that there is a clear inverse relationship between physical activity and all-cause mortality. So, people should perform at least a certain level of regular physical activities in their everyday life. Activity recognition system can be used to monitor individuals daily physical activities and so as to estimate the consumed calories in every day. Recognition can be performed by exploiting the data collected either from various sources such as environmental [147, 148] or body-worn sensors [149, 150, 151, 152]. The first method works activity recognition using high dimensional and densely sam-

pled video streams. The major drawback of this solution is that to monitor a person, large number of cameras are required to deploy with high costs. Another disadvantage is the privacy is violated. So, activity recognition based on sensory data acquired through one or more accelerometers[4] are reported in many recent works [64, 153, 154]. Accelerometers have been widely used due to their low cost, compact size, low power requirement, no privacy violation, non-intrusiveness and direct acquisition of data related to the motion of a person.

Activity recognition using accelerometer sensors has become popular in the last decade. Extensive research works [150, 155, 156, 157, 154] have been done for classifying postures and activities including sitting, standing, walking, running and so on with a high degree of accuracy . Most of the activity recognition works have focused on the use of multiple accelerometers attached to different sites on a human body and under a controlled environment [155, 158, 159, 160]. However, the use of multiple accelerometers at predefined positions is cumbersome, and not suitable also for long term activity monitoring because of two or more sensors attachments and cable connections [157]. Recently, a small number of works have been focused on the use of a single accelerometer mounted at wrist, waist or chest, with still good recognition accuracies of some basic activities. The level of accuracy using the methods based on single accelerometer is reduced in case of some static activities such as sitting, standing, etc.

Nishkam and Nikhil in [161] have reported recognition of accuracy of 8 daily activities including walking, running, vacuuming, brushing teeth, stairs-Up and stairs-Down using single accelerometer, worn near the pelvic region. The data were collected from two subjects, and have considered four different settings of the dataset: (1) data collected from a single subject over different days, (ii) data collected from multiple subjects over different days, (iii) data collected from a single subject on one day as training and another day for testing from the same subject, and (iv) training data collected from a subject on day and more data from another subject on another day as testing. They have evaluated the accuracies of a number of classifiers for each settings. They reported the maximum recognition with accuracies of 99.57%, 99.81%, and 90.61%, and 65.33%, respectively for settings 1, 2, and 3 using plurality voting[5] classifier. However, the classification accuracy for the 4th setting was 65.33%. The highest recognition accuracy 73.33% for the 4th data setting has been achieved using boosted SVM classifier. Khan et al. [157] have reported a physical activity recognition using a single three dimension (3D) accelerometer, the three dimensions reflect the direction of physical movement in each axis, i.e., forward/backward, left/right and up/down. They have considered static (such as standing, sitting, lying), dynamic (such as running, walking upstairs and downstairs), and transitions such as lie to stand, stand to lie, walk to stand, etc in their methods. They have reported the recognition of 15 activities with an average accuracy of 97.9% using a single triaxial accelerometer attached to the subject's chest.

4 An accelerometer is a sensor that measures the physical acceleration, experienced by an object due to some inertial forces or mechanical excitations. These forces may be static, like the constant force of gravity pulling at our feet, or they could be dynamic- caused by moving or vibrating the accelerometer. They can be used to measure a variety of things such as rotation, vibration, collision, etc. They are measured in terms of acceleration gravity, g=9.81 m/sec^2.

5 Plurality voting selects the class that has been selected by the majority of the base level classifiers such as decision table, decision trees, naive Bayes, SVM as the final predicted class

The dataset used in their experimental study was collected from 6 subjects under a controlled condition. The data collected under this protocol was structured and a fixed time was assigned to perform an activity by the subject. The recognition of physical activities have been reported in the literature with a fairly high accuracy, even with single triaxial accelerometer. The major problem with single accelerometer is that the obtained accuracies are reduced when there is no movement like standing straight or the movement is limited to a certain part of body like hand, mouth.

Many pieces of work have shown that the position of an accelerometer on the human body has a significant impact on the results obtained while measuring physical activity [158, 162]. There is still no dedicated position where the measurements of an accelerometer are able to provide globally best results that are independent of the set of activities; The most frequently referred positions found in the literature are wrist, chest, waist, and ankle. There are also some recent studies that use commercially available mobile devices to collect data for activity recognition [144, 163, 164]. The use of accelerometers embedded into the Samrtphone emphasizes the use of single instead of multiples accelerometer sensors.

Regarding the features, various kinds of time and / or frequency domain features of the accelerometry data have been mentioned in the literature including mean, standard deviation (STD), energy, spectral entropy, cross correlations, autoregressive (AR) coefficients, minimum, maximum, median, percentiles, signal magnitude area (SMA), angle between the vectors, etc [157, 63, 165]. The contribution of a feature may change with changing the application scenario. Among these, the most commonly recommended with acceptable high accuracy features are: AR coefficients, SMA, tilt angle (TA), and cross correlations. The use of more features may increase the recognition accuracy [63]. However, it is expected to avoid the features that need complex computing overload, as they consume much of computing resources and energy, when methods are implemented inside power limited devices such as Smartphones. To the best of our knowledge, though spectral entropy has been used as a member of the features vector for activity recognition, but there is no use of sample entropy for this purpose. On the other hand, we have derived an analytical formula for sample entropy (SampEn) [86] of an AR model, and the parametric estimation of SampEn of an AR model has been proposed for HRV analysis. The feasibility of the parametric estimation of SampEn has been justified for short series, when there is nonstationarity and / or nonGaussianity. It is again remembered that the theoretical value of SampEn (SampEn_{TH}), is defined only by the coefficients of the AR model. Thus, instead of using a number of AR coefficients, a single value of SampEn can be used to keep smaller the feature set dimension.

As for the classifiers, a number of classification methods including k-nearest neighbor (kNN), decision trees, support vector machines (SVM), and artificial neural network (ANN) have been investigated in the literature. Among them, ANN and SVM have been proved to provide higher accuracy compare to others [144, 157, 161]. Khan and coauthors [157] reported that the use of hierarchical classifiers[6].

6 In the first stage, the activities are first divided into static and dynamic classes using a classifier. Then in the second stage one classifier is used for classifying static activities and another for classifying dynamic activities.

The researchers of each group have developed their own methods for a specific set of movements, employing their own devices. This highly specific set of movements and methodologies make it difficult or even impossible to make explicit comparisons between the approaches of different researchers. The majority of the common people are not habituated in regular exercise like jogging, swimming, running, etc. We have considered only those activities that people do in common during their daily life. So, we have considered walking, bicycling, sitting, standing, and lying.

Actigraph GT3X+ [166] and GENEActiv [167, 168] are two famous accelerometer tools that provide triaxial measured acceleration data. The acceleration measured by the two brands was correlated by ($r=0.93$, $p<0.001$) [169] in case of both laboratory and free-living environments. In our study, We have considered GENEActiv for recording acceleration data.

We have extracted a set of features, mentioned in the literature including the SampEn from the accelerometers worn on the wrist and waist, separately from the same subjects. The most relevant features have been selected from the extracted features. Then, the relevant features are used for training ANN and SVM, as they are mostly used with higher accuracy. The accuracy of them was also compared on the testing dataset.

The main objectives of this study are to determine, (i) if use of SampEn instead of AR coefficients provides the comparable accuracy of physical activity recognition, (ii) which of the two classifiers (ANN or SVM) gives maximum accuracy.

5.6.1 SENSOR DATA ACQUISITION

We have used a triaxial accelerometer called GENEActiv, which has been proven effective in both small scale studies and large international cohorts of over 10,000 subjects [167]. It is a micro-electromechanical system (MEMS) sensor with dynamic range (*i.e.* the maximum amplitude vibration that can be detected by the sensor) ± 8G and resolution 12 bit. The sampling frequency can be varied from 10 to 100 Hz. The output of an accelerometer actually depends on its position of the human body. In this experimental setup, we have placed the accelerometer on the subject's chest as we are interested on the whole body movement [157]. The polarity of the GENEActiv accelerometer and it's position on the chest are shown in figure 5.14.

The accelerometer data have been collected from four adults (1 woman and 3 men; age range 35 ± 5) for five physical activities: (i) Bicycling, (ii) Walking, (iii) Sitting, (iv) Lying, and (v) Standing with around 30 ± 5 minutes for each activity. Thus, a total of around 10 hours of recording with 50 Hz sampling frequency. During sitting, the subjects were working on their desktop and during standing they were talking to their colleagues. The subjects have been well informed about the purpose of recording. During lying, all subjects remain supine in supine position. They have been walking and cycling on free streets without facing any traffic or unwanted halting. The data for each activity from each subject have been collected continuously without any pause. A sample of the activity signals for each axis of the accelerometer is shown in figure 5.15.

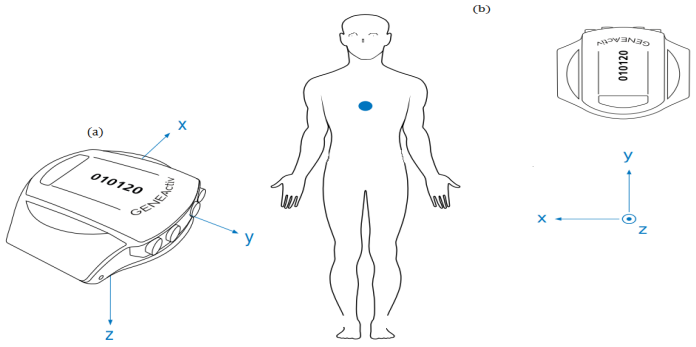


Figure 5.14: The polarity of GENEActiv accelerometer and it's position on the chest. Panel (a) shows the general polarity of 3 axes of the accelerometer, panel(b) shows the position of the accelerometer marked by the blue dot on the chest of human body and the respective orientation during data acquisition.

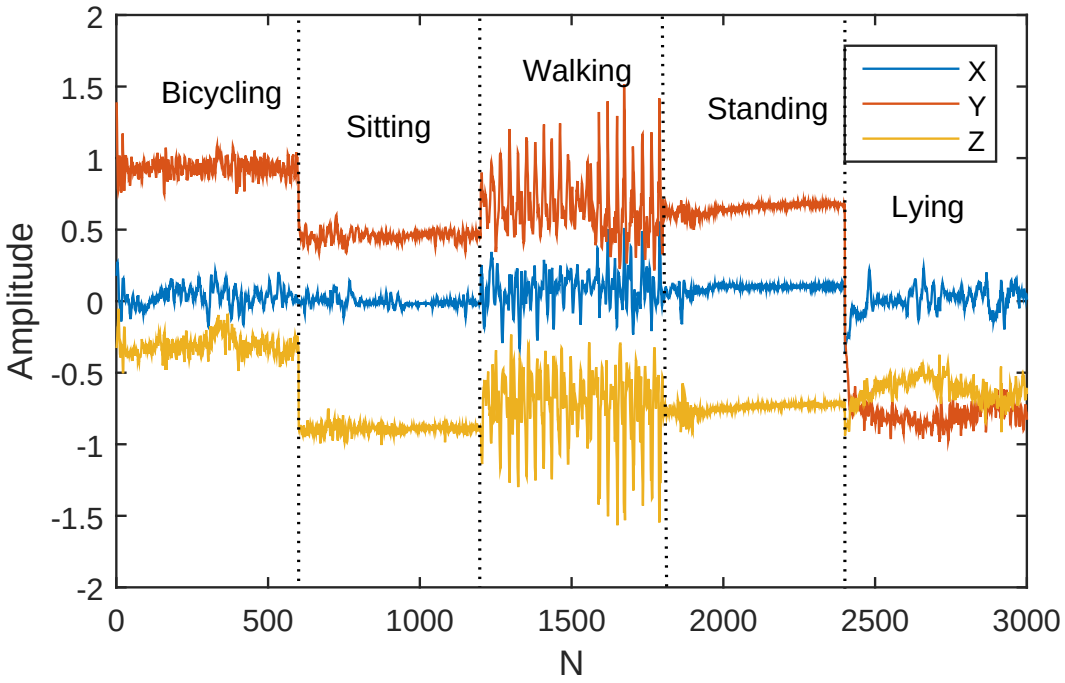


Figure 5.15: Set of sample aceleration signals of five types of human activities for 3 axes of the traxial accelerometer.

5.6.2 METHODS

The general structure of the proposed physical activity recognition method is represented in figure 5.16. The physical activity recognition method consists of the following steps:

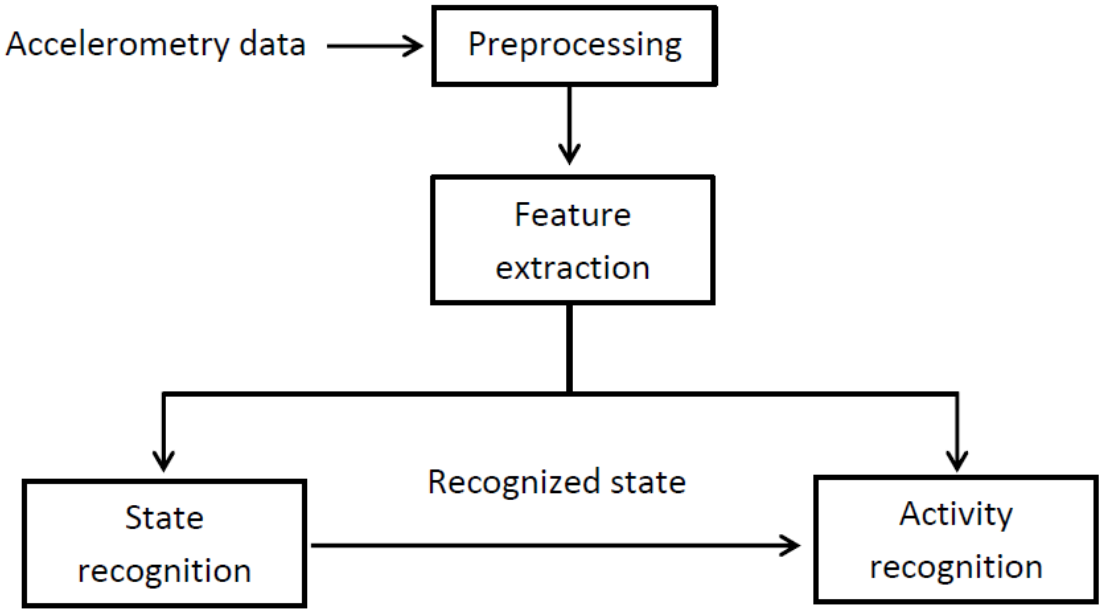


Figure 5.16: The general block diagram of the physical activity recognition system.

5.6.3 PREPROCESSING

The accelerometer's real time output may contain random noise that should be filtered out before it is used for activity recognition. A 5 point moving average filter (used previously by other researchers [157]) has been applied for filtering out the noise from the accelerometery data. The noise eliminated data are then divided in fixed-width sliding windows of N=75 samples (1.5 sec) with no overlapping.

5.6.4 FEATURE EXTRACTION

The signal of each axis is processed simultaneously, and an augmented feature vector of 61 features (*i.e.* the set of almost all features reported in the literature for activity recognition including some new features) have been extracted from the series of each axis. These features include time domain features such as mean, STD, minimum, maximum values, correlation between the axes, energy, signal magnitude area (SMA), tilt angle (TA), and vector magnitude (VM); frequency domain such as dominant frequency and it's magnitude, the first 3 AR coefficients, the powers in very low frequency (VLF : [0.003 0.04] Hz) band, low frequency band (LF: [0.04 0.15] Hz), high frequency band (HF: [0.15 0.4] Hz), and the ratio between LF and HF (LF/HF). Besides time and frequency domain parameters, the entropy features such as SampEn_{RR} (*i.e.* the numerical estimation of SampEn from the actual series), SampEn_{TH} of an AR model fitted to the series, and the expected value of SampEn (SampEn_{μ}) of the model have been obtained through K=200 realizations of Monte Carlo simulations, the probability of agreement between SampEn_{RR} and distribution of estimations obtained through MonteCarlo simulations for each axis. The extraction of all features except energy, dominant

frequency and it's magnitude, correlation, SMA, and TA have been explained before. So, the description of only these features is given here.

- "Energy (EN_T)". The energy of a triaxial accelerometer signal of length N:

$$E_{NT} = \sum_{n=1}^{N} |x[n]|^2 + |y[n]|^2 + |z[n]|^2$$

  where, x, y, and z represents respectively, the acceleration signal for X, Y, and Z axes of the accelerometer.

- "Dominant frequency ($domFreq$) and it's magnitude". There are situations when an observed signal show a periodic behavior due to the presence of dominant frequency, (*i.e.* the frequency at which the signal carries the highest energy among all frequencies). The notion of dominant frequency is similar to fundamental frequency that is the smallest frequency having a peak among all frequencies in the spectrum. Thus, the $domFreq$ and the highest magnitude at this frequency are determined for each axis of the accelerometer series.

- "Inter-axis correlation". The Inter-axis correlation defines the correlation between each pair of accelerometer series, *e.g.* the correlation, between accelerometer signals of X and Y axes is defined as

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x[i] - \mu_x) \star (y[i] - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x[i] - \mu_x)^2} \star \sqrt{\sum_{i=1}^{N}(y[i] - \mu_y)^2}},$$

  where μ_x and μ_y are the mean values of X and Y axes, respectively. Similarly, the correlations ρ_{yz} and ρ_{zx} are determined.

- "AR coefficients". The data from each axis has been fitted to an AR model. The model order is determined by satisfying the Akaike information criterion (AIC) and the Anderson's whiteness test. Then the first 3 coefficients are considered for each axis.

- "SampEn". The three measures of SampEn: $SampEn_{RR}$, $SampEn_{TH}$, and $SampEn_{\mu}$ are estimated for acceleration signals of each axis. The values of parameters m=1 and r=0.2×STD are used in their estimations. The value of $SampEn_{TH}$ is obtained using equation 3.4. The values of $SampEn_{RR}$ and $SampEn_{\mu}$ are obtained using the procedures described in sections 2.4.1.4 and 3.2.4, respectively. The value K=200 is used in estimating $SampEn_{\mu}$.

- "Normalized Signal magnitude area (SMA_{norm})". The SMA, which has been reported as a significant feature in many works [150, 157, 170, 171] is defined by the sum of the absolute values for all axes of the series. The normalized (with respect to the length of the window) SMA is computed by

$$SMA_{norm} = \frac{1}{N}\Sigma_{i=1}^{N}(|\, x[i] \,| + |\, y[i] \,| + |\, z[i] \,|)$$

- "Tilt angle (TA)". The tilt angle (TA) refers to the relative tilt of the body in space [157], which is determined by the average angle (θ) between the vector of gravitation (**G** and the positive Y axis towards the ground). The angle $\theta(i)$ (in degree) is computed for each sample. Then TA is computed by the average of θ. Thus, tilt angle

$$TA = \frac{1}{N}\sum_{i=1}^{N} \mid arccos(y(i)) \mid \tag{5.3}$$

- "Vector magnitude (VM)". The vector magnitude of the triaxial accelerometer is determined by the square root of the sum of squares of magnitudes of each axis. Thus,

$$VM = \frac{1}{N}.\sqrt{\Sigma_{i=1}^{N}(x[i]^2 + y[i]^2 + z[i]^2)}$$

5.6.5 BEST RELEVANT FEATURES SELECTION

Thus, we have extracted a set of 61 features from the triaxial acceleration data. All of these features might not equally important for classifying the physical activity. Some of them describe the same properties, *i.e.* they are highly correlated. On the other hand, some features may be insignificant in classifying the activities. The best relevant among 61 features have been selected using the strategy, which has been described in section 5.5.5 based on the performance of FFNN. The same features have been used for SVM training and testing.

5.6.6 CLASSIFICATION

After computing the features from each level of physical activity, the relevant features have been fed into FFNN and SVM. FFNN has been described in section 5.5.4. Vapnik et al. [172] generalized a new class of learning machine called support vector machine (SVM), even though the concept was originally initiated in 1982 for case of linearly separable classes with no errors. SVM maps the input vectors into some dimensional feature space through some apriori chosen nonlinear mappings. It is a binary classifier, whose job is to find an optimal hyperplane to separate the training data into the given classes. An optimal hyperplane is defined as the linear decision function with maximal margin between the vectors of the two classes, as depicted in figure 5.17. It is observed from figure 5.17 that the small set of training data, called support vectors are used to determine such optimal hyperplane.

In this study, we have used both FFNN and SVM as classifiers. Due to the study data collected from a few number of subjects, the train and test sets have been prepared using 10 fold cross validation technique, *i.e.* the total features have been divided equally into 10 random folds; among which 9 folds have been used for training and the remaining 1 fold for testing. The training and testing cycles are repeated for 10 times.

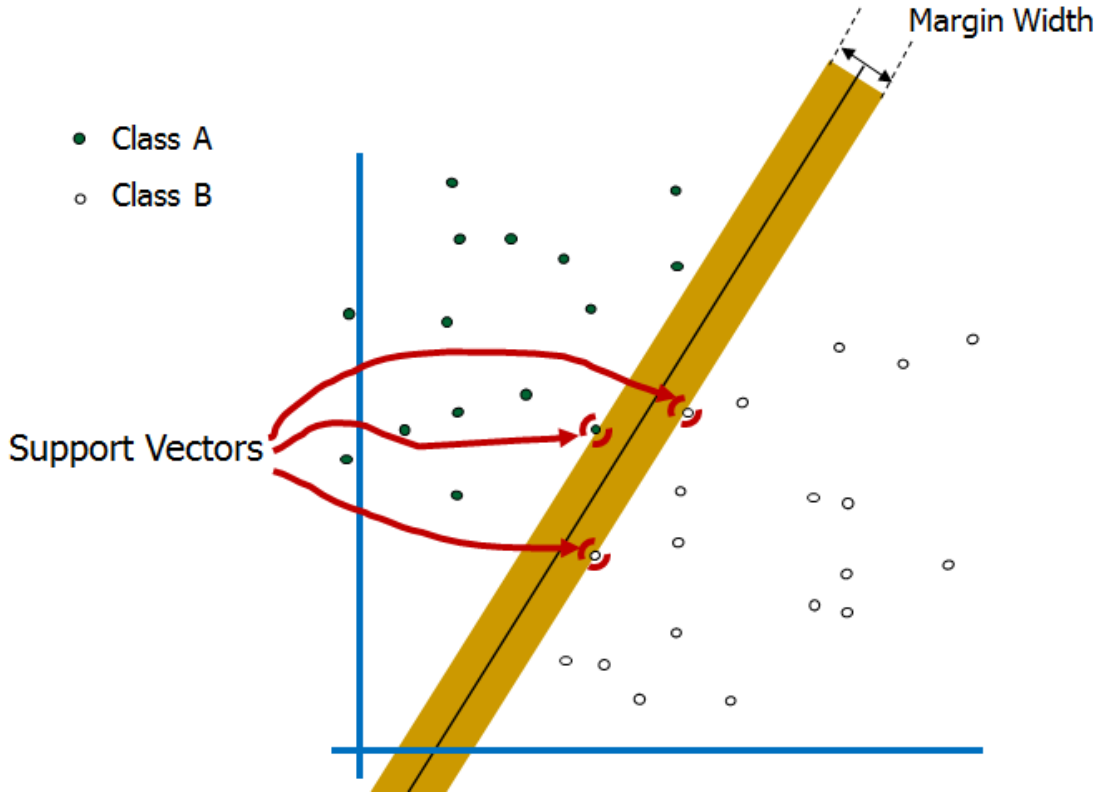


Figure 5.17: Support vectors for two linearly separable classes of objects A and B. It is called a linear SVM. This is a little modified version of SVM given in http://www.iro.umontreal.ca/ pift6080/H09 /documents/papers/svmtutorial.ppt

The classifiers are initialized in each iteration. Finally, the accuracy is determined by the average of the accuracies obtained in 10 repetitions.

5.6.7 RESULTS ON PHYSICAL ACTIVITY CLASSIFICATION

A set of nine different features: the mean ($Mean_Z$), minimum (Min_Z) and maximum (Max_Z) values of Z-axis, the standard deviation (STD_X) of X-axis, the theoretically derive value of SampEn of the AR model ($SampEn_{TH}(Y)$), the ratio of low frequency power to high frequency power (LF/HF_Y), the magnitude of dominant frequency ($|dominantFreq_Y|$) of the Y-axis acceleration signals, and the tilt angle (TA), have been found as the best relevant ones. The inclusion of other features do not increase the accuracy. The number of hidden neurons, which gives maximum accuracy by the FFNN was 9.

The distributions of the best relevant features are shown in figure 5.18. It is obvious from visual observation of the features distribution that some features should be more useful for distinguishing the static activities from the dynamic activities. On the other hand, some features are useful for classifying the inter-static activities, other subset of

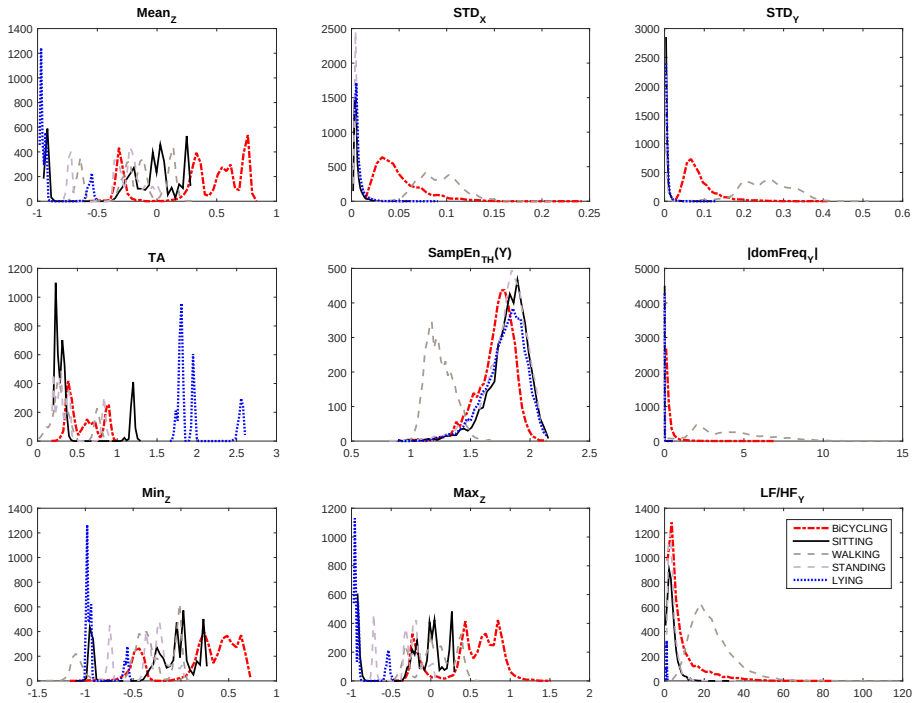


Figure 5.18: The distribution of some selected features for five physical activities recognition.

features seem to be effective for classifying dynamic activities themselves. So, in case of hierarchical classifiers, only the set of $Mean_Z$, STD_X, STD_Y, and TA have been chosen in the first stage for classifying the activities into STATIC (SITTIN, STANDING, and LYING) and DYNAMIC (WALKING and CYCLING) states. Then in the second stage, the set of three features consist of STD_Y, $SampEn_{TH}(Y)$, and LF/HF_Y for classifying dynamic, and the set of $Mean_Z$, STD_Y, TA, $|dominantFreq_Y|$, and Max_Z for classifying the static activities have been chosen. In hierarchical FFNN classifier, the number of hidden neurons was set equal to the number of input features.

The accuracy of the system for physical activity recognition has been evaluated for the following cases:

5.6.7.1 CLASSIFICATION OF FIVE PHYSICAL ACTIVITIES

We get average accuracies of 88.16% and 87.73%, respectively using ANN and SVM for five physical activities. The classification accuracy of CYCLING, WALAKING, and LYING are very high (average is more than 98%)in using both classifiers. However, the accuracies of SITTING and STANDING are low (average is less than 65%), because they confused each other. The acceleration signals during these two activities are very similar as shown in figure 5.15. So, we have discarded STANDING activity from the dataset, and hence the the following results are reported on the classification of 4 instead of 5 activities.

5.6.7.2 CLASSIFICATION OF FOUR PHYSICAL ACTIVITIES

The accuracy, we obtained for the classification of four activities using the vector of features STD_Y, TA, LF/HF_Y, and $SampEn_{TH}$ of 3-axes (using SVM: 99.36%, ANN:98.67%), even with only one value of SampEn (*i.e.* $SampEn_{TH}$ of *Y*-axis), the accuracy is still very high (SVM:99.25%, ANN:94.11%) which is better than any results reported in the literature. The details of the classification of 4 activities are illustrated by table 5.10.

The classification accuracy of four activities using single stage SVM is very high for any activity, even though the recognition accuracy of dynamic activities (CYCLING and WALKING) using ANN is comparatively small (Please see column # 4 of Table 5.10. So, our next target is to investigate if hierarchical classifier using ANN can improve the recognition accuracy than single (linear) classifier. The use of hierarchical structure of SVM does not change the results anymore. However, the use of hierarchical classifier based on ANN improves the classification accuracy. The details classification results of 4 activities using hierarchical ANN classifier is shown in table 5.11. The obtained results have been explained for three classifiers: the classification of STATIC states and DYNAMIC states by the first classifier, the results for classifying static activities, and finally the results for classifying the dynamic activities. Thus, the use of hierarchical classifier increases the classification accuracy of CYCLING from 87.89% to 89.41% and WALKING from 89.13% to 98.32%.

5.6.7.3 COMPARISON OF CLASSIFICATION ACCURACY USING AR COEFFICIENTS AND SAMPEN

The classification accuracy of four classes: SITTING, LYING; WALKING, and CYCLING using only the set of 9 AR coefficients (3 for each axis) is (ANN:62.68%, SVM:52.04%). On the other hand, using only the theoretical values of SampEn of each axis ($SampEn_{TH}$(X), $SampEn_{TH}$(Y),$SampEn_{TH}$(Z)), we get the recognition accuracy of (ANN: 63.21%, SVM:64.49%). Thus, we obtain about the same classification accuracy using only 3 values of SampEn, instead of 9 values of AR coefficients. However, neither SampEn nor AR coefficients can alone classify the activities for more than 64%. To achieve higher accuracy, other relevant features should be augmented with either AR coefficients or SampEn. In our following investigation, we will consider other relevant features with $SampEn_{TH}$.

5.6.8 EVALUATION ON PHYSICAL ACTIVITY CLASSIFICATION SYSTEM

The development of the system has been started with five physical activities and the accuracy of each activity has been computed using both ANN and SVM classifiers. Neither SVM nor ANN is able to distinguish STANDING and SITTING with considerable accuracy. However, the recognition of other activities are comparable with the existing methods, even with less number of features. A subject spends a very small portion of her/his daily activities by standing situation. So, the standing activity has been discarded finally. The similar accuracy is obtained using either AR coefficients (9 features, each for each axis of acceleration signals) or $SampEn_{TH}$ (3 features, one for each axis of acceleration signals).

The developed system incorporates the use of a single triaxial accelerometer on the center position of human chest. It is feasible to wear by the free living subjects as it relies on the single point of sensors attachment to their body. It is significantly effective in a sense that it is able to recognize the main common daily physical activities with average accuracy more than 99%.

Although many systems have been reported in the literature for monitoring daily physical activities from signals acquired through a triaxial accelerometer, this system appears promising in some regards. At first, the performance obtained from this system compares better for same activities [157] reported in the literature, even with less number of features. The use of theoretically derived SampEn of the AR model that are actually based on the model coefficients capture more or less the same information represented by the model coefficients. It is mentioned that the AR coefficients have been proved effective features for physical activity recognition. The use of single value just, $\text{SampEn}_{\text{TH}}$, instead of a set of AR coefficients really reduced the feature dimension.

The use of SVM always provides better accuracy than ANN. The hierarchical structure of the classifier based on ANN improves the accuracy, even though no changes are achieved in case of SVM. Although, the sensors and the subjects used for collecting acceleration signals in [157] are different than those used in this system, the results may be compare in some sense for similar activities, same position of the sensor, and same classifier. The developed system gives better results than those provided in [157] for single (linear) classifier. A comparison of the results for similar activities are provided in table 5.12.

The increased accuracy might be due to the use of new features or accelerometer sensor (higher sensitivity), and/ or both. In the developed system, data have been collected from 4 subjects (1 woman and 3 men), a total of about 10 hours of recording.

5.7 SUMMARY

Innovative methods for extracting features from physiological signals have been described. The methods are based on the analysis of HRV, ABP variability, sleep stages classification from HRV and physical activity recognition. A set of new features in addition to the existing ones for each of the considered classification purposes have been described. With respect to the features in the literature, the proposed entropy based features are more relevant for the considered classification purposes. The parametric estimations of entropy has been applied for HRV regularity analysis during persistent atrial fibrillation, and also for nonlinear regularity analysis of arterial blood pressure variability.

A set new features with the existing time and frequency domain parameters have been explained. Feature selection strategy has been described for selecting the best relevant features, and hence reducing the feature's set dimension. Entropy features have been selected as one of the best relevant features for classification or recognition purposes.

An automatic sleep stage classification, based on heart rate variability analysis, with a focus on the distinction of WAKE from sleep, and REM from NREM sleep has been

studied. SampEn measures have been found significant for NREM vs REM classification, while the probability of agreement (or disagreement) between numerical and parametric estimates of SampEn has been found as one of the relevant features for WAKE vs SLEEP classification. The best relevant features have been used for development of the an automatic sleep classification system.

The same procedure has been followed for the development of an automatic classification of physical activities from acceleration data. In physical activity recognition, support vector machines, and also the hierarchical structure of the classifiers have been considered.

Table 5.6: Sleep stage classification using the full features set. Table (a): results for WAKE vs SLEEP classification; Table (b): NREM vs REM classification results. The length of RR series is represented in epochs. The type of distribution is denoted by Distr.; the terms 'Bal.' and 'Unbal.' have been used to mean the distribution with equal or unequal number of samples of each class.

a) WAKE vs SLEEP

		10 Fold				LOO			
Epochs	Distr	ACC (%)	SENS (%)	SPEC (%)	k	ACC (%)	SENS (%)	SPEC (%)	k
2	Unbal.	77.16±0.14	40.40±0.71	92.62±0.25	0.38±0.01	71.65±13.77	41.39±20.78	89.61±8.75	0.28±0.18
	Bal.	69.93±0.69	66.22±0.71	73.64±1.10	0.40± 0.01	68.37±11.79	65.56±21.27	70.93±17.00	0.27±0.21
6	Unbal.	77.22±0.68	47.20±2.80	89.99±0.62	0.41 ± 0.02	69.66 ± 16.69	42.35 ± 25.58	88.28 ± 10.11	0.26 ± 0.22
	Bal.	72.40 ± 0.95	73.51 ± 1.36	71.29 ± 1.14	0.45 ± 0.02	67.59 ± 12.09	71.97 ± 22.28	70.26 ± 17.12	0.31 ± 0.24
10	Unbal.	77.91 ± 0.27	52.17 ± 1.42	88.90 ± 0.76	0.44 ± 0.01	71.92 ± 18.24	43.68 ± 27.34	88.96 ± 7.78	0.29 ± 0.24
	Bal.	75.09 ± 1.10	76.17 ± 2.18	74.00 ± 2.15	0.50 ± 0.02	70.76 ± 10.68	73.28 ± 23.74	72.10 ± 15.31	0.32 ± 0.24

b) NREM vs REM

Epochs	Distr	ACC (%)	SENS (%)	SPEC (%)	k	ACC (%)	SENS (%)	SPEC (%)	k
2	Unbal.	83.17 ± 0.14	96.02 ± 0.22	29.33 ± 1.19	0.32 ± 0.01	82.07 ± 5.14	94.42 ± 6.33	30.30 ± 22.68	0.27 ± 0.16
	Bal.	71.96 ± 0.86	72.48 ± 2.02	71.44 ± 1.39	0.44 ± 0.02	68.69 ± 16.55	69.69 ± 22.88	67.59 ± 23.85	0.29 ± 0.19
6	Unbal.	86.74 ± 0.46	94.78 ± 0.36	51.38 ± 3.35	0.51 ± 0.02	84.63 ± 6.51	93.13 ± 9.33	46.39 ± 26.96	0.41 ± 0.21
	Bal.	80.50 ± 1.02	80.58 ± 2.04	80.42 ± 1.75	0.61 ± 0.02	75.83 ± 16.92	75.57 ± 22.85	78.20 ± 23.00	0.44 ± 0.24
10	Unbal.	88.21 ± 0.63	94.91 ± 0.54	57.88 ± 2.72	0.57 ± 0.02	84.62 ± 8.12	91.47 ± 11.41	52.27 ± 33.69	0.42 ± 0.25
	Bal.	82.79 ± 2.17	84.00 ± 2.18	81.59 ± 2.92	0.66 ± 0.04	79.39 ± 15.65	79.40 ± 19.98	80.17 ± 22.38	0.49 ± 0.25

Table 5.7: Results of the features selection procedure for WAKE vs SLEEP classification, using windows of 6 epochs with balanced datasets. Table (a): classification performances after removing one feature at a time (the feature removed is indicated in each row). Table (b): classification performances after adding one feature at a time.

a) Removing a feature

	ACC (%)	SENS (%)	SPEC (%)	k
All	74.40	73.80	74.90	0.49
VLF	76.10	76.60	75.60	0.52
DFA_{α_1}	74.90	75.90	74.00	0.50
$SampEn_{TH}$	75.00	75.30	74.70	0.50
$SampEn_{\mu}$	74.70	74.80	74.50	0.49
LF/HF	74.80	74.40	75.20	0.50
$Pole_{HF}$	75.80	77.40	74.20	0.52
$SampEn_{RR}$	74.90	74.40	75.50	0.50
STD	73.30	74.20	72.30	0.47
$Prob_{Agree}$	71.60	71.20	72.10	0.43
HF	68.90	62.50	75.30	0.38
LF	67.10	58.90	75.30	0.34

b) Adding a features

	ACC (%)	SENS (%)	SPEC (%)	k
$Mean_{RR}$	67.10	58.90	75.30	0.34
VLF	70.10	69.90	70.30	0.40
DFA_{α_1}	72.10	71.20	72.90	0.44
$Prob_{Agree}$	72.50	73.20	71.80	0.50
STD	72.90	74.20	71.50	0.46
LF/HF	73.30	72.90	73.70	0.47
$SampEn_{\mu}$	74.80	75.90	73.70	0.50
$SampEn_{TH}$	74.30	74.50	74.10	0.49
HF	75.00	74.50	75.50	0.50
$SampEn_{RR}$	74.60	75.30	73.80	0.49
LF	74.20	74.00	74.40	0.48
$Pole_{HF}$	74.40	73.80	74.90	0.49

Table 5.8: Results of the features selection procedure for NREM vs REM classification, using windows of 6 epochs with balanced datasets. Table (a): classification performances after removing one feature at a time (the feature removed is indicated in each row). Table (b): classification performances after adding one feature at a time.

a) Removing a feature

	ACC (%)	SENS (%)	SPEC (%)	k
All	84.10	83.70	67.70	0.68
VLF	84.40	85.40	83.40	0.69
HF	84.20	84.60	83.90	0.69
$SampEn_{TH}$	84.50	85.50	83.40	0.70
DFA_{α_1}	84.90	85.70	84.20	0.70
LF/HF	84.50	84.80	84.10	0.69
STD	84.80	85.60	84.10	0.67
$SampEn_{\mu}$	84.40	83.90	84.80	0.69
$Prob_{Agree}$	84.70	84.80	84.50	0.69
$SampEn_{RR}$	83.60	82.60	84.60	0.67
$Mean_{RR}$	80.90	79.10	82.70	0.62
LF	75.90	74.20	77.60	0.52

b) Adding a features

	ACC (%)	SENS (%)	SPEC (%)	k
$Pole_{HF}$	75.90	74.20	77.60	0.52
LF	80.50	79.50	81.60	0.61
$Mean_{RR}$	83.30	81.80	84.80	0.67
$SampEn_{\mu}$	84.60	84.60	84.50	0.70
$SampEn_{RR}$	85.10	85.10	85.20	0.70
STD	84.70	85.20	84.30	0.70
DFA_{α_1}	85.00	85.60	84.40	0.70
VLF	85.00	85.50	84.50	0.70
LF/HF	84.60	84.80	84.30	0.69
HF	84.60	84.70	84.50	0.69
$Prob_{Agree}$	83.50	83.80	83.30	0.67
$SampEn_{TH}$	84.16	83.72	67.71	0.68

Table 5.9: Sleep stages classification using 4 relevant features only. Data Table (a): results (mean±std) for WAKE vs SLEEP classification; Table (b): NREM vs REM classification

a) WAKE vs SLEEP

		10 Fold				LOO			
Epochs	Distr	ACC (%)	SENS (%)	SPEC (%)	k	ACC (%)	SENS (%)	SPEC (%)	k
2	Unbal.	70.67 ± 0.29	1.98 ± 2.09	99.57 ± 0.47	0.02 ± 0.02	69.81 ± 16.34	23.04 ± 15.21	94.40 ± 6.33	0.17 ± 0.16
	Bal.	67.69 ± 0.44	62.89 ± 0.91	72.48 ± 0.77	0.35 ± 0.01	67.55 ± 9.93	65.29 ± 17.72	70.37 ± 13.81	0.26 ± 0.22
6	Unbal.	72.38 ± 0.74	16.64 ± 4.97	96.12 ± 1.22	0.16 ± 0.04	70.85 ± 14.89	37.56 ± 23.28	90.94 ± 7.75	0.24 ± 0.22
	Bal.	70.79 ± 0.81	72.99 ± 1.49	68.60 ± 2.20	0.42 ± 0.02	69.02 ± 11.34	75.13 ± 18.99	68.57 ± 15.92	0.31 ± 0.24
10	Unbal.	74.23 ± 0.69	31.22 ± 4.62	92.60 ± 1.08	0.28 ± 0.04	71.02 ± 15.17	42.64 ± 27.83	88.95 ± 8.78	0.25 ± 0.26
	Bal.	73.30 ± 0.94	76.39 ± 2.21	70.21 ± 1.20	0.47 ± 0.02	71.34 ± 11.73	77.15 ± 17.00	69.27 ± 14.02	0.33 ± 0.23

b) NREM vs REM

Epochs	Distr	ACC (%)	SENS (%)	SPEC (%)	k	ACC (%)	SENS (%)	SPEC (%)	k
2	Unbal.	81.64 ± 0.12	98.45 ± 0.27	11.26 ± 1.49	0.14 ± 0.02	82.12 ± 5.33	96.63 ± 3.65	19.30 ± 20.86	0.18 ± 0.16
	Bal.	71.73 ± 0.66	74.08 ± 0.78	69.38 ± 0.83	0.43 ± 0.01	68.35 ± 15.32	68.34 ± 21.57	71.11 ± 22.58	0.29 ± 0.18
6	Unbal.	85.68 ± 0.43	95.65 ± 0.45	41.81 ± 3.84	0.44 ± 0.03	83.07 ± 6.79	93.98 ± 9.92	37.70 ± 31.03	0.32 ± 0.21
	Bal.	80.28 ± 1.33	79.40 ± 2.29	81.16 ± 0.77	0.61 ± 0.03	74.54 ± 19.72	73.74 ± 25.54	80.24 ± 20.34	0.44 ± 0.24
10	Unbal.	88.22 ± 0.44	95.67 ± 0.42	54.52 ± 3.69	0.56 ± 0.03	85.78 ± 7.27	93.43 ± 9.87	51.90 ± 33.90	0.45 ± 0.26
	Bal.	83.78 ± 2.06	82.41 ± 2.60	85.15 ± 2.73	0.68 ± 0.04	79.77 ± 15.72	79.68 ± 20.07	81.13 ± 22.88	0.51 ± 0.25

Table 5.10: The classification accuracy (%) of 4 physical activities using linear classifiers

			ANN
Activity	[AR Coeffs]	[$SampEn_{TH}$]	[STD_Y,$SampEn_{TH}(Y)$,TA,LF/HF_Y]
CYCLING	57.06	56.22	87.89
SITTING	56.41	57.41	99.42
WALKING	96.38	98.38	89.13
LYING	40.87	40.83	100
Average	62.68	63.21	94.11
			SVM
Activity	[AR Coeffs]	[$SampEn_{TH}$]	[STD_Y,$SampEn_{TH}(Y)$,TA,LF/HF_Y]
CYCLING	72.15	86.48	99.47
SITTING	58.09	68.76	99.61
WALKING	72.23	96.75	97.90
LYING	5.70	5.98	100
Average	52.04	64.49	99.25

Table 5.11: The classification accuracy (%) of four physical activities using hierarchical ANN

(a) Classification accuracy in the first classifier	
Activity	Accuracy
STATIC	99.74
DYNAMIC	99.81
Average	99.77
(b) Classification accuracy in the 2nd classifier	
Activity	Accuracy
SITTING	99.96
LYING	99.95
Average	99.95
(c) Classification accuracy in the 3rd classifier	
Activity	Accuracy
CYCLING	89.41
WALKING	98.32
Average	93.87

Table 5.12: The comparison of accuracy (%) for similar activities reported by Khan et al. [157] and the developed system using ANN

Activity	Linear classifier		Hierachical classifier)	
	[157]	Our method	[157]	Our method
Sitting	74	99.42	95	99.96
Walking	74	89.13	99.00	98.32
Lying	95	100	99	99.95

6

ENTROPY FEATURE FOR BENGALI NUMERALS RECOGNITION

6.1 INTRODUCTION

Handwritten digit (numeral) recognition (or classification) is an active topic of optical character recognition (OCR). OCR is an active field of research in pattern recognition, artificial intelligence, and computer vision, and it is a common method of digitizing texts of hard copy into soft copy so that they can be electronically edited, searched, stored more compactly, displayed on-line, used in machine processes such as text-to-speech translation, text data extraction, text mining [173]. The text may be composed of only alphabet and/or numerals.

In OCR applications, the recognition of numerals deals with postal code reading for automatic mail sorting, reading amounts from bank check, number plate identification, extracting numeric data from filled in forms, etc [174]. The typical requirements [175] of an acceptable recognition system are summarized as follows:

- "Writer independent". The system should be able to recognize the writing of any person independently on age, sex, style of writings.
- "Size and shape independent": The system should recognize numerals of any size and shape.
- "Less noise sensitive": The system should be highly robust to the presence of noise or varying background.
- "Low error rate": The system should have very low rate of errors.
- "High speed": The system should have very small response time for commercial applications.

Bengali [176] (or Bangla) is the native language of the people of Bengal, which is comprised of Bangladesh, the Indian states of Westbengal, Tripura, and Southern Assam,

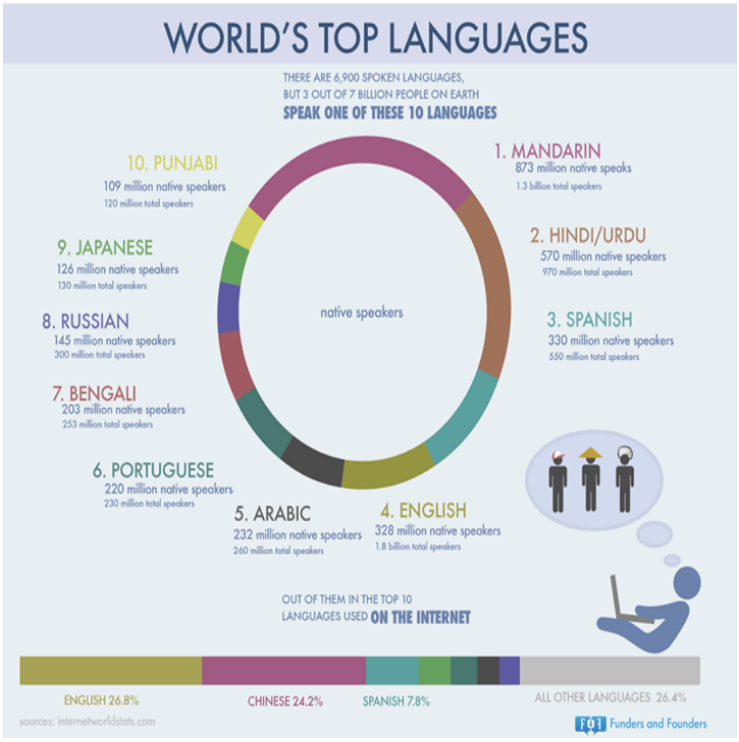


Figure 6.1: The ranking of worlds top languages based on native speakers. Source: https://www.facebook.com/bangladesh.usembassy/photos

and is also the official language of these states. UNESCO has declared 21st February as the "international mother language day" [176] in recognition of the deaths of people scarifying their lives for the sake of their Bengali language in 1952. It is 7th (*i.e* how many people speaks) most popular language in the World [177] with nearly 200 million people speaking in Bengali. The ranking of Worlds top ten languages based on their native speakers is given in figure 6.1.

Like Arabic, Bangla has its own own number system, consisting of 10 basic symbols. Some samples of Bengali handwritten numerals are given in figure 6.2.

In spite of its polarity, unfortunately, researches in Bengali character recognition did not achieve 100% accuracy with reliability so far, in particular on handwritten recognition issue. Some research works on handwritten Bengali numerals recognition have been reported in the literature [178, 179, 180, 181, 182]. To care for the special properties of the numerals and writing styles in Bengali, some researchers have designed specific methods, while others have used the existing generic character recognition methods [183].

Due to the high variability in handwritten styles, extracting robust features with respect to the variation of character shapes and sizes are the most important task for getting higher accuracy. Using large set of features provides higher accuracy, but makes the system computationally expensive. In practice, we cannot reject the variability of handwriting style. The challenge is to extract features such that they overlook the intra-variability (*i.e.* the difference among the samples of same class) and are neverthe-

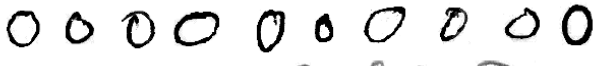
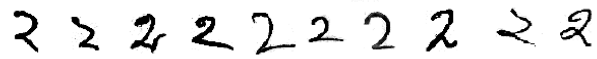
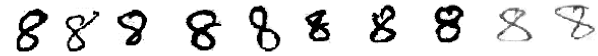
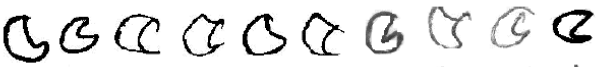
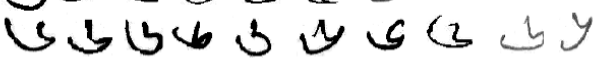
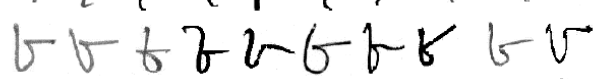
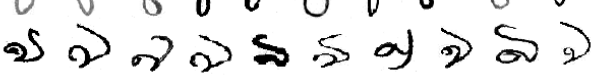
Arabic digits (numerals)	Bengali numerals (word)	Bengali numerals	Some samples of handwritten Bengali numerals
0	Shunya	০	
1	Ek	১	
2	Dui	২	
3	Tin	৩	
4	Chaar	৪	
5	Panch	৫	
6	Chhoi	৬	
7	Shat	৭	
8	Aat	৮	
9	Noi	৯	

Figure 6.2: Samples of printed and handwritten Bengali numerals. The symbols of Arabic digits (or numerals) are shown in the left most column. The Bengali numerals corresponding to each Arabic numeral are shown in word and symbols in columns second and third , respectively from the left. The right most column contains 10 samples of each handwritten Bengali numerals.

less sensitive to the inter-variability (*i.e.* the difference among the samples of different classes) with respect to the writing styles, thickness and size of the handwritten Bengali numerals.

However, in order for a recognition system to be acceptable in practice, the response time besides the accuracy of the developed system needs to be considered. The existing methods in the literature have given emphasis on recognition accuracy, without considering computation cost. In this paper, we have focused on the methods for extracting shape outline [184] based features, giving emphasis on reducing the feature space dimension and computational cost, as well as increasing the recognition accuracy so that the system can be implemented in the low power computers and smart-phones.

The emphasis has been given on reducing feature space dimension and computational costs besides the recognition accuracy. Instead of looking for which classifier may give maximum accuracy, artificial neural network (ANN), which has been used mostly in the literature has also been used here for classification. Most of the published works in this research area have considered discrete individual databases of different sizes. The authors of [180] have considered computational cost besides accuracy, but without evaluating their method on any common database.

Human recognizes characters as images of certain shapes. The learning of human do not depend on any mathematical feature derivation. In this study, our goal is to derive features that mostly represent the shape outline of numerals. The more difficulty

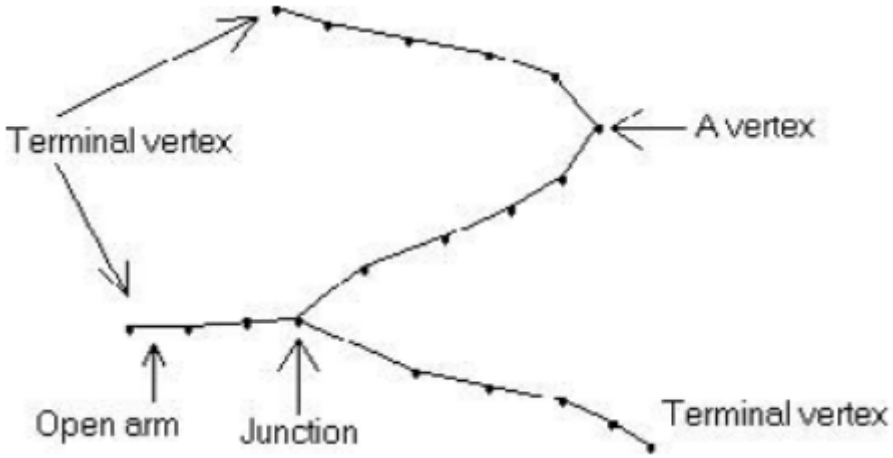


Figure 6.3: The graph representing the topology of numeral 2 and its relevant parts. A junction is a vertex with 3 or more neighbors. A terminal is the vertex with only one neighbor. An open arm is a link between terminal vertex and its neighbor.

with handwriting recognition is the variability of shape, size, as well as inclusion or deletion of extra strokes, implies confusion in their recognition. Our objective is to extract simple and small set of features with less computational costs considering all constraints, and evaluating their performance on a common database, on which some works with higher accuracy have already been reported recently.

6.2 EXISTING FEATURES

The techniques for extracting features from the numerals can be broadly classified into structural and global analysis [185]. In structural analysis, topological and geometrical features are mostly investigated by the researchers. These features include loops, junctions, directions, strokes, shadow, longest runs. On the other hand, the global analysis directly takes into account the shape matrix to find the features of the numeral. An example of this method is template matching. This type of technique suffers from the sensitivity to noise, and is not adaptive to differences in writing style [185].

- “Structural features". Structural features are obtained from a graph representing the topology of the numeral. Some structural features of numeral 2 (Dui), described by [178] is shown in figure 6.3.

  From this graph, structural features like the horizontal and vertical distances between junction and terminal (top, bottom, left, or right) nodes,the presence or volume of cycles (or loops), the slope of open arms are computed.

- “Morphological features". The features obtained by applying different morphological operations [186] (*e.g.* opening and closing on the image [186]). The effect of opening is to preserve that parts of the foreground that matches with the shape of structure element and removing the other parts of the foreground pixels. Similarly, the function of closing operator is to preserve the background pixels that have the same shape of the structuring element and removing the other back-

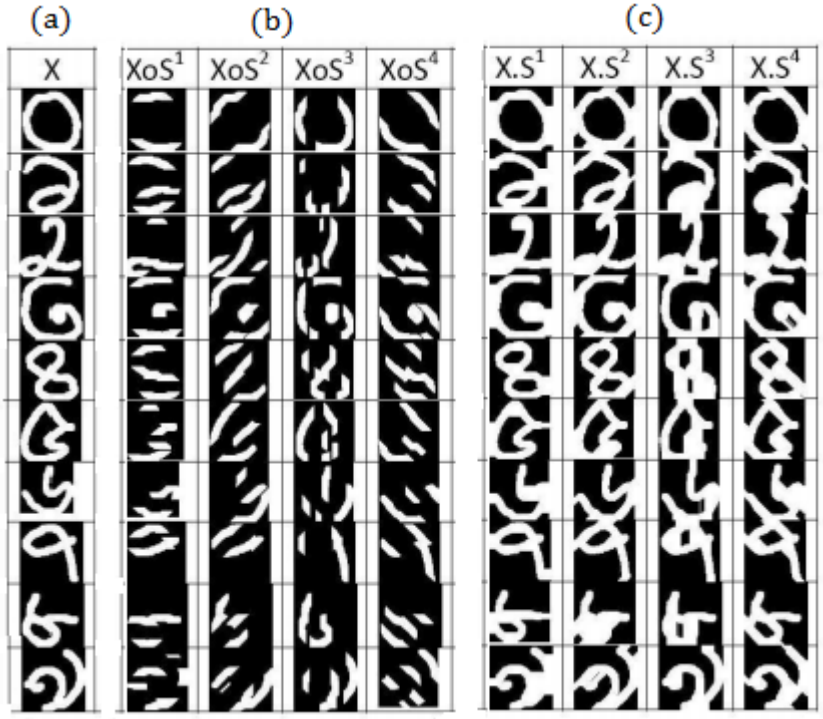


Figure 6.4: The effects of applying directional morphological opening and closing on handwritten numerals. Panel (a): handwritten numerals, panel (b) after opening, and panel (c) after closing operations in four directions: horizontal, left diagonal, vertical, and right diagonal.

ground pixels. Purkait et al., [187] has described the effect opening and closing operations (figure 6.4) on 10 samples of Bengali numerals.

- "Topological features". Topological features represent the information characterizing the topology of an object [188]. The usual topological features for Bangla numeral are the number of close loops, the length of connected components, the position of loops, etc.

- "Contour features". A contour of an image is defined as the foreground pixel, p, such that any of its $n \times n$ neighbor is a background pixel. Each neighbor (to the right (horizontal), slanted by 45 degree, vertical, and slanted by 135 degree) of the contour is assigned a code, called chain code. The frequency of chain code is considered as contour features.

- "Gradient strength features". To obtain gradient strength features, gradient image is generated from a gray scale image by applying a Roberts filter [189]. Then the arc tangent of the gradient is quantized into a number of directions, and the strength of the gradient is accumulated with each of the quantized directions [190]. For any gray scale image g(x,y), the strength of gradient, *i.e.*

$$E_{\Delta} = \sqrt{\{g(x+1,y+1) - g(x,y)\}^2 + \{g(x+1,y) - g(x,y+1)\}^2} \tag{6.1}$$

6.3 EXISTING METHODS

Several algorithms with many features, which make the system complex, have been reported in the literature, even though 100% accuracy is not achieved yet. A few research works with high accuracy on handwritten Bengali (Bangla) numeral recognition system is briefly explained here.

Bhattacharya et *al.*, [178] proposed recognition of handwritten Bangla numerals using neural network models. A skeletal shape, (represented as a graph) was first extracted from the numeral pattern using topology adaptive self organizing neural network. Features like loops, junctions,$\cdots$, etc were extracted and multilayer perceptron (MLP) neural networks was then used to classify the numerals. They obtained about 90% recognition accuracy on a test set of 3440 samples.

Pal et *al.*,, [179] proposed a system for Bangla handwritten numeral recognition. They considered features (like direction of overflow, height, position of the reservoir with respect to bounding box of the numeral) based on the concept of water overflow from the reservoir along with structural and topological features (such as the number of close loops, their position, the ratio of close loop height to component height, $\cdots$, etc) of the numerals. The overall recognition accuracy of the system achieved was about 92.8% on a test dataset of 12000 samples.

In 2007, Pal and co-authors [190] further published results of handwritten numeral recognition of Bangla with five other Indian scripts. They used quadratic classifiers on 16-direction gradient histogram features using Robert masks. The highest recognition accuracy was 98.99% on a test set of 14650 samples.

Recognition of handwritten Bangla numerals using hierarchical Bayesian network was proposed by [181]. They used the original images of the numerals, instead of extracted features, directly at the input of the network. They reported an average recognition accuracy of about 87.50% on 2000 untrained images.

The research works on Bangla handwritten numerals found in the literature, seldom used common sample database, due to its unavailability. Even after a database (ISI Bangla handwritten numerals database) [184] was publicly available, a few researchers [183, 191, 187] have evaluated their methods using it.

Bhattacharya and his colleague [191] have developed a pioneer system for handwritten numeral recognition evaluated on the ISI Bangla handwritten numeral database. They used multiresolution wavelet analysis for feature extraction.

Daubechies [192] wavelet filter is applied to a binary image of size $L \times L$, where L must be a power of 2. The first application of Daubechies wavelet-4 filter, produces four image components L|L, L|H, H|L, and H|H each of size $\frac{L}{2} \times \frac{L}{2}$, correspond to low frequencies in both horizontal and vertical directions, low frequencies in horizontal and high frequencies in vertical, high frequencies in horizontal and low frequencies in vertical, and high frequencies in both horizontal and vertical directions, respectively. The Daubechies filter is successively (k times) applied on the L|L components. Thus k sets of four image components corresponding to k fine to coarse resolution levels are obtained. Now, chain code histogram features [193] are extracted from each of the detail (L|L) image components for each level of resolution.

To compute features from the three other components ($L^k|H^k$, $H^k|L^k$, and $H^k|H^k$) at each resolution level k, the bounding box of each of them is divided into ($l \times l | l$ is a power of 2) equal size blocks. Then the ratio of number of black pixels to the total number of pixels for each block is quantified. Finally, a feature vector is constructed by concatenating the features for all components of the wavelet filtered image. For L=128, k=3, l=3, a feature vector of 256 values is obtained. They used a distinct multilayer perception classifier (MLP) for each resolution level. They reported recognition accuracies of 99.14% and 98.20% for training and test samples, respectively on ISI database.

Liu et *al.,* [183] have provided the new benchmark for comparison of Bangla handwritten recognition methods on ISI standard database. They have used three normalization techniques on both binary and gray-scale images. Hence after normalization, they considered gray-scale, normalized binary, and a binary image normalized to gray-scale (i.e. pseudo-gray). They formed a feature vector of size 200 from the normalized numeral using 8-direction gradient feature. A MLP with one hidden layer of 100 nodes was then used for training. In this method, they reported an average accuracy 98.69%. The highest average recognition accuracy of 99.16% was obtained using both class-specific feature polynomial classifier (CFPC) [194] and support vector machine (SVM) [195]. However, they got an average accuracy for normalized binary image of about 98.46%.

Purkait et *al.,* [187] performed another recognizable work on the same ISI database. They used some morphological (i.e., directional opening, directional closing, directional erosion obtained by applying morphological operations) as well as K-curvature features for handwritten numeral recognition using MLP classifier. "The left and right K-slopes at any point P on a curve are defined as the slopes of the line joining P to the points K steps away along the curve on each side of P, and K-curvature of P is the difference between its left and right K-slopes" [196]. The best recognition performance (96.25%) was obtained for morphological opening feature set. However, this figure of recognition was increased to 97.75%, when the classifiers were fused using a modified naive -Bayes combination.

A list of features with higher accuracy (ACC) reported for Bengali digit recognition is summarized in table 6.1.

There are tremendous progress accuracy in the recognition of handwritten Bengali numerals, but with large set of features. The methods with high accuracy, in the literature have used a relatively high dimension (more than 200) features that might take more computing time. A system required less computing power is really helpful for its implementation on Smartphones or low-cost computers in real time environment.

This study is focusing on methods for extracting features, which are able to capture the more representative information of the same class, regarding the variability of the numerals, with less computational cost. The accuracy of a recognition system depends on the test dataset and it is difficult to compare the performance of two systems on two different test database. In this study, the accuracy of the developed system will be evaluated by the Bengali handwritten numerals on ISI database [184], which is the largest publicly available database (to the best of our knowledge) and have already been used by some other researchers. So, objective of this study is to extract robust

Methods	Features	Feature size	Database	Size	ACC(%)
Bhattacharya [178]	Structural features: cycles, junctions, and number of terminal nodes	-	Personally collected	5320	90.56
Pal et al. [179]	Topological features: number of close loops, center of gravity, and the ratio of close loop height to the height of the component	-	Personal collection	12000	92.8
Liu et al. [183]	Direction gradient feature	200	ISI	23392	98.46
Pal et al. [190]	Contour and gradient-strength	400	Personal collection	14650	98.69
Bhattacharya & Chaudhury [191]	Wavelet based chain-code features	256	ISI	23392	98.20
Purkait and Chanda [187]	morphological and structural features	500	ISI	23392	97.75

Table 6.1: Common features for Bengali digit recognition, where ('-' means that the size is not mentioned.)

features and evaluate its performance on the publicly available large database. The features are based on the shape outline of the numerals [184], giving emphasis on the recognition accuracy, reliability, as well as computational cost.

6.4 DATA AND METHODS

The ISI handwritten Bangla numeral database [184], consists of 23392 samples for training and 4000 samples for testing has been used in this study. The number of training samples of each class varies slightly, but the number test samples for each class is fixed to 400. The testing samples were randomly selected. The sample images are gray scaled, with noisy background and considerable variation in foreground (stroke regions). Some samples are shown in figure 6.2.

The proposed classification system is summarized by the block-diagram in figure 6.5

6.4.1 THRESHOLDING & PRE-PROCESSING

The raw images in the database are gray scaled with 256 levels, and they are contaminated mostly by peeper noise during scanning. The raw digit image is first converted to a binary image using Otsu's [197] thresholding method. Then median filter has been

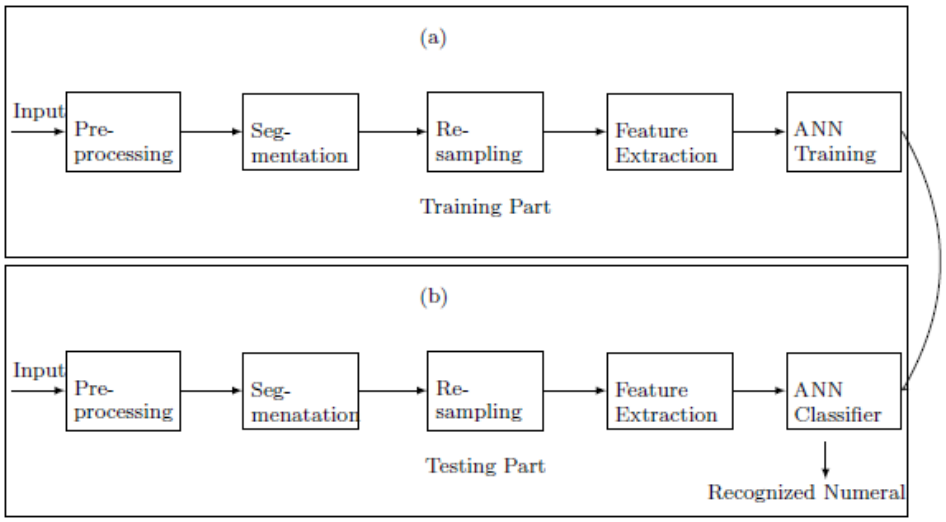


Figure 6.5: Block diagram of the classification system. Panel (a): the training of ANN using train dataset. After training, the recognition accuracy is tested using the samples from test dataset in panel (b).

applied in both horizontal and vertical directions to remove peeper noise present in the binary image.

6.4.2 SEGMENTATION

After pre-processing, the boundary of the binary image is extracted using the horizontal and vertical pixel scanning method [198]. In horizontal scanning, the first row containing any black pixel is considered as the top boundary and hence indicating the starting of the region of interest (*i.e.* the foreground image of the numeral). This horizontal scanning is continued until a row of all white pixels are found. The immediate previous one of the row of all white pixels is the bottom boundary of the region of interest. Now the region between top and bottom boundary is scanned vertically for any column contains at least one black pixel and the left boundary is detected. This scanning is continued until a column of all white pixels is detected. The immediate previous column of the column of all white pixels is defined as the right boundary of the region of interest. In this way, the foreground image is separated from the unwanted region.

6.4.3 RESAMPLING

Resampling is an important step of any pattern recognition algorithm. Handwriting numerals have no specific size. To extract features, every pattern should have same dimension, and hence the binary image of black (0) and white (1) pixels has been resampled to a fixed resolution of size $R \times R$. The raw digit and its resampled (32×32 form of a sample of Bengali digit five (Panch) is shown in figure 6.6.

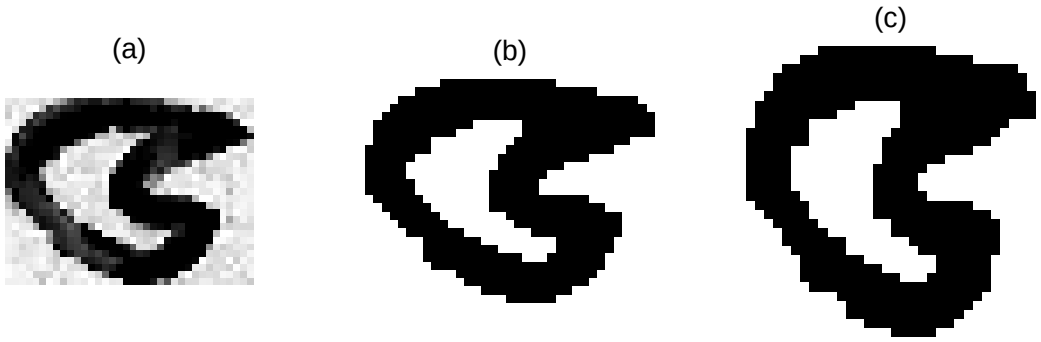


Figure 6.6: Sample of digit five (Panch) in different steps of processing. Panel (a): the original digit. Panel (b): the digit after pre-processing panel (c): the digit after resampled to 32×32. The pepper error in panel (a) has been completely removed after filtering without any major shape distortion of the digit.

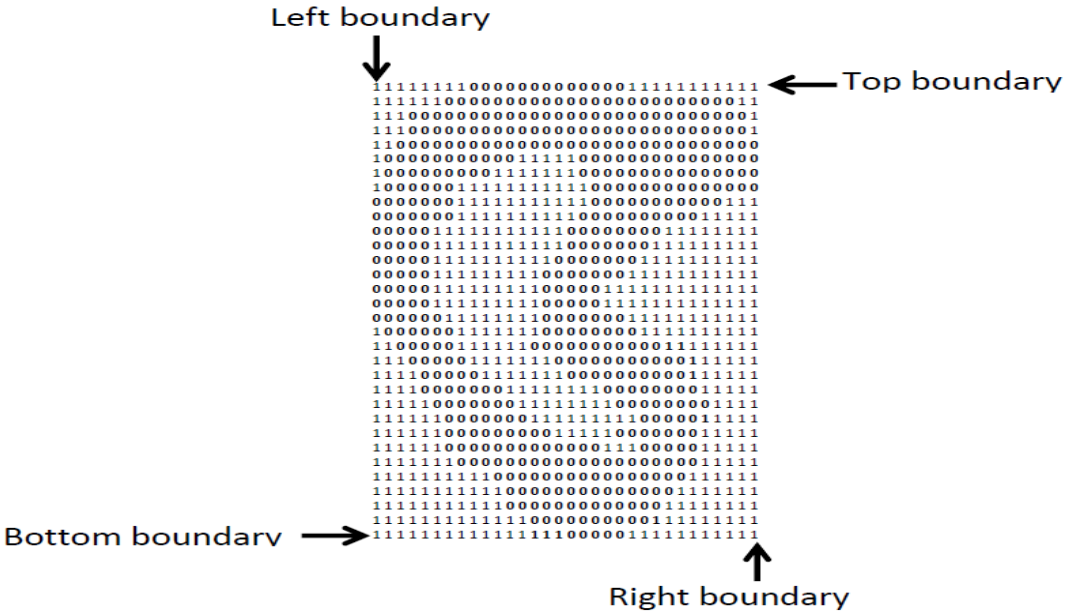


Figure 6.7: The binary form of the resampled digit of figure 6.6. The 0's and 1's represent the black and white pixels, respectively. The distance of the surface edge (black) black pixels ('0') from the bottom boundary are [15 ,14, 13, 12, 10, 9, 6, 5, 5, 5, 4, 2, 2,1, 1, 1,0,0,0,0,0, 1,1,2,3, 4, 5, 9, 23, 24, 24, 24] (from left to right).

6.4.4 FEATURE EXTRACTION

We have introduced a set of new features in this study. The features are extracted from the binary image. A binary image of a sample of digit five (Panch) is shown in figure 6.7 in order to ease the way of explaining the features. The features and their extraction methods are described below:

6.4.4.1 BLACK RUNS

A black run [199] in any binary pattern is defined as the consecutive sequence of black pixels. When a binary pattern is traversed from one side to opposite side, a

Figure 6.8: The negative of digit five (Panch)

series of 1's and 0's are faced. The sequence of such continuous 0 (zero) forms a black run. The number of black runs were computed for each row and each column of the normalized binary digit. In other words, if a horizontal straight line is passing from top to the bottom of a numeral, the number of intersections of the line with the black pixel for each row, represents the black runs of that row. Similarly, the number of intersections of a vertical line with the numeral at each column represents the black runs for that column. The row-wise black runs (Row_{BR}) of figure 6.7 is $Row_{BR} = [1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,2,2,3,3,2,2,2,2,2,2,2,1,1,1,1]$. Similarly, a vector of column-wise black runs (Col_{BR}) is obtained. Thus a feature vector of 64 black runs is formed by combining Row_{BR} and Col_{BR}.

6.4.4.2 EDGE DISTANCE

Each numeral has it's individual basic structure, which may be deformed by adding (or removing) stroke, hole, curves, etc due to writing styles, shown in Fig. 6.2. Our target was to extract features which trace the basic shape irrespective of this deformation. If a resampled binary numeral is bounded by a virtual square, such that the most outer rows and columns fall over the boundary of the square, shown in Fig. 6.6, then the distance (*i.e.* the number of white pixels in a row (or column) from the adjacent boundary to the surface black pixel (*i.e.* the edge pixel) is defined as the edge distance of that blak pixel. The edge distance of a black pixel from the adjacent vertical boundary (leftmost and rightmost columns) and horizontal boundary (topmost and bottommost rows) are referred to as vertical and horizontal edge distances, respectively. Thus, for a numeral of size $R \times R$ there are $4 \times R$ edge distance. The negative image of digit five (Panch) and its edge distances are shown in figure 6.8. The edge distances are shown by the black strips from the adjacent boundary. The edge distance for adjacent surface pixels do not differ so much. That is why, to keep smaller feature dimension only pixels on the odd numbered rows and columns were considered for measuring the boundary distance. Thus, we have counted a set of $4 \times R/2$, instead of $4 \times R$ edge distances for a numeral of size $R \times R$. This edge distance might have less intraclass and more interclass variations for Bengali numerals.

6.4.4.3 SLASH(/)-BAKSLASH(\) STROKES

This feature extraction technique is based on approximating a side of the numeral by a set of straight line segments. In this study, the straight line segment whose slope is more than 90 degree is defined as *backslash* (\) and whose slope is less than 90 degree is defined as *slash* (/). Each side of the Bengali numeral is represented by a collection of slash (/) and backslash (\) straight line segments, which is extracted from the edge distance (ED_{side}, where $side$ can be either bottom, left, top or right) of the numeral.

To extract slash-backslash feature for any side of the Bengali numeral from the binary image, the derivative (ED'_{side}) of the ED_{side} is computed by taking the subtraction of the successive edge distance at each row (or column) first, *i.e.* $ED'_{side}(i)=ED_{side}(i+1)-ED_{side}(i)$, for $1 \leqslant i \leqslant R-1$. The values of ED'_{side} are either positive ('+VE', i.e., any integer greater than 0), negative ('-VE', i.e., any integer less 0), or constant (0). The vector ED'_{side} is traversed from one end to another to detect the positions of changing (POS_{change}) their values from either (0 or -VE to +VE) or (0 or +VE to -VE. When a POS_{change} is detected at a row (or column), the immediate previous rows (or columns) having ED'_{side} values with same sign (+VE and 0 or -VE and 0) upto this POS_{change} defines a line segment (slash or backslash). The continuous part of ED_{side} for which $ED'_{side} \leqslant 0$ is approximated by backslash (\) and $ED'_{side} \geqslant 0$ is approximated by slash. Thus, the backslash and slash can be approximated from ED'_{side} by the regular expressions -VE(-VE|0)* and +VE(+VE|0)*, respectively. The extraction of slash-backslash is explained in figure 6.9.

6.4.5 CORRECTED CONDITIONAL ENTROPY

CcEn is a measure of entropy of a time series, which is of dimension 1. On the other hand, an image is a 2D signal. To compute entropy of a numeral a trick has been adopted here. We know, the estimation of corrected conditional entropy (CcEn) requires converting the time series into a symbolic sequence of two or more symbols. The binary numeral is already a pattern of 0's and 1's. Now the trick is to convert a 2D binary pattern into a 1D symbolic sequence. To do this, the symbols of each row (from top to bottom) are concatenated into a vector. Thus we get a symbolic sequence of 0's and 1's. Then the CcEn of this binary sequence is estimated, as described in section 2.4.1.5 of chapter 2 with the value of free parameter m=2.

Thus for a resampled image of size $R \times R$, we get an augmented feature vector of $2 \times R$ blackruns, $2 \times R$ edge distance, 8 slash (/)- backslash (\), one CcEn features.

6.4.6 FEATURE DIMENSION REDUCTION

Feature extraction may consist another additional step of feature selection. In the feature extraction step, information relevant to the signal classification is extracted from the input data first and form a D-dimensional feature vector V. In the feature selection step, the vector V is transformed into a vector, which has the dimensionality D_T(D_T<D). If the feature extractor is properly designed so that the feature vector is matched to the

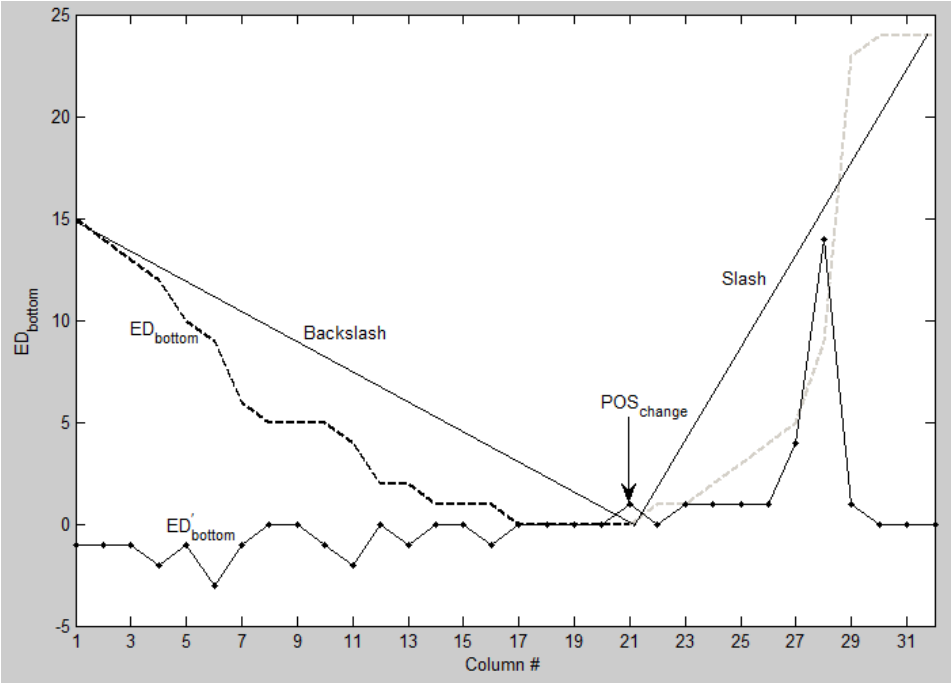


Figure 6.9: The slash and backslash for the bottom side of Bengali numeral Panch. The dashed line shows the values of ED_{bottom} and dot-dashed line represents the values of ED'_{bottom}. POS_{change} is the column (column# 21), at which ED'_{bottom} is +VE. The values of ED'_{bottom} upto column 20 is either 0 or -VE. This segment of the edge as shown by the black dashed line is approximated by the backslash (\). The values of ED'_{bottom} at and after 21st column is either 0 or +VE. This segment of the edge as shown by the light dashed line is approximated by the slash (/).

pattern classifier with low dimension. Then, there is no need of feature selection. However, the feature vectors have to be decorrelated before feeding them into the classifier. Principal component analysis [200] is a popular algorithm for feature dimension reduction. It extracts relevant features by projecting the feature vector into a new feature space through a linear transformation matrix. Principal component analysis (PCA) optimizes the transformation matrix T, by finding the largest variations in the original feature space. In this study, we have applied PCA to reduce the feature space.

6.4.7 CLASSIFICATION

A FFNN has been trained by the aggregate set of features to classify the input numerals. The extracted features have been used for training the FFNN with 4 nodes in the output layer. The number of hidden nodes, which maximized the accuracy was selected. The recognition accuracies on both training and testing sets of ISI database have been observed. The FFNN has been trained using 10 fold cross validation (*i.e.* the whole data set is divided into 10 equal partitions, where the samples in each partition are selected randomly. Then, the samples of one partition is used for testing and the samples from 9 other partitions are used for training the neural network. This procedure is repeated for training and testing the neural network with 10 possible combinations.

In every repetition, the neural network is reinitialized during training. In addition, the recognition accuracy is also tested using the test dataset of ISI database, when training is performed using the training dataset of ISI database.

6.4.8 RESULTS FOR BENGALI HANDWRITTEN NUMERALS CLASSIFICATION

This section is dedicated to the results obtained for feature extraction and classification of Bengali handwritten numerals. The results regarding the number of neurons is presented first, then the accuracy of the system is illustrated.

6.4.8.1 NUMBER OF HIDDEN NEURONS

The choice of an optimal number of neurons in the hidden layer has been made through an extensive simulations. The accuracy is observed for a range of neurons in the hidden layer, which gives the maximum value of accuracy. The number of nodes used in the hidden layer is varied from 10 to 40 (at an interval of 5). The optimal number of nodes found for the hidden layer is 30 with maximum accuracy.

6.4.8.2 CLASSIFICATION ACCURACY

The database is divided into two parts: training set and test set. The recognition accuracy is observed for both training and testing. The performance of the system is evaluated by two parameters: accuracy (ACC) and reliability (k). The performance of the recognition system on training dataset is evaluated using 10-fold cross validation technique. On the other hand, the FFNN is trained using the entire training data, and then its performance is evaluated on the test dataset. The effects of resolution of the numerals obtained after re-sampling has also been investigated.

- “Using the full features set". The recognition accuracies obtained on training database are 95.64%, 97.43%, and 97.10%, respectively for resolutions 16×16, 32×32, and 64×64, giving the recognition accuracies for test database 95.48%, 97.69%, and 97.06% using full set of features. A confusion matrix of the recognition accuracy on test database with R=32 is given in table 6.2 to show misclassification besides the true recognition of numerals.

- “Using PCA features". The application of principal component analysis (PCA) reduces the feature dimension from 73, 137, 265 to 31, 47, and 168, respectively for numerals of size 16×16, 32×32, and 64×64. Thus the use of PCA reduces the feature dimension by more than 50%. The accuracy obtained with these reduced set of features are 93.21%, 95.97%, and 97.06%.

- “Reliability of the classification". The reliability of the classification is evaluated by the value of k. The reliabilities obtained with 31, 73, and 168 features are 92.76%, 97.30%, and 97.20%, respectively. Thus, the best reliability (also the highest accuracy: 97.69%) is achieved with resolution 32×32.

Table 6.2: Confusion matrix of the classifier. Columns correspond to target class and row corresponds to the target class.

	(Recognized class)									
-	0	1	2	3	4	5	6	7	8	9
0	395	0	0	0	1	2	1	0	0	0
1	0	381	0	0	2	1	1	0	0	15
2	0	0	396	0	0	0	4	0	0	0
3	1	0	0	388	2	5	3	0	0	
4	8	0	1	0	389	2	0	0	0	0
5	1	0	1	0	5	388	3	0	0	0
6	0	1	2	3	1	6	387	0	0	0
7	1	1	1	1	0	0	0	395	0	1
8	2	0	1	1	0	0	0	0	396	0
9	0	13	0	0	0	0	0	3	0	384

To observe the influence of the entropy feature (CcEn) in the recognition accuracy, the FFNN was trained and tested using all but CcEn features. In this case, the accuracy obtained was 95.64% (using features except CcEn) instead of 97.69% (including CcEn in the feature set).

6.4.9 EVALUATION ON BENGALI HANDWRITTEN NUMERALS CLASSIFICATION

A set of new features has been introduced in Bengali numeral recognition. To the best of our knowledge, the entropy concept and balck runs have been used for image segmentation. The slash (/)- backslash\ and CcEn in optical character recognition are completely new concept.

The classification method based on this new set of features correctly classifies 97.69% of test samples of Bengali handwritten numeral database. Out of 4000 test samples, the number of correctly classified, mis-classified, and unrecognized is 3899, 97, and 4, respectively. It is observed from Table **??** that the maximum number of mis-classifications is happened between digits 1 (Ek) and 9 (Noy).

The use of CcEn increased the overall classification accuracy with reducing the frequency of mis-classifications between one ad nine, followed by a little increase in the mis-classification between 0 (Shunnya) and 4 (Chaar). The highest accuracy of 99% is obtained for digits 2 (Dui) and 8 (aat). The worse recognition rates of 95.25% and (96%) are found for digits 1, and 9, respectively.

The accuracy increases with increasing the resolution upto 32×32. The less accuracy at smaller resolution is reported, this is due to the fact that the features are dependent on the numeral's shape outline, which is affected by down sampling the numeral (figure 6.10).

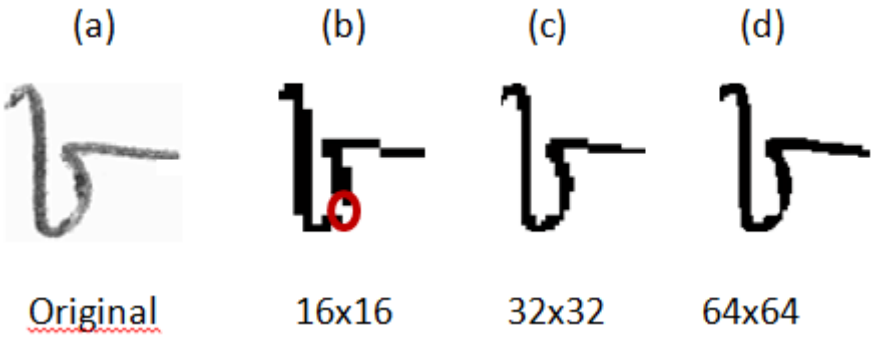


Figure 6.10: The digit eight (aat) with 3 different resolutions. The original digit is dispalyed in panel (a). Panels (b), (c), and (d) represents the resampling of digit eight. The red circle in panel (b) marks the missing of a pixel due to under sampling.

6.5 SUMMARY

Handwritten numerals recognition is an important topic in pattern recognition and optical character recognition research areas. Bengali is one of the mostly spoken (by the native speakers) languages in the world. Research in Bengali numeral recognition has been progressed, but still away from 100% accuracy. Those methods with high accuracy are using high dimension of features, with more computation power. Some methods with high accuracy are reported on discrete (personally collected) database.

In this chapter, the literature on handwritten Bengali numerals recognition has been reviewed briefly. A set of new features including the CcEn, have been proposed in this study. After feature extraction and feature selection, FFNN has been trained and tested on a large publicly available database. The performance of the proposed system has been compared with the existing methods on the same database with high accuracy reported in the literature.

7

CONCLUSION AND FUTURE WORKS

In this chapter, a short summary and conclusion of the described works are presented, along with a dictation of possible future works.

7.1 CONCLUSION

The work described through this study has the objective of researching innovative methods for extracting entropy and related features, and then using them in real classification problems.

Since many measures of entropy are available in the literature, a preliminary study on entropy has been performed. A set of most commonly used measures of entropy has been selected for this research. Even though they are popular metrics, their estimations are extremely sensitive to the series length. Unfortunately short series are generally used in real applications. A related problem arises in spectral analysis, and parametric spectrum analysis using AR models is commonly performed in this regard with reasonable stationary. In this research, parametric estimations of entropy through AR models have been proposed. Analytical expressions for ApEn, SampEn, and CEn of an AR model have been derived as well. Since the numerical estimates of these measures are sensitive to the series length, we can get the deviation of numerical estimation from the theoretical one. Besides series length, the estimations of these metrics also require the selection of embedding dimension 'm' and tolerance of mismatch 'r' between the templates constructed from the series. Although, Lake et *al.*, already have tackled the problem of deriving analytical formulas of ApEn and SampEn of an AR process in the limit $m \to \infty$. We have derived analytical expressions for entropy of an AR process for finite m in our method. The theoretical values of entropy for any m larger than the model order can also be derived. The theoretical values of SampEn (derived by our method) converges with the Lake's estimation for any m greater than the model order (M).

The numerical estimation of SampEn is undefined for very short series. On the other hand, ApEn is defined at any series length with accepting the biasing of it's estimate. The estimates of ApEn and SampEn for short series may be far away from what expected for long series. The feasibility of parametric estimations of entropy on both synthetic series and real data suggests that the parametric estimation of entropy is possible for very short and artifacts free series, for which numerical estimates are unreliable or even undefined. The entropy of an AR process is directly related to its autocorrelation function. The comparison of parametric estimates of entropy with numerical ones provides some more information about the signal characteristics. When numerical and parametric estimates of the entropy do agree, it means that entropy is mainly influenced by the linear properties of series. On other hand, when the AR model is fitted well, any disagreement between numerical and parametric estimates of the metric implies that the entropy is truly offering some information that cannot be captured by the traditional temporal or spectral parameters. Thus, the method also offers a tool for statistical analysis in addition.

Preliminary methods for extracting features, in particular for sleep classification and handwritten numerals recognition have been studied, since prospective research in these fields are advanced robust feature extraction, and classification. A set of entropy features in addition to some related existing features have been studied in this research to solve a number of classification problems, and the relevance of entropy features has been verified for these cases, especially in sleep state classification from HRV analysis, physical activity recognition, where the states are changed in very short period.

In particular, the researched methods for the automatic sleep stage classification from only HRV signal analysis appears prospective with a focus on the distinction of wakefulness from sleep, and NREM from REM. The regularity based (SampEn) features are found as one of the most significant among the features extracted from the HRV series for sleep classification. Apart from increasing the overall classification performances, they give also information about the physiology of sleep, in particular as regards NREM stages. These findings pave the way to further investigations of the behavior of the autonomic nervous system during sleep. Besides this, the parametric estimates of SampEn have been selected as one of the best relevant features for physical activity classification.

With regard to the Bengali handwritten numeral recognition, a set of completely new features have been derived for Bengali handwritten numerals. The use of entropy (in particular CcEn) feature has been shown effective in this classification problem, which is completely a new application of CcEn. The developed system has obtained more or less same recognition accuracy using very small and computationally inexpensive systems.

The novelty of this research are developing methods for parametric estimations of entropy, extracting entropy based features, which are proved effective in some classification problems, such that the use of these features with some existing ones give more accuracy with very small number of features, and hence reduced cost of the systems.

7.2 FUTURE WORKS

Many different aspects of the researched methods could be considered in order to increase further the accuracy and usability.

First, only three measures of entropy has been considered, the parametric estimation of other entropy measures like spectral entropy and transfer entropy can be considered. The use of entropy related features have been shown only in HRV, acceleration signals, and Bengali handwritten numeral recognitions. The usability of these features can be verified in other physiological signals such as electroencephalogram (EEG) and speech signals processing.

A few experiments about the complexity behavior on real data (*e.g.* HRV) have been performed in this study. In future, the impact of the proposed complexity analysis method on more experimental data should be considered to establish a procedure which might be able to properly synthesize their complexity behavior of real data (both short and long series) by distinguishing linear versus nonlinear mechanisms with a reasonable accuracy.

In many cases of the dynamic analysis in chaotic systems, entropy contains much relevant information which can be mostly obtained using other relevant measures such as detrended fluctuation analysis, empirical mode decomposition, Lempel Ziv complexity analysis, etc. In fact, we have compared sample entropy measure with Lempel-Ziv complexity on a small experimental setup in this work. A deep comparison between entropy and the related metrics such as can be studied on a large scale of experiments in future.

Segmentation has become one of the most important problems that must be solved before classification or recognition of objects or signals (*e.g.* voice recognition from the continuous array of recordings, digit or character recognition from continuous scripts, etc). Although entropy based segmentation methods have been proved robust in texture segmentation, text classification, and image segmentation. In future, the use of entropy metrics can be extended in the segmentation of patterns of short segments of data form long recordings such as HRV tracings in Holter-tape recordings.

Only a few states from few subjects have been considered for physical activity recognition. To justify the true effectiveness of entropy features in physical activity classification large data should be considered. The system can be enlarged for all daily activities. The selected features can be applied for developing systems for e-healthcare of elderly or physically impaired persons.

With respect to the character recognition, only handwritten Bengali numerals have been considered. The research can be extended to handwritten Bengali script recognition, along with other related Indian languages like Hindi or Devanagari. The concept of entropy feature can be extended to apply other related pattern recognition problems. Due to less computational powers required by the developed systems, applications using these features can be developed for implementing on Smartphones.

[1] C. Cachin and P. D. U. Maurer, *Entropy Measures and Unconditional Security in Cryptography*. Konstanz: Hartung-Gorre (1 Jan 1997), 1997. *(Cited on page 2)*

[2] M. Ezhilarasan, P. Thambidurai, K. Praveena, S. Srinivasan, and N. Sumathi, "A new entropy encoding technique for multimedia data compression," in *International Conference on Computational Intelligence and Multimedia Applications, 2007.*, vol. 4, Dec 2007, pp. 157–161. *(Cited on page 2)*

[3] T. M. and J. A. Thomas, *Elements of Information Theory 2nd Edition*, 2nd ed. Hoboken, N.J: Wiley-Interscience, Jul. 2006. *(Cited on pages 2, 11, 12, 15, 21, 23, and 35)*

[4] I. Bialynicki-Birula and J. Mycielski, "Uncertainty relations for information entropy in wave mechanics," *Commun.Math. Phys.*, vol. 44, no. 2, pp. 129–132, 1975. *(Cited on page 2)*

[5] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957. *(Cited on page 2)*

[6] A. Renyi, "On measures of entropy and information," in *Proc. Fourth Berkely Symp. Math. Stat- Prob.*, vol. 1, University of California Press, 1961, pp. 547–561. *(Cited on pages 2 and 20)*

[7] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991, vol. 98. *(Cited on page 2)*

[8] P. Grassberger and I. Procaccia, "Estimation of the kolomogorov entropy from a chaotic signal," *Phys. Rev. A*, vol. 28, pp. 2591–2593, 1983. *(Cited on pages 2 and 17)*

[9] X. Chen, I. Solomon, and K. Chon, "Comparison of the use of approximate entropy and sample entropy: applications to neural respiratory signal," *Conf Proc IEEE Eng Med Biol Soc*, vol. 4, pp. 4212–4215, 2005. *(Cited on page 2)*

[10] K. Nigam, "Using maximum entropy for text classification," in *In IJCAI-99 Workshop on Machine Learning for Information Filtering*, 1999, pp. 61–67. *(Cited on page 2)*

[11] V. E. Kosmidou and L. I. Hadjileontiadis, "Using sample entropy for automated sign language recognition on semg and accelerometer data," *Med. Biol. Eng. Comput.*, vol. 48, no. 3, pp. 255–267, Mar. 2010. *(Cited on page 2)*

[12] A. Porta, S. Guzzetti, N. Montano, R. Furlan, M. Pagani, A. Malliani, and S. Cerutti, "Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1282–1291, 2001. *(Cited on page 2)*

[13] Z. X. L. Min, "Pattern recognition of surface electromyography signal based on multi-scale fuzzy entropy," *J. Biomed. Eng.*, vol. 6, p. 032, 2012. *(Cited on page 2)*

[14] C.-I. Chang, Y. Du, J. Wang, S.-M. Guo, and P. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques," in *IEE Proceedings in Vision, Image and Signal Processing*, vol. 153, no. 6, 2006, pp. 837–850. *(Cited on page 2)*

[15] G. J. Jung and Y.-H. Oh, "Information distance-based subvector clustering for asr parameter quantization," *IEEE Signal Process. Lett.*, vol. 15, pp. 209–212, 2008. *(Cited on page 2)*

[16] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 75–81, 1976. *(Cited on pages 2 and 11)*

[17] E. N. Bruce, *Biomedical signal processing and signal modeling*. Wiley, 2001. *(Cited on page 2)*

[18] J. G. Proakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, N.J: Prentice Hall, 2006. *(Cited on page 3)*

[19] S. J. Redmond, M. E. Scalzi, M. R. Narayanan, S. R. Lord, S. Cerutti, and N. H. Lovell, "Automatic segmentation of triaxial accelerometry signals for falls risk estimation," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2010, pp. 2234–2237, 2010. *(Cited on page 3)*

[20] I. Guyon and A. Elisseeff, in *Feature Extraction*, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds. Springer Berlin Heidelberg, Jan. 2006, pp. 1–25. *(Cited on page 4)*

[21] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 11, no. 6, pp. 601–617, Nov. 1967. *(Cited on page 4)*

[22] C. Alippi, M. M. Polycarpou, C. Panayiotou, and G. Ellinas, "Proceedings of 19th international conference on artificial neural networks," in *Springer Science & Business Media*, Limassol, Cyprus, Sep. 2009. *(Cited on page 4)*

[23] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys.*, vol. 13, no. 1, pp. 183–200, 1975. *(Cited on page 5)*

[24] S. Kay and J. Marple, S.L., "Spectrum analysis-a modern perspective," *P. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981. *(Cited on page 5)*

[25] C. V. J. F. Bercher, "Estimating the entropy of a signal with applications," *IEEE Trans. Signal Process.*, no. 6, pp. 1687 – 1694, 2000. *(Cited on page 5)*

[26] A. Holzinger, C. Stocker, B. Peischl, and K.-M. Simonic, "On using entropy for enhancing handwriting preprocessing," *Entropy*, vol. 14, no. 11, pp. 2324–2350, Nov. 2012. *(Cited on page 6)*

[27] C.-K. Leung and F.-K. Lam, "Image segmentation using maximum entropy method," in *1994 International Symposium on Speech, Image Processing and Neural Networks, 1994. Proceedings, ISSIPNN '94*, 1994, pp. 29–32. *(Cited on page 6)*

[28] A. Lempel, M. Cohn, and W. Eastman, "A class of balanced binary sequences with optimal autocorrelation properties," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 38–42, 1977. *(Cited on page 11)*

[29] E. Plotnik, M. Weinberger, and J. Ziv, "Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the lempel-ziv algorithm," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 66–72, 1992. *(Cited on page 11)*

[30] J. Ziv and N. Merhav, "A measure of relative entropy between individual sequences with application to universal classification," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1270–1279, 1993. *(Cited on page 11)*

[31] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530–536, 1978. *(Cited on page 11)*

[32] E.-h. Yang and J. Kieffer, "Simple universal lossy data compression schemes derived from the lempel-ziv algorithm," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 239–245, 1996. *(Cited on page 11)*

[33] P. Grassberger, "Estimating the information content of symbol sequences and efficient codes," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 669–675, 1989. *(Cited on page 11)*

[34] C. Keith, "Probability distributions and maximum entropy." [Online]. Available: http://www.math.uconn.edu/~kconrad/blurbs/entropy.pdf *(Cited on page 12)*

[35] D. E. Lake, "Renyi entropy measures of heart rate gaussianity." *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 21–27, 2006. *(Cited on pages 13, 20, 27, and 31)*

[36] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, *Handbook of Differential Entropy*. Boca Raton: Chapman and Hall/CRC, 2013. *(Cited on pages 14 and 19)*

[37] Y. Sinai, "Kolmogorov-sinai entropy," *Scholarpedia*, vol. 4, no. 3, p. 2034, 2009. *(Cited on page 16)*

[38] J. Shynk, *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications*. John Wiley & Sons, Sep. 2012. *(Cited on page 16)*

[39] A. N. Kolmogorov, "A new metric invariant of transient dynamical systems and automorphisms in lebesgue space," *Dokl. Akad. Nauk. SSSR.*, vol. 119, pp. 861–864, 1958. *(Cited on page 16)*

[40] Y. Sinai, "On the notion of entropy for a dynamic system," *Dokl. Akad. Nauk. SSSR.*, vol. 124, pp. 768–771, 1959. *(Cited on page 16)*

[41] *(Cited on page 16)*

[42] "Measure-preserving dynamical system," Jun. 2014. *(Cited on page 17)*

[43] S. M. Pincus and A. M. Goldberger, "Physiological time-series analysis: what does regularity quantify," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 266, pp. 1643–1656, 1994. *(Cited on pages 17, 18, 27, 28, and 47)*

[44] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *J Clin Monit*, vol. 7, no. 4, pp. 335–345, 1991. *(Cited on pages 18, 33, and 47)*

[45] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 71, no. 2 Pt 1, p. 021906, 2005. *(Cited on page 18)*

[46] S. M. Pincus, "Approximate entropy as a measure of system complexity," in *Proc. Natl. Acad. Sci. USA*, vol. 88, USA, March 1991, pp. 2297–2301. *(Cited on page 18)*

[47] J. S. Richman and J. R. Moorman, "Physiological time series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Heart Circ Physiol.*, vol. 278, pp. 2039–2049, 2000. *(Cited on page 18)*

[48] S. M. Pincus and W.-M. Huang, "Approximate entropy: Statistical properties and applications," *Commun. Stat. Theor-M.*, vol. 21, no. 11, pp. 3061–3077, 1992. *(Cited on pages 18 and 31)*

[49] A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati, T. Gnecchi-Ruscone, A. Malliani, and S. Cerutti, "Measuring regularity by means of a corrected conditional entropy in sympathetic outflow," *Biol. Cybern.*, vol. 78, no. 1, pp. 71–78, Jan. 1998. *(Cited on pages 19, 25, and 31)*

[50] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000. *(Cited on page 19)*

[51] A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D: Nonlinear Phenomena*, vol. 166, no. 1–2, pp. 43–62, 2002. *(Cited on pages 19 and 20)*

[52] G. A. Darbellay, "An estimator of the mutual information based on a criterion for independence," *Comput. Stat. Data Anal.*, vol. 32, no. 1, pp. 1–17, 1999. *(Cited on page 19)*

[53] C. E. Shannon, "A mathematical theory of communication," *Bell. Syst. Tech. J.*, vol. 27, pp. 379–423, 1948. *(Cited on page 20)*

[54] B.-l. Hao, "Symbolic dynamics and characterization of complexity," *Physica D: Nonlinear Phenomena*, vol. 51, no. 1-3, pp. 161–176, 1991. *(Cited on page 21)*

[55] X.-S. Zhang, Y.-S. Zhu, N. Thakor, and Z.-Z. Wang, "Detecting ventricular tachycardia and fibrillation by complexity measure," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 5, pp. 548–555, 1999. *(Cited on page 21)*

[56] K. Chakrabarti, E. Keogh, S. Mehrotra, and M. Pazzani, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM Trans. Database Syst.*, vol. 27, no. 2, pp. 188–228, 2002. *(Cited on page 21)*

[57] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: a novel symbolic representation of time series," *Data Min Knowl Disc*, vol. 15, no. 2, pp. 107–144, 2007. *(Cited on pages 21 and 22)*

[58] R. J. Larsen and M. L. Marx, *An introduction to mathematical statistics and its applications; 5th ed.* Boston, MA: Prentice Hall, 2012. *(Cited on page 22)*

[59] M. R. Schroeder, "Linear prediction, entropy and signal analysis," *IEEE ASSP Magazine*, vol. 1, no. 3, pp. 3–11, 1984. *(Cited on page 25)*

[60] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Vision Graph.*, vol. 29, no. 3, pp. 273–285, Mar. 1985. *(Cited on page 25)*

[61] E. Czogala and J. Leski, "Application of entropy and energy measures of fuzziness to processing of ECG signal," *Fuzzy Set. Syst.*, vol. 97, no. 1, pp. 9–18, 1998. *(Cited on page 25)*

[62] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2004. *(Cited on page 25)*

[63] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," in *Ubiquitous intelligence and computing*. Springer, 2010, pp. 548–562. *(Cited on pages 25 and 85)*

[64] L. Myong-Woo and M. K. Adil, "A single tri-axial accelerometer-based real-time personal life log system capable of activity classification and exercise information generation." *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2010, pp. 1390–3, 2010. *(Cited on pages 25 and 84)*

[65] R. M. Gray, "Toeplitz and circulant matrices: A review," Tech. Rep., 2001. *(Cited on pages 30 and 31)*

[66] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol-Heart C.*, vol. 278, no. 6, pp. H2039–H2049, 2000. *(Cited on pages 33, 47, and 50)*

[67] "Monte carlo simulation definition." [Online]. Available: http://www.investopedia.com/terms/m/montecarlosimulation.asp *(Cited on page 33)*

[68] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and

A. Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, no. 6, pp. 537–553, 2001. *(Cited on pages 40 and 62)*

[69] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. *(Cited on page 40)*

[70] "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996. *(Cited on pages 40, 50, 61, and 63)*

[71] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, 1974. *(Cited on page 41)*

[72] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 4th ed. Hoboken, N.J: Wiley, 2008. *(Cited on pages 41 and 75)*

[73] X. Zhang, R. Roy, and E. Jensen, "Eeg complexity as a measure of depth of anesthesia for patients," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 12, pp. 1424–1433, 2001. *(Cited on page 42)*

[74] D. Abasolo, R. Hornero, C. Gomez, M. Garcia, and M. Lopez, "Analysis of eeg background activity in alzheimers disease patients with lempelziv complexity and central tendency measure," *Med Eng Phys*, vol. 28, no. 4, pp. 315–322, 2006. *(Cited on page 42)*

[75] J. Hu, J. Gao, and J. Principe, "Analysis of biomedical signals by the lempel-ziv complexity: the effect of finite data size," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2606–2609, 2006. *(Cited on page 42)*

[76] M. Aboy, R. Hornero, D. Abasolo, and D. Alvarez, "Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2282–2288, 2006. *(Cited on page 42)*

[77] J.-L. Blanc, N. Schmidt, L. Bonnier, L. Pezard, and A. Lesne, "Quantifying neural correlations using lempel-ziv complexity," in *Proceedings of the second french conference on Computational Neuroscience, Marseille*, Marseille, France, 2008. *(Cited on page 42)*

[78] J. M. Amigo, J. Szczepanski, E. Wajnryb, and M. V. Sanchez, "Estimating the entropy rate of spike trains via lempel-ziv complexity," *Neural Comput*, vol. 16, no. 4, pp. 717–736, 2004. *(Cited on page 42)*

[79] M. Migliorini, M. O. Mendez, and A. M. Bianchi, "Study of heart rate variability in bipolar disorder: Linear and non-linear parameters during sleep," *Front Neuroeng*, vol. 4, no. 22, pp. 1–7, 2012. *(Cited on page 42)*

[80] A. M. Bianchi, M. O. Mendez, M. Ferrario, L. Ferini-Strambi, and S. Cerutti, "Long-term correlations and complexity analysis of the heart rate variability signal during sleep. comparing normal and pathologic subjects," *Methods Inf Med*, vol. 49, no. 5, pp. 479–483, 2010. *(Cited on page 42)*

[81] R. Cabiddu, S. Cerutti, G. Viardot, S. Werner, and A. M. Bianchi, "Modulation of the sympatho-vagal balance during sleep: Frequency domain study of heart rate variability and respiration," *Front Physiol*, vol. 3, 2012. *(Cited on page 42)*

[82] A. Kales, A. Rechtschaffen, L. A. University of California, Brain Information Service, and NINDB Neurological Information Network (U.S.), *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Allan Rechtschaffen and Anthony Kales, editors.* Bethesda, Md.: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968. *(Cited on pages 42 and 62)*

[83] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, M. M. Tangredi, and American Academy of Sleep Medicine, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the american academy of sleep medicine," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, 2012. *(Cited on page 42)*

[84] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 283, no. 3, pp. R789–797, 2002. *(Cited on page 46)*

[85] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value r for approximate entropy," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1966–1972, 2008. *(Cited on pages XV, 47, and 48)*

[86] M. Aktaruzzaman and R. Sassi, "Parametric estimation of sample entropy in heart rate variability analysis," *Biomed. Signal Process. Control.*, vol. 14, pp. 141–147, 2014. *(Cited on pages 51 and 85)*

[87] D. P. Zipes and H. J. J. Wellens, "Sudden cardiac death," *Circulation*, vol. 98, no. 21, pp. 2334–2351, Nov. 1998. *(Cited on pages 56 and 59)*

[88] A. Gacek, "An introduction to ecg signal processing and analysis," in *ECG Signal Processing, Classification and Interpretation*, A. Gacek and W. Pedrycz, Eds. Springer London, 2012, pp. 21–46. *(Cited on pages 57, 58, 60, and 61)*

[89] T. Gaziano, K. S. Reddy, F. Paccaud, S. Horton, and V. Chaturvedi, "Cardiovascular disease," in *Disease Control Priorities in Developing Countries*, 2nd ed., D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, and P. Musgrove, Eds. Washington (DC): World Bank, 2006. *(Cited on page 58)*

[90] Writing Group Members, V. L. Roger, A. S. Go, D. M. Lloyd-Jones, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox, H. J. Fullerton, C. Gillespie, S. M. Hailpern, J. A. Heit, V. J. Howard, B. M. Kissela, S. J. Kittner, D. T. Lackland, J. H. Lichtman, L. D. Lisabeth, D. M. Makuc, G. M. Marcus, A. Marelli, D. B. Matchar, C. S. Moy, D. Mozaffarian, M. E. Mussolino, G. Nichol, N. P. Paynter, E. Z. Soliman, P. D. Sorlie, N. Sotoodehnia, T. N. Turan, S. S. Virani, N. D. Wong, D. Woo, M. B. Turner, on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee, and On behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee, "Heart disease and stroke statistics–2012 update: A report from the american heart association," *Circulation*, vol. 125, no. 1, pp. e2–e220, Jan. 2012. *(Cited on page 59)*

[91] D. Dubin, *Rapid Interpretation of EKG's, Sixth Edition*. *(Cited on page 60)*

[92] A. Martinez, R. Alcaraz, and J. J. Rieta, "A new method for automatic delineation of ECG fiducial points based on the phasor transform," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2010, pp. 4586–4589, 2010. *(Cited on page 60)*

[93] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans Biomed Eng*, vol. 32, no. 3, pp. 230–236, Mar. 1985. *(Cited on page 60)*

[94] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Med Biol Eng Comput*, vol. 44, no. 12, pp. 1031–1051, Dec. 2006. *(Cited on page 61)*

[95] R. Ferri, L. Parrino, A. Smerieri, M. G. Terzano, M. Elia, S. A. Musumeci, and S. Pettinato, "Cyclic alternating pattern and spectral analysis of heart rate variability during normal sleep," *J Sleep Res*, vol. 9, no. 1, pp. 13–18, Mar. 2000. *(Cited on page 62)*

[96] H. R. Colten, C. o. S. M. Bruce M. Altevogt, Editors, and Research, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. The National Academies Press, 2006. [Online]. Available: http://www.nap.edu/openbook.php?record_id=11617 *(Cited on page 62)*

[97] A. Kales, A. Rechtschaffen, L. A. University of California, and N. N. I. N. (U.S.), *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Allan Rechtschaffen and Anthony Kales, editors*. U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network Bethesda, Md, 1968. *(Cited on page 62)*

[98] *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine, 2007. *(Cited on page 62)*

[99] D. Mendel, "The mechanics of the circulation," *J R Soc Med*, vol. 71, no. 8, pp. 629–629, 1978. *(Cited on page 62)*

[100] S. R. Ulimoen, S. Enger, J. Carlson, P. G. Platonov, A. H. Pripp, M. Abdelnoor, H. Arnesen, K. Gjesdal, and A. Tveit, "Comparison of four single-drug regimens on ventricular rate and arrhythmia-related symptoms in patients with permanent atrial fibrillation," *Am. J. Cardiol.*, vol. 111, no. 2, pp. 225–230, 2013. *(Cited on page 63)*

[101] E. K. Heist, M. Mansour, and J. N. Ruskin, "Rate control in atrial fibrillation targets, methods, resynchronization considerations," *Circulation*, vol. 124, no. 24, pp. 2746–2755, 2011. *(Cited on page 63)*

[102] J. McDonald, *Handbook of Biological Statistics*. Sparky House Publishing,Baltimore, Maryland, 2009. *(Cited on pages 64 and 70)*

[103] K. T. Dolan and M. L. Spano, "Surrogate for nonlinear time series analysis," *Phys. Rev. E.*, vol. 64, no. 4, pp. 0 461 281–6, Sep. 2001. *(Cited on page 66)*

[104] M. V. Pitzalis, F. Massari, C. Forleo, A. Fioretti, R. Colombo, C. Balducci, F. Mastropasqua, and P. Rizzon, "Respiratory systolic pressure variability during atrial fibrillation and sinus rhythm," *Hypertension*, vol. 34, no. 5, pp. 1060–1065, 1999. *(Cited on page 67)*

[105] L. Mainardi, V. Corino, S. Belletti, P. Terranova, and F. Lombardi, "Low frequency component in systolic arterial pressure variability in patients with persistent atrial fibrillation," *Auton. Neurosci.*, vol. 151, no. 2, pp. 147–153, 2009. *(Cited on pages 67 and 68)*

[106] V. Da Corino, L. Mainardi, and F. Lombardi, "Spectral analysis of blood pressure variability in atrial fibrillation: The effect of tilting," in *Computing in Cardiology Conference (CinC), 2013*, 2013, pp. 1203–1206. *(Cited on page 67)*

[107] H. Arai, H. Sato, M. Yamamoto, M. Uchida, T. Nakamachi, T. Kaneko, T. Arikawa, Y. Tsuboko, T. Aizawa, H. Iinuma, and K. Kato, "[changes in the fractal component of spectral analysis of heart rate variability and systolic blood pressure variability during the head-up tilt test]," *J. Cardiol*, vol. 34, no. 4, pp. 211–217, 1999. *(Cited on page 67)*

[108] R. Maestri, G. D. Pinna, A. Accardo, P. Allegrini, R. Balocchi, G. D'Addio, M. Ferrario, D. Menicucci, A. Porta, R. Sassi, M. G. Signorini, M. T. La Rovere, and S. Cerutti, "Nonlinear indices of heart rate variability in chronic heart failure patients: redundancy and comparative clinical value," *J. Cardiovasc. Electrophysiol.*, vol. 18, no. 4, pp. 425–433, 2007. *(Cited on page 67)*

[109] A. Porta, T. Gnecchi-Ruscone, E. Tobaldini, S. Guzzetti, R. Furlan, and N. Montano, "Progressive decrease of heart period variability entropy-based complexity during graded head-up tilt," *J. Appl. Physiol.*, vol. 103, no. 4, pp. 1143–1149, 2007. *(Cited on page 67)*

[110] K. M. Stein, J. Walden, N. Lippman, and B. B. Lerman, "Ventricular response in atrial fibrillation: random or deterministic?" *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 277, no. 2, pp. H452–H458, 1999. *(Cited on page 67)*

[111] V. D. A. Corino, I. Cygankiewicz, L. T. Mainardi, M. Stridh, R. Vasquez, A. Bayes de Luna, F. Holmqvist, W. Zareba, and P. G. Platonov, "Association between atrial fibrillatory rate and heart rate variability in patients with atrial fibrillation and congestive heart failure," *Ann Noninvasive Electrocardiol*, vol. 18, no. 1, pp. 41–50, 2013. *(Cited on page 67)*

[112] V. D. A. Corino, S. Belletti, P. Terranova, F. Lombardi, and L. T. Mainardi, "Heart rate and systolic blood pressure in patients with persistent atrial fibrillation. a linguistic analysis," *Methods Inf Med*, vol. 49, no. 5, pp. 516–520, 2010. *(Cited on page 67)*

[113] D. A. Porta, S. Guzzetti, N. Montano, M. Pagani, V. Somers, A. Malliani, G. Baselli, and S. Cerutti, "Information domain analysis of cardiovascular variability signals: Evaluation of regularity, synchronisation and co-ordination," *Med. Biol. Eng. Comput.*, vol. 38, no. 2, pp. 180–188, 2000. *(Cited on page 67)*

[114] T. A. Kuusela, T. T. Jartti, K. U. O. Tahvanainen, and T. J. Kaila, "Nonlinear methods of biosignal analysis in assessing terbutaline-induced heart rate and blood pressure changes," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 282, no. 2, pp. H773–783, 2002. *(Cited on page 67)*

[115] European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery, A. J. Camm, P. Kirchhof, G. Y. H. Lip, U. Schotten, I. Savelieva, S. Ernst, I. C. Van Gelder, N. Al-Attar, G. Hindricks, B. Prendergast, H. Heidbuchel, O. Alfieri, A. Angelini, D. Atar, P. Colonna, R. De Caterina, J. De Sutter, A. Goette, B. Gorenek, M. Heldal, S. H. Hohloser, P. Kolh, J.-Y. Le Heuzey, P. Ponikowski, and F. H. Rutten, "Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (ESC)," *Eur. Heart J.*, vol. 31, no. 19, pp. 2369–2429, 2010. *(Cited on page 68)*

[116] N. Stanley, "The physiology of sleep and the impact of ageing," *Eur. Urol. Supplements*, vol. 3, no. 6, pp. 17–23, 2005. *(Cited on page 72)*

[117] F. Pizza, S. Contardi, A. B. Antognini, M. Zagoraiou, M. Borrotti, B. Mostacci, S. Mondini, and F. Cirignotta, "Sleep quality and motor vehicle crashes in adolescents," *J Clin Sleep Med*, vol. 6, no. 1, pp. 41–45, 2010. *(Cited on page 73)*

[118] M. H. Araghi, A. Jagielski, I. Neira, A. Brown, S. Higgs, G. N. Thomas, and S. Taheri, "The complex associations among sleep quality, anxiety-depression, and quality of life in patients with extreme obesity," *SLEEP*, 2013. *(Cited on page 73)*

[119] A. R. Cass, W. J. Alonso, J. Islam, and S. C. Weller, "Risk of obstructive sleep apnea in patients with type 2 diabetes mellitus," *Fam Med*, vol. 45, no. 7, pp. 492–500, 2013. *(Cited on page 73)*

[120] N. Covassin, M. de Zambotti, N. Cellini, M. Sarlo, and L. Stegagno, "Cardiovascular down-regulation in essential hypotension: relationships with autonomic control and sleep," *Psychophysiology*, vol. 50, no. 8, pp. 767–776, 2013. *(Cited on page 73)*

[121] J. Engeda, B. Mezuk, S. Ratliff, and Y. Ning, "Association between duration and quality of sleep and the risk of pre-diabetes: evidence from NHANES," *Diabet. Med.*, vol. 30, no. 6, pp. 676–680, 2013. *(Cited on page 73)*

[122] C. Zamarròn, L. Cuadrado, and R. Alvarez-Sala, "Pathophysiologic mechanisms of cardiovascular disease in obstructive sleep apnea syndrome," *Pulm. Med.*, vol. 2013, p. e521087, 2013. *(Cited on page 73)*

[123] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, M. M. Tangredi, and American Academy of Sleep Medicine, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the american academy of sleep medicine," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, 2012. *(Cited on page 73)*

[124] U. J. Scholz, A. M. Bianchi, S. Cerutti, and S. Kubicki, "Vegetative background of sleep: spectral analysis of the heart rate variability," *Physiol. Behav.*, vol. 62, no. 5, pp. 1037–1043, 1997. *(Cited on page 73)*

[125] T. Penzel, A. Bunde, L. Grote, J. W. Kantelhardt, J. H. Peter, and K. Voigt, "Heart rate variability during sleep stages in normals and in patients with sleep apnea," *Stud Health Technol Inform*, vol. 77, pp. 1256–1260, 2000. *(Cited on page 73)*

[126] H. Kondo, M. Ozone, N. Ohki, Y. Sagawa, K. Yamamichi, M. Fukuju, T. Yoshida, C. Nishi, A. Kawasaki, K. Mori, T. Kanbayashi, M. Izumi, Y. Hishikawa, S. Nishino, and T. Shimizu, "Association between heart rate variability, blood pressure and autonomic activity in cyclic alternating pattern during sleep," *Sleep*, vol. 37, no. 1, pp. 187–194, 2014. *(Cited on page 73)*

[127] E. Sforza, V. Pichot, J. C. Barthelemy, J. Haba-Rubio, and F. Roche, "Cardiovascular variability during periodic leg movements: a spectral analysis approach," *Clin Neurophysiol*, vol. 116, no. 5, pp. 1096–1104, May 2005. *(Cited on page 73)*

[128] L. Ferini-Strambi, A. Bianchi, M. Zucconi, A. Oldani, V. Castronovo, and S. Smirne, "The impact of cyclic alternating pattern on heart rate variability during sleep in healthy young adults," *Clin Neurophysiol*, vol. 111, no. 1, pp. 99–101, 2000. *(Cited on page 73)*

[129] S. Telser, M. Staudacher, Y. Ploner, A. Amann, H. Hinterhuber, and M. Ritsch-Marte, "Can one detect sleep stage transitions for on-line sleep scoring by monitoring the heart rate variability," *Somnologie*, vol. 8, no. 2, pp. 33–41, 2004. *(Cited on page 73)*

[130] J. M. Kortelainen, M. O. Mendez, A. M. Bianchi, M. Matteucci, and S. Cerutti, "Sleep staging based on signals acquired through bed sensor," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 3, pp. 776–785, 2010. *(Cited on pages 73 and 74)*

[131] A. M. Bianchi, M. O. Mendez, and S. Cerutti, "Processing of signals recorded through smart devices: sleep-quality assessment," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 3, pp. 741–747, 2010. *(Cited on page 73)*

[132] D. Zemaityte, G. Varoneckas, and E. Sokolov, "Heart rhythm control during sleep," *Psychophysiology*, vol. 21, no. 3, pp. 279–289, 1984. *(Cited on page 73)*

[133] B. V. Vaughn, S. R. Quint, J. A. Messenheimer, and K. R. Robertson, "Heart period variability in sleep," *Electroencephalogr Clin Neurophysiol*, vol. 94, no. 3, pp. 155–162, 1995. *(Cited on page 73)*

[134] M. O. Mendez, M. Matteucci, V. Castronovo, L. F. Strambi, S. Cerutti, and A. M. Bianchi, "Sleep staging from heart rate variability: time-varying spectral features and hidden markov models," *Int. J. Biomed. Eng. Tech.*, vol. 3, no. 3/4, pp. 246–263, 2010. *(Cited on pages 73, 75, 82, and 83)*

[135] S. J. Redmond, P. d. Chazal, C. OḂrien, S. Ryan, W. T. McNicholas, and C. Heneghan, "Sleep staging using cardiorespiratory signals," *Somnologie*, vol. 11, no. 4, pp. 245–256, 2007. *(Cited on pages 73, 82, and 83)*

[136] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, "Sleep stages classification based on heart rate variability and random forest," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 624–633, 2013. *(Cited on page 73)*

[137] F. Ebrahimi, S.-K. Setarehdan, J. Ayala-Moyeda, and H. Nazeran, "Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals," *Comput. Meth. Prog. Bio.*, vol. 112, no. 1, pp. 47–57, 2013. *(Cited on page 73)*

[138] G. Baselli, A. Porta, O. Rimoldi, M. Pagani, and S. Cerutti, "Spectral decomposition in multichannel recordings based on multivariate parametric identification," *IEEE Trans Biomed Eng*, vol. 44, no. 11, pp. 1092–1101, 1997. *(Cited on page 75)*

[139] C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos*, vol. 5, no. 1, pp. 82–87, 1995. *(Cited on page 76)*

[140] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943. *(Cited on page 77)*

[141] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006. *(Cited on page 78)*

[142] J. Cohen, "Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit," *Psychol Bull*, vol. 70, no. 4, pp. 213–220, 1968. *(Cited on page 79)*

[143] P. G. H. Golub and D. C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, vol. 14, no. 5, pp. 403–420, 1970. *(Cited on page 81)*

[144] S. Dernbach, B. Das, N. C. Krishnan, B. Thomas, and D. Cook, "Simple and complex activity recognition through smart phones," in *2012 8th International Conference on Intelligent Environments (IE)*, 2012, pp. 214–221. *(Cited on pages 83 and 85)*

[145] "World health organization: Move for health." [Online]. Available: http://www.who.int/moveforhealth/en/ *(Cited on page 83)*

[146] "Physical activity guidelines advisory committee report, 2008. washington, dc: U.s. department of health and human services, 2008." [Online]. Available: http://www.health.gov/paguidelines/report/pdf/committeereport.pdf *(Cited on page 83)*

[147] R. Poppe, "Vision-based human motion analysis: An overview," *Comput. Vis. Image Underst.*, vol. 108, no. 1-2, pp. 4–18, 2007. *(Cited on page 83)*

[148] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008. *(Cited on page 83)*

[149] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," in *Pervasive Computing*. Springer, 2004, pp. 18–32. *(Cited on page 83)*

[150] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster, "Classification of basic daily movements using a triaxial accelerometer," *Med. Biol. Eng. Comput.*, vol. 42, no. 5, pp. 679–687, 2004. *(Cited on pages 83, 84, and 89)*

[151] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Pattern Recognition and Image Analysis*, J. Vitria, J. M. Sanches, and M. Hernandez, Eds. Springer Berlin Heidelberg, 2011, pp. 289–296. *(Cited on page 83)*

[152] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson, "Multisensor data fusion for physical activity assessment," *IEEE Trans Biomed Eng*, vol. 59, no. 3, pp. 687–696, 2012. *(Cited on page 83)*

[153] D. M. Pober, J. Staudenmayer, C. Raphael, and P. S. Freedson, "Development of novel techniques to classify physical activity mode using accelerometers," *Med Sci Sports Exerc*, vol. 38, no. 9, pp. 1626–1634, 2006. *(Cited on page 84)*

[154] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," *Procedia Computer Science*, vol. 34, pp. 450–457, 2014. *(Cited on page 84)*

[155] P. H. Veltink, H. B. Bussmann, W. de Vries, W. L. Martens, and R. C. Van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," *IEEE Trans Rehabil Eng*, vol. 4, no. 4, pp. 375–385, 1996. *(Cited on page 84)*

[156] J. Fahrenberg, F. Foerster, M. Smeja, and W. Muller, "Assessment of posture and motion by multichannel piezoresistive accelerometer recordings," *Psychophysiology*, vol. 34, no. 5, pp. 607–612, 1997. *(Cited on page 84)*

[157] A. M. Khan, Y. Lee, S. Lee, and T. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010. *(Cited on pages XX, 84, 85, 86, 88, 89, 90, 94, and 101)*

[158] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks, 2006*, 2006, pp. 113–116. *(Cited on pages 84 and 85)*

[159] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Pervasive Computing*, K. P. Fishkin, B. Schiele, P. Nixon, and A. Quigley, Eds. Springer Berlin Heidelberg, 2006, pp. 1–16. *(Cited on page 84)*

[160] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *IEEE International Conference on Systems, Man, and Cybernetics 2001*, vol. 2, 2001, pp. 747–752. *(Cited on page 84)*

[161] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence*, ser. IAAI'05, vol. 3. AAAI Press, 2005, pp. 1541–1546. *(Cited on pages 84 and 85)*

[162] S. Preece, J. Goulermas, L. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009. *(Cited on page 85)*

[163] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor Newsl*, vol. 12, no. 2, pp. 74–82, 2011. *(Cited on page 85)*

[164] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013*, 2013, pp. 437–442. *(Cited on page 85)*

[165] "What is the feature extraction tool and how does it work?" [Online]. Available: https://help.theactigraph.com/entries/22119735-What-is-the-feature-extraction-tool-and-how-does-it-work- *(Cited on page 85)*

[166] L. A. Kelly, D. G. McMillan, A. Anderson, M. Fippinger, G. Fillerup, and J. Rider, "Validity of actigraphs uniaxial and triaxial accelerometers for assessment of

physical activity in adults in laboratory conditions," *BMC Medical Physics*, vol. 13, no. 1, p. 5, 2013. *(Cited on page 86)*

[167] "Wrist-worn accelerometer research watches." [Online]. Available: http://www.geneactiv.org/ *(Cited on page 86)*

[168] D. W. Esliger, A. V. Rowlands, T. L. Hurst, M. Catt, P. Murray, and R. G. Eston, "Validation of the GENEA accelerometer," *Medicine & Science in Sports & Exercise*, vol. 43, no. 6, pp. 1085–1093, 2011. *(Cited on page 86)*

[169] A. V. Rowlands, F. Fraysse, M. Catt, V. H. Stiles, R. M. Stanley, R. G. Eston, and T. S. Olds, "Comparability of measured acceleration from accelerometry-based activity monitors," *Medicine & Science in Sports & Exercise*, p. 1, 2014. *(Cited on page 86)*

[170] T. Bernecker, F. Graf, H. Kriegel, C. Moennig, and C. Tuermer, "Activity recognition on 3d accelerometer data (technical report)," pp. 1–22, 2012. *(Cited on page 89)*

[171] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 3, pp. 136–147, 1997. *(Cited on page 89)*

[172] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. *(Cited on page 90)*

[173] "Optical character recognition." [Online]. Available: http://en.wikipedia.org/wiki/Optical_character_recognition *(Cited on page 103)*

[174] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognition*, vol. 36, no. 10, pp. 2271–2285, 2003. *(Cited on page 103)*

[175] F. Kimura and M. Shridhar, "Handwritten numerical recognition based on multiple algorithms," *Pattern Recogn.*, vol. 24, no. 10, pp. 969–983, 1991. *(Cited on page 103)*

[176] "Bengali language." [Online]. Available: http://en.wikipedia.org/wiki/Bengali_language *(Cited on pages 103 and 104)*

[177] "Summary by language size." [Online]. Available: http://www.ethnologue.com/statistics/size *(Cited on page 104)*

[178] U. Bhattacharya, T. K. Das, A. Datta, S. K. Parui, and B. B. Chaudhuri, "Recognition of handprinted bangla numerals using neural network models," in *Proceedings of the 2002 AFSS International Conference on Fuzzy Systems. Calcutta: Advances in Soft Computing*. London, UK, UK: Springer-Verlag, 2002, pp. 228–235. *(Cited on pages 104, 106, 108, and 110)*

[179] U. Pal, A. Belaid, and B. B. Chaudhuri, "A system for bangla handwritten numeral recognition," *IETE J Res*, vol. 3, pp. 444–457, 2006. *(Cited on pages 104, 108, and 110)*

[180] Y. Wen, Y. Lu, and P. Shi, "Handwritten bangla numeral recognition system and its application to postal automation," *Pattern Recognition*, vol. 40, no. 1, pp. 99–107, 2007. *(Cited on pages 104 and 105)*

[181] J. Xu, J. Xu, and Y. Lu, "Handwritten bangla digit recognition using hierarchical bayesian network," *ISKE*, pp. 1096–1099, 2008. *(Cited on pages 104 and 108)*

[182] M. Aktaruzzaman, M. F. Khan, and A.-U. Ambia, "A new technique for segmentation of handwritten numerical strings of bangla language," *International Journal of Information Technology and Computer Science*, vol. 5, no. 5, pp. 38–43, 2013. *(Cited on page 104)*

[183] C. Liu and C. Y. Suen, "A new benchmark on the recognition of handwritten bangla and farsi numeral characters," *Pattern Recogn*, vol. 42, pp. 3287–295, 2009. *(Cited on pages 104, 108, 109, and 110)*

[184] B. Chaudhuri, "A Complete Handwritten Numeral Database of Bangla – A Major Indic Script," in *Proc. 10th IWFHR*. La Baule (France): Université de Rennes 1, 2006, pp. 379–384. *(Cited on pages 105, 108, 109, and 110)*

[185] N. Li, *An Implementation of OCR System Based on Skeleton Matching*. University of Kent, Canterbury, UK: University of Kent Computing Laboratory, 1993. *(Cited on page 106)*

[186] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. *(Cited on page 106)*

[187] P. Purkait and B. Chanda, "Off-line recognition of hand-written bengali numerals using morphological features," in *ICFHR*, 2010, pp. 363–368. *(Cited on pages 107, 108, 109, and 110)*

[188] M. Allili and D. Ziou, "Topological feature extraction in binary images," in *Signal Processing and its Applications, Sixth International, Symposium on. 2001*, vol. 2, 2001, pp. 651–654. *(Cited on page 107)*

[189] "Roberts cross," Aug. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Roberts_cross&oldid=620521499 *(Cited on page 107)*

[190] U. Pal, N. Sharma, T. Wakabayashi, and F. Kimura, "Handwritten numeral recognition of six popular indian scripts," *Proc 12th ICDAR*, vol. 2, pp. 749–753, 2007. *(Cited on pages 107, 108, and 110)*

[191] U. Bhattacharya and B. Chaudhuri, "Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 444–457, 2009. *(Cited on pages 108 and 110)*

[192] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990. *(Cited on page 108)*

[193] F. Kimura, Y. Miyake, and M. Shridhar, "Handwritten ZIP code recognition using lexicon free word recognition algorithm," in *Proceedings of the Third International Conference on Document Analysis and Recognition, 1995*, vol. 2, Aug. 1995, pp. 906–910 vol.2. *(Cited on page 108)*

[194] C. Liu and H. Sako, "Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition," *Pattern Recogn*, vol. 39, pp. 669–681, 2006. *(Cited on page 109)*

[195] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *KDDM*, vol. 2, pp. 1–43, 1998. *(Cited on page 109)*

[196] A. Rosenfeld and A. Kak, "Digital picture processing," *Acad Press, 24/28 Oval Road, London, NW1 7DX*, vol. 2, pp. 1–43, 1982. *(Cited on page 109)*

[197] N. Otsu, "A threshld selection method from grey-level histograms," *IEEE Trans. Systems Man Cybernetics.*, vol. 9, pp. 377–393, 1979. *(Cited on page 110)*

[198] M. I. Mojahidul, M. Aktaruzzaman, M. K. Farukuzzaman, and M. I. Shohidul, "Handwritten character recognition system for bangla text containing modifiers and overlapped characters," *ACM Transactions on Computer Systems*, vol. 01, no. 05, pp. 15–19, 2011. *(Cited on page 111)*

[199] B. Chaudhuri, U. Pal, and M. Mitra, "Automatic recognition of printed oriya script," *Sadhana*, vol. 27, no. 1, pp. 23–34, 2002. *(Cited on page 112)*

[200] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1, pp. 37–52, 1987. *(Cited on page 115)*

A

PUBLICATIONS

Some ideas and significant results those have been published (or accepted) from this thesis:

A.1 LIST OF REFEREED JOURNAL PAPERS

1. "*Parametric estimation of sample entropy in heart rate variability analysis*". **M. Aktaruzzaman** and R. Sassi, Biomedical Signal Processing and Control, vol. 14, pp. 141-147, 2014, ISSN:1746-8094, DOI: 10.1016/j.bspc.2014.07.011.

   Abstract:

   In this paper, a detailed study on the possibility and significance of performing a parametric estimation of sample entropy (SampEn) was proposed. SampEn is a non-linear metric, meant to quantify regularity of a time series. It is widely employed on biomedical signal (*i.e.* heart rate variability). Results relevant to approximate entropy, a related index, were also reported.

   An analytical expression for SampEn of an autoregressive (AR) model was derived first. Then we studied the feasibility of a parametric estimation of SampEn through AR models, both on synthetic and real series. RR series of different lengths were fitted to an AR model and expected values (SE_μ) estimated.

   Values of SampEn, computed from real beat-to-beat interval time series (obtained from 72 normal subjects and 29 congestive heart failure patients), with $m = 1$ and $r = 0.2$, were within the standard range of SE_μ more than 83% (for series length N=75) and 28% (for N=1500) of the cases. Surrogate data were employed to verify if departures from Gaussianity were to account for the mismatch.

   The work supported the finding that when numerical and parametric estimates of SampEn agree, SampEn is mainly influenced by linear properties of the series. A disagreement, on the contrary, might point those cases where SampEn is truly

offering new information, not readily available with traditional temporal and spectral parameters.

2. "*Non-linear regularity of arterial blood pressure variability in patient with Atrial Fibrillation in tilt-test procedure.*" S. Cerutti, V. D. A. Corino, L. T. Mainardi, **M. Aktaruzzaman**, and R. Sassi, Europace, vol. 16 suppl issue: 4, pp. iv141-iv147, 2014, ISSN:1099-5129, DOI: 10.1093/europace/euu262 .

   Abstract:

   Aims Dynamics of cardiovascular series may be explored with non-linear techniques. It is unknown if the arterial pressure irregularity commonly observed in patients with AF might be further increased by a sympathetic stimulus such as orthostatic tilt.

   Methods Twenty patients (62±14 years, 15 men) were recruited for the study. Continuous beat-to-beat non-invasive arterial pressure was acquired at rest and during a passive orthostatic stimulus ("tilt-test"). Systolic (SAP) and diastolic (DAP) arterial pressure series of 300-samples were analyzed in both conditions. Approximate ($ApEn_{RR}$) and sample entropy ($SampEn_{RR}$) were computed, as irregularity measures. Equivalent metrics ($ApEn_{\mu}$ and $SampEn_{\mu}$) derived from an autoregressive model of the series were also obtained through numerical simulations, to further elucidate the nonlinear mechanisms present in the series.

   Results In 11 patients (group A), SAP significantly increased during tilt (from 103±13 to 114±17 mmHg, p<0.001 rest vs. tilt), whereas in 9 patients (group B) SAP remained almost unchanged (SAP: 110±18 vs. 106±19 mmHg, ns, rest vs. tilt). No clinical differences were found between group A and B. When analyzing group A, all irregularity measures significantly increased in SAP ($ApEn_{RR}$: 1.75±0.20 vs. 1.88±0.16, p<0.05; $SampEn_{RR}$: 1.71±0.30 vs. 1.88±0.27, p<0.05; $ApEn_{\mu}$: 1.87±0.20 vs. 1.96±0.18, p<0.05; $SampEn_{\mu}$: 1.94±0.27 vs. 2.06±0.18, p<0.05; rest vs. tilt), whereas no differences were found in DAP series. No significant differences were found in group B for either SAP or DAP.

   Conclusion The alterations of SAP during tilt in AF patients are not uniform and seem associated with different regularity patterns. The pressor response to sympathetic stimulation was also associated with an increase of SAP series irregularity.

3. "*The addition of entropy based regularity parameters improves sleep stage classification based on heart rate variability.*". **M. Aktaruzzaman**, M. Migliorini, M. Tenhunen, S. L. Himanen, R. Sassi, and A. M. Bianchi, Journal of Medical & Biologal Engineering & Computing, vol.-, pp.-, 2015 (article in press), ISSN: 0140-0118, DOI: 10.1007/s11517-015-1249-z).

   Abstract:

   This work considers automatic sleep stage classification, based on heart rate variability analysis, with a focus on the distinction of wakefulness (WAKE) from sleep, and rapid eye movement (REM) from non-REM (NREM) sleep. A set of

automatically annotated 20 one-night polysomnographic recordings was considered and artificial neural networks (ANN) were selected for classification. For each inter-heartbeat (RR) series, beside features previously presented, we introduced a set of 4 parameters related to signalś regularity. RR series of 3 different lengths were considered (corresponding to 2, 6, and 10 successive epochs in the same sleep stage). A set of four features alone captured 99% of the data variance in each classification problem and both contained one of the new features proposed. The accuracy of classification for REM vs NREM (68.4%, 2 epochs; 83.8%, 10 epochs) was higher than when distinguishing WAKE vs SLEEP (67.6%, 2 epochs; 71.3%, 10 epochs). Also, the reliability parameter (Cohenś Kappa) was higher (0.68 and 0.45 respectively). Sleep staging classification based on HRV, was still less precise than other staging methods, employing a larger variety of signals collected during polysomnographic studies. However, cheap and unobtrusive HRV-only sleep classification proved sufficiently precise for a wide range of applications.

A.2 INTERNATIONAL CONFERENCE PAPERS

4. "*Sample entropy parametric estimation for heart rate variability analysis*". **M. Aktaruzzaman** and R. Sassi, Computing in Cardiology 2013, Zaragoza, 22-25th Sep,Spain, 2013.

   Sample Entropy (SampEn) is a powerful approach for characterizing heart rate variability regularity. On the other hand, autoregressive (AR) models have been employed for maximum-entropy spectral estimation for more than 40 years. The aim of this study is to explore the feasibility of a parametric approach for SampEn estimation through AR models. We re-analyze the Physionet paroxysmal Atrial Fibrillation (AF) database, where RR series are provided before and after an AF episode, for 25 patients. In particular, we selected short RR series, close to AF episodes, to fit an AR model. Then, theoretical values of SampEn, based on each AR model, were analytically derived (SE_{th}) and also estimated numerically (SE_{syn}). The value of SampEn (SE_{rr}), computed on the 50 RR series with r=0.2×STD, m=1 and N=120, were within the standard range of SE_{syn} in 30 cases (39 for SE_{th}). This figure increased to 82% of cases, if shorter series were selected (N=75), and if RR series were replaced by surrogates with Gaussian amplitude distribution. Interestingly, without removing ectopic beats, every estimate of SampEn considered was significantly different between pre- and post- AF (SE_{rr}: p=0.02; SE_{syn}: p=0.0024; SE_{th}: p=0.023). When an AR model is appropriate and theoretical estimates differ from numerical ones, a parametric approach might enlighten additional information brought by SampEn.

5. "*HRV Regularity during Persistent Atrial Fibrillation: a Parametric Assessment using Sample Entropy.*".

M. Aktaruzzaman, V. Corino, L. T. Mainardi, S. R. Ulimoen, P. G. Platonov, A. Tveit, S. Enger, and R. Sassi.

8th conference of the European study group on cardiovascular oscillations,ESGCO 2014, Trento, Italy.

Abstract:

In this study, we investigated the relation between sample entropy (SampEn) of HRV series and the connected theoretical value ($SampEn_{TH}$), obtained for the autoregressive (AR) models fitted to the same sequences. AR models are commonly used for parametrical spectral analysis and classical HRV spectral parameters were considered as well. The analysis was performed on a subpopulation of the Rate Control in Atrial Fibrillation (RATAF) study, where RR series were collected before and after a β-blocker, Carvedilol, was administered.

SampEn, $SamEn_{TH}$ and the spectral parameters were significantly different after drug administration. However while SampEn is sensible to nonlinearities or non-Gaussianity in the series, the other parameters are not. To investigate further the changes in the series induced by the drug, both synthetic series generated by the fitted AR models and IAAFT surrogates were employed. The results suggest a reduction in non-Gaussianity as long as a relatively smaller increase in regularity.

6. "*Analysis of the effects of series length on Lempel-Ziv complexity during sleep.*".

   M. W. Rivolta, M. Migliorini, **M. Aktaruzzaman**, R. Sassi. and A. M. Bianchi.

   The 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC-14), Chicago, Illinois, USA:

   Abstract:

   Lempel-Ziv Complexity (LZC) has been demonstrated to be a powerful complexity measure in several biomedical applications. During sleep, it is still not clear how many samples are required to ensure robustness of its estimate when computed on beat-to-beat interval series (RR). The aims of this study were: i) evaluation of the number of necessary samples in different sleep stages for a reliable estimation of LZC; ii) evaluation of the LZC when considering inter-subject variability; and iii) comparison between LZC and Sample Entropy (SampEn). Both synthetic and real data were employed. In particular, synthetic RR signals were generated by means of AR models fitted on real data.

   The minimum number of samples required by LZC for having no changes in its average value, for both NREM and REM sleep periods, was 10^4 ($p<0.01$) when using a binary quantization. However, LZC can be computed with $N > 1000$ when a tolerance of 5% is considered satisfying.

   The influence of the inter-subject variability on the LZC was first assessed on model generated data confirming what found ($>10^4$; $p<0.01$) for both NREM and REM stage. However, on real data, without differentiate between sleep stages, the minimum number of samples required was 1.8×10^4.

The linear correlation between LZC and SampEn was computed on a synthetic dataset. We obtained a correlation higher than 0.75 ($p<0.01$) when considering sleep stages separately, and higher than 0.90 ($p<0.01$) when stages were not differentiated.

Summarizing, we suggest to use LZC with the binary quantization and at least 1000 samples when a variation smaller than 5% is considered satisfying, or at least 10^4 for maximal accuracy. The use of more than 2 levels of quantization is not recommended.