THE UNIVERSITY
of EDINBURGH

Design and Application of Dispersion Entropy Algorithms for Physiological Time-Series Analysis

Evangelos Kafantaris

Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2022

To the family and friends who shared this journey

To those who spread around the pond and crossed the Indus

Abstract

Changes in the variability of recorded physiological time-series have been connected with transitions in the state of the monitored physiological system. The two primary paradigms describing this connection are the Critical Slow Down (CSD) and the Loss of Complexity (LoC) paradigms. The CSD paradigm considers that during frail or pathological states, a slowing down is observed in the capacity of the system to recover from external stressors resulting in increased output complexity for certain regulated variables. The LoC paradigm suggests that when the equilibrium of a system is disrupted, certain effector variables that displayed multi-scale complexity produce output measurements of reduced variability indicating a loss in the system's flexibility and capacity to adapt in the presence of external stressors.

For this purpose, entropy has emerged as a prominent nonlinear metric capable of assessing the non-linear dynamics and variability of time-series. Consequently, multiple entropy quantification algorithms have been developed for the analysis of time-series. These algorithms are based on Shannon Entropy such as the Permutation Entropy and Dispersion Entropy (DisEn) algorithms; and on Conditional Entropy such as the Sample Entropy and Fuzzy Entropy algorithms.

Within the scope of this study, the univariate and multivariate DisEn algorithms, first introduced in 2016 and in 2019 respectively, are used as the foundation and benchmark for the introduction of novel algorithmic variations. The selection of the DisEn algorithms is made due to their capability of producing features with significant discrimination capacity taking into consideration amplitude-based information while maintaining a linear computational complexity and having a functional multivariate variation capable of quantifying cross-channel dynamics.

To initially ensure the effective quantification of DisEn during univariate physiological time-series analysis, the effect of missing and outlier samples, which are common occurrence in physiological recordings, is studied and quantified. To improve algorithmic robustness, novel variations of the univariate DisEn algorithm are introduced for the analysis of low recording quality time-series. The original algorithm and its variations are tested under different experimental setups that are replicated across heart rate variability, electroencephalogram, and respiratory impedance time-series. The analysis indicates that missing samples have a reduced effect on the output DisEn and the error percentage can be maintained at values lower than 8% with the introduction of a variation that skips invalid values. Contrary to missing samples, outliers have a major disruptive effect with error percentages in the range of 57% to 73% for the original DisEn algorithm that is limited in values lower than 22% with the introduction of respective variations.

To expand the study from univariate to multivariate analysis, the multivariate DisEn algorithm is applied to physiological network segments formulated from multi-channel recordings of synchronized electroencephalogram, nasal respiratory, blood pressure, and electrocardiogram signals. The effect of outliers, present across different channels, is quantified for both univariate and multivariate DisEn features. The sensitivity of DisEn features to outliers is utilized for the detection of artifactual network segments using logistic regression classifiers. Two variations of the classifier are deployed in several experimental setups, with the first utilizing solely univariate and the second both univariate and multivariate DisEn features. Noteworthy performance is achieved, with the percentage of correct network segment classifications surpassing 95% in a number of experimental setups, for both configurations.

Finally, to improve DisEn quantification during the analysis of multivariate systems for physiological monitoring applications, the framework of Stratified Entropy is introduced. Based on the framework, a set of strata with a clear hierarchy of prioritization are defined. Each channel of an input multi-channel time-series is allocated to a stratum and their contribution to the output DisEn value is determined by their allocation. Three novel Stratified DisEn algorithms are presented, as implementations of the framework, allowing multivariate analysis with controllable contribution from each channel to the output DisEn value. The original algorithm and the novel variations are implemented on synthetic time-series consisting of 1/f and white Gaussian noise, waveform physiological time-series and derived physiological data. The introduced Stratified DisEn variations operate as expected and correctly prioritize the channels allocated to the primary stratum of the hierarchy across all synthetic time-series setups. The results of waveform physiological time-series indicate that certain of the novel features extracted through Stratified DisEn achieve effect size increases in the range of 0.2 to 1.4 when separating between states of healthy sleep and sleep with obstructive sleep apnea. The derived physiological data results further highlight the increased discrimination capacity of the novel features with increases in the range of 5% to 30% in the mean absolute difference between values extracted during steady versus stressful physiological states. Furthermore, an example of decrease in the output DisEn values when moving from a steady to a stressful physiological state is highlighted during the prioritization of the heart rate channel, in alignment with LoC, providing an example of how Stratified Entropy could be used to test hypothesis based on the CSD and LoC paradigms.

By making steps towards addressing the challenge of low data quality and providing a new framework of analysis, this thesis aims to improve the process of assessing and measuring the variability of physiological time-series, leading to the consequent extraction of viable physiological information.

Lay Summary

To acquire a better understanding of the functionalities of human physiology multiple devices have been developed that allow the recording of physiological signals. These signals are usually composed of the electrical activity of different organ systems of the human body and can provide important insights with regards to the physiology of the monitored individual.

Manual inspection of these signals from medical personnel allows the prognosis and diagnosis of disease and also supports medical decision making. However, due to the immense amounts of signals being recorded there is a limit as to how much information can be retrieved manually. Consequently, to fully utilize these recordings, there is growing interest of developing computer programs – algorithms, capable of analyzing these vast amounts of data. The development of such algorithms is a challenging task due to specific characteristics of those signals.

Physiological signals are complex and their nature can change rapidly, based on the operation of the corresponding organ system. Furthermore, different organ systems in the human body are in constant interaction with each other and therefore during the analysis of such signals it is important to consider not only the signal itself but also the interactions observed between the particular signal and signals of other organ systems. Finally, the quality of the recordings is usually low, since an individual has to be able to function normally throughout the day if the recording occurs under normal conditions, or they might be in interaction with medical personnel when in a clinical environment.

This work studies algorithms that are capable of measuring the variability of patterns needed to describe a particular signal. The more patterns required for the description of the signal the higher the value of variability assigned to that signal. Measuring the variability of physiological signals allows the detection of changes in the physiology of individuals, providing valuable insights for their health.

Therefore, the presented work introduces novel variations of these algorithms, developed to address the aforementioned characteristics in order to improve their viability for the analysis of physiological signals in the following three ways. First, by improving their effectiveness during the analysis of signals of low quality when applied to one signal at a time. Second, by partially automating the separation between signal segments of acceptable quality that are ready for analysis and those that are of low quality and might require further preprocessing. Third, by developing a new methodology for the analysis of multiple signals at a time, that allows the patterns of certain signals to be prioritized over others when this would allow a more accurate understanding of the interactions between the respective organ systems.

Acknowledgements

I would like to express my gratitude to my supervisor Dr. Javier Escudero Rodriguez for his keen interest, invaluable guidance and continuous support. I am truly grateful for having the opportunity to work with him and will always remember our in-depth discussions.

I would like to also thank Prof. Sotirios Tsaftaris and Dr. Tsz-Yan Milly Lo for co-supervising this project and Dr. Ian Piper from the National Health Service for his technical insights.

Furthermore, I would like to thank Prof. John Thompson and Dr. James Hopgood from the Institute for Digital Communications and Dr. Martin Reekie from the Institute for Bioengineering for our exchange of ideas over the many years of my undergraduate and postgraduate studies.

Throughout this journey I was truly fortunate to meet and converse with people who helped me acquire a long-term perspective of my work. I am grateful to Prof. Joseph Sifakis, Dr. Costas Bekas and Dr. Efi Kokiopoulou.

Above all, the completion of this endeavour would not be possible without the support of my family. I am immensely grateful to have them in my life.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Evangelos Kafantaris
14th November 2022

Contents

Figures and Tables

Figures

Tables

Abbreviations

AltMetDisEn	Alternative Statistical Metrics Dispersion Entropy
ApEn	Approximate Entropy
BP	blood pressure
BVP	blood volume pulse
CEn	Conditional Entropy
CSD	Critical Slow Down
DisEn	Dispersion Entropy
DynSkipDisEn	Dynamic Skip Sample Dispersion Entropy
ECG	electrocardiogram
EEG	electroencephalogram
EMG	electromyographic
FuzzyEn	Fuzzy Entropy
HR	heart rate
LinInterDisEn	Linearly Interpolated Dispersion Entropy
LoC	Loss of Complexity
LogSig	Logarithm Sigmoid Function
MEG	magnetoencephalogram
mvMDE	Multivariate Multiscale Dispersion Entropy
mvMPE	Multivariate Multiscale Permutation Entropy
NCDF	Normal Cumulative Distribution Function
OSA	Obstructive Sleep Apnea
PEn	Permutation Entropy
P-SmvMDE	Proportional SmvMDE
ReR	respiration rate
RESP	respiratory
RI	respiratory impedance
RR	heart rate variability
SampEn	Sample Entropy
ShEn	Shannon Entropy
SkipDisEn	Skip Sample Dispersion Entropy
SmvMDE	Stratified Multivariate Multiscale Dispersion Entropy
ST-SmvMDE	Soft Threshold SmvMDE
TDS	Time Delay Stability
T-SmvMDE	Threshold SmvMDE
VCO2	carbon dioxide production

VE	pulmonary ventilation
VO2	oxygen consumption

Symbols

Table 1: Symbols used in the Thesis

Symbol	Description
m	Embedding Dimension
c	Number of Classes
d	Time Delay
τ	Scale Factor
t	Threshold Parameter
w	Reduced Weight Parameter
μ	Mean
σ	Standard Deviation
P	Artifact Percentage Factor Value
G	Artifact Grouping Factor Value

Chapter 1

Introduction

1.1 Motivation for this Research

With the advancement of physiological recording technology deployed across a broad spectrum of applications, from wearable devices to intensive care units, increased amounts of data are becoming available for analysis. Their appropriate utilization could lead to effective prognosis, early stage intervention, personalised treatments, and improved clinical decision making [1–8].

For this purpose, the development of respective algorithms requires certain characteristics of the physiological time-series to be considered. These consist of the potential non-linear nature of their dynamics [9–14], their multivariate nature due to the interaction of multiple organ systems in human physiology [15–19], and the low data-quality arising from typical recording conditions [20–22].

Furthermore, the utilization of such algorithms could aid in addressing dangerous phenomena such as that of "alarm fatigue" observed in clinical environments [23, 24]. Algorithms currently deployed in intensive care units display excessive amounts of false positive alarms that are disruptive to the operation of clinical staff and can lead them to ignore alarms that are perceived as false even when they are accurate, leading to potential errors. [2, 25].

The variability[1] of recorded physiological time-series has been highlighted as an important feature that can provide clinically viable insights with regards to the physiological state of the monitored individual. Changes in the variability of the recorded physiological time-series have been connected with transitions in the state of the monitored system. Based on the Critical Slow Down (CSD) paradigm [26–28] certain variables follow a homeostatic behavior with low levels of variability being prevalent during healthy physiological states and increased levels of variability observed when the transition to a pathological state has occurred, indicating a collapse of the regulatory mechanisms that were meant to nullify perturbations. A second

1. A detailed description of the context within which the term variability is used in this Thesis is provided in Section 2.1.3.

group of variables follows non-homeostatic behavior as described from the Loss of Complexity (LoC) [9, 10, 29, 30] paradigm, with increased levels of variability observed during healthy states and low levels of variability during pathological states indicating a collapse of underlying complex mechanisms and adaptive capacity.

Entropy quantification algorithms have been extensively used for the measurement of variability in time-series [31], including the analysis of physiological recordings. These algorithms can be broadly characterised into those based on Shannon Entropy (ShEn) [32] or on Conditional Entropy (CEn), which is defined as the quantity of information observed in a sample at a time-point n that cannot be explained based on previous samples up to time point $n-1$ [33]. Examples of such algorithms include Permutation Entropy (PEn) [34] and DisEn [35,36] based on ShEn, and Approximate Entropy (ApEn) [37] and Sample Entropy (SampEn) [38] based on CEn.

This thesis utilizes the univariate Dispersion Entropy (DisEn) [35,36] and Multivariate Multiscale Dispersion Entropy (mvMDE) [39] algorithms, first introduced in 2016 and in 2019, respectively, to study how the characteristics of physiological time-series and their respective recordings influence algorithmic design and the implementation during different forms of analysis. The selection of the DisEn algorithms is made due to their capability of producing features with significant discrimination capacity taking into consideration amplitude-based information while maintaining a computational complexity that scales linearly with the size of the input data, and having a functional multivariate variation capable of quantifying cross-channel dynamics [39].

1.2 Aims and Objectives

The presented research has two main aims. The first aim is to investigate the current capacity of DisEn algorithms to conduct effective physiological time-series analysis when faced with the challenges of low data quality and implementation of multivariate analysis within the context of human physiology. The second aim is to expand on the initial findings and introduce novel variations of DisEn algorithms and frameworks to address these challenges and expand the research tools available for analysis.

To achieve these aims, the following three objectives are defined:

- To analyse the robustness of univariate DisEn to artifactual samples and develop variations that mitigate the disruptive capacity of artifacts.
- To extend the assessment of the impact of artifacts in the performance of mvMDE during the analysis of multi-channel time-series and to implement a process for the partial automation of separating between valid and artifactual segments of data.
- To expand multivariate analysis for the extraction of novel and informative features taking into consideration the LoC and CSD paradigms.

1.3 Contributions

Consequently, the main contributions of this Thesis are:

Univariate Dispersion Entropy Analysis for the Assessment of Missing and Outlier Samples and Development of Robust Variations - Chapter Three: Quantification of the effects of artifactual samples in the performance of univariate DisEn and the proposal of novel variations of DisEn for increased robustness during the analysis of time-series segments containing artifactual samples.

The research corresponding to this Chapter has been published in "*Entropy*" [40].

Multivariate Dispersion Entropy Analysis for the Assessment of Outliers and Detection of Artifactual Network Segments - Chapter Four: Quantification of the effects of artifactual outliers in the performance of mvMDE and the development of an artifactual network segment detection tool using a logistic regression classifier that utilizes univariate and multivariate DisEn features, for the partial automation of data cleaning.

The research corresponding to this Chapter has been published in "*Entropy*" [41].

The Stratified Entropy Framework and its Implementation using Dispersion Entropy - Chapter Five: Introduction of the Stratified Entropy framework that allows the extraction of novel features capable of measuring information that was previously inaccessible through traditional multivariate analysis. Three novel Stratified Multivariate Multiscale Dispersion Entropy (SmvMDE) algorithms are designed and tested as implementations of the framework.

The research corresponding to this Chapter has been published in "*IEEE Transactions on Biomedical Engineering*" [42].

1.4 Structure of the Thesis

The structure of the Thesis is as follows:

- Chapter 1 – Introduction: This Chapter introduced the motivation, objectives, and contributions of this Thesis.

- Chapter 2 – Background: The Chapter starts by introducing key characteristics of physiological systems, challenges arising by the quality of physiological recordings, and opportunities for the extraction of viable physiological information through the measurement of physiological time-series' variability. It continues with the description of entropy quantification algorithms as a prominent tool for measuring variability and reviews state of the art variations capable of multivariate and multiscale analysis. It concludes by discussing the framework of Network Physiology and novel forms of analysis that have been developed since its initial introduction.

- Chapter 3 – Univariate Dispersion Entropy Analysis for the Assessment of Missing and Outlier Samples and Development of Robust Variations: The Chapter presents the work conducted using the univariate DisEn algorithm. Missing and outlier samples are simulated across different physiological time-series to study the effect of artifactual samples in the operation of univariate DisEn. Novel DisEn variations are introduced for improved robustness to artifactual samples and their performance is benchmarked to that of the original algorithm.

- Chapter 4 – Multivariate Dispersion Entropy Analysis for the Assessment of Outliers and Detection of Artifactual Network Segments: The Chapter presents the experiments conducted using mvMDE on networks of multi-channel time-series formulated from synchronised physiological recordings. Artifactual outliers are simulated on all channels of the network and their disruptive capacity in the output features is quantified and differences based on the channel containing the artifacts are analysed. Finally, the utilization of both univariate and multivariate DisEn features for the detection of artifactual network segments is studied through the deployment of a logistic regression classifier.

- Chapter 5 – The Stratified Entropy Framework and its Implementation using Dispersion Entropy: The Chapter introduces the framework of Stratified Entropy and presents three algorithmic variations of the Stratified mvMDE algorithm as implementations of the framework. The operation of the novel algorithms is compared to that of mvMDE and is tested and analysed using multi-channel synthetic time-series, waveform physiological time-series and derived physiological data.

- Chapter 6 – Summary, Limitations of Study and Future Work: This Chapter summarises the presented work, discusses limitations of the study and highlights opportunities for future work and concludes the Thesis.

Chapter 2

Background

2.1 Characteristics of Physiological Systems and Time-Series

This section discusses topics that have to be considered during the analysis of physiological time-series. These topics consist of:

1. Characteristics arising from the nature of physiological systems.
2. Data quality issues arising due to the recording conditions.
3. The variability of physiological time-series and how increases or reductions of variability are associated with changes in the state of the monitored physiological system.

2.1.1 The Dynamic, Nonlinear, and Multivariate Nature of Physiological Systems

Physiological systems have been abstracted as dynamic systems that can be described by two core elements, their state and their dynamics. The state of the system at a particular point in time can be determined by the values of the variables that represent it [43, 44]. Consequently, when a system is represented by V variables, its current state can be represented as a point in an V-dimensional space [45] that is termed as the state or phase space of a dynamic system. The dynamics of a system constitute the laws that describe how the state of the system changes over time [46].

With regards to the dynamics that describe physiological systems, it is important to consider that a significant number of them are characterized by non-linearity [13, 45, 47–49]. This implies that the homogeneity and additivity properties, which are core properties of linear systems, are not preserved. Therefore, the outputs of physiological systems and their state transitions cannot be predicted through the simple linear combination of interactions of their internal components. Instead, due to non-linear effects such as feedback and/or multiplicative effects between their components, or with components of external systems, state transitions become significantly unpredictable. In some cases small changes in values of the state variables can lead to rapid state transitions while in other cases large input deviations can have negligible state changes when saturation effects are observed [46, 49].

Finally, physiological systems do not exist independently but instead are in constant and intricate interactions with other systems in the organism they formulate [15, 17] in what could potentially be described as a system of systems. Consequently, the analysis of physiological systems is a multivariate analysis process that would benefit from the consideration of interactions both between internal components of the same system but also between components of distinct systems [18, 19, 50]. This multivariate nature becomes even more crucial when also taking into account the potential non-linear dynamics of these systems which indicates that state transitions can occur from small changes in both their internal or external interactions [51–53].

Consequently, the analysis of data retrieved from physiological sensors requires the careful consideration of the non-linear and multivariate nature of the dynamics of the monitored systems.

2.1.2 Physiological Recordings and Data Quality

The amount of physiological time-series data available for analysis has significantly increased through the advancement and deployment of physiological recording technologies ranging from wearable devices to novel monitoring systems deployed in clinical environments. The extraction of viable physiological information from such data could lead to effective prognosis, early stage intervention, personalized treatments, refined clinical study design, and improved clinical decision making [1–8].

However, physiological time-series contain significant amounts of artifactual samples. These artifacts can be generated from physiological events such as in the case of artifacts generated in EEG signals due to eye movement [54], recording events such as loose equipment attachment, user movement and electromagnetic interference [20–22] and algorithmic errors occurring during the preprocessing of data such as in the formulation of a RR time-series from an input ECG signal [55].

Therefore, the analysis of such datasets should carefully consider the existence of artifactual samples and have processes in place to mitigate their disruptive capacity. This can be achieved either through the modification of feature extraction algorithms to increase their robustness to artifactual samples [56], the appropriate deployment of data cleaning and quality control preprocessing procedures [57, 58] or the utilization of end-to-end machine learning pipelines that are robust to artifacts [59]. Within the scope of this Thesis two types of artifactual samples are considered in alignment with prior research, artifactual missing samples [60] and artifactual outlier samples [61]. Missing samples are considered any samples whose value is invalid either due to a fault during the recording of the respective signal or from their removal through data preprocessing steps. Artifactual outliers are considered samples whose values have been significantly altered from their original values.

Failure to address low data quality will limit the capacity of algorithmic implementations to extract useful information when deployed in physiological monitoring applications. From an effectiveness standpoint, the performance of disease prognosis and clinical support algorithms can be severely limited when trained on low quality datasets [62,63]. Notably, inaccuracy in the extracted features such as in the case of SpO_2 sensors deployed in intensive care units [25], can lead to the already threatening phenomenon of "alarm fatigue" [2, 23, 24] described in Section 1.1.

The combination of the non-linear and multivariate dynamics arising from the nature of the monitored physiological systems and the low data quality arising from the nature of physiological recording conditions formulate a combined challenge that calls for appropriate algorithms to be used in the analysis of such time-series.

2.1.3 Complexity, Irregularity and Variability

The terms complexity, irregularity and variability are commonly used within the context of physiological time-series analysis. In many occasions they have been used in a manner that could be considered interchangeable [64–66] and can lead to confusion particularly when considering that the use of these terms can change across different publications of academic research.

Within the scope of this Thesis and after careful reviewing of the literature presented in this Chapter the terms are used in the following manner:

- The term complexity of a physiological signal is used to describe changes in the value of the signal that arise from consistent structural dynamics of the monitored physiological system [67,68]. When measured it is expected to follow a stable profile of values across multiple temporal scales.
- The term irregularity is used to describe random changes in the signal that do not arise from underlying dynamics in the monitored system [39, 69]. When measured it is expected to follow a declining profile of values across increasing temporal scales.
- Finally the term variability is the main term used in this Thesis to describe changes in the values of physiological time-series. The term variability was selected as a less restrictive term [29, 70] that will be used to describe in general changes of interest in physiological time-series. It should be noted that within this scope, the term variability is used to describe the amplitude fluctuations observed across a set of samples as typically taken in entropy quantification algorithms and is not used in association with second order moments of a random variable such as that of variance.

2.1.4 Variability of Physiological Time-Series

For the effective extraction of clinically viable information from recorded physiological time-series, the quantification of their variability has arisen as a prominent feature that can provide significant insights with regards to the physiological state of the monitored individual [9,29,70]. This has occurred because changes in the variability of recorded time-series have been connected with transitions in the state of the monitored physiological system. The two primary paradigms describing this connection are the CSD [26–28] and the LoC [9,10,29,30] paradigms.

Based on the CSD paradigm, during healthy physiological states, certain variables follow a homeostatic pattern. They are characterized by low variability and are maintained at a physiologically beneficial setpoint with any perturbations being corrected via appropriate recovery mechanisms [27, 71]. A slowing down of their recovery rate is an indication that the monitored individual is reaching a breaking point after which they have transitioned into a pathological state with the regulated variables eventually losing their capacity to recover and therefore being characterized by increased levels of variability [26, 27, 71].

The LoC paradigm highlights that for some variables increased variability across multiple temporal scales is observed during healthy physiological states [70]. These variables follow a non-homeostatic pattern, their increased variability is indicative of complex underlying dynamics of the monitored system and is reflective of the capacity of the system to adapt to a variety of external stressors [9, 10, 29, 30]. Consequently decreases in the variability of such variables indicate the transition from a healthy to a pathological state with a reduction of the individual's adaptive capacity [9, 70].

A recent hypothesis that aims to combine these two paradigms is the "Entropy Pump" (EP) hypothesis that separates physiological variables in two categories. Regulated variables that display homeostatic behavior in alignment with the CSD paradigm and effector variables that display non-homeostatic behavior following the LoC paradigm [19]. Based on this hypothesis, an entropy pump is observed thanks to which homeostasis is achieved by maintaining a stable, low complexity output for regulated variables through the complex and variable outputs of effector variables that counteract any perturbations affecting the regulated variables through negative feedback loops [19, 51, 72]. Therefore, the transition to a pathological state would be indicated by the collapse of the directionality of the pump and a consequent increase of the regulated variables' variability in alignment with the CSD paradigm and a reduction of the effector variables' variability in alignment with the LoC paradigm. As an example of this directionality during healthy physiological states and its collapse, the respective research indicates the case of the systolic blood pressure maintained at a steady setpoint through increased heart rate variability in healthy individuals as opposed to increased systolic blood pressure variability and reduced heart rate variability recorded for individuals with type-2 diabetes [19].

A tool that has been increasingly applied for the measurement of variability in physiological time-series is entropy quantification algorithms that are discussed in detail in Sections 2.2 - 2.5.

2.2 Entropy as a Measure of Variability

This section discusses the mathematical concepts that have led to the utilization of entropy for the analysis of time-series and the measurement of their variability. The core concepts of ShEn and CEn are discussed in detail alongside the processes of binning estimates and the formulation of embedded vectors that have aided in the implementation of algorithms for their estimation.

2.2.1 Shannon Entropy

Shannon's initial definition of Entropy within the context of Information Theory was made with the aim of quantifying the amount of information contained within a signal [31, 32]. The mathematical definition of ShEn defined the information of a signal as the amount of unexpected data contained in a message (time-series segment). Consequently, for a time-series $X = \{x_i, i = 1, \ldots, N\}$ and the probability of each $x_i, i = 1, \ldots, N$ denoted as p_i with $p_i > 0$ and $\sum_{i=1}^{N} p_i = 1$, ShEn is defined as:

$$ShEn(X) = -\sum_{i=1}^{N} p_i \cdot \log(p_i). \tag{2.1}$$

It is important to consider that, the range of i plays an important role in the applicability of an algorithm that utilizes ShEn. When a time-series is extracted from a system whose samples can have a value retrieved from a limited set of potential values and consequently i being a relatively small number, the direct application of ShEn is possible as long as the probability of each value p_i can be calculated for the analyzed time-series segment.

However, within the context of a time-series where each sample can have a potentially unique value, or a value with a significantly low probability of repetition, the direct application of ShEn would lead to values that are consistently close to its maximum value and are unable to track potential changes in the state of the system based on its output entropy values. The analysis of physiological time-series is an important example of this phenomenon. In this case, the spectrum of potential sample values is large both due to the complexity of the system that generates the time-series but also due to the noise present in the recording

process itself. Consequently, during such analysis, the ShEn features extracted from the sample values would not be capable of providing meaningful insights with regards to the state of the monitored system. Since almost every sample value will be considered unique leading to a strong bias of increased ShEn.

2.2.2 Binning Estimates

For this reason, estimation approaches have been defined for the quantification of entropy measures during time-series analysis with one of the most widely deployed methods being the binning estimator. Based on this approach approach a time-series X that contains values in a range of $R_X = [X_{min}, X_{max}]$ undergoes a quantization process that results in the formulation of a time-series X^q that consists of samples that belong in the alphabet $A_X = \{1, \ldots, Q\}$. Consequently each sample $x_i^q, i = 1, \ldots, Q$ is assigned to one of the Q quantization levels (or "bins"). The bins can then be utilized in a variety of different approaches for the estimation of an entropy metric. The most direct approach is the following: for each respective bin, its probability of occurrence is calculated and then the ShEn of the input time-series can be estimated using the probabilities of each bin ($A_X = \{1, \ldots, Q\}$) as opposed to those of each unique sample:

$$ShEn(X) = -\sum_{A_X=1}^{Q} p_{A_X} \cdot \log(p_{A_X}) \tag{2.2}$$

When considering this process, an important question that has to be answered with regards to its applicability is how the bins are defined so that they can provide an accurate estimation of the time-series dynamics. A straightforward and simple approach is the even spread of the Q bins in the amplitude range of the the time-series, resulting in bins with equal amplitude range: $r = (X_{max} - X_{min}/Q)$ [73]. An alternative approach which is utilized when the bins are not directly used for the calculation of the output entropy value and are instead an intermediate step, is the implementation of a variable bin size to ensure an equal number of samples been allocated to each bin [74, 75].

Sections 2.4.1 - 2.4.2 will discuss in detail the PEn [34, 76] and DisEn [35, 36, 39] algorithms that follow a more advanced approach to binning estimation based on symbolic dynamics, with the DisEn algorithm formulating the algorithmic foundation of the work presented in this Thesis.

2.2.3 Embedded Vectors

An additional tool that is extensively utilized in entropy quantification algorithms is the formulation of embedded vectors. For a time-series $X = \{x_i, i = 1,\ldots,N\}$, groups of samples (x_i) can be extracted for the formulation of m-dimensional embedded vectors following a sliding window step of 1 sample at a time and a time delay d between picked samples, following Takens' embedding theorem [77], resulting in: $x_i = \{x_i, x_{i+d}, \ldots, x_{i+(m-1)d}\}$, for each $i = 1,2,...,N-(m-1)d$.

For example, assuming $X = \{1,2,3,4,5,6,7,8\}$, $m = 2$ and $d = 1$ the embedded vectors that would be formulated are: $\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{5,6\},\{6,7\},\{7,8\}$.

The formulation of embedded vectors is a fundamental process that is commonly combined with the formulation of bins, for the implementation of algorithms based on symbolic dynamics such as PEn and DisEn as well as the estimation of CEn through the utilization of algorithms such as ApEn and SampEn.

2.2.4 Conditional Entropy

Conditional Entropy (CEn) is defined as a the quantity of information observed in a sample at a time-point n that cannot be explained based on previous samples up to time point $n-1$ [33,78]. For a time-series $X = \{x_i, i = 1,\ldots,N\}$ and a second time-series $Y = \{x_i, i = 1,\ldots,N+1\}$ which is a copy of X up to sample $i = N$, a mathematical definition of CEn is given as:

$$CEn(Y|X) = -\sum_{x\varepsilon X, y\varepsilon Y} p(x,y)\cdot \log\frac{p(x,y)}{p(x)} \tag{2.3}$$

If the definition of CEn is applied directly to a time-series, the same challenges arise as discussed for ShEn since minor fluctuations in the amplitude of the input signal would lead to disproportionately high estimations of novel information. However, algorithms such as ApEn and SampEn utilize algorithmic steps that combine the binning estimation process with the formulation of embedded vectors to implement a computational approach capable of estimating CEn with improved accuracy.

2.3 Approximate and Sample Entropy Algorithms

2.3.1 Approximate Entropy

Introduced in 1991 [37], the ApEn algorithm utilizes a two step process for the computational estimation of an input time-series CEn.

As a first step, embedded vectors of dimension m are formulated and a similarity distance between the vectors is calculated to define whether the vectors are considered equivalent based on a predefined tolerance r. The same process is then repeated for embedded vectors of dimension $m+1$ allowing the calculation of two similarity measures among embedded vectors, one between m samples and a second between $m+1$ samples which is in alignment with the definition of CEn. With the two similarity matrices for m and $m+1$ vectors, the second step is implemented, as it is now possible to compare the difference in the amount of information available when accessing samples up to time points $n-1$ (represented by embedded dimension m) and n (represented by embedded dimension $m+1$).

Consequently, for a time-series X, embedding vectors x_i^m with $(1 \leq i \leq N-m+1)$ are formulated, and the probability $C_i^m(r)$ of finding any other vector whose similarity distance is $\leq r$ is calculated as follows [37, 79]:

$$C_i^m(r) = \frac{\text{\# of } x_j^m \text{ such that } d[x_i^m, x_j^m] \leq r}{N-m+1}, (1 \leq j \leq N-m+1) \tag{2.4}$$

where r is a value defined by the user of the algorithm and usually corresponds to a percentage of the standard deviation (σ) of the time-series in the range of 10% to 25%. Following the two step process, $m+1$ embedded vectors are also formulated and the same process is applied for the calculation of the respective metric:

$$C_i^{m+1}(r) = \frac{\text{\# of } x_j^{m+1} \text{ such that } d[x_i^{m+1}, x_j^{m+1}] \leq r}{N-m}, (1 \leq j \leq N-m) \tag{2.5}$$

With $C_i^m(r)$ and $C_i^{m+1}(r)$ calculated, an additional metric $\Phi^m(r)$, $\Phi^{m+1}(r)$ is defined as;

$$\Phi^m(r) = \frac{1}{(N-m+1)} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{2.6}$$

$$\Phi^{m+1}(r) = \frac{1}{(N-m)} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r) \tag{2.7}$$

This metric corresponds to the average logarithmic probability of finding any matching m-dimensional embedded vectors in the time-series X and operates as a summary metric of vector similarity for the m-dimensional case. With the metrics for both groups, one with regards to similarity observed during the accessing of m $(n-1)$ and $m+1$ (n) samples, CEn can now be estimated through the calculation of ApEn using the following equation:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{2.8}$$

The equation for the calculation of ApEn allows the derivation of a metric that describes the amount of novel information retrieved when increasing the samples accessed from the input time-series X. For example if the m dimensional vectors remain similar in the $m+1$ case then this leads to the conclusion that no increase in novel information is observed through the inclusion of the additional sample in the vectors. Consequently the calculated ApEn for this case will be zero. In the opposite case, where all the similar m dimensional vectors become dissimilar in the $m+1$ case it means that the new information introduced by the additional sample cannot be predicted by the previous samples in each vector leading to a maximum value of ApEn.

While ApEn managed to introduce an operational estimation of the CEn concept, there is an important bias that has to be considered [38]. The calculation of the $C_i^m(r)$ and $C_i^{m+1}(r)$ metrics takes into consideration self-matches, matches occurring between the same embedded vector in the special case of $i = j$. Consequently, ApEn is biased towards the calculation of a higher degree of similarity than what is actually present in the time-series X. Furthermore, this bias is size dependent, with shorter time-series facing a larger bias of similarity and therefore a lower than expected calculated value of ApEn, leading to inconsistency when trying to compare the ApEn values of time-series with different sample length N. The bias exists in the algorithm because if the self-matching cases of the vector are omitted than there is a chance that the natural logarithm in equations 2.6-2.7 will have a $\ln(0)$ value which is undefined and would therefore result in a computational error. To address this limitation, the SampEn algorithm was introduced which would later supersede ApEn.

Despite its limitations, ApEn found extensive use for the analysis of physiological time-series in applications such as the investigation of abnormalities in respiratory function caused by panic disorders [80] and the analysis of EEG signals to measure the effectiveness of anesthesia drugs [81], to estimate sleep stages [82] and to detect changes in the EEG's variability that could be associated with Alzheimer's disease [83].

2.3.2 Sample Entropy

The SampEn algorithm modifies the step for the calculation of $C_i^m(r)$ and $C_i^{m+1}(r)$ metrics [38], for separation purposes the respective variables in SampEn will be signified as $U_i^m(r)$ and $U_i^{m+1}(r)$ respectively. The equations for their calculation are the following:

$$U_i^m(r) = \frac{\text{\# of } x_j^m \text{ such that } d[x_i^m, x_j^m] \leq r}{N-m-1}, (1 \leq i \leq N-m, j \neq i) \tag{2.9}$$

$$U_i^{m+1}(r) = \frac{\text{\# of } x_j^{m+1} \text{ such that } d[x_i^{m+1}, x_j^{m+1}] \leq r}{N-m-1}, (1 \leq i \leq N-m, j \neq i) \tag{2.10}$$

When comparing the ApEn equations 2.4-2.5 to the respective SampEn 2.9-2.10 three changes can be noted. The first is the rule that the cases where $i = j$ are discounted from the calculation of vector similarity to remove the bias present in ApEn. The second change is the modification of the range for the values of j to be the same for both equations to ensure that the same number of vectors are evaluated during the calculations of U for both m and $m+1$ cases. The third change, which is a direct result of the previous two, is that the denominator for both cases has been updated to $N-m-1$, with the $N-m$ been determined by the common j range and the additional -1 factor added to account for the $j \neq i$ restriction.

Furthermore, instead of the Φ metrics of ApEn, SampEn calculates the summary values $U^m(r)$ and $U^{m+1}(r)$ as follows:

$$U^m(r) = \frac{1}{(N-m)} \sum_{i=1}^{N-m} U_i^m(r) \tag{2.11}$$

$$U^{m+1}(r) = \frac{1}{(N-m)} \sum_{i=1}^{N-m} U_i^{m+1}(r) \tag{2.12}$$

The $\ln(\cdot)$ function within the summation of the ApEn equations 2.6-2.7 is removed in the respective SampEn equations 2.11-2.12 to avoid the potential $\ln(0)$ value error and instead is utilized during the calculation of the SampEn value itself as follows:

$$SampEn(n,r,N) = -\ln \frac{U^{m+1}(r)}{U^m(r)} \tag{2.13}$$

With these modifications SampEn is capable of following a similar pattern as ApEn with regards to measuring the difference in matching vectors when moving from m to $m+1$ dimensionality while not being affected by ApEn's bias resulting in the effective replacement of the algorithm for the estimation of CEn.

SampEn has been extensively utilized in physiological time-series analysis. Examples of its application include the analysis of neonatal heart rate variability for the detection of sepsis [84],the analysis of center of pressure time-series data for the detection of postural differences when walking with open versus closed eyes [85], the analysis of EEG signals for the detection of epileptic seizures [86], and the analysis fMRI data for the detection of age related changes in dynamic functional connectivity [87].

2.3.3 Implementation considerations for ApEn and SampEn

Parameter Selection

Both the ApEn and the SampEn algorithms operate utilizing the same parameters of r that corresponds to the percentage of statistical deviation utilized as the tolerance threshold for the classification of two vectors as similar and m which defines the number of samples contained in each embedded vector.

The selection of the value of r is a balancing act and depends on the temporal characteristics of the dynamics included in the analyzed time-series. A lower than viable value of r would lead to increased sensitivity in fluctuations caused by noise and therefore would incorrectly label vectors as being different. Contrariwise, an overly increased value of r would significantly reduce the sensitivity of detecting differences among vectors and would therefore incorrectly label them as similar when they are not [37, 79]. As stated in Section 2.3.1 the recommended range for r is between 10% to 25% of the time-series σ for both ApEn and SampEn. This is an empirically derived range recommended in the original publication [37] that introduced the ApEn algorithm in 1991 and was further reinforced in 2000 based on experiments conducted in the study that introduced SampEn [38]. It should be noted that this range corresponds to time-series with relatively slow dynamics such as measurements of hormonal release and would therefore require adjusting when applied to other signals with faster dynamics such as EEG [88].

The recommendations for the selection of value for the m parameter are also empirical. As a rule of thumb for both ApEn and SampEn it is recommended that the value of m should be selected based on the restriction posed by the length of the analyzed time-series N. It is stated that the condition $N > 10^m$ should be treated as a harsh minimum with a recommended range of 20^m to 30^m [38, 79]. Consequently, common values utilized for m are those of 2 and 3.

Computational Complexity

The computational complexity of both the ApEn and SampEn algorithms is quadratic $O(N^2)$ with regards to the input time-series length. This occurs due to the operational requirement of both algorithms to calculate the distances between all vectors with the same dimension (m and $m+1$) [37, 38]. Furthermore an additional factor that can affect computational time is that of m since it affects the equation used to calculate the distances between two vectors with higher values for m leading to more complex distance equations. The quadratic scaling, is a significant limitation that posed an additional motivation for the development of novel entropy quantification algorithms that could achieve lower computational times such as in the case of PEn and DisEn algorithms which aim to estimate ShEn.

2.4 Permutation and Dispersion Entropy Algorithms

This section describes the PEn and DisEn algorithms that combine the utilization of binning estimates and embedded vectors for the estimation of ShEn in input time-series. Both algorithms utilize a similar approach of matching the embedded vectors to symbolic patterns that are treated as separate bins whose relative frequency of occurrence is calculated. However, the algorithms differ to the extend that the PEn algorithm disregards information from the amplitude value of the samples while the DisEn algorithm contains additional steps to include that information in the formulation of respective bins. At this point it is important to clarify that the version of the DisEn algorithm discussed in this Thesis is the version first described in the original publication of 2016 [35] and the one referred to as "Amplitude Based Dispersion Entropy" algorithm in the follow up publication of 2018 [36]. In the 2018 publication there is an additional variation titled "Fluctuation Based Dispersion Entropy" whose functionality is closer to the PEn algorithm.

2.4.1 Permutation Entropy

PEn is based on ShEn and utilizes the formulation of embedded vectors, with each vector being mapped to an ordinal pattern. Each unique ordinal pattern is treated as a bin whose relative frequency is used to calculate the output PEn value. The algorithm is designed with the aim of having low temporal complexity and relative robustness to noise for the extraction of viable features describing the variability of the input time-series segment.

For an input univariate time-series $x_j(j = 1,2,\ldots,N)$ of length N the steps followed by the PEn algorithm are as follows:

1. Formulation of embedded vectors: An embedding dimension (m) and a time delay (d) are set for the creation of embedded vectors, $x_i = \{x_i, x_{i+d}, \ldots, x_{i+(m-1)d}\}$, of length m, for each $i = 1,2,...,N-(m-1)d$.
2. Embedded vector reordering: Withing each embedded vector, each sample is given a positioning index in the range of: $0 \ldots m-1$, the samples are then arranged in increasing amplitude order. The newly ordered index positions correspond to the ordinal pattern of that embedded vector. There are $m!$ unique ordinal patterns based on the embedding dimension used. The number of times a particular ordinal pattern arises during the reordering of an embedded vector is recorded.
3. Calculation of Ordinal Pattern Relative Frequency: For each of the $m!$ unique ordinal patterns, their relative frequency (p) is calculated based on the number of instances that an embedded vector has being assigned to that ordinal pattern divided by the total number of embedded vectors, as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{i \,|\, i \leq N-(m-1)d, x_i^m \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)} \tag{2.14}$$

4. Calculation of Univariate Permutation Entropy: Utilizing the relative frequencies of each ordinal pattern the time-series' output PEn is calculated based on Shannon's definition of Entropy:

$$PEn(\mathbf{X}, m, d) = -\sum_{\pi=1}^{m!} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right). \tag{2.15}$$

Consequently, an input time-series that could be described by a singular ordinal pattern would result in the minimum output PEn value (i.e., 0) as opposed to one requiring the utilization of all possible ordinal patterns in equal probability which would result in a maximum output value.

It is important to consider the case where two samples within an embedded vector share the same value. In that case, the designers of the algorithm [34] suggested ranking the equal samples based on their order of emergence with the first equal sample taking an ordinal value that is one level lower than the second. The addition of noise was also suggested as a potential solution however that could lead to in-precise patterns. Furthermore, due to the utilization of ordinal patterns, amplitude based information is disregarded. Such information could correspond to the magnitude of the difference between samples within an embedded vector. An embedded vector with samples values of $\{100, 150, 200\}$ and a vector with values $\{100, 101, 102\}$ would both be mapped to an ordinal pattern of $\{1, 2, 3\}$ despite the differences between the samples of the first vector being larger than the respective differences of the second. Information with regards to the mean (μ) value of samples in an embedded vector is also disregarded since an embedded vector with sample values of $\{30, 20, 10\}$ and one with values of $\{1030, 1020, 1010\}$ would both be mapped to an ordinal pattern of $\{3, 2, 1\}$ despite the significant difference of their μ values. Finally, in the case of significantly small difference between sample values, PEn becomes susceptible to noise since small amplitude fluctuations can lead to the mapping of an incorrect ordinal pattern.

Recent research has produced algorithmic variations for PEn that aim to utilize amplitude based information to partially address these limitations. The Weighted PEn algorithm introduced in 2013 [89] utilizes the variance of the samples in each embedded vector to define a weight coefficient with regards to its contribution in the calculation of the relative frequency for its mapped ordinal pattern, while the Amplitude Aware PEn presented in 2016 [90] utilizes information with regards to the μ value of the samples included in each embedded vector and the amplitude difference between consecutive samples for the definition of a respective coefficient. In addition to these variations, the utilization of amplitude information is also addressed in a different manner in the DisEn algorithm discussed in Section 2.4.2.

PEn has found extensive utilization in the analysis of physiological time-series in applications such as the analysis of EEG signals to measure the effectiveness anesthesia [91], the detection of absence seizures [92] and the characterization of sleep stages [93]. Additionally heart rate variability time-series have been analysed for the detection of cardiovascular autonomic neuropathy in patients with type 1 diabetes mellitus [94] and the detection of individuals that had cardiodepressive vasovagal syncope during a head-up tilt test procedure [95]. Finally, PEn has also been utilized in the study of gene expression time-series for the detection of temporal gene expression profiles [96].

2.4.2 Dispersion Entropy

The DisEn algorithm, similarly to PEn, is also based on ShEn and utilizes embedded vectors that are mapped to unique dispersion patterns which are then utilized as bins for the calculation of relative frequencies. Unlike PEn, the DisEn algorithm is capable of utilizing amplitude based information during the mapping of each embedded vector to one of the unique dispersion patterns [35, 36].

Prior to the application of DisEn, an optional but recommended preprocessing step is the application of a mapping function to the input time-series. In the 2018 study a number of mapping functions were tested, with the recommendation being the utilization of a non-linear mapping function such as the Logarithm Sigmoid Function (LogSig) [36]. The selection of a non-linear over a linear mapping function seeks to ensure that maximum and minimum amplitude values, that can be significantly larger or smaller than the μ value of the channel, do not disrupt the allocation of samples to classes by forcing the majority of samples to be assigned to a small number of classes [36]. It also allows the formulation of an amplitude range that would facilitate the effective allocation of classes as described in the first of the following algorithmic steps.

The process followed by the algorithm for the analysis of either the original or the mapped input univariate time-series $x_j(j = 1,2,\ldots,N)$ of length N is the following:

1. Production of a "quantised" time-series: A number of classes (c) are distributed along the amplitude range of the time-series, and each sample is allocated to the nearest respective class based on its amplitude. This results in the production of a "quantised" time-series $u_j(j = 1,2,\ldots,N)$.
2. Formulation of embedded vectors: An embedding dimension (m) and a time delay (d) are set for the creation of embedded vectors, $u_i^{m,c} = \{u_i^c, u_{i+d}^c, \ldots, u_{i+(m-1)d}^c\}$, of length m, for each $i = 1,2,...,N-(m-1)d$.
3. Mapping to dispersion patterns: Each embedded vector $u_i^{m,c}$ is mapped to a respective dispersion pattern $\pi_{v_0\ldots v_{m-1}}$ based on its corresponding classified samples. The number of potential unique dispersion patterns is c^m, as defined by the number of classes and the embedding dimension.

4. Calculation of Dispersion Pattern Relative Frequency: For each of the c^m unique dispersion patterns, their relative frequency is calculated as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{i \,\big|\, i \leq N-(m-1)d, u_i^{m,c} \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)} \tag{2.16}$$

5. Calculation of Univariate Dispersion Entropy: Utilizing the obtained relative frequencies, the time-series' output DisEn value is calculated using the following equation [35, 36], based on Shannon's definition of entropy:

$$DisEn(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right) \tag{2.17}$$

Following the aforementioned steps, an input time-series described by a single dispersion pattern would result in a minimum output DisEn value (i.e., 0) as opposed to one requiring the utilization of all possible dispersion patterns in equal probability, which would result in a maximum output value.

Examples of physiological time-series analysis using DisEn include the analysis of magnetoencephalogram (MEG) signals for the detection of variability reductions in the prevalence of Alzheimer's disease when compared to healthy individuals [97] and the analysis of ECG data for the detection of distribution differences in the values of DisEn when comparing between healthy heart beats, atrial premature beats and beats corresponding to premature ventricular contractions [98].

2.4.3 Implementation considerations for PEn and DisEn

Parameter Selection

Both the PEn and the DisEn algorithms utilize the parameter d which corresponds to the time-delay applied between the selection of samples utilized for the formulation of embedded vectors and m which corresponds to the length of each embedded vector (similarly to the ApEn and SampEn algorithms). In the case of the DisEn algorithm an additional parameter c is utilized corresponding to the number of classes that are spread throughout the amplitude range of the input signal after the optional application of a mapping function.

With regards to d the most commonly utilized value is that of 1 since usually all samples are utilized in the order of occurrence, however higher values can be used if the sampling rate of the input signal is considered significantly higher than the temporal profile of the target dynamics to be captured.

For DisEn the selection of value for c is a balancing act with reduced values of c leading to loss of signal information and increased values of c leading to reduced robustness to noise. For very low value of c, multiple samples will be allocated to the same class despite being representative of different dynamics. For example in the extreme and not sustainable case of $c = 1$ all embedded vectors will be mapped to the same dispersion pattern leading to a minimum value of DisEn = 0 regardless of the input time-series. For a significantly increased value of c even fluctuations caused by noise would lead to the formulation of different dispersion patterns, leading to biased high values of output DisEn. As an empirical recommendation, the range of $3 \leq c \leq 9$ is recommended as practical values for the majority of applications [35, 36].

Furthermore an additional consideration which affects the selection of values for both c and m is associated with number of input samples N. As discussed in the respective study [36], in order to acquire a representative set of relative frequencies for the formulated dispersion patterns, the number of input samples should follow the condition: $N > c^m$ with an empirical recommendation for trustworthy results being the condition that: $N > c^{m+1}$. Furthermore with regards m a minimum value of 2 is required to allow for each embedded vector to be formulated from at least two input samples.

For the PEn algorithm the selection of value for m is less restrictive since there is no c parameter as in the case of DisEn. As an empirical recommendation from the respective study [34] the value range of $3 \leq m \leq 7$ is recommended.

Computational Complexity

For both PEn and DisEn their computational complexity scales linearly with the number of input samples ($O(N)$). This occurs since each embedded vector is accessed only once to be mapped to an ordinal or dispersion pattern. This is an important advantage over the ApEn and SampEn algorithms whose scaling is quadratic ($O(N^2)$) and consequently their computational time can be prohibitively long for an array of applications.

2.5 Multivariate Multiscale Entropy Quantification Algorithms

This section describes in detail the Multivariate Multiscale Permutation Entropy (mvMPE) and the Multivariate Multiscale Dispersion Entropy (mvMDE) algorithms. Both of these algorithms follow similar approaches with regards to the coarse graining process used for multiscale analysis but have differences with regards to the utilization of samples from different channels of the input time-series for the extraction of multivariate features.

2.5.1 Multivariate and Multiscale Analysis

As discussed in Subsection 2.1.1, physiological systems display multivariate dynamics that occur due to interactions that are both internal, within components of the same system and external with components of other systems. Consequently, a comprehensive measurement of physiological variability through the use of entropy metrics would require the development of multivariate analysis techniques to allow the assessment of dynamics occurring across the channels of multi-channel time-series that are formulated from the combination of multiple recorded variables. For this purpose, recent research has produced algorithmic variations capable of multivariate analysis such as the mvMPE and the mvMDE algorithms presented in Sections 2.5.2 and 2.5.3 respectively.

The necessity for multivariate analysis due to the nature of physiological systems is complemented by the requirement for analysis across multiple temporal scales due to their dynamics, particularly when considering the multiscale complexity displayed by effector variables during healthy physiological states [9, 19, 68]. Consequently, for a complete assessment of physiological variability and appropriate utilization of multivariate data, when available, entropy quantification algorithms need to be capable of both multivariate and multiscale analysis.

Coarse Graining Process

For the successful quantification of a time-series' complexity across multiple time scales, it is common to utilize a coarse graining process to modify the input time-series based on the time scale factor τ. The mvMPE and the mvMDE algorithms utilize the moving average coarse graining approach as follows [67, 99, 100]:

Given a p-channel time-series $\mathbf{Y} = \{y_{k,b}\}_{k=1,2,\cdots,p}^{b=1,2,\cdots,N}$, each channel is processed separately and divided into non-overlapping segments of length equal to the defined τ. For each segment, an average value is calculated and used to derive the coarse-grained multi-channel time-series as follows:

$$x_{k,i}(\tau) = \frac{1}{\tau} \sum_{b=(i-1)\tau+1}^{i\tau} y_{k,b}, 1 \leq i \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N, 1 \leq k \leq p \tag{2.18}$$

where L is the original channel length and N the resulting coarse-grained channel length.

While this process is commonly used for multiscale analysis, it is important to consider that the variance of the time-series generated is reduced during the averaging of the original samples. Consequently, this reduction should be taken into consideration when comparing between the output multiscale profiles of two distinct time-series, with variance occurring from random fluctuations expected to follow strong pattern of reduction as the τ value increases [67] . Finally, the reduction of variance should also be taken into consideration during the design

of respective algorithmic steps such as in the case of mvMDE during the application of a mapping function, where the μ and σ of the original non coarse-grained time-series are used for the mapping across all τ values. This ensures that the mapping function applied to the time-series is not affected by the averaging process [39].

The original versions of both mvMPE and mvMDE utilize the moving average approach, they can however, be modified to utilize more complex processes such as the low-pass Butterworth filtering [100, 101], and empirical mode decomposition [101]. Finally, while coarse graining is commonly used for the implementation of multiscale analysis, novel and more complex methods have been introduced that can provide a more accurate multiscale profile of the input time-series at the cost of increased computational complexity. Examples include the composite multiscale analysis [102] and the refined composite multiscale analysis [103].

2.5.2 Multivariate Multiscale Permutation Entropy

The original Multivariate Multiscale Permutation Entropy (mvMPE) algorithm was introduced in 2012 [76] with the algorithmic operations that allow for multiscale analysis being introduced at the first algorithmic steps of mvMPE and the operations corresponding to multivariate analysis at the final steps.

The steps for the formulation of embedded vectors and their mapping to ordinal patterns, described in Section 2.4.1, are applied to each of the p channels separately following the same algorithmic operations. However the step that calculates the relative frequency of each ordinal pattern is modified. The number of cases in which an embedded vector has been mapped to a certain ordinal pattern are counted across all p-channels and consequently the relative frequency for each pattern is also divided by the number of channels p, as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{i \,|\, i \leq N-(m-1)d, x_i^m \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{p(N-(m-1)d)}. \tag{2.19}$$

The output mvMPE value is then calculated in the same manner as for univariate PEn but this time using the relative frequencies calculated based on all p-channels:

$$mvMPE(\mathbf{X}, m, d) = -\sum_{\pi=1}^{m!} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right). \tag{2.20}$$

An important point to highlight with regards to the presented implementation of mvMPE is the fact that while the relative frequencies of each unique ordinal pattern are calculated based on the mapping that occurs across all p-channels of the time-series, the embedded vectors that are mapped to the ordinal patterns are formulated from each channel independently.

Consequently no cross-channel dynamics are represented in the embedded vectors since there is no mixing of samples from different channels. Contrariwise, the mvMDE algorithm discussed in Section 2.5.3 formulates patterns that are based on cross-channel dynamics while the recently proposed graph-based mvMPE algorithm [104] utilizes a user defined cartesian product of two graphs to model relations between the different channels of the input time-series during the calculation of the output entropy value.

The mvMPE algorithm was initially utilized for the analysis of EEG signals for the detection of differences between healthy elderly individuals and individuals suffering from Mild Cognitive Impairment and Alzheimer's Disease [76].

2.5.3 Multivariate Multiscale Dispersion Entropy

Multivariate Multiscale Dispersion Entropy (mvMDE) allows the multivariate quantification of DisEn from multi-channel time-series. Similarly to its univariate equivalent, the preprocessing of each individual channel using a mapping function is recommended. For a set of p-channel time-series $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\cdots,p}^{i=1,2,\cdots,N}$ of length N each, the computational steps of mvMDE are the following [39]:

1. Production of univariate quantised time-series: A number of classes $(1,2,\ldots,c)$ are distributed along the amplitude range of each channel separately. Their samples are allocated to their nearest respective class based on their amplitude. As a result, a quantized channel $u_j (j = 1,2,\ldots,N)$ is produced for each respective input channel, resulting in a set of p-quantized channels $\mathbf{U} = \{u_{k,i}\}_{k=1,2,\cdots,p}^{i=1,2,\cdots,N}$.
2. Formulation of multivariate embedded vectors: From $\{u_{k,i}\}$, the quantized samples are embedded into univariate vectors of length m (with a time delay d) for each channel. The univariate embedded vectors are then combined in sets of p-synchronised vectors, one from each channel. The vectors within each synchronised set are serially concatenated for the production of a respective multivariate embedded vector $Z(j)$, of length $m \cdot p$, for each $j = 1,2,\ldots,N-(m-1)d$.
3. Mapping to multiple dispersion patterns: In mvMDE, each embedded vector is mapped to multiple dispersion patterns to effectively evaluate patterns both temporally within the same channel as well as across channels. Each subset of m elements in $Z(j)$ is accessed, following all possible $\binom{m \cdot p}{m}$ combinations. This formulates $\phi_q(j)(q = 1,\ldots\binom{m \cdot p}{m})$ embedded subvectors that are then mapped to their corresponding $\pi_{v_0 \ldots v_{m-1}}$ dispersion pattern. The total number of dispersion pattern instances is $(N-(m-1)d)\binom{m \cdot p}{m}$ and the number of unique dispersion patterns is c^m.
4. Calculation of Dispersion Pattern Relative Frequency: For each of the c^m unique dispersion patterns, their relative frequency is calculated as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{j \,\big|\, j \leq N-(m-1)d, \phi_q(j) \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)\binom{mp}{m}} \tag{2.21}$$

5. Calculation of Multivariate Dispersion Entropy: Utilizing the relative frequencies of the dispersion patterns considering both temporal and spatial domains as above, the output entropy value for X is calculated based on Shannon's definition of entropy:

$$mvMDE(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right) \tag{2.22}$$

The publication that introduced mvMDE in 2019 presented four different variations of the algorithm. The variation described in this section is the fourth variation which is also the one specified by its designers as the final and recommended variation. The mvMDE algorithm has been applied to 148-channel MEG data for distinguishing between healthy individuals and individuals suffering from Alzheimer's disease and to EEG data for the detection of variability differences between focal versus non-focal recordings [39].

2.6 Network Physiology

As it has been emphasized in this Chapter, the potential multivariate dynamics of physiological systems have been an important consideration of recent research aiming to provide a comprehensive analysis of their operation. A framework associated with a significant number of recent multivariate approaches is that of Network Physiology, first introduced in 2012 [105].

The framework of Network Physiology aims to describe the interactions occurring across diverse organ systems by representing them as a physiological network and extracting features from within that network. For this purpose two different processes are utilized, the formulation of a graph structure and the utilization of a respective feature extraction algorithm such as the Time Delay Stability (TDS) algorithm [18, 105].

2.6.1 Utilization of Graphs

Graphs are structures that contain a set of objects that are related to each other. The two core components that formulate a graph are nodes and edges. Each node of the graph is associated with an object while the edges of the graph connect the different nodes with each other based on the existence of interaction between the respective objects [106]. A core distinction between graph architectures is that graphs can contain either directed or undirected edges. In the case of an undirected graph [107], an edge described as one connecting two nodes, for example A to B, is the same edge as the one described to connect B to A. Contrariwise in directed graphs the direction of the edge matters therefore an edge connecting A to B would not be the same as one connecting B to A since they would be reverse with regards to their directionality [108].

In the initial implementation of Network Physiology undirected graphs were utilized due to the symmetrical nature of the TDS algorithm utilized for the extraction of features. A physiological network is constructed by representing each monitored physiological system as a node of the graph and projecting the interactions between systems as edges between the nodes [105, 109]. At this point, it is important to clarify that within the initial definition of the framework, graphs are not used with the aim of utilizing graph specific algorithms for the extraction of features but as a visualization and meta analysis tool for the tracking of topological changes and the matching of different graph topologies with physiological states of interest [18, 105]. The features used for the measurement of topological changes are instead extracted through the use of the TDS algorithm.

2.6.2 Time Delay Stability Algorithm

The TDS algorithm was developed with the aim of identifying and quantifying the pair-wise coupling and network interactions of diverse dynamical systems based on prior observations of coordinated bursting activity in the outputs of signals such as EEG, ECG, and respiratory rate [18, 110, 111]. For this purpose the algorithm aims to track the time delay with which bursts of activation in the output signal of a given system are followed by burst in the signals of other systems.

Periods of TDS with constant time delay between cross signal bursts indicate stable interactions and stronger coupling between the respective systems. Moving this into the graph representation of physiological networks during the analysis of multi-channel time-series with each channel corresponding to an input signal, the value given to a specific edge between two nodes, with each node representing a signal, is associated with the percentage of the time were TDS is observed [18, 105]. Consequently, the TDS algorithm is a bivariate feature extraction algorithm that measures the coupling stability between two input signals (A and B) based on the following steps [18, 105, 112]:

1. Both input signals (A and B) are divided into overlapping window segments, the length and overlap of each window is selected based on the target application and usually the overlap is 50% of the overall window length.
2. Each segment is normalized to a μ of 0 and unit σ to ensure that the coupling measured between the signals is not affected by their relative amplitudes.
3. The cross-correlation of each pair of synchronised segments from the two signals is calculated using periodic boundary conditions.
4. For each pair of windows the maximum absolute value of the cross-correlation function is used as an estimate of the time delay (γ) for that pair. The resulting γ values formulate a time-series with length equal to the total number of window pairs between the two signals.

5. The resulting time-series is then scanned in groups of 5 windows at a time. When the γ of the first window in each group remains relatively stable across the 4 consecutive windows with fluctuations that do not surpass the range of $[\gamma-1, \gamma+1]$ that group of windows is considered stable and the time period covered by that group is marked as a period of TDS. If any γ value in a group surpasses the range, then that group of windows remains unmarked.
6. The final TDS value is calculated as the percentage of points that have been designated as having TDS in the overall γ time-series.

When the TDS algorithm is implemented within the framework of Network Physiology, a TDS matrix is constructed that contains all the TDS values measuring the coupling strength associated with the edges that connect the different nodes in the constructed physiological network.

2.6.3 Implementations of the Framework

Network Physiology was first implemented for the detection of differences during different sleep stages such as deep, light, and rapid eye movement sleep through the analysis of EEG, ECG, and RESP signals [18, 105]. Furthermore, a recent study utilized the framework to study the coupling between synchronous bursts in cortical rhythms and peripheral muscle activation during states of sleep and wakefulness through the analysis of EEG and electromyographic (EMG) signals [112].

While the first implementations of Network Physiology were based on the utilization of graphs and the TDS algorithm, novel implementations with different forms of analysis have been introduced. For example the framework was extended to the analysis of thermal imaging data, with the aim of separating between subjects telling the truth versus lying in a simulated experiment, through the use of a modified version of the multivariate Granger causality algorithm instead of TDS [113]. Recently, Network Physiology was combined with the testing of the EP hypothesis for the detection of physiological disruptions caused by COVID-19 using three forms of analysis: principal component analysis, descriptive statistics, and network analysis [72].

Of particular interest is the utilization of entropy quantification algorithms within the framework of Network Physiology. Within this scope the Conditional Transfer Entropy algorithm was applied to EEG, ECG, RESP, and blood volume pulse (BVP) signals, for the detection of edges that displayed statistical significance when comparing across networks of individuals that were being monitored during states of rest and mental stress [114]. Furthermore, self entropy, mutual information, and conditional mutual information metrics were applied to time-series formulated from EEG, ECG, RESP, and BVP signals for the extraction of network based

features that were used for the classification of different levels of mental stress using support vector, random forest, and logistic regression classifiers in a comparative study [115]. Within the scope of this Thesis, the framework of Network Physiology poses an interesting domain to contextualise the work in Chapters 4 and 5.

2.7 Conclusion of the Chapter

This Chapter provided the background and literature review that is associated with the work presented in Chapter 3,4 and 5. Section 2.1 discussed key characteristics of physiological systems that have to be considered during the design and implementation of algorithms that aim to analyse respective time-series and introduced paradigms and hypothesis that connect the changes in variability of physiological time-series with changes in the state of physiological systems. Section 2.2 introduced Entropy as a measure of variability, outlined fundamental concepts and described processes that are utilized for the quantification of Entropy in time-series. Sections 2.3 and 2.4 introduced the ApEn, SampEn, PEn and DisEn algorithms by discussing in detail their mathematical definitions, algorithmic steps and implementations. Section 2.5 expanded into the domain of multivariate and multiscale analysis and presented in detail the mvMPE and mvMDE algorithms. Finally Section 2.6 presented the framework of Network Physiology discussing its structure and components and gave examples of implementations that expanded upon the initial framework. Moving forward the DisEn algorithm is utilized as the foundation of the work presented in the following Chapters.

Chapter 3

Univariate Dispersion Entropy Analysis for the Assessment of Missing and Outlier Samples and Development of Robust Variations

Note: This Chapter's contributions have been published in *Entropy*, 2020 [40].

3.1 Introduction

This Chapter presents the experiments conducted utilising the univariate DisEn algorithm for the analysis of physiological time-series containing simulated artifactual samples. The experiments are designed to measure the effects of artifactual samples in the operation of DisEn and develop variations that have improved performance during the analysis of time-series with increased amounts of artifacts, by mitigating their disruptive capacity.

Prior research tested the robustness of ApEn, SampEn, and Fuzzy Entropy (FuzzyEn) during the analysis of time-series containing missing samples. The results indicate that, while the classification capacity of the algorithms can be preserved under certain conditions, the fluctuations of entropy values can be large, affecting the accuracy of the results extracted for each analysed time-series segment [60]. Furthermore, recent research has provided new variations of SampEn leading to improved performance when analysing time-series with missing samples [56]. Concerning the effect of artifactual samples that have the form of outliers, ApEn and SampEn have been tested, and the results indicate that outlier samples can disrupt the process of entropy quantification to a much greater extent than missing samples and should therefore be a key consideration when testing the robustness of respective algorithms [61].

As discussed in Chapter 2 the DisEn algorithm was selected as the foundational algorithmic platform for this Thesis due to its favorable characteristics such as increased discrimination capacity and low computation time [36, 116]. However, while its performance is promising, it is important to ensure that the algorithm is capable of being deployed in scenarios where significant amounts of missing and outlier samples are contained in the extracted recordings. Consequently the research presented in this Chapter seeks to contribute the following:

- The quantification of the effect of missing and outlier samples on the performance of DisEn.
- The introduction of new variations of the DisEn algorithm to improve its performance when applied to time-series with missing and outlier samples.
- The assessment of the performance of the original algorithm and its variations, across different physiological time-series with simulated artifacts and under separate experimental setups defined by the percentage of missing or outlier samples and the degree to which these samples are grouped together or exist individually.

The Chapter is structured in the following manner. The Methods section presents the algorithmic variations of DisEn developed and tested in this set of experiments. It continues by presenting the datasets used, the process for producing time-series containing missing and outlier samples, the metrics used for performance assessment, and the statistical analysis applied to the results of the designed experimental setups. The Results section provides a summary of the results of the implemented statistical tests, continues by presenting the performance of DisEn variations, for each physiological type separately, applied to time-series with missing samples, and closes with the respective performance for time-series with outlier samples. In the Discussion section, important insights from the study are reviewed, and performance patterns are examined and interpreted considering the interplay between normal samples and missing or outlier samples, the rapidity of amplitude fluctuations of the analysed time-series, and the operation of the respective DisEn variation. Finally, limitations of the current study are addressed, and opportunities for future work are highlighted.

3.2 Methods

3.2.1 Dispersion Entropy (DisEn)

As discussed in Section 2.4.2, the DisEn algorithm has a set of steps prior to the formulation of a “quantised” time-series that focus on the mapping of the input time-series with a selected mapping function. The variations of DisEn introduced in this Chapter contain additional algorithmic steps that affect these initial processes of the algorithm. A summary of the original DisEn’ algorithmic steps is shown in 3.1 while the respective algorithmic diagrams for the variations are displayed in Figures 3.2-3.5.

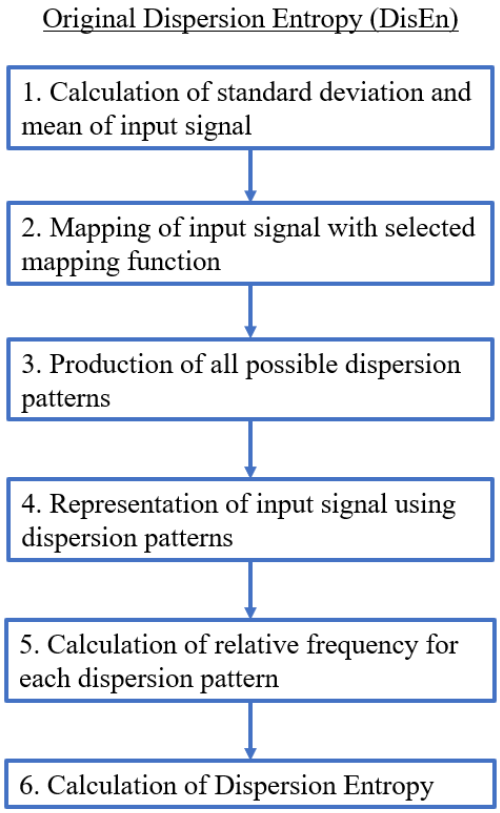

Figure 3.1: Algorithmic diagram of the original Dispersion Entropy algorithm.

3.2.2 Dispersion Entropy Variations

The following variations of the algorithm are developed and tested for improved performance when used for the analysis of time-series that contain missing and outlier samples. The first two variations aim at improving the robustness of DisEn by mitigating the effect of missing samples while the latter two by mitigating the effect of outlier samples.

Skip Sample Dispersion Entropy (SkipDisEn)

The SkipDisEn variation removes samples marked as missing and connects the remaining samples in a continuous time-series based on the computational steps shown in Figure 3.2. This is the default approach followed in prior entropy quantification algorithms [60, 84]; in this study, the aim is to assess the effectiveness of the respective DisEn variation when applied on time-series with missing samples.

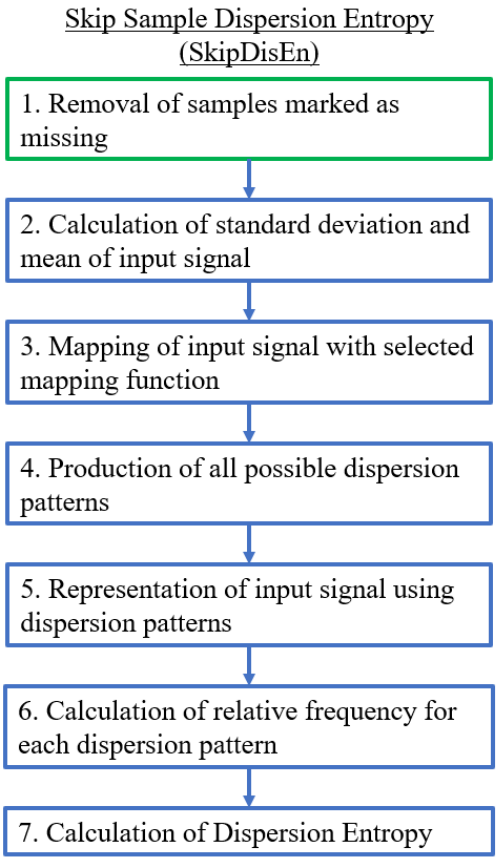

Figure 3.2: Algorithmic diagram of the Skip Sample Dispersion Entropy (SkipDisEn) variation. The added step is outlined in green.

Linearly Interpolated Dispersion Entropy (LinInterDisEn)

The LinInterDisEn variation uses linear interpolation to replace samples tagged as missing based on the equation $y(x) = \frac{y_o(x_1-x)+y_1(x-x_o)}{x_1-x_o}$, where y_0, y_1 are the amplitudes, and x_0, x_1 are the locations of the nearest available samples. In this variation, linear interpolation is being implemented due to promising results of performance improvement for entropy quantification algorithms in previous research [56, 117]. Similarly to SkipDisEn, this variation focuses on having improved performance when dealing with missing samples. Its computational steps are shown in Figure 3.3.

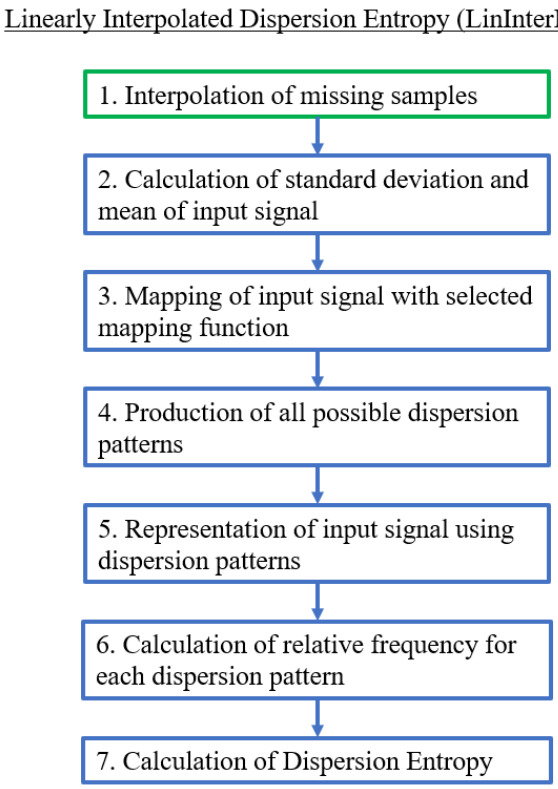

Figure 3.3: Algorithmic diagram of the Linearly Interpolated Dispersion Entropy (LinInterDisEn) variation. The added step is outlined in green.

Alternative Statistical Metrics Dispersion Entropy (AltMetDisEn)

The AltMetDisEn variation uses alternative statistical metrics for the implementation of mapping functions. The originally used μ is replaced with a median and σ is estimated using the median absolute deviation multiplied by the scaling factor of 1.4826 [118]. The new statistical metrics are chosen for their robustness to outliers in order to reduce the disruption of classes allocation due to the increases in the amplitude range of the input time-series caused by outliers [119]. Furthermore, AltMetDisEn is modified in the same manner as SkipDisEn in order to skip any samples tagged as missing and is therefore expected to have an analogous performance on the analysis of time-series with missing samples. The algorithmic diagram of AltMetDisEn is shown in Figure 3.4.

Dynamic Skip Sample Dispersion Entropy (DynSkipDisEn)

The DynSkipDisEn variation implements a dynamic skipping approach through the use of an additional parameter named cutoff. DynSkipDisEn aims at replicating the performance of SkipDisEn when applied to time-series with outlier samples by automatically discarding any samples with values that deviate more than a certain degree of σ, defined by the cutoff parameter, from the μ of each analysed time-series segment, as shown in Figure 3.5. The utilization of the σ and μ of each respective time-series segment for the removal of outliers is a necessary step when considering that the majority of physiological time-series are non-stationary [120,121] and therefore the respective σ and μ values are expected to change over time and consequently from one analyzed input segment to the next. Furthermore, since the cutoff threshold of this algorithm is a scaled version of the σ of its input time-series segment, the effect of outlier samples in the calculated σ should be taken into consideration when selecting the value of the cutoff parameter, as discussed in Section 3.4.5.

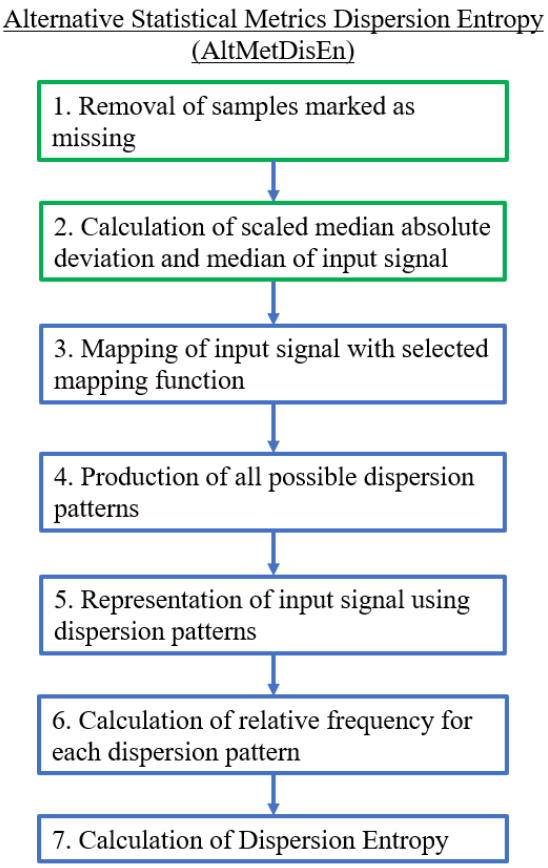

Figure 3.4: Algorithmic diagram of the Alternative Statistical Metrics Dispersion Entropy (AltMetDisEn) variation. The added and modified steps are outlined in green.

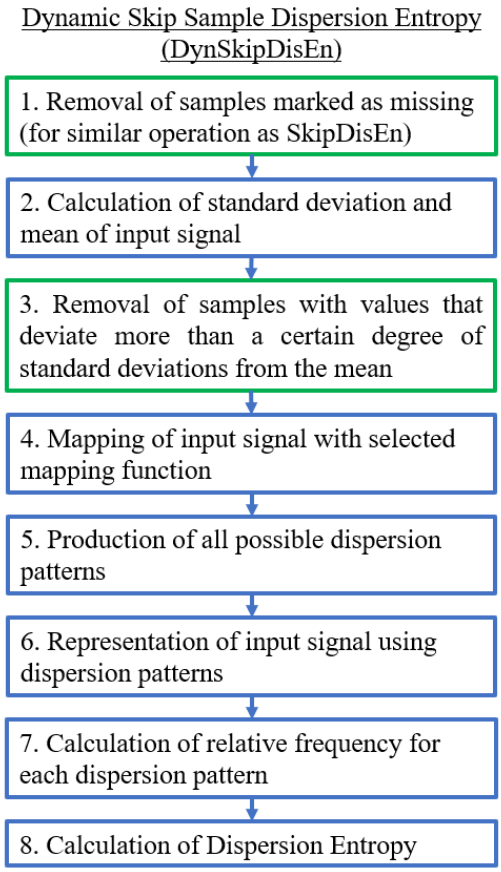

Figure 3.5: Algorithmic diagram of the Dynamic Skip Sample Dispersion Entropy (DynSkipDisEn) variation. The added and modified steps are outlined in green.

3.2.3 Experimental Datasets

Aiming to develop variations of the DisEn algorithm with robust performance across a spectrum of monitoring applications, the following physiological time-series are chosen for the study.

Heart rate variability (RR) data are commonly monitored in a range of biometric applications from wearable devices to patient monitoring in intensive care units for monitoring the cardiovascular system of individuals [122, 123]. The Fantasia Database [124] publicly available in Physionet [125] contains 40 EEG recordings of healthy adults sampled at 250 Hz, while also providing the respective RR data sampled at the rate of occurrence of each interval, which are used for this study. In total, RR data of 20 young adults are chosen for analysis.

In addition, EEG are chosen as a representative and commonly analysed signal for monitoring the nervous system of individuals. The EEG signals used for this study are retrieved from the publicly available CHB-MIT Scalp EEG Database [125, 126] which contains recordings from children that had intractable seizures. The data used in this study consist of 13 records of bipolar FP1-F7 channel recordings sampled at 256 Hz.

To measure the performance of the algorithm in monitoring the operation of the respiratory systems of individuals, the RI signal is chosen. A total of 15 recordings are retrieved by the publicly available BIDMC PPG and Respiration Dataset [125, 127] which contains recordings from critically-ill patients during hospital care at the Beth Israel Deaconess Medical Centre. Each recording contains 8 minutes of RI signal sampled at 125 Hz.

In this Chapter, the choice to utilize both a derived time-series such as the RR and signals such as the EEG and RI is made to ensure that differences in the effect of missing and outlier samples are measured in both types of time-series that can be analyzed across different applications.

3.2.4 Generation of Disrupted Time-Series

The production of "disrupted" time-series containing missing (disruptedM) or outlier (disruptedO) samples, used for measuring the performance of different variations of the DisEn algorithms is done through the following steps. For disruptedM time-series:

1. Extraction of ground truth DisEn values. Each original time-series containing N points without missing samples is separated in non-overlapping windows of 360 samples each. The selected window length is applicable for the analysis of both waveform physiological time-series at a high temporal resolution [36, 98] such as the EEG and RI time-series, and the analysis of derived physiological time-series, such as the RR, which can have a lower sampling frequency and therefore require windows of small sample length for effective temporal analysis [128]. It is important to note that the capacity to retrieve a viable output value from relatively small sample lengths is considered one of the advantages of the original DisEn algorithm [36, 98] compared to algorithms such as SampEn as discussed in Sections 2.3.3 and 2.4.3. The original algorithm DisEn is used to calculate the ground truth DisEn value of each respective window.
2. Segmentation of time-series. Copies of the initial time-series are segmented in groups of 1–5 samples, as defined by the grouping factor G. The G factor values used are 1, 2, 3, 4, and 5 samples.
3. Marking of missing samples. Based on the percentage factor P, a percentage of segments are uniformly drawn from each time-series and their samples are marked as missing. The P factor values used in this study are 10%, 20%, 30%, 40%, and 50%.
4. Production of random variations. Finally, the above process is replicated 10 times for each combination of P and G values producing different random variations for each experimental setup.
5. Total number of disrupted time-series. As a result from each initial time-series, $5 \times 5 \times 10 = 250$ "disruptedM" versions are produced, and these are used to assess the performance of the DisEn algorithm variations.

The rationale of choice for the above setups of "disruption" is to acquire a clear perspective of how increases in the number of missing samples affect the performance of DisEn algorithms and whether that performance changes when these missing samples are distributed individually or clustered together in groups considering that both events are common in physiological monitoring applications.

For the production of "disruptedO" time-series, an almost identical process is used. However, in this case, the modified samples are not marked as missing; instead, their amplitude is replaced with a value outside the physiological range of the original time-series. Similarly to previous experiments, testing the robustness of ApEn and SampEn to outliers [61], the amplitude of each outlier sample is obtained from a Gaussian distribution. For each physiological time-series the outliers' μ is defined as: $outliermean = \pm 4 \times max|amplitude|$ and σ as: $0.5 \times max|amplitude|$. Half of the modified samples are given a positive value, and half of them are given a negative value. For G factors higher than 1, all modified samples within a group share the same sign and value. The choice of setting the μ of the distribution to be the maximum absolute amplitude observed in the input time-series multiplied by a factor of 4 is made to ensure that outlier samples are outside the physiological range of the recorded time-series, and at the same time simulate the limitation of the maximum amplitude of the recording equipment. The value of σ for each time-series is chosen to allow outlier values to vary, as it is expected, while at the same time not allow their range of values to spread within physiological range. Similarly to time-series with missing samples for each original time-series, $5 \times 5 \times 10 = 250$ "disruptedO" versions are produced.

3.2.5 Performance Assessment

As mentioned in Section 3.2.4, the initial time-series is separated in windows and the ground truth DisEn value, for each window, is computed and stored using the original DisEn algorithm. The same process takes place for each disrupted time-series, and the absolute percentage deviation is calculated using the ground truth DisEn value of a specific window versus the equivalent DisEn value calculated from the "disrupted" version of the same window. This assessment is applied to each physiological dataset separately. To summarise the performance of the selected algorithm for a setup of P and G values, a single μ absolute percentage deviation value is acquired and presented alongside its respective σ. This is achieved by averaging across

1. the windows of each time-series,
2. the 10 different "disrupted" editions of each time-series, and
3. the total number of time-series that have been chosen from the respective dataset records.

Furthermore, in the case of the AltMetDisEn variation, if the variation is applied to the original time-series, the calculated DisEn values will differ from those of the original DisEn algorithm. This occurs because the changes implemented in the AltMetDisEn variation occur prior to mapping the input time-series with a selected mapping function and therefore have an effect even when no missing or outlier samples exist. To maintain consistency with the rest of the performance measurements, the values of the original DisEn algorithm are used as ground truth in the calculation of error percentages for the AltMetDisEn similarly to the rest of the variations. To measure the error introduced by the AltMetDisEn variation when applied to the original time-series, the AltMetDisEn variation is also applied to each of the original time-series, and the mean absolute percentage deviation from the original DisEn values is calculated and reported in the respective parts of Section 3.3 to compare it with the error occurring due to missing or outlier samples. Finally, for all variations of DisEn tested, including the original algorithm, the parameter values chosen are as follows:

- Embedding dimension: $m = 2$ samples.
- Number of classes: $c = 6$ classes.
- Mapping approach: logarithm sigmoid function.
- Time delay: 1 sample.
- Cutoff: 0.7 σ (used only by DynSkipDisEn).

The parameter values are selected after consulting the respective literature [35, 36] and considering that each input window used in this study has a length of 360 samples [98].

3.2.6 Statistical Testing

The following statistical analysis is applied for the error percentage distributions produced by DisEn variations during each experimental setup.

- Kolmogorov-Smirnov Test. Each separate distribution is standardised and compared to a standard normal distribution using a Kolmogorov-Smirnov test.
- Mann-Whitney U Test. Based on the results of the Kolmogorov-Smirnov test, a Mann-Whitney U test is chosen and applied to all distribution pairs produced within the same experimental setup to test whether the distribution of error percentages produced by one DisEn variation is significantly different from the distributions produced from the other variations tested under the same experimental setup.

3.3 Results

3.3.1 Summary of Statistical Testing Results

The results of the Kolomogorov-Smirnov test applied to all error percentage distributions after they have been standarised indicate that all distributions reject the null hypothesis at the 1% significance level. Therefore, their error percentages do not come from a Gaussian distribution.

The Mann-Whitney U test, which was chosen after taking into consideration the non-Gaussian nature of the distributions, indicates that out of the total 450 distribution pairs tested, 441 U tests reject the null hypothesis with a strict threshold of a p-value lower than 10^{-3}. Actually, 97% of p-values are lower than 10^{-7}. The nine error percentage distribution pairs that do not display statistically significant difference are signified in their respective sections that follow.

3.3.2 Experimental Setups for Time-Series with Missing Samples

The variations SkipDisEn, AltMetDisEn, and LinInterDisEn were tested on the three separate physiological datasets of RR, EEG, and RI time-series that have been modified to contain missing samples, as described in Section 3.2.4. Their performance was assessed under 25 different experimental configurations as defined by the percentage of missing samples, P factor, and the grouping of missing samples, G factor. The original version of the DisEn algorithm would return an invalid output if a single sample within the input time-series was marked as missing, resulting in very low performance when dealing with time-series containing missing samples. Therefore, in this part of the analysis, only the performances of new variations are presented and compared.

Performance for RR Time-Series with Missing Samples

As shown in Figure 3.6, SkipDisEn and AltMetDisEn display similar performance when analysing RR time-series. The μ percentage error for SkipDisEn is within the range of 0.98% and 4.70%, with minimum values at $P = 10\%$, $G = 5$ and maximum values at $P = 50\%$, $G = 1$, respectively. The mean percentage error for AltMetDisEn is within the range of 2% and 4.12% observed at $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 1$, respectively. Furthermore, the mean percentage deviation of the ground truth values for AltMetDisEn from the original ground truth values is calculated at 2.41% with a σ of 1.23%.

LinInterDisEn displays a significantly higher average error rate within the range of 1.15% and 16.29% with minimum deviation at $P = 10\%$, $G = 1$ increasing significantly especially in cases of “clustered” missing samples (higher G-Factor values) to reach a maximum average error rate of 16.29% observed at $P = 50\%$, $G = 5$. The effect of “clustered” missing samples in the performance of LinInterDisEn is expected due to the reduced accuracy of synthetic samples produced using linear interpolation when a higher number of adjacent samples are missing.

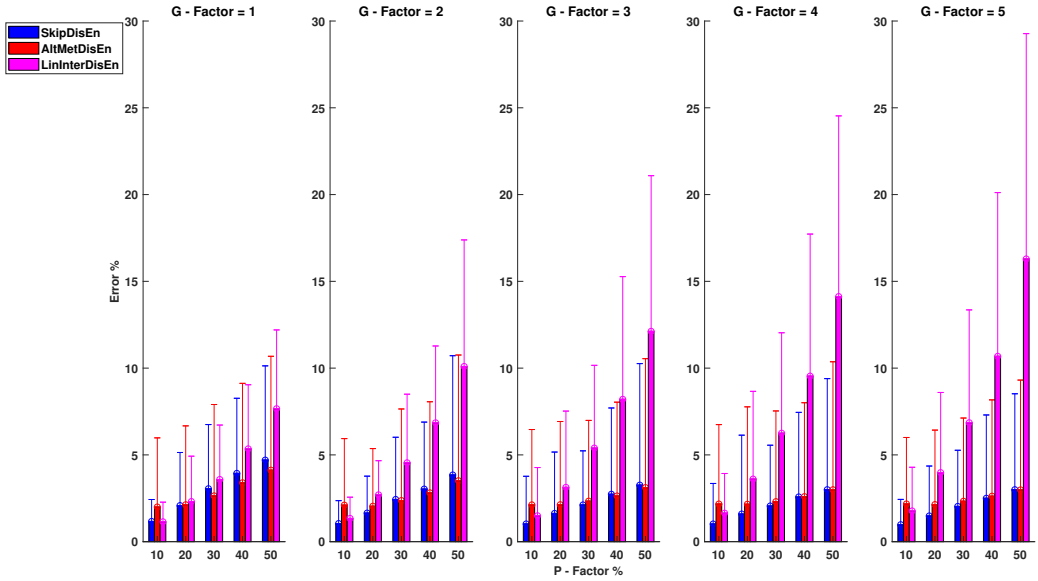

Figure 3.6: Performance of DisEn variations on RR time-series with missing samples. The μ and σ of the percentage error are shown for each tested variation. The distribution pairs of SkipDisEn and AltMetDisEn for: $P = 40\%$, $G = 4$; $P = 50\%$, $G = 4$; $P = 40\%$, $G = 5$; $P = 50\%$, $G = 5$ do not display statistically significant differences based on the Mann-Whitney U test.

Out of the 75 U test results retrieved in this group of experimental setups, 4 distribution pairs do not display statistically significant difference. These consist of the error percentage distributions acquired from the SkipDisEn and the AltMetDisEn variations for the experimental setups of:

- $P = 40\%$, $G = 4$ with a p-value of 0.19.
- $P = 50\%$, $G = 4$ with a p-value of 0.03.
- $P = 40\%$, $G = 5$ with a p-value of 0.01.
- $P = 50\%$, $G = 5$ with a p-value of 0.19.

Performance for EEG Time-Series with Missing Samples

As shown in Figure 3.7, SkipDisEn and AltMetDisEn maintain similar levels of performance. However, in this analysis, LinInterDisEn displays better performance for lower P and G factor values. SkipDisEn's error is within the range of 1.38% and 7.59% observed at $P = 10\%$, $G = 3$ and $P = 50\%$, $G = 1$ and similarly AltMeDisEn's error is within the range of 2.28% and 7.39% observed at $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 1$, respectively. The mean absolute deviation of ground truth values of AltMetDisEn from the original ground truth values is calculated at 2.42% with a σ of 0.60%.

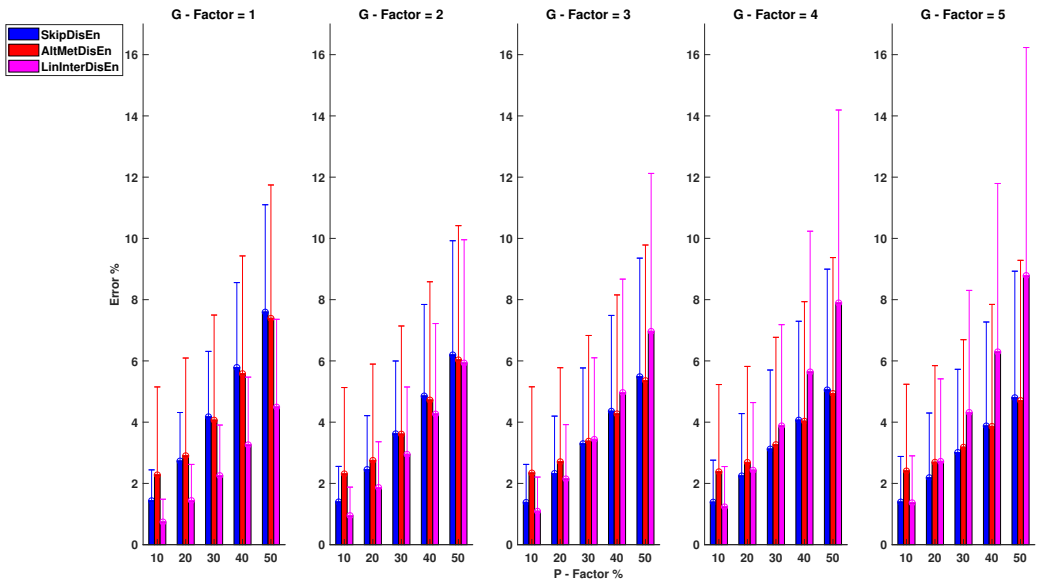

Figure 3.7: Performance of DisEn variations on EEG time-series with missing samples. μ and σ of the percentage error are shown for each tested variation. The distribution pairs of AltMetDisEn and LinInterDisEn for $P = 50\%$, $G = 2$ and SkipDisEn and LintInterDisEn for $P = 10\%$, $G = 5$ do not display statistically significant difference based on the Mann-Whitney U test.

LinInterDisEn achieves improved performance for lower values of P,G compared to SkipDisEn and AltMetDisEn. Its error is within the range of 0.74% and 8.79% observed at $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 5$, respectively. With increases in the values of experimental factors, particularly that of G, its initially superior performance eventually drops to lower than that of the aforementioned variations in the analysis of EEG time-series.

From the 75 U tests retrieved from this experimental setup, only two do not display statistical significance: the U test between the error percentage distributions of AltMetDisEn and LinInterDisEn for $P = 50\%$, $G = 2$ with a p-value of 0.59 and the U test between SkipDisEn and LinInterDisEn distributions for $P = 10\%$, $G = 5$ with a p-value of 0.01.

Performance for RI Time-Series with Missing Samples

As shown in Figure 3.8, there is a significant performance difference between SkipDisEn and AltMetDisEn. SkipDisEn has a mean percentage error in the range of 0.93% and 5.72% for $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 5$, respectively, while AltMetDisEn displays inferior performance with a mean percentage error in the range of 3.64% and 8.14% for $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 1$. The mean absolute deviation between the ground truth values of AltMetDisEn and the original ones is calculated at 3.03% with a σ of 1.47%. Both SkipDisEn and AltMetDisEn variations continue to follow a pattern of low error percentages that increase for higher P,G values across all tested physiological time-series.

In the analysis of RI time-series, LinInterDisEn significantly outperforms SkipDisEn and AltMetDisEn, unlike in the case of RR and EEG time-series. Its mean percentage error is limited in the range of 0.03% and 1.11% with minimum and maximum values observed at $P = 10\%$, $G = 1$ and $P = 50\%$, $G = 5$, respectively. Unlike in the cases of SkipDisEn and AltMetDisEn, it can be seen that the performance of LinInterDisEn changes based on the physiological time-series analysed. All U test results in this group of experimental setups indicate statistical significance between the distributions.

3.3.3 Experimental Setups for Time-Series with Outlier Samples

In contrast with the case of missing samples, the original version of the DisEn algorithm returns a valid DisEn value when applied to a window containing multiple outlier samples. Therefore, in this part of the study, the original version of the DisEn algorithm is used and its performance results are reported providing a starting point for measuring the effects of outliers on the calculation of DisEn values. The original DisEn algorithm and its AltMetDisEn and DynSkipDisEn variations were tested on the RR, EEG, and RI datasets under the same experimental configurations for factors P and G described in Section 3.3.2 and using the same values for the DisEn parameters as defined in Section 3.2.5. In this experimental setup, the analysed time-series have been modified to contain outlier samples outside the physiological range of each time-series as described in Section 3.2.4.

Performance on RR Time-Series with Outlier Samples

Figure 3.9 shows that the original DisEn displays poor performance on the analysis of RR time-series, especially for lower values of the P factor. Its mean absolute error is in the range of 24.35% and 72.58% for the configurations $P = 50\%$, $G = 1$ and $P = 10\%$, $G = 5$, respectively. AltMetDisEn displays improved performance in the cases of low P values. However, for higher P values, its performance is similar to that of the original DisEn. Its percentage error is in the range of 22.64% and 55.78% observed at $P = 50\%$, $G = 1$ and $P = 30\%$, $G = 5$, respectively. DynSkipDisEn achieves the best performance with an error percentage in the

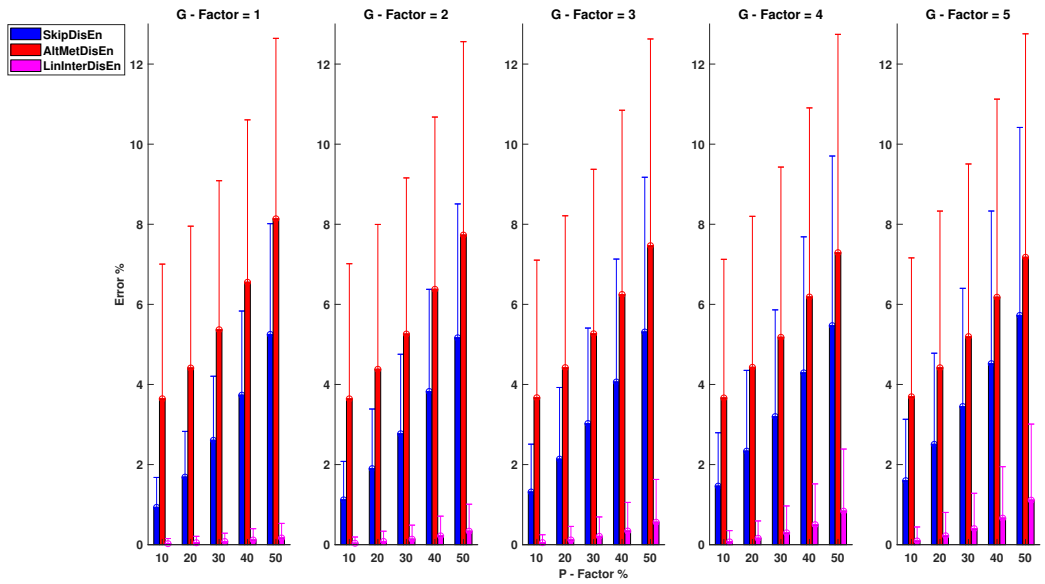

Figure 3.8: Performance of DisEn variations on RI time-series with missing samples. The μ and σ of the percentage error are shown for each tested variation. All distribution pairs display a statistically significant difference based on the Mann-Whitney U test.

range of 14.58% and 17.84% for $P = 40\%$, $G = 1$ and $P = 10\%$, $G = 3$. For the original DisEn and the AltMetDisEn variation, a certain amount of performance improvement is noticed as the percentage of outlier samples increases, which indicates that a deeper analysis on the effect of outlier values on the performance of DisEn is required. This is discussed in Section 3.4.4 of the study.

The only distribution pair that does not display a statistically significant difference for this group of experimental setups consists of the original DisEn and AltMetDisEn distributions for $P = 40\%$ and $G = 5$ with a p-value of 0.20.

Performance for EEG Time-Series with Outlier Samples

In the case of EEG time-series, the performance of all variations seems to improve compared to RR time-series. As shown in Figure 3.10, the original DisEn displays an error rate in the range of 16.37% and 62.55% for $P = 50\%$, $G = 2$ and $P = 1$, $G = 5$. AltMetDisEn achieves an improved performance for lower P and G values with a percentage error rate in the range of 14.11% and 40.47% for $P = 30\%$, $G = 1$ and $P = 10\%$, $G = 5$. Once more, the best performance is achieved by DynSkipDisEn with a percentage error limited in the range of 8.34% and 11.89% for $P = 40\%$, $G = 1$ and $P = 10\%$, $G = 5$. All distribution pairs for this group of experimental setups display a statistically significant difference.

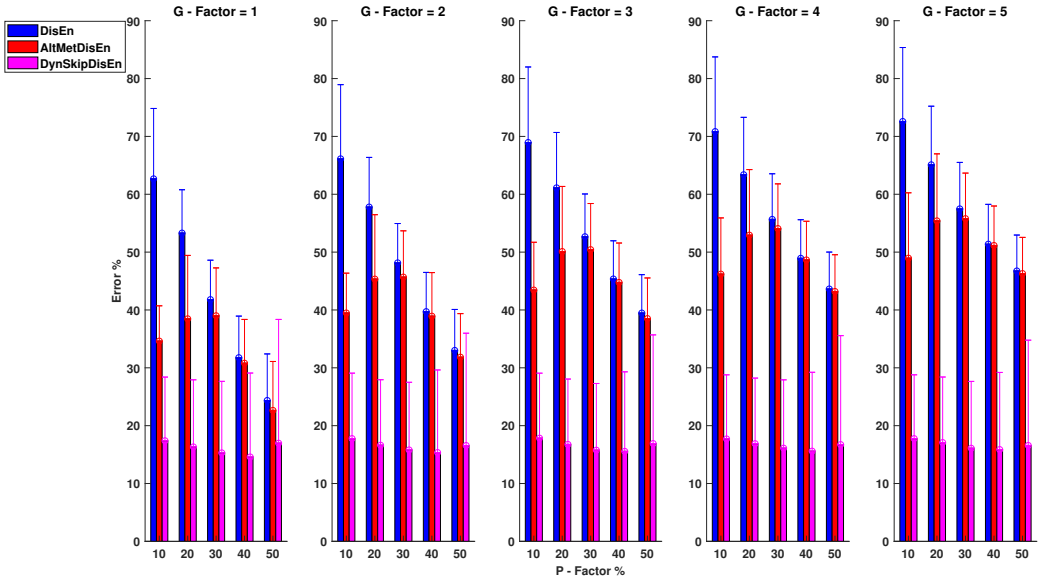

Figure 3.9: Performance of DisEn variations on RR time-series with outlier samples. The μ and σ of the percentage error are shown for each tested variation. The distribution pair of the original DisEn and AltMetDisEn distributions for $P = 40\%$, $G = 5$ does not display a statistically significant difference based on the Mann-Whitney U test.

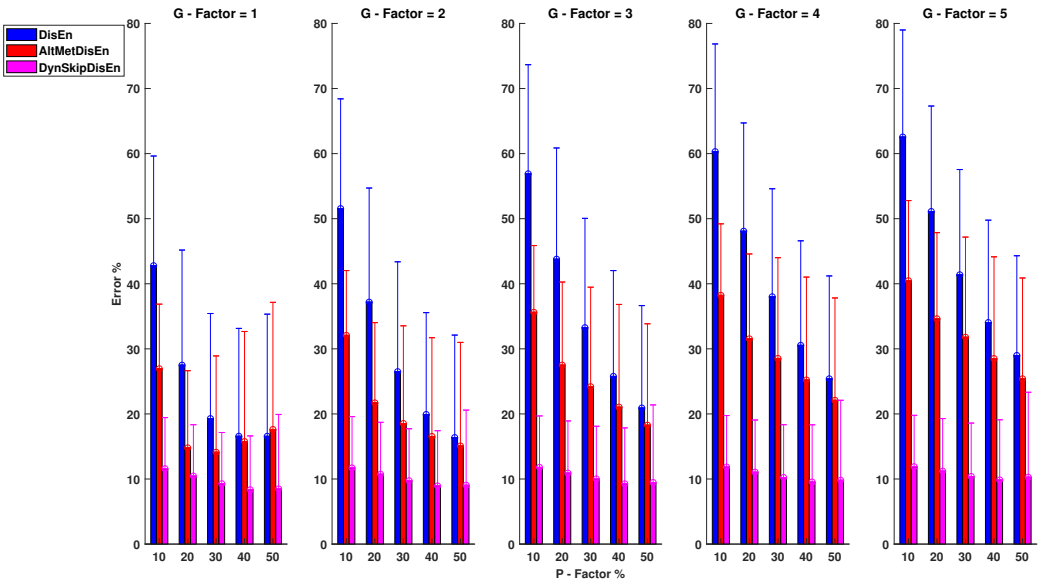

Figure 3.10: Performance of DisEn variations on EEG time-series with outlier samples. The μ and σ of the percentage error are shown for each tested variation. All distribution pairs display statistically significant difference based on the Mann-Whitney U test.

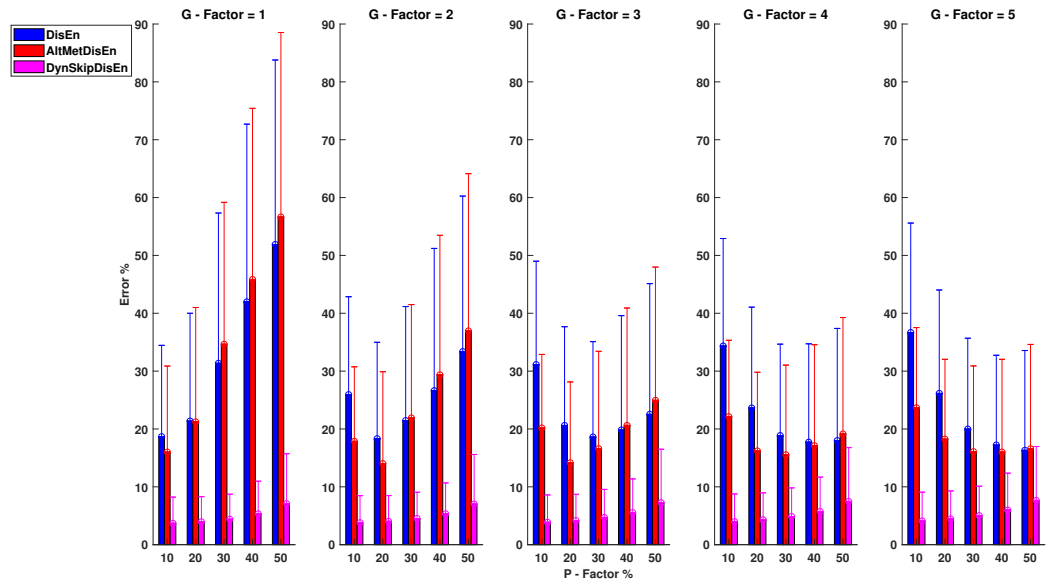

Figure 3.11: Performance of DisEn variations on RI time-series with outlier samples. μ and σ of the percentage error are shown for each tested variation. The distribution pairs of the original DisEn and AtlMetDisEn for $P = 50\%$, $G = 5$ and AltMetDisEn and DynSkipDisEn for $P = 50\%$, $G = 5$ do not display statistically significant differences based on the Mann-Whitney U test.

Performance RI Time-Series with Outlier Samples

As shown in Figure 3.11, when applied to RI time-series, the original DisEn algorithm percentage error is in the range of 16.31% and 51.89% for $P = 50\%$, $G = 5$ and $P = 50\%$, $G = 1$. The AltMetDisEn performance shows a percentage error range of 14.02% and 56.65% for $P = 20\%$, $G = 2$ and $P = 50\%$, $G = 1$, respectively. DynSkipDisEn achieves a significantly improved performance with a percentage error limited in the range of 3.70% and 7.65% for $P = 10\%$, $G = 1$ and $P = 50$, $G = 5$. For this group of experimental setups, two distribution pairs do not display any statistically significant difference. These consist of the original DisEn and the AltMetDisEn distributions for $P = 50$, $G = 5$ with a p-value of 0.01 and the AltMetDisEn and DynSkipDisEn distributions for $P = 50$, $G = 5$ with a p-value of 0.01.

3.3.4 Computation Time

To ensure that the variations presented and tested in this study preserve the low computation time of the original DisEn algorithm [36], their computation time was measured for the analysis of time-series with a length of 360 and with 9000 samples on randomly selected time-series segments from all three physiological datasets, and the results are presented in Tables 4.1 and 3.2, respectively. The computations are carried out using a PC with Intel(R) Core(TM) i7-8750H CPU @ 2.2GHZ, 16 GB RAM running MATLAB R2018b. The computation time of

Table 3.1: Computation time in milliseconds for the time-series segments of 360 samples.

	RR	EEG	RI
DisEn	1.6	1.5	1.9
SkipDisEn	1.5	1.4	1.9
AltMetDisEn	1.7	1.8	1.9
LinInterDisEn	1.6	1.7	2
DynSkipDisEn	1.5	1.4	1.5

Table 3.2: Computation time in seconds for the time-series segments of 9,000 samples.

	RR	EEG	RI
DisEn	2.1	2.3	2.3
SkipDisEn	2.4	2.5	2.2
AltMetDisEn	2.6	2.8	2.7
LinInterDisEn	3.1	3.3	3.2
DynSkipDisEn	2.5	2.5	2.6

the original DisEn algorithm is measured when applied to randomly selected segments of the original time-series, SkipDisEn and LinInterDisEn are applied to disruptedM time-series, while the AltMetDisEn and DynSkipDisEn are applied to disruptedO time-series. As the results indicate, no significant difference in the computation time is noted across the algorithmic variations apart from a small expected increase in the case of the LinInterDisEn variation observed at the time-series segments with a length of 9000 samples due to the additional linear interpolation mechanism introduced.

3.4 Discussion

As part of this study, novel variations of the DisEn algorithm are introduced to improve its performance when applied to time-series with missing and outlier samples. Time-series from three different physiological time-series— RR, EEG, and RI —are modified to produce multiple variations of time-series containing missing samples (disruptedM) and time-series containing outlier samples (disruptedO). Each produced variation of disruptedM and disruptedO time-series corresponds to a different experimental setup in order to assess the performance of algorithmic variations under different percentages of missing or outlier samples and under different degrees of grouping of these samples. The results of the analysis indicate that, while low-data quality, especially when it arises from artifactual outlier samples, can cause disruption in the entropy quantification mechanisms of the DisEn algorithm, significant improvements in its performance can be achieved with corresponding modifications.

3.4.1 Differences in the Effect of Missing versus Outlier Samples

An initial finding of the study concerns the effectiveness of DisEn when applied to disruptedM time-series versus its limited performance during the analysis of disruptedO time-series. In the case of disruptedM time-series, the SkipDisEn variation, which requires minimal modification of the initial DisEn algorithm, is capable of achieving a mean percentage error that remains lower than 7.6% across all examined time-series, even when up to 50% of the original samples are missing. Furthermore, the LinInterDisEn variation's performance can surpass that of SkipDisEn, as shown in the analysis of disruptedM RI time-series. However, it is important to consider that LinInterDisEn's performance is significantly affected by the rapidity of amplitude fluctuations in the investigated time-series and should therefore only be used when the respective information is available. For example as seen from the presented results, while LinInterDisEn outperforms SkipDisEn for RI time-series, it underperforms in the case of RR and EEG time-series which contain dynamics that are represented by more rapid amplitude fluctuations and therefore the accuracy of the imputation is reduced.

On the other hand, from the disruptedO time-series results, it can be verified that, in the case of DisEn, similarly to ApEn and SampEn [61, 129], outlier samples have a much more disruptive effect than missing samples. Taking into consideration the effectiveness of SkipDisEn in acquiring viable DisEn values, it is recommended when possible to label outlier samples as missing in order to achieve a performance close to that observed in the analysis of disruptedM time-series. However, when the removal of all outlier samples is not guaranteed, the DynSkipDisEn variation is recommended. DynSkipDisEn is designed to tackle the disruption of the class allocation process, through the removal of samples that deviate from the μ more than a certain degree of σ, as defined by the additional cutoff parameter.

3.4.2 Factors Affecting the Performance of LinInterDisEn

As shown in Section 3.3.2, the performance of LinInterDisEn is primarily affected by two factors: the clustering of missing samples on the analysed time-series, controlled by the G factor of the defined experimental setups, and the rapidity of amplitude fluctuations in each time-series analysed. The clustering of missing samples has an expected negative effect on the performance of LinInterDisEn due to the reduced quality of synthetic samples produced using linear interpolation when a larger amount of adjacent samples are missing.

Furthermore, the characteristics of the time-series analysed are expected to affect the performance of the linear interpolation mechanism when considering that RR time-series contain rapid amplitude fluctuations that are harder to estimate using linear interpolation [130–132]. RI time-series are dominated primarily by low-frequency components [133, 134], leading to a

larger number of linear time-series segments that can be more accurately estimated. Finally, EEG time-series fall in between with a significant amount of high-frequency components [135], leading to amplitude fluctuations that can challenge the LinInterDispEn algorithm, especially in experimental setups with high P and G factor values.

3.4.3 Standard Deviations of Performance Measurements

The σ values recorded throughout the presented experimental setups signify fluctuations in the performance of each tested variation on a window-by-window basis. This deviation occurs primarily due to two factors, the first one being that the disrupted time-series were formulated by introducing missing and outlier samples on the original time-series randomly, at the entire length of the time-series in order to more realistically simulate the phenomenon instead of equally distributing them across each window. Therefore, some windows would have more missing or outlier samples than others, leading to inevitable fluctuations in the tested performance.

However, the second factor that leads to increased σ values of the mean performance error is the small sample length of the analysed windows. An important advantage of the original DisEn algorithm is its capacity to acquire valuable insights, even when applied to time-series windows with small sample lengths [36,98]; for that reason, it is chosen the performance of the original DisEn algorithm and its variations using 360 samples per window, which is a considerably smaller sample length than what was commonly used in similar studies concerning the performance of ApEn and SampEn when applied to time-series containing missing and outlier samples [56, 60, 61]. Considering the observed fluctuations in the algorithmic performance recorded in the study, it is recommended that, for field applications where the DisEn value of each window is considered individually, a larger sample length is used (if possible) when the analysed time-series are expected to contain missing and outlier samples.

3.4.4 Effect of Outlier Sample Percentage across Physiological Time-Series

In order to acquire a better perspective on the effect of outlier samples in the performance of DisEn variations across different physiological time-series, it is important to consider the mechanism through which outlier samples disrupt the DisEn calculation process. As mentioned in Section 2.4.2, during the second operational step of DisEn, a number of classes ($c = 6$ in these experiments) are allocated across the amplitude range of the mapped input time-series. With the introduction of outlier samples, this range expands significantly, resulting in fewer classes being allocated within the range of the original time-series. Instead, the majority of classes are allocated in the extended amplitude range. As a result, amplitude dynamics existing in the original time-series that would previously be represented using mul-

tiple dispersion patterns are now classified under a single dispersion pattern category, leading to a much lower output DisEn value. This phenomenon is shown in Figures A.1–A.6, where examples of disrupted dispersion patterns for $P = 10\%$, $G = 1$ are shown in green, and those for $P = 50\%$, $G = 1$ are shown in red.

For high percentages of outlier samples, new dispersion patterns arise, and these do not represent physiological dynamics that occur within the original samples of the time-series but instead represent the amplitude dynamics that occur between original and outlier samples. This can lead to an increase in the irregularity of the input time-series and therefore to an increase in the calculated DisEn value for disruptedO time-series with high P factor values. This is an important phenomenon to consider during the analysis of the performance of DisEn variations when tested on disruptedO time-series.

It is observed that, in the case of RR and EEG disruptedO time-series, with the effect being more prevalent in the case of RR, the performance of the original DisEn and the AltMetDisEn is actually increasing as the percentage of outlier samples increases, which at first can seem counterintuitive. However, in the case of RI time-series, the performance of DisEn variations does not follow a clear pattern. Taking into consideration the existence of rapid amplitude fluctuations in RR and EEG time-series, as opposed to the RI time-series, which contain primarily gradual changes in amplitude, as discussed in Section 3.4.2, the number of unique dispersion patterns used to describe each window of the original time-series is expected to be higher for RR followed by EEG and then by RI time-series. Therefore, their respective DisEn values are expected to follow a similar pattern.

As mentioned previously for small values of the P factor, the mean DisEn value drops significantly in all three cases due to the time-series apparently becoming more regular as shown in green in Figures A.1–A.6. As the outlier percentage increases, more dispersion patterns are introduced, as shown in red in Figures A.1–A.6, in order to describe the now multiple amplitude fluctuations that occur between normal and outlier samples. Consequently, the DisEn values of all three time-series increase for higher P factor values. In the cases of RR and EEG time-series, the increase in DisEn values that occurs brings them closer to their respective ground truth values, resulting in the performance “increase” observed in Sections 3.3.3 and 3.3.3. Therefore, this increase in performance is not achieved due to an internal mechanism of the algorithm but rather from the acquisition of DisEn values closer to the ground truth arising from amplitude fluctuations occurring between outlier and normal samples.

In the case of the original RI time-series, their corresponding DisEn are lower values compared to those for RR and EEG time-series. Therefore, increases in DisEn values for disruptedO time-series occurring from the aforementioned phenomenon do not necessarily bring the calculated DisEn values closer to the ground truth; therefore, algorithmic performance does not follow a pattern similar to that of RR and EEG disruptedO time-series.

Finally, in the case of DynSkipDisEn, the calculation of DisEn is not affected significantly by dispersion patterns arising from the interaction between original and outlier samples due to the significant amount of outlier samples that are removed.

3.4.5 Setting the Cutoff Parameter of DynSkipDisEn

When setting the value of the cutoff parameter for the DynSkipDisEn variation, it is important to balance two opposing sources of error. A high cutoff value, e.g., close to 2 σ from the μ, can allow an extensive amount of outliers within the range of samples analysed by the algorithm, leading to a significant reduction in its performance. On the other hand, a strict low cutoff value can lead to the false positive removal of valid samples. Considering the more disruptive nature of outliers as opposed to that of missing valid samples, a more conservative approach towards choosing lower values for the cutoff parameter is recommended when considering multiple options. Therefore, when setting the value of the cutoff, the quality of the data to be analysed should be taken into consideration when corresponding information is available.

Within the scope of this study, the cutoff parameter is set to the strict value of 0.7 σ from the μ due to the high percentage of outlier samples introduced in the majority of the experimental setups. Furthermore, due to the range of values that outliers can cover, those with values further from the physiological range increase the calculated σ of the input window, while those with values closer to physiological range have a higher probability of passing through the cutoff threshold, making a strict cutoff value a necessity. As shown in Figures A.7–A.15, when comparing the capacity of DynSkipDisEn with a cutoff of 0.7 versus a cutoff of 1 to reconstruct the class allocation pattern of the original time-series from its disruptedO versions, having a strict cutoff value of 0.7 leads to significantly improved performance, especially in the cases of higher outlier percentages where a cutoff of 1 leads to highly disrupted class allocation patterns. However, for applications where DynSkipDisEn is combined with preprocessing for the removal of outlier samples, a higher cutoff between 1 and 2 is recommended, since in that case the percentage of remaining outliers in the time-series should be significantly lower, and a higher cutoff value would therefore provide improved performance.

It is important to highlight that while the value of 0.7 σ was selected for this study, the calibration of the cutoff parameter through any form of iterative fine tuning, would require the appropriate separation of the available data into training, validation and testing sets to avoid overfitting. Consequently, an iterative derivation of the cutoff parameter would have to utilize appropriate output labels for supervised training which might not be available in certain applications.

3.4.6 Limitations of Current Study and Future Work

As indicated by the results of this study, the rapidity of amplitude of the investigated physiological time-series have a direct effect on the performance of DisEn and its variations. Therefore, while the physiological time-series used in the study— RR, EEG and RI —are commonly used for health monitoring applications, the study can be expanded to verify the observations in additional physiological time-series such as EMG, blood pressure, and potentially intracranial pressure signals. Therefore, it is suggested that similar experimental setups are used to assess the performance of DisEn variations prior to their deployment in respective applications.

Furthermore, the DisEn algorithm can be implemented using a wide variety of mapping functions. Within the scope of this study, the logarithm sigmoid function is used due to its successful implementation in previous studies of physiological time-serie analysis [36, 98]. However, as mentioned in Section 3.4.4, outlier samples tend to disrupt the process of class allocation, which follows the mapping of the original time-series with the chosen mapping function. It would therefore be valuable to expand the study on measuring the robustness of different mapping functions to outliers such as the normal cumulative distribution function. Consequently, when optimising a DisEn variation for a specified implementation, the mapping function should be chosen by considering both the characteristics of the input time-series and the mapping function's robustness to outlier samples.

Furthermore, while DynSkipDisEn is a promising variation trying to automatically remove outlier samples, there are two points that should be taken into consideration. The first one is that, even if the samples of the original time-series follow a Gaussian distribution, the existence of outliers will change the distribution in a non-Gaussian form, this should be taken into consideration, since it will affect the calculated μ and σ based on which the DynSkipDisEn filters the samples of the input window. Finally, as suggested in Section 3.4.5, the correct choice of value for the cutoff parameter should consider the amount of outliers located in the analysed time-series. This information might not be available in certain applications. When this is the case, the choice of a relatively low cutoff value is recommended, considering the more disruptive nature of outliers when compared to missing samples.

3.5 Conclusion of the Chapter

This study investigates the effect of missing and outlier samples in the operation of DisEn and presents algorithmic variations to minimise their effect and improve its performance. The results indicate that the effect of missing samples can be effectively reduced with the addition of a skipping step in the operations of DisEn, while linear interpolation can further improve its performance when operating on time-series containing primarily low-frequency components. Outlier samples affect to a larger extent the performance of DisEn by disrupting the amplitude

range during the class allocation step of the algorithm. A significant mitigation of the disruptive effect of outliers is achieved with the introduction of a cutoff parameter in the DynSkipDisEn variation. The presented DisEn variations operate using information only from within the time-series segment that is used as input at a time to allow for a real-time entropy quantification process. However, upon availability, information concerning a time-series' dominant frequency components and estimations of missing and outlier samples' percentages can aid in the selection of the appropriate DisEn variation and the optimisation of its parameter values. The insights and algorithmic variations presented in this study could aid the implementation of DisEn in physiological monitoring applications.

Chapter 4

Multivariate Dispersion Entropy Analysis for the Assessment of Outliers and Detection of Artifactual Network Segments

Note: This Chapter's contributions have been published in *Entropy*, 2021 [41].

4.1 Introduction

This Chapter presents a set of experiments that study the challenges posed by artifactual outliers during multivariate entropy quantification. As mentioned in Chapter 2, multiple multivariate entropy quantification algorithms have been developed to measure the dynamics developed across different recorded variables. Examples include multivariate versions of SampEn [69], FuzzyEn [136, 137], PEn [76], and DisEn [39]. Furthermore, the application of such algorithms for the extraction of multivariate entropy-based features can be utilized within the scope of Network Physiology [18, 105] to produce a network-based representation of the system's state [114, 115].

However, for the accurate detection of topological transitions occurring within the network and their correct association with transitions in the state of the monitored system, it is crucial to ensure that the features extracted with these kind of algorithms are accurately measuring the dynamics of the recorded variables as opposed to quantifying disruptions that arise from artifactual samples [62, 63, 138, 139]. Currently, there is a lack of studies testing the disruptive capacity of artifactual outliers in the performance of multivariate entropy quantification algorithms. Therefore, the motivation behind the work of this Chapter is to extend the study presented in Chapter 3 from univariate to multivariate analysis, to assess differences and similarities with regards to the disruptive capacity of artifacts in the performance of DisEn.

The design of the experiments in this Chapter were based on the insights provided by the experiments of Chapter 3. Based on those results, it was noted that while the effects of missing samples can be minimised significantly, outlier samples can have a disruptive impact in the output values of DisEn features in alignment with prior research conducted for ApEn and SampEn [61]. Consequently, the work in this Chapter was conducted with the following aims in mind. To study the effect of artifactual outliers on multivariate DisEn features and compare it to the effect observed on univariate features; and to utilize the sensitivity of univariate and multivariate DisEn features to artifactual outliers for the detection of artifactual network segments. The approach taken with regards to the utilization of the DisEn algorithms in this Chapter is the reverse to that followed in Chapter 3. In Chapter 3 the challenge of low data quality was approached through the modification of the univariate DisEn algorithm to improve its robustness to artifactual samples. In this Chapter, the algorithm is used as a sensor [140], using the sensitivity of its features to outliers as an advantage to classify between viable and artifactual network segments.

For this purpose, network segments are formulated from multi-channel time-series. Within the context of Network Physiology each channel would correspond to a node of the network representing the monitored system and each network segment to a temporal slice of the system. Each channel consists of one of the following physiological signals: EEG, nasal RESP, arterial BP, and ECG. Artifactual outliers are simulated with one channel being “disrupted” at a time to allow for the study of differences in the effect of outliers based on the morphology of the signal corresponding to the channel. Multiple experiments are conducted with varying percentages of outlier samples.

The values of features extracted from network segments containing artifactual outliers are compared with the respective values of features extracted from the original network segments to quantify the disruptive capacity of the outliers. Finally, the logistic regression classifier is tested in two configurations—a univariate, and a network-based multivariate configuration. The two configurations are deployed to allow for comparisons between the two approaches and identify benefits, as well as potential challenges when moving from univariate to multivariate analysis for the classification of network segments.

Consequently, the contributions of the work presented in this Chapter are:

- The quantification of the effect of artifactual outliers in the accuracy of both univariate and multivariate DisEn feature values extracted from the physiological network segments formulated form multi-channel time-series.
- The assessment of whether a simple logistic regression classifier could be effectively trained on distributions of features extracted from “normal” and “artifactual” network segments to differentiate between the two.

4.2 Methods

4.2.1 Stages of the Study

The research presented in this manuscript is conducted in the following stages:

- Network Formulation and Feature Extraction: The available multi-channel time-series recordings are segmented in multivariate network segments. For each normal network segment, 16 different variants of artifactual network segments are produced, one for each experimental setup of interest, as described in Section 4.2.4. From each segment, a total of 15 univariate and multivariate features are extracted.
- Statistical Analysis: Conducted separately in each experimental setup, statistical analysis is applied to each extracted feature at the level of: separate feature distributions, pairs of feature distributions, and individual feature values, as described in Section 4.2.5.
- Artifactual Network Segment Detection: A univariate and multivariate network-based logistic regression classifier is trained and tested in each experimental setup to assess the capacity of both configurations in detecting artifactual network segments when outliers are present in different channels and at varying percentages of occurrence.

4.2.2 Experimental Data and Preprocessing

For the formulation of a network based on multi-channel time-series, the publicly available MIT-BIH Polysomnographic Database is chosen, which contains a total of 18 records of synchronised physiological signals initially recorded for the evaluation of chronic Obstructive Sleep Apnea (OSA) syndrome and digitized at a sampling interval of 250 Hz [125, 141]. For the purposes of this study, 11 records are selected based on the availability of complete and synchronised recordings of EEG, RESP, BP, and ECG signals. Within the utilized records there were occurrences of both healthy sleep and pathological sleep phenomena such as OSA. The inclusion of data corresponding to both healthy and pathological states is done to ensure that the comparison is made between all forms of valid data versus artifactual ones. Furthermore, the utilization of the ECG signal for the formulation of one of the channels, as opposed to the use of RR series utilized in Chapter 3, was made to ensure that all four channels of the time-series contained samples synchronized to the same sampling interval of 250 Hz.

Each multi-channel recording is segmented in non-overlapping windows of 7500 samples (30 seconds), resulting in a total of 1463 multi-channel network segments, 133 per record. The length of the window is chosen, after consulting the respective literature [36, 39] to ensure that it is long enough as to allow a sufficient amount of samples for the effective calculation of the output DisEn values, while at the same time being short enough to provide an adequate temporal resolution for effective monitoring of the system.

Within the framework of Network Physiology, the recordings are analyzed in sets of 1463 multi-channel "network segments" consisting of windows of four synchronised channels, one from each physiological signal, following the process described in Section 4.2.3. No further preprocessing is applied to the the time-series prior to the application of the respective DisEn algorithms.

4.2.3 Extraction of DisEn Features

For each normal network segment, a total of 15 features are extracted, four of which are univariate, one from each channel, and 11 of which are multivariate.

In contrast to the Network Physiology framework, the features extracted in this study are not associated with an edge in a graph architecture. This is done to avoid the limitation of the original graph architecture to only consider bivariate interactions between the nodes of the network [18, 105]. Instead one feature is extracted by including all available nodes in the accessed network of EEG-RESP-ECG-BP channels and to retrieve the other 10 possible multivariate features, the following subnetwork combinations are utilized: EEG-RESP, EEG-BP, EEG-ECG, RESP-BP, RESP-ECG, BP-ECG, EEG-BP-ECG, RESP-BP-ECG, EEG-RESP-BP, EEG-RESP-ECG.

In the case of univariate features, for each channel, its respective 7500 sample window is fed as input to the univariate DisEn algorithm [35,36]. For each multivariate feature, the respective combination of synchronised 7500 sample windows is fed as input to the multivariate DisEn algorithm for its calculation [39].

Table 4.1 displays the parameter values which are the same for univariate and multivariate DisEn. The values were selected after consulting the performance benchmarks provided in the respective studies [36,39] and considering the empirical rules discussed in Section 2.4.3. Taking into consideration the number of 7500 samples contained within each window the number of classes (c) was set to 9 to allow for an effective quantization of the amplitude range present in each window. With the value of c defined and based on the recommendation for univariate DisEn [36] that $c^{m+1} < 7500$ the value of the embedding dimension (m) was set to 3. Finally, the time-delay (d) was set to 1 to ensure that no samples are skipped in the formulation of embedded vectors. As a preprocessing step, each individual time-series is mapped using the Normal Cumulative Distribution Function (NCDF).

Table 4.1: Parameter values for univariate and multivariate DisEn.

Parameter	Symbol	Value
Embedding Dimension	m	3
Number of Classes	c	9
Time Delay	d	1

4.2.4 Production of Artifactual Network Segments

Within the scope of this study, artifactual outlier samples are simulated across all four channels of – EEG, RESP, BP, and ECG – with one channel being "disrupted" at a time. Furthermore, the percentage of samples being outliers is determined by the percentage factor P whose value varies across experimental setups in the levels of 0.1%, 0.5%, 1%, and 5%. As a result, a total of 16 experimental setups are formulated, containing the 1463 normal network segments and a corresponding variation of 1463 artifactual network segments.

The process through which the 1463 artifactual network segments of each experimental setup are produced is the following:

1. Marking of Outlier Samples: Based on the percentage factor P, a percentage of samples are uniformly drawn from each 7500-sample window, and their amplitude is replaced with a value in the outlier amplitude range.
2. Setting the Value of Outlier Samples: The amplitude of each outlier is obtained from a Gaussian distribution with a standard deviation (σ) equal to the absolute maximum amplitude of each channel: $\sigma = max|amplitude|$. Concerning the distribution mean (μ), there are two choices to be considered. The first choice, which is in alignment with the simulation processes followed by previous studies testing the effect of outlier samples in the performance of ApEn, SampEn, and univariate DisEn [40, 61, 129], would be to ensure that outlier values are outside the physiological range of each recorded signal while maintaining the amplitude boundaries of respective sensing equipment, leading to a distribution μ equal to $outliermean = \pm 4 \times max|amplitude|$. The second choice would be to set a lower distribution μ to ensure that a minority of outliers remain within physiological range to also cover certain scenarios where sensor miscalibration and recording interferences could produce outliers of the respective magnitude, and for that purpose, a μ equal to $outliermean = \pm 2 \times max|amplitude|$ is chosen. In order to provide results that are comparable to previous studies while at the same time covering all possible scenarios, each experimental setup is replicated for both outlier μ with the results of the second case, $outliermean = \pm 2 \times max|amplitude|$, presented first and discussed in more detail since it covers a wider range of scenarios. In all experimental setups, the sign of half the outliers is set to positive and the other half to negative, following random assignment.

3. Calculation of Artifactual DisEn Features: For each experimental setup, the corresponding 1463 artifactual network segments are used to calculate the respective univariate and multivariate artifactual DisEn Features based on the same process that was followed for the normal network segments in Section 4.2.3.

4.2.5 Statistical Analysis

To quantify the disruptive capacity of outlier samples in the accuracy of extracted network features, the following three-stage statistical analysis is applied.

1. Kolmogorov–Smirnov Test: Initially, each feature distribution is standardised and compared to a standard normal distribution using the Kolmogorov–Smirnov Test.
2. Mann–Whitney U Test: At the second stage, and after consulting the results of the Kolmogorov–Smirnov test, the Mann–Whitney U test is chosen to compare each feature distribution extracted from artifactual network segments with its corresponding feature distribution extracted from the respective normal network segments, to verify statistically significant differences between the distributions of each pair.
3. Mean Percentage Difference: Finally, for each DisEn feature extracted from an artifactual network segment, the absolute percentage difference from its original value, the one calculated from the respective network segment without outliers, is calculated. To provide a summary for every feature extracted during each experimental setup, its mean percentage difference (MPD) and σ of the percentage difference are calculated.

4.2.6 Artifactual Network Segment Detection

For the detection of artifactual network segments, a logistic regression classifier is applied in two configurations, a univariate and a multivariate one. The univariate configuration is utilising, from each network segment, only the four DisEn features that are extracted using the respective channel as separate input to the univariate DisEn algorithm, while the multivariate configuration utilises all the available 15 features of each network segment. The choice to implement two separate algorithmic configurations is made with the following aims in mind:

- To derive insights concerning the potential benefits but also challenges that arise when moving from univariate to multivariate analysis for network segment classification.
- To identify differences in classification performance, for both configurations, based on the channel containing outlier samples in each experimental setup.

For this purpose, both algorithmic configurations were tested under the same 16 experimental setups. Each setup contains features extracted from a total of 2926 segments, out of which 1463 are the original network segments, and the other 1463 are their artifactual variations, as determined by the parameters of the experiment. As mentioned in Section 4.2.2, the 1463 network segments correspond to 11 records, or 133 segments per record. The segments selected for training and testing during each experimental setup are selected in the following two data splits.

The first data split is done at the record level, with the first nine records used for training and the last two for testing purposes. This is done to ensure that the training of the classifier is done on different patients than the ones it is tested on, and therefore, its recorded performance is patient-agnostic. This leads to the feature sets of a total of 2394 segments (1197 normal vs. 1197 artifactual ones) being used for training, and a total of 532 feature sets (266 normal vs. 266 artifactual ones) being used for testing.

At this point, a second data split is introduced in the training set. It is important to consider that in a field application, the classifier would never have access to the exact same network segments in both normal and artifactual variations. For this reason, only half, or 1197 training sets are used, the first 599 of which correspond to feature sets of a normal network segment, while the other 598 correspond to different artifactual ones. As a result, for each experimental setup, both classifier configurations are trained on 1197 distinct training feature sets and tested on 532 testing feature sets.

Finally, the performance of each configuration for a certain experimental setup is calculated as the percentage of correct network segment classifications observed for each configuration when applied to the respective testing dataset.

4.3 Results

4.3.1 Kolmogorov–Smirnov and Mann–Whitney U Test Results

With 16 experimental setups and 15 features extracted from the network segments of each setup, a total of 240 feature distributions are produced corresponding to artifactual network segments, alongside 15 feature distributions corresponding to normal network segments. For the outliers with a μ of: $outliermean = \pm 2 \times max|amplitude|$, 211 out of the total 255 distributions displayed a statistically significant difference from a normal distribution after being standarised, rejecting the null hypothesis with p-value < 0.05. For outliers with a μ of $outliermean = \pm 4 \times max|amplitude|$, 212 out of the 255 distributions rejected the null

hypothesis at p-value < 0.05. Taking into consideration the non-Gaussian nature of most feature distributions, the Mann–Whitney U test is selected to compare in pairs, the feature distributions extracted from artifactual network segments with the feature distributions extracted from the corresponding normal network segments.

Since in each experimental setup, only one of the channels contains outliers, out of the 15 total features extracted per network segment, only eight of these features were extracted from channel combinations that include the “artifactual” channel. As a result, it is expected that in each experimental setup these eight feature distributions will display a statistically significant difference when compared to the distributions of features extracted from the respective normal segments. No significant difference is expected for the seven feature distributions that do not include the "artifactual" channel.

As expected, for both categories of outlier distributions with different μ magnitude, and for all experimental setups, all feature distributions extracted from a combination of channels containing outliers have a statistically significant difference to the original feature distributions, rejecting the null hypothesis with p values $< 10^{-12}$. No statistical difference is observed between pairs corresponding to channel combinations that did not include artifactual outliers.

4.3.2 Disruption of DisEn values across Experimental Setups

The experimental setups presented in this study contain networks within which one channel contains artifactual outliers. As a result, while network features that are not extracted from network segments containing outliers remain unaffected, the values of rest of the features display significant MPD, as highlighted by the following results. The Sections 4.3.2–4.3.2 present the results for outliers with both $outliermean = \pm 2 \times max|amplitude|$ and $outliermean = \pm 4 \times max|amplitude|$. The results from the first outlier configuration are discussed in detail while second configuration tends to follow similar patterns, but with increase in the overall values of MPD, which is expected considering the increased deviation of the mean outlier amplitude from the physiological amplitude range.

Setups with EEG Outliers

The MPD values for setups with EEG outliers are shown in Figure 4.1. For the univariate feature, the MPD ranges from a minimum value of 17.9% with a σ of 15.8% observed at a P factor of 0.1%, to a maximum value of 60.6% with a σ of 20% observed at a P factor of 5%. Bivariate features display a maximum MPD of 31.5% and a σ of 11.8% observed for the feature extracted from synchronised segments of the EEG and RESP channels at a P factor of 5%, while for the rest of the features, the MPD values do not surpass 16.6%. It is important to note the continuous decrease of MPD observed as the number of channels forming the network segments from which a feature is extracted, increases.

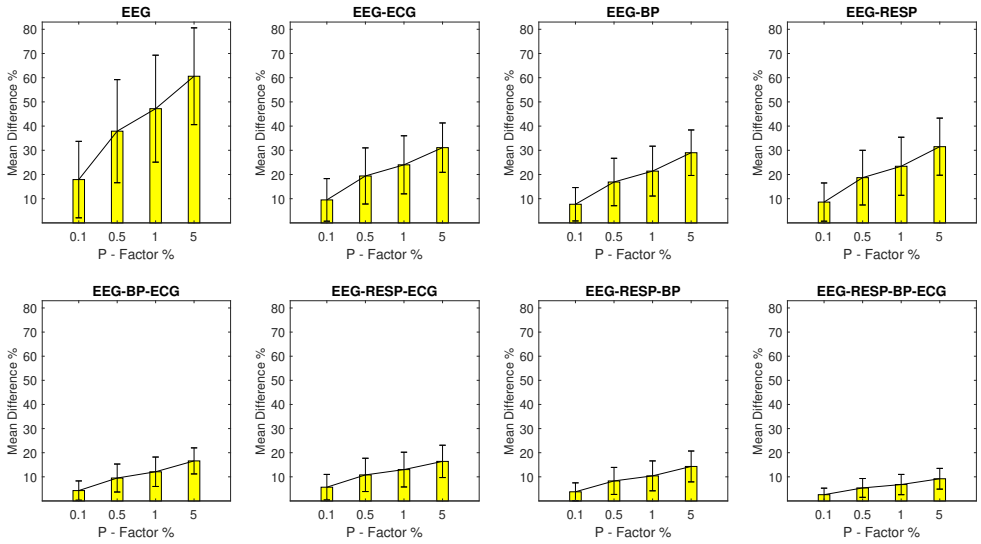

Figure 4.1: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the EEG channel of the network contains a percentage of outliers determined by the corresponding P-factor. These results correspond to experimental setups with $outliermean = \pm 2 \times max|amplitude|$.

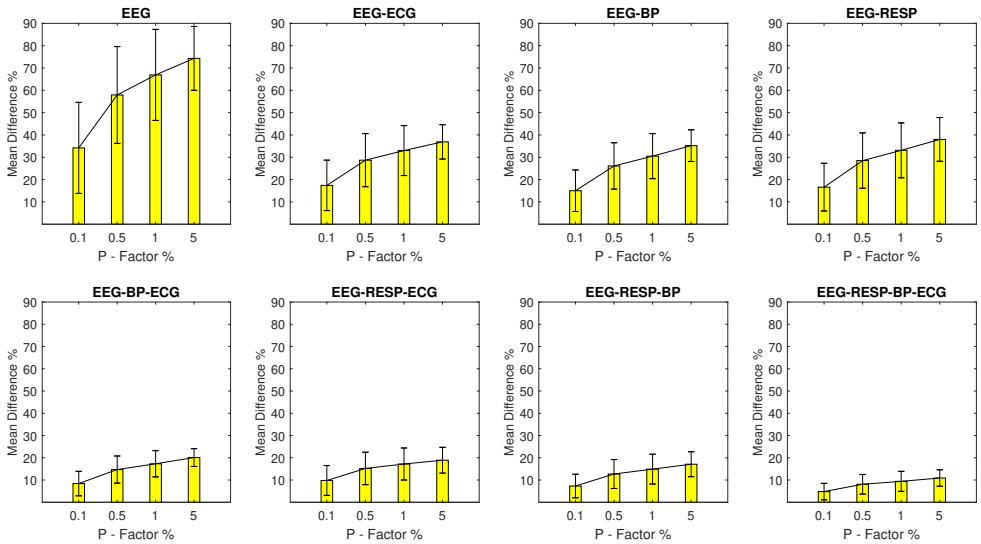

Figure 4.2: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the EEG channel of the network contains a percentage of outliers determined by the corresponding P-Factor. These results correspond to experimental setups with $outliermean = \pm 4 \times max|amplitude|$.

Setups with RESP Outliers

For setups with RESP outliers, a noteworthy change is observed in the results shown in Figure 4.3. Significant but substantially smaller MPD values are observed for the univariate feature when compared to the EEG outlier setups, with a minimum MPD of 5% and a σ of 7.5% observed for a P factor 0.1% and a maximum MPD of 20.5% and a σ of 14.2% observed for a P factor of 5%. The bivariate features follow closely with a maximum MPD of 14.7% and a σ of 8.2% observed for the feature extracted from synchronised segments of the RESP and ECG channels, at a P factor of 5%, while the MPD values for the rest of the features do not surpass 7.6%.

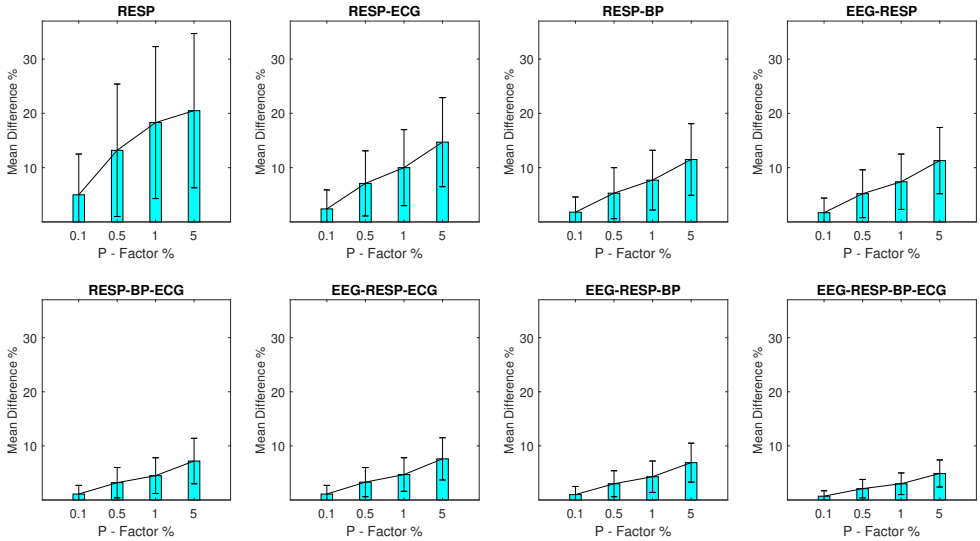

Figure 4.3: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the RESP channel of the network contains a percentage of outliers determined by the corresponding P-factor. These results correspond to experimental setups with $outliermean = \pm 2 \times max|amplitude|$.

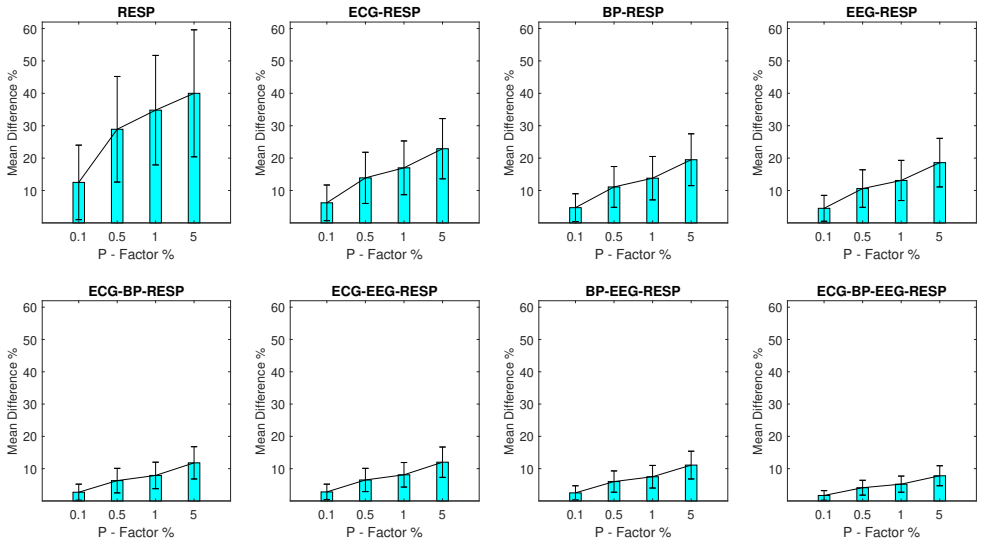

Figure 4.4: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the RESP channel of the network contains a percentage of outliers determined by the corresponding P-Factor. These results correspond to experimental setups with $outliermean = \pm 4 \times max|amplitude|$.

Setups with BP Outliers

The MPD values for experimental setups with BP outliers are shown in Figure 4.5 and seem to follow a similar pattern to the one observed for the EEG and RESP setups. The BP univariate feature displays significant value disruption with a minimum MPD of 11.3% and a σ of 4.6% for a P factor of 0.1% increasing to a MPD of 48.2% with a σ of 6.4% for a P factor of 5%. For bivariate features, a maximum MPD of 24.9% and a σ of 3.6% are observed for the feature extracted from synchronised segments of the BP and RESP channels at a P factor of 5%, while the rest of the features follow with MPD values that do not exceed 13.1%.

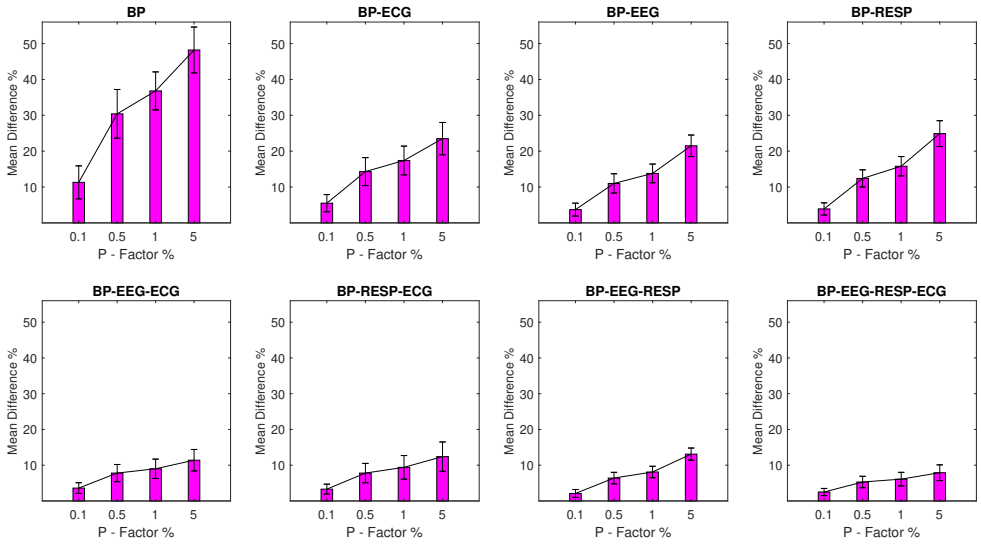

Figure 4.5: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups, where the BP channel of the network contains a percentage of outliers determined by the corresponding P-factor.These results correspond to experimental setups with $outliermean = \pm 2 \times max|amplitude|$.

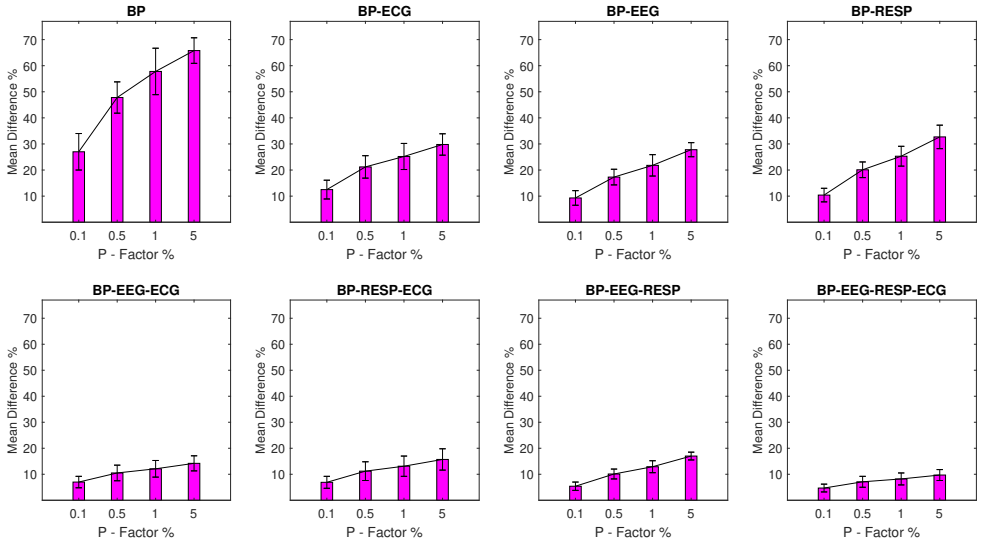

Figure 4.6: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the BP channel of the network contains a percentage of outliers determined by the corresponding P-Factor. These results correspond to experimental setups with $outliermean = \pm 4 \times max|amplitude|$.

Setups with ECG Outliers

Finally, the MPD values for experimental setups with outlier samples contained in the ECG channel are shown in Figure 4.7 following the same pattern. The univariate feature extracted from the ECG channel contains the highest MPD range with a minimum value of 23.2% and a σ of 11.7% for a P factor of 0.1%, increasing to 60.1% with a σ of 9.1% for a P factor of 5%. Bivariate features follow with a significant reduction in disruption. Their maximum MPD is observed in the case of the feature extracted from synchronised segments of the ECG and RESP channels with value of 29% and a σ of 7.1% observed at a P factor of 5%. The rest of the features follow with MPD values that are lower than 14.2% across all respective setups. Finally, of particular interest is the profile followed from the univariate ECG feature for experiments with $outliermean = \pm 4 \times max|amplitude|$. Moving from a P factor value of 1% to 5% shows a decline in the MPD value. This pattern is similar to the results reported in Sections 3.3.3 and 3.3.3 and discussed in 3.4.4 with regards to dispersion patterns that are mapped to fluctuations arising between outlier and normal samples.

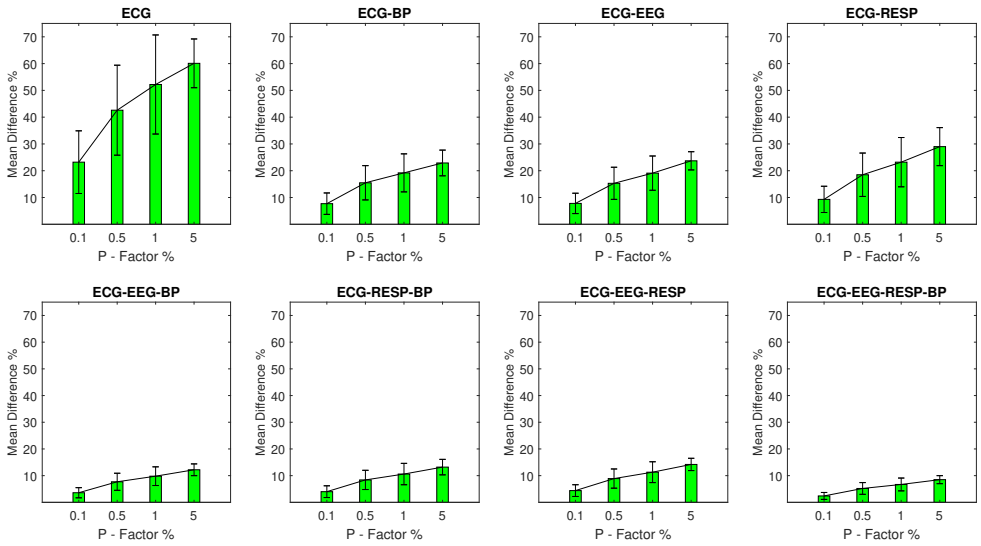

Figure 4.7: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the ECG channel of the network contains a percentage of outliers determined by the corresponding P-factor. These results correspond to experimental setups with $outliermean = \pm 2 \times max|amplitude|$.

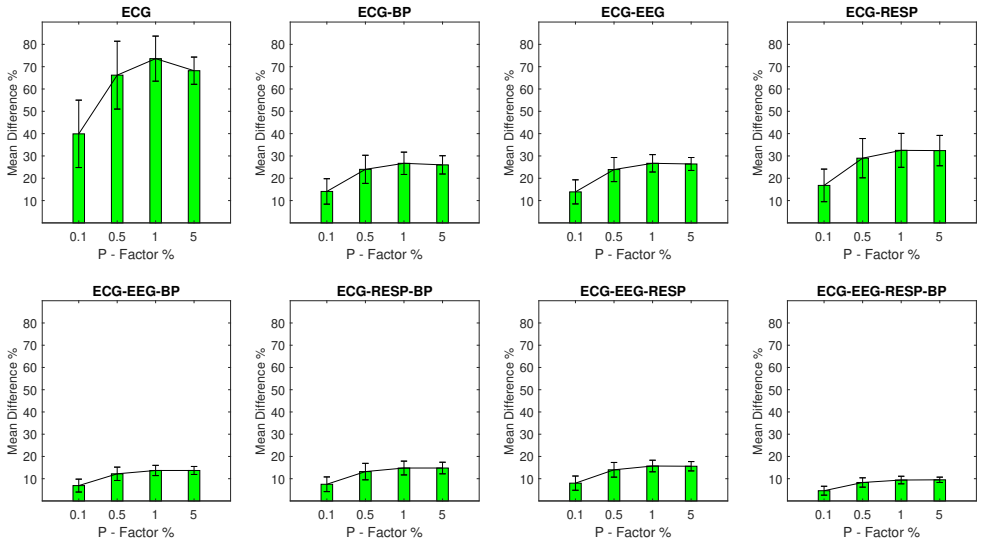

Figure 4.8: The μ and σ of the percentage difference are shown for each artifactual feature distribution across experimental setups where the ECG channel of the network contains a percentage of outliers determined by the corresponding P-Factor. These results correspond to experimental setups with $outliermean = \pm 4 \times max|amplitude|$.

4.3.3 Network Segment Classification Results

As indicated by the MPD results, outliers have a significant effect in the values of the extracted features, with the univariate feature having the largest deviations from the original values. It would therefore be important to verify whether these deviations can be used to detect artifactual network segments, in the case of the univariate classifier, and whether the inclusion of multivariate features improves performance or introduces disruptive noise. The performances of both the univariate and multivariate classifiers are reported for the respective experimental setups which are grouped together based on the channel of the network containing outliers, following the same format as Section 4.3.2 with the results for outliers with $outliermean = \pm 2 \times max|amplitude|$ discussed in detail. The results for $outliermean = \pm 4 \times max|amplitude|$ indicate that the increased amplitude difference leads to increased percentages of correct classification across all experimental setups.

Classification Performance with EEG Outliers

In the case of EEG outliers, a pattern of performance improvement is observed when moving from univariate to multivariate classification, as shown in Table 4.2. Initially, the multivariate classifier significantly outperforms the univariate one with correct classification percentages of 88.7% compared to the univariate 70.3% for a P factor 0.1% and of 97.2% compared to 88.5% for a P factor of 0.5%. Eventually, the univariate and multivariate classifiers reach equivalent performance levels for P factors of 1% and 5%.

Table 4.2: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups, with outliers located in the EEG channel of the network $outliermean = \pm 2 \times max|amplitude|$.

EEG		
P-factor	**Univariate**	**Multivariate**
0.1%	70.3%	88.7%
0.5%	88.5%	97.2%
1%	97.7%	98.5%
5%	99.1%	99.1%

Table 4.3: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups, with outliers located in the EEG channel of the network and $outliermean = \pm 4 \times max|amplitude|$.

EEG		
P-factor	**Univariate**	**Multivariate**
0.1%	84.2%	96.6%
0.5%	98.9%	99.2%
1%	99.2%	98.9%
5%	99.6%	99.6%

Classification Performance with RESP Outliers

For experimental setups with RESP outliers, limited effectiveness is initially observed for both classifiers, while a significant classification improvement (20% increase) is achieved when moving from the univariate to the multivariate model for a P factor of 5%. As shown in Table 4.4, both classifiers display performance that does not surpass a percentage of correct classifications of 56% for a P factor of 0.1%. However, for the P factor of 5%, the multivariate model achieves a percentage of 96.2% of correct classifications, significantly outperforming the univariate one with a respective percentage of 76.5%.

Table 4.4: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the RESP channel of the network and $outliermean = \pm 2 \times max|amplitude|$.

RESP		
P-factor	**Univariate**	**Multivariate**
0.1%	55.8%	55.6%
0.5%	64.1%	67.5%
1%	72.6%	71.2%
5%	76.5%	96.2%

Table 4.5: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the RESP channel of the network and $outliermean = \pm 4 \times max|amplitude|$.

RESP		
P-factor	**Univariate**	**Multivariate**
0.1%	63.7%	67.1%
0.5%	83.8%	76.3%
1%	89.1%	78.8%
5%	90.8%	95.1%

Table 4.6: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the BP channel of the network and $outliermean = \pm 2 \times max|amplitude|$..

BP		
P-factor	**Univariate**	**Multivariate**
0.1%	94%	99.1%
0.5%	100%	99.6%
1%	99.4%	100%
5%	100%	100%

Classification Performance with BP Outliers

For experimental setups with BP outliers, both univariate and multivariate classifiers achieve similar and effective performance as displayed in Table 4.6. The correct classification percentages are in the range of 94% to 100% for the univariate classifier and above 99% for the multivariate one. In this case, artifactual segments are detected even in significantly low percentages of outliers, indicating that the channel in which the outliers are located plays an important role in the correct classification of the corresponding network segments.

Table 4.7: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the BP channel of the network and $outliermean = \pm 4 \times max|amplitude|$.

BP		
P-factor	**Univariate**	**Multivariate**
0.1%	99.4%	99.4%
0.5%	100%	100%
1%	100%	100%
5%	100%	100%

Table 4.8: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the ECG channel of the network and $outliermean = \pm 2 \times max|amplitude|$.

ECG		
P-factor	**Univariate**	**Multivariate**
0.1%	61.8%	68.6%
0.5%	94.9%	74.4%
1%	97%	80.8%
5%	100%	95.7%

Table 4.9: Percentage of correct network segment classifications for univariate and multivariate classifiers when tested on experimental setups with outliers located in the ECG channel of the network and $outliermean = \pm 4 \times max|amplitude|$.

ECG		
P-factor	**Univariate**	**Multivariate**
0.1%	95.5%	73.1%
0.5%	100%	91%
1%	100%	97.2%
5%	99.8%	100%

Classification Performance with ECG Outliers

Finally, in the case of ECG outliers, a different pattern is observed when moving from a univariate to a multivariate classifier. As shown in Table 4.8, for a P factor of 0.1%, a performance boost is noted from 61.8% of correct classifications for the univariate classifier, increasing to 68.6% for the multivariate one. However, for larger P factors, the univariate classifier is constantly outperforming the multivariate one with substantial effectiveness, considering the correct classification percentage range of 94.9% to 100% as opposed to the multivariate performance range of 74.4% to 95.7%. This indicates that after a certain threshold of outliers in the network, the univariate classifier is effective, while the addition of multivariate features adds noise that substantially reduces the corresponding performance of the multivariate one.

4.4 Discussion

As part of this study, the disruptive capacity of channel-specific outliers in the values of all possible univariate and multivariate DisEn features extracted from corresponding network segments was quantified. Each network segment consists of four synchronised channel segments: EEG, RESP, BP, and ECG, resulting in a total of 16 experimental setups, with each setup being defined by the channel containing the artifactual outliers and the percentage of samples set as outliers as specified by the corresponding P factor, with possible values being: 0.1%, 0.5%, 1%, and 5%. Furthermore, for all 16 experimental setups, a univariate and a multivariate logistic regression classifier is trained and tested for the detection of artifactual network segments, with the percentage of correct segment classifications being reported for each setup (in Tables 4.2–4.9).

4.4.1 Robustness of Multivariate Network Features to Univariate Outliers

Based on the results presented in Section 4.3.2, a pattern can be observed in the recorded MPD values. The multivariate features have significantly lower values of MPD from the correct feature values, when compared to the corresponding MPD of the univariate feature for each experimental setup. This was particularly noticeable for setups with high P factor values, such as 1% or 5%, when outliers are present in the EEG and ECG channels, with the univariate MPD surpassing the corresponding multivariate MPD by at least 23%. Furthermore, the MPD becomes lower as the number of channels forming the network segments from which a feature is extracted, increases. The feature extracted from all four available channels in a network segment has consistently the lowest MPD values across all experimental setups. Finally, in the case of experimental setups with outliers in the RESP channel, while the same pattern persists, the reduction of the value of the MPD when moving from univariate to multivariate features is significantly smaller. This further indicates that signal-specific characteristics affect the disruptive capacity of outliers, a subject that is discussed in further detail in Section 4.4.2.

When comparing the operation of univariate and multivariate DisEn, a core difference that provides a relative robustness to multivariate features when compared to univariate ones concerns the number of quantised samples formulating the embedded vectors that are mapped to dispersion patterns. For the univariate DisEn algorithm, as indicated in Section 2.4.2, the embedding dimension (m) defines the number of quantised samples used to create dispersion patterns [35, 36]. In this study, $m = 3$ for both the univariate and multivariate algorithms. However, as mentioned in Section 2.5.3, in the case of multivariate features, the number of quantised samples formulating a multivariate embedded vector that is then mapped to multiple dispersion patterns increases with the number of channels used for its extraction. As a result, the effect of outlier samples in the formulation of the corresponding dispersion patterns is reduced, leading to a reduction in their capacity to disrupt the calculation of the respective DisEn value.

Therefore, while the outliers still led to a significant disruption in the distribution of values for the corresponding multivariate features, as indicated by the results of the Mann–Whitney U test reported in Section 4.3.1, the relative robustness of multivariate features when compared to univariate ones indicates a potential advantage of multivariate methodologies for respective applications. It would be worthwhile to expand on this study with the replication of equivalent experiments utilising other entropy quantification algorithms, such as the univariate and multivariate variations of PEn [34, 76] and SampEn [38, 69] to verify whether the robustness of multivariate features to univariate outliers is persistent across different entropy quantification methodologies.

4.4.2 Disruptive Effect of Outliers Across Physiological Signals

As indicated by the results presented in Section 4.3.2 and mentioned in Section 4.4.1, significant changes in the effect of outliers on the value of the extracted univariate features are observed, based on the channel containing the corresponding outliers. The disruption is more significant when outliers are present in the EEG and ECG channels, with slightly smaller disruption observed when outliers are present in the BP channel and significantly smaller for outliers in the RESP channel. A similar pattern was observed in the experiments presented in Chapter 3 as discussed in Section 3.4.4. When outliers are present, they tend to disrupt the process of allocating classes across the amplitude range of the segment [35, 39] by significantly expanding it based on the outliers' amplitudes. As a result, the amount of classes allocated in the physiological range is significantly reduced, leading to a reduction of dispersion patterns representing physiological dynamics in the input segment, and therefore an overall reduction of the calculated DisEn values. Signals such as the ECG [142] and EEG [135] contain higher-frequency components leading to rapid fluctuations in the amplitude of each signal, relative to the BP [143] and especially the RESP [133, 134] signals. As a result, the observed decrease in DisEn values differs from channel to channel, based on the signal formulating it, resulting in the observed difference in MPD magnitudes.

It is important to notice that the same pattern is not observed to the same degree in the case of multivariate features. As discussed in Section 4.4.1, the multivariate features display a certain robustness to the effect of outliers based on the number of channels formulating the network segments from which they are extracted. No major deviations in the values of MPD are observed among the same multivariate features across different experimental setups where the channel containing the outliers changes. These findings indicate that the potential combination of univariate DisEn variations modified to be robust to outliers, similarly to the ones presented in Chapter 3, with the current multivariate DisEn variation (mvMDE) [39] could provide an effective entropy quantification interface for the extraction of features from multi-channel time-series that contain artifactual outliers.

4.4.3 Performance Comparison in Artifactual Segment Detection

The classification performance observed across experimental setups and presented in Section 4.3.3, indicates promising results for the detection of artifactual network segments. However, strong differences that should be taken into consideration are observed in the performance of the univariate and multivariate classifiers based on the channel containing the artifactual outliers. In the case of EEG and RESP experimental setups, the multivariate classifier outperformed the univariate one, while significant performance differences were observed between the two groups of experimental setups. In the case of the EEG channel, significant performance was achieved by the multivariate classifier even for low P-factor values, such as 0.5%, while the univariate classifier reached similar levels of performance for P-factor values of 1% and 5%. However, the RESP outliers proved much more challenging in the detection of the corresponding network segments, with only the multivariate classifier achieving effective performance at a P factor of 5%. It is important to consider that the reduced performance in the case of the experimental setups with RESP outliers was expected, especially in the case of the univariate classifier, when considering that the corresponding outliers had significantly lower disruptive capacity when compared to experimental setups with outliers in other channels, as quantified in Section 4.3.3. In the case of BP outliers, no significant performance differences were noted between the univariate and multivariate models with both classifiers displaying effective performance, as mentioned in Section 4.3.2.

However, the case of ECG outliers highlights an important challenge when deploying architectures that utilise multiple multivariate features. For P factor values of 0.5% and higher, the univariate model is not just highly effective at classifying network segments, but also outperforms the multivariate one, indicating that in this case, the multivariate features add noise that reduces the performance of the multivariate model. This highlights the necessity of utilising a machine-learning architecture that would be robust to potential noise added by the utilisation of multiple features, through effective feature selection [144–146]. This is particularly important for future applications of Network Physiology aiming not just at the assessment of data quality, but also at the extraction of physiological insights from networks [114, 147, 148], since in those types of applications, the most informative features will be harder to detect due to the dynamics of interest having the potential to occur at any level of multivariate interaction, as opposed to starting from outliers which initially occur at one of the channels during recording.

4.4.4 Limitations of Current Study and Future Work

Throughout the experimental setups of this study, it is important to note that, as mentioned in Section 4.2.4, only one channel at a time contained artifactual outliers. This design choice for the experiments of the study was made in order to prioritise experimental setups that would provide insights, not just about the effect of artifactual outliers in the quantification processes of DisEn, but also about how these effects differ based on which channel of the time-series contains the outliers. Consequently, this study could be further expanded through the conduction of experiments where outliers are present in more than one channel at a time, this is expected to lead to higher MPD observed for multivariate features, and therefore, even better classification performance in the detection of artifactual network segments when utilising a multivariate classifier.

Furthermore, while the simulated artifactual outliers presented in this study were in alignment with previous research in the field [40, 61], it would be important to replicate this study and especially the classification of artifactual network segments, utilising data sets with annotated real-world artifactual segments to further assess the applicability of the method as a deployed data quality control tool.

Finally, as indicated by the results presented in Section 4.3.3 and discussed in Section 4.4.3, the logistic regression classifier was not able to appropriately utilise the multivariate features in the case of experimental setups with ECG outliers, leading to reduced performance when compared to the univariate one. Therefore, both for the purpose of artifactual segment detection and for future applications of Network Physiology, it would be important to expand on this study by designing and implementing classification architectures capable of effective feature selection.

4.5 Conclusion of the Chapter

This study investigated and quantified the effect of artifactual outlier samples in the accuracy of univariate and multivariate DisEn features extracted from network segments consisting of four synchronised channels. Furthermore, it presented a proof-of-concept artifactual segment detection tool deployed in univariate and multivariate configurations using the corresponding extracted features.

The results indicate that the distribution of values for each feature extracted from a network segment containing artifactual outliers is significantly altered. The largest magnitude of disruption is observed in univariate features with an MDP value in the range of 20–48% for most experimental setups, while the multivariate features display a relative robustness, which increases based on the number of channels from which they are extracted. The feature extracted from all four available channels in a network segment displays an MDP value that remains

lower than 10% across experimental setups. The classification results of the study indicate that the univariate classifier performance surpasses 90% of correct segment classifications for the majority of experimental setups. A strong exception are the setups with outliers in the RESP channel where a performance of 90% correct segment classifications is surpassed only with the multivariate classifier for a P factor of 5%. The multivariate classifier outperforms the univariate one in setups with EEG and RESP outliers, but underperforms when compared to the univariate one in the case of ECG outliers. These results highlight the importance of using a machine learning architecture capable of effective feature selection when moving from univariate to multivariate analysis.

Finally, the changes observed both in terms of the percentage differences and the classification effectiveness when comparing across experimental setups in which outliers are present in different channels indicate that, in alignment with prior research [40], the characteristics of each physiological signal should be taken into consideration when assessing the impact of outlier samples in the process of entropy quantification.

Chapter 5

The Stratified Entropy Framework and its implementation using Dispersion Entropy

Note: This Chapter's contribution have been published in *IEEE Transactions on Biomedical Engineering*, 2022 [42].

5.1 Introduction

This Chapter introduces the framework of Stratified Entropy analysis and presents three novel algorithmic variations of SmvMDE as implementations of the framework.

In Section 2.1.4, it was mentioned that the quantification of entropy – as a measure of physiological signal's variability – is of direct interest for the extraction of physiologically viable information and the potential monitoring of a system's states, particularly when considering the CSD [26–28] and LoC [10, 70] paradigms that associate changes in the measured variability with state transitions.

Chapters 3 and 4 presented research that focused on the challenge of low data quality arising from artifactual samples that are common occurrence in physiological recordings. The effect of artifacts during both univariate and multivariate entropy quantification were studied and two different approaches, that of modifying the feature extraction algorithm itself and that of partially automating the process of data cleaning, were considered [40, 41].

The research presented in this Chapter focuses on the challenge that arises during multivariate analysis due to the differences that are present between the multiple channels formulating the time-series. As discussed in Section 2.1.1, for the effective analysis of physiological dynamics, approaches should consider both univariate and multivariate analysis of multi-channel time-series. This is a necessary step to ensure that cross-channel dynamics can be quantified to allow the study of dynamics developed across different components of the same organ system as well as across distinct systems [15–19].

However, while multivariate algorithms can extract an output feature from a multi-channel time-series, the approach is limited with regards to the total information retrieved. The dynamics of certain input channels may overshadow those of others due to the potentially different dominant frequencies amongst the physiological signals in each channel. This becomes apparent when multi-channel time-series are comprised of signals that arise from heterogeneous organ systems such as the combination of ECG [142], EEG [135], arterial BP [143], and nasal RESP signals [133, 134], whose dominant frequencies and temporal structures display clear differences.

As a step towards addressing this challenge, recent studies have suggested non-uniform multiscale embedding between the input channels [149]. This approach aims to find the optimal combination of scales for the analysis of the multi-channel time-series so that each channel is analyzed at the scale where most of its dynamics would arise, by modifying the time-delay [149] or the embedding dimension used for each channel [150]. While this approach offers an interesting and modular configuration of analysis, it faces challenges that limit its applicability. These are the potential mismatch of each channel's data length with the optimal scale values, the limitation of multiscale analysis to specific scales for each channel resulting in an incomplete multiscale output, the instability of the method for increased number of channels and the potential for overshadowing to occur even at optimal scale combinations.

A different approach for the analysis of interdependencies within a group of multi-channel time-series arises from the utilization of Cross-Entropy algorithms, developed for ApEn, SampEn [38], FuzzyEn [151], and PEn [152]. With them, an entropy based feature quantifies the coupling between two channels. The variations of SampEn and FuzzyEn are non-directional, while the variations of ApEn and PEn are directional. In the latter cases, one of the two channels acts as a "designated" channel in the measurement. Thus, the potential overshadowing of each channels' dynamics could be avoided since each channel has the opportunity to be designated. However, this approach is limited to bivariate measurements between two channels. Therefore, it cannot capture higher-order dynamics arising jointly from three or more channels. A second limitation is that, by definition, it measures the coupling between the two channels and is not a measurement of their combined dynamics.

With the aim of addressing these challenges, the framework of Stratified Entropy is proposed, combining positive elements of both the Multivariate and Cross-Entropy algorithms to augment the information that can be extracted from a set of multi-channel time-series by allowing each channel's dynamics to have a different level of prioritization during the quantification of the output entropy value based on its allocation to a respective stratum. Namely, the main contributions of this chapter are:

- The introduction of the Stratified-Entropy framework as a new form of multivariate and multiscale analysis that increases the amount of information extracted from a multi-channel time-series via entropy quantification algorithms.

- The implementation of the Stratified-Entropy framework through the introduction of three novel algorithms of SmvMDE that prioritize channels during the calculation of the output entropy value based on their allocation to hierarchical strata.
- The analysis and benchmarking of the SmvMDE algorithms through experiments applied to synthetic time-series, waveform physiological time-series, and derived physiological data.

5.2 Methods

5.2.1 Stratified Entropy Framework

Within the framework of Stratified Entropy, strata are defined with a clear hierarchy of prioritization. The number of strata can vary based on the implementation of the framework. Each channel is allocated to one of the available strata and every channel has a weighted contribution in the calculation of the output entropy feature based on their allocated stratum.

Building upon the existing mvMDE algorithm [39] three novel variations of the SmvMDE are introduced: The T-SmvMDE, Soft Threshold (ST-SmvMDE), and Proportional (P-SmvMDE) variations. For both the original algorithm and the novel variations the output values are normalized in the range of 0 to 1 by dividing equation 2.22 with the term $\ln c^m$ as follows:

$$mvMDE(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^m} \frac{p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right)}{\ln c^m} \tag{5.1}$$

For the purposes of this study, all three variations have been designed based on a two strata configuration, a *core* stratum (prioritized) and a *periphery* stratum. The potential extension to configurations with higher numbers of strata is discussed in Subsection 5.3.5. The variations differ in how the dynamics of channels allocated to the core stratum are prioritised over the periphery channels.

The following subsections start with a description of the SmvMDE variations and the changes they introduce to mvMDE, and describe the experiments conducted to analyse and benchmark their operation.

5.2.2 Stratified-Dispersion Entropy Variations

Building on the original mvMDE, three variations of SmvMDE are introduced as implementations of the Stratified Entropy Framework. With their two strata configuration, the SmvMDE variations separate the channels in two sets. The set of one or more designated channels, which are allocated to the “core” stratum, and the set of secondary channels which are allocated to the “periphery” stratum.

The original mvMDE treats all embedded subvectors as equal [39]. Instead, the SmvMDE variations prioritise subvectors that contain samples retrieved from designated channels. The Threshold (T-SmvMDE), Soft Threshold (ST-SmvMDE), and Proportional (P-SmvMDE) variations use distinct approaches for adjusting the contribution of each combination by modifying the third and fourth steps of the original mvMDE algorithm, as described in Subsection 2.5.

Threshold Variation

T-SmvMDE defines the minimum number of samples extracted from designated channels that each subvector should contain in order to be considered. This is achieved through a new input parameter: the threshold (t). The initially $\binom{m \cdot p}{m}$ subvectors utilized in the case of the original mvMDE are reduced to a subset of length l_t that only includes subvectors that meet or surpass the threshold of having t or more samples in the patterns of length m. As a result, for each multivariate embedded vector $Z(j)$ only $\phi_q(j)(q = 1, \dots l_t)$ subvectors are mapped to dispersion patterns. This results in the reduction of dispersion pattern instances to $(N-(m-1)d)l_t$. Fig. 5.1 displays a diagram illustrating the T-SmvMDE subvector selection process.

For each unique dispersion pattern, their relative frequency is calculated with a modified denominator to match the reduced number of dispersion patterns:

$$p(\pi_{v_0 \dots v_{m-1}}) = \frac{\#\{j \,\big|\, j \leq N-(m-1)d, \phi_q(j) \text{ has type } \pi_{v_0 \dots v_{m-1}}\}}{(N-(m-1)d)l_t} \tag{5.2}$$

Soft Threshold Variation

As an intermediate algorithm between T-SmvMDE and mvMDE, ST-SmvMDE combines the t input parameter with the additional reduced weight (w) parameter to reduce the contribution of subvectors that do not meet the threshold of t, without removing them completely. The possible values of the w parameter range from a minimum value of 0, where the output value will match that of T-SmvMDE, and a maximum value of 1, where the output will match that of the original mvMDE algorithm, since no reduction of contribution will occur.

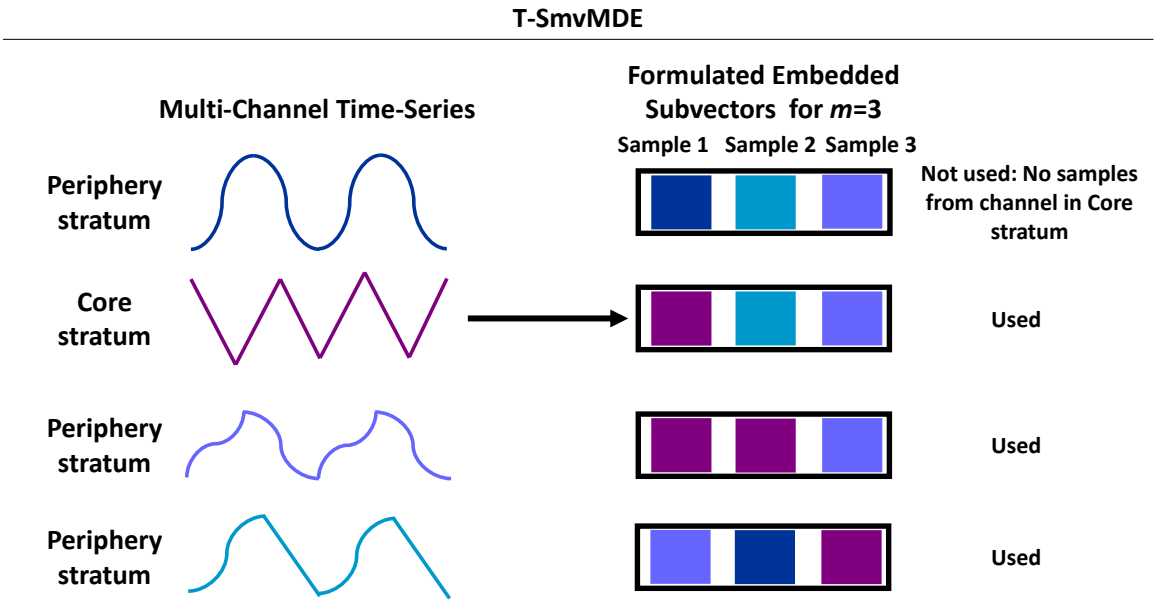

Figure 5.1: A representation of the embedded subvector inclusion and exclusion process for the T-SmvMDE variation with $m = 3$ and $t = 1$.

Based on t, the subvectors are split into two subsets: A primary subset with length l_p whose contribution to the calculation of a dispersion pattern's frequency remains unchanged; and a secondary subset with length l_s whose impact is reduced by multiplying the number of respective dispersion pattern instances with w. Consequently, for each $Z(j)$: $\phi_p(j)(p = 1, \dots l_p)$ subvectors are formulated from the primary and $\phi_s(j)(s = 1, \dots l_s)$ from the secondary subset, respectively.

Therefore, the maximum value of instances for a dispersion pattern becomes $(N - (m - 1)d)(l_p + (l_s w))$. As a result, for each unique dispersion pattern, their relative frequency is:

$$\begin{aligned} &p(\pi_{v_0 \dots v_{m-1}}) = \\ &\frac{\#\{j \,|\, j \leq N - (m-1)d, \phi_p(j) \text{ has type } \pi_{v_0 \dots v_{m-1}}\}}{(N - (m-1)d)\ (l_p + l_s w)} \\ &+ \frac{\#\{j \,|\, j \leq N - (m-1)d, \phi_s(j) \text{ has type } \pi_{v_0 \dots v_{m-1}}\}}{(N - (m-1)d)\ (l_p + l_s w)} \cdot w \end{aligned} \tag{5.3}$$

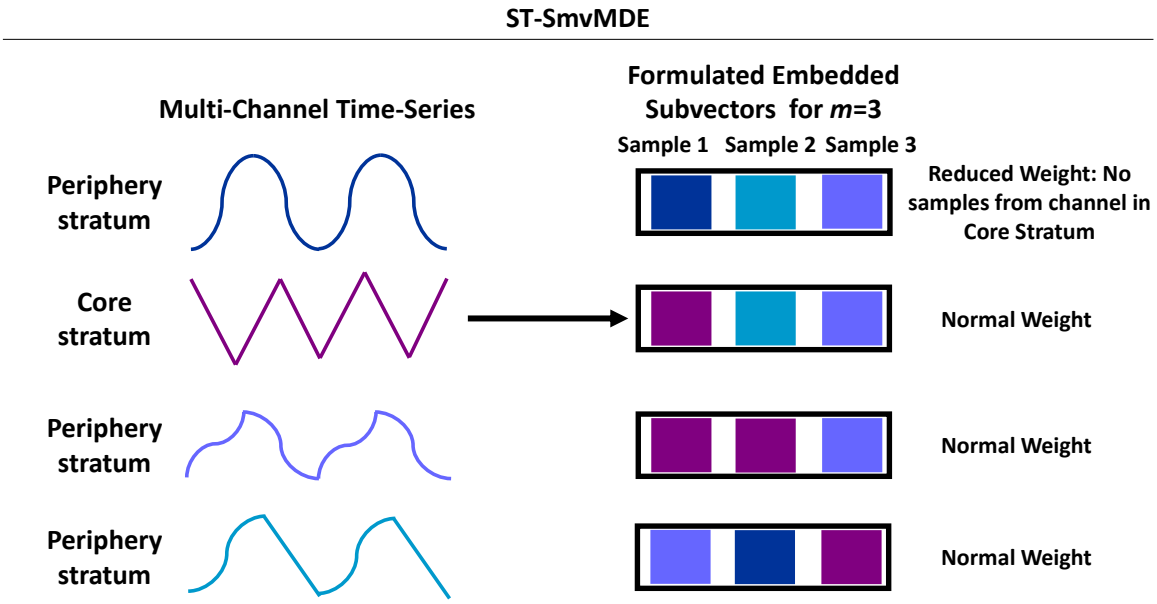

Figure 5.2: A representation of the embedded subvector inclusion and exclusion process for the ST-SmvMDE variation with $m = 3$, $t = 1$ and $w = 0.5$.

Proportional Variation

The third variation, P-SmvMDE, requires no additional parameters. Instead of utilizing a threshold to filter subvectors, it allocates them in subsets based on the number of samples contained in each combination that are retrieved from designated channels and applies a proportional factor to each category. With m being the length of each subvector and h being the number of samples extracted from designated channels, this factor is defined as $\frac{h}{m}$.

Therefore, the values of the proportional factor range from a minimum of 0 to a maximum of 1 and the total number of subsets in which the subvectors are allocated is equal to $m+1$. Consequently, for each $Z(j)$: $\phi_h(j)(h = 1, \dots l_h)$ subvectors are formulated from each subset with l_h being the length of the respective subset. Hence, the maximum value of instances (α) for a dispersion pattern becomes $\alpha = \sum_{h=0}^{m} (N-(m-1)d)\ (l_h \frac{h}{m})$.

The relative frequency of each unique dispersion pattern is calculated by counting dispersion pattern instances in subvectors of each subset multiplied by their respective ($\frac{h}{m}$) factor, divided by the maximum value of instances:

$$p(\pi_{v_0 \dots v_{m-1}}) = \frac{1}{\alpha} \cdot \sum_{h=0}^{m} \#\{j \,|\, j \leq N-(m-1)d, \phi_h(j) \text{ has type } \pi_{v_0 \dots v_{m-1}}\} \cdot \frac{h}{m} \tag{5.4}$$

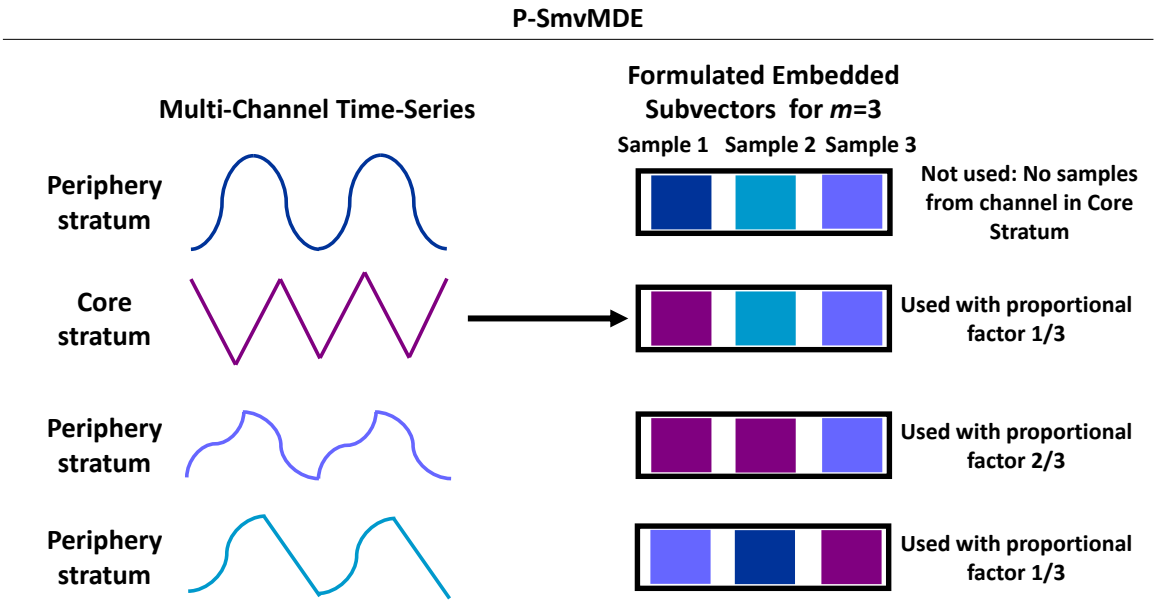

Figure 5.3: A representation of the embedded subvector inclusion and exclusion process for the T-SmvMDE variation with $m = 3$.

5.2.3 Synthetic Time-Series Experiments

The SmvMDE variations and the original mvMDE are applied to synthetic time-series, to study the differences in their operation and their multiscale outputs.

Uncorrelated white Gaussian and $1/f$ noise

Combinations of uncorrelated white Gaussian noise (WGN) and $1/f$ noise are used due to their differences in complexity and irregularity. As discussed in Subsection 2.1 complexity is expected to follow a stable multiscale profile [67, 68] and irregularity is expected to have a decreasing multiscale profile. The complexity of $1/f$ noise is higher than WGN while the irregularity of WGN is higher than $1/f$ [69, 153]. Thus, multivariate combinations of WGN and $1/f$ time-series have been used in previous research to test multiscale entropy quantification algorithms [103, 154].

To test the operation of mvMDE, all possible combinations of WGN and $1/f$ noise are formulated in 3-channel time-series, resulting in the following inputs for the experiments:

1. Three WGN channels.
2. Two WGN and one $1/f$ channels.
3. One WGN and two $1/f$ channels.
4. Three $1/f$ channels.

Table 5.1: Parameters values for Synthetic Time-Series (ST) Waveform Physiological Time-Series (WPT) and Derived Physiological Data (DPD) Experiments

Parameter (Symbol)	ST	WPT	DPD
Embedding Dimension (m)	2	3	3
Number of classes (c)	5	6	4
Time Delay (d)	1	1	1
Scale Factor Range (τ)	1 to 20	1 to 10	1
Threshold (T and ST SmvMDE) (t)	1	2	2
Reduced Weight (ST SmvMDE) (w)	0.5	0.5	0.5

Considering the operation of SmvMDE, the output entropy value will be affected to a larger degree by channels allocated to the core stratum over the periphery. This would not affect experimental setups 1) and 4). However, it would lead to different results for setups 2) and 3) which contain both WGN and $1/f$ channels based on their allocation to strata. Therefore, for SmvMDE variations, experimental setups 2) and 3) are expanded. In a first iteration, the designated channel assigned to the core is one of the WGN channels, followed by a second iteration where a $1/f$ channel is designated. This results in a total of six experimental setups for SmvMDE:

1. Three WGN channels.
2. Two WGN and one $1/f$ channels with WGN designated.
3. One WGN and two $1/f$ channels with WGN designated.
4. Two WGN and one $1/f$ channels with $1/f$ designated.
5. One WGN and two $1/f$ channels with $1/f$ designated.
6. Three $1/f$ channels.

Statistical Analysis

Each experimental setup is repeated 40 times independently and the respective μ and σ are calculated for each τ value from 1 to 10. All experimental setups are replicated for channel lengths of 15,000 and 300 samples to assess potential differences due to long versus short time-series. The parameter values used for mvMDE and SmvMDE are chosen based on the limitations introduced by the short length time-series and match those used in the original mvMDE study to allow for easy comparison between both studies [39]. They are displayed in Table 5.1.

Computational Time Experiments

To ensure that SmvMDE variations maintain the low computation time properties of the original mvMDE, 2-channel, 5-channel, and 8-channel time-series are formulated from uncorrelated WGN with channel lengths ranging from 1,000 up to 100,000 samples. Each experimental setup is repeated over 20 independent realizations and the average computation time is calculated and reported for the mvMDE and SmvMDE algorithms. For the implementation of SmvMDE algorithms, an arbitrary designated channel is selected. The computations are carried out using a PC with Intel(R) Core(TM) i7-8750H CPU @ 2.2 GHZ, 16 GB RAM running MATLAB R2018b. The parameter values of mvMDE and SmvMDE remain the same with the exception of τ_max being reduced from 20 to 10 to be consistent with [39].

5.2.4 Waveform Physiological Time-series Experiments

Experiments are conducted on waveform physiological time-series to study the extend to which the SmvMDE algorithms have increased discrimination capacity between physiological states. For this purpose, the effect size difference of output distributions extracted using SmvMDE are benchmarked to those extracted using mvMDE.

MIT-BIH Polysomnographic Database

To access multi-channel time-series formulated from high sampling rate signals recorded from different organs, the publicly available MIT-BIH Polysomnographic Database is used. It contains a total of 18 records of multiple physiological waveforms, initially recorded for the evaluation of chronic obstructive sleep apnea (OSA) syndrome and sampled at 250 Hz [125, 141].

For the purpose of this study, the records slp41 and slp45 are selected due to the availability of extensive sections of healthy stage 2 sleep; and the records slp04 and slp16 due to the existence of multiple incidents of OSA with arousal during stage 2 sleep. All records contain complete and synchronized recordings of EEG, ECG, BP, and RESP signals. The EEG signal is split into the frequency bands of: delta (0.5-3.5 Hz), theta (4-7.5 Hz), alpha (8-11.5 Hz), sigma (12-15.5 Hz), and beta (16-19.5 Hz) [105]. The selection of ranges for each frequency band is made in alignment with prior research utilizing similar polysomnographic datasets for the analysis of sleep stages [105, 111]. Hence, 8-channel time-series are extracted from each record consisting of the channels: Delta, Theta, Alpha, Sigma, Beta, ECG, BP, and RESP.

Formulation and Selection of Analysis Windows

These time-series are split into 8-channel non-overlapping windows with 7,500 samples per channel corresponding to the 30-second annotation interval of the database. Based on the annotations, 235 multi-channel "healthy" windows are extracted corresponding to healthy stage 2 sleep (slp41 = 96 windows, slp45 = 139 windows), and 235 multi-channel "apnea" windows corresponding to OSA with arousal during stage 2 sleep (slp04 = 140 windows, slp16 = 95 windows).

Calculation of DisEn

The parameter values for the extraction of multiscale entropy distributions from the 235 "healthy" and 235 "apnea" windows are chosen based on the considerations discussed in Subsection 5.3.5 and displayed in Table 5.1 under the waveform physiological time-series (PT) column. Per window, ten values are extracted, one for each τ (1 to 10).

The mvMDE is used to obtain one multiscale distribution from the "healthy" and one from the "apnea" datasets. For the effective study of SmvMDE variations (T, ST, and P), the variations are applied in eight iterations each per dataset. During each iteration a different channel is designated. This leads to the extraction of eight multiscale distributions from each dataset to study how the prioritization of each channel's dynamics affects the output entropy values and the physiological differentiation capacity of SmvMDE.

Statistical Analysis

To effectively benchmark the differentiation capacity of SmvMDE variations to mvMDE, the following steps are completed for each τ separately:

1. The Hedges'g effect size is computed [155] for the "healthy" versus "apnea" output distributions.
2. The effect size difference is calculated when moving from mvMDE to a certain SmvMDE variation with a particular designated channel.
3. The confidence intervals are estimated for each calculated effect size difference to verify their significance.

In Step 3, bootstrapping is applied to the "healthy" and "apnea" output distributions to estimate the confidence intervals. The bootstrapping is implemented by sampling with replacement the sets of 235 multiscale entropy values in each output distribution. For each output distribution of the SmvMDE variations, 40 independent realizations of bootstrapped distributions are generated. No bootstrapping is applied to the output distribution of mvMDE since the aim is to benchmark the SmvMDE distributions to the same, original mvMDE results.

To implement this analysis, the bootstrapped distributions of each SmvMDE and the original distribution of mvMDE are used in the following steps, which are applied for each designated channel selection and at each τ (1 to 10):

1. Each of the 40 bootstrapped "healthy" distributions is paired at random with one of the 40 bootstrapped "apnea" distributions. (This pairing is kept the same across all SmvMDE variations for consistency.)
2. The Hedges' g effect size is calculated between the two distributions of each pair, resulting in 40 sets of Hedges'g effect size values.
3. Hedges'g effect size values are also computed between the "healthy" and "apnea" distributions of mvMDE.
4. The benchmarking effect size values of mvMDE are subtracted from the effect size values extracted from each pair of boostrapped distributions. This results in 40 multiscale sets of effect size differences whose μ and 95% confidence intervals are calculated.

The μ and 95% confidence intervals of the effect size difference are plotted separately for each designated channel selection and τ value from 1 to 10.

5.2.5 Derived Physiological Data Experiments

The operation of SmvMDE is also studied for low-temporal resolution, derived data. The performance of SmvMDE variations is benchmarked to that of mvMDE via the difference in output entropy for separate individuals, when moving from physiological states of low to high external stress.

Maximal Exercise Dataset

For the application of SmvMDE to derived physiological data the publicly available Treadmill Maximal Exercise Test Dataset is used [22, 156]. This dataset was collected, curated, and published by the Exercise Physiology and Human Performance Lab of the University of Málaga. The recordings include five cardiorespiratory variables: heart rate (HR) (in beats per min), oxygen consumption (VO2) (in mL/min), carbon dioxide production (VCO2) (in mL/min), respiration rate (ReR) (in respiration/min), and pulmonary ventilation (VE) (in L/min). All variables were recorded in a synchronized manner with the sampling event being each breath measurement, resulting in a varied sampling period (usually in the range of 1-4 s).

Each test consisted of an individual walking and running on a treadmill, starting with a warm-up period of treadmill speeds close to 5 km/h, followed by a period of gradual speed increase that reached speeds in the range of 14 to 17 km/h, and completed with a cool-down period with speeds close to 5km/h. A total of 857 individuals participated in the study with some people having more than one test, resulting in 992 recordings. The participants' ages ranged from 10 to 63 y.o.

Formulation and Selection of Analysis Windows

Two physiological state classes are formulated: a low speed (LS) class that corresponds to data recorded during warm-up until the speed reached 7 km/h; and a high speed (HS) class that corresponds to data recorded while the treadmill speed was higher than 15 km/h. Recordings with at least 120 synchronised samples for each class were selected to ensure an adequate window size for analysis. In the few cases where an individual had more than one eligible test, the first one was selected. A total of 98 eligible recordings with age range 14 to 50 y.o. are extracted.

Calculation of DisEn

The extracted data are 98 pairs of multivariate 120-sample windows, with each pair including one segment from the LS class and one from the HS. Due to the low temporal resolution of the data and the consequent small window size, the analysis is done only at temporal scale $\tau = 1$. The selected parameter values are displayed in Table 5.1. All SmvMDE variations (T, ST, and P) are applied in five iterations each, during which a different channel is designated. Consequently, for each algorithm and designated channel selection 98 pairs of DisEn values are extracted.

Statistical Analysis

For each experimental setup and within each of the 98 pairs of DisEn values, the entropy difference observed when moving from the LS to the HS state is recorded. Boxplots are generated to compare the output difference distributions between SmvMDE and mvMDE for each designated channel selection. Additionally, the μ absolute difference observed in each difference distribution and the number of entries that displayed an increased absolute value of difference during each SmvMDE configuration, compared to their mvMDE values, are reported. Finally, to highlight a potential directionality that could match the EP hypothesis [19, 72], the number of entries with a higher entropy value in the LS state than in the HS state are also reported for each configuration.

5.3 Results and Discussion

5.3.1 Synthetic Time-Series Experiments

The results of the application of mvMDE and SmvMDE on 3-channel time-series of WGN and $1/f$ noise, are presented in Fig. 5.4 and Fig. 5.5 for univariate length of 15,000 and 300 samples, respectively. For each experimental setup, replicated for 40 independent iterations, the μ and σ of DisEn values are plotted for each τ (1 to 20).

mvMDE Operation

The operation of the mvMDE matches the patterns that have been verified by prior research [39]. As τ increases, the output entropy value has a stronger decline for the 3-channel WGN time-series. As the number of $1/f$ channels increase, the output entropy value follows a more stable profile with the 3-channel $1/f$ time-series being the most stable.

SmvMDE Operation

For experimental setups that contain solely WGN channels and $1/f$ channels respectively, the operation of all three SmvMDE variations is identical to mvMDE, as expected. In contrast in the other experiments, a stronger decline of output entropy is observed as τ increases when a WGN channel is designated. Instead when a $1/f$ channel is designated, the output follows a more stable profile for increasing values of τ.

When comparing the results of the three SmvMDE variations for the same experimental setup:

1. Using the mvMDE output values as reference, the largest deviations are observed by the P-SmvMDE variation, followed by the T-SmvMDE, and then the ST-SmvMDE variation.
2. The ST-SmvMDE outputs are between those of T-SmvMDE and mvMDE as expected by its design and the w value set to 0.5.
3. The higher deviation of the P-SmvMDE outputs from T-SmvMDE is expected when considering that for an $m = 2$ the P-SmvMDE variation gives a higher prioritization to the core stratum than the respective implementation of T-SmvMDE with $m = 2$ and $t = 1$.

Short Length Time-Series

Fig. 5.5 displays the results for the 300 sample length experiments. For all tested algorithms, the outputs follow the same patterns as their 15,000 sample length equivalent, indicating that the behaviour of SmvMDE remains the same regardless of time-series length. However, for all experimental setups, the σ values are increased, with the increase being stronger for larger τ values, as expected. Consequently, between the outputs of SmvMDE variations, overlapping can be observed between experimental setups that combine WGN and $1/f$. This indicates that during the analysis of multi-channel time-series, the sample size of the window being analyzed should be larger than the respective minimum size for mvMDE.

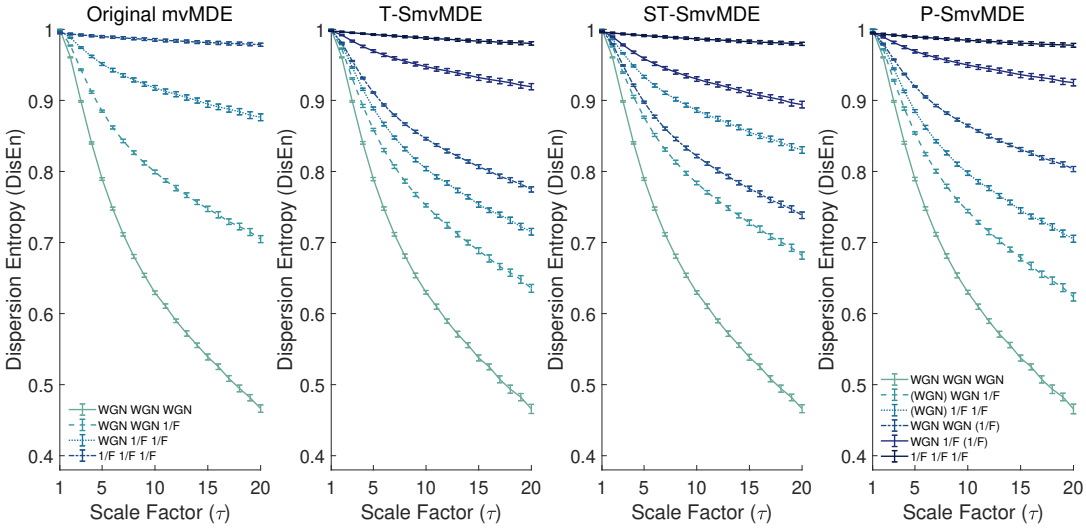

Figure 5.4: The μ and σ of output DisEn are plotted for τ values of 1 to 20 for the four experimental setups of mvMDE and the six experimental setups of SmvMDE with time-series length of 15,000 samples. For SmvMDE, the designated channel is displayed within ().

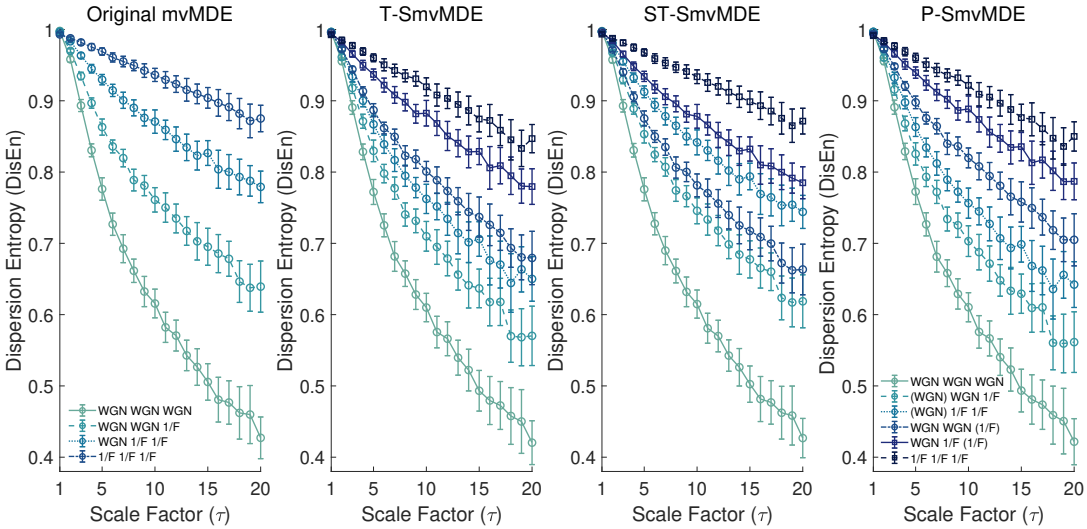

Figure 5.5: The μ and σ of output DisEn are plotted for τ values of 1 to 20 for the four experimental setups of mvMDE and the six experimental setups of SmvMDE with time-series length of 300 samples. For SmvMDE, the designated channel is displayed within ().

Table 5.2: Computational time of mvMDE and SmvMDE in seconds

Samples and Channels	mvMDE	T	ST	P
1,000 samples and 2-channels	0.024	0.025	0.026	0.026
1,000 samples and 5-channels	0.061	0.059	0.064	0.065
1,000 samples and 8-channels	0.100	0.0095	0.117	0.126
3,000 samples and 2-channels	0.067	0.069	0.069	0.070
3,000 samples and 5-channels	0.177	0.171	0.183	0.187
3,000 samples and 8-channels	0.302	0.285	0.339	0.356
10,000 samples and 2-channels	0.241	0.251	0.250	0.255
10,000 samples and 5-channels	0.649	0.631	0.663	0.685
10,000 samples and 8-channels	1.182	1.029	1.240	1.254
30,000 samples and 2-channels	1.056	1.091	1.101	1.108
30,000 samples and 5-channels	2.839	2.612	2.879	2.870
30,000 samples and 8-channels	4.789	4.526	5.044	5.044
100,000 samples and 2-channels	8.048	8.067	8.202	8.181
100,000 samples and 5-channels	20.550	20.157	20.820	20.967
100,000 samples and 8-channels	33.571	32.436	35.243	35.133

5.3.2 Computational Time

The results in Table 5.2 indicate that SmvMDE variations maintain the low computational time of the original mvMDE, as expected, since no computationally critical operations have been modified and the linear time complexity is maintained. Across all variations the main factor affecting the computation time is the univariate length of the time-series. When comparing the results for experimental setups with the same univariate length, the differences in computation time between the original mvMDE and the SmvMDE variations become more noticeable for higher number of channels.

The maximum differences in computation time are noted in the experimental setup with a time-series length of 100,000 samples and 8-channels. The maximum increase of 1.672 seconds (4.98%) is noted when moving from the mvMDE to the ST-SmvMDE algorithm while the maximum decrease of 1.135 seconds (3.38%) is noted when moving from mvMDE to T-SmvMDE. The decrease of computation time in the case of T-SmvMDE is an expected benefit due to the lower number of subvectors utilised in that variation.

5.3.3 Waveform Physiological Time-Series Experiments

The results of the statistical analysis implemented on the output entropy distributions extracted from the 235 “healthy” and 235 “apnea” 8-channel windows using T-SmvMDE and P-SmvMDE, are presented in Fig. 5.6, with each subplot corresponding to a different designated channel selection. The mean Hedges’ g effect size difference and the 95% confidence intervals are plotted for each τ (1 to 10). For clarity, only the confidence intervals that do not overlap with 0 are plotted. The benchmarking values of the mvMDE are shown in Table 5.3.

ST-SmvMDE is, by design, an intermediary variation between T-SmvMDE and mvMDE. Thus, its outputs also follow an intermediary pattern, closer to the operation of mvMDE, leading to smaller effect size differences that are available in Table 5.4.

Table 5.3: mvMDE Hedges' g effect sizes for Waveform Physiological Time-Series

	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=7$	$\tau=8$	$\tau=9$	$\tau=10$
Effect Size	0.225	0.233	0.242	0.247	0.246	0.236	0.220	0.193	0.159	0.105

Table 5.4: Mean ST-SmvMDE Hedges' G effect size difference for Waveform Physiological Time-Series. The designated channel is noted on the first column.

	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=7$	$\tau=8$	$\tau=9$	$\tau=10$
Delta	−0.002	−0.002	−0.002	−0.001	0.000	0.001	0.003	0.005	0.010	0.025
Theta	0.019	0.019	0.019	0.019	0.020	0.021	0.023	0.024	0.028	0.032
Alpha	0.003	0.003	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.011
Sigma	−0.009	−0.009	−0.009	−0.009	−0.009	−0.010	−0.011	−0.011	−0.004	0.007
Beta	0.017	0.019	0.020	0.020	0.021	0.021	0.021	0.020	0.020	0.028
ECG	0.003	0.004	0.002	−0.002	−0.005	−0.010	−0.014	−0.019	−0.022	−0.017
BP	0.016	0.017	0.018	0.018	0.016	0.017	0.018	0.017	0.016	0.020
RESP	0.022	0.022	0.025	0.026	0.027	0.028	0.029	0.029	0.033	0.038

T-SmvMDE Operation

The benchmarking of T-SmvMDE, indicates that the prioritization of the following channels leads to consistent increases in differences between the output entropy distributions extracted from "healthy" vs "apnea" windows when moving from the application of mvMDE to T-SmvMDE:

- ECG, RESP, and Alpha channels across all values of τ.
- Beta channel for τ values of 2 to 10.
- BP channel for τ values of 5 to 10.

The multiple cases of increase in effect size indicate that the T-SmvMDE variation may quantify differences between the two states that the direct application of mvMDE was not able to highlight.

P-SmvMDE Operation

The respective benchmarking results for P-SmvMDE indicate that increases in difference between the output entropy distributions are observed when prioritizing one of the following channels:

- Alpha across all values of τ.
- Beta for τ values of 3 to 8.
- ECG for τ values of 7 to 10.

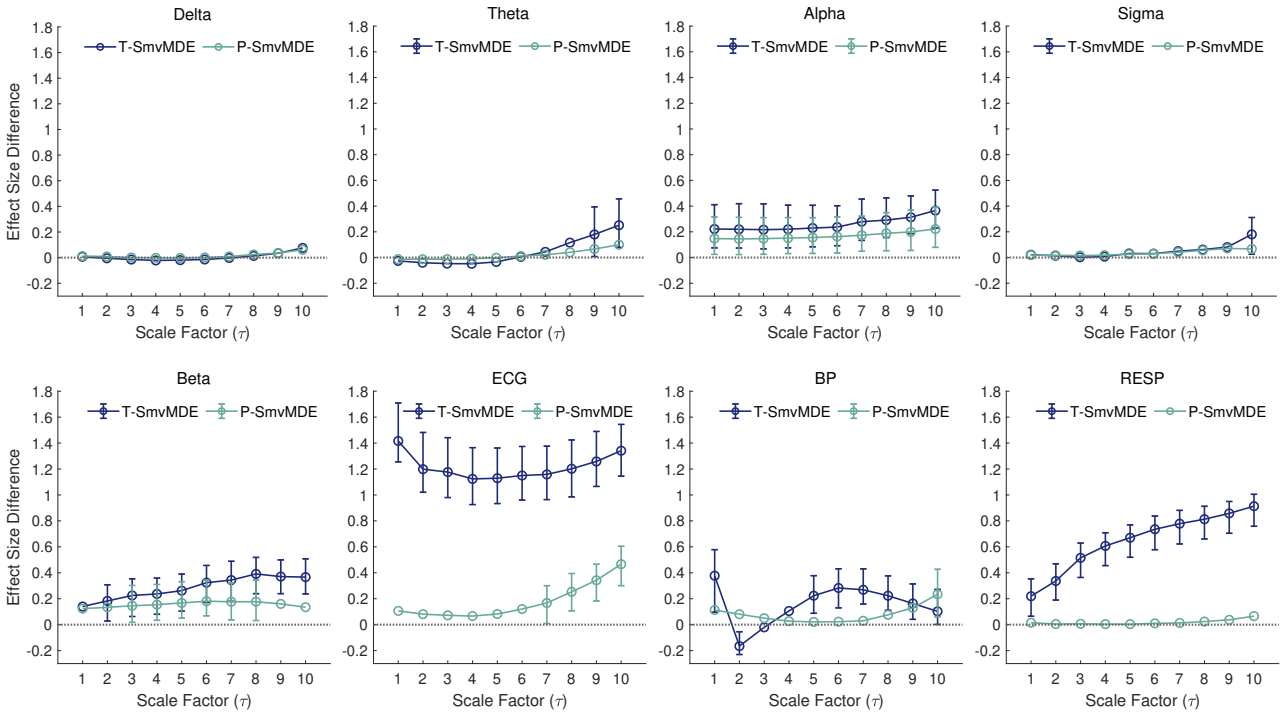

Figure 5.6: The μ and the 95% confidence intervals of the effect size difference calculated from subtracting the multiscale mvMDE Hedges' g effect sizes from those of T-SmvMDE and P-SmvMDE. Each subplot corresponds to a different designated channel selection.

Consequently, the designated channels displaying increased discrimination capacity for P-SmvMDE were also highlighted by T-SmvMDE. However, increases observed by moving to P-SmvMDE were smaller in magnitude and for fewer designated channel selections compared to T-SmvMDE. Considering the parameter values used for the SmvMDE variations in this setup, the T-SmvMDE sets a higher prioritization to the core stratum over the periphery compared to P-SmvMDE. This may indicate that this particular application benefited from implementations that defined stronger prioritization. Furthermore, within the framework of Stratified Entropy, the detection of certain prioritization cases as more effective in extracting distinct feature distributions between physiological states highlights the potential for the development of feature selection methodologies that would aim to optimize physiological classification tasks.

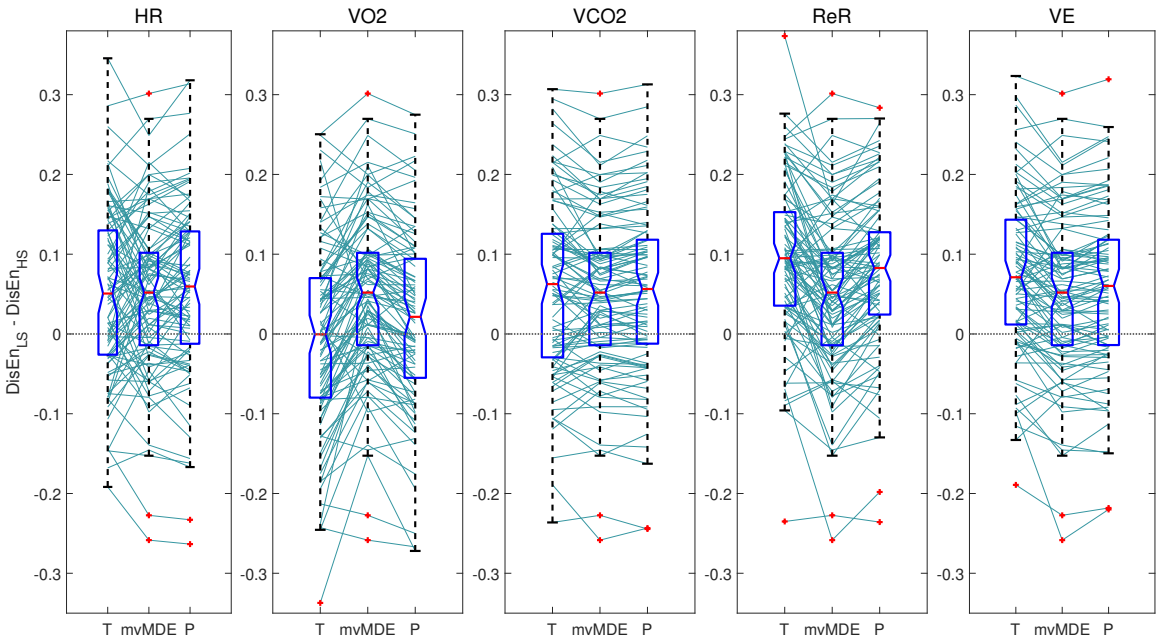

Figure 5.7: Boxplots comparing the DisEn Entropy difference calculated when moving from LS to HS using T-SmvMDE, mvMDE and P-SmvMDE. The plots correspond to different designated channel selections with the benchmarking values of mvMDE remaining the same. The lines connect the difference of the same experimental pair across the output differences of each algorithm.

5.3.4 Derived Physiological Data Experiments

The DisEn differences observed when moving from the LS to the HS state for each of the 98 exercise tests are displayed in the boxplots of Fig. 5.7. Each panel corresponds to a different designated channel selection and includes the boxplots with the distributions of differences observed through the application of mvMDE, T-SmvMDE, and P-SmvMDE. The mean absolute difference observed during the application of mvMDE is equal to 0.0894 while the respective mean absolute differences, number of entries with increased entropy difference compared to mvMDE and number of entries with larger entropy during LS versus HS are displayed for all SmvMDE and designated channel selection in Table 5.5.

Table 5.5: Mean absolute difference, Number of Entries with increased entropy difference compared to mvMDE, and Number of entries with larger entropy during LS versus HS.

T-SmvMDE	Mean Absolute Difference	Improved	Positive
HR	0.093	46	64
VO2	0.093	51	49
VCO2	0.100	68	64
ReR	0.115	66	87
VE	0.097	54	78
P-SmvMDE	Mean Absolute Difference	Improved	Positive
HR	0.097	57	71
VO2	0.086	40	51
VCO2	0.095	69	67
ReR	0.103	64	81
VE	0.093	53	69
ST-SmvMDE	Mean Absolute Difference	Improved	Positive
HR	0.089	47	71
VO2	0.087	37	67
VCO2	0.090	66	69
ReR	0.090	52	70
VE	0.090	51	69

SmvMDE Operation

When benchmarking the operation of SmvMDE to mvMDE, an improvement in differentiation capacity is noted when the VCO2, ReR, and VE channels are designated. This improvement is consistent for both T-SmvMDE and P-SmvMDE with increases in the mean absolute difference and the entries with increased LS-HS difference. The selection of HR as a designated channel displayed increased differentiation capacity for P-SmvMDE.

Similarly to Subsection 5.3.3, the majority of designated channels for which an increase in the discrimination capacity of SmvMDE is noted are common between T-SmvMDE and P-SmvMDE, indicating that while the two variations provide different ways to prioritise strata, they have the capacity of highlighting similar dynamics that were overshadowed by traditional multivariate analysis.

It is important to note that when designating the ReR channel, the largest mean absolute difference is observed for both T-SmvMDE and P-SmvMDE as well as the largest number of entries where the LS DisEn values are higher than the HS ones. This points towards a LoC process [10, 70] when moving from a steady state to a state that induces increased stress in the system, in alignment with the EP hypothesis [19, 72].

5.3.5 On the implementation of Stratified Entropy

Input Window Length

For the implementation of entropy quantification, the selection of the c and m parameters defines the minimum length of each channel within the input window. The univariate DisEn algorithm is capable of analysing short-length time-series [36] with minimum length (N) being $N > c^m \cdot \tau_{max}$. The mvMDE variation of the algorithm further improved its capacity to operate on short-length time-series due to the utilization of larger-multivariate embedding vectors compared to their univariate counterparts [39] leading to a minimum length of: $N > \frac{c^m \cdot \tau_{max}}{\binom{m \cdot p}{m}}$. For SmvMDE variations, the minimum input length is between the limits of univariate DisEn and mvMDE. As shown in Subsection 5.3.1, overlapping is observed in the short-length time-series among the large τ value outputs while analyzing the same time-series with different channels being prioritized. Consequently, the utilization of a stricter minimum is recommended, closer to the univariate DisEn: $N > c^m \cdot \tau_{max}$ when deploying SmvMDE variations.

Number of designated channels

With m being an exponent in defining the minimum input window length, it is important to consider that during the implementation of Stratified Entropy, an increased value of m might be needed when increasing the number of designated channels. In such case, it is important ensure that there are not multiple subvectors consisting entirely of samples retrieved from designated channels which would lead to them being treated equally and result to an output profile that would resemble that of mvMDE. Furthermore, while an increase in the value of m would allow additional designated channels this might not be an optimal approach, since an overshadowing of dynamics would now be possible to occur within the core stratum itself. Thus, it is recommended that the majority of Stratified Entropy applications follow a conservative approach when allocating channels to the core stratum.

Number of Strata

The total number of strata defined in a Stratified Entropy implementation affects both its design and its implementation since appropriate algorithmic steps have to be formulated for the prioritization of channels based on their strata allocation, while proper selection of parameter values is required for effective operation. For example, in the case of expanding the presented SmvMDE variations from a two to a three strata configuration, the T-SmvMDE and the ST-SmvMDE variations could be modified to operate with two different t and w (in the case of ST) values based on which strata are prioritized, while P-SmvMDE could be modified with having two tiers of proportional factors respectively. This design modification could be complemented with an appropriate increase of the m value to allow samples of varied prioritization to be included in the same subvectors similarly to the process discussed for having multiple designated channels.

However, while the expansion of the total number of strata is possible, it increases algorithmic complexity and restricts the range of effective parameter values, particularly of m as discussed above. Therefore, configurations with increased numbers of strata might be more relevant for applications that would clearly benefit from their utilization despite the increased complexity, such as for example when *a priori* knowledge exists with regards to a hierarchy of channels.

5.3.6 Limitations and Future Work

The algorithmic variations illustrate successful implementations of the Stratified Entropy framework, with effective prioritization of the channels allocated to the core stratum over the periphery, and the extraction of novel features. However, it is important to expand its implementation using additional entropy quantification algorithms, such as PEn, to acquire a more complete perspective on the utility that the framework offers. Furthermore, due to its capacity to be applied in a modular manner and with low computational cost, it would be worthwhile to combine Stratified Entropy with other variations of entropy algorithms to target specific applications. Examples include its integration with the aforementioned non-uniform multiscale embedding to incorporate *a priori* knowledge, with optimal scale selection for each channel, or the utilization of a fuzzy membership function in DisEn [157].

The results in the experiments of derived physiological data indicate a directionality in agreement with the LoC paradigm and the EP hypothesis. Hence, it would be important to replicate the analyses in other datasets and study the capacity of SmvMDE to quantify the directionality of EP phenomena. Moreover, the combination of SmvMDE with machine learning would allow for physiological state classification and prediction tasks. Consequently, strata allocations should be selected with appropriate justification or through effective feature selection processes to avoid data dredging or overfitting.

5.4 Conclusion of the Chapter

The framework of Stratified Entropy was introduced and three algorithmic variations for its implementation were presented. Stratified Entropy allows the extraction of features that would not be accessible through traditional multivariate entropy analysis by allowing the prioritization of certain channels' dynamics over others' based on the allocation of channels to different strata. The SmvMDE variations significantly extend mvMDE through the inclusion of algorithmic steps that prioritize samples extracted from channels in the prioritized core stratum during the calculation of the entropy value.

The results from the application of SmvMDE to time-series consisting of uncorrelated WGN and $1/f$ noise indicate that the variations successfully prioritize the dynamics of the designated channel. The low computation time profile of the original mvMDE variation is maintained due to no computationally critical steps being modified. When applying the SmvMDE vari-

ations to 8-channel waveform physiological time-series, certain SmvMDE features produce distributions with higher statistical difference between healthy versus OSA sleep of stage 2. While increases in group-wise statistical differences do not imply improved discrimination capacity at the subject level, the results indicate that certain applications favor the use of SmvMDE over mvMDE particularly when the stratification of a time-series' channels is of interest. The respective results from low temporal resolution derived physiological data further highlight the increased discrimination capacity of SmvMDE and its potential use to detect the directionality of "entropy pump phenomena".

The presented framework is flexible with regards to the number of channels allocated to the prioritized stratum and the total number of strata. Furthermore, it can be extended to other entropy quantification algorithms and combined with machine learning. Consequently, with appropriate algorithmic design and parameter configuration, the framework of Stratified Entropy could provide novel and effective methodologies for the extraction of viable physiological information.

Chapter 6

Summary, Limitations of Study and Future Work

6.1 Summary

The increased availability of physiological time-series due to advancements in physiological recording technology provides a unique opportunity for the extraction of viable information that could aid both in the effective prognosis as well as the treatment of pathology through early stage intervention, personalised treatment, and improved clinical decision making [1–8].

When considering the characteristics of these time-series it is important to note the potential non-linear nature of their dynamics [9–14], the existence of multivariate dynamics arising from interactions of multiple components from different organ systems [15–19], and the artifactual samples produced due to the recording conditions [2, 20–22, 25]. Consequently, the effective extraction of information from these time-series requires the utilization of algorithms that have been designed while taking into consideration these characteristics.

For this purpose, entropy quantification algorithms have been utilized in a number of applications due to their capacity of measuring the variability of physiological time-series [35, 36, 84, 158], which has been associated with the detection of physiological state changes based on the CSD [26–28] and LoC [9, 10, 29, 30] paradigms as well as the EP hypothesis [19, 51, 72]. Within the scope of this Thesis the DisEn algorithm is selected as the foundation of the study due to its favorable characteristics and the existence of both univariate and multivariate multiscale variations of the algorithm [35, 36, 39].

The work of this Thesis was conducted with the aims of investigating how the aforementioned characteristics of physiological time-series affect the operation of the DisEn algorithms and developing novel algorithmic variations and frameworks for their improved utilization. The analysis of physiological time-series was conducted in three levels. That of univariate, multivariate, and stratified multivariate analysis as presented in Chapters 3, 4, and 5, respectively.

The work presented in Chapter 3 [40] provided insights and quantified the disruptive capacity of artifactual samples in the calculation of DisEn values from RR, EEG and RI time-series through a series of experiments, with different percentages and groupings of missing and outlier samples. The results of this work highlighted outlier artifactual samples as having the highest disruptive capacity in alignment with prior research [61]. Differences were detected with regards to the effect of artifactual samples based on the type of artifact being studied and the time-series in which the artifacts were simulated. Furthermore, novel variations of the univariate DisEn algorithm were introduced and their performance was tested and benchmarked to that of the original DisEn algorithm in an effort to make preliminary steps towards addressing the challenge of low-data quality through the introduction of algorithmic steps that aim to reduce their disruptive capacity. The effect of missing samples was maintained at values lower than 8% with the introduction of SkipDisEn while the significant disruptive effect of outliers was limited to error percentages values lower than 22% with the introduction of the DynSkipDisEn variation as opposed to the range of 57% to 73% for the original DisEn algorithm.

The work of Chapter 4 [41] extended the investigation with regards to the effect of outlier artifactual samples to the level of multivariate analysis. The original mvMDE algorithm was tested on multi-channel network segments formulated from temporal windows of synchronised physiological recordings assimilated from sensors monitoring different organ systems (EEG, ECG, BP, and RESP). Compared to univariate features, multivariate DisEn features displayed a certain robustness to outliers due to the multivariate embedding vector formulation process, while the channel in which outliers were located affected the outliers' disruptive capacity. The disruptive capacity of outliers was increased when the channel containing outliers had higher frequency components such as the ECG and EEG compared to BP and RESP were outliers had a relatively reduced impact. Furthermore, an artifactual network segment detection tool using a simple logistic regression classifier was implemented and tested in two configurations. The first configuration utilized solely univariate DisEn features while the second utilized both univariate and multivariate features to study differences in the classification accuracy. Again, the channel containing the artifactual outliers determined which of the two configurations displayed improved performance. The classification accuracy surpassing 95%, for a number of experimental setups, indicated the potential for utilizing the original DisEn algorithm's sensitivity to outliers as a detection measure to partially automate the data cleaning process.

Finally, Chapter 5 [42] introduced the Stratified Entropy framework and three novel SmvMDE algorithms for its implementation. The framework was introduced with the aim of providing a new form of analysis that would allow the prioritization of certain channels' dynamics over others' based on their strata allocation, in order to extract information that was inaccessible through multivariate analysis. In traditional multivariate analysis the dynamics of certain input channels may overshadow those of others due to the potentially different dominant frequen-

cies amongst the physiological signals in each channel. By allowing the prioritization of certain channels over others, SmvMDE algorithms allow each channel to have its dynamics being prioritized based on which strata allocation configuration is used at a time. The SmvMDE algorithms were tested and benchmarked to the original mvMDE algorithm through experiments conducted in synthetic time-series, waveform physiological time-series, and derived physiological data. The results of the presented work indicated that the algorithms were successfully prioritizing samples based on the strata allocation of their respective channel, were capable of extracting novel features some of which had increased discrimination capacity when comparing between different physiological states and could potentially be used for hypothesis testing through the utilization of respective strata allocations based on a priori knowledge, such as in the case of the EP hypothesis.

6.2 Limitations of Study and Future Work

This section describes limitations and opportunities for future work that correspond to the entire scope of the Thesis. Chapter 3, 4, and 5 also contain chapter specific subsections that discuss cases that are specific to the scope of each Chapter.

The work conducted in this Thesis utilized publicly available datasets. This is beneficial with regards to the ease of experimental reproducibility; however it is limited from the fact that the utilized datasets were not tailor made for the designed experiments. With regards to the work of Chapters 3 and 4, the datasets used contained a variety of physiological time-series from different organ systems and the artifactual samples were simulated in alignment with the simulation practices of prior research [60, 61]. However, it would be beneficial to replicate the experiments in datasets that contain real world artifacts and have been designed with the aim of developing solutions to address this challenge as mentioned in the respective Sections 3.4.6 and 4.4.4. An example of such dataset is the one provided by the Great Ormond Street Hospital for the Alan Turing Data Study Group [159] under the prerequisite that appropriate labels of artifactual sections are provided. In Chapter 5, a variety of synthetic and real world data were utilized to both analyse the operation of the SmvMDE algorithms and test them in different applications for the extraction of features with increased discrimination capacity. In this case, further insights with regards to the EP hypothesis could be extracted by appropriate strata allocations during the analysis of a dataset containing both regulated and effector variables.

Moving from univariate (Chapter 3) to multivariate analysis (Chapter 4) and later in this Thesis to stratified multivariate (Chapter 5) analysis, a significant increase is noted with regards to the potential features that can be extracted from the channels of an input time-series. For example in the case of Chapter 4, where the input time-series contained four different channels, using solely univariate analysis allowed the extraction of four features, one per

channel, while multivariate analysis increased that number to fifteen, since features from all possible combinations could be extracted. In the case of Chapter 5 the potential features increase even more when considering that at the preprocessing level, the EEG channel was split into different frequency bands moving from four to eight channels but also for each multivariate feature, multiple variations of the same feature could potentially be extracted based on which channel was allocated to the prioritized stratum. This provides opportunities for the measurement of newly accessible dynamics and the improvement of physiological state classification. At the same time however, it increases feature dimensionality [160]. Consequently, further development and implementation of an artifactual segment detection tool as the one presented in Chapter 4 or the utilization of the Stratified Framework presented in Chapter 5 for the deployment of a physiological state classification architecture would benefit from the separation of the available data in training, validation, and testing sets and the utilization of an appropriate feature selection process. A study with this focus would have to consider the choice of the feature selection algorithm in alignment with the utilized classifier, similarly to the process followed by recent research which combined: features extracted using the refined multiscale SampEn and Bubble Entropy algorithms with adaptive lasso regression and a neural network classifier for the formulation of a predictive model for pressure injury based on abdominal temperature signals [161].

As stated in 2.6, Network Physiology has become a prominent framework for the analysis of interactions between different organ systems [18, 105]. However, the originally suggested utilization of graphs limits the framework in the utilization of bivariate features that are produced based on the combination of two input channels, since the edges of a graph are drawn between two nodes with each node representing one channel [106]. The current utilization of features extracted from more than two channels, within the scope of Network Physiology, is kept separate from its graph based architecture since there is no respective structure for their representation. To fully integrate multivariate and stratified multivariate features with Network Physiology, future research can expand the framework to utilize hypergraphs that can contain edges which connect more than two nodes [162, 163]. Such an extension would allow the direct integration of multivariate features into a respective hypergraph structure, while it would allow the study of further extensions such as the potential utilization of a multilayer hypergraph structure [164] formulated based on Stratified Entropy. Within that context each hypergraph layer can be used to represent a particular strata allocation structure with multiple strata allocations being used within a study.

As discussed in Section 2.4.2 one of the main advantages of the PEn and DisEn algorithms is that their computational time scales linearly with the size of the input data $O(N)$. This is a significant improvement over algorithms with quadratic scaling $O(N^2)$ such as ApEn and SampEn. However, when considering the utilization of DisEn within the scope of multivariate multiscale analysis it becomes clear that multiple iterations of the algorithm are required to

extract the multiple features of interest. Consequently, further computational optimizations can become a necessity for applications requiring close to real time monitoring. For this reason, the utilization of parallel computing processes is an important direction of future work, both with regards to the parallel computation of features which can be done at the implementation level but also through the optimization of the algorithm itself to run certain processes in parallel when respective hardware is available.

Finally, the work presented in this Thesis utilized the DisEn algorithms as a consistent platform that was tested, modified and expanded upon within the scope of the presented research. However, the resulting findings provide useful insights that can be implemented in different platforms of entropy quantification algorithms. Of direct interest is the implementation of the Stratified Entropy framework to the mvMPE algorithm due its similarities to the mvMDE algorithm under the prerequisite that a variation capable of formulating multivariate embedded vectors containing samples from multiple input channels is utilized. Furthermore, it would be important to investigate the implementation of the framework using the multivariate multiscale variation of SampEn since despite the quadratic scaling of its computational time, it remains one of the most widely applied entropy quantification algorithms [31].

6.3 Conclusion of the Thesis

This Thesis provided an in depth assessment on the effects of artifactual samples in the performance of univariate and multivariate entropy quantification through the lens of the DisEn algorithm. It made steps towards addressing the challenge of low data quality by investigating and providing solutions both through the approach of improving the robustness of the entropy quantification algorithm and the approach of partially automating the process of separating between valid and artifactual data segments. Furthermore, it expanded upon the existing forms of analysis by introducing the framework of Stratified Entropy and presenting an implementation of the framework through the novel SmvMDE algorithms with the aim of improving the process of measuring the variability of multi-channel physiological time-series.

Appendix A

Supplementary Figures for Chapter 3

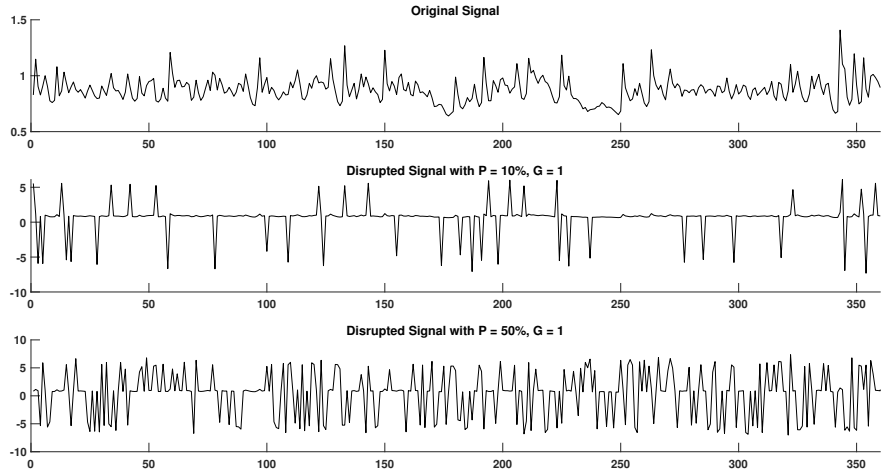

Figure A.1: Original and disrupted signal segments of RR in support of Section 3.4.4.

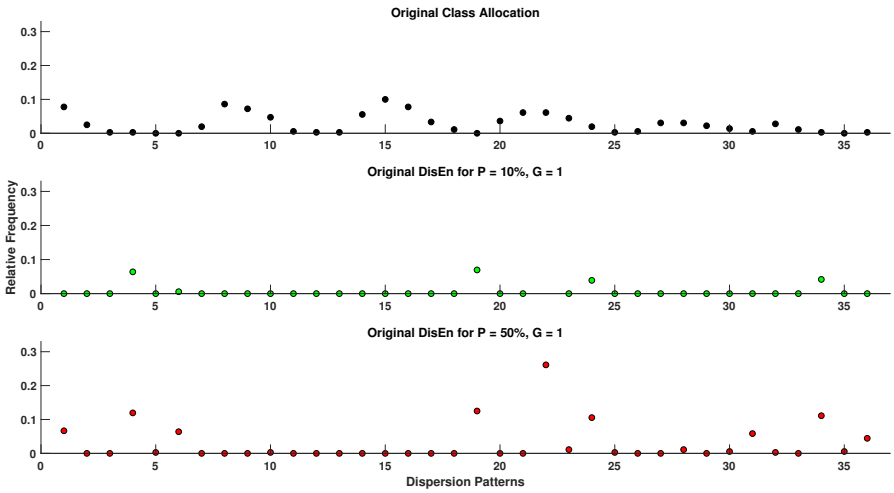

Figure A.2: Original versus disrupted dispersion patterns of RR using the original DisEn algorithm in support of Section 3.4.4.

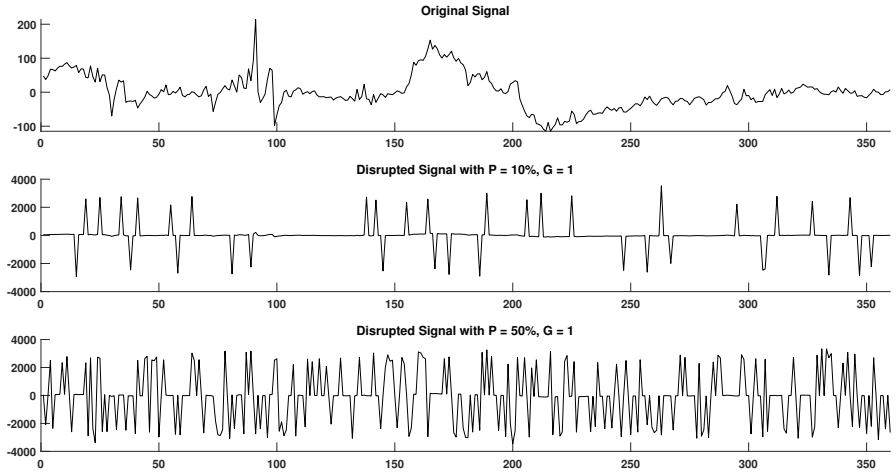

Figure A.3: Original and disrupted signal segments of EEG in support of Section 3.4.4.

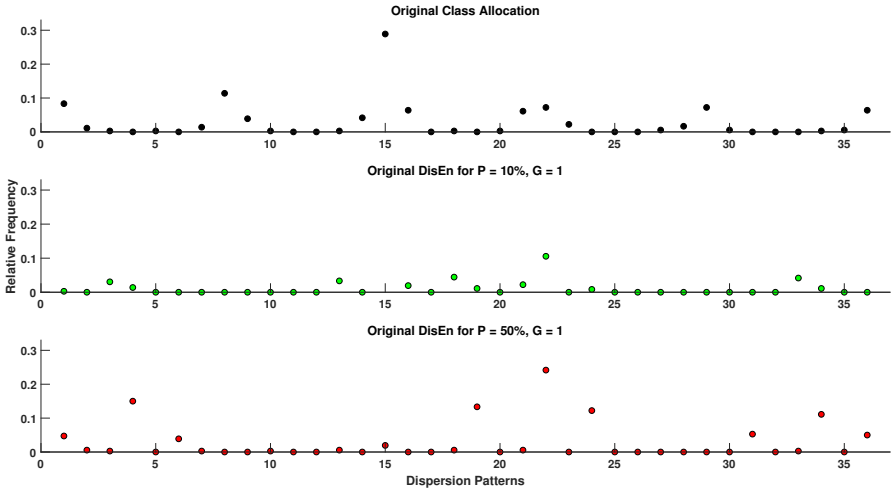

Figure A.4: Original versus disrupted dispersion patterns of EEG using the original DisEn algorithm in support of Section 3.4.4.

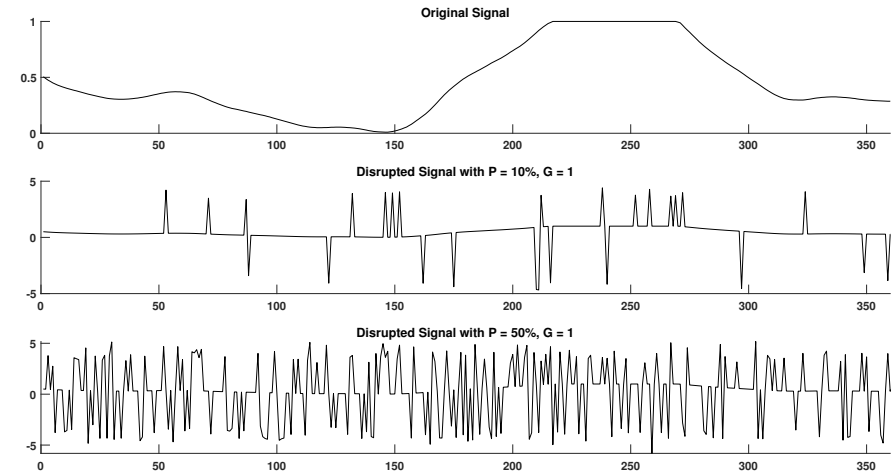

Figure A.5: Original and disrupted signal segments of RI in support of Section 3.4.4.

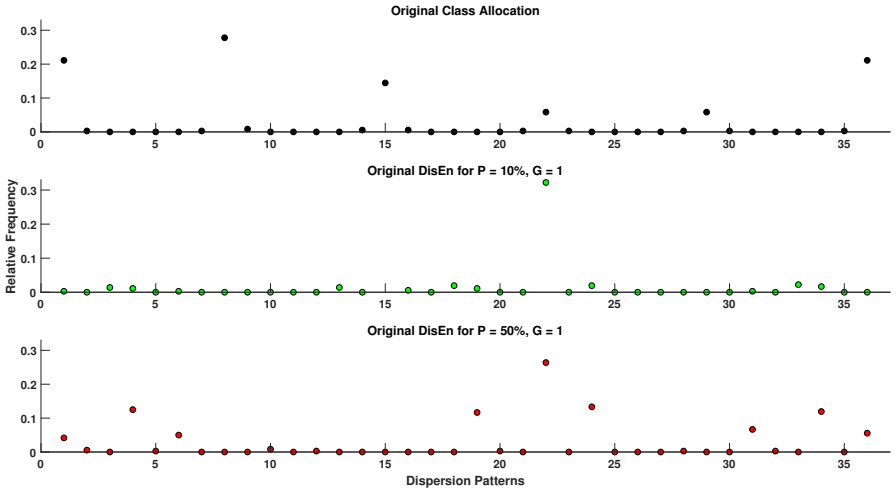

Figure A.6: Original versus disrupted dispersion patterns of RI using the original DisEn algorithm in support of Section 3.4.4.

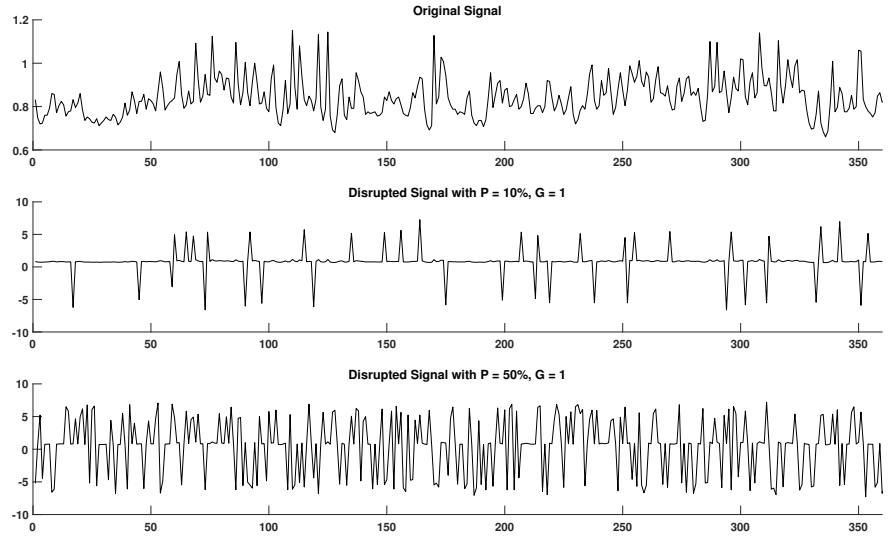

Figure A.7: Original and disrupted signal segments of RR in support of Section 3.4.5.

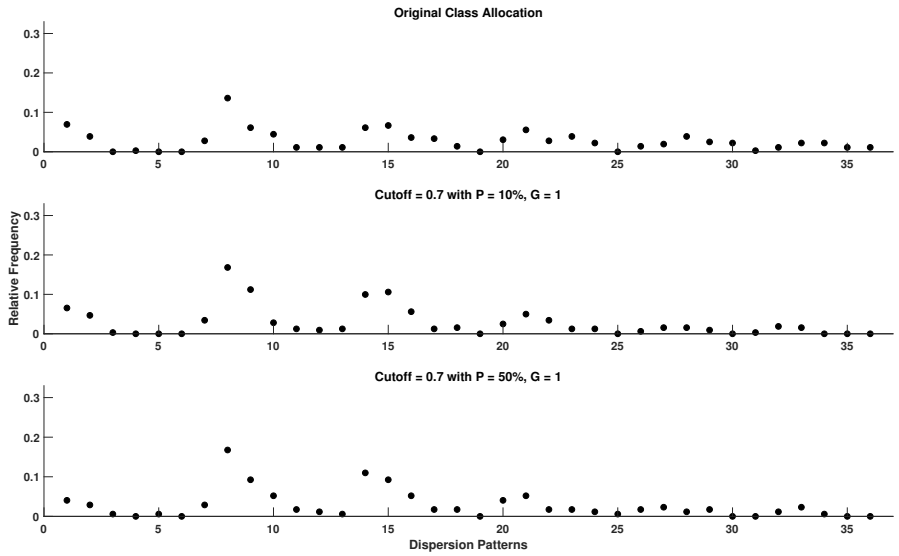

Figure A.8: Original versus disrupted dispersion patterns of RR using DynSkipDisEn with a cutoff = 0.7 in support of Section 3.4.5.

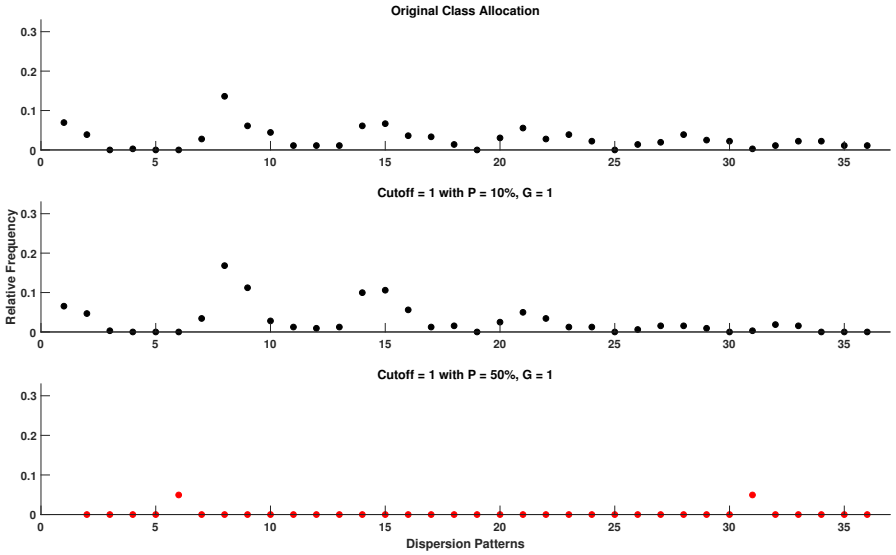

Figure A.9: Original versus disrupted dispersion patterns of RR using DynSkipDisEn with a cutoff = 1 in support of Section 3.4.5.

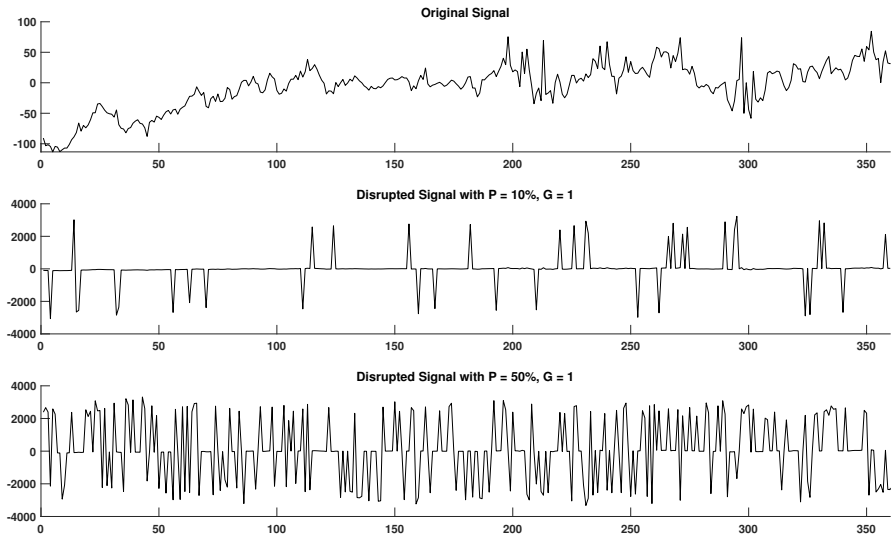

Figure A.10: Original and disrupted signal segments of EEG.

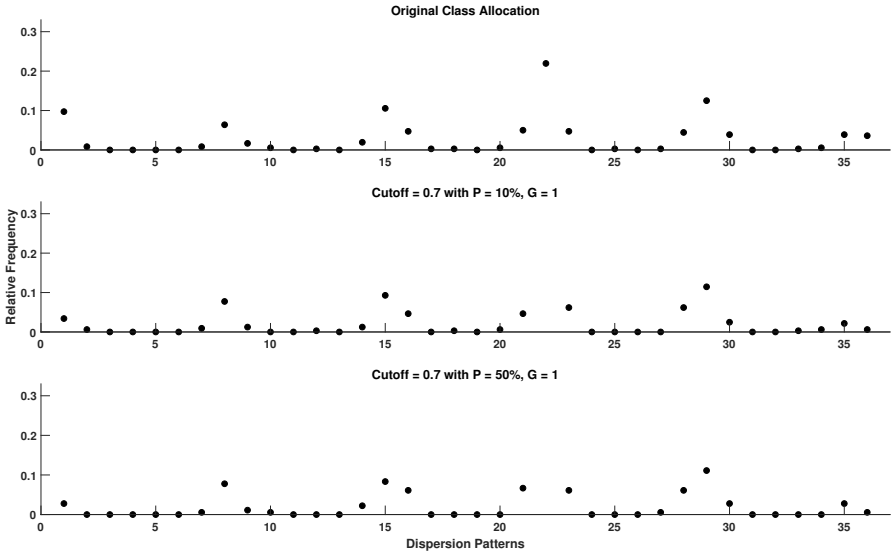

Figure A.11: Original versus disrupted dispersion patterns of EEG using DynSkipDisEn with a cutoff = 0.7 in support of Section 3.4.5.

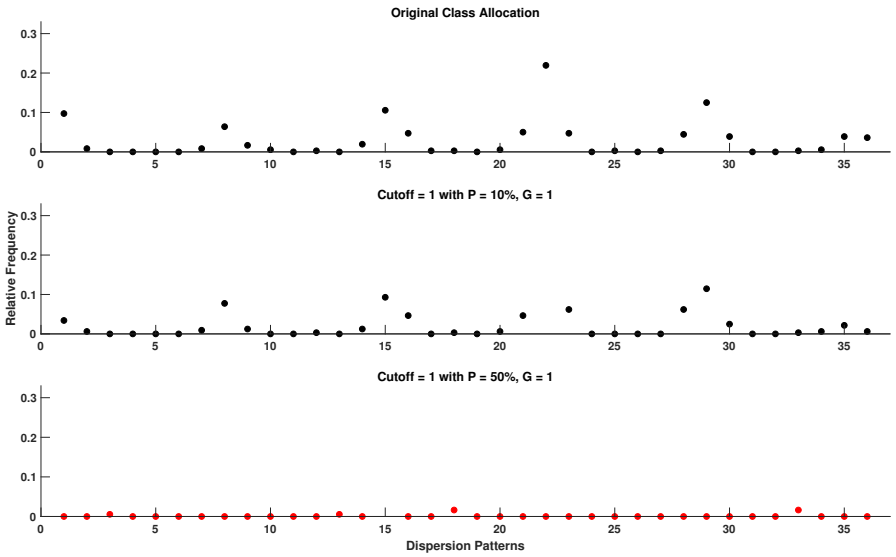

Figure A.12: Original versus disrupted dispersion patterns of EEG using DynSkipDisEn with a cutoff = 1 in support of Section 3.4.5.

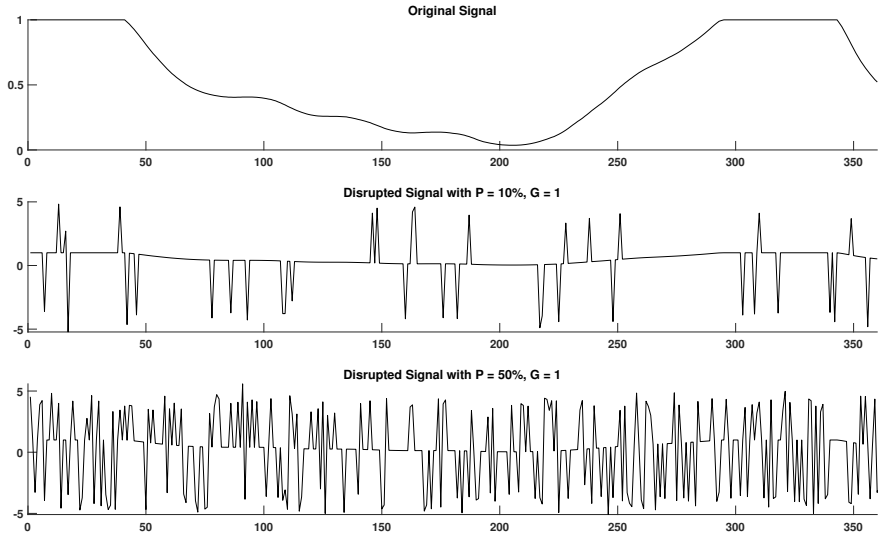

Figure A.13: Original and disrupted signal segments of RI in support of Section 3.4.5.

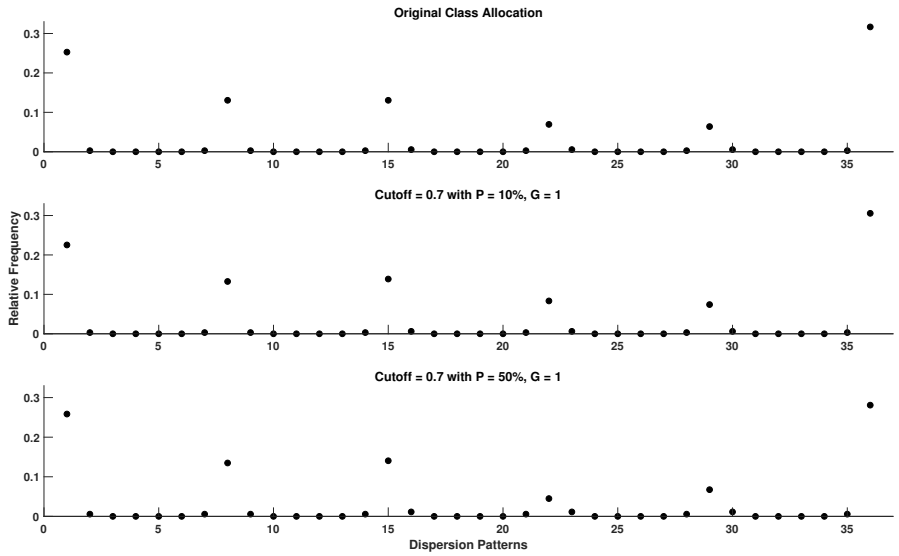

Figure A.14: Original versus disrupted dispersion patterns of RI using DynSkipDisEn with a cutoff = 0.7 in support of Section 3.4.5.

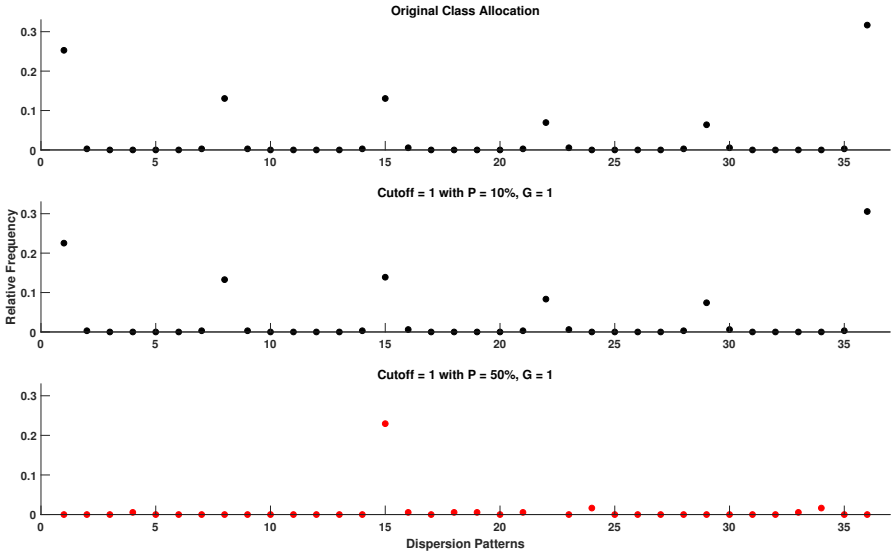

Figure A.15: Original versus disrupted dispersion patterns of RI using DynSkipDisEn with a cutoff = 1 in support of Section 3.4.5.

Bibliography

[1] D. R. Witt, R. A. Kellogg, M. P. Snyder, and J. Dunn, "Windows into human health through wearables data analytics," *Current Opinion in Biomedical Engineering*, vol. 9, pp. 28–46, Mar. 2019.

[2] C. W. Paine, V. V. Goel, E. Ely, C. D. Stave, S. Stemler, M. Zander, and C. P. Bonafide, "Systematic Review of Physiologic Monitor Alarm Characteristics and Pragmatic Interventions to Reduce Alarm Frequency: Review of Physiologic Monitor Alarms," *Journal of Hospital Medicine*, vol. 11, no. 2, pp. 136–144, Feb. 2016.

[3] J. Keim-Malpass, K. B. Enfield, J. F. Calland, D. E. Lake, and M. T. Clark, "Dynamic data monitoring improves predictive analytics for failed extubation in the ICU," *Physiological Measurement*, vol. 39, no. 7, p. 075005, Jul. 2018.

[4] B. Conroy, I. Silva, G. Mehraei, R. Damiano, B. Gross, E. Salvati, T. Feng, J. Schneider, N. Olson, A. G. Rizzo, C. M. Curtin, J. Frassica, and D. C. McFarlane, "Real-time infection prediction with wearable physiological monitoring and AI to aid military workforce readiness during COVID-19," *Scientific Reports*, vol. 12, no. 1, p. 3797, Dec. 2022.

[5] C. D. Smallwood, "Monitoring Big Data During Mechanical Ventilation in the ICU," *Respiratory Care*, vol. 65, no. 6, pp. 894–910, Jun. 2020.

[6] A. L. Fortmann, S. R. Spierling Bagsic, L. Talavera, I. M. Garcia, H. Sandoval, A. Hottinger, and A. Philis-Tsimikas, "Glucose as the Fifth Vital Sign: A Randomized Controlled Trial of Continuous Glucose Monitoring in a Non-ICU Hospital Setting," *Diabetes Care*, vol. 43, no. 11, pp. 2873–2877, Nov. 2020.

[7] D. Bustos, J. C. Guedes, M. P. Vaz, E. Pombo, R. J. Fernandes, J. T. Costa, and J. S. Baptista, "Non-Invasive Physiological Monitoring for Physical Exertion and Fatigue Assessment in Military Personnel: A Systematic Review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8815, Aug. 2021.

[8] A. Ometov, V. Shubina, L. Klus, J. Skibińska, S. Saafi, P. Pascacio, L. Flueratoru, D. Q. Gaibor, N. Chukhno, O. Chukhno, A. Ali, A. Channa, E. Svertoka, W. B. Qaim, R. Casanova-Marqués, S. Holcer, J. Torres-Sospedra, S. Casteleyn, G. Ruggeri, G. Araniti, R. Burget, J. Hosek, and E. S. Lohan, "A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges," *Computer Networks*, vol. 193, p. 108074, Jul. 2021.

[9] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," *Proceedings of the National Academy of Sciences*, vol. 99, no. Supplement 1, pp. 2466–2472, Feb. 2002.

[10] A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?" *Neurobiology of Aging*, vol. 23, no. 1, pp. 23–26, Jan. 2002.

[11] N. Scafetta, R. E. Moon, and B. J. West, "Fractal response of physiological signals to stress conditions, environmental changes, and neurodegenerative diseases," *Complexity*, vol. 12, no. 5, pp. 12–17, May 2007.

[12] R. J. Martis, U. R. Acharya, A. K. Ray, and C. Chakraborty, "Application of higher order cumulants to ECG signals for the cardiac health diagnosis," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA: IEEE, Aug. 2011, pp. 1697–1700.

[13] C. Pradhan, S. K. Jena, S. R. Nadar, and N. Pradhan, "Higher-Order Spectrum in Understanding Nonlinearity in EEG Rhythms," *Computational and Mathematical Methods in Medicine*, vol. 2012, pp. 1–8, 2012.

[14] A. Müller, J. F. Kraemer, T. Penzel, H. Bonnemeier, J. Kurths, and N. Wessel, "Causality in physiological signals," *Physiological Measurement*, vol. 37, no. 5, pp. R46–R72, May 2016.

[15] S. Cerutti, "Multivariate and multiscale analysis of biomedical signals: Towards a comprehensive approach to medical diagnosis," in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*. Rome, Italy: IEEE, Jun. 2012, pp. 1–5.

[16] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1–37, Sep. 2005.

[17] S. Cerutti, D. Hoyer, and A. Voss, "Multiscale, multiorgan and multivariate complexity analyses of cardiovascular regulation," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1892, pp. 1337–1358, Apr. 2009.

[18] R. P. Bartsch, K. K. L. Liu, A. Bashan, and P. C. Ivanov, "Network Physiology: How Organ Systems Dynamically Interact," *PLOS ONE*, vol. 10, no. 11, p. e0142143, Nov. 2015.

[19] R. Fossion, A. L. Rivera, and B. Estañol, "A physicist's view of homeostasis: how time series of continuous monitoring reflect the function of physiological variables in regulatory mechanisms," *Physiological Measurement*, vol. 39, no. 8, p. 084007, Aug. 2018.

[20] I. Azimi, T. Pahikkala, A. M. Rahmani, H. Niela-Vilén, A. Axelin, and P. Liljeberg, "Missing data resilient decision-making for healthcare IoT through personalization: A case study on maternal health," *Future Generation Computer Systems*, vol. 96, pp. 297–308, Jul. 2019.

[21] R. B. Kumar, N. D. Goren, D. E. Stark, D. P. Wall, and C. A. Longhurst, "Automated integration of continuous glucose monitor data in the electronic health record using consumer technology," *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 532–537, May 2016.

[22] G. B. Moody, "The PhysioNet/Computing in Cardiology Challenge 2010: Mind the Gap," *Comput Cardiol*, p. 13, 2010.

[23] J. P. Shivers, L. Mackowiak, H. Anhalt, and H. Zisser, ""Turn it Off!": Diabetes Device Alarm Fatigue Considerations for the Present and the Future," *Journal of Diabetes Science and Technology*, vol. 7, no. 3, pp. 789–794, May 2013.

[24] J. P. Keller, "Clinical alarm hazards: a "top ten" health technology safety concern," *Journal of Electrocardiology*, vol. 45, no. 6, pp. 588–591, Nov. 2012.

[25] K. R. Johnson, J. I. Hagadorn, and D. W. Sink, "Alarm Safety and Alarm Fatigue," *Clinics in Perinatology*, vol. 44, no. 3, pp. 713–728, Sep. 2017.

[26] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," *Nature*, vol. 461, no. 7260, pp. 53–59, Sep. 2009.

[27] M. G. M. Olde Rikkert, V. Dakos, T. G. Buchman, R. d. Boer, L. Glass, A. O. J. Cramer, S. Levin, E. van Nes, G. Sugihara, M. D. Ferrari, E. A. Tolner, I. van de Leemput, J. Lagro, R. Melis, and M. Scheffer, "Slowing Down of Recovery as Generic Risk Marker for Acute Severity Transitions in Chronic Diseases:," *Critical Care Medicine*, vol. 44, no. 3, pp. 601–606, Mar. 2016.

[28] Y. Nakazato, T. Sugiyama, R. Ohno, H. Shimoyama, D. L. Leung, A. A. Cohen, R. Kurane, S. Hirose, A. Watanabe, and H. Shimoyama, "Estimation of homeostatic dysregulation and frailty using biomarker variability: a principal component analysis of hemodialysis patients," *Scientific Reports*, vol. 10, no. 1, p. 10314, Dec. 2020.

[29] B. Manor and L. A. Lipsitz, "Physiologic complexity and aging: Implications for physical function and rehabilitation," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 45, pp. 287–293, Aug. 2013.

[30] R. Sleimen-Malkoun, J.-J. Temprado, and S. L. Hong, "Aging induced loss of complexity and dedifferentiation: consequences for coordination dynamics within and between brain, muscular and behavioral levels," *Frontiers in Aging Neuroscience*, vol. 6, Jun. 2014.

[31] M. Ribeiro, T. Henriques, L. Castro, A. Souto, L. Antunes, C. Costa-Santos, and A. Teixeira, "The Entropy Universe," *Entropy*, vol. 23, no. 2, p. 222, Feb. 2021.

[32] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. vol. 27, pp. 623–656, 1948.

[33] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*, 4th ed. Boston: McGraw-Hill, 2002.

[34] C. Bandt and B. Pompe, “Permutation Entropy: A Natural Complexity Measure for Time Series,” *Physical Review Letters*, vol. 88, no. 17, p. 174102, Apr. 2002.

[35] M. Rostaghi and H. Azami, “Dispersion Entropy: A Measure for Time-Series Analysis,” *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, May 2016.

[36] H. Azami and J. Escudero, “Amplitude- and Fluctuation-Based Dispersion Entropy,” *Entropy*, vol. 20, no. 3, p. 210, Mar. 2018.

[37] 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, Mar. 1991.

[38] 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, Jun. 2000.

[39] H. Azami, A. Fernández, and J. Escudero, “Multivariate Multiscale Dispersion Entropy of Biomedical Times Series,” *Entropy*, vol. 21, no. 9, p. 913, Sep. 2019.

[40] E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero, “Augmentation of Dispersion Entropy for Handling Missing and Outlier Samples in Physiological Signal Monitoring,” *Entropy*, vol. 22, no. 3, p. 319, Mar. 2020.

[41] E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero, “Assessment of Outliers and Detection of Artifactual Network Segments Using Univariate and Multivariate Dispersion Entropy on Physiological Signals,” *Entropy*, vol. 23, no. 2, p. 244, Feb. 2021.

[42] E. Kafantaris, T.-Y. M. Lo, and J. Escudero, “Stratified Multivariate Multiscale Dispersion Entropy for Physiological Signal Analysis,” *IEEE Transactions on Biomedical Engineering*, pp. 1–12, 2022.

[43] V. Z. Marmarelis, “Modeling methology for nonlinear physiological systems,” *Annals of Biomedical Engineering*, vol. 25, no. 2, pp. 239–251, Mar. 1997.

[44] G. D. Mitsis and V. Z. Marmarelis, “Modeling of Nonlinear Physiological Systems with Fast and Slow Dynamics. I. Methodology,” *Annals of Biomedical Engineering*, vol. 30, no. 2, pp. 272–281, Feb. 2002.

[45] C. Stam, “Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field,” *Clinical Neurophysiology*, vol. 116, no. 10, pp. 2266–2301, Oct. 2005.

[46] G. Boeing, “Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction,” *Systems*, vol. 4, no. 4, p. 37, Nov. 2016.

[47] A. Beuter, L. Glass, M. C. Mackey, M. S. Titcombe, S. S. Antman, J. E. Marsden, L. Sirovich, and S. Wiggins, Eds., *Nonlinear Dynamics in Physiology and Medicine*, ser. Interdisciplinary Applied Mathematics. New York, NY: Springer New York, 2003, vol. 25.

[48] L. Glass, “Synchronization and rhythmic processes in physiology,” *Nature*, vol. 410, no. 6825, pp. 277–284, Mar. 2001.

[49] K. Sunagawa, T. Kawada, and T. Nakahara, "Dynamic nonlinear vago-sympathetic interaction in regulating heart rate," *Heart and Vessels*, vol. 13, no. 4, pp. 157–174, Jul. 1998.

[50] L. Faes, A. Porta, G. Nollo, and M. Javorka, "Information Decomposition in Multivariate Systems: Definitions, Implementation and Application to Cardiovascular Networks," *Entropy*, vol. 19, no. 1, p. 5, Dec. 2016.

[51] H. Modell, W. Cliff, J. Michael, J. McFarland, M. P. Wenderoth, and A. Wright, "A physiologist's view of homeostasis," *Advances in Physiology Education*, vol. 39, no. 4, pp. 259–266, Dec. 2015.

[52] L. Faes, A. Porta, and G. Nollo, "Information Decomposition in Bivariate Systems: Theory and Application to Cardiorespiratory Dynamics," *Entropy*, vol. 17, no. 1, pp. 277–303, Jan. 2015.

[53] M. Zanetti, L. Faes, G. Nollo, M. De Cecco, R. Pernice, L. Maule, M. Pertile, and A. Fornaser, "Information Dynamics of the Brain, Cardiovascular and Respiratory Network during Different Levels of Mental Stress," *Entropy*, vol. 21, no. 3, p. 275, Mar. 2019.

[54] R. Croft and R. Barry, "Removal of ocular artifact from the EEG: a review," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 30, no. 1, pp. 5–19, Feb. 2000.

[55] M. Manikandan and K. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signal," *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 118–128, Mar. 2012.

[56] X. Dong, C. Chen, Q. Geng, Z. Cao, X. Chen, J. Lin, Y. Jin, Z. Zhang, Y. Shi, and X. D. Zhang, "An Improved Method of Handling Missing Values in the Analysis of Sample Entropy for Continuous Monitoring of Physiological Signals," *Entropy*, vol. 21, no. 3, p. 274, Mar. 2019.

[57] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data Cleaning: Overview and Emerging Challenges," in *Proceedings of the 2016 International Conference on Management of Data*. San Francisco California USA: ACM, Jun. 2016, pp. 2201–2206.

[58] F. Nait-Abdesselam and C. Titouna, "Data Quality Improvements for Internet of Things Using Artificial Neural Networks," in *2020 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. New Delhi, India: IEEE, Dec. 2020, pp. 1–6.

[59] C. Zhou and R. C. Paffenroth, "Anomaly Detection with Robust Deep Autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax NS Canada: ACM, Aug. 2017, pp. 665–674.

[60] E. Cirugeda-Roldan, D. Cuesta-Frau, P. Miro-Martinez, and S. Oltra-Crespo, "Comparative Study of Entropy Sensitivity to Missing Biosignal Data," *Entropy*, vol. 16, no. 11, pp. 5901–5918, Nov. 2014.

[61] A. Molina-Picó, D. Cuesta-Frau, M. Aboy, C. Crespo, P. Miró-Martínez, and S. Oltra-Crespo, "Comparative study of approximate entropy and sample entropy robustness to spikes," *Artificial Intelligence in Medicine*, vol. 53, no. 2, pp. 97–106, Oct. 2011.

[62] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artificial Intelligence Review*, vol. 33, no. 4, pp. 275–306, Apr. 2010.

[63] B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 845–869, May 2014.

[64] D. Kaplan, M. Furman, S. Pincus, S. Ryan, L. Lipsitz, and A. Goldberger, "Aging and the complexity of cardiovascular dynamics," *Biophysical Journal*, vol. 59, no. 4, pp. 945–949, Apr. 1991.

[65] C. Chen, Y. Jin, I. L. Lo, H. Zhao, B. Sun, Q. Zhao, J. Zheng, and X. D. Zhang, "Complexity Change in Cardiovascular Disease," *International Journal of Biological Sciences*, vol. 13, no. 10, pp. 1320–1328, 2017.

[66] A. d. l. Torre Luque and X. Bornas, "Complexity and Irregularity in the Brain Oscillations of Depressive Patients: A Systematic Review," *Neuropsychiatry*, vol. 07, no. 05, 2017.

[67] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Physical Review Letters*, vol. 89, no. 6, p. 068102, Jul. 2002.

[68] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, Feb. 2005.

[69] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, p. 061918, Dec. 2011.

[70] L. A. Lipsitz, "Dynamics of Stability: The Physiologic Basis of Functional Health and Frailty," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 57, no. 3, pp. B115–B125, Mar. 2002.

[71] M. Scheffer, J. E. Bolhuis, D. Borsboom, T. G. Buchman, S. M. W. Gijzel, D. Goulson, J. E. Kammenga, B. Kemp, I. A. van de Leemput, S. Levin, C. M. Martin, R. J. F. Melis, E. H. van Nes, L. M. Romero, and M. G. M. Olde Rikkert, "Quantifying resilience of humans and other animals," *Proceedings of the National Academy of Sciences*, vol. 115, no. 47, pp. 11 883–11 890, Nov. 2018.

[72] A. Barajas-Martínez, R. Mehta, E. Ibarra-Coronado, R. Fossion, V. J. Martínez Garcés, M. R. Arellano, I. A. González Alvarez, Y. V. M. Bautista, O. Y. Bello-Chavolla, N. R. Pedraza, B. R. Encinas, C. I. P. Carrión, M. I. J. Ávila, J. C. Valladares-García, P. E. Vanegas-Cedillo, D. H. Juárez, A. Vargas-Vázquez, N. E. Antonio-Villa, P. Almeda-Valdes, O. Resendis-Antonio, M. Hiriart, A. Frank, C. A. Aguilar-Salinas, and A. L. Rivera, "Physiological Network Is Disrupted in Severe COVID-19," *Frontiers in Physiology*, vol. 13, p. 848172, Mar. 2022.

[73] M. Wibral, R. Vicente, and J. T. Lizier, Eds., *Directed Information Measures in Neuroscience*, ser. Understanding Complex Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[74] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.

[75] G. A. Darbellay, "An estimator of the mutual information based on a criterion for independence," *Computational Statistics*, p. 17, 1999.

[76] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, Jul. 2012.

[77] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, vol. 898, pp. 366–381, series Title: Lecture Notes in Mathematics.

[78] M. Valente, M. Javorka, A. Porta, V. Bari, J. Krohova, B. Czippelova, Z. Turianikova, G. Nollo, and L. Faes, "Univariate and multivariate conditional entropy measures for the characterization of short-term cardiovascular complexity under physiological stress," *Physiological Measurement*, vol. 39, no. 1, p. 014002, Jan. 2018.

[79] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: what does regularity quantify?" *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 266, no. 4, pp. H1643–H1656, Apr. 1994.

[80] D. Caldirola, L. Bellodi, A. Caumo, G. Migliarese, and G. Perna, "Approximate Entropy of Respiratory Patterns in Panic Disorder," *American Journal of Psychiatry*, vol. 161, no. 1, pp. 79–87, Jan. 2004.

[81] J. Bruhn, H. Röpcke, and A. Hoeft, "Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during Desflurane Anesthesia," *Anesthesiology*, vol. 92, no. 3, pp. 715–726, Mar. 2000.

[82] N. Burioka, M. Miyata, G. Cornélissen, F. Halberg, T. Takeshima, D. T. Kaplan, H. Suyama, M. Endo, Y. Maegaki, T. Nomura, Y. Tomita, K. Nakashima, and E. Shimizu, "Approximate Entropy in the Electroencephalogram during Wake and Sleep," *Clinical EEG and Neuroscience*, vol. 36, no. 1, pp. 21–24, Jan. 2005.

[83] D. Abásolo, R. Hornero, P. Espino, J. Poza, C. I. Sánchez, and R. de la Rosa, "Analysis of regularity in the EEG background activity of Alzheimer's disease patients with Approximate Entropy," *Clinical Neurophysiology*, vol. 116, no. 8, pp. 1826–1834, Aug. 2005.

[84] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, Sep. 2002.

[85] S. Ramdani, B. Seigle, J. Lagarde, F. Bouchara, and P. L. Bernard, "On the use of sample entropy to analyze human postural sway data," *Medical Engineering & Physics*, vol. 31, no. 8, pp. 1023–1031, Oct. 2009.

[86] Y. Song, J. Crowcroft, and J. Zhang, "Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine," *Journal of Neuroscience Methods*, vol. 210, no. 2, pp. 132–146, Sep. 2012.

[87] Y. Jia, H. Gu, and Q. Luo, "Sample entropy reveals an age-related reduction in the complexity of dynamic brain," *Scientific Reports*, vol. 7, no. 1, p. 7990, Dec. 2017.

[88] A. Delgado-Bonal and A. Marshak, "Approximate Entropy and Sample Entropy: A Comprehensive Tutorial," *Entropy*, vol. 21, no. 6, p. 541, May 2019.

[89] B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Physical Review E*, vol. 87, no. 2, p. 022911, Feb. 2013.

[90] H. Azami, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, May 2016.

[91] X. Li, S. Cui, and L. Voss, "Using Permutation Entropy to Measure the Electroencephalographic Effects of Sevoflurane," *Anesthesiology*, vol. 109, no. 3, pp. 448–456, Sep. 2008.

[92] J. Li, J. Yan, X. Liu, and G. Ouyang, "Using Permutation Entropy to Measure the Changes in EEG Signals During Absence Seizures," *Entropy*, vol. 16, no. 6, pp. 3049–3061, May 2014.

[93] N. Nicolaou and J. Georgiou, "The Use of Permutation Entropy to Characterize Sleep Electroencephalograms," *Clinical EEG and Neuroscience*, vol. 42, no. 1, pp. 24–28, Jan. 2011.

[94] C. Carricarte Naranjo, L. M. Sanchez-Rodriguez, M. Brown Martínez, M. Estévez Báez, and A. Machado García, "Permutation entropy analysis of heart rate variability for the assessment of cardiovascular autonomic neuropathy in type 1 diabetes mellitus," *Computers in Biology and Medicine*, vol. 86, pp. 90–97, Jul. 2017.

[95] B. Graff, G. Graff, D. Makowiec, A. Kaczkowska, D. Wejer, S. Budrejko, D. Kozłowski, and K. Narkiewicz, "Entropy Measures in the Assessment of Heart Rate Variability in Patients with Cardiodepressive Vasovagal Syncope," *Entropy*, vol. 17, no. 3, pp. 1007–1022, Mar. 2015.

[96] X. Sun, Y. Zou, V. Nikiforova, J. Kurths, and D. Walther, "The complexity of gene expression dynamics revealed by permutation entropy," *BMC Bioinformatics*, vol. 11, no. 1, p. 607, Dec. 2010.

[97] H. Azami, M. Rostaghi, A. Fernandez, and J. Escudero, "Dispersion entropy for the analysis of resting-state MEG regularity in Alzheimer's disease," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL, USA: IEEE, Aug. 2016, pp. 6417–6420.

[98] E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero, "Application of Dispersion Entropy to Healthy and Pathological Heartbeat ECG Segments," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany: IEEE, Jul. 2019, pp. 2269–2272.

[99] M. U. Ahmed and D. P. Mandic, "Multivariate Multiscale Entropy Analysis," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, Feb. 2012.

[100] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, "Refined Multiscale Entropy: Application to 24-h Holter Recordings of Heart Period Variability in Healthy and Aortic Stenosis Subjects," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 9, pp. 2202–2213, Sep. 2009.

[101] Hamed Azami and Javier Escudero, "Coarse-Graining Approaches in Univariate Multiscale Sample and Dispersion Entropy," *Entropy*, vol. 20, no. 2, p. 138, Feb. 2018.

[102] S.-D. Wu, C.-W. Wu, S.-G. Lin, C.-C. Wang, and K.-Y. Lee, "Time Series Analysis Using Composite Multiscale Entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, Mar. 2013.

[103] H. Azami, A. Fernández, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis," *Medical & Biological Engineering & Computing*, vol. 55, no. 11, pp. 2037–2052, Nov. 2017.

[104] J. S. Fabila-Carrasco, C. Tan, and J. Escudero, "Multivariate permutation entropy, a Cartesian graph product approach," Mar. 2022, arXiv:2203.00550 [cs, math].

[105] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin, and P. C. Ivanov, "Network physiology reveals relations between network topology and physiological function," *Nature Communications*, vol. 3, no. 1, Jan. 2012.

[106] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, Mar. 2009.

[107] J. Goñi, A. Avena-Koenigsberger, N. Velez de Mendizabal, M. P. van den Heuvel, R. F. Betzel, and O. Sporns, "Exploring the Morphospace of Communication Efficiency in Complex Networks," *PLoS ONE*, vol. 8, no. 3, p. e58070, Mar. 2013.

[108] O. Sporns and J. D. Zwi, "The Small World of the Cerebral Cortex," *Neuroinformatics*, vol. 2, no. 2, pp. 145–162, 2004.

[109] P. C. Ivanov, K. K. L. Liu, and R. P. Bartsch, "Focus on the emerging new fields of network physiology and network medicine," *New Journal of Physics*, vol. 18, no. 10, p. 100201, Oct. 2016.

[110] M. Massimini, F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi, "Breakdown of Cortical Effective Connectivity During Sleep," *Science*, vol. 309, no. 5744, pp. 2228–2232, Sep. 2005.

[111] A. Lin, K. K. L. Liu, R. P. Bartsch, and P. C. Ivanov, "Delay-correlation landscape reveals characteristic time delays of brain rhythms and heart interactions," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2067, p. 20150182, May 2016.

[112] R. Rizzo, X. Zhang, J. W. J. L. Wang, F. Lombardi, and P. C. Ivanov, "Network Physiology of Cortico–Muscular Interactions," *Frontiers in Physiology*, vol. 11, p. 558070, Nov. 2020.

[113] A. Derakhshan, M. Mikaeili, A. M. Nasrabadi, and T. Gedeon, "Network physiology of 'fight or flight' response in facial superficial blood vessels," *Physiological Measurement*, vol. 40, no. 1, p. 014002, Jan. 2019.

[114] Y. Antonacci, L. Astolfi, G. Nollo, and L. Faes, "Information Transfer in Linear Multivariate Processes Assessed through Penalized Regression Techniques: Validation and Application to Physiological Networks," *Entropy*, vol. 22, no. 7, p. 732, Jul. 2020.

[115] M. Zanetti, T. Mizumoto, L. Faes, A. Fornaser, M. De Cecco, L. Maule, M. Valente, and G. Nollo, "Multilevel assessment of mental stress via network physiology paradigm using consumer wearable devices," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 4, pp. 4409–4418, Apr. 2021.

[116] M. Rostaghi, M. R. Ashory, and H. Azami, "Application of dispersion entropy to status characterization of rotary machines," *Journal of Sound and Vibration*, vol. 438, pp. 291–308, Jan. 2019.

[117] K. K. Kim, H. J. Baek, Y. G. Lim, and K. S. Park, "Effect of missing RR-interval data on nonlinear heart rate variability analysis," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 210–218, Jun. 2012.

[118] P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, Dec. 1993.

[119] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, Jul. 2013.

[120] E. J. Weber, P. C. Molenaar, and M. W. Molen, "A Nonstationarity Test for the Spectral Analysis of Physiological Time Series with an Application to Respiratory Sinus Arrhythmia," *Psychophysiology*, vol. 29, no. 1, pp. 55–62, Jan. 1992.

[121] D. Guarin, E. Delgado, and A. Orozco, "Testing for nonlinearity in non-stationary physiological time series," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA: IEEE, Aug. 2011, pp. 2671–2674.

[122] J. Pontet, P. Contreras, A. Curbelo, J. Medina, S. Noveri, S. Bentancourt, and E. R. Migliaro, "Heart rate variability as early marker of multiple organ dysfunction syndrome in septic patients," *Journal of Critical Care*, vol. 18, no. 3, pp. 156–163, Sep. 2003.

[123] P. Augustyniak, "Wearable wireless heart rate monitor for continuous long-term variability studies," *Journal of Electrocardiology*, vol. 44, no. 2, pp. 195–200, Mar. 2011.

[124] 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. 4, pp. R1078–R1084, Oct. 1996.

[125] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, Jun. 2000.

[126] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Harvard University–MIT Division of Health Sciences and Technology, Boston, MA, 2009.

[127] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, Aug. 2017.

[128] R. Mahajan, T. Viangteeravat, and O. Akbilgic, "Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics," *International Journal of Medical Informatics*, vol. 108, pp. 55–63, Dec. 2017.

[129] M. Garcia-Gonzalez, M. Fernandez-Chimeno, and J. Ramos-Castro, "Errors in the Estimation of Approximate Entropy and Other Recurrence-Plot-Derived Indices Due to the Finite Resolution of RR Time Series," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 345–351, Feb. 2009.

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

[131] R. Gilgen-Ammann, T. Schweizer, and T. Wyss, "RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise," *European Journal of Applied Physiology*, vol. 119, no. 7, pp. 1525–1532, Jul. 2019.

[132] H. Mølgaard, K. Hermansen, and P. Bjerregaard, "Spectral components of short-term RR interval variability in healthy subjects and effects of risk factors," *European Heart Journal*, vol. 15, no. 9, pp. 1174–1183, Sep. 1994.

[133] D. W. Kaczka, G. M. Barnas, B. Suki, and K. R. Lutchen, "Assessment of time-domain analyses for estimation of low-frequency respiratory mechanical properties and impedance spectra," *Annals of Biomedical Engineering*, vol. 23, no. 2, pp. 135–151, Mar. 1995.

[134] B. Diong, H. Nazeran, P. Nava, and M. Goldman, "Modeling Human Respiratory Impedance," *IEEE Engineering in Medicine and Biology Magazine*, vol. 26, no. 1, pp. 48–55, Jan. 2007.

[135] O. Dressler, G. Schneider, G. Stockmanns, and E. Kochs, "Awareness and the EEG power spectrum: analysis of frequencies," *British Journal of Anaesthesia*, vol. 93, no. 6, pp. 806–809, Dec. 2004.

[136] Weiting Chen, Zhizhong Wang, Hongbo Xie, and Wangxin Yu, "Characterization of Surface EMG Signal Based on Fuzzy Entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, Jun. 2007.

[137] H. Azami, K. Smith, and J. Escudero, "MEMD-enhanced multivariate fuzzy entropy for the evaluation of complexity in biomedical signals," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL, USA: IEEE, Aug. 2016, pp. 3761–3764.

[138] S. Bao, L. Luo, J. Mao, D. Tang, and Z. Ding, "Robust Monitoring of Industrial Processes in the Presence of Outliers in Training Data," *Industrial & Engineering Chemistry Research*, vol. 57, no. 24, pp. 8230–8239, Jun. 2018.

[139] M.-T. Lo, Y.-C. Chang, C. Lin, H.-W. V. Young, Y.-H. Lin, Y.-L. Ho, C.-K. Peng, and K. Hu, "Outlier-resilient complexity analysis of heartbeat dynamics," *Scientific Reports*, vol. 5, no. 1, p. 8836, Aug. 2015.

[140] B. Saneja and R. Rani, "An efficient approach for outlier detection in big sensor data of health care: An Efficient Approach for Outlier Detection," *International Journal of Communication Systems*, vol. 30, no. 17, p. e3352, Nov. 2017.

[141] Y. Ichimaru and G. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry and Clinical Neurosciences*, vol. 53, no. 2, pp. 175–177, Apr. 1999.

[142] S. Luo, "A review of electrocardiogram filtering," *Journal of Electrocardiology*, p. 11, 2010.

[143] H. M. Stauss, "Identification of blood pressure control mechanisms by power spectral analysis," *Clinical and Experimental Pharmacology and Physiology*, vol. 34, no. 4, pp. 362–368, Feb. 2007.

[144] Y. Gu, S. Tan, K. Wong, M. R. Ho, and L. Qu, "Using GA-based feature selection for emotion recognition from physiological signals," in *2008 International Symposium on Intelligent Signal Processing and Communications Systems*. Bangkok, Thailand: IEEE, Feb. 2009, pp. 1–4.

[145] J. Domínguez-Jiménez, K. Campo-Landines, J. Martínez-Santos, E. Delahoz, and S. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomedical Signal Processing and Control*, vol. 55, p. 101646, Jan. 2020.

[146] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[147] P. Gong, H. T. Ma, and Y. Wang, "Emotion recognition based on the multiple physiological signals," in *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. Angkor Wat, Cambodia: IEEE, Jun. 2016, pp. 140–143.

[148] Y. Chu, X. Zhao, J. Han, and Y. Su, "Physiological Signal-Based Method for Measurement of Pain Intensity," *Frontiers in Neuroscience*, vol. 11, p. 279, May 2017.

[149] H. Gu and C.-A. Chou, "Optimizing non-uniform multivariate embedding for multiscale entropy analysis of complex systems," *Biomedical Signal Processing and Control*, vol. 71, p. 103206, Jan. 2022.

[150] H. Xiao and D. P. Mandic, "Variational Embedding Multiscale Sample Entropy: A Tool for Complexity Analysis of Multichannel Systems," *Entropy*, vol. 24, no. 1, p. 26, Dec. 2021.

[151] H.-B. Xie, Y.-P. Zheng, J.-Y. Guo, and X. Chen, "Cross-fuzzy entropy: A new method to test pattern synchrony of bivariate time series," *Information Sciences*, vol. 180, no. 9, pp. 1715–1724, May 2010.

[152] W. Shi, P. Shang, and A. Lin, "The coupling analysis of stock market indices based on cross-permutation entropy," *Nonlinear Dynamics*, vol. 79, no. 4, pp. 2439–2447, Mar. 2015.

[153] H. C. Fogedby, "On the phase space approach to complexity," *Journal of Statistical Physics*, vol. 69, no. 1-2, pp. 411–425, Oct. 1992.

[154] A. Humeau-Heurtier, "Multivariate Generalized Multiscale Entropy Analysis," *Entropy*, vol. 18, no. 11, p. 411, Nov. 2016.

[155] L. V. Hedges, "Distribution Theory for Glass's Estimator of Effect size and Related Estimators," *Journal of Educational Statistics*, vol. 6, no. 2, pp. 107–128, Jun. 1981.

[156] D. Mongin, C. Chabert, D. S. Courvoisier, J. García-Romero, and J. R. Alvero-Cruz, "Heart rate recovery to assess fitness: comparison of different calculation methods in a large cross-sectional study," *Research in Sports Medicine*, pp. 1–14, Jul. 2021.

[157] M. Rostaghi, M. M. Khatibi, M. R. Ashory, and H. Azami, "Fuzzy Dispersion Entropy: A Nonlinear Measure for Signal Analysis," *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2021.

[158] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, Aug. 2012.

[159] E. Kafantaris, F. Hatami, N. Hareem, J. Yi, J. French, L. Shlomovich, M. Bucholc, M. Loning, L. Ming, O. Crook, P. L. Villagra, and P. Oleskiewicz, "Data Study Group Final Report: Great Ormond Street Hospital," Alan Turing Institute, Tech. Rep., Feb. 2020.

[160] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, Jun. 2022.

[161] N. Padhye, D. Rios, V. Fay, and S. K. Hanneman, "Pressure Injury Link to Entropy of Abdominal Temperature," *Entropy*, vol. 24, no. 8, p. 1127, Aug. 2022.

[162] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, "Directed hypergraphs and applications," *Discrete Applied Mathematics*, vol. 42, no. 2-3, pp. 177–201, Apr. 1993.

[163] A. Ritz, A. N. Tegge, H. Kim, C. L. Poirel, and T. Murali, "Signaling hypergraphs," *Trends in Biotechnology*, vol. 32, no. 7, pp. 356–362, Jul. 2014.

[164] E. Vasilyeva, A. Kozlov, K. Alfaro-Bittner, D. Musatov, A. M. Raigorodskii, M. Perc, and S. Boccaletti, "Multilayer representation of collaboration networks with higher-order interactions," *Scientific Reports*, vol. 11, no. 1, p. 5666, Dec. 2021.