Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

Anno Accademico 2025–2026

Comparative Analysis of Entropy-Based Features for Multi-Domain Signal Classification

Relatore
Prof. L. Mesin

Candidato
Michele Fioretti

Correlatore
Prof. G. Manis

Abstract

Complexity Analysis is a subset of Signal Analysis which focuses on chaotic systems and the study of their internal dynamics. Many biomedical systems can be considered as chaotic, and the understanding of their hidden laws has become a topic of interest in literature. Entropy is a tool created to evaluate the complexity of a signal, biomedical or not. Since the introduction of Shannon Entropy in 1948, many Entropy definitions were developed creating a wide set of Entropy-based features being available for Complexity Analysis.

This thesis evaluated the performance of these complexity features compared with some linear features, in order to understand whether Entropy can provide a statistically significant improvement to the Analysis of three different databases.

The first two databases involved two different conditions of diseased cardiac contractility, respectively Congestive Heart Failure (CHF) and aging. In each one epochs of pathological HRV signal were analyzed together with epochs from subjects with regular cardiac activity, to prove features' potential to monitor and detect the worsening of clinical conditions.

The third database was focused on epilepsy, with epochs of EEG signals from different phases of epileptic events, particularly interictal and preictal conditions. The switch between interictal and preictal phase remains a difficult task in the treatment of epileptic patients, and the analysis worked into assessing whether features could provide an improvement in prediction of the seizure onset.

The subset of features involved on each database was obtained using a Sequential Forward Selection (SFS) approach, combined with a Gaussian Naive Bayesian (GNB) classifier. This method provided an accuracy evaluation from both Complexity and Time/Frequency approaches , creating the opportunity to evaluate the contribution of Entropy to the Analysis.

Results showed that Entropy-based features introduced an increase in accuracy , yet not statistically significant in some databases. This remains a starting point for future works on Complexity Analysis, showing possible applications of it into topics like cardiac monitoring, or real time diseases detection.

Contents

List of Figures

List of Tables

Introduction

Entropy has emerged as a powerful tool in the analysis of complex signals, offering a quantitative approach to understanding their underlying structure, variability, and unpredictability.

Since Fourier laid the foundations of Signal Theory at the beginning of the 19th century, the research in Signal Analysis has evolved continuously in the next decades, with other authors giving their contribution to increase the knowledge around it.

While traditional signal analysis methods have been effective for linear and stationary signals, they often struggle with chaotic or nonlinear signals. There are complex architectures behind the nature of these signals, which are hard to consider, and can be related to some physiological properties, if they are biomedical signals, or to other motivations if they are not biomedical. EEG, HRV are examples of biomedical signals with deep complexity , due to human body dynamics, which can not be easily analyzed with classical Signal Analysis methods.

Complexity Analysis emerged in order to understand the architecture of these signals, and Entropy in particular. It can be considered as a measure of disorder, complexity and unpredictability of a generic signal. Since the first definitions of Entropy provided by Shannon in 1948, many more were introduced, producing a big range of features to be applied in order to characterize complex systems.

The relevance of Entropy-based Analysis is related to the detection of possible subtle changes behind the dynamic of a certain signal. In biomedical fields, it can be very useful to discern pathological or physiological conditions in subjects. A real time Signal Analysis could detect the onset of a certain clinical disease, and countering it immediately. It could also be applied to monitor the progression of a certain disease. Clinical decision-making relies on Signal Analysis as a relevant tool to understand the subject condition, and to eventually alter the medical therapy.

Not all signals give a good performance if analyzed with Entropy. Also, not all Entropy definitions produce the same performances, with some of them giving the best results when applied on data suited for them. Length of data, noise, or particular conditions on signals can always alter the performance of the Complexity Analysis. Therefore,

understanding the suitability of an entropic feature on certain type of data could lead to a better Complexity Analysis, and consequently to a better detection of possible critical conditions. In biomedical applications, this could significantly improve the possibilities of a quick diagnosis, or improve the responsiveness in monitoring a diseased subject.
The work in this Thesis focused on the comparison between performances obtained from classification with Entropy-based or Time/Frequency features. The aim was to understand if Complexity Analysis could add a statistically significant increase in the accuracy from classification of biomedical signals' epochs. Analysis focused on HRV and EEG databases, involving recordings from subjects with clinical conditions such as for example heart failure, or epilepsy.
A positive outcome could boost future applications of Entropy in systems for real-time detection of pathological conditions, for example in a Brain Computer Interface (BCI) applied on epileptic subjects, or to detect possible worsening of heart functionality and contractility in a diseased subject. Complexity Analysis represents a tool with still boundaries yet to define. This work focused on understanding whether it could possibly get more applications and usability in fields such as Biomedical Signal Analysis, with possible further developments.

1. Entropy concept

Nature offers a wide distribution of systems. Some systems can be considered as predictable, some as random, and some as chaotic. These three definitions are not equivalent:

- a **predictable system** is ruled by some simple mathematical models. The result is a predictable evolution of it, which can be translated into a predictable signal. A classical example of a predictable signal is a sinusoidal one, which has some simple laws to describe it.
- A **random system** has a non deterministic nature. There are no rules behind it, leading to a stochastic behavior, and an impossible prediction of it. Also, initial condition of the system is totally unrelevant to the future evolution of it. Signals with a big randomness contribute as for example a gaussian distribution can be considered as related to a random system.
- Even though they could look similar, a **chaotic system** is very different from a random one. Many complex and deterministic dynamics are the base for a chaotic system. Lack of predictability of a signal considered as chaotic can be related to the difficulty into the comprehension of the complex laws of it. But the future evolution of a chaotic signals can not be considered as stochastic, having some deterministic dynamics to define it. A Lorenz attractor is a classical example of a complex and chaotic signal, which is not random.

Since these definitions were introduced, Complexity took a significant position into Signal Analysis. Whether a predictable system can be easily described and predicted with linear features, it is not the same for chaotic signals. Complexity Analysis committed itself into trying to understand the laws ruling a chaotic system, and to possibly predict future behavior of chaotic signals. Many Complexity features were introduced with the aim to get a better overview of the problem, and Entropy is one of them.
Entropy is a fundamental concept originally rooted in thermodynamics and information theory, broadly interpreted as a measure of disorder, uncertainty, or complexity within

a system. It is a concept that can be easily moved to the analysis of many different signals, from biomedical recordings, to earthquakes, to sounds. Entropy can be applied to signals of different nature, even though its performance on them is always related to some signal characteristics.

Over the past few decades, the development of entropy methods in signal analysis has paralleled advanced in nonlinear dynamics and complexity science. As early as the 1990s, researchers recognized that many signals exhibit characteristics of chaotic systems, leading to the adaptation of entropy measures to detect subtle changes in physiological states [1]. Since then, Entropy-based signal analysis continues to increase its relevance,driven by its ability to reveal hidden patterns and capture meaningful information from the complex variability inherent in systems (especially living systems). One of the biggest applications of entropy is the Biomedical Signal analysis. The intrinsic variability observed in signals such as electrocardiograms (ECG), electroencephalograms (EEG), and electromyograms (EMG) reflects underlying physiological mechanisms and regulatory processes. Traditional linear methods often fail to capture the complex, nonlinear dynamics present in these signals. Entropy-based measures, by contrast, are capable of assessing the signal's structural richness and its underlying dynamical behavior without the need for predefined models [2]. The application of entropy in biomedical contexts is particularly relevant for detecting pathological conditions, monitoring disease progression, and evaluating treatment efficacy. For example, alterations in signal complexity can indicate a breakdown in the autonomic regulation of the heart or abnormal neural synchronization. Moreover, entropy measures are largely non-invasive, computationally efficient, and adaptable to real-time analysis, making them well-suited for both clinical and wearable technologies [3]. Entropy serves as a conceptual and computational bridge between the raw data of biomedical signals and the interpretation of physiological or pathological states.

A variety of entropy formulations have been proposed and applied. Each one presents its own peculiarities, which give them advantages and disadvantages if applied to evaluate data of different nature,length, and origin.

In this chapter, the definitions involved will be explored, and described in their peculiarity and notable signs.

1.1 Shannon Entropy

$$H(X) \ = -\sum_{i=1}^{n} p_i \log_b(p_i) \tag{1.1}$$

Introduced by Claude E. Shannon in 1948 [4], is a fundamental concept in information theory used to quantify the amount of uncertainty or information contained in a probability distribution. It measures the average amount of information required to describe the outcome of a random variable. In the context of time series analysis, Shannon Entropy assesses the unpredictability of the system generating the signal. A high entropy value indicates that the system exhibits high complexity or randomness, whereas a low entropy value reflects regularity or predictability.
Mathematically, given a discrete set of outcomes with associated probabilities (*p1, p2, . . . , pN*), the Shannon Entropy is defined as Equation 1.1.

- (*pi*) represents the probability of the (*i*)-th outcome,
- (*b*) denotes the base of the logarithm (commonly $b = 2$ to express entropy in bits).

The entropy is way higher when all outcomes are equally probable, and has its minimum when one outcome is certain (i.e., $pi = 1$ for some i , and 0 for all others). This leads to a lower limit for Entropy score, corresponding to Entropy=0. Instead, there is no upper limitation for Entropy value.
This measure forms the theoretical foundation of many entropy-based complexity metrics used in biomedical signal analysis, data compression, and machine learning, including Permutation Entropy, Dispersion Entropy, and Sample Entropy.

1.2 Renyi Entropy

The *Rényi Entropy*, introduced by Alfréd Rényi in 1961 [5] , is a generalization of the Shannon entropy that introduces a parameter to control the sensitivity of the entropy measure to the distribution of probabilities. It is particularly useful for analyzing

systems with varying degrees of randomness or concentration in their probability distributions.

Given a discrete probability distribution (*p1, p2, . . . , pn*) and a non-negative real parameter alpha different from 1 , the Rényi entropy of order *alpha* is defined as

$$\underset{Entropy(\alpha)}{Renyi} = \frac{1}{1-\alpha} \log_b \left(\sum_i^n p_i^\alpha \right) \tag{1.2}$$

Rényi entropy is particularly useful in signal processing, complexity analysis, and information-theoretic learning, where one wishes to explore the structure of probability distributions beyond the standard Shannon framework. It allows for a more flexible evaluation of uncertainty depending on the context and desired emphasis on rare or frequent events.
The parameter *alpha* determines how the entropy weighs the contribution of different probabilities:

- When *alpha* greater than 1 , the entropy becomes more sensitive to high-probability events.
- When *alpha* is lower than 1 it becomes more sensitive to low-probability events. Do not forget that *alpha* is always a non negative parameter.
- In the limit as *alpha* equal to 1 , Rényi entropy converges to the Shannon entropy.

1.3 Approximate Entropy

Approximate Entropy (ApEn), introduced by S.M.Pincus in 1991 [1], is a statistical measure used to quantify the *regularity* and *unpredictability* of fluctuations within time-series data. It is particularly effective for analyzing *short and noisy* datasets, where traditional entropy-based methods may be less reliable. ApEn estimates the likelihood that similar patterns in the data will remain similar at the next time step. In essence, it provides a numerical description of the complexity or irregularity of a signal.

The mathematical definition of Approximate Entropy is given by:

$$\mathrm{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{1.3}$$

where:

- m is the embedding dimension, defining the length of the vectors being compared;
- r is the similarity tolerance (or threshold);
- N is the total length of the time series.
- *tau* is the time delay between every element of the embedding vector.

The function $\phi^m(r)$, also known as the Correlation Integral, measures the average logarithmic frequency of similar vectors of length m:

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \tag{1.4}$$

Each quantity $C_i^m(r)$ represents the proportion of vectors x_j^m that are within a distance r of x_i^m (including self-matches):

$$C_i^m(r) = \frac{\left|\left\{j \mid j \neq i,\ d(x_i^m, x_j^m) < r\right\}\right|}{N-m} \tag{1.5}$$

Since self-matches are included, the distance between the vector x_j^m and itself is always zero. As a result, there is always at least one match for each $C_i^m(r)$, which is therefore always at least equal to $\frac{1}{N-m}$. This introduces a slight bias in the estimation, because self-matches artificially increase the similarity count, thus affecting the evaluation of the signal's complexity.

The vectors x_i^m are constructed by embedding the time series u(i) into m-dimensional space:

$$x_i^m = [u(i), u(i+1), \ldots, u(i+m-1)] \tag{1.6}$$

The distance function d(x_i^m, x_j^m) is typically defined using the maximum norm (also

known as the Chebyshev distance):

$$d(x_i^m, x_j^m) = \max_{k=1,\ldots,m} |u(i+k-1) - u(j+k-1)| \tag{1.7}$$

The fundamental idea of Approximate Entropy is that in a highly regular time series, sequences that are similar for m consecutive points are likely to remain similar when extended to $m+1$ points. Therefore, the difference $\phi^m(r) - \phi^{m+1}(r)$ will be small, resulting in a low ApEn value, which indicates a predictable, low-complexity signal.
In contrast, a big difference $\phi^m(r) - \phi^{m+1}(r)$ can be the result of a complex, chaotic signal, leading to an higher value of ApEn.

- **Low ApEn values** indicate more regular and predictable behavior.
- **High ApEn values** suggest major complexity or unpredictability.

Since ApEn compares the similarity between vectors of length m and $m+1$, the two main parameters, m (embedding dimension) and r (similarity threshold), are crucial in defining the behavior of the measure. They respectively describe the length of each vector, and the threshold for the similarity between them. The value of r is always then multiplied for the standard deviation of the signal itself.
The threshold is evaluated as *r*std(signal)*.
By varying the parameters m and r, it is possible to generate different Approximate Entropy features from the same time series. Each combination captures different aspects of the signal's temporal dynamics, and these features may exhibit different performance depending on the application (e.g., classification, prediction, anomaly detection).
Approximate Entropy involves both amplitude and time analysis in the signals. Each element in the vectors has a time delay of tau between them: with this approach, the embedding vector becomes a temporal section of the signal. The distance evaluation in the phase space is also sensible to possible shuffling of the vectors: a different order of the elements inside each of them could lead to different distance evaluations. It could be possible to obtain a different number of recurrences from vectors shuffled, when compared with recurrences obtained from the original vectors. This makes the whole Approximate Entropy evaluation sensible to both temporal ordering and amplitude values inside the vectors.
Compared with the Shannon and Renyi Entropy definitions, which were mainly based on probability of a certain outcome, Approximate Entropy provides a more complete analysis of the complexity of the signal, analyzing also temporal and amplitude dynamics on it.

1.3.1 Modified Approximate Entropy

In this work was also included a correction of the Approximate Entropy, following the indications provided by L. Mesin in the related paper. [6]. This definition involves some adaptation and indications related to the values of parameters correlated to this complexity index. The original definition from Pincus [1] creates a quite stable complexity evaluation, and suggests some values of the parameters m and r. The embedding dimension was suggested to be kept on low values, for example m=2 or 3, and the value of r set to a suggested 0.2 *standard deviation (of the signal). A more complete analysis needs to consider also other factors that could alter the evaluation:

- Approximate Entropy is sensitive to the sampling frequency [6]. If a signal gets oversampled, the result would be a linearization of it, with consequently a more predictable pattern, leading to biased lower Entropy.

- When the value of r is set (for example r=*0.2*std*) the evaluation is very sensible to the epoch length. In particular, having a value of m greater than 2 , in a short epoch, could provide few recurrences in the evaluation, both for vectors of dimension (m) and (m+*1*).

- The logarithm in the definition could prioritize rare events, giving them a more informative role. [6]This is related to the definition itself of Entropy, but having short epochs could possibly create unstable evaluations.

- For a complexity evaluation with Approximate Entropy a complete removal of low-frequency trends is mandatory. This preliminary processing needs to be related to the nature of the signal itself, and to be applied carefully, to avoid any possible alteration of the analysis.

- Values of signal way too close in time get considered as neighbors in the Approximate Entropy estimation. This leads to always have a minimum number of recurrences in both embedding dimensions. The result is a bias towards lower Approximate Entropy values.

Modified Approximate Entropy [6] faces the previous described issues with the following corrections:

- A signal needs always to get resampled to a value *fs* which needs to be at least the double of the signal bandwidth. This is needed in order to avoid any possible problems with oversampled signals, and Entropy biased to lower values.

- The time delay between values could be set to values different from 1, in order to get a better relation with the sampling frequency from the signal [6];
- The tolerance value r is chosen in order to percentage of points found when considering the embedding dimension m. This avoids the problem of getting too few recurrences when the epoch length is short. [6]
- The total number of embedded vectors is the same, both in dimension m and $m+1$. In lower dimension, the evaluation usually would provide a greater number of vectors. In this index, the last vectors in lower dimension are discarded, avoiding the possibility of obtaining negative index values.
- Low frequency trends are carefully removed, always considering the sampling frequency applied in each situation.
- A Theiler window is introduced.[6] This application removes points closer in time than a delay (*tau*), avoiding both self-recurrences, and recurrences driven by points way too close in time.
- Points with no recurrences in dimension m get discarded, being not meaningful for the analysis.
- Points with recurrence in dimension m and no recurrence in dimension $m+1$ get enhanced, by adding one recurrence to the integral. This helps the evaluation. Approximate Entropy relies on this kind of differences, whether the embedding dimension provides different number of recurrences. Adding a value gives further more relevance to these situations.

These corrections can provide a more stable evaluation of Approximate Entropy, less sensible to the choice of parameters, like for example the value of the sampling frequency. [6] This index provided good results in many applications, and it was decided to include it in the subset of involved Entropy definitions to assess also its applicability to these datasets.

1.4 Sample Entropy

Sample Entropy (SampEn) is a widely used measure, introduced by Joshua S. Richman and J. Randall Moorman in 2000. [7] It was introduced to overcome certain limitations

of Approximate Entropy (ApEn), such as the inclusion of self-matches and its sensitivity to the length of the data set.

Formally, SampEn is defined as the negative natural logarithm of the conditional probability that two sequences similar for m consecutive points remain similar at the next point, within a tolerance r, as the data length N tends to infinity. This is expressed as

$$\text{SampEn}(m, r) = \lim_{N \to \infty} -\log \left(\frac{\sum_{i=1}^{N-m} C_i^{m+1}(r)}{\sum_{i=1}^{N-m} C_i^{m}(r)} \right) \tag{1.8}$$

where $C_i^m(r)$ represents the count of template vectors of length m that lie within a distance r from the *i-th* template vector, excluding self-matches. Both SampEn and ApEn utilize the parameters m (embedding dimension) and r (similarity tolerance), in order to evaluate similarities between vectors. As for ApEn, the value of r needs to be multiplied for the standard deviation of the signal in order to obtain the threshold level used to count similarities. As it is possible to understand from the two formulations, there are many similarities between ApEn and SampEn.
The principal distinction between SampEn and ApEn lies in the exclusion of self-matches in the computation of SampEn, which effectively reduces bias and enhances the consistency of the entropy estimates, particularly for shorter time series. In contrast, ApEn includes self-matches when counting similar patterns, which may lead to overestimation of regularity.

Moreover, SampEn demonstrates reduced dependence on the length of the time series, thereby providing more reliable complexity estimates in practical scenarios where data may be limited. It shows a good performance with short and noisy signals, where Approximate Entropy may struggle with short time series, where there are a limited number of vectors obtained .
In conclusion, Sample Entropy refines the approach to complexity measurement by improving the probabilistic evaluation of repeated patterns within a time series, thus addressing key drawbacks of Approximate Entropy and enhancing its applicability across diverse domains such as physiology, finance, and engineering.

1.5 Permutation Entropy

Permutation Entropy (PE) is an entropy-based measure introduced by C. Bandt and B.Pompe in 2002 to quantify the complexity of time-series data by analyzing the ordinal patterns within the series [8]. Unlike other entropy measures that rely on amplitude information, PE focuses on the temporal ordering of values, making it particularly robust to noise and invariant to nonlinear monotonic transformations.

Mathematically, given an embedding dimension m, PE is defined as the Shannon entropy of the distribution of ordinal patterns of length *m* extracted from the time series. This is expressed as

$$PE(m) = -\sum_{i=1}^{m!} p_i * \log_b(p_i) \tag{1.9}$$

where *pi* denotes the probability of the *i-th* ordinal pattern occurring in the time series, *m!* represents the total number of possible ordinal patterns of length *m*, and *b* is the base of the logarithm, typically chosen as 2 for expressing entropy in bits. The embedding vectors are used also in this definition, but with a different approach.
Signal amplitude is used to assign each embedding vector to a certain ordinal pattern. For example, if the vector is [4;7;9], comparing the amplitude of values inside the vector it could be assigned to the ordinal pattern [0;1;2]. Amplitude comparison is applied just inside the vector, and just to understand the match of the vector to one of the *m!* different ordinal patterns. Temporal order is crucial because a potential shuffling of the values inside the vector could assign it to a different ordinal pattern. For example, considering the previous vector, if after the shuffling we had [7;4;9], it would have been pattern [1;0;2], which is totally different from the original one. This makes Permutational Entropy strongly dependent on temporal ordering.
Considering the whole signal, it is possible to evaluate the probability of each permutational pattern, and from that to evaluate the Shannon Entropy of the *m!* possible patterns. The key advantage of Permutation Entropy lies in its ability to capture the intrinsic dynamical properties of the system by examining the relative ordering of data points rather than their actual values. This characteristic renders PE highly effective in detecting changes in complexity and dynamics, especially in noisy or non-stationary environments.

Compared to other entropy measures such as Approximate Entropy and Sample Entropy, PE is computationally efficient and less sensitive to parameter choices. Its reliance on ordinal patterns rather than amplitude thresholds allows it to provide a

complementary perspective on time-series complexity, often revealing features that other methods might overlook.

In summary, Permutation Entropy offers a powerful and versatile tool for analyzing complex signals across various disciplines, including physics, biology, and economics, by leveraging the concept of ordinal pattern distributions to quantify temporal complexity.

The embedding dimension m is the main parameter considered into the Permutation Entropy definition.

1.6 Dispersion Entropy

Dispersion Entropy (DispEn) is a relatively recent complexity measure designed to quantify the irregularity of time-series data by analyzing the dispersion patterns of normalized embedding vectors [9], introduced by M.Rostaghi and H.Azami in 2016. It extends traditional entropy methods by incorporating amplitude information , providing a sensitive and robust evaluation of signal complexity. Temporary information is less involved, only for the creation of the embedding vectors.

Formally, Dispersion Entropy is defined as the Shannon entropy of the distribution of dispersion patterns derived from the normalized embedding vectors. DispEn is expressed as

$$\text{DispEn}(m, c, \tau) = -\sum_{k=1}^{c^m} p_k \log(p_k) \tag{1.10}$$

where m is the embedding dimension, c denotes the number of classes used for mapping the normalized time series values, and *tau* is the time delay.

The embedding vectors are constructed as

$$\mathbf{x}_i^m = \left[x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}\right] \tag{1.11}$$

To reduce the influence of amplitude variations and to facilitate the classification of values into discrete categories, these embedding vectors are normalized.
A common approach consists of applying the *z-score* normalization:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1.12}$$

where μ and σ represent the mean and standard deviation of the original time series x, respectively [10].

Following normalization, the continuous values *zi* are discretized into *c* classes by partitioning the range of normalized values into equal intervals.
Each normalized embedding vector

$$\mathbf{z}_i^m = \left[z_i, z_{i+\tau}, z_{i+2\tau}, \ldots, z_{i+(m-1)\tau}\right] \tag{1.13}$$

is mapped into one of the c^m possible dispersion patterns *xk*. The class assignment is typically performed as

$$\tilde{z}_i = \text{floor}\left(c \times \frac{z_i - z_{\min}}{z_{\max} - z_{\min}}\right) + 1, \quad \tilde{z}_i \in \{1, 2, \ldots, c\} \tag{1.14}$$

where *z min* and *z max* denote the minimum and maximum values of the normalized data, ensuring all values are mapped within the predefined classes.The floor function, returns the greatest integer less than or equal to a given real number. This operation is commonly used in quantization and discretization procedures, such as those required in the computation of Dispersion Entropy. In this context, real-valued normalized data are mapped into discrete categories or classes.
For instance, if a normalized value *zi* is to be assigned to one of *c* equally spaced classes, the floor operation is applied, and this ensures that *zi* is mapped to an integer class label within the range (1, 2, . . . , c). The addition of 1 shifts the range of the result to start from 1 instead of 0, which is often more convenient for indexing purposes in entropy calculations.
The probability of each dispersion pattern occurring is then computed as:

$$p_k = \frac{\left|\left\{i \mid \mathbf{z}_i^m = x_k\right\}\right|}{N - (m-1)\tau} \tag{1.15}$$

where N is the length of the original time series. The strength of Dispersion Entropy lies in its ability to capture both amplitude variations by classifying normalized values into discrete categories before entropy calculation. This approach enhances the sensitivity to subtle dynamical changes while maintaining computational efficiency. It is a brand new approach, because as we saw with previous definitions (Shannon, Renyi, Approximate, Sample, Permutation) in none of them was involved a classification of the values obtained.
The temporal information is involved just in the creation of embedding vectors. This makes it very different from other Entropy definitions, like Permutation Entropy for example, where temporal order of values was an important feature for it. Dispersion Entropy works way more on the classifications of the amplitude values from the em-

bedding vector. In this Entropy definition, *m, c* and *tau* are the 3 parameters involved.

Where *m* is always the embedding dimension, and *tau* the time delay of the embedding vector, *c* corresponds to the number of classes involved in this evaluation. It is possible to obtain different features from Dispersion Entropy, just by combining different values of these parameters.

1.7 Distribution Entropy

Distribution Entropy (DistEn) is a statistical measure designed to quantify the complexity and randomness of a time series by analyzing the distribution of distances between embedding vectors, introduced in 2019 by M.Rostaghi,H. Azami . [11].

Given a time series $x = (x1, x2, \ldots, xN)$, the first step involves constructing the embedding vectors as:

$$x_i^m = \left[x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}\right] \tag{1.16}$$

where *m* is the embedding dimension and *tau* is the time delay.

Next, for each pair of embedding vectors x_i^m and x_j^m, a distance metric is computed to evaluate their similarity. Two commonly used distance functions in the context of Distribution Entropy are the Euclidean and Chebyshev distances:

Euclidean Distance:

$$d_E(x_i^m, x_j^m) = \sqrt{\sum_{k=0}^{m-1} \left(x_{i+k\tau} - x_{j+k\tau}\right)^2} \tag{1.17}$$

Chebyshev Distance:

$$d_C(x_i^m, x_j^m) = \max_{k=0,\ldots,m-1} \left|x_{i+k\tau} - x_{j+k\tau}\right| \tag{1.18}$$

Once distances are computed for all valid vector pairs, they are discretized into *c* bins or classes. The Distribution Entropy is then calculated as the Shannon entropy of the resulting histogram of distances:

$$\mathrm{DistEn}(m, c, \tau) = -\sum_{k=1}^{c} p_k \log p_k \tag{1.19}$$

where *pk* is the empirical probability of the distance falling into the *k -th* bin.

1.7.1 Comparison with Approximate Entropy

While conceptually related to other entropy measures, such as Approximate Entropy (ApEn), DistEn differs because it doesn't work with a distance threshold (r for ApEn) but evaluates the Shannon Entropy of distances, discretized in c bins.
There are many similarities with ApEn, both involving amplitude and temporal evaluations. This makes them very sensible to possible shuffling of values inside the embedding vector, because alters the distance of it from the neighbors.

1.7.2 Comparison with Dispersion Entropy

DistEn shares structural similarities with Dispersion Entropy, such as the use of embedding vectors and classes. However, there are key differences.
While DispEn categorizes the values of embedding vectors directly into discrete classes based on normalized amplitude values, DistEn focuses on the distribution of distances between these vectors, capturing a more global view of the time series' structure. Dispersion Entropy provides only an amplitude evaluation of the vector (normalized and classified) while Distribution Entropy gives also a temporal evaluation of the vector itself, by working with distances(classified).
As for Dispersion Entropy, *m, c* and *tau* are the 3 parameters involved. The difference with Dispersion Entropy is that in the Distribution Entropy c is the number of classes where distances are classified. In Dispersion , classes were referred to the amplitude of vectors obtained.
This difference makes DistEn more sensitive to subtle dynamical changes and long-range correlations in the data, whereas DispEn may be more suitable for detecting local variations and abrupt shifts. Moreover, DispEn requires a pre processing step for normalization and class assignment, whereas DistEn relies solely on inter-vector distances, potentially simplifying parameter tuning. In summary, Distribution Entropy provides a complementary perspective to Dispersion Entropy by utilizing a distance-based approach, allowing for an alternative and potentially more robust characterization of complexity in time-series data.

1.8 Bubble Entropy

Bubble Entropy (*BubbEn*) is a recently proposed entropy-based complexity measure designed to quantify the unpredictability of time series by analyzing the ordering of

elements. It was introduced by Manis, M. Aktaruzzaman and R.Sassi in 2017, and represents an improvement on existing permutation-based entropy metrics by leveraging the number of switches required to sort segments of the signal via the Bubble Sort algorithm [12].

Formally, the Bubble Entropy for a given embedding dimension *m* and time delay *tau* is defined as:

$$\text{BubbEn}(x; m, \tau) = H_m - H_{m+1} \tag{1.20}$$

Here, *H(m)* and *H(m+1)* represent the Shannon entropy of the switching count distributions for embedding dimensions *m* and (*m+1*), respectively. Each entropy value is calculated as:

$$H_m = -\sum_k p_k^{(m)} \log p_k^{(m)} \tag{1.21}$$

where $p_k^{(m)}$ denotes the empirical probability of observing k switches (i.e., swaps) in the Bubble Sort process when sorting vectors of length m extracted from the time series.

The core idea of Bubble Entropy lies in applying the Bubble Sort algorithm to local patterns (subsequences) of the time series, obtained via phase-space reconstruction. Given a time series *x = (x1, x2, . . . , xN)*, one first constructs a sequence of embedded vectors:

$$\mathbf{v}_i = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}]$$

Each embedded vector *vi* is then sorted using Bubble Sort, and the number of swaps required to fully sort each vector is recorded. The distribution of these switch counts across all embedded vectors of dimension *m* provides a basis for calculating the entropy *Hm*.

1.8.1 Bubble Sort

Bubble Sort is a sorting algorithm which sorts the vector in rising order, considering the amplitude of the vector elements. For example, if the vector has values [10,6,8], after the sort the order will be [6,8,10] Amplitude is used to sort, but it is not applied in absolute sense, but just related to the vector: if we had [100,60,80] the result would have been the same. The number of switches needed in this example would be 2:

- we start with the vector [10,6,8]. The first comparison involves 10 with 6, which get switched. The result is [6,10,8].

- now 6 gets compared with 10, but they are sorted, so no switch is happening. Then, 10 gets compared with 8, and the second switch happens. The result is [6,8,10], and the vector is now sorted.

Bubble Sort is particularly suited for this purpose due to its simplicity and deterministic switch behavior, which allows for consistent characterization of the time series' local ordering patterns. Although Bubble Sort is inefficient for large-scale sorting, in this context it serves purely as a tool to extract structural information from short subsequences. A highly regular signal will require a predictable number of switches to sort its vectors, leading to a lower entropy. In contrast, a more chaotic or irregular signal will yield a more varied switching distribution, resulting in higher entropy values.

This makes Bubble Entropy particularly useful in fields such as biomedical signal analysis, where it has been applied to distinguish healthy from pathological signals in ECG and EEG datasets [12, 13]. Compared to other entropy measures like Permutation Entropy or Sample Entropy, Bubble Entropy is more sensitive to subtle changes in local ordering without relying on amplitude information or arbitrary thresholds. Since it relies on the relative order rather than exact values, it is less sensitive to high-frequency noise.

Its only parameters are the embedding vector dimension (m) and time delay(tau). There are no classes, or threshold involved into this definition. It also involves both temporal and amplitude approaches. A shuffled vector would need a different number of switches to be sorted, leading to a totally different value of Bubble Entropy. The amplitude is also crucial, due to Bubble Sort algorithm using it in order to obtain the sorted vector.

2. Databases

Entropy, as already explained, is a concept that can be applied for signal analysis of signals of different origins. Studies have shown that can be useful to asses chaotic patterns especially in biomedical signals [3] , but not only with them. Many works,showed as entropy can be used also into seismic or sounds analysis[14]. It is indeed an analysis of the disorder, unpredictability and uncertainty of a signal. Some signals can be more suited than others to Entropy Analysis, showing sensible difference between values from different subgroups of subjects. For example, HRV signals shown good performance in Entropy Analysis, highlighting pathological condition and progression of diseases. Many other signals don't have the same performance with this Complexity Analysis, not being able to discern between signals related to different subgroups.

In this work, were taken in consideration different nature signals, in order to evaluate Entropy Analysis performance on all of them. The databases included were :

- **CHF.**[2] HRV data from subjects Congestive Heart Failure (CHF).
- **NSR.**[2] An amount of HRV recordings from subjects with normal heart activity, and no pathological conditions.
- **FANTASIA.**[15] A study involving HRV recordings from young and old subjects.
- **EEG SIENA.** [16] EEG recordings from epileptic subjects.

2.1 CHF and NSR

The two databases were analyzed and compared with each other, being considered as an unique database. The reason behind this choice is because they both have HRV recordings of similar length, and acquisition method. This led to an easier comparison and evaluation in different values of Entropy, between the two subgroups.

Each subgroup was corresponding to the original CHF or NSR database, creating a pathological subgroup (CHF) and an healthy one (NSR).

- **CHF**[2]: this database can be found on Physionet, and includes long-term ECG recordings from 15 subjects (11 men, aged from 22 to 71, and 4 women, aged 54 to 63) with sever congestive heart failure (NYHA class 3-4). From these 15 subjects, were obtained 29 recordings[17] . This group of subjects was part of a previous study [18], focused into analyzing the possible advantages of a therapy with an oral inotropic agent, called milrinone. The original aim was to test if this therapy could lead to some improvements in cardiac contractility in subjects with Heart Failure. The individual recordings in the database are each about 20 hours long, recorded in 1986 at Boston's Beth Israel Hospital, after an oral dose of milrinone. From the ECG recordings, it was possible to extract the HRV recordings.

- **NSR**[2]: this database includes beat annotations from 54 long term ECG recordings of subjects in Normal Sinus Rhythm (NSR).Phyllis Stein (Washington University School of Medicine, St. Louis), and Rochelle Goldsmith (Columbia-Presbyterian Medical Center, New York) made the data available to everyone on Physionet. The group involves 30 men, aged 28 to 76, and 24 women, aged 58 to 73. They were clinically healthy, and their ECG recordings can be considered as normal. From the ECG recordings, were then extracted the HRV signals.

The similarities between data in these 2 databases made the comparison of Entropy values easier, since they are both long-term ECG recordings. It was decided to work on the HRV signal, preferred over the ECG, because it was more likely to appreciate larger differences in Entropy produced from the two subsets.[19] This choice was based on several important factors:

- HRV reflects the variations in time intervals between successive R-peaks (RR intervals) in the ECG, which are closely associated with autonomic nervous system activity. Since Entropy Analysis aims to capture the dynamic complexity of physiological regulation, the RR intervals provide a direct and relevant representation of such regulatory mechanisms. In contrast, the raw ECG signal contains redundant morphological information (*P wave, QRS complex, T wave*) that may not contribute meaningfully to entropy-based metrics and may even introduce noise.

- HRV is a low-dimensional, derived signal that simplifies the data to focus on temporal dynamics. This reduction makes Entropy calculations more computationally efficient and less sensitive to high-frequency noise and baseline fluctuations commonly found in ECG recordings.

- Several studies [19] have shown that Entropy measures applied to HRV are more effective in detecting autonomic dysfunction and predicting adverse clinical outcomes than when applied to the full ECG signal.

The use of HRV over ECG for Entropy Analysis is in the end motivated by its physiological relevance, simplicity, robustness to noise, and proven efficacy in clinical research. These factors make HRV the preferred choice for quantifying heart-related complexity using Entropy metrics.

2.1.1 HRV signal

Heart Rate Variability (HRV) is a biomedical signal that describes the time interval between consecutive heartbeats. It can be derived by recording the time difference between each beat, obtaining a series of time values , typically expressed in milliseconds (ms). This measurement can also be obtained considering the time between RR peaks in ECG waveforms, as illustrated in Figure 2.1. Although this method is quite easy, automatic implementation can be noise sensible, altering the peak detection. This highlights the importance of applying preprocessing techniques to the ECG signal, in order to ensure the most correct results possible.
An R-peaks detection algorithm is then involved, on preprocessed ECG signals. The most common are:

- Pam-Tompkins algorithm;[20]
- Wavelet Stationary Transform (SWT) ;[21]
- Hilbert transform; [22]
- Combination of both SWT and Hilbert trasform;[23]
- Deep Learning approach. It has the aim to improve the recognition of the ECG waveforms, especially of the R peaks. The most used for this detection is RpNet, which provided a good quality performance when applied to ECG recordings.[24]

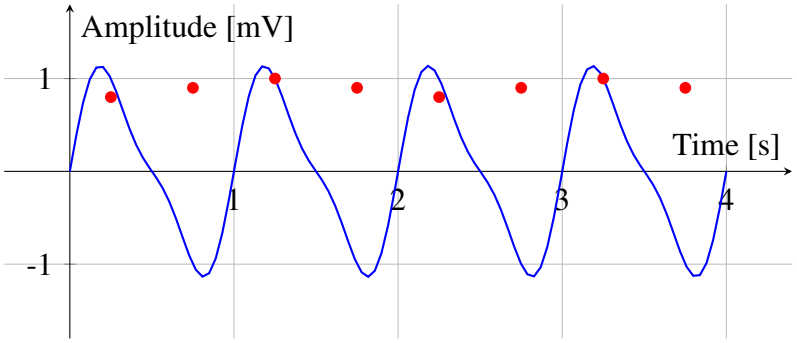


Figure 2.1: ECG signal with annotated R-peaks.

Once all R-peaks have been detected with one of the previous algorithms, the RR interval can be computed as the time-distance between consecutive peaks. Sometimes it is possible to get some ectopic values, or some outliers in the series of RR intervals. It is needed an artifact removal process, in order to keep just RR intervals with reasonable values in terms of Amplitude and time-distance from the closest peaks.

As shown in Figure 2.2, the steps to extract HRV from ECG are not that many, and can be computed quite easily.

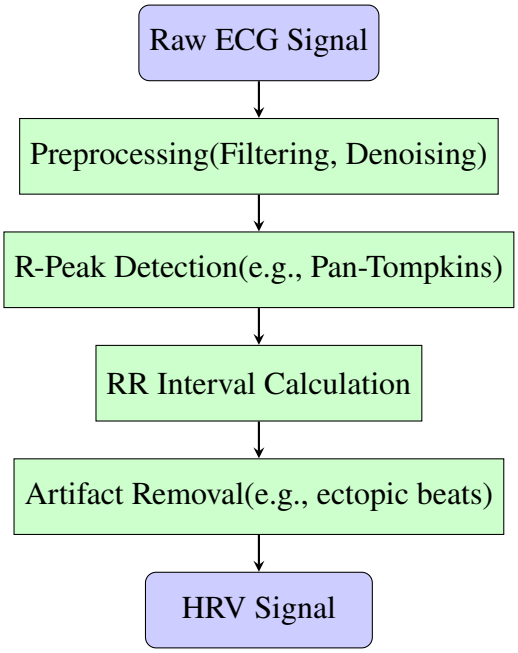


Figure 2.2: Flowchart for extracting HRV from ECG signal

The signal obtained is made of time values, usually expressed in milliseconds(ms).

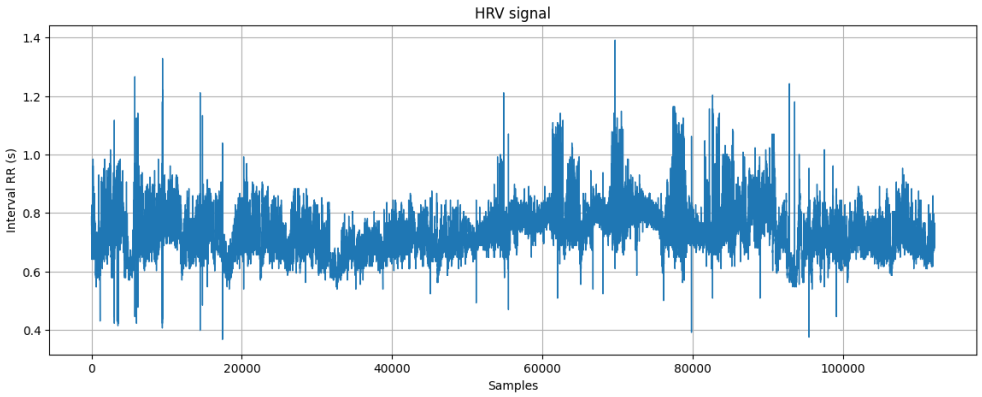


Figure 2.3: HRV signal timeplot from database CHF

As shown in Figure 2.3, the values are not always so regular. This variability comes from the fact that Heart Beat is very sensible to every possible alteration factor, and the human auto-regulation system produces corrections and adjustments in a continuous way. An healthy subject shows big variability in HRV. It is a sign that his body, and his

cardiac system, can produce modulations to the rhythm in a quick way.
Every single action, movement, or also emotional response can generate an alteration of the HRV signal, making it extremely unpredictable, and very interesting for an Entropy-based Analysis.
A subject with health disease, or an elderly, can show a reduction in HRV: it could get loosen the ability of their body to adjust the heart beat in answer to stimulations. The result would be an HRV signal much more stable, with an Entropy value very different from the one produced from an healthy HRV recording.

The frequency-domain analysis of Heart Rate Variability (HRV) is commonly used to assess autonomic nervous system (ANS) activity. It involves decomposing the RR interval time series into specific frequency bands, each associated with distinct physiological mechanisms. [25] [26] [27] [28] The typical frequency bands used in HRV analysis are:

- **Ultra Low Frequency (ULF) – < 0.003 Hz:** This band reflects extremely slow oscillations and is typically only measurable in long-term (e.g., 24-hour) recordings. It is believed to be influenced by circadian rhythms, core body temperature, metabolism, and long-term regulatory mechanisms. Its physiological origin remains less well understood compared to other bands.
- **Very Low Frequency (VLF) – 0.003–0.04 Hz:** The VLF band is associated with thermoregulatory and hormonal mechanisms, including the renin–angiotensin system. Although less frequently analyzed in short-term recordings, it is considered a marker of sympathetic activity and underlying physiological stress.
- **Low Frequency (LF) – 0.04–0.15 Hz:** The LF band reflects both sympathetic and parasympathetic activity, although its interpretation remains debated. It is influenced by baroreflex activity and is often considered an indicator of blood pressure regulation. When normalized, LF is sometimes used as a marker of sympathetic modulation.
- **High Frequency (HF) – 0.15–0.4 Hz:** The HF band corresponds to respiratory sinus arrhythmia and is strongly associated with parasympathetic (vagal) activity. It reflects short-term variability linked to the respiratory cycle and is considered a robust indicator of parasympathetic nervous system function. [26]

Understanding the distribution of spectral power within these bands enables researchers and clinicians to infer the balance between sympathetic and parasympathetic influences on cardiac function.

HRV Frequency Band	Frequency Range (Hz)
Ultra Low Frequency (ULF)	0- 0.003
Very Low Frequency (VLF)	0.003 – 0.04
Low Frequency (LF)	0.04 – 0.15
High Frequency (HF)	0.15 – 0.4

Table 2.1: Typical HRV frequency bands and their respective ranges.

The different activity of the subject can be analyzed from the Power Spectral Distribution of the HRV signal. Different activities from sympathetic and parasympathetic systems can produce alterations in the power distribution in LF and HF band. They are two sides of the autonomous nervous . [29] Sympathetic system is involved when a quick response is needed and triggered from external or internal stimulations. The main outcomes of its activity are :

- **Cardiac Frequency Raise;**
- **Pupils dilatation;**
- **Lungs dilatation;**
- **Digestion slowing;**
- **Glucose blood release.**

All those effects contribute to a better handling of stressing situations, where the subject needs to produce a quick response and action [29]. Danger, sport activity and stressing tasks can produce an increase of the sympathetic system activity. In the HRV spectral distribution, this produces a Power increase in the LF band.
The other side of the autonomous nervous system is the parasympathetic system, which is involved in relaxing activities. Its outcomes are in opposition to the sympathetic system ones. It can produce:

- **Cardiac Frequency Slowing;**
- **Pupils constriction;**
- **Breathing deceleration;**
- **Digestion stimulation;**
- **Energetic recovery.**

It gets triggered when the body is resting, or sleeping. It provides the recovery after some stressing activities. [29] In the HRV spectral distribution, resting activities and parasympathetic involvement produce an increase in the HF band power.
When sympathetic system is triggered, parasympathetic gets reduced , and this provides a balance in the autonomous system activity.

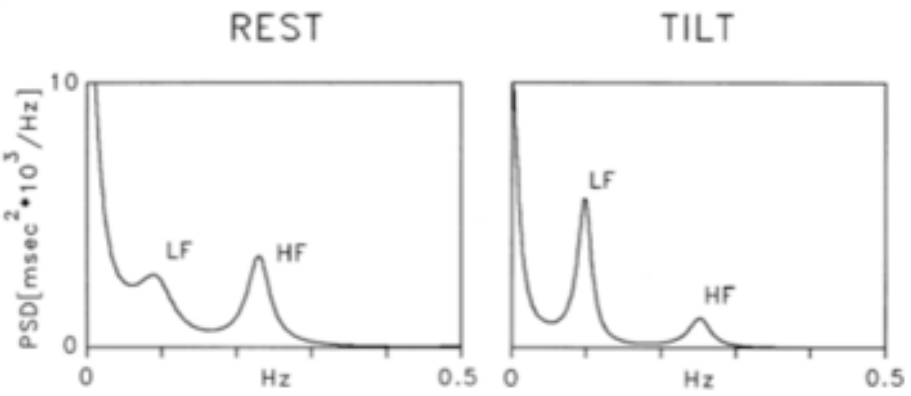


Figure 2.4: HRV Power Spectral Distribution in different activites

Figure 2.4 provides a graphic representation of the Power Spectral Distribution when the subject is performing different activities. As we can see, a stressing tool produces a raise in the LF band, as a consequence of a major activation of the sympathetic system. A drop in HF band is correlated to a minor activation of the parasympathetic system.

2.1.2 HRV medical applications

HRV provides the possibility to track and monitor the Heart Variability, representing a powerful tool in order to evaluate the aging of a subject, or the evolution of a cardiac disease. Many pathologies and clinical conditions can interact with the functionality of the autonomous system, and it can be observed from the frequency bands of the HRV signal. This makes the HRV analysis meaningful and informative for many medical applications.

- **Monitoring Physiological aging**: this is not a pathological conditions if it is considered as an autonomous phenomenon. Aging is natural ,and the improvement of healthy conditions lead to a major life expectation. The prevalence of elderly population increased in the past years, bringing the necessity to handle a major number of aging subjects. Cardiac system is one of the most influenced by Physiological aging, where the normal performances can be altered if compared to a younger subject. HRV can be used to monitor and control the progress, giving

the opportunity to check and eventually provide interventions where it is needed. Doctors rely on it as a powerful monitoring tool that can be applied on subjects progressively aging and possibly altering their physiological performances.

- **Early detection of abnormal cardiovascular pathologies:** the monitoring of the HRV data recorded on a certain subject could provide an early intervention and clinical correction to avoid further worsening of the cardiac condition. A wearable device could possibly record hours of cardiac activity, producing alerts when detecting abnormal trends. This could be applied to continuous modifications, which can get observed only when performing long-term recordings. HRV provides an overview of the cardiac variability on both short and long term, giving the opportunity to have an early detection. Consequently, it would be possible to analyze this anomalous situation as soon as possible, and give more success possibilities to the intervention from the medical equipe.

- **Decision Making Support:** every decision and modification to the clinical treatment of a subject needs to be data-driven, and justified by the analysis of these data. HRV measurements give the opportunity to support doctors in the Decision Making Process, and provide the possibility to create a Database available for EBM (Evidence Based Medicine), where doctors can compare the clinical situation of the patient to those from other past subjects. In particular, HRV recordings could be compared in order to understand which stage is the subject experiencing, when focusing on a certain pathology development. These recordings could represent a powerful tool for the doctors, and their following decisions.

- **Treatment Evaluation:** every patient's treatment needs always some later monitoring, in order to understand the real efficacy of it. The comparison between HRV recording before and after the treatment could help to assess possible weaknesses, or critical situations emerging from it. The clinical overview could be supported also from blood analysis, and other tests to provide a more complete monitoring of the patient. HRV could support the treatment evaluation even for experimental treatments, to understand the influence on the cardiac functionalities, and on the autonomous system.

- **Early detection of Cardiac Events:** wearable devices could work with HRV signal to detect dangerous situations for the patient, and provide an immediate alert to the subject, and possibly even to his doctor. These devices need to record, and provide an early comparison of the signal to others included in the database.

When the difference between safe condition HRV signals and live recorded one is too high, the device could start an alert procedure to notice this situation to the patient, and start the intervention protocol forwarding it to the doctors, and the medical staff possibly taking care of the subject.

- **Remote monitoring:** Telemedicine relies heavily on wearable devices providing recording of the cardiac activity of the patients. When doctors and patients are not in the same place, remote monitoring gives the opportunity to always control and assess any possible issues on the subjects. Internet and Bluetooth technologies can translate the recordings from hardware devices towards the software devices, and then make them available for the medical staff. This gives the doctor the opportunity to check HRV signals from his patients, even without planning a live meeting with them in his studio. The doctor could then detect and notice dangerous situations to the patients, producing corrections to the clinical treatment. Remote monitoring gives the opportunity to accelerate the evaluations, and obtain a faster procedure for the clinical analysis and intervention.

All these applications can be included in the **Personalized Medicine**. This approach took always more relevance in the past few years due to the new introductions of informatics and technological tools related to to Medical data. The aim is to obtain an Evidence Based approach for the treatment of patients, which could follow them in an observation period. HRV data could help to correct the model, which is usually more general, and adapt it to the patient's particularities and clinical overview. This adaptation could lead to a more Patient Oriented, and Personalized approach for the treatment of pathologies. Considering HRV measuring, it is not impossible to hypothesize Clinical Analysis of patients with cardiac diseases, which could create a more realistic and data-based modeling of it, with the opportunity to correct it with the new information provided by data recorded from the HRV devices.

2.1.3 Heart Failure

Heart Failure (HF) is a complex clinical syndrome resulting from any structural or functional impairment of ventricular filling or ejection of blood. It is characterized by the heart's inability to pump sufficient blood to meet the body's metabolic demands [30]. Common causes of heart failure include coronary artery disease, hypertension, cardiomyopathy, and valvular heart disease. Symptoms typically include dyspnea, fatigue, fluid retention, and reduced exercise tolerance [31].

Heart failure is commonly categorized into two types based on the ejection fraction (EF): Heart Failure with Reduced Ejection Fraction (HFrEF), where EF is below

40%, and Heart Failure with Preserved Ejection Fraction (HFpEF), where EF is 50% or higher [32]. Both types are associated with significant morbidity, mortality, and decreased quality of life.
Heart failure is becoming an increasingly prevalent and pressing public health issue worldwide. Its incidence and prevalence are rising due to an aging population, improved survival rates following acute cardiovascular events, and the growing prevalence of risk factors such as hypertension, diabetes, and obesity [33]. As of recent estimates, more than 64 million people globally are living with heart failure, and the numbers are projected to increase substantially in the coming decades [33].

This chronic condition is associated with high rates of hospitalization and rehospitalization, significant healthcare costs, and reduced quality of life. In high-income countries, heart failure represents one of the leading causes of hospitalization in individuals over 65 years old. The economic burden is also considerable, accounting for a substantial portion of national healthcare expenditures [32].
During the years, many drug therapies were tested in order to improve heart contractility, and Ejection Fraction. The work produced in 1986 [18] ,from where the database CHF comes from, is about the testing of oral dosage of a drug named milrinone on heart failure subjects. Milrinone is a phosphodiesterase-3 (PDE3) inhibitor with both inotropic and vasodilatory properties, commonly referred to as an ”inodilator”. By inhibiting the breakdown of cyclic adenosine monophosphate (cAMP), milrinone increases intracellular calcium levels in cardiac myocytes, thereby enhancing myocardial contractility [34]. In parallel, it induces vascular smooth muscle relaxation, resulting in decreased afterload and preload, which can improve cardiac output in patients with acute decompensated heart failure.

Milrinone is typically administered intravenously and is often used in acute settings, such as in intensive care units, particularly in patients with low-output heart failure or cardiogenic shock [35]. Sometimes can also be administered by oral dosage, even though it is less effective. The subjects in CHF database were given milrinone in oral dosage, because in 1986, when the Boston university work [18] was made, many things about milrinone were still unknown, and many test were conduced with different dosage methods, in order to get a general overview of drug therapy performance on heart failure subjects.
In those years,it was one of the most interesting research topics about improving heart contractility. However, it was then discovered that its use is associated with several important risks, including arrhythmias, hypotension, and increased mortality in long-term use. For these reasons, milrinone is now generally reserved for short-term hemodynamic support or as a bridge to more definitive therapies such as cardiac transplantation

or mechanical circulatory support.

In recent years, other pharmacological treatments that offer better long-term outcomes and fewer adverse effects have been adopted in heart failure treatment. Guideline-directed medical therapy (GDMT) for chronic heart failure with reduced ejection fraction (HFrEF) now primarily includes beta-blockers, angiotensin receptor-neprilysin inhibitors (ARNIs), mineralocorticoid receptor antagonists (MRAs), and sodium-glucose cotransporter 2 inhibitors (SGLT2is) [32]. These agents have shown substantial improvements in survival, reduction in hospitalizations, and symptom relief in large randomized clinical trials. As a result, while milrinone may still have a role in acute management, it is no longer considered a cornerstone of chronic heart failure therapy.

The growing impact of heart failure underscores the urgent need for improved prevention, early detection, and effective long-term management strategies. In this context, digital health tools and biomarkers—such as Heart Rate Variability (HRV)—are gaining attention for their potential to support non-invasive monitoring and individualized care approaches.
To evaluate the severity and functional impact of heart failure, the New York Heart Association (NYHA) classification is widely used. This classification stratifies patients into four classes according to the degree of physical activity limitation and symptom severity:

- **Class I:** No limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, or dyspnea.
- **Class II:** Slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, or dyspnea.
- **Class III:** Marked limitation of physical activity. Comfortable at rest. Less than ordinary activity causes fatigue, palpitation, or dyspnea.
- **Class IV:** Unable to carry on any physical activity without discomfort. Symptoms of heart failure are present even at rest.

The NYHA classification provides a simple yet effective method to assess functional status and guide treatment decisions. However, it is subjective and can vary between clinicians, making it important to use in conjunction with objective diagnostic tools such as echocardiography and biomarkers (e.g., NT-proBNP) [36].
CHF database [2] involves only subjects with heart failure evaluated in Class III or IV, of different ages.

2.1.4 Heart Failure and HRV

Congestive Heart Failure (CHF) is associated with profound alterations in the autonomic regulation of the cardiovascular system, which are clearly reflected in Heart Rate Variability (HRV). Numerous studies have shown that patients affected by CHF exhibit a significant reduction in HRV compared to healthy individuals, indicating an impaired autonomic control of cardiac activity [37, 38].

The reduction in HRV observed in CHF patients is generally attributed to a marked worsening of the autonomous system regulation. This altered autonomic balance leads to a diminished dynamic adaptability towards external stimulation and homeostasis. The general Power Spectral Distribution get reduced in power, leading to an altered distribution in the HRV spectral bands. Both LF an HF bands get reduced, as a consequence of the alterations in the sympathetic and parasympathetic systems. Heartbeat dynamics become consequently more regular and less complex, reflecting a reduced responsiveness of the cardiac control mechanisms [39, 40]. Beyond the reduction in overall variability, CHF has also been associated with a loss of nonlinear and complex dynamics in HRV signals. Several studies have reported that pathological conditions such as heart failure are accompanied by a breakdown in the multiscale regulatory mechanisms that normally govern cardiac rhythm, leading to a decrease in signal complexity [2, 1]. This loss of complexity may reflect the reduced ability of the cardiovascular system to respond to internal and external alterations. Reduced HRV in CHF patients has been associated with disease severity and with an increased risk of adverse cardiovascular outcomes, including arrhythmic events and mortality [38]. For this reason, HRV analysis represents a valuable non-invasive tool for assessing autonomic dysfunction in heart failure and for supporting risk stratification and clinical monitoring of affected patients.

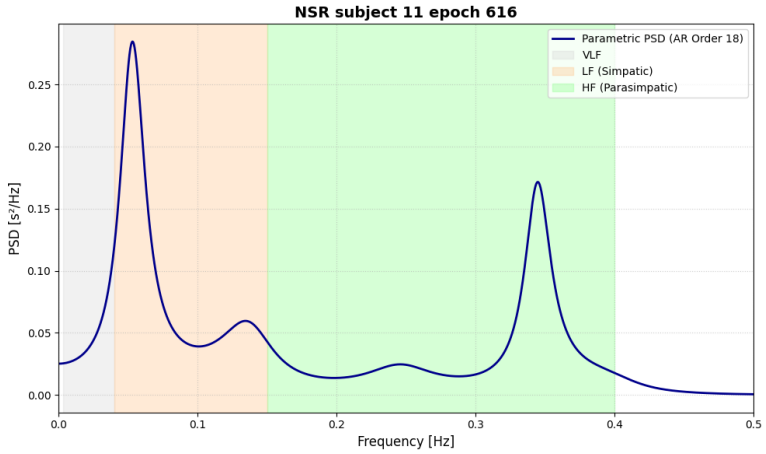


Figure 2.5: Power Spectral Density on a epoch of signal from NSR database

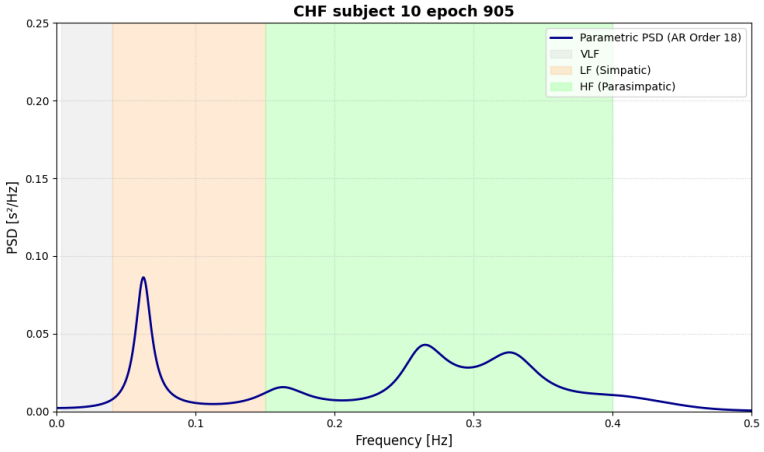


Figure 2.6: Power Spectral Density on an epoch of signal from CHF database

In figures 2.6 and 2.5 the Power Spectral Densities (PSDs) of HRV signals from both the CHF and NSR databases are shown. Subjects with congestive heart failure (CHF) exhibit a very different spectral profile, characterized by a substantial reduction in the Power across the whole Spectral Distribution.
These differences are consistent with the known impact of cardiac dysfunction on autonomic regulation. Advanced heart failure (NYHA Class III-IV), as in CHF database, is associated with severely reduced parasympathetic activity and overall reduced HRV, which are clearly reflected in the spectral domain. Both LF and HF bands get power reductions, and the LF/HF rate can be altered in different ways. Entropy measures were then applied on the resulting recordings.

2.1.5 data description

- **CHF.** This database was involving 29 recordings, from 15 subjects in the original group. The individual recordings from CHF database [2] are each about 20 hours in duration, and contains two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ±10 millivolts. The original analog recordings were made at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center) using ambulatory ECG recorders with a typical recording bandwidth of approximately 0.1 Hz to 40 Hz.
  HRV data were already available. These recordings had different lengths, but were about 100 000 samples each. Each sample contains the time distance between R-peaks.

- **NSR.** This database includes beat annotation files for 54 long-term ECG record-

ings [2] of subjects in normal sinus rhythm (30 men, aged 28.5 to 76, and 24 women, aged 58 to 73). The ECG original data (not available) were digitized at 128 samples per second, and the beat annotations were obtained by automated analysis with manual review and correction.
HRV data from beat annotations were available, as for the CHF ones. Also on them, the recordings were of different lengths, but about 100000 samples each.

2.2 FANTASIA

This database , available on Physionet, involves twenty young (21 - 34 years old) and twenty elderly (68 - 85 years old) rigorously-screened healthy subjects. [15] Each subgroup of subjects includes equal numbers of men and women. The group underwent 120 minutes of continuous supine resting while continuous electrocardiographic (ECG), and respiration signals were collected; in half of each group, the recordings also include an uncalibrated continuous non-invasive blood pressure signal. All subjects remained in a resting state in sinus rhythm while watching the movie Fantasia (Disney, 1940) to help maintain wakefulness.
As for the previous database, it was decided to work on HRV signals, instead than on ECG recordings. On this database, all of the subjects are in healthy condition and the only difference between subjects in different subgroups is the age, which could affect the HRV signal.

2.2.1 Aging

Aging is a progressive biological process characterized by gradual structural and functional changes occurring across multiple physiological systems. These changes arise from the cumulative effects of molecular damage, cellular senescence, and long-term environmental influences, leading to a progressive decline in the efficiency of regulatory mechanisms within the body [41]. One of the systems most affected by aging is the cardiovascular system. Structural and functional alterations include increased arterial stiffness, reduced vascular elasticity, and modifications in cardiac electrophysiological properties. In particular, the pacemaking cells of the sinoatrial node show a reduced responsiveness to autonomic regulation and external physiological stimuli, contributing to a less flexible cardiovascular response [42, 43]. These changes are associated with a general reduction in physiological adaptability and a diminished capacity to maintain homeostasis in response to internal or external stimulations. As a consequence, older individuals may experience a decreased ability to adapt to environmental changes, physical activity, or pathological conditions [44]. Furthermore, aging is strongly associated with an increased prevalence of chronic diseases and functional decline. Conditions such as cardiovascular disease, metabolic disorders, and neurodegenerative pathologies become more common with advancing age, often contributing to reduced independence and overall health status in the elderly population [45].
This problem became more relevant in the modern days, when the life expectancy grew up massively due to the improvements in clinical treatments, and general improved

life conditions. The number of people composing the elderly population is now way higher, and this leads to a major number of clinical conditions, and elderly subjects to monitor. The Global Sanity needs to face a growing number of older subjects who require medical treatment, assistance and monitoring from medical staff and the whole healthcare industry is now exposed to a major number of requirements from both the population and hospitals to manage keeping an high level of quality in the clinical support.

2.2.2 Aging and HRV

Heart Rate Variability (HRV) reflects the autonomic regulation of the cardiovascular system and is known to be influenced by several physiological and pathological factors. One of the most consistently observed influences on HRV is aging. Numerous studies have demonstrated that advancing age is associated with a reduction in HRV, reflecting a decline in autonomic flexibility and adaptability [46, 15].

Age-related decreases in HRV are generally attributed to a reduction both in parasympathetic (vagal) tone and sympathetic systems. This shift in autonomic balance is associated with structural and functional changes in the cardiovascular and nervous systems, including decreased baroreflex sensitivity, reduced responsiveness of sinoatrial node pacemaker cells, and increased arterial stiffness [47, 48]. This leads to a less responsive regulation of the cardiac activity, producing an HRV signal more stable. The Power Spectral Distribution of the HRV signal gets generally reduced across the whole Spectrum, especially in the HF band. The parasympathetic system is the one that gets more impaired from Aging, leading to a predominance of the LF band, which is also reduced, if compared to that obtained from an healthy young subject. The ratio LF/HF can often get increased because of it, but this is always related to the nature of the subject himself, and his clinical overview.

Importantly, aging does not just reduce the overall amplitude of HRV, but also alters its temporal structure. In particular, fractal and nonlinear analyses have revealed that the complex, self-similar patterns typically found in young, healthy individuals become more regular and predictable with age. Iyengar and his group [15] demonstrated that elderly subjects exhibit a loss of fractal scaling in their interbeat interval dynamics, and they demonstrated this by working on the 'Fantasia' Database.[15] The age-related loss of fractal organization in heartbeat dynamics may reflect the degradation of integrated physiological regulatory systems and may impair an individual's ability to adapt to stress.

These changes have important clinical implications, as lower HRV in older adults has been associated with increased risk of cardiovascular events, frailty, and all-cause

mortality [49, 50]. Therefore, HRV analysis may serve as a valuable tool to assess age-related decline in autonomic function and to identify individuals at higher risk of adverse outcomes.

The two group of subjects in this work were both resting, and both watching the same movie (fantasia) , so basically the external stimulations that their bodies were receiving were the same. The adaptation ability that they were producing, in answer to them, was only depending on their individual autonomic system.

Age is expected, from previous studies, to reduce responsivity in elderly,and to possibly create a less dynamic and more stable signal. In Entropy-based analysis, this could lead to a very different value of entropy, being the elderly's HRV way less adaptive. This could be reflected on a lower value of Entropy on elderly results, if compared with Entropy from young subgroup.

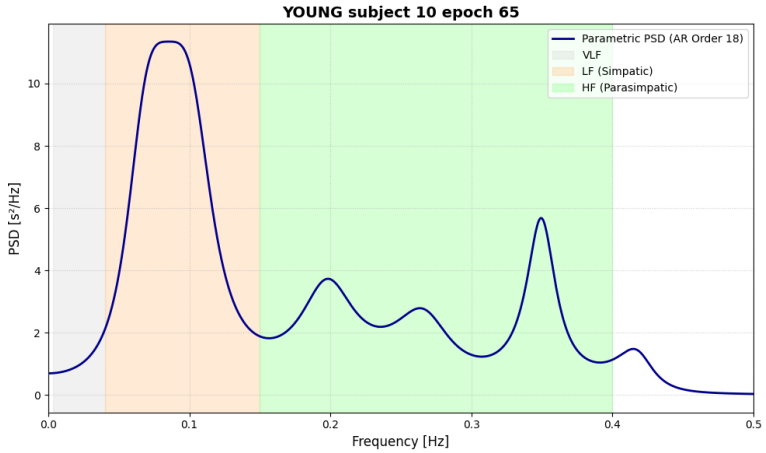


Figure 2.7: Power Spectral Density on HRV from a young subject

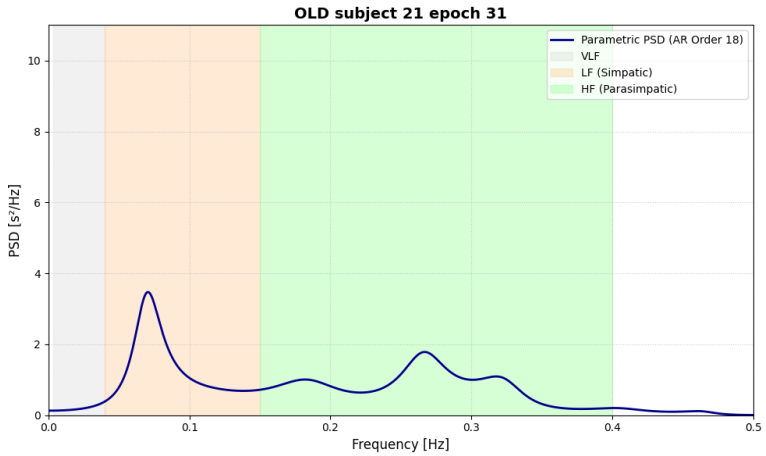


Figure 2.8: Power Spectral Density on HRV from an elderly subject

In figures 2.8 and 2.7 the Power Spectral Densities (PSDs) of HRV signals from

both young and elderly subjects are presented. The overall power distribution across the spectrum reveals noticeable differences between the two populations. Elderly subjects show a general Power Spectral reduction across all different frequencies. Aging alters especially the regulation and the control of the Parasympathetic System, which has reflection in the HF band. The plot confirms this trend, showing a relevant reduction in the Power associated to the High Frequencies, significant when compared with the one from on an epoch of HRV signal recorded on a young subject.
These spectral alterations are consistent with age-related changes in autonomic regulation, which are known to affect both the magnitude and distribution of power within the LF and HF bands of the HRV spectrum.

2.2.3 data description

The continuous ECG, respiration, and (where available) blood pressure signals were digitized at 250 Hz.[15] Each heartbeat was annotated using an automated arrhythmia detection algorithm, and each beat annotation was verified by visual inspection. This work was already made by the group that released the database.
The HRV data were already available, and recordings had different length. The medium length of the recording is about 7000 samples each.

2.3 EEG SIENA

The database , available on Physionet, consists of EEG recordings of 14 patients acquired at the Unit of Neurology and Neurophysiology of the University of Siena [16]. Subjects include 9 males (ages 25-71) and 5 females (ages 20-58). Patients were asked to stay in the bed as much as possible, either asleep or awake.
The data were collected by the Unit of Neurology and Neurophysiology at University of Siena, Italy, during a regional research project, called PANACEE , aiming at the development of noninvasive patient-specific monitoring/control low-cost devices for the prediction of epileptic seizures. The 14 patients did not show all the same nature of epileptic seizure:

- 12 of them showed epileptic seizures which can be classied as IAS (Focal onset Impaired Awareness);
- Patient PN10 showed FBTC seizures (Focal to Bilateral Tonic-Clonic);
- Patient PN14 showed WIAS seizures (Focal Onset Without Impaired Awareness).

The diagnosis of epilepsy and the classification of seizures according to the criteria of the International League Against Epilepsy were performed by an expert clinician after a careful review of the clinical and electrophysiological data of each patient.

2.3.1 EEG signal

Electroencephalography (EEG) is a non-invasive neurophysiological technique that measures the electrical activity of the brain through electrodes placed on the scalp. The recorded signal primarily reflects the postsynaptic potentials of thousands to millions of neurons, especially the pyramidal cells located in the cerebral cortex. Due to volume conduction, the EEG captures a summation of electrical fields generated by neuronal populations. This results in a composite and interferent signal, rather than isolated neurons activity[51].

The EEG signal is typically acquired using a differential configuration, meaning that the potential recorded at a given electrode is measured with respect to a reference electrode. The choice of reference (e.g., linked ears, average reference, mastoid) can influence the appearance and interpretation of the EEG waveform [52]. This differential configuration enables the identification of spatial patterns of brain activity and helps

reduce the effects of common-mode noise and external interferences.

The brain electrical activity is a mirror of the activities of the subject, and his actions, also involving emotions and sensorial perceptions. Involving so many different inputs, the signal is very different from one moment to another, making it a chaotic signal, and without a morphological pattern on it.
The EEG spectrum is traditionally divided into specific frequency bands, each reflecting different types of brain activity. These bands range from slow oscillations such as delta waves to faster rhythms like gamma waves.

- **Delta (0.5–4 Hz):** Associated with deep sleep stages (particularly NREM sleep) and unconscious states. Predominant in infants and during slow-wave sleep in adults.
- **Theta (4–8 Hz):** Linked to drowsiness, light sleep, meditation, and creative or emotional processes. Often observed during transition from wakefulness to sleep.
- **Alpha (8–13 Hz):** Most prominent when a subject is awake but relaxed, especially with closed eyes. Commonly seen in the occipital region.
- **Beta (13–30 Hz):** Related to active thinking, attention, problem-solving, and focused mental activity. High beta levels can also be associated with anxiety or tension.
- **Gamma (30–100 Hz):** Associated with high-level cognitive functions such as memory, perception, and consciousness. Gamma rhythms are also linked to information binding across brain regions.

Frequency Band	**Frequency Range (Hz)**
Delta	0.5 – 4
Theta	4 – 8
Alpha	8 – 13
Beta	13 – 30
Gamma	30 – 100

Table 2.2: Main frequency bands in EEG signals.

In figure 2.9 the Power Spectral Density (PSD) from an epoch the EEG signal signal recorded from different channels is shown .
The PSDs plots reveal slightly changes across each phase, with higher spectral power concentrated in the lower frequency bands. Most of the signal energy is localized within Delta (0.5-4 Hz) , Theta (4-8 Hz) , and Alpha (8-13 Hz) band frequencies.

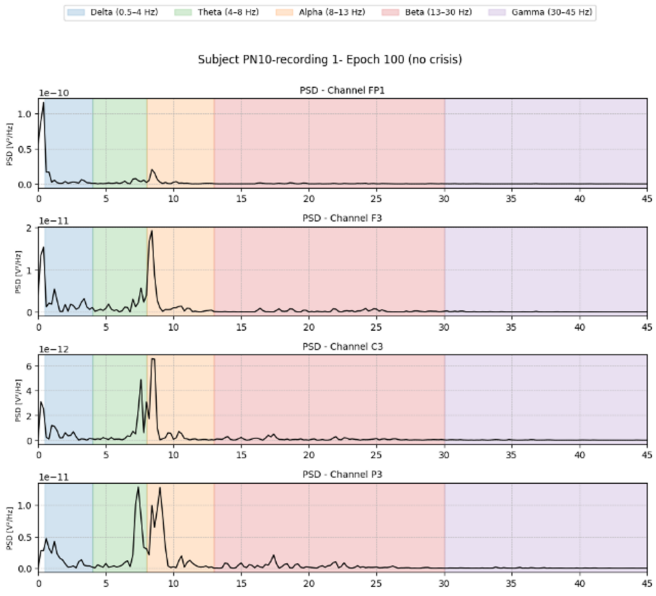


Figure 2.9: Power Spectral Density on EEG segment

The human cerebral cortex is functionally specialized, with different regions supporting distinct cognitive and motor functions. For instance, the occipital lobes are primarily involved in visual processing, the parietal regions integrate sensory information, the temporal lobes are essential for auditory processing and memory, and the frontal lobes govern executive functions, motor control, and higher-order cognition [53]. In figure 2.10 are shown the main brain areas of the cerebral cortex.

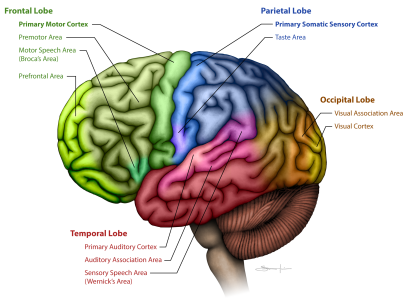


Figure 2.10: Disposal of the areas in the Cerebral cortex

Accordingly, EEG activity recorded from these regions reflects these functional spe-

cializations; for example, beta rhythms [14-30 Hz] over the sensorimotor cortex are associated with movement planning and execution, while alpha rhythms [7-13 Hz] over occipital regions are linked to visual rest and attentional states. [54] The electrical activity in these regions is controlateral: the left emisphere controls the right side of the body, for both sensorial and motor activity. In the same way, the right emisphere controls the left side of the body.

The international 10–20 system, introduced by Jasper in 1958, provides a standardized method for electrode placement on the scalp, ensuring reproducibility and comparability in studies [55]. The system divides the scalp into regions corresponding to the underlying cortical areas, with electrodes labeled according to their location: frontal (F), central (C), temporal (T), parietal (P), and occipital (O). Odd numbers are used for the left hemisphere, even numbers for the right, and 'Z' denotes midline electrodes [56]. The nomenclature "10–20" refers to the fact that electrodes are placed at intervals of 10% or 20% of the distance between specific anatomical landmarks: the nasion, the inion, and the pre auricular points [57]. This method ensures that electrode positions are proportional to head size, making it applicable across different individuals.

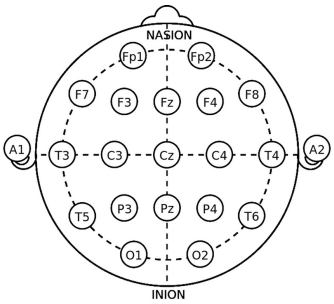


Figure 2.11: Electrodes disposal in 10-20 configuration

The number of electrodes involved can change, but tipically are used about 19 electrodes, plus the reference electrodes. For higher spatial resolution, the 10–20 system has been extended to the 10–10 and 10–5 systems, increasing the number of electrodes to 74 and over 300, respectively [58]. These extended systems are particularly useful in research settings requiring detailed cortical mapping and are supported by the International Federation of Clinical Neurophysiology (IFCN) [56].
Despite its widespread use, the 10–20 system has limitations. Variability in skull anatomy can lead to deviations between electrode positions and the underlying cortical regions, which may affect the interpretation of EEG data [59]. Therefore, while the 10–20 system provides a practical and standardized approach to electrode placement,

considerations of individual anatomical differences are important in both clinical and research applications.

2.3.2 Epilepsy

Epilepsy is a chronic neurological disorder characterized by recurrent, unprovoked seizures resulting from abnormal electrical discharges in the brain. Seizures are broadly classified into two main categories: focal and generalized. Focal seizures originate within a specific area of one hemisphere of the brain and can remain localized or spread to other regions, potentially evolving into generalized seizures. Generalized seizures involve both hemispheres from the onset and include types such as tonic-clonic, absence, and myoclonic seizures [60, 61].
Many type of seizures can happen, each of them involving a different electrical activity, and specific features. Among these,

- *Focal to Bilateral Tonic-Clonic* (FBTC) seizures are focal in onset but spread to involve both hemispheres, leading to generalized convulsive activity [62].
- *Impaired Awareness Seizures* (IAS) are focal onset seizures in which consciousness is altered or lost during the event, and the subject loses control of himself during the event.
- *Without Impaired Awareness Seizures* (WIAS) are focal onset seizures in which awareness remains preserved throughout the episode [63]. Can be considered similar to an IAS event from an electrical point of view, but the preservation of awareness keeps a signicant difference between them.

Accurate differentiation between these seizure types is essential for correct diagnosis, treatment planning, and prognosis [64]. Diagnosis phase is critical, because each zone of the brain can be involved in different ways from the seizure, and damages can be very different from the type of electrical activity of the epilepsy event. And if the diagnosis can help the treatment after seizure event, the unpredictability of epileptic seizures poses significant challenges for patients and clinicians. Despite advances in medical imaging and monitoring techniques, accurately predicting the occurrence of seizures remains elusive. This unpredictability is partly due to the complex and dynamic nature of brain activity, which varies among individuals and over time. Factors such as stress, sleep deprivation, and hormonal changes can influence seizure thresholds, further complicating prediction efforts [60].
Electroencephalography (EEG) plays a pivotal role in the diagnosis and management of epilepsy. It records the brain's electrical activity, allowing clinicians to identify

abnormal patterns associated with seizures.
The analysis of EEG patterns provides valuable insights into the functional organization of the brain and the localization of epileptogenic zones. When the epileptic seizure starts, the pattern of the EEG changes, and becomes synchronized across all the EEG channels. Usually, the amplitude of the signal increases, because of the improved synchronization of the eletrical activity in the brain. This leads to an higher amplitude across all the signals. In figure 2.12 it is possible to see the recordings of an EEG signal during the start of an epileptic seizure. It is possible also to see the zone where the seizure is placed: it is a focal seizure event, being it very limited to a single zone in the cerebral cortex.

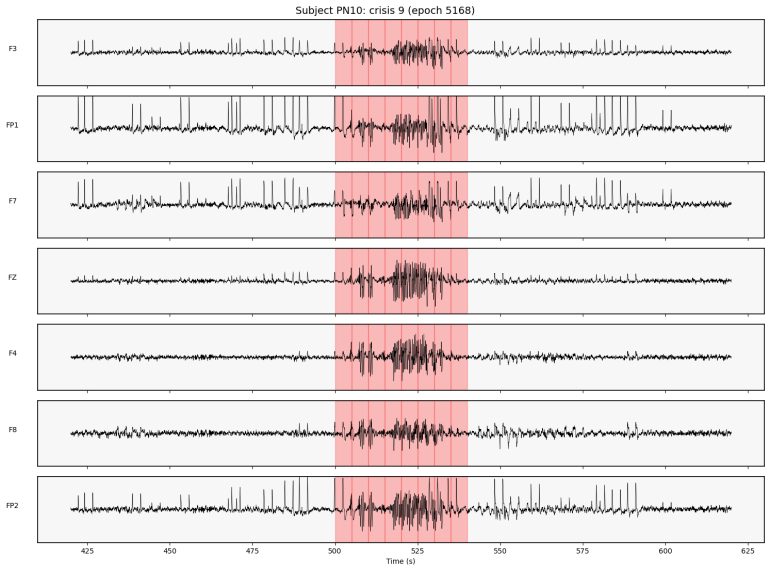


Figure 2.12: Epilepsy seizure in EEG recording

For instance, focal seizures often exhibit localized abnormalities in EEG recordings, which can aid in pinpointing the origin of the seizures. Conversely, generalized seizures typically present with widespread, synchronous discharges across both hemispheres [61].
Despite the utility of EEG in identifying seizure-related activity, predicting the exact timing of seizures remains a significant challenge. Research efforts continue to explore advanced signal processing techniques and machine learning algorithms to enhance seizure prediction capabilities. However, as of now, reliable and consistent prediction methods are still under development [60]. An EEG signal Analysis can show a drop in the complexity of the signal in the Ictal phase, which can easily be highlighted with a plot of the Entropy values as the one in the Figure 2.13.

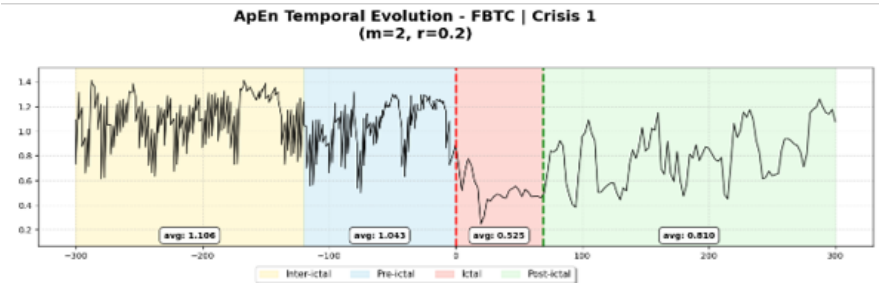


Figure 2.13: Entropy values of the EEG signal in an epileptic seizure

The challenge remains to understand the dynamics behind the onset. If a method could possibly catch the moment when the interictal normal EEG signal switches to a preictal pattern, it would be possible to provide an alarm, and notify the onset to the patient. This would be crucial to the subject and could give him the opportunity to take anti-epileptic drugs, and to prevent the onset of epileptic symptoms.
The positive outcomes of a similar result could lead to the development of Brain Computer Interfaces to monitor, notify and handle epileptic seizures by analyzing in real time the EEG signal. With actual technologies, this result is yet to be achieved, and this goal still brings many issues with it, correlated also to the need of a real-time analysis, and consequently a computational efficient algorithm.

2.3.3 EEG applications for Epilepsy

EEG signal is the main tool in order to record electrical activity from the brain . It can be applied for many different purposes, correlated to a various set of diseases. Epilepsy events can be monitored with an EEG analysis, and many applications can be obtained from the involvement of EEG recording systems.

- **Seizure detection and classification:** as exposed in the previous paragraph, many types of epilepsy exist, and each one is related to a certain clinical treatment. It is very important to assess as soon as possible the nature of the epileptic event, in order to understand possible damages and consequences for the subject. A symptoms analysis is also needed, in order to get a more complete overview of his clinical situation. Not all epilepsy events lead to the same outcomes on the patient, and EEG recordings need to be included in the assistance provided to the subject, to provide more data and inputs to the doctor/neurologist following the subject.
  The prediction of epileptic events remains very challenging, and the solution is often to induce them by external stimulations, to assess the responsiveness of the patient. This facilitates the onset of the electrical synchronization, and

makes the seizure event recordable during the observation window. When the epileptic event is happening, EEG represents a fundamental tool to understand the electrical activity and synchronization that the subject is experiencing. The classification of the epilepsy event is the final outcome of this process.

- **Wearable monitoring:** EEG electrodes could be placed on the head of the subject. This guarantees a continuous recording of the electrical activity of the brain, for long term monitoring. It is not possible to predict epilepsy events, and this is the triggering motivation for a wearable EEG device. This approach gives the opportunity to detect any alterations in the EEG signal in the vicinity of the seizure: localization and time analysis can contribute to get a better understanding of the dynamics of the electrical synchronization.
  Data from wearable devices can be used as Telemedicine tools, translating these recordings to the medical staff related to the subject. Data can be shared by Bluetooth or Internet, giving the opportunity to consult epochs of a long term recording close to the seizure events.

- **Real Time Alert:** the previous described wearable EEG devices can include a Real Time Alert protocol, which can notify both the patient and the medical staff when seizure events are taking place. This gives the opportunity to the subjects to take their own anti-epileptic drugs and face the crysis in the correct way. The Medical staff can try to contact the patient, in order to monitor the development of the event, and to track any symptoms on him.
  The goal would be to obtain a wearable EEG device which records the electrical activity of the brain, and when anomalous preictal patterns are detected, triggers the Alerting protocol to the subject even before the epilepsy onset. This requires a training of a classifier based on preictal epochs, with features capable of recognizing preictal conditions. Nowadays, this functioning is still yet to be fullfilled, and the work on this Thesis focuses on the possibility of getting more informative inputs from Entropy-based features. This could help to develop a wearable device providing Real Time Alerts when Preictal patterns are detected: the positive outcomes would be massive for epileptic patients.

- **Neurostimulation:** EEG recordings can also be integrated with neurostimulation systems designed to reduce the occurrence or severity of epileptic seizures. In these approaches, EEG signals are continuously monitored in order to detect abnormal patterns associated with seizure onset. When such patterns are identified, a stimulation device can deliver electrical impulses to specific brain regions or to peripheral nerves, aiming to disrupt the pathological synchronization responsible

for the seizure. Examples of such approaches include responsive neurostimulation and vagus nerve stimulation. These systems represent an important step toward closed-loop therapeutic strategies, where neural activity is monitored and modulated in real time.
The delivery of stimulation impulses is always related to the nature of the epileptic disease, and especially when the onset is Focal the placement of the electrodes can be planned by observing the zone of the electrical triggering synchronization event. This makes the EEG recording crucial in order to provide a correct Neurostimulation.

- **Treatment monitoring:** EEG analysis can also support the evaluation of therapeutic interventions in epileptic patients. Long-term EEG recordings allow clinicians to monitor how seizure frequency, duration, and electrical patterns evolve over time in response to pharmacological treatments or other therapeutic approaches. This information can help neurologists assess the effectiveness of anti-epileptic drugs and adjust treatment plans accordingly. In this context, EEG provides an objective tool to support clinical decision making and to track disease progression.
  This application can also be considered as part of Telemedicine, where doctors and medical staff can remotely suggest alterations of the treatment plan to the patient, by consulting EEG recordings, frequency and intensity of seizure events. The comunication between the patient and the medical staff is central in order to provide a Personalized and Evidence Based Approach suited for the individualities of the subject. Treatment needs to be calibrated, and EEG provides support to the decision making and the progression of the Medical Plan.

- **Brain Computer Interface:** EEG-based Brain Computer Interfaces represent systems capable of translating brain activity into commands that can control external devices or trigger specific responses. In the context of epilepsy, BCI approaches may be used to detect seizure-related patterns in real time and automatically activate warning systems or therapeutic interventions. For example, a BCI could identify preictal EEG patterns and trigger an alert or activate a neurostimulation protocol. These systems rely on signal processing and machine learning techniques to extract meaningful features from EEG recordings and represent a promising research direction for improving seizure management.
  Brain Computer Interfaces summarizes some of the previous EEG applications in a single tool, which provides an autonomous and real time support to the subject. The development of these applications needs to lower the Computational

Complexity of the data processing, and to reduce the number of False Negatives. This remains the most challenging and clinically testing application, which could really improve the quality of life of an epileptic subject. The dimension of the possible positive outcome is proportional to the complexity in the development of these applications, and this task is yet to be solved by modern technology.

- **Personalized epilepsy monitoring:** Epileptic activity can vary significantly across different patients, both in terms of EEG patterns and seizure manifestations. For this reason, recent research has focused on the development of personalized monitoring systems that adapt to the specific characteristics of each patient. Machine learning models can be trained on individual EEG recordings in order to learn subject-specific patterns associated with seizures. This personalized approach may improve the accuracy of detection and prediction systems and represents an important step toward patient-specific epilepsy management.

Among these applications, automated seizure detection and early identification of abnormal EEG patterns represent particularly active research areas, where advanced signal processing techniques and entropy-based features may provide additional informative descriptors of brain dynamics.

2.3.4 data description

The database consists of EEG recordings of 14 patients monitored with a Video-EEG with a sampling rate of 512 Hz, with electrodes arranged on the basis of the international 10-20 System. All the recordings contain 1 or 2 EKG signals, too. The data were acquired employing EB Neuro and Natus Quantum LTM amplifiers, and reusable silver/gold cup electrodes [16]. Each subject could have more than one recording, to a maximum of 5 files per each subject. The files contain signals recorded on the same or different days and the seizure events are chronologically ordered. The table 2.3 contains all the information on the subjects.

In total, the database was containing 47 seizures on about 128 recording hours. Not all the seizures are obtained from the same recording on each subject, because as said, each subject could have more than one recording. The order of the channels is not exactly the same in each file.
Each file was containing also different data, like HR (Heart Rate) SpO2 (Periferic Oxigen saturation) and PLET (Pletismographic evaluation). These channels usually contain a stable value set as zero in the files involved in this work, which leads to a

ID	Age	Gender	Seizure	Localization	Lateralization	EEG ch.	Seizures	Rec. time (min)
PN00	55	Male	IAS	T	R	29	5	198
PN01	46	Male	IAS	T	L	29	2	809
PN03	54	Male	IAS	T	R	29	2	752
PN05	51	Female	IAS	T	L	29	3	359
PN06	36	Male	IAS	T	L	29	5	722
PN07	20	Female	IAS	T	L	29	1	523
PN09	27	Female	IAS	T	L	29	3	410
PN10	25	Male	FBTC	F	Bilateral	20	10	1002
PN11	58	Female	IAS	T	R	29	1	145
PN12	71	Male	IAS	T	L	29	4	246
PN13	34	Female	IAS	T	L	29	3	519
PN14	49	Male	WIAS	T	L	29	4	1408
PN16	41	Female	IAS	T	L	29	2	303
PN17	42	Male	IAS	T	R	29	2	308

Table 2.3: Summary of Patient Seizure Data

corrispondent Entropy value of zero.

It was also possible to find other channels, with other data (probably trigger events, or some other evaluations not clarified in the database). For example, the file corresponding to PN17, has 49 channels: 29 of them are EEG channels, 2 of them are EKG, then there is HR, SpO2,PLET. The rest of the channels are named :1,2,61 ,62 ,63 ,64 ,3 ,23 ,33, 31 ,A ,B ,C ,D ,MK. It is not clear what was the meaning of these channels.

3. Methods

3.1 Signal processing

Signal processing is a crucial step to proceed with a correct Signal Analysis. Each signal involved in this work was processed in order to remove any possible noise, artifact, low and high frequency trends. Especially in Entropic evaluations, a signal not correctly filtered can create bias on values produced.

3.1.1 Epoch Segmentation

It was introduced on each database in order to evaluate the evolution of the Complexity of the Signal during the recording. This approach can provide many positive outcomes:

- **Stationarity of the signal:** on a epoch of a certain length it is more probable to get a condition of stationarity. On a long-term recording this hypothesis is difficult to keep, especially if signals are extremely related to variations or have very complex hidden dynamics.

- **Spectral Resolution**: the choice of the epoch length is always related to the desired spectral resolution. This parameter provides the smaller difference in the frequency axis that can be observed.From the formula, it can be obtained as

$$\Delta F = \left\lfloor \frac{1}{Ts} \right\rfloor \tag{3.1}$$

  where **Ts** is the time length of the epoch. Having a very long epoch could provide the possibility of having small values of Spectral Resolution, and a more focused analysis on Spectral evolution of the signal.
  The dimension of the epoch can't be chosen as way too large, because the stationarity condition could get lost. The epoch length needs to be chosen in order to keep both stationarity and an acceptable value of Spectral Resolution.

- **Complexity evolution:** working with a signal segmented in epochs consent

to understand the behavior and the trending of complexity inside the temporal window of the recording. This could be helpful to relate possible variations in Entropy to the onset of clinical conditions (epileptic seizures for example).

- **Machine Learning approach:** having many epochs of a single signal can represent an improvement to the deployment of a meaningful Machine Learning algorithm. Classifiers needs a big amount of data, and epochs segmentation can increase the number of inputs that could be provided.

- **Computational Complexity:** Entropy will now be evaluated on a shorter epoch of signal. With this approach, it is possible to get faster analysis of the complexity of a signal, dividing it into epochs. The number of epochs can be huge, but the general evaluation time will get reduced. The outcome is a general reduction of the Computational Complexity of the Entropy-based approach, which gets comparable and competitive with one involving time/Frequency features.

The number N of the epochs can be obtained as

$$N = \left\lfloor \frac{T - L}{S} \right\rfloor + 1 \tag{3.2}$$

where N is the length of the recording. L is the length of the epoch, and S is the stride, which is the temporal length of the overlapped epoch. For example, in the EEG database the signal was divided in segments of 2 minutes, leading to a T=120 s, with an epoch length of 5 s (L). With an overlap=50 % , leads to S=2.5 s, and a total N= 47 number of epochs.
Every database has its own signals, and every signals its specific features. It is not possible to assess which is the correct length of an epoch, without analyzing the nature of it. Choosing the wrong dimension of the epochs could create an evaluation not valuable, and a big bias in data.
Overlap between consecutive epochs is another factor that needs to be chosen wisely. In this work, on all three databases was chosen to apply a 50 % overlap. It is not impossible to use a lower percentage of overlap, while instead major values are not suggested, due to the alterations that they could bring. Literature suggests 50 % as the maximum value.

3.1.2 CHF database

The processing applied on HRV recordings from both CHF and NSR databases included the following steps:

1. **Outlier removal:** all RR intervals outside the range [0.2-1.5] seconds were removed. It was decided to apply this filtering on the values in order to discard values unreasonable long or short for an RR interval. Only a few values were removed, as the signals had likely already been preprocessed by the database creators.

2. **Epochs segmentation:** the whole recording was divided in epochs of 200 samples. It is mandatory to talk of samples and not of time-length, being each sample a value of the RR interval. The epoch contains 200 values of RR intervals, and the segmentation obtained with a 50% overlap.

3. **Mean Value removal:** From each recording, mean value was subtracted. This step is useful to get a clearer overview on the Power Spectral Density (PSD) of signals, and to evaluate whether further filtering is needed. It removes low frequency trends.

3.1.3 FANTASIA database

Processing applied on FANTASIA HRV recordings was the same as the one applied in CHF and NSR database.

3.1.4 EEG SIENA database

The work on the EEG database followed a different approach. Time indications of the onset of epileptic seizures gave the opportunity to extract different segments of the EEG recording:

- **ictal segment**;
- **preictal segment**: a segment of length comparable to the time length of the ictal event. Being the longest ictal event of less than 2 minutes, this was the dimension of the preictal segment involved.
- **postictal segment**: a segment of 2 minutes following the ictal phase.
- **interictal segment**: this nomenclature identifies a segment external to the preictal and postictal condition. It is possible to identify it as a segment of EEG signal where there is no relation to the seizure event. Thi segment was also 2 minutes long.

From the whole database, it was chosen to consider only the subject having FBTC (Focal to Bilateral Tonic-Clonic) seizures. This subject had 10 different epileptic seizure across 6 different recordings, leading to 10 ictal, preictal and postical segments, and 6 interictal segments (one for each recording). From those 6 recordings, the processing was involving the following steps:

1. **Segments Extraction**: the indications in the files were used to extract segments from the previous described phases of the seizure events. Segments were 2 minutes long each, in order to obtain comparable length segments.

2. **Resampling**. In the original database, data were obtained with a 512 Hz Video-EEG [16]. EEG bandwidth is usually considered to a maximum of 100 Hz, but in general the most informative content is kept under 50 Hz. It was decided to resample EEG data at a new sampling frequency of 128 Hz, in order to have a sampling frequency which can be considered as at least the double of the maximum informative frequency (50 Hz). This approach satisfy the requirements from Nyquist's criterion.

3. **Bandpass Filtering**. It was decided to apply a bandpass filtering, with a passband of 0.5-40 Hz.It was chosen to keep bands until the Beta ryhthm and to exclude major frequencies, which could include noise related motion artifact, or Power-Line interferences.
   The filter was zero-phase, to remove any possible alterations in the signals.

4. **Epoch Segmentation**. Segments were divided in epochs of 5 seconds, with 50 % overlap. This length was chosen in order to ensure stability of the evaluations. Having segments of about 2 minutes, it was possible to obtain about 47 epochs from each segment.
   This had to be multiplied for the number of channels involved in the analysis. In the considered database, data were obtained from a 19 channels EEG [16].

5. **Mean Value Removal**. Mean value of the signal was removed from each epoch, to exclude any possible trendings.

The work in this Thesis focused on the classification between epochs of EEG signal respectively in the :

- **PREICTAL** phase;
- **INTERICTAL** phase.

The aim is to understand whether an Entropy-based Analysis could provide an optimal classification and detect the changes between preictal and interictal EEG signal. This result could lead to a further investigation on the possible application of Entropy to the training of algorithms to detect the onset of the preictal phase.
The database included time information of the onset of the epileptic events, and were used to extract segments of EEG signal from the different phases. Channels relatable to the Focal zone were also added as metadata in the database, and these information were used to get an Analysis with two different approaches:

- **ALL CHANNELS** involved. The classifier was receiving epochs of EEG signals from the whole set of EEG channels available.
- only the **FOCAL CHANNELS**. In this second approach were involved only epochs of EEG signals from the channels above the Focal zone.

On each of the two approaches, the aim was also to obtain a clear classification and recognition of the phase related to each epoch of the EEG signal. In this thesis, from the original database [16] it was decided to consider only the subject named PN10, experiencing 10 seizure events, classified as FBTC from the medical equipment. In the original file [16] the Frontal zone was the one where seizures where originated. FOCAL CHANNELS approach was involving EEG recordings from only channels above the Focal zone: *F3", "FP1", "F7", "FZ", "F4", "F8", "FP2"* were the channels selected for it, and all the features were evaluated exclusively from these channels.

3.2 Features

In the context of scientific data analysis and artificial intelligence (AI), a *feature* represents an individual measurable property or characteristic of a phenomenon under observation. Features serve as the fundamental building blocks for machine learning models, as they encode relevant information extracted from raw data, enabling algorithms to perform tasks such as classification, regression, or clustering effectively [65]. The process of identifying and recognizing appropriate features—known as feature extraction and feature selection-has a deep impact on the accuracy and generalizability of AI systems [66]. In many real-world applications, such as biomedical signal processing or seismic event detection, carefully designed features can significantly enhance model performance while reducing computational complexity [67]. Consequently, the selection of informative, discriminative, and non-redundant features remains a crucial step in developing robust AI-based solutions.

Feature selection is a crucial step in the design of machine learning pipelines, particularly when working with high-dimensional datasets. Its primary goal is to identify and retain the most informative and non-redundant features, thereby improving model performance, reducing overfitting, and enhancing interpretability.

Feature selection methods can be broadly categorized into three classes:

- **filter methods**, which evaluate features based on intrinsic properties such as statistical correlation or mutual information;
- **wrapper methods**, which evaluate subsets of features through the performance of a predictive model;
- **embedded methods**, which perform feature selection during the model training process itself [68, 66].

In domains such as biomedical signal analysis, where features may be derived from various mathematical formulations or transformations, feature selection is essential to distill the most discriminative characteristics from potentially redundant feature groups.

3.2.1 Time Frequency Features

The work in this Thesis focused into assessing whether Entropic Features could provide an improvement into Signal Analysis. To evaluate the possible improvements, a group of Linear Features were selected to make possible the comparison. The subset was composed in order to work on both Time and Frequency Domain. All these features were evaluated on each epoch of the signal, considering the distribution of values inside it. With respect to the Time domain, it was chosen to include in this work the following features:

Feature	**Definition**
Mean	Average value of the samples within the epoch.
Median	Central value of the ordered samples distribution.
Standard Deviation	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Minimum	Minimum value within the epoch.
Maximum	Maximum value within the epoch.
RMS	$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$
Skewness	$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$
Kurtosis	$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$
Interquartil Range (IQR)	Difference between the third and first quartile: $Q_3 - Q_1$.

Table 3.1: Time-domain features extracted from each signal epoch

Not only time-domain features were evaluated, but also frequency-domain characteristics were analyzed.

Feature	**Definition**
Peak Frequency	Frequency corresponding to the maximum value of the Power Spectral Density $S(f_k)$.
Spectral Centroid	$f_c = \frac{\sum_{k=1}^{N} f_k S(f_k)}{\sum_{k=1}^{N} S(f_k)}$ Average frequency of the spectral power distribution.
Spectral Bandwidth	$\sqrt{\frac{\sum_{k=1}^{N} (f_k - f_c)^2 S(f_k)}{\sum_{k=1}^{N} S(f_k)}}$ Measure of the spread of the spectrum around the spectral centroid.
RMS of PSD	$\mathrm{RMS}_{S(f_k)} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} S(f_k)^2}$

Table 3.2: Frequency-domain features extracted from the Power Spectral Density

This set of linear features was applied to each database. When working with HRV signals, additional HRV-specific features were included.

Feature	**Definition**
RMSSD	$$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2}$$ Root Mean Square of successive differences between RR intervals.
pNN50	Percentage of successive RR interval differences greater than 50 ms.
LF Power	Spectral power within the Low Frequency band of the HRV signal.
HF Power	Spectral power within the High Frequency band of the HRV signal.

Table 3.3: Additional HRV-specific features

All these features in tables 3.1, 3.2, 3.3 were considered in order to create a meaningful analysis, with features not redundant to each other. The workflow was then applying a successive phase of Feature Selection in order to select the best subset of Linear Frequencies for each database.

The Time Frequency Features were considered as a gold standard: the performance of Entropic Features had to prove to represent an improvement with respect to the information provided by this set of features. This approach was applied on each of the involved databases. Linear Features involved in this work are basic and computational efficient features, which can be easily evaluated in real time. If a new approach (as the Entropic) aims to be applied, it needs to guarantee an improvement with respect to the methods already available. This is the reason why Time/Frequency features were considered , and performances from every different approach was compared and statistically analyzed with it.

3.2.2 Entropic Features

The second approach involved Entropic Features, obtained from different combination of parameters from the definitions introduced in the previous chapters.

Entropy Definitions	**m**	**r**	**c**	**a**	**p**	**Features**
Shannon						1
Renyi				2		1
Approximate	2–4	[0.15; 0.20; 0.25] * std				9
ApEn_mod	2	p percentage of recurrence points			0.1	1
ApEn_mod	2	[p/100] * std			10	1
Sample	2–4	[0.15; 0.20; 0.25] * std				9
Permutation	2–10					9
Dispersion	2–5		3–9			28
Distribution	2–5		3–8			24
Bubble	2–30					29
Total						**112**

Table 3.4: Entropy-based feature definitions and parameter configurations

The previous table 3.4 shows all the Entropy definitions involved, with all the relative parameters combinations. The Modified Approximate Entropy [6] was applied with a Theiler Window set to a value of 0. In the related code this means that only self-recurrences are excluded from the evaluation. The time delay *tau* is always set to a value of 1 for all the definitions involving it.
This lead to a various set of 112 Entropic features. Each of them was applied on each epoch of the signals, in order to evaluate complexity of the epochs with different Entropic definitions. A following phase of Feature Selection was involved in order to select a subset of features which could provide a better performance on the considered database. This reduced the dimensionality of the data, leading to work with a more informative subset.
Classification results obtained from Entropic Features were compared with those provided by Time/Frequency Features. A statistical analysis lead to understand if the improvement eventually obtained could have represented a statistically significant vari-

ation.

3.2.3 Mixed Features

A third approach involved a mixed set of Time/Frequency and Entropic Features. This combination focused on assessing whether the addition of Entropic Features could improve the performance of the set. As for the previous two approaches, a step of Feature Selection was applied. The features were selected only considering the informative content that they could have provided, not considering their nature (Time/Freq or Entropic). The subset could have indeed been composed by either just Linear Frequencies, just Entropic, or a mix of the most performative ones.
This Mixed approach gave us the opportunity to analyze in a different way the features' performance on each database, working with a subset of the best ones. Results obtained from it were compared with results provided from the Time/Frequency Features alone. The aim was indeed to always create a comparison with the Linear approach, to understand if an improvement was obtained.

3.3 Classifier

A classifier is a machine learning model designed to assign input data to one of several predefined categories or classes. The process typically involves training the model on a labeled dataset, where each input is associated with a known class label. The classifier learns to map input features to class labels by minimizing a loss Function, often through optimization techniques such as gradient descent [69]. Once trained, the classifier can predict the class label of new, unseen data based on the patterns it has learned.

3.3.1 Machine Learning

Machine learning (ML) is a subset of artificial intelligence that enables systems to learn from data and improve their performance over time without explicit programming. The core idea is to develop algorithms that can identify patterns within data and make predictions or decisions based on these patterns. ML models are typically trained on a dataset, where they learn to map input features to desired outputs by minimizing a loss function, often using optimization techniques such as gradient descent [69].
In this work the Machine Learning approach was used by Classifier, in order to evaluate classification performances of each Entropy definition. Classifiers were used with all different databases, in order to analyze performances on to different type of data, and different number of classes.

Machine Learning approaches can suffer from limitations on the number on data. This could lead to bad Accuracy values from the Classifier, or to problems like Overfitting. Overfitting occurs when a model learns not only the underlying patterns in the training data but also the noise and details that do not generalize to new data. This leads to a model that performs well on the training set but poorly on test data, indicating a lack of generalization [70].

Overfitting is often a result of several factors:

- **Model Complexity**: Complex models with many parameters can fit the training data very closely, capturing even the minor fluctuations that do not represent the true underlying patterns [70].
- **Insufficient Training Data**: A small or unrepresentative training dataset may not provide enough information for the model to learn the general patterns, leading it to memorize the specific examples [70].
- **Excessive Training Time**: Allowing the model to train for too many iterations can lead it to fit the noise in the data, rather than the actual signal [70].

All these possibilities could lead to a good, even optimal performance of the classifier on the Training Set. The performance could then get way worse when applying the Classifier to new data, unseen from the Classifier. The Classifier has been trained a lot, and maybe too much (or too well) just on a Training Set, which could also be quite small. This leads to a difficulty into classifying new data, generating misclassifications, and bad Accuracy values.

In the figure 3.1 are shown the Overfitting and Underfitting situations. The graphs

Figure 3.1: Underfitting, Overfitting on Training and Validation Set

represents the number of misclassifications produced (y axis) in relation to the number of iterations (x axis).

- In an Optimal situation, the classifier can produce a large number of errors at the start, but with progressive iterations, the performance improves, getting better both on Training and Validation Set.

- In the Overfitting situation, the number of errors gets reduced only on Training set, where the performance becomes more than optimal, with very few errors. If the performance on Training Set is good, the one on the Validation doesn't improve with time. This leads to a possible bad performance of the classifier, with new data.

- In the Underfitting situation , the performance remains always stable, in both Training and Validation Set, even with increasing number of iterations. This means that the Classifier has bad performances in general, but at least doesn't create differences between Training and Validation Set.

Overfitting is in general the situation to avoid, because it can suggest that our classifier is working good, but it is just because it is working on the Training Set. To mitigate overfitting, several strategies can be applied:

- **Cross-Validation**: This technique involves partitioning the data into multiple subsets and training the model on some subsets while testing it on others. This helps in assessing the model's performance and ensuring it generalizes well to unseen data [70].

- **Regularization**: Methods like L1 (Lasso) and L2 (Ridge) regularization add a penalty to the loss function to constrain the model's complexity, discouraging it from fitting the noise in the data [70].

- **Early Stopping**: During training, the model's performance is monitored on a validation set, and training is halted when performance starts to deteriorate, preventing the model from overfitting [71].

- **Dropout**: In neural networks, dropout involves randomly setting a fraction of input units to zero at each update during training time, which helps prevent overfitting by reducing complex co-adaptations of hidden units [72].

Understanding and addressing overfitting is crucial for developing robust machine learning models that perform well on new, unseen data. Another typical Artificial Intelligence approach is **Deep Learning**, which has both similiarities and differences with Machine Learning. One of the major differences is that Machine Learning works with a subset of features already extracted from the user. This is called Feature

Extraction. Then, a following phase of Feature Selection could de applied in order to reduce the dimensionality of the Features Set. In Deep Learning, is the classifier itself which works with received data and extracts its own Features Set.
In this work, a Machine Learning approach was used, because we were providing the Features Set for the Classifier: it was the user the one performing the Feature Extraction.

3.3.2 Gaussian Naive Bayesian

The Gaussian Naive Bayesian (GNB) classifier is a probabilistic model that applies Bayes' theorem with the assumption that features are conditionally independent given the class label [73]. This assumption greatly simplifies the estimation of probabilities and allows the model to scale well to high-dimensional datasets with limited training data [74].

Bayes' theorem provides a principled way to compute the posterior probability of a class C given an observed feature vector (x). It is defined as:

$$P(C|\mathbf{x}) = \frac{P(\mathbf{x}|C)P(C)}{P(\mathbf{x})} \tag{3.3}$$

where

- $P(C|\mathbf{x})$ is the posterior probability of class C given the feature x. It is the probability of x belonging to the class C.
- $P(\mathbf{x}|C)$ is the likelihood of observing **x** under class C. In this situation, the class C is already there, and this is the likelihood that inside that class is considered x.
- $P(C)$ is the prior probability of the class: can be considered as the general probability that a data is considered in class C.
- $P(\mathbf{x})$ is the evidence, which acts as a normalizing constant. It is the general probability of observing x, independently of the classes.

In the figure 3.2 is shown the possibility of having two classes (A and B) for the classifier.

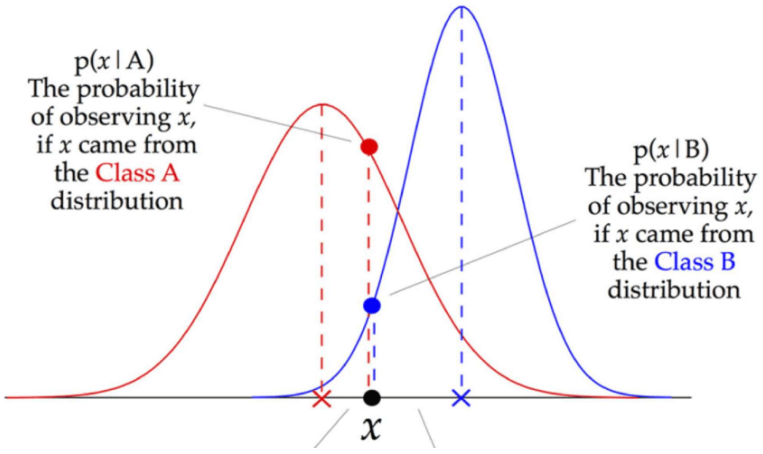


Figure 3.2: Gaussian Naive Bayesian classifier

As we can see, for a certain data (x) are compared the probabilities of x belonging to the class A , or to the class B. The classifier compares the 2 probabilities values, and then assigns the data to the class which has the higher probability value on it. In the Gaussian Naive Bayesian variant, each feature is assumed to follow a Gaussian distribution, and the likelihood $P(\mathbf{x}|C)$ is computed as the product of the individual Gaussian densities for each feature. The model expressed in the figure 3.2 is referred to a 2-classes situation, but can easily be adapted for an approach involving way more classes. Each class has to always be represented with a Gaussian distribution, in order to be applied into this classifier.
In the end, the GNB needs two assumptions to be made, in order to be applied:

- features are independent one from the other: **Naive assumption**.
- Each class has a Gaussian distribution of data : **Gaussian assumption**.

These are two big assumptions , that are not so common to find in data distribution, and that could create some inductive biases if the classifier is applied on data not supporting these two hypothesis. Despite its strong independence assumption—which may not hold in real-world scenarios—GNB classifiers often provide robust and interpretable results, especially when feature dependencies are weak or the dataset is noisy [75].

3.4 Sequential Forward Selection

3.4.1 General overview

The three different approaches previously introduced considered a various set of features. Especially when working with Entropic features, the number of features involved was much major than the number of features on the Time/Frequency approach. In order to obtain a more balanced Analysis, it was chosen to work with a selected subset of features on each dataset.
Sequential Forward Selection (SFS) [76] was chosen to obtain the feature selection on each database.Sequential Forward Selection is a wrapper-based feature selection method, as it evaluates the contribution of each feature subset by directly optimizing the performance of a given classifier. It was chosen to select a subset composed by 5 different features when working on each database. These features were chosen based with their Accuracy provided when working with the GNB classifier. SFS and GNB were used as a unity in the Feature Selection phase. The selection of features was provided by the Inner-Cross Validation method [77], always working with the GNB.

3.4.2 Inner Cross-Validation

In this work the whole dataset was divided into 5 different folds, of the same data dimension:

- 4 folds were used to obtain the **Outer Training Set**;
- the 5th fold was used as a **Test Set**.

This operation was repeated for 5 different times, rotating the folds of data, in order to always have a different fold of data as the Test Set. The Outer Training Set was the one considered in order to provide the Feature Selection with the Sequential Forward Selection method. It was divided into three Inner folds:

- 2 of these inner folds were composing the **Inner Training Set**;
- the 3rd Fold was used as an **Inner Validation Set**.

This inner division was repeated 3 times on each Outer Training Set, in order that every fold could have been used as an Inner Validation Set.
The Sequential Forward Selection [76] was applied on the Inner Folds, in order to select a subset of features which could have been significant and provide the most informative content to the Analysis. This method aimed to create a subset of 5 features, which worked as a unit together with GNB to create the best Analysis on each database.

1. All features got considered at first alone, to evaluated their classification performance on the Inner Validation Set. Subsequently, the Inner Validation Set was rotated, and a new Inner Fold was used. Having 3 Inner Folds, this process was applied 3 times, and each feature produced a different Accuracy on each Validation Fold. These 3 values were averaged, leading to having a single Accuracy value for each autonomous feature. The feature providing the best averaged Accuracy was selected as the 1st to add the Feature Selection Subset.

2. This means that from a theoretical starting set of N features, 1 was included in the 5 selected, and the other (N-1) were still under analysis. In the following step each of the (N-1) features was tested to work together with the previous selected feature. The performance of them was evaluated on the Inner Validation Test: this lead to a value of Accuracy for each couple of features (composed by the 1st already selected, and in rotation the other tested). If an hypothetical feature X was the one classifying better coupled with the feature previously selected, that feature X would have became the 2nd feature added to the Selected Features Subset. Using this approach avoids the possibility of selecting two redundant features, or

to include a couple of features providing the same informative content. A feature gets added to the Subset only if it can provide an increase in the Accuracy. It is not mandatory that if Feature Y was the 2nd best feature (when analyzed as autonomous) it would then be added as the 2nd feature of the subset. The aim is to obtain a final subset which provides a complete informative content, and not redundant. The addition aims to add new information and features that can add different information to the ones already selected.

3. This work of addition of features and iterative testing was conduced until a final subset of 5 features was obtained. It is not mandatory that this final set is composed by the top 5 features which provided the best Averaged Accuracy when analyzed autonomously. The aim is indeed to produce a subset which can produce complete information, avoiding redundancy. Each feature gets added if it can show that adds information and can become an asset for the classification.

GNB was the classifier providing Accuracy on the Inner Validation Set, and giving the opportunity to select features. It is impossible to consider this approach without mentioning the contribution of the Classifier. At the end of this Inner Cross Validation process [77], a final subset of 5 Selected Features was obtained. This was the subset which performed in the best way in this Outer Training Set. But as we previously mentioned, the whole dataset was divided into 5 folds, with each fold used as the Test Set on each different iteration. This leads to having 5 different Outer Training Sets, and each Outer Training has its respective Subset of Features Selected, obtained with the Inner Cross Validation Method.
The Figure 3.3 resumes the whole process of Feature Selection obtained with this Sequential Forward Selection approach, leading to 5 different subsets of features. This rotation of folds gave the opportunity to validate the performances on an always different inner folds.

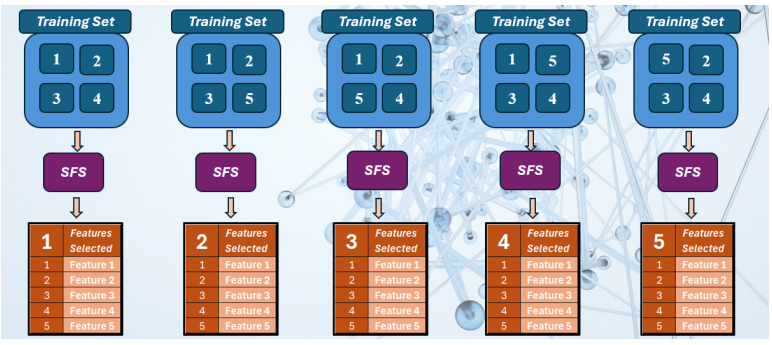


Figure 3.3: Inner Cross validation

3.4.3 Outer Cross-Validation

From the previous phase, it was possible to obtain 5 different Subsets of Features Selected, one for each Outer Training Set. These Features Subsets were then tested, always with the contribution of the GNB classifier, to classify the data in the Test Set. This external process is called Outer Cross Validation [76] and provided an Analysis of the data in the Test Set with always the more performative features. The Test Set was unknown for the classifier, and Accuracy was evaluated using the respective Subset of Features. Having 5 different folds, each one needed to be used once as the Test Set, leading to 5 iterations. On each iteration, the data on the Test Set were always different, giving the opportunity to test and evaluate the performance of the classifier on an always different number of unknown epochs of signal.
The Figure 3.4 resumes this approach, and exposes the rotation of folds used as the Test Sets.

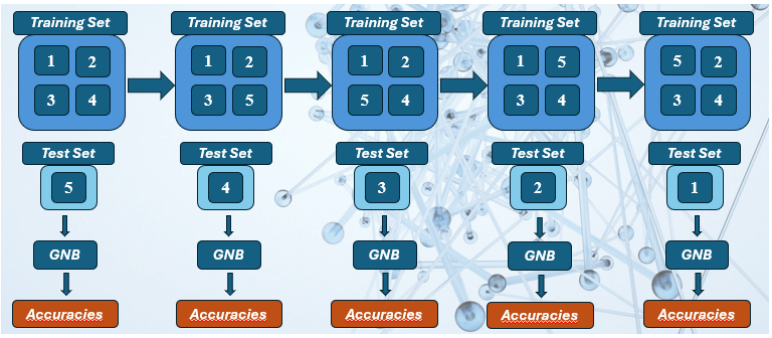


Figure 3.4: Outer Cross validation

All the epochs from the same recording were analyzed and classified.From each recording was obtained a certain Accuracy value, and a whole dataset was producing a number of Accuracies proportional to the number of recordings composing the dataset itself. In this section Accuracy was the metric involved to explain the method, but in the end Sensitivity and Specificity were the chosen metrics.

3.4.4 Data Partitioning

The Sequential Forward Selection approach worked on a division of the whole dataset in 5 different folds, with an always different Test Set. The aim was to propose always new data to the classifier, to provide a realistic and valuable classification. Working with different epochs and different signals, it was chosen to not divide epochs of the same signal: for example, if the epoch 54 of the Signal 24 was in the Test Set , it is not

possible to find an epoch of the Signal 24 in the Outer Training Set. All the epochs of a certain signal get included either in the Test, or in the Outer Training Set.
This avoids to create a situation where for example, Signal 24 has some epochs in the Training and some in the Test Set: this would have facilitated the classification for the GNB, and created a bias in the results. The decision was to not divide a signal, and assign all of its epochs either to Training or Test Set .
The rotation of Folds was also ensuring that in the Test Set there were always new and unknown data. Classification would have been facilitated if epochs of a certain signal get tested twice , for example in iteration 3 and 5. This rotation avoids this bias, and each epoch from each signal get tested once during the work on a certain database.
The division of data into Training and Test Set was always aiming to keep the proportion of a Training Set being 4 times more numerous than the Test Set. It was not always possible to keep this exact proportion, even because of the impossibility to divide epochs of the same signal. The figure 3.5 shows an example of how Training and Test Set were composed in the CHF database.

Subgroups	CHF		NSR	
Number of recordings	29		54	
Sets	*TRAINING SET*	*TEST SET*	*TRAINING SET*	*TEST SET*
Number of recordings	23	6	43	11

Figure 3.5: Training and Test Set composition

Each Training and Test Set showed epochs of signals from both NSR and CHF groups. With this approach, the classifier was always testing unknown data, which could have possibly been epochs of signals from either pathological or healthy subjects. Each fold was composed with epochs of not divided signals. The figure 3.5 shows the number of recordings involved in a certain fold, and it is noticeable that for example another iteration could have produced a slight different number of recordings . The aim was to keep the rate between Training and Test Set always as 4 to 1, but it was not always possible.

	TRAINING SET		TEST SET	
Fold	CHF	NSR	CHF	NSR
1	23	43	6	11
2	23	43	6	11
3	23	43	6	11
4	23	43	6	11
5	24	44	5	10
Total tested recordings			29	54

Figure 3.6: Training and Test Set composition

As we can see from image 3.6, the 5th fold provides a different distribution of recordings, but this is not an issue, because as we can see observe all the recordings from both subgroups gets tested once, coherently with our approach.
This image shows the number of recordings, but the Classifier was always working with Epochs: in order to get a more representative table it was chosen to include the number of signals, but in the end the classifier was always working with epochs.
Similar approaches were used also with the FANTASIA database, involving signals from either the group of elderly and young subjects.
In the EEG SIENA database the fold set was always composed by epochs of EEG signal from both interictal and preictal phases, with a proportion between Training and Test Set similar to the one applied in the other databases. This approach was kept while using data from all channels, and when considering only data from channels above the focal zone.

3.5 Metrics

Different metrics were involved in order to understand and evaluate the classification performances from the different approaches. Sensitivity and Specificity were considered as the main two metrics, for their ability in detecting False Negative and False Positive rate . Accuracy and Balanced Accuracy were also involved in this work, in order to produce added value to the evaluation.

3.5.1 Sensitivity

Sensitivity, also known as *recall* or *true positive (TP) rate*, measures the ability of a classifier to correctly identify epochs belonging to the positive class. In a binary classification problem, it quantifies the proportion of correct positive samples detection. Sensitivity is particularly important in applications where missing a positive event may

have critical consequences. In the context of biomedical signal analysis, for example, a low sensitivity would correspond to a high number of false negatives, meaning that relevant physiological events may remain undetected [78, 79].

Formally, sensitivity is defined as the ratio between the number of true positives (TP) and the total number of actual positive instances, which includes both true positives and false negatives (FN):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.4}$$

A high sensitivity value indicates that the classifier is effective in detecting the positive class, minimizing the occurrence of false negative errors. For this reason, sensitivity is often considered a key metric when evaluating classification models in medical diagnostic systems and anomaly detection tasks [80]. In this work, Sensitivity was involved as a powerful metric capable to detect FN, which is a result important to consider, especially when two of the databases focus on Detection problems. The FN rate is crucial and the analysis needs to consider the rate of missed pathological conditions. An improvement in Sensitivity value could possible highlight a method more suited for real time analysis, and providing a more efficient alerting system.
A Positive value in the classification represents an epoch detected as Pathological, where a Negative an epoch considered as healthy. If the classifier doesn't reckon the epoch as pathological (producing a False Negative) the consequences could be dramatic for the subject. This is the reason why Sensitivity was involved in this work and considered as the main metric in order to evaluated the performance of the different approaches. Being the Analysis focused on monitoring of Biomedical Signals, the rate of False Negatives produced is crucial to assess the possible translatability of the method to real applications and devices.

3.5.2 Specificity

Specificity, also referred to as the *true negative rate*, measures the ability of a classifier to correctly identify instances belonging to the negative class. In binary classification problems, it represents the proportion of correct negative samples classifications. This metric is particularly useful to evaluate the false alarms avoiding, since a low specificity would correspond to a high number of false positive predictions [79, 78].(false alarms)

Formally, specificity is defined as the ratio between the number of true negatives (TN) and the total number of actual negative instances, which includes both true negatives and false positives (FP):

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.5}$$

A high specificity value indicates that the classifier is able to correctly recognize negative samples while minimizing false positive errors. In biomedical signal analysis and medical diagnostic systems, specificity is often considered together with sensitivity in order to obtain a more complete evaluation of classification performance [80].
A classifier which provides an high FP rate creates an anomalous number of alarms. Too many events are considered as Positive, and pathological. The subject receives alerts even when the event was referred to a clinical normal condition. In order to provide an efficient and valuable tool, the classification needs to reduce the number of FP produced, and Specificity can monitor it. In this work, this metric is applied with this purpose, and with the aim to reduce the un-needed alarms.

3.5.3 Accuracy

In classification problems, the performance of a classifier is commonly evaluated using the *Accuracy* metric. Accuracy is defined as the ratio of the number of correct predictions to the total number of predictions made. Mathematically, it is expressed as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathbb{I}(y_i = \hat{y}_i)}{N} \tag{3.6}$$

where N is the total number of instances, yi is the true class label of the i-th instance, $\hat{y}_i$ is the predicted class label, and $\mathbb{I}(\cdot)$ is the indicator function that equals 1 if the condition is true and 0 otherwise [81]. In the context of multi-class classification, accuracy is computed by counting the number of instances where the predicted class matches the true class and dividing by the total number of instances. However, its interpretation strongly depends on the number of classes involved in the task. In a balanced classification problem with K classes, a random classifier would achieve a baseline accuracy equal to $\frac{1}{K}$. Therefore, an obtained accuracy should always be compared to this chance level to assess its significance. As the number of classes increases, the chance level decreases, making the classification task inherently more challenging. Consequently, an accuracy value that may appear low in absolute terms can still be considered satisfactory if it significantly exceeds the random baseline. For this reason, it is essential to relate the reported accuracy to the number of classes and the intrinsic complexity of the dataset when evaluating the effectiveness of a classification algorithm. It can also suffer from class imbalance, where the number of data in classes is very different.

In summary, it should always be analyzed the context, the number and numerosity of classes involved before judging the value of Accuracy produced from a classifier.
In this work Accuracy values were produced for all the epochs of a certain recording. So, for example, if the database CHF was composed by 29 recordings, we were producing 29 different Accuracy values. Each one was evaluating the classification provided by the classifier on each epoch of the recording itself.

3.5.4 Balanced Accuracy

Balanced Accuracy is a performance metric designed to provide a more reliable evaluation of classification models when dealing with imbalanced datasets. Unlike standard accuracy, which may be biased toward the majority class, Balanced Accuracy accounts for the classifier's performance on both positive and negative classes by averaging sensitivity and specificity. This allows a more balanced assessment of the model's ability to correctly identify instances from each class [82, 79].

Formally, Balanced Accuracy is defined as:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{3.7}$$

where sensitivity represents the true positive rate and specificity represents the true negative rate. This metric is particularly useful in classification tasks where the class distribution is uneven, as it prevents inflated performance estimates that may arise when a classifier favors the majority class [82]. Other metrics, for example the F1-score, focus more on the positive class, and having unbalances in its number of epochs could bias the evaluation of the ML approach.
In this work, many databases were unbalanced, and Balanced Accuracy was introduced in order to keep a realistic evaluation of the classifier performance regarding both FP and FN. It resumes both Specificity and Sensitivity values in an unique numerical value.

3.6 Statistical evaluation

The comparison between performances produced by Time/Frequency features compared with Entropic or Mixed Features was evaluated with a Statistical Analysis. The distribution of the Metrics values provided by the classifier was analyzed in order to assess whether the eventual improvements obtained could have been confirmed as Statistical significant. Statistic provides tools such as Parametric [83] and Non Parametric tests [84]. These tests work on assumptions of Gaussian distribution on the values. This assumption needs to be confirmed, to understand which is the best test to use to

continue with the Statistical evaluation.
Shapiro test [85] gives the opportunity to analyze the distributions, and it is a previous step needed to subsequently proceed with the correct test, and then evaluate the results obtained. The Shapiro test was applied on the distribution of the differences between the Accuracies provided by each approach. For example, if the Time/Frequency approach was providing a distribution X of accuracies on the epochs, it was then compared with the distribution Y of the accuracies provided by the Entropic approach. The Statistical evaluation was conduced considering the distribution of differences *(di = Xi - Yi)* between values from the two compared approaches.
The focus was indeed the evaluation of the differences, and the assessing of a possible statistical improvement in the comparison. Being Sensitivity and Specificity the main metrics involved in order to assess the rate of False Negatives and False Positives produced by the different approaches, Statistical evaluations was applied on the distribution of the values from these two metrics.

3.6.1 Shapiro test

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.8}$$

The Shapiro test is a statistical tool that can be used to analyze a distribution of values, and to understand if it can be considered as normal.In its formula , we can find:

- *x(i)* as the ordered values of the distribution . The values are ordered in a increased way, where for example x(1) is the least value.
- $\bar{x}$ is the mean value of the distribution;
- a are coefficients related to a covariance matrix. The values of these coefficients are proportioned to the order of the values inside the distribution:
- n is the dimension of the distribution itself.

The value produced from the Shapiro test can be used to understand if the data are normally distributed. In particular, the relative **p-value** was applied with the following approach:

- **p-value >0.05** suggests a Normal distribution of values. Parametric tests should be subsequently applied on the distribution.
- **p-value <=0.05** suggests a distribution with a major deviation from normality. In this situation, Non Parametric tests are more suited for a Statistical Analysis.

This preliminary test was applied on to all the Sensitivity and Specificity values distribution from all the databases, in order to understand the nature of them , and to correctly use the correct test .
Both parametric and non parametric tests work with the Differences

3.6.2 Parametric tests

If the Shapiro test confirmed a normal distribution of the differences, the statistical evaluation was then conducted using Parametric tests. [83] In this thesis it was chosen to apply the Paired t-test, expressed with the following formula.

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{3.9}$$

As for the Shapiro test, the distribution of the differences was the one considered, and in the formula we can observe:

- $\bar{d}$ as the mean of these differences;
- s_d as the standard deviation of the distribution of the differences;
- n as the number of differences included.

The value produced from the Paired t-test was then analyzed with the relative p-value, to assess the significance of the result:

- **p-value >=0.05** suggests a non significant difference between the two distributions of compared values.
- **p-value <0.05** suggests a statistical significant difference between the two distributions values.

This parametric test was indeed a powerful tool to assess the results provided, if the distribution of differences could be considered as Normal.

3.6.3 Non-Parametric tests

These non parametric tests [84] were applied when the preliminary Shapiro test was assessing a distribution of differences non normal. It was chosen to involve the Wilcoxon Signed Rank test as the Non Parametric test.

$$W = \min(W+, W-) \tag{3.10}$$

in this formula , we can observe $W+$ and $W-$ which can be evaluated in the following manner:

$$W+ = \sum R_i^+ \tag{3.11}$$

$$W- = \sum R_i^- \tag{3.12}$$

Ri are the ranked values of the differences between the two compared distribution.The ranking involves the absolute values of differences *(di)* which then get increasingly ordered, starting with *R=1* for the lowest difference. If two absolute values are coincident, for example both *i=2* (in absolute value), they get assigned to mean rank. If they should have been assigned to rank 2 and 3, they will get both a rank *R=2.5*.
At this point, the ranking procedure considers if the difference was positive of negative, and we get the Signed Ranks:

- **di>0**: the signed rank is considered as positive, and gets assigned to R_i^+, and considered for the evaluation of W+.
- **di<0**: the signed rank is considered as negative, and gets assigned to R_i^-. It will then be considered in the evaluation of W-.

From the value of W provided from the Non Parametric test, it is possible to continue the analysis considering the associated p-value. The approach is similar to the one for the Parametric test:

- **p-value >=0.05** suggests a non significant difference between the two distributions of compared values.
- **p-value <0.05** suggests a statistical significant difference between the two distributions of compared values.

4. Results

4.1 CHF database

Results in figures 4.1 show the Sensitivity values obtained in this work from the three different approaches.

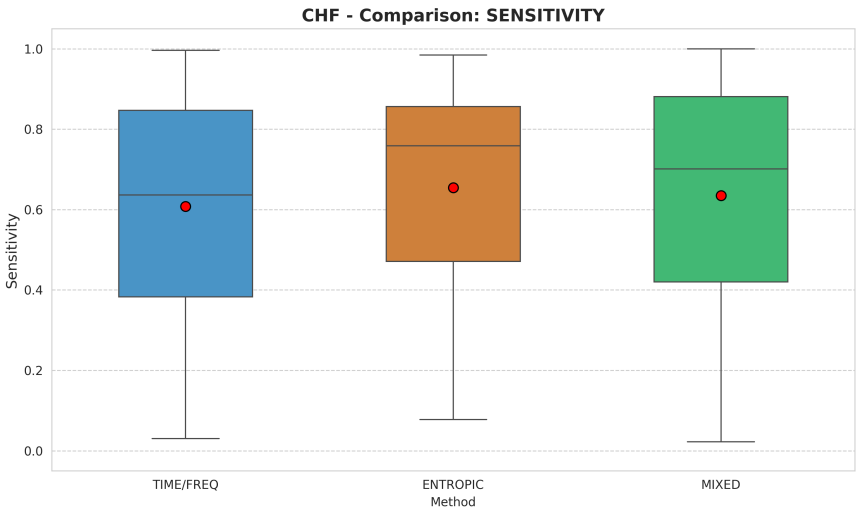


Figure 4.1: Sensitivity boxplot CHF

Figure 4.2 shows the results of the statistical evaluation of the Sensitivity values distribution. The Comparison always keeps Time/Frequency approach as the gold standard, and Mixed and Entropic results are compared with it.

Comparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0.02135	NO	Wilcoxon	0.00479	YES	ENTROPIC
TIME/FREQ vs MIXED	0.02093	NO	Wilcoxon	0.00988	YES	MIXED

Figure 4.2: Sensitivity statistical evaluation on CHF database

Statistical Analysis confirmed Entropic and Mixed approach as capable to improve

the Sensitivity on CHF database, reducing the number of False Negative produced by the Classifier.
Figure 4.3 shows the distribution of the Specificity values in the CHF database.

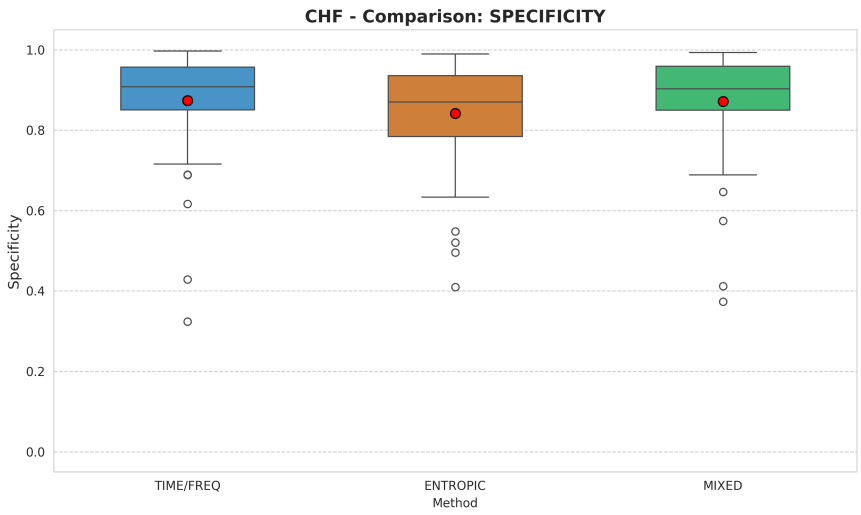


Figure 4.3: Specificity boxplot CHF

Figure 4.4 resumes the Statistical Analysis on the Specificity values from the CHF database.

Comparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0,00E+00	NO	Wilcoxon	0.00008	YES	TIME/FREQ
TIME/FREQ vs MIXED	0,00E+00	NO	Wilcoxon	0.23388	NO	TIME/FREQ

Figure 4.4: Specificity statistical evaluation on CHF database

Entropic features didn't provide any improvements in the reduction of the number of FP produced from the classifier. Time/Frequency approach was still considered as the best possible in order to obtain a better Specificity.
The tables in figure 4.5 and 4.6 show mean values and standard deviations of the Metrics involved in this Analysis.

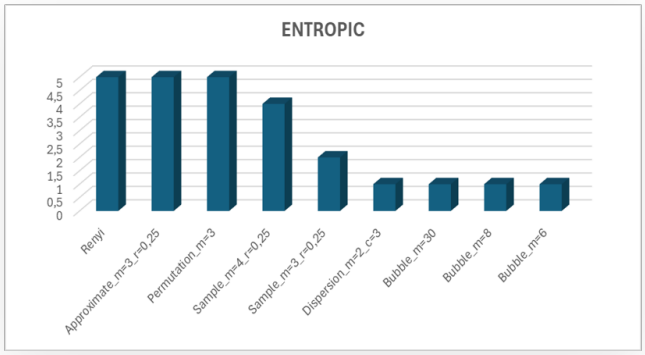
Metric	TIME/FREQ	ENTROPIC	MIXED
SENSITIVITY	0.608 ± 0.299	**0.655 ± 0.273**	**0.635 ± 0.305**
SPECIFICITY	0.874 ± 0.132	0.842 ± 0.130	0.872 ± 0.132

Figure 4.5: Sensitivity and Specificity values

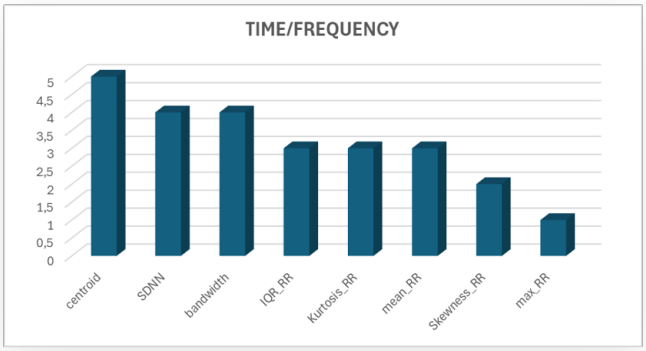
Metric	TIME/FREQ	ENTROPIC	MIXED
ACCURACY	0.781 ± 0.249	0.776 ± 0.218	**0.789 ± 0.237**
BALANCED ACCURACY	0.750 ± 0.130	**0.768 ± 0.131**	0.763 ± 0.136

Figure 4.6: Accuracy and Balanced Accuracy values

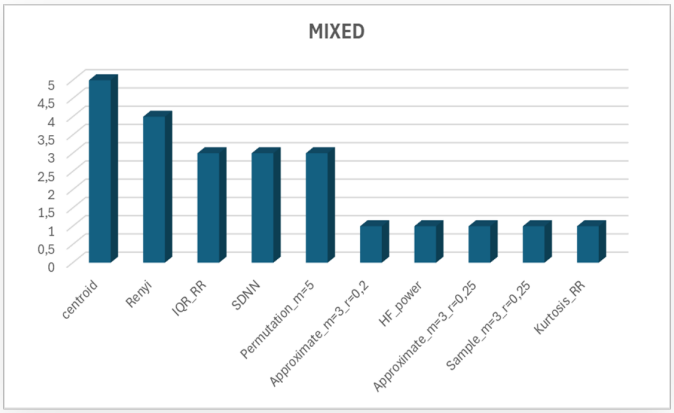
Accuracy and Balanced Accuracy were both improved by the involvement of Entropic features in the Analysis. Together with Sensitivity, all these metrics confirm the added value provided to the classification by Entropic features in the CHF database. In figure 4.7 are provided histograms of the features more frequently selected by the Classifier with the SFS approach. Each histogram shows the selected features on Entropic, Time/frequency and Mixed approaches. 5 different subsets of features were applied over the 5 different folds of the database. Centroid and Renyi provided themselves as the features selected most frequently.



(a) Entropy-based features



(b) Time/Frequency features



(c) Mixed features

Figure 4.7: Selected features in CHF database analysis

4.2 FANTASIA database

Results in figures 4.8 show the Sensitivity values obtained in this database from the three different approaches.

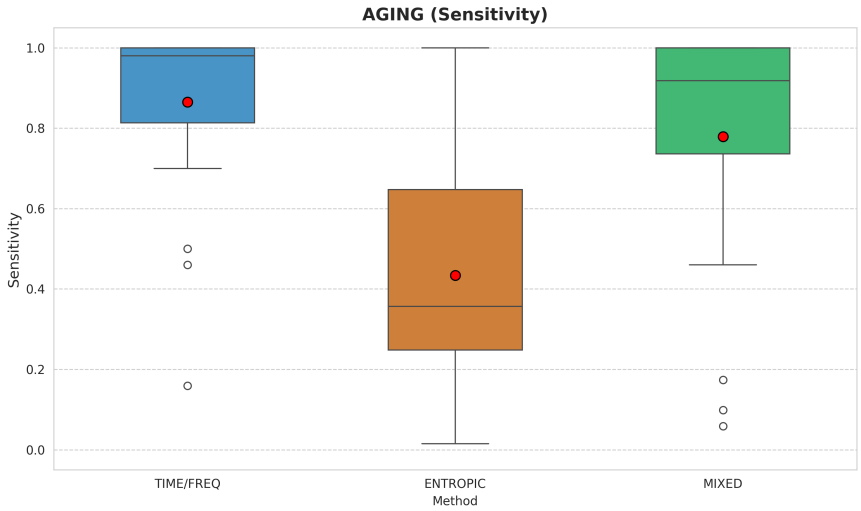


Figure 4.8: Sensitivity boxplot FANTASIA

Figure 4.9 shows the results of the statistical evaluation of the Sensitivity values distribution.

Comparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0.00001	NO	Wilcoxon	0.00017	YES	TIME/FREQ
TIME/FREQ vs MIXED	0.00001	NO	Wilcoxon	0.286	NO	TIME/FREQ

Figure 4.9: Sensitivity statistical evaluation on FANTASIA database

Entropic features didn't improve the Sensitivity in the classification. Time/Frequency approach remained the most reliable in order to obtain a minor number of FN produced in this database.

Figure 4.10 shows the distribution of the Specificity values in the FANTASIA database.

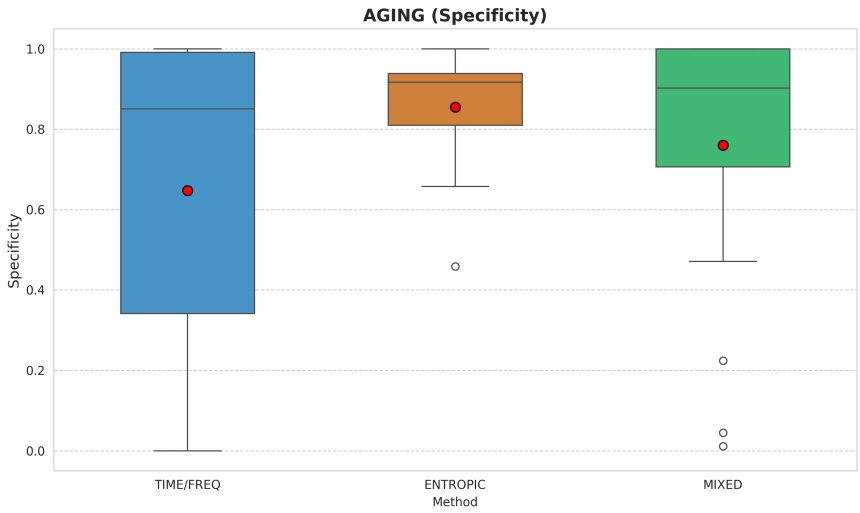


Figure 4.10: Specificity boxplot FANTASIA

Figure 4.11 resumes the Statistical Analysis on the Specificity values from the FANTASIA database.

omparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0.00053	NO	Wilcoxon	0.11193	NO	TIME/FREQ
TIME/FREQ vs MIXED	0.00019	NO	Wilcoxon	0.08428	NO	TIME/FREQ

Figure 4.11: Specificity statistical evaluation on FANTASIA database

Entropic features didn't provide significant improvement in the Specificity. From the boxplot, the values distribution seems more narrow, but Time/Frequency were always considered as the best approach considering also the reduced complexity of the method.
The tables in figure 4.12 and 4.13 show mean values and standard deviations of the Metrics involved in this Analysis.

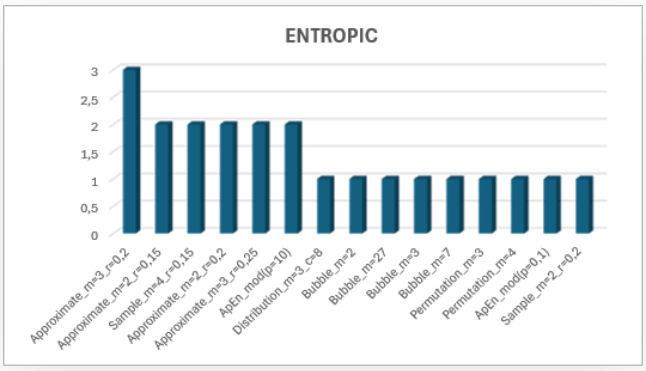
Metric	TIME/FREQ	ENTROPIC	MIXED
Sensitivity	0.865 ± 0.235	0.434 ± 0.274	0.779 ± 0.325
Specificity	0.648 ± 0.403	0.855 ± 0.135	0.760 ± 0.326

Figure 4.12: Sensitivity and Specificity values

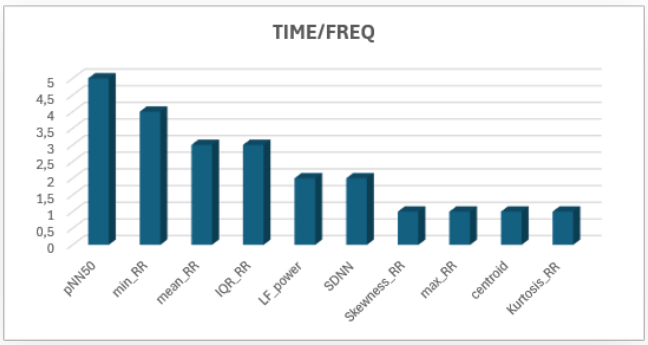
Metric	TIME/FREQ	ENTROPIC	MIXED
ACCURACY	**0.781 ± 0.331**	0.646 ± 0.284	0.769 ± 0.322
BALANCED ACCURACY	0.756 ± 0.236	0.644 ± 0.109	**0.769 ± 0.236**

Figure 4.13: Accuracy and Balanced Accuracy values

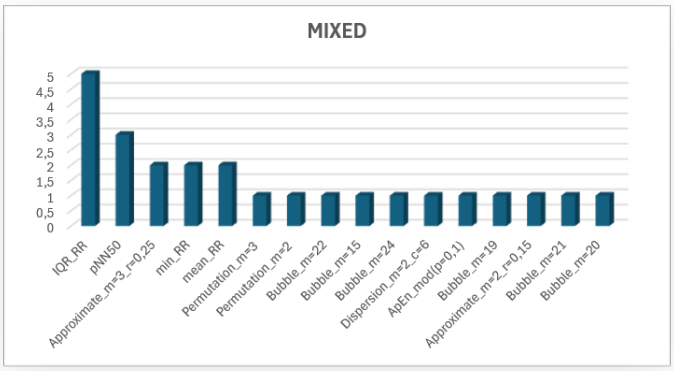
Mean values improved in Specificity when considering the Entropic and Mixed Approaches, but the Statistical Analysis confirmed Time/Frequency as the best approach in order to obtain a lower number of FP from the classifier. Balanced Accuracy was also improved by the introduction of Entropic features.
Figure 4.14 shows the various features selection in Entropic method, with 13 different features selected across the 5 folds. Time/frequency approach showed a more narrow subset of features, and pNN50 provided itself as the more selected. In the mixed approach, IQR RR was the most selected.



(a) Entropy-based features



(b) Time/Frequency features



(c) Mixed features

Figure 4.14: Selected features in FANTASIA database analysis

4.3 EEG SIENA database

Results in figures 4.15 show the Sensitivity values obtained in the EEG database from the three different approaches.

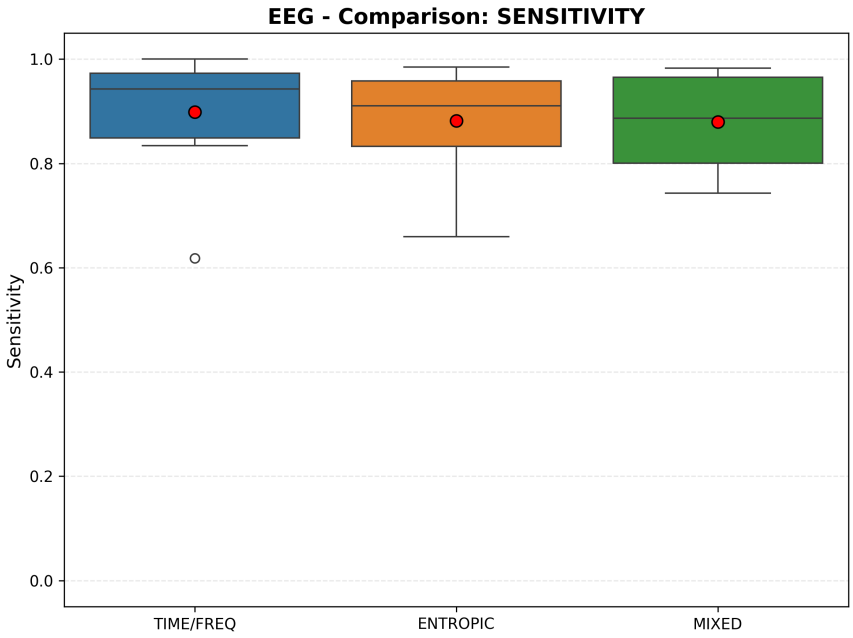


Figure 4.15: Sensitivity boxplot EEG

Figure 4.16 shows the results of the statistical evaluation of the Sensitivity values distribution.

nparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0.5858598675	SI	Paired t-test	0.7163979483	NO	TIME/FREQ
TIME/FREQ vs MIXED	0.3286163561	SI	Paired t-test	0.5980753896	NO	TIME/FREQ

Figure 4.16: Sensitivity statistical evaluation on EEG database

Entropic features didn't improve the Sensitivity in the classification. Time/Frequency approach remained the most reliable in order to obtain a minor number of FN produced in this database.

Figure 4.10 shows the distribution of the Specificity values in the EEG database.

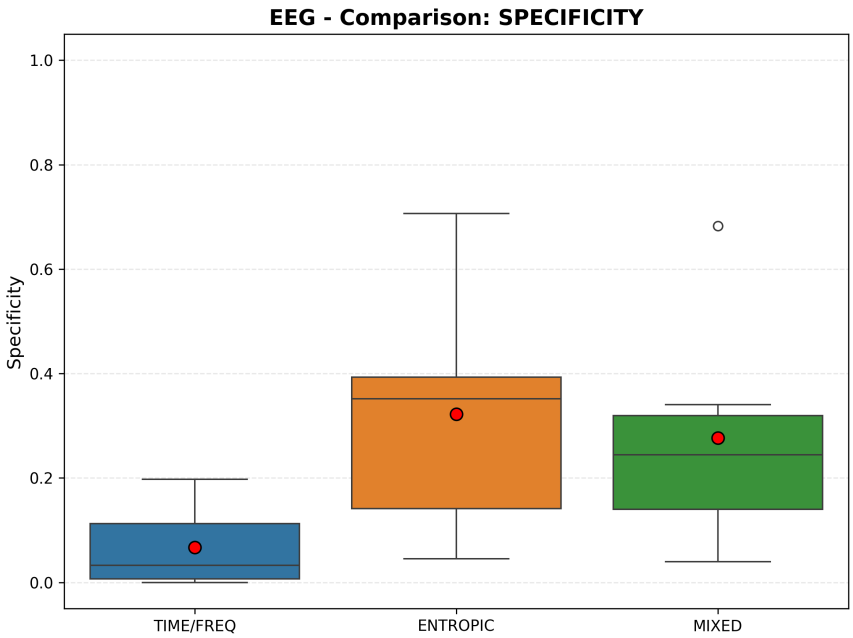


Figure 4.17: Specificity boxplot EEG

Figure 4.18 resumes the Statistical Analysis on the Specificity values from the EEG database.

Comparison	Shapiro p-value	Gaussian Distribution?	Statistical test	p-value test	Significative?	Best method
TIME/FREQ vs ENTROPIC	0.3369452132	SI	Paired t-test	0.03614389333	YES	ENTROPIC
TIME/FREQ vs MIXED	0.0523871798	SI	Paired t-test	0.07523605639	NO	MIXED

Figure 4.18: Specificity statistical evaluation on EEG database

Specificity remains quite low across the three different approaches, but Entropic features provided a statistically significant improvement in this metric. They managed to reduce the number of FP produced from the classifier.
The tables in figure 4.19 and 4.20 show mean values and standard deviations of the Metrics involved in this Analysis.

Metric	TIME/FREQ	ENTROPIC	MIXED
SENSITIVITY	0.898 ± 0.116	0.881 ± 0.103	0.880 ± 0.092
SPECIFICITY	0.067 ± 0.081	**0.322 ± 0.242**	0.277 ± 0.226

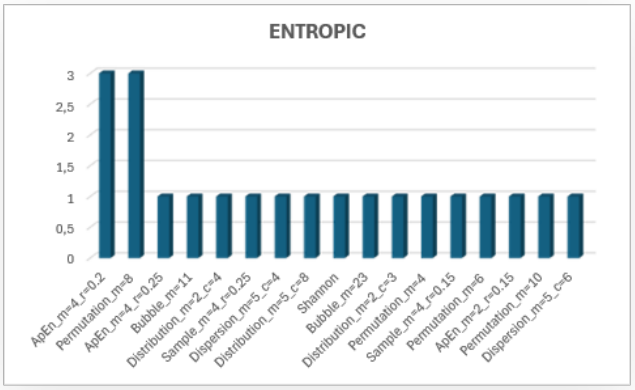
Figure 4.19: Sensitivity and Specificity values

Metric	TIME/FREQ	ENTROPIC	MIXED
ACCURACY	0.587± 0.428	**0.672 ± 0.323**	**0.653 ± 0.336**
BALANCED ACCURACY	0.293 ± 0.214	**0.336 ± 0.161**	**0.327 ± 0.168**

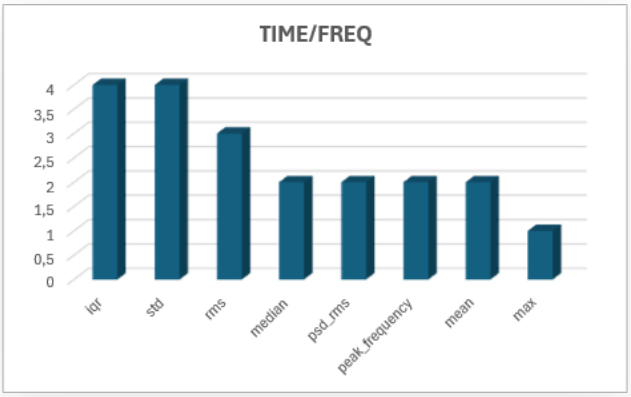
Figure 4.20: Accuracy and Balanced Accuracy values

Accuracy and Balanced Accuracy were two metrics both improved from the Entropic features.
Figure 4.21 shows the different histograms of features selected. Many different features were selected, especially when considering entropic ones. The Permutation Entropy , with a value of m=8 proved to be the most selected, in both Entropic and Mixed approach.

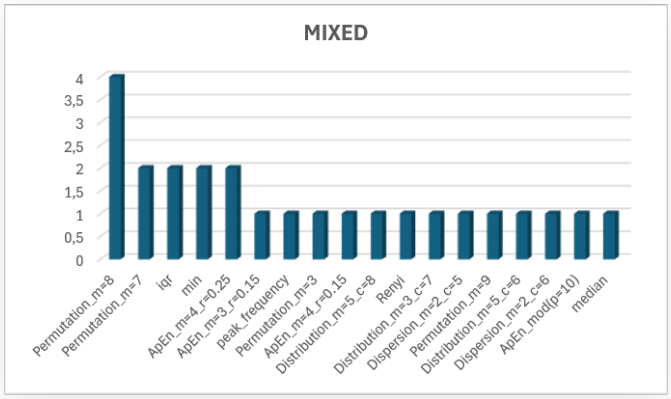
These results are always related to the involvement of all the EEG channels. Those provided by the FOCAL channels approach aren't shown, being them very similar to those obtained from considering the whole set of 19 EEG channels. This similarity can be related to the nature of the events: an FBTC event involves epileptic seizures spreading from the focal zone towards the whole brain. Having similar results can confirm the onset of generalized epileptic patterns already in the PREICTAL phase.
It is not impossible to hypothesize that if the events were FOCAL, working only with FOCAL channels could have provided a more precise detection of PREICTAL epochs.



(a) Entropy-based features



(b) Time/Frequency features



(c) Mixed features

Figure 4.21: Selected features in analysis involving ALL EEG channels

5. Discussion

Complexity Analysis is a subset of Signal Analysis which provides interesting research topics, and still many boundaries to define. The studies around it are various, and its scientific attractiveness is still intact. Entropy, since the first definitions provided in the 20th century, remains the main tool in order to assess and evaluate the complexity of a signal. In 1948 Shannon provided a first formula for its evaluation, and from it many more scientist developed many more Entropy definitions in the following years. The introduction of new Entropy definitions lead to a various features subset, giving the opportunity to analyze with new tools the complexity of signals.
Biomedical signals were the focus of this work, which aimed to assess if Entropy-based features could possibly improve the Signal Analysis.
Working with three different databases, with different signals (HRV and EEG) and different topics on each database was a choice taken in order to consider different biomedical situations. A database like the one involving epochs of HRV signals focused on the possible application of Entropic features in order to monitor the evolution of a cardiac disease as the heart failure. The relevance of this condition is always more prominent in our society, together with the increasing percentage of aging population. The work aimed to identify possible improvements in HRV analysis in subjects related to these conditions, making the monitoring easier for hypothetical Personalized Medicine applications.
EEG database focused instead of the possibility to get improvement, using Entropy based features, in the detection of the onset of epileptic events. This remains a challenging task in medicine, and it is still very difficult to predict the onset of the pathological condition. . A good result could lead to future application of Entropy in possibly some Brain Computer Interface, to monitor the electrical activity of the brain, and report to the patient the incoming epileptic event.
As described, Entropic features were applied with different objectives, and in different databases, in order to get a wider overview of their chances to improve Signal Analysis. Sensitivity and Specificity were considered as the main two metrics in order to evaluate the performances of the method, because of their ability to detect the FN and

FP produced. False Negatives are the most dangerous condition for a biomedical device, being related to epochs considered as healthy (Negative) where instead they were pathological (Positive). The results could be dangerous and tragic, and the reduction in the number of FN produced is needed. Sensitivity allows an evaluation of this aspect of the classification. False Positives are related to the False Alarm rate provided by the method. A false alarm isn't dangerous, but it is not good, and needs to be reduced to obtain a valuable tool for Signal Analysis. Specificity is the more suited metric to evaluate the production of FP. Accuracy and Balanced Accuracy were added as two metrics capable to provide further information to the Analysis.

Results were various in the different datasets. The addition of Entropic features to the Time Linear ones provided an improvement in the classification of epochs from the CHF database. Sensitivity, Accuracy and Balanced Accuracy were improved, with a statistical significant improvement in the Sensitivity values distribution in both Entropic and Mixed approaches. This result assesses how Entropic features can reduce the number of False Negatives produced by the classifier. This metric is very important, especially when considering HRV signal as a tool for monitoring devices. The number of False Negative produced is crucial in order to understand the possible translatability of the method to Real-Time Applications. An enhanced number of FN could expose the subject to dangerous conditions, where the device doesn't detect the onset of cardiac anomalies.Except from Specificity, all these three metrics (Sensitivity, Accuracy and Balanced Accuracy) were improved by the introduction of Entropic features.

Opposite results were reported from the other database with HRV signals, where Time/Frequency features widely assessed themselves to be the best method to monitor the effects of aging in cardiac variability. Time/Frequency approach provided the best results across the whole set of metrics, even though some non significant improvements were observed. For example, the increase in the mean value of Specificity from the Entropic and Mixed approaches, which couldn't produce a statistical significant improvement in the Analysis. The Statistical tests highlighted Time/Frequency as the best approach in Sensitivity . The number of FN produced by Entropic and Mixed approaches stood way too high, leading to a low Sensitivity value. For real time analysis, this result doesn't suggest the inclusion of Entropic features in monitoring of Physiological Aging.

The analysis of the histogram showed also how there wasn't a predominance of a certain entropic features, leading to a very various features subset. This result can also be related to the reduced number of subjects involved, which was also more limited than the CHF database. . This variability could be overcomed if a larger number of data was involved. Having opposite outcomes in the two databases with HRV signals could

also be related to the nature of the pathology. Heart Failure is associated with a clinical loss of responsiveness of the cardiac contractility. The related loss in complexity could provide a condition easier to detect for Entropic features.They indeed focus on the evaluation of irregularity and complexity of the signal, and CHF epochs of signals could be more suited for them.
It is not impossible to assume that aging provides instead a slight loss in complexity, and Time/Frequency can still provide a better classification of the signals. All the elderly subjects were considered as healthy from an analysis conduced by medical staff, so it is not impossible that the drop in complexity was not so relevant to make Entropic features the more suited for this kind of database.
Considering the EEG database, results obtained can be considered as positive. Entropic features and Mixed features provided themselves to be better suited than Time/Frequency features on the classification of EEG epochs. Sensitivity wasn't improved by the introduction of Entropic features,but the performances were high across all three approaches, highlighting the ability of the Classifier to detect Preictal epochs. Accuracy and Balanced Accuracy were both improved, with Entropic and Mixed Features producing a better performance when compared to Linear Features.
The most interesting result regards the Specificity, which evaluates the production of False Positive from the method. All the approaches produced a low value of Specificity: this means that the Classifier triggers very often false alarms, calling most of epochs as Preictal, even when they are related to a Interictal segment. Entropic features managed to improve the Specificity on a statistical significant level, reducing the number of FP, and consequently the false allarm rate.
This topic remains one of the most improvable sides of this preictal detection. If the aim is to provide an Epilepsy EEG-based monitoring device, it is needed to obtain a reasonable Specificity value, not alerting the patient with this frequency. Entropic features showed improvements, and good results in order to provide less Classification errors.
The limitation in the number of data could have also biased the Specificity metric, having a major number of Preictal epochs. The performance of this classifier should be tested again with a major and more balanced epochs distribution, in order to understand whether Entropy could really improve the detection of Preictal events.

Resuming, out of 3 different evaluations, Entropic features provided an improvement in 2 of them, with FANTASIA database being the only one where Time/Frequency features showed a better performance in classification. This positive trend suggests that Entropy-based features could detect and catch hidden dynamics that a classical analysis based on Time/Frequency features couldn't assess. These results were obtained with a

limited number of data, and with a supervised GNB classifier. Having a wider set of available data could have possibly provided a statistically significant improvement from the addition of Entropic features to the Signal Analysis. This suggests the possibility to continue this work including a major number of data, in order to assess the real added value provided by Entropic-based features.
This work could also be expanded including other classifiers, and working with a non supervised approach. The introduction of non linear classifiers, and a more solid deep learning approach could definitely assess whether Entropy-based features represent an added value to many biomedical topics. GNB was also a classifier assuming independency between features, which is something that it is not certain, and probably the involvement of other classifiers could help to understand if the statistical significance was missing because of data, or because of the classifier itself. SFS approach was involved in order to avoid the redundancy of information provided by different features. This lead to the opportunity of selecting features that could create a more complete and non redundant evaluation, creating meaningful features subset. This selection was correlated with the classifier itself; GNB and SFS were working as a unit, and the selection was obtained in order to get a better result on this specific classifier. Applying other classifiers could select other features, which could provide more information.
It is also needed a further evaluation to understand the computational complexity of the Entropic features, if the aim is to introduce them to real-time evaluations,for example in epileptic subjects. Entropy definitions involved in this work were not so computationally heavy, but the translation to a possible Brain Computer Interface needs to provide an elite computational velocity. If more complex ML approaches will be included in this method, it is not impossible to expect higher Computational Complexity, and a further Analysis is needed in order to get a better understanding of the translatability of the method.
This work provides many positive outcomes, and bring some suggestions for future clinical investigation around Complexity Analysis and their possible applications. Entropy-based features provided improvements, also statistically significant in some metrics. expressing the chance to investigate more and in a deeper way the performance of these features in scenarios as those described in this Thesis. The positive results could be worth the work related to this further analysis, and could create many improvements for pathological conditions.

Bibliography

[1] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[2] A. L. Goldberger, L. A. 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.

[3] G. Valenza, L. Citi, and R. Barbieri, "Interpretable multiscale entropy for biomedical signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1612–1621, 2017.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[5] A. Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[6] L. Mesin, "Estimation of complexity of sampled biomedical continuous time signals using approximate entropy," *Frontiers in Physiology*, vol. 9, p. 710, 2018.

[7] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[8] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

[9] M. Rostaghi and H. Azami, "Dispersion entropy: A complexity measure for time series," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.

[10] X. Li, C. Ouyang, and C. Li, "Sample entropy analysis of eeg signals under different pathological conditions," *Journal of Neuroscience Methods*, vol. 174, no. 1, pp. 99–107, 2015.

[11] M. Rostaghi and H. Azami, "Distribution entropy: A complexity measure for time-series data," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 1011–1021, 2018.

[12] G. Manis, M. Aktaruzzaman, and R. Sassi, "Bubble entropy: An entropy almost free of parameters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 2711–2718, 2017.

[13] H. Xie, W. He, H. Liu, and H. Zhao, "Improved bubble entropy and its application to biomedical signal analysis," *Biomedical Signal Processing and Control*, vol. 52, pp. 181–188, 2019.

[14] M. Liuni, A. Röbel, M. Romito, and X. Rodet, "Rényi information measures for spectral change detection," *arXiv preprint arXiv:1109.5876*, 2011.

[15] N. Iyengar, C.-K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 6, pp. R1078–R1084, 1996.

[16] P. Detti, "Siena scalp eeg database (version 1.0.0)." Data set, 2020. RRID:SCR_007345.

[17] T. C. C. of the New York Heart Association, *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*. Boston: Little, Brown Co., 6th ed., 1964.

[18] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, pp. 661–670, Mar. 1986.

[19] V. Alcan, "Sample entropy analysis of heart rate variability in rr interval detection," *Journal of Engineering Sciences and Design*, vol. 8, no. 3, pp. 783–790, 2020. Accesso online.

[20] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[21] N. Merah *et al.*, "R-peaks detection based on stationary wavelet transform," *Biomedical Signal Processing and Control*, vol. 18, pp. 1–8, 2015.

[22] N. Makwana *et al.*, "Hilbert transform based adaptive ecg r-peak detection technique," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 2, no. 5, pp. 539–544, 2012.

[23] M. K. Ojha *et al.*, "Efficient r peak detection algorithm from ecg using combination stationary wavelet transform and hilbert transform," *Solid State Technology*, vol. 63, no. 5, pp. 7791–7799, 2020.

[24] S. Vijayarangan *et al.*, "Rpnet: A deep learning approach for robust r peak detection in noisy ecg," *arXiv preprint arXiv:2004.08103*, 2020.

[25] T. F. of the European Society of Cardiology, the North American Society of Pacing, and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[26] M. Malik and A. J. Camm, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996.

[27] U. R. Acharya, K. Joseph, N. Kannathal, C. Lim, and J. S. Suri, *Heart rate variability: a review*, vol. 44. Springer, 2006.

[28] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, p. 258, 2017.

[29] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*. Philadelphia: Elsevier, 14 ed., 2021.

[30] J. J. McMurray *et al.*, "Esc guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Heart Journal*, vol. 42, no. 36, pp. 3599–3726, 2021.

[31] C. W. Yancy, M. Jessup, B. Bozkurt, *et al.*, "2013 accf/aha guideline for the management of heart failure," *Journal of the American College of Cardiology*, vol. 62, no. 16, pp. e147–e239, 2013.

[32] P. A. Heidenreich, B. Bozkurt, D. Aguilar, *et al.*, "2022 aha/acc/hfsa guideline for the management of heart failure," *Journal of the American College of Cardiology*, vol. 79, no. 17, pp. e263–e421, 2022.

[33] B. R. Saville, R. J. Mentz, and E. J. Velazquez, "Heart failure: epidemiology, pathophysiology, and clinical outcomes," *The Lancet*, vol. 401, no. 10391, pp. 2281–2294, 2023.

[34] P. Poirier, P. Voisine, J. G. Dumesnil, and et al., "Milrinone," *The Annals of Pharmacotherapy*, vol. 37, no. 4, pp. 580–588, 2003.

[35] G. Felker and R. Mentz, "Management of acute decompensated heart failure," *The Lancet*, vol. 396, no. 10244, pp. 1699–1712, 2020.

[36] P. Ponikowski, A. A. Voors, *et al.*, "2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Heart Journal*, vol. 37, no. 27, pp. 2129–2200, 2016.

[37] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[38] P. K. Stein and R. E. Kleiger, "Heart rate variability and cardiovascular risk," *Progress in Cardiovascular Diseases*, vol. 47, no. 5, pp. 331–346, 2005.

[39] A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti, "Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms," *British Heart Journal*, vol. 71, pp. 1–2, 1991.

[40] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[41] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "The hallmarks of aging," *Cell*, vol. 153, no. 6, pp. 1194–1217, 2013.

[42] E. G. Lakatta and D. Levy, "Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises," *Circulation*, vol. 107, no. 1, pp. 139–146, 2003.

[43] D. G. Strauss and R. H. Selvester, "Aging and the sinoatrial node," *Journal of Electrocardiology*, vol. 44, no. 6, pp. 700–704, 2011.

[44] D. R. Seals and M. D. Esler, "Aging and cardiovascular function," *Circulation Research*, vol. 105, no. 6, pp. 522–534, 2009.

[45] T. Niccoli and L. Partridge, "Ageing as a risk factor for disease," *Current Biology*, vol. 22, no. 17, pp. R741–R752, 2012.

[46] P. J. Uhlhaas and W. Singer, "Abnormal neural oscillations and synchrony in schizophrenia," *Nature Reviews Neuroscience*, vol. 11, pp. 100–113, 2009. Used here as general background on neural complexity decline with age.

[47] N. Montano, T. Gnecchi-Ruscone, A. Porta, F. Lombardi, and A. Malliani, "Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during aging," *Journal of Autonomic Nervous System*, vol. 49, no. 3, pp. 241–248, 1994.

[48] J. A. Lipponen and M. P. Tarvainen, "A review of heart rate variability analysis methods with clinical applications," *Frontiers in Physiology*, vol. 10, pp. 1–16, 2019.

[49] H. Tsuji, M. G. Larson, F. J. Venditti, E. S. Manders, J. C. Evans, C. L. Feldman, and D. Levy, "Reduced heart rate variability and mortality risk in an elderly cohort: The framingham heart study," *Circulation*, vol. 90, no. 2, pp. 878–883, 1994.

[50] M. A. Almeida-Santos, J. A. Barreto-Filho, J. L. Oliveira, and F. P. Reis, "A systematic review of the effects of age on heart rate variability in healthy individuals," *International Journal of Cardiology*, vol. 226, pp. 103–110, 2017.

[51] E. Niedermeyer and F. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 5th ed., 2004.

[52] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2nd ed., 2006.

[53] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, 4th ed., 2015.

[54] G. Pfurtscheller and F. Lopes da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[55] H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, p. 371–375, 1958.

[56] G. Klem, H. Lüders, H. Jasper, and C. Elger, "The ten–twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 52, no. 1, pp. 3–6, 1999.

[57] G. Chatrian, E. Lettich, and P. Nelson, "Ten percent electrode system for topographic studies of spontaneous and evoked eeg activity," *American Journal of EEG Technology*, vol. 25, no. 2, pp. 83–92, 1985.

[58] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[59] M. Okamoto, H. Dan, K. Sakamoto, and K. Takeo, "Three-dimensional probabilistic anatomical correlation of functional brain images using mni mri single-subject brain template," *NeuroImage*, vol. 21, no. 1, pp. 131–139, 2004.

[60] S. Beniczky, "Updated classification of epileptic seizures: Position paper of the international league against epilepsy," *Epilepsia*, vol. 66, no. 4, pp. 753–759, 2025.

[61] T. D. Nascimento, "Focal epilepsies: Update on diagnosis and classification," *Epileptic Disorders*, vol. 25, no. 1, pp. 1–13, 2023.

[62] P. Kwan, S. C. Schachter, and M. J. Brodie, "Drug-resistant epilepsy," *New England Journal of Medicine*, vol. 365, no. 10, pp. 919–926, 2011.

[63] R. S. Fisher, J. H. Cross, J. A. French, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. Roulet Perez, I. E. Scheffer, and S. M. Zuberi, "Operational classification of seizure types by the international league against epilepsy: Position paper of the ilae commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, 2017.

[64] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman, "The epileptogenic zone: general principles," *Epileptic disorders: international epilepsy journal with videotape*, vol. 8, pp. S1–9, 2006.

[65] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[66] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, pp. 1157–1182, 2003.

[67] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.

[68] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[69] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[70] IBM, "What is overfitting?," *IBM*, 2021.

[71] W. Contributors, "Early stopping," *Wikipedia*, 2024.

[72] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[73] I. Rish, "An empirical study of the naive bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, pp. 41–46, 2001.

[74] H. Zhang, "The optimality of naive bayes," *AA*, vol. 1, no. 2, p. 3, 2004.

[75] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[76] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1100–1103, 1971.

[77] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.

[78] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[79] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

[80] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[81] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756*, 2020.

[82] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," *Proceedings of the International Conference on Pattern Recognition*, pp. 3121–3124, 2010.

[83] D. C. Montgomery, *Design and Analysis of Experiments*. John Wiley & Sons, 9 ed., 2017.

[84] W. J. Conover, *Practical Nonparametric Statistics*. John Wiley & Sons, 3 ed., 1999.

[85] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.