Tampere University

MORTEZA ZABIHI

Contributions to Biomedical Signal Analysis Using Nonlinear Dynamics and Machine Learning

MORTEZA ZABIHI

Contributions to Biomedical Signal Analysis Using Nonlinear Dynamics and Machine Learning

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Technology and Communication Sciences
of Tampere University,
for online public discussion
on Friday, 25 September 2020, at 13 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos	Professor Moncef Gabbouj Tampere University Finland	
Pre-examiners	Doctor Reza Firoozabadi Advanced Algorithm Research Center Philips Healthcare USA	Professor Kouhyar Tavakolian University of North Dakota USA
Opponent	Professor Abdulnasir Hossen Sultan Qaboos University Oman	

Abstract

Machine Learning and Signal Processing have myriad applications in healthcare from automating the administrative tasks to speeding up medical diagnosis and drug discovery. This domain brings computer and medical sciences in a single thread, which can pull together a massive amount of techniques to improve the efficiency and cost of healthcare in the sector. In this interdisciplinary field, Machine Learning solutions are needed to be designed explicitly for each task. Thus, the developed solutions in this area must be more prudent about compliance compared to other domains. It is crucial to know how the clinical data are collected or for what purposes the data are gathered to deliver reliable solutions.

This study focuses on two primary objectives. The first objective is to propose a novel set of nonlinear descriptors for biosignals using nonlinear dynamics. The second objective is to develop feature extraction and classification schemes to help medical experts improve the diagnosis of prevalent medical conditions.

The first contribution of the thesis is to introduce a novel set of nonlinear descriptors for capturing the dynamic signature of time series. The proposed features are inspired by the Poincaré section and nullclines concepts in nonlinear dynamics. The discriminative power of the proposed features is evaluated in the epileptic seizure detection task. For this purpose, a two-layer classification approach is developed to detect seizure events using multichannel electroencephalogram signals. The extensive comparative analysis shows the superiority of the proposed framework over other state-of-the-art methods, although further research is needed for reducing the high false alarm rate.

The second contribution is to develop three novel feature extraction and classification schemes to tackle clinically relevant conditions. These medical applications encompass the identification of heart anomalies using heart sounds, automatic classification of atrial fibrillation using hand-held electrocardiogram devices, and sepsis prediction from 6 to 12 hours before clinical recognition. The experimental results show the feasibility of these assistive diagnostic tools. The developed solutions are assessed against a variety of solutions proposed by several teams from academia and industry and ranked among the top three methods.

The proposed feature extraction and classification techniques are stand-alone contributions to the field. The proposed nonlinear dynamics techniques can be used in a variety of domains due to their generic nature. Moreover, the developed automatic classification and predictive models in this thesis showed significant improvement in the identification of challenging clinical events. According to the conducted evaluation analysis, the pro-

posed methods exhibit relatively robust performance over the collected data across different institutions. We conclude that Machine Learning assistive diagnostic models designed with sufficient care can provide viable help for expert clinicians.

Preface

I wish to express my most profound appreciation to my supervisors Prof. Moncef Gabbouj and Prof. Serkan Kiranyaz, for their support, criticism, and advice. I cannot thank Prof. Moncef Gabbouj enough. He gave me valuable insights and an enormous amount of freedom to conduct my research. The successes that I have had over my Ph.D. studies have been due to his influence. I would also like to extend my deepest gratitude to Prof. Serkan Kiranyaz. My research has benefited greatly from discussions with him, his technical supervision, and infinite patience. He taught me how to write structured and coherent papers. It has been a great privilege to work with him.

I am indebted to my collaborator and close friend, Dr. Ali Bahrami Rad, for providing invaluable support and insights into my life inside and outside of academia. It was with him that I discovered my passion for scientific competitions. Our friendship and scientific collaboration over the past years have been priceless to me, and I look forward to continuing those in the years to come.

I am especially thankful to my awesome office mate and friend, Dr. Adamantios Ntakaris. I have appreciated his support and friendship over the past three years.

I am thankful to my co-authors, Prof. Ville Jäntti, Prof. Tarmo lipping, Prof. Aggelos K. Katsaggelos, and Prof. Turker Ince, for their thoughtful comments and insightful suggestions. My gratitude also goes to the former and current members of our research group Dr. Iftikhar Ahmad, Dr. Honglei Zhang, Dr. Guanqun Cao, Dr. Stefan Gerd Uhlmann, Fahad Sohrab, Mete Ahishali, and Aysen Degerli who have been an excellent source of discourse, ideas, and friendship.

I am very grateful to Virve Larmila, Kari Suomela, and Elina Orava for the fantastic administrative and technical assistance they provided in the laboratory at various stages of our work.

I would like to express special thanks to my great friends, in particular, Dr. Pooya Skaeti, Milad Mostofizadeh, and Aaron for their friendship, spiritual support, and endless encouragement during the last five years. And also to my wonderful friends in Iran not named here. You all have made these years truly special.

I am forever grateful to my parents. I would like to thank them for their endless love and support. I have always felt their presence, and I could not ask more devoted parents. All the success I have in my life, I owe to them. Without my parents' love and support, I would not have been able to pursue my academic interests freely. I dedicate this thesis to them.

Thanks to my younger brother Mojtaba, who is always the first one to support me. I am forever in debt for his constant support and encouragement. I am also grateful to my older sisters and brother for their emotional support along the way.

Tampere 1.11. 2019

Morteza Zabihi

Contents

List of Figures

List of Tables

List of Symbols

g	Activation function
a	Approximation coefficient wavelet transform
b	Bias
D_c	Correlation dimension
C	Correlation sum
J	Cost function
Σ	Covariance
d	Detail coefficients of wavelet transform
d	Dimension
λ	Eigenvalue
ω	Eigenvector
m	Embedding dimension
X^*	Equilibria
β	Expansion coefficient
$\boldsymbol{s}$	Feature matrix
$\mathfrak{M}$	Floquet multiplier
$\mathfrak{f}$	Frequency
G	Gini index
θ	Heaviside function
I	Identity matrix
s	Intersection point on the Poincaré section
A	Jacobian matrix
y	Label
r	Logistic map driving parameter
L	Loss function
$\mathcal{M}$	Manifold
μ	Mean
$\mathcal{C}$	MFCC coefficients
$\mathscr{s}$	Number of features

$\mathcal{N}$	Normal distribution
f	Nonlinear function
ℓ	Number of leaves in a decision tree
F	Objective function XGBoost
P	Poincaré map
$\mathcal{P}$	Probability
$\wp$	Probability density function
r	Radius
λ	Regularization parameter
Res	Residual
Ω	Regularization function
ω	Score of leaves in a decision tree
$\mathcal{T}$	Set of observations in a node
η	Small value, $|\eta| \ll 1$
h	Smooth observation function
Φ	Smooth time map
x	State in phase space
τ	Time lag
x	Time series
t	Timestamp
v	Values in the feature matrix
k	Weak learner
W	Weight

List of Abbreviations

AF	Atrial Fibrillation
ANN	Artificial Neural Network
ApEn	Approximate Entropy
AUROC	Area Under the Receiver Operating Characteristics
AUPRC	Area Under the Precision-Recall Curve
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EEG	Electroencephalogram
FNN	False Nearest Neighbors
LDA	Linear Discriminant Analysis
MDA	Mahalanobis Discriminant Analysis
MFCC	Mel Frequency Cepstral Coefficients
ML	Machine Learning
PCG	Phonocardiogram
QDA	Quadratic Discriminant Analysis

List of Publications

[P1]   M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M. Gabbouj, and T. Ince, " Analysis of high-dimensional phase space via Poincaré section for patient-specific seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 386 – 398, March 2016.

[P2]   M. Zabihi, S. Kiranyaz, V. Jäntti, T. Lipping, and M. Gabbouj, "Patient-Specific Seizure Detection Using Nonlinear Dynamics and Nullclines." *IEEE. Journal of Biomedical and Health Informatics*, March 2019.

[P3]   M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, A. K. Katsaggelos, "Heart Sound Anomaly and Quality Detection Using Ensemble of Neural Networks without Segmentation," *Computing in Cardiology Conference (CinC)*, IEEE, Vancouver, BC, Canada, 11-14 Sept, 2016.

[P4]   M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, M. Gabbouj, "Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier," *Computing in Cardiology Conference (CinC)*, IEEE, Rennes, France, 24-27 Sept, 2017.

[P5]   M. Zabihi, S. Kiranyaz, and M. Gabbouj, "Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models," *Computing in Cardiology Conference (CinC)*, IEEE, Singapore, 8-11 Sept, 2019.

1 Introduction

The unprecedented advancement rate of the Machine Learning (ML) era has transformed the world of healthcare. The clinical outcomes are getting more effective and accurate when they are empowered by a combination of experts and machine-powered solutions. A wide variety of ML solutions have been successfully applied in the healthcare industry, from the prediction and assistive diagnosis of different medical conditions to lifestyle and behavioral modifications. One of the top applications of these models in healthcare is medical assistive diagnosis. This can include anything from identifying neurological diseases and cardiovascular anomalies to the early detection of cancers.

In medical diagnosis, physicians review the symptoms and medical history of an individual and prescribe several tests. These tests collect measurable properties from biological systems and organs to reflect their electrophysiological activities, mechanical functions, or biochemical processes. These measurements facilitate the quantitative assessment and monitoring of the physiological function. However, interpreting biosignals and clinical data is a challenging task for human experts. For instance, sometimes the clinical symptoms do not have a clear cause, or the clinical manifestations suggest multiple distinct conditions. Additionally, the medical conditions can evolve within or between individuals, which causes a profound variability in the characteristics of physiological measurements. The other challenge that is often overlooked is information overload. With the rise in the amount of clinical information and long-term measurements, often relevant and credible data are not considered. These challenges can lead to misdiagnosis or delayed diagnosis. Moreover, these complexities can cause interobserver variation, which refers to disagreements between experts in identifying a medical event.

All these challenges motivate the research on the deployment of practical ML models, which can adequately help experts interpret clinical data and detect certain health events.

These ML solutions are equipped with Signal Processing and ML techniques, which are explicitly designed for the medical problems they aim to solve. The primary purpose of these algorithms is to optimize the performance of the assistive diagnosis tools (referred therein as diagnosis tools for short) and to be used as a reliable recommendation system rather than replacing physicians in the diagnosis process. Such solutions accelerate the diagnosis procedure and reduce human error. Although several models have been developed in the medical diagnostic domain, there is much room for improvement.

This thesis aims to tackle two major objectives. The first objective is to propose a new set of dynamical signatures for electroencephalogram (EEG) signals. The fundamental assumption of nonlinear time series analysis is that the physical measurement is generated by a nonlinear deterministic process. This is not a realistic assumption for complex multicellular systems, such as the brain. It is too ideal to model the function of biological systems with a set of differential equations. However, in many biological systems, the footprint of deterministic nonlinear behaviors is evident [1] [2] [3] [4] [5].

Nonlinear time series techniques model the behavior of a system using conceptually deterministic states. It uses the time series to reconstruct a multi-dimensional representation of the *actual* system states. Such reconstruction is performed by assuming that the physical measurement conveys the essential dynamical information of the original system. Nonlinear time series analysis offers revealing properties that are hidden from the perspective of methods such as Fourier or wavelet transform.

One example of such shortcomings in frequency analysis is the interpretation of noise using the ubiquitous power spectral density approach. A (purely) stochastic process has a decaying autocorrelation since the samples with higher time-lag are getting less correlated. This decaying rate depends on the amount of the signal stochasticity. For instance, the white noise characteristics in power spectral density (Fourier transform of autocorrelation function) emerge as an almost uniformly distributed pattern over the entire frequency band, i.e., flat spectrum. However, time series generated from a purely deterministic chaotic system can have a similar property. Let's consider white noise and a signal generated by a deterministic map (logistic map - see Section 2) with almost the same power. In Figure 1.1, the time series with their power spectrums are illustrated (Figure 1.1.a, b).

As can be seen, both signals have the same characteristics in power spectral density (except in the zero frequency, i.e., DC component). Despite these similarities, the two signals are generated from entirely different mechanisms. The reconstructed phase spaces (see section 2.3) of the time series are provided in Figure 1.1.c to visualize the underlying dynamics of the signals. This representation unveils information that cannot

be offered by frequency analysis. Thus, time series analysis should be carried out by choosing a proper method; otherwise, it can lead to invalid conclusions.

Therefore, nonlinear methods can provide discriminative features of biosignals and appear equally appropriate (compared to stochastic processes) for biosignal analysis [6]. Besides, due to the nonlinearity involved in all the known biological systems, nonlinear dynamics is a relevant field to be considered when studying biosignals. In this thesis, we aim to answer two main research questions: a) can we develop a set of discriminative nonlinear descriptors for Electroencephalogram (EEG) signals? and b) can we build a framework for epileptic seizure detection based on these nonlinear features? The proposed features are then compared with other established methods in this domain. It has been shown that features inspired by nonlinear dynamics achieved state-of-the-art performance compared to various features in frequency and time-frequency domains.
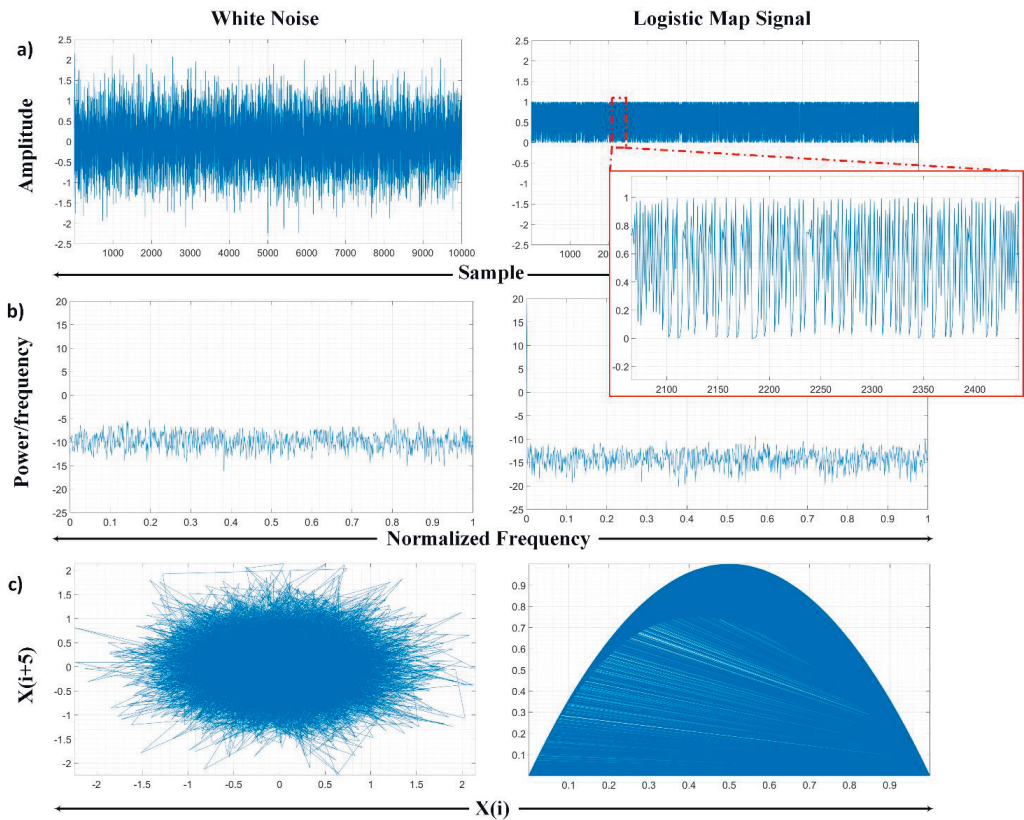

Figure 1.1 Sample illustration of nonlinear time series analysis application. *a)* White noise and logistic map signal ($x_{n+1} = 4x_n(1 - x_n)$, with $x_0 = 0.1$). *b)* Welch's power spectral density, *c)* reconstructed phase space with $\tau = 5$.

The second objective is to develop Signal Processing and ML techniques for three challenging diagnosis-focused applications. In this thesis, we aim to answer the following research questions: which classification approaches are suitable for (a) detecting heart anomalies using short-term phonocardiogram (PCG) signals, (b) detecting atrial fibrillation using one-lead electrocardiogram (ECG) signals, and (c) predicting sepsis for intensive care unit patients early enough? We propose to utilize a combination of nonlinear, spectral, and other well-established methods to tackle these applications. The primary motivations for tackling the medical conditions are the following. These clinical applications have not been well-resolved due to several challenges, such as limited availability of datasets, noise, artifacts, and lack of standard regulations. Additionally, these medical conditions have become attractive in the interdisciplinary field of biosignal analysis due to their prevalence in the global population. Furthermore, ML models in healthcare suffer from the lack of evaluations with the datasets collected by different institutions.

To facilitate finding solutions to these problems, the worldwide forum, Physionet, organizes annual scientific competitions in which the world's largest open-access medical datasets are provided. Physionet serves as a global platform where interested researchers are able to test their solutions across the collected data from various medical environments and conduct a fair comparative evaluation against the other developed models. The ML models proposed in this thesis are ranked in the top three solutions among a large set of international teams in the Physionet competitions.

1.1 Objectives and Thesis Overview

In this thesis, our focus is particularly drawn on investigating and proposing advanced biosignal and clinical data analysis to improve the classification performance of certain health events. The specific objectives of this thesis are:

- Introducing a new set of nonlinear descriptors for Electroencephalogram (EEG) signals and proposing a patient-specific seizure detection framework for multi-channel EEG recordings.
- Developing a feature extraction and classification scheme for the detection of heart anomalies using phonocardiogram (PCG) signals and their quality assessment.
- Investigating a comprehensive set of features and developing a classification scheme for atrial fibrillation detection using hand-held one-lead Electrocardiogram (ECG) signals.

▪ Proposing a novel set of features to model the missingness of clinical data and developing a predictive model for early prediction of sepsis in intensive care units.

The thesis is organized as follows: In Chapter 2, the fundamental and critical concepts of nonlinear time series analysis are presented. The chapter starts with the introduction of nonlinear dynamics in time series analysis (Section 2.1), with a special focus on the necessary steps to the nonlinear dynamics, i.e., phase space (Section 2.2), phase space reconstruction (Section 2.3), Poincaré section (Section 2.4), and directivity curve (Section 2.5). In Chapter 3, a brief overview of the applied supervised classification models is presented. These classifiers consist of naïve Bayes, linear, and quadratic discriminant analysis (Section 3.1), random forest (Section 3.2), Gradient Boosting (Section 3.3), and Artificial Neural Networks (Section 3.4). Chapter 4 summarizes the five publications included in this thesis. This section presents the novel solutions and their state-of-the-art results for the aforementioned research objectives. Due to the variety of the covered applications, the proposed feature engineering methods and classification developments via their experimental results and benchmark datasets are explained separately for each application. First, the proposed nonlinear features for EEG signals and their applications on epileptic seizure detection are described in Section 4.2. Then, the developed feature extraction and classification schemes for the detection of heart anomalies, atrial fibrillation, and early prediction of sepsis are summarized in Sections 4.3.1, 4.3.2, and 4.3.3, respectively. Finally, the general and specific conclusive remarks are provided in the conclusion Chapter.

1.2 Author's Contributions

The five publications covered in this thesis aim to answer to the above research questions and meet the set objectives. The author's contribution to the field of biosignal analysis are summarized below.

- Proposing a novel set of Electroencephalogram (EEG) descriptors based on nonlinear dynamics (publications [P1] and [P2]).
- Proposing a classification framework for epileptic seizure detection (publication [P1]).
- Classification of cardiac anomalies using heart sounds (publication [P3]).
- Developing a classification method and investigating a comprehensive set of discriminative features for atrial fibrillation detection (publication [P4]).
- Proposing a new set of features for missing samples in clinical data (publication [P5]).

- Developing a predictive model for early sepsis prediction for intensive care unit patients (publication [P5]).
- Implementing the methods proposed in publications [P3], [P4], [P5], and making them available as open-source software.

2 Nonlinear Dynamics and Time Series Analysis

Time series analysis, in its broadest form, infers patterns and characteristics that can improve the prediction of an event. Stochastic and deterministic models are the two standard approaches for time series analysis. Stochastic models assign probabilities to the state of the system based on the random variable at the current time and its earlier values (i.e., conditional probability distribution). On the contrary, in the deterministic approach, the outputs of the system can be described precisely using the input and a set of equations. Thus, randomness has no effects on the future behavior of the system.

However, uncertainty can be generated by a purely deterministic system. In such cases, the irregularities in the behavior of the system are explained by Chaos theory [7]. Chaotic solutions can only be generated from nonlinear difference (or differential) equations, even though they can be relatively simple. The Logistic difference equation or logistic map is a classical demonstration of such cases [8]. The logistic map is a simple population model and is given by

$$x_{n+1} = r x_n (1 - x_n), \tag{2.1}$$

where $x \in [0,1]$ is the population size in percentage, and r is a driving parameter. The term $(1 - x_n)$ acts as a mathematical damper. If $r > 1$, it drives the population x higher and makes $(1 - x_n)$ smaller. The system growth suppresses further population growth, and finally stabilizes the system to reach a state of equilibrium. To visualize the strange behavior of the logistic map, we simulate the equation for r from 0 to 4 and plot its corresponding x values (see Figure 2.1).
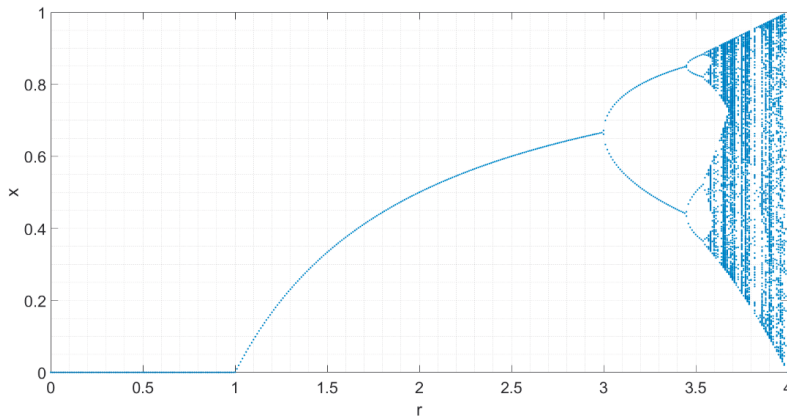
Figure 2.1 Bifurcation diagram of the logistic map

As can be seen in Figure 2.1, for $0 < r < 1$, x finally reaches 0 (i.e., the population dies) independent of the initial value. At $r = 1$, there is a significant change in the behavior of the solution (bifurcation). When $1 < r < 3$, the population eventually approaches $1 - \frac{1}{r}$ and settles into a stable population. Another dominant behavior happens when $3 < r < 3.44$, where x bounces between two solutions (periodic orbit). As r increases, $r > 3.44$, the solution oscillates between four values. The periodic doubling continues until r reaches around 3.569, which is the onset of a chaotic behavior. As can be seen, a small change in the initial values leads to erratically different x values through time. This shows that increasing r makes the stabilization more difficult and eventually causes the emergence of irregular behaviors while the system is still purely deterministic. It is interesting how this rich dynamical response is generated only from the governing equation and not from variations caused by the environment, such as noise.

The main difference between such irregularities and randomness can be explained by their predictability. Random behaviors are unpredictable and cannot be reproduced. On the contrary, the mentioned irregularities are deterministic and can be regenerated if the identical initial conditions are known. It needs to be mentioned that the sensitivity on initial conditions exists in linear systems. However, in linear systems, this sensitivity can only drive the solution to infinity. This is contrary to deterministic chaos, where the solution trajectories need to be bounded (stretching and folding of the trajectories, i.e., non-linearity) [7]. In addition to the conditions such as bounded solution, determinism, and sensitive dependence on initial conditions, chaotic solutions should have aperiodic time-asymptotic behavior [7]. This implies that the trajectories of the solution do not sink to

equilibriums, e.g., fixed-point or periodic orbits, such as the logistic map behavior for $r >$ 3.569.

Nonlinear dynamics covers not only chaotic behaviors but also studies numerous types of behaviors, such as periodicity and aperiodicity. Therefore, it is a relevant field to consider when studying biomedical signals. This is motivated by the fact that all known biological systems are nonlinear. Nonlinear dynamics introduces an alternative perspective and set of tools to analyze physical measurements and time series. As an example, the transition between normal heartbeats and ventricular tachycardia in ECG signals can be described by bifurcation theory [9]. Another example is the changeover from a non-seizure to a seizure state in Electroencephalogram (EEG) of an epileptic brain which can be characterized using nonlinear dynamics [10].

It is worth mentioning that irregularities in time series cannot be considered deterministic only by a casual inspection. Such behavior may be due to the superposition of different types of noise into the signal. Thus, surrogating data [11] [12] is needed to test the deterministic and stochastic characteristics of a time series. In this thesis, we do not attempt to prove the existence of chaos in biomedical signals. Moreover, we do not claim that deterministic approaches are preferred over stochastic ones. Here we use nonlinear dynamics as a tool to capture salient features of medical events and use them to boost the accuracy of our regression and predictive models. This chapter covers a brief overview of nonlinear dynamics and its fundamental techniques, which are the basis of our main contributions.

2.1 Nonlinear Dynamics and Differential Equations

The nonlinear dynamics studies the evolution of nonlinear systems through time by an autonomous differential equation of the form

$$\frac{dX}{dt} = \dot{X} = f(X), \tag{2.2}$$

where f is a nonlinear function. Studying the properties of a nonlinear system is equivalent to solving the differential equations described in Eq. 2.2. Here, we are looking for a trajectory $X(t)$ whose derivative, $\dot{X}$, is $f(X)$. An interesting subset of the solution trajectory $X(t)$ is named equilibrium or steady-state of the system. Equilibria are invariant under the dynamical evolution of the system through time and can have different forms based on the system's degree of freedom. The degree of freedom represents the number

of variables needed to explain the system dynamics. The equilibria of a system are de-
fined when the time derivatives of all variables are zero,

$$\frac{dX}{dt} = \dot{X} = 0. \tag{2.3}$$

As expected, the complexity of the solution and equilibria can vary with respect to the
degree of freedom. For example, one dimensional (first-order) systems can exhibit only
two possible behaviors. The trajectory either goes to infinity, or it approaches a fixed
point. Therefore, in one-dimensional systems, oscillation or periodic solutions do not ex-
ist. Equilibria are fundamental characteristics of a dynamical system not only because
they provide salient features of the system but because they can be used to estimate the
entire behavior of the system. In the following sections, we discuss these concepts in
more detail.

2.2 Phase Space

Often, instead of a single function in Eq. 2.2, there is a system of equations where $X \in \mathbb{R}^n$ is a vector. Such systems are represented as coupled ordinary differential equations,

$$\begin{cases} \dot{X}_1 = f_1(X_1, \dots, X_n) \\ \dot{X}_2 = f_2(X_1, \dots, X_n) \\ \quad\quad \dots \\ \dot{X}_n = f_2(X_1, \dots, X_n) \end{cases}, \tag{2.4}$$

where $\mathbb{R}^n$ is the phase space or state space of the system. The phase space can be
described as a graphical representation of the system behavior that is formed by a set
of trajectories with different initial points. Therefore, each point in the phase space is
associated with variables of the system and describes the corresponding state.

Phase space is a concept that connects symplectic geometry, the Hamiltonian formula-
tion in classical mechanics, and partial differential equations together. Hence, it plays a
crucial role as a graphical tool to describe the system dynamics. Phase space is useful
in many applications, e.g., when computing the integral of the right-hand side of Eq. 2.4
is too complicated, and consequently finding the analytical solution is almost impossible
[13].

Due to the determinism assumption, system behavior can be precisely determined by
the phase space. This can be perceived as a result of the no-intersection theorem [14].

The no-intersection theorem states that in a finite period, two distinct trajectories in the phase space cannot intersect, and a single trajectory cannot cross itself. In other words, this theorem emphasizes the fact that the future state of each trajectory can be determined only based on its current location in the phase space. To illustrate this, imagine a phase space in which two different trajectories have the same state (i.e., intersection). If the intersection point is considered as an initial point, then the same initial point follows two different evolutions in the future, which is a contradiction of determinism. This theorem is useful for constructing the phase space as it constrains the evolution of trajectories (see section 2.3).

The first step in creating the n-dimensional phase space is to find the equilibria of the given function f by setting the time derivatives equal to 0. If the solution starts at the equilibrium, it stays at the same point. However, the critical question is that if the solution starts at another point in the phase space, whether the trajectory reaches the equilibrium (stable) or the equilibrium is repulsive (unstable). The way to find out the stability type is to examine the dynamics of the system in the vicinity of the equilibria, X^*, with a small deviation η from X^*,

$$\dot{X} = f(X^* + \eta), \tag{2.5}$$

where $|\eta| \ll 1$. Using Taylor's formula in the neighborhood of X^*,

$$\dot{X} = f(X) = f(X^* + \eta) = f(X^*) + \eta \dot{f}(X^*) + \frac{\eta^2}{2} \ddot{f}(X^*) + \cdots \tag{2.6}$$

Based on the definition, the function at equilibria equals to 0, i.e., $f(X^*) = 0$. The term $\dot{f}(X^*)$ in Eq. 2.6 yields the slope of f at X^*, which means that it is a constant number. As η is a small value the quadratic term of $\frac{\eta^2}{2}\ddot{f}(X^*)$ and the rest of the higher-order terms are neglected. Thus, this yields the linearization of $\dot{X} = f(X)$ at X^*. After linearization, the system is a linear differential equation,

$$\frac{dX}{dt} = AX, \tag{2.7}$$

where A is the Jacobian matrix consisting of the partial derivative of f in all directions of $\mathbb{R}^n$ at equilibria. The components of the Jacobian matrix are arranged as

$$A_{ij} = \left[\frac{\partial f}{\partial X_1} \quad \cdots \quad \frac{\partial f}{\partial X_n} \right]. \tag{2.8}$$

For a linear autonomous first-order differential equation, the solution is in the form of $x(t) = e^{\lambda t}\omega$. If the solution $e^{\lambda t}\omega$ is placed into Eq. 2.7, then $\lambda e^{\lambda t}\omega = Ae^{\lambda t}\omega$, which yields $\lambda\omega = A\omega$, where λ is the eigenvalue, and ω is the eigenvector. Thus,

$$(A - \lambda I)\omega = 0 \;\rightarrow\; \det(A - \lambda I) = 0, \tag{2.9}$$

where I is the identity matrix. Then, the eigenvalues are calculated using the Jacobian determinant. Once the eigenvalues are known, then the type of their stability can be determined. For example, for a two-dimensional system, if the real part of the eigenvalues are negative (i.e., $\lambda_{1,2} < 0$, or if $\lambda_{1,2} = a \pm i\Psi$, $a < 0$) then the solution $x(t) = e^{\lambda t}\omega$ goes to zero, and the system is stable at the equilibria. If both eigenvalues are positive, then the equilibria are unstable. If the eigenvalues have opposite signs, then the equilibrium is a saddle point. Moreover, if the eigenvalues have imaginary parts, then the system would have a spiral-shape trajectory in the phase space. Once the system equilibria and the corresponding eigenvalues and eigenvectors are found, the geometric shape of the trajectories in the phase space is estimated qualitatively.

It is worth mentioning that there is an exception in Eq. 2.6. This linearization is valid only if the term $\dot{f}(X^*)$ is not zero. If $\dot{f}(X^*) = 0$, then $\left|\eta\dot{f}(X^*)\right| \leq \left|\frac{\eta^2}{2}\ddot{f}(X^*)\right|$. This states that the quadratic term cannot be neglected, and consequently, the linearization cannot be applied. If $\dot{f}(X^*) = 0$, then the equilibria can be stable, nonstable, or a saddle point. It should be noted that function f is assumed to have a convergent Taylor series. It is considered that the solution to $\dot{X} = f(X)$ exists and is unique by assuming that $f(X)$ is continuously differentiable, i.e., $f(X)$ and its first derivate, $\dot{f}(X)$, are both continuous.

2.3 Phase Space Reconstruction

Phase space is constructed based on the governing differential equation. However, in real systems, only physical measurements are available in the form of time series. By reconstructing the phase space from the measured time series, we are looking for some general characterizations of the system, which are independent of a specific trajectory [14]. In other words, the main question here is whether or not it is possible to have a mapping from the unknown (actual) phase space to an $\mathbb{R}^m$ using a time series such that $\mathbb{R}^m$ preserves the dynamic structure.

To answer this question, we shall first describe an adequate topological mapping for the idea of structure-preserving. Such a mapping function should be injective (i.e., one-to-one) and homomorphism (i.e., a continuous function with a continuous inverse function).

By using this type of mapping, the two spaces are topologically equivalent, and this is the essence of the embedding concept. The injective property of embedding is a necessary constraint for determinism. Recall the non-intersection theorem discussed in Section 2.2. The one-to-one property ensures that each distinct state of the original phase space is mapped to a unique point in the reconstructed phase space.

Takens' delay embedding theorem [15] provides a mathematical treatment for such a mapping. Takens proved that with a time-delayed version of a single measured time series, a d-dimensional space can be reconstructed that preserves the main dynamic properties of the original phase space. This theorem is based on the Whitney-Embedding theorem [16], which provides a connection between phase space and the measurements. The Whitney-Embedding theorem states that every Hausdorff, smooth d-manifold can be smoothly embedded into a $2d + 1$ Euclidean space, $\mathbb{R}^{2d+1}$.

Suppose that the physical measurement (or time series) is generated by a smooth time map Φ, which is diffeomorphism on the smooth manifold $\mathcal{M}$ [17]. Additionally, suppose there is a smooth observation function h: $\mathcal{M} \to \mathbb{R}$, which maps the actual flow of the dynamic system Φ to our one-dimensional measurement (i.e., time series). The delay embedding theorem states that there is a $2d + 1$-fold map $H[\Phi, \text{h}, \tau]: \mathcal{M} \to \mathbb{R}^{2d+1}$, that is immersion. Immersion means that both H and H^{-1} (its inverse) are differentiable, and both H and its derivative are injective everywhere. The proof is provided in [18]. This means that H is an embedding of $\mathcal{M}$ in $\mathbb{R}^{2d+1}$. The map H is defined as

$$x \to \left(h\big(x(t_n)\big), h\left(\Phi\big(x(t_n)\big)\right), h\left(\Phi^2\big(x(t_n)\big)\right), \dots, h\left(\Phi^{m-1}\big(x(t_n)\big)\right) \right), \quad (2.10)$$

where $x(t_n)$ denotes the actual state in manifold $\mathcal{M}$ at time t_n. Since it is assumed that a deterministic system rules the underlying dynamic, then

$$\Phi\big(x(t_n)\big) = x(t_{n-\tau}) \to \Phi\big(x(t_{n-\tau})\big) = x(t_{n-2\tau})$$

$$\to \Phi\big(x(t_{n-2\tau})\big) = \Phi\left(\Phi\big(x(t_n)\big)\right) = \Phi^2\big(x(t_n)\big). \quad (2.11)$$

Therefore,

$$x\big(t_{n-(m-1)\tau}\big) = \Phi^{m-1}\big(x(t_n)\big), \quad (2.12)$$

where τ is the time-lag, and m is the embedding dimension. Thus, Eq. 2.10 can be written as

$$x \to \left(h\big(x(t_n)\big), h\big(x(t_{n-\tau})\big), \dots, h\left(x\big(t_{n-(m-2)\tau}\big)\right), h\left(x\big(t_{n-(m-1)\tau}\big)\right) \right). \quad (2.13)$$

Often, the observation function h is related to the time series x as

$$x(t_n) = h\big(x(t_n)\big) + \epsilon(t_n), \tag{2.14}$$

where noise ϵ is assumed to be zero.

Estimating the embedding dimension m and the time lag τ are the main challenges of reconstruction. Although several methods are proposed to estimate these two parameters, they depend strongly on the application and the time series. The main approaches for finding the minimum embedding dimension m are using false nearest neighbors (FNN) [19] and correlation dimension. The basic idea of FNN is an iterative search for the distance of states (points in the reconstructed phase space) in m and $m + 1$ dimensions. This process continues until the distance ratio of the two successive dimensions is not larger than a constant threshold. In other words, in this process, we are looking for actual neighbors that are not separating by adding a coordinate to the reconstructed phase space [20] [21]. As it is evident, there are some issues with this method, such as which neighbor or threshold should be chosen.

In the correlation dimension method, we shall first introduce the definition of correlation sum C as

$$C(r) = \frac{1}{N(N-1)} \sum_{i,j=1,i\neq j}^{N} \theta\big[r - |x_i - x_j|\big], \tag{2.15}$$

where r is the radius of a neighborhood of points, N is the number of points in the (reconstructed) phase space, θ is Heaviside function, and $|x_i - x_j|$ is the distance between two points in the d-dimensional phase space,

$$|x_i - x_j| = \sqrt{\sum_{k=0}^{d-1}(x_{i+k\tau} - x_{j+k\tau})^2}. \tag{2.16}$$

The correlation sum shows the number of points located in the region with radius r. If r is too small, it means that no pair of points can be found inside the neighbor of r and $C(r) = 0$. On the contrary, if r is too large, then all possible points can be found in the region, and therefore $C = 1$. Hence, C varies between 0 and 1. The correlation dimension, D_c, is then defined as the rate of change of C from 0 to 1,

$$D_c = \lim_{r \to 0} \frac{\log C(r)}{\log r}. \tag{2.17}$$

For calculating the minimum embedding dimension m, first D_c is calculated for different $d = 1, 2, 3, ...$ until D_c becomes independent of d. Then, the corresponding d (or saturated d) is chosen as the embedding dimension m. The idea behind the correlation dimension is that the calculated saturated d is higher than twice the real dimension of the original phase space [16].

The other needed parameter for phase space reconstruction is τ. For time lag τ different methods are proposed to find the statistical dependence between successive lagged time series. For this purpose, the most straightforward approach is using the autocorrelation function. The delay, where the autocorrelation function has zero for the first time, is chosen as τ. The main idea is to find the delay such that the linear correlation between a point and a point τ ahead is zero. Another approach is using mutual information. In this approach, the delay corresponds to the first minimum of the mutual information is used as τ. The main distinction between the two approaches is that the latter one considers the nonlinear dependencies as well.

Although the idea of reconstructing the phase space is powerful and useful in many disciplines, there are some practical limitations. The variables of a nonlinear system intertwine with each other and cannot be separated. In phase space reconstruction, using this characteristic of nonlinear systems, we assume that the physical measurement contains all the information of the underlying dynamic. However, in practice, it cannot be guaranteed whether the observed time series is an accurate representation of the dynamics. Moreover, it is assumed that the measured signal is noiseless, while in practical measurements, the presence of noise is unavoidable. Besides, the numerical algorithms used to find the proper embedding dimension and time lag involve with several approximations.

2.4 Poincaré Section

Poincaré section provides the possibility to encode the dynamics of the system by focusing on a particular section of the phase space rather than observing the whole trajectory [14]. Poincaré section is a hyperplane that traverses the trajectory of the solution and cuts through it in the phase space. Depending on the evolution of the system, the trajectory intersects with the hyperplane in different locations. Figure 2.2 shows an example of a Poincaré section and Rössler attractor. The function that rules the relation between these successive intersections is called Poincaré map, P, and is defined as,

$$x_{k+1} = P(x_k), \qquad (2.18)$$

where k denotes the kth intersection on the hyperplane. The d-dimensional phase space is replaced with a $d-1$-order map using this method. Besides reducing the problem by a dimension, this technique offers other advantages as well.
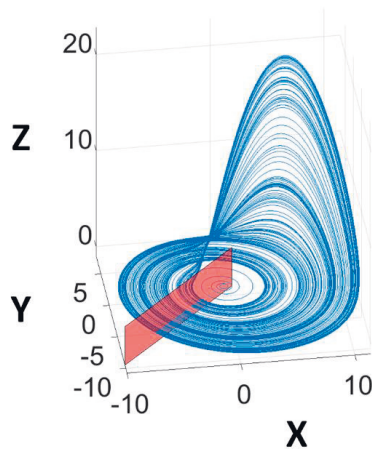
Figure 2.2 An example of a Poincaré section in Rössler attractor [22]

The patterns of the intersection points can reveal information about the periodicity of the trajectory. Let us imagine a periodic orbit in the phase space. Based on the Poincaré-Bendixon theorem [23], if a trajectory enters and does not exit a closed and bounded region of the phase space where no equilibrium exists in the region, then the trajectory must approach a periodic orbit. Such trajectories in the phase space intersect with the Poincaré section at the same point at different times. Thus, one can investigate the periodicity of a periodic orbit in the original system by only looking at the intersection points. For a closed orbit, $x^* = P(x^*)$, which is the definition of equilibrium of the Poincaré map.

Additionally, the Poincaré section can be used for stability analysis. The basic idea of stability analysis of an orbit is to answer the question of how a trajectory evolves if a perturbation is added to the initial condition. In other words, we are investigating the sensitivity of the original trajectory to the introduced perturbation. Consider the first-order differential equation

$$\dot{X} = \mathcal{A}(t)X. \tag{2.19}$$

Suppose that $\mathcal{A}(t)$, a time-varying matrix $\in \mathbb{R}^{n \times n}$, is periodic with a period T_0. Then monodromy matrix is defined as $x(T_0)$, where $x(t)$ (with $x(0) = I$) is the solution of Eq. 2.19. The eigenvalues of the monodromy matrix are known as Floquet exponents. If each of

the Floquet exponents has a modulus less than one, then the origin is exponentially stable. Similarly, if at least one exponent is greater than one, then the system is unstable. Similarly, in the Poincaré section, the evolution of a periodic orbit is investigated in the vicinity of the equilibrium s^*,

$$s_2 - s^* = P(s_1) - P(s^*), \tag{2.20}$$

where s is an intersection point on the Poincaré section. Using Taylor series in the neighborhood of s^* (i.e., s_2),

$$s_2 = P(s_1) = P(s^*) + \dot{P}(s^*)(s_1 - s^*) + \cdots \tag{2.21}$$

Thus, Eq. 2.20 can be rewritten as

$$s_2 - s^* = P(s^*) + \dot{P}(s^*)(s_1 - s^*) + \cdots - P(s^*) \tag{2.22}$$

where $\dot{P}(s^*)$ is called the Floquet multiplier. If $d_i = s_i - s^*$, and $\mathfrak{M} = \dot{P}(s^*)$, then

$$d_{i+1} = \mathfrak{M}^i d_i. \tag{2.23}$$

This shows that if $\mathfrak{M} < 1$, then $d_{i+1} < d_i$, which means that the intersection points are getting close to the equilibrium s_*, and therefore, it is a stable periodic orbit. Likewise, if $\mathfrak{M} > 1$, then the periodic orbit is repelling and unstable [14].

It is worth mentioning that there are some limitations to this method. First, finding the Poincaré map may be difficult or impossible in real time series. Second, the Poincaré map depends strongly on the hyperplane as different cross-sections with the phase space can result in different intersection points [24]. Therefore, for example, a periodic orbit is going to be missed entirely. Despite the mentioned limitations, several studies [25] [26] [27] [28] [29] have successfully applied techniques that are inspired by the Poincaré section and achieved reliable results.

2.5 Directivity Curves

For a two-dimensional differential equation, directivity curve [30] or nullcline is defined as a set of points in the phase space where they satisfy at least one of the following conditions

$$\begin{cases} \frac{dX_1}{dt} = 0 \\ \frac{dX_2}{dt} = 0 \end{cases}. \tag{2.24}$$

Geometrically, it means that X_1- or X_2-nullclines are curves in the phase space where they do not evolve along the X_1- or X_2-coordinate, respectively. The intersection of the nullclines represents the equilibrium of the system, i.e., $\frac{dX_1}{dt} = \frac{dX_2}{dt} = 0$. These curves provide boundaries in the phase space that divide up the phase space into different regions. For instance, X_1-nullclines ($\frac{dX_1}{dt} = 0$) divide the phase space into regions where $\frac{dX_1}{dt}$ is either positive or negative. Therefore, if we cross over the X_1-nullclines from one region to another, the direction of $\frac{dX_1}{dt}$ will change while the direction of $\frac{dX_2}{dt}$ remains unaffected. By estimating the directions of the vector fields in these regions, one can visualize and sketch the motion of the trajectory without solving the differential equations. The definition can also be applied to higher dimensions.

Nullclines are useful for sketching the phase space and analyze the behavior of trajectories qualitatively. However, to the best of our knowledge, nullclines have not been used when the original differential equation is unknown. In [P2], we introduce an empirical approach to obtain the nullclines on the reconstructed phase space (see Section 4.2.2).

3 Supervised Machine Learning Models

Supervised learning methods in machine learning aim to classify a new observation by inducing a decision algorithm from a set of labeled training data. In general, the best choice of the classifier is unclear and depends predominantly on the nature of the input (i.e., the underlying data production process) and the data size. Thus, choosing the right classifier often requires trial and error [31]. Therefore, we use different supervised learning techniques to tackle our classification tasks. This chapter provides an overview of the machine learning models used in the thesis.

In this thesis, we use four types of classifiers, namely, Bayesian models, random forest, Gradient boosting, and Artificial Neural Network (ANN) models. Bayes-based models such as naïve Bayes, linear, and quadratic discriminant analysis (LDA and QDA) are based on the Bayes decision rule. These classifiers are considered to be simplistic, with only a few parameters to be learned. However, despite their naïve assumptions and simplicity, they have shown surprisingly promising performances in various medical problems [32] [33].

The second and third types of classifiers covered in this thesis are decision tree models. Once they are trained, they can be seen as a set of if-then rules. Hence, their results are simple to comprehend. There are different variants of decision tree models. In the more advanced alternatives, ensemble learning techniques are applied to decision tree models, including bagging and boosting techniques, to make more robust decisions. As in many versatile applications, decision tree models have been successfully utilized in the medical field [34] [35] [36] [37].

The fourth type of classifier is the feed-forward fully-connected ANNs. An ANN is composed of interconnected layers of neurons mapping the input observations to their asso-

ciated labels. Currently, ANNs, especially deep ANNs such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) recurrent networks, are the popular approach of ML [38] [39] [40]. However, in this thesis, we have only used conventional ANNs.

3.1 Bayes Decision Models

Naïve Bayes

Naïve Bayes [41] models the probability of a class using Bayes' rule,

$$\mathcal{P}(y|\boldsymbol{s}) = \frac{p(\boldsymbol{s}|y)\mathcal{P}(y)}{p(\boldsymbol{s})}, \tag{3.1}$$

while assuming the features are conditionally independent. Under this assumption, naïve Bayes treats the posterior probability as the product of a univariate normal distribution. Thus,

$$\mathcal{P}(y|\boldsymbol{s}) = \frac{\mathcal{P}(y) \prod_{i=1}^{n} p(s_i|y)}{\sum_{k=1}^{K} \mathcal{P}(y=k) \prod_{i=1}^{n} p(s_i|y=k)}, \tag{3.2}$$

where y, K, $\boldsymbol{s}$, and n are the class index, the number of classes, the feature vector, and the number of features, respectively. Once the posterior probabilities of the observation for all classes are calculated, the observation is assigned to the class having the maximum posterior probability. In this thesis, we used a multinomial distribution [42] for $p(s_i|y)$ instead of the univariate normal distribution as we are dealing with categorical features (for more details, see Section 4.2.1).

Linear and Quadratic Discriminant Analysis

In Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) classifiers, the prediction is carried out by Bayes' rule as,

$$\mathcal{P}(y|\boldsymbol{s}) = \frac{p(\boldsymbol{s}|y)\,\mathcal{P}(y)}{\sum_{k=1}^{K} \mathcal{P}(\boldsymbol{s}|y=k)\mathcal{P}(y=k)}. \tag{3.3}$$

Here, for the likelihood $p(\boldsymbol{s}|y)$, a multivariate normal distribution $\mathcal{N}(\mu_y, \Sigma)$ with mean μ_y and covariance Σ is used as follows:

$$p(s|y) = \frac{e^{-\frac{1}{2}(s-\mu_y)^T \Sigma^{-1}(s-\mu_y)}}{\sqrt{(2\pi)^y |\Sigma|}},$$ (3.4)

where $|\Sigma|$ and Σ^{-1} are the determinant and its inverse, respectively. Often μ_y (the class means) and Σ ($\Sigma = \frac{1}{N}(s-\overline{s})(s-\overline{s})^T$) are estimated empirically using the training data, where s denotes an $s * N$ matrix with N observations and s features. We used the uniform prior probability, $P(y) = \frac{1}{\Sigma_{l=1}^{k} y}$.

The objective is then to find the highest posterior probability [43]

$$argmax_y P(y|s) = \ argmin_y (s-\mu_y)^T \Sigma^{-1}(s-\mu_y).$$ (3.5)

The difference between LDA and QDA is that LDA assume a common covariance matrix Σ for all classes, while in QDA the covariance matrix is estimated for each class separately, Σ_y.

3.2 Random Forest

Random forest, as its name implies, is an ensemble of decision trees. Thus, decision tree learning is first described in this section, and then the random forest is briefly introduced. A decision tree forms a tree structure by splitting a root node into child nodes. This process is carried out repeatedly and tries to minimize the training error in each leaf node (final nodes). The learned model can then be interpreted as a set of if-then rules, which makes the decision easy to understand. There are different tree-based algorithms, such as ID3 [44], C4.5 [45], and CART [46]. However, all of them have been developed based on the same principle, hence we briefly describe only the classification and regression trees (CART) algorithm.

CART is a decision-tree-building algorithm that can handle both classification and regression tasks. The model constructs a tree through successively binary partitioning of observations. For each node, it calculates the impurity for all the given features. Then, the feature with minimum impurity will be chosen for that node. In fact, at each node, it looks for a feature, which can predict the target with the highest accuracy. For this purpose, CART uses the Gini index [47]. The Gini index is obtained for each feature by subtracting the sum of the squared probability of each class k from one,

$$G = 1 - \Sigma_{k=1}^{K}(P_k)^2.$$ (3.6)

Then, the feature with the lowest Gini index is selected for the node. The main idea is that each child only includes a set of observations that belongs to the same class. Thus, CART suggests partitioning the observations into subsets in a way that they are purer in the children nodes than the ones in their parent nodes. In particular, the tree is built using the following three main steps:

1) Find the best split for each feature, i.e., cutting points,
2) Find the best split for the current node using the previous step,
3) Split the node using the best split in the previous step.

The pseudocode for the CART algorithm is provided below.

CART pseudocode (node splitting) algorithm with continuous values

- **Start** at the root node
- **While** (*Stopping Rule = False*) **do**
    - **For** s_i $(i = 1,2, ..., \mathcal{s})$ **do**
        - **Sort** the values $v \in s_i$, in ascending order
        - **For** $v \in s_i$ **do**
            - **Calculate** the impurity gain:

$$\Delta G = \mathcal{P}(\mathcal{T})G_{i,T} - \mathcal{P}(\mathcal{T}_L)G_{i,\mathcal{T}_L} - \mathcal{P}(\mathcal{T}_R)G_{i,\mathcal{T}_R}$$

            ($\mathcal{T}$ is the set of observations in the node, $\mathcal{T}_L$ and $\mathcal{T}_R$ are the sets of observations in left and right nodes, respectively. $\mathcal{P}$ denotes the sum of probability of observations in the current node)

        - **Find** the v with $max(\Delta G)$ *(i.e., cut point)*
    - **Find** s_i with $max(\Delta G)$ for the current node.

The fitted tree with the CART algorithm is too specific for the training data, which rarely generalizes well to unseen data thus increasing the risk of overfitting. The reason is that the Gini index, compared to entropy, produces more pure nodes with smaller partitions. This leads to a tree with more depth. Moreover, the CART algorithm is designed so that the tree growth continues until it reaches its full depth. Therefore, pruning is performed to avoid overfitting. The tree can be pruned based on stopping rules (pre-pruning). In pre-pruning, the tree is forced to stop splitting if the leaf node contains no observations, if further splitting might not improve the performance significantly, or if after partitioning, the distributions of the classes in the observations remain the same.

An alternative solution for overfitting is post-pruning. This means that pruning starts after the tree is built into its full depth. In this approach, despite the pre-pruning, a validation set is needed to prune the tree in a cross-validation scheme. For this purpose, first, the

training data is randomly split into $\mathcal{K}$ disjoint folds. Then, $\mathcal{K} - 1$ folds are used to construct different pruned trees and then estimate their performance on the last fold. This process repeats until the performances of the designed trees are evaluated on all the folds. Finally, the tree that yields the minimum average error is chosen as the pruned tree.

As mentioned earlier, decision trees are easy to interpret. They can handle missing data, categorical and numerical features, and require minimum preprocessing. Decision trees can also be used implicitly for feature selection. However, a single tree may not have robust performance. This happens because introducing a perturbation into the dataset may decrease the generalization power of the tree. Additionally, due to the greedy strategy used for learning, they may not find the best splitting rules. Often, the tree-ensemble methods such as random forest and gradient boosting approaches are employed to somewhat alleviate these issues.

In a random forest, each tree votes for each class, and the final decision is made based on the aggregation of these votes (see Figure 3.1). To decrease the correlation between different trees and diversify them, a bootstrap algorithm is used in both observation and feature levels. To be more specific, the training data with N observations is randomly partitioned into n subsets, usually with size N ($n \ll N$) with replacement. Thus, the same observation can fall into different partitions with repetition. On the other hand, it is likely that an observation may not appear in some partitions.
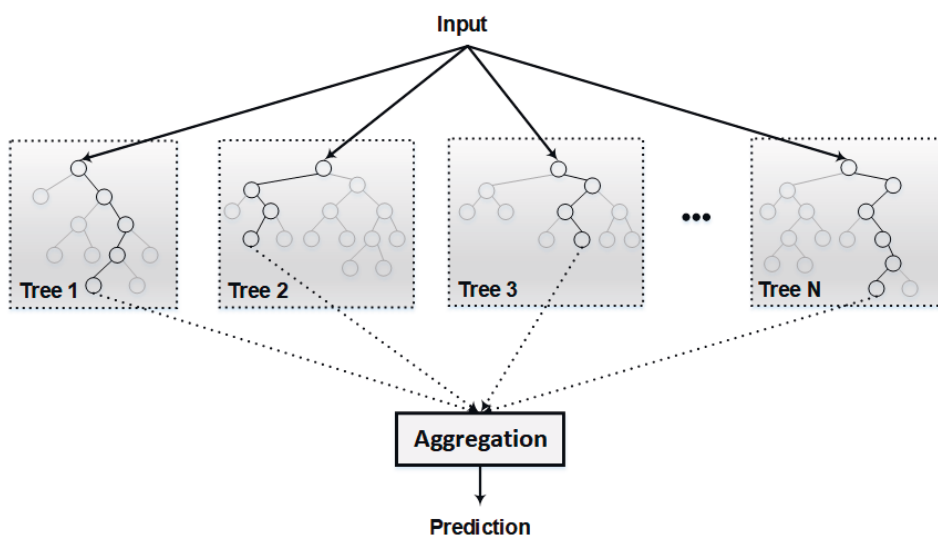

Figure 3.1 Random forest classifier illustration during the testing phase.

Moreover, while constructing each tree, instead of considering all s features, only a subset of them with size m ($m \ll s$, m is usually the square root of s) are randomly chosen and used for each node. Therefore, the same feature can be selected multiple times in different nodes of a tree. In this algorithm, all the trees are trained independently. Finally, during the test, the aggregation of the trees' predictions will be used as the final decision.

Imposing randomness at the feature level and observation level creates almost uncorrelated trees. This often makes a tree-committee that produces more accurate and robust results in comparison to a single tree. Moreover, using multiple trees leads to a higher-order interaction between features, which improves the classification performance.

3.3 Extreme Gradient Boost

Gradient Boosting

Gradient Boosting [48], similar to random forests, is an ensemble method based on the boosting algorithm. The boosting algorithm is a sequential ensemble technique where the performance of the model is improved by assigning a higher weight to the previous incorrectly classified samples (AdaBoost [49]). Unlike random forest, the weak learners (usually decision trees) are trained sequentially in this model. The process continues until all observations fall into the right class, or a specified constraint is met. During the testing phase, the prediction of unseen data is made based on the weighted sum of the predictions made by all the learners. The major disadvantage of AdaBoost is that it is unstable in the presence of outliers in the training data.

In gradient boosting, instead of weighting the observations, the loss function of the previous learner is optimized. Therefore, the current tree is more accurate than the previous one. In other words, in gradient boosting, the errors made by the previous learner are learned. One of the main advantages of gradient boosting is that the loss function can be specified according to the problem of interest. Let us assume a simple binary classification problem. The gradient boosting process starts with a single leaf where an initial prediction is made for all the observations in the training data using the logistic function as follows:

$$\mathcal{P} = \frac{e^{log(odds)}}{1 + e^{log(odds)}},\tag{3.7}$$

where $odds$ reflects the likelihood that an event takes place.

Then, the residual is achieved for each observation and is defined as the difference between the actual and the current prediction,

$$Res = observed - \mathcal{P}. \tag{3.8}$$

Afterward, the given features are used to create a decision tree to predict the obtained residual from the previous step. The output of each leaf in the trained tree is calculated as

$$Output\ of\ leaf = \frac{\sum_i Res_i}{\sum_i (\mathcal{P}_i(1-\mathcal{P}_i))}, \tag{3.9}$$

where i is the number of observations that fall in the current leaf. Res_i is the ith residual that falls in the leaf, and $\mathcal{P}_i$ is the last prediction made for the corresponding observation. Next, the predictions for each observation are updated by combining the initial predication with the output of the built tree,

$$\text{Pr}^{t+1} = \frac{e^{\mathcal{P}t + \alpha(Output\ of\ leaf)}}{1+e^{\mathcal{P}t + \alpha(Output\ of\ leaf)}}, \tag{3.10}$$

where α is the learning rate. This process iteratively continues until a specified (maximum) number of trees is reached, or Res becomes smaller than a given threshold.

In a general form, the algorithm can be formulated as follows. Suppose the given data is $\{(s_i, y_i)\}$, where $i = 1,2,...,n$ is the i'th observation, and s_i and y_i denote the feature vector and its corresponding label, respectively. The objective is to learn an ensemble of learners as

$$F(s) = \sum_{t=0}^{T} B_t \hbar_t(s), \tag{3.11}$$

where $\hbar_t(s)$ is a weak learner and β_t is the expansion coefficient at iteration t. In the first iteration, $t = 0$, using a constant initial prediction γ (recall that γ refers to the logarithm of the $odds$ in Eq. 3.10), the prediction function is calculated as,

$$F_0 = argmin_\gamma \sum_{i=1}^{N} L(y_i, \gamma), \tag{3.12}$$

where F_0 denotes the prediction function of the first iteration. To find F_0, the derivative of the loss function L with respect to γ is calculated. Then for the remaining iterations $t = 1:T$, the subsequent F_t is recursively calculated as,

$$F_t(s) = F_{t-1}(s) + \beta_t \hbar_t(s). \tag{3.13}$$

As mentioned, at each iteration leaner h_t is trained to predict the residual of the previous learner's loss function L. Thus, the gradient of the loss function in Eq. 3.12 is computed as,

$$Res_{i,t} = -[\frac{\partial L(y_i, F_{t-1}(s_i))}{\partial F_{t-1}(s_i)}]. \tag{3.14}$$

This demonstrates that the residual corresponds to the running gradient descent on the loss function. β_t in Eq. 3.13 is estimated at each iteration in the same manner,

$$\beta_t = argmin_\beta \sum_{i=1}^{N} L(y_i, F_{t-1}(s) + \beta k_t(s)). \tag{3.15}$$

The critical point is that in gradient boosting, only a subset of the training data is randomly chosen and used to obtain F_t at each iteration t. This introduces randomization to the process, which improves both the accuracy and computational speed.

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) [50] is a variant of the Gradient Boosting algorithm that improves the model by adding a regularization term to the objective function. Here we briefly summarize the regularization framework. In XGBoost, the objective function is defined as

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(k_t), \tag{3.16}$$

where the first term is the training loss and shows how well the model is fitted. The second term denotes the regularization, which penalizes the complexity of the trees. XGBoost uses the following function as regularization

$$\Omega(k) = \ell L + \frac{1}{2}\lambda\|\omega\|^2, \tag{3.17}$$

where L is the number of leaves in tree k. ω denotes the score of the leaves. ℓ and λ are the regularization parameters. This encourages the model to have a smaller number of leaves and more smooth weights to create a robust tree.

$$\hat{y}_i^{\,t} = \hat{y}_i^{\,t-1} + F_t(s_i) \rightarrow Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{\,t-1} + F_t(s_i)) + \sum_{t=1}^{T} \Omega(k_t). \tag{3.18}$$

In fact, by adding the regularization term, the pruning process is embedded into the objective function. By expanding the loss function L using the second-order Taylor expansion, the optimization problem changes to

$$Obj = \sum_{i=1}^{n}[L\left(y_i, \hat{y}_i^{t-1}\right) + F_t(\boldsymbol{s}_i) \frac{\partial L(y_i, \hat{y}_i^{t-1})}{\partial \hat{y}_i^{t-1}} + \frac{F_t(\boldsymbol{s}_i)^2}{2} \frac{\partial^2 L(y_i, \hat{y}_i^{t-1})}{\partial^2 \hat{y}_i^{t-1}}] + \sum_{t=1}^{T} \Omega(\mathcal{k}_t). \quad (3.19)$$

The motivation for using Taylor expansion is to achieve the above quadratic form. This makes the objective function not only more efficient but also increases the flexibility to use different weak learners and loss functions.

Besides the regularization, instead of searching greedily for the cut point for each continuous feature, it proposes some candidates based on the percentiles of the feature's distributions making the process more efficient. Similar to random forest, XGBoost subsamples the features to decrease the chance of overfitting and also increase the speed of computations. Moreover, XGBoost has other features such as memory efficiency, optimization of the hard disk space for big data, and parallel learning.

3.4 Artificial Neural Networks

An Artificial Neural Network (ANN) can be considered as a universal function approximator [51]. ANNs are inspired by the human brain and the nervous system and designed using interconnected layers of neurons. Each neuron in layer l is connected to (some of) the neurons in the previous layer l-1. Therefore, as the input flows within the hidden layers, more salient and abstract level features are expected to be extracted. An ANN with one input feature, one hidden and one output node with a nonlinear sigmoid activation function corresponds to logistic regression (see Figure 3.2) [52]. However, more complex ANNs are often needed to be able to learn the pattern of interest in a given dataset.

The main idea of learning here means to adjust the parameter space (weights and biases) in such a way to minimize the cost function. For this purpose, the back-propagation learning technique is applied. It recursively uses the chain rule to compute the gradient in the parameter space and consists of two passes through the network layers. In the forward propagation, the prediction is made using the given set of parameters, while in the backward propagation, the parameters are updated to minimize the cost function.

To be more specific, in the forward propagation, the intermediate variables are calculated sequentially from the input layer to the output layer. In this process, each neuron computes the weighted sum of the neuron's inputs plus the bias. Then, the activation function is applied to the obtained values, in layer l, to produce the neuron's output, $\boldsymbol{x}^l$, as

$$\boldsymbol{z}^l = W\boldsymbol{x}^{l-1} + b, \quad (3.20)$$

$$x^l = g(z), \tag{3.21}$$

where W is the weight parameter, b denotes the bias, and g indicates the activation function. Once the final layer's output is calculated, the loss function L is obtained to measure how well the output value is predicted for the given input. This operation is performed for every observation sample to obtain the cost function J as,

$$J = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \hat{y}_i), \tag{3.22}$$

where y_i and $\hat{y}_i$ are the class label and prediction, respectively. In the backward propagation, the gradients of the initial weights and biases are computed for optimizing (i.e., minimizing) the cost function. For this purpose, the gradient descent is used to compute the gradients form the output layer and propagate them backward. Once the gradients are computed, each weight and bias is updated as follows:

$$W^l = W^l - \alpha \frac{\partial J(W,b)}{\partial W^l} \tag{3.23}$$

$$b^l = b^l - \alpha \frac{\partial J(W,b)}{\partial b^l} \tag{3.24}$$

where α is the learning rate. Partial derivatives make it possible to measure the change in the cost function with respect to a specific weight or bias, which is essential to minimize the cost function. To do this, the chain rule is utilized as follows:

$$\frac{\partial J(W,b)}{\partial W^l} = \frac{\partial J(W,b)}{\partial x^l} \frac{\partial x^l}{\partial z^l} \frac{\partial z^l}{\partial W^l} \tag{3.25}$$

$$\frac{\partial J(W,b)}{\partial b^l} = \frac{\partial J(W,b)}{\partial x^l} \frac{\partial x^l}{\partial z^l} \frac{\partial z^l}{\partial b^l} \tag{3.26}$$

In this thesis, we used the simplest form of ANNs, i.e., the fully-connected and feedforward ANN with one or more hidden layers and neurons. Additionally, the hyperbolic tangent function is used as the activation function. For training, we used Bayesian regularization backpropagation [53], which is based on the Levenberg–Marquardt algorithm [54].
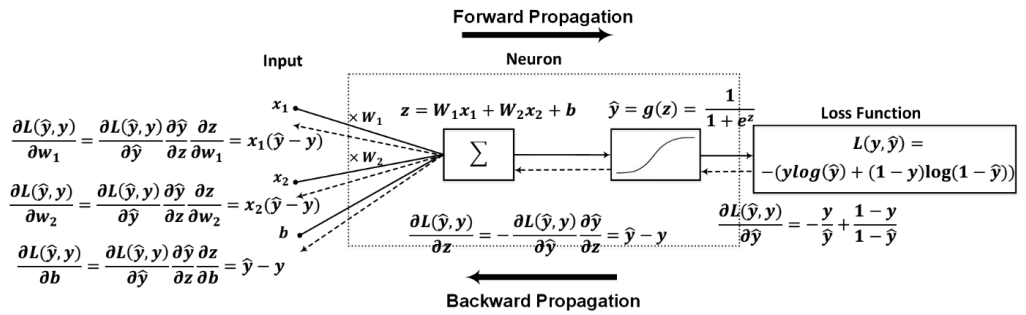
Figure 3.2 The gradient descent for logistic regression

4 Contributions

This chapter summarizes the main contributions in the publications included in the thesis. These contributions can be grouped into three categories:

- Proposing a novel set of descriptors based on nonlinear dynamics and its application in epileptic seizure detection ([P1] and [P2]).
- Developing several classification methods and investigating a comprehensive set of discriminative features for Atrial Fibrillation (AF) detection ([P4]), identification of cardiac anomalies ([P3]), and early sepsis prediction ([P5]).
- Implementing the methods proposed in publications [P3], [P4], [P5], and making them available as open-source software.

This chapter is organized as follows: First the performance metrics are described in Section 4.1. Then, the proposed nonlinear dynamics Electroencephalogram (EEG) analysis and a classification framework for epileptic seizures are discussed in Section 4.2. Next, in Section 4.3, the developed solutions for three biomedical applications are presented. These tasks include AF detection using single-lead Electrocardiogram (ECG) signals (Section 4.3.1), cardiac anomaly detection using phonocardiogram (PCG) signals (Section 4.3.1), and early sepsis prediction using clinical data (Section 4.3.3).

4.1 Performance Metrics

Evaluation of learning algorithms plays a vital role in ML. A fair assessment can provide insight into the performance of the model and how it can be improved. In this section, the used evaluation metrics in this thesis are summarized.

The standard performance measures used in [P1] and [P2] are $Sensitivity$, $Specificity$, and $Accuracy$, which are defined as follows:

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.2}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.3}$$

where TP (True Positive) and FP (False Positive) are the number of positive samples, which are classified correctly and incorrectly, respectively. Similarly, TN (True Negative) and FN (False Negative) are the number of negative samples classified correctly and incorrectly, respectively.

In [P3], the $Overall\ Score$ is defined based on a modified version of $Sensitivity$ and $Specificity$ to consider the weight of noisy samples as well.

$$Overall\ Score = \frac{Sensitivity_{modified}+Specificity_{modified}}{2} \tag{4.4}$$

$$Sensitivity_{modified} = w_{a_1} \frac{Aa_1}{Aa_1+Aq_1+An_1} + w_{a_2} \frac{Aa_2+Aq_2}{Aa_2+Aq_2+An_2} \tag{4.5}$$

$$Specificity_{modified} = w_{n_1} \frac{Nn_1}{Na_1+Nq_1+Nn_1} + w_{n_2} \frac{Nn_2+Nq_2}{Na_2+Nq_2+Nn_2} \tag{4.6}$$

where w_{a_1}, w_{a_2}, w_{n_1}, w_{n_2}, Nn_1, Nn_2, Aa_1, Aa_2, Aq_1, Aq_2, Nq_1, and Nq_2 are defined as,

$$w_{a_1} = \frac{clean\ positive\ samples}{total\ positive\ samples} \tag{4.7}$$

$$w_{a_2} = \frac{noisy\ positive\ samples}{total\ positive\ samples} \tag{4.8}$$

$$w_{n_1} = \frac{clean\ negative\ samples}{total\ negetive\ samples} \tag{4.9}$$

$$w_{n_2} = \frac{noisy\ negative\ samples}{total\ negetive\ samples} \tag{4.10}$$

Table 4.1 The confusion matrix defined in [P3]

		Predictions		
		Negative (normal samples)	**Uncertain**	**Positive** (abnormal sampels)
	Negative, clean	Nn_1	Nq_1	Na_1
	Negative, noisy	Nn_2	Nq_2	Na_2
Ground Truth	**Positive, clean**	An_1	Aq_1	Aa_1
	Positive, noisy	An_2	Aq_2	Aa_2

In [P4], the $F\_measure$ is defined as the average of $F1$ score for each class.

$$F_{measure} = \frac{F_{1n}+F_{1a}+F_{1o}+F_{1p}}{4} \tag{4.11}$$

where,

$$F_{1n} = \frac{2(normal\ samples\ classified\ correctly\ )}{Total\ normal\ samples} \tag{4.12}$$

$$F_{1a} = \frac{2(Atrial\ fibrillation\ samples\ classified\ correctly\ )}{Total\ Atrial\ fibrillation\ samples} \tag{4.13}$$

$$F_{1o} = \frac{2(Other\ samples\ classified\ correctly\ )}{Total\ other\ samples} \tag{4.14}$$

$$F_{1p} = \frac{2(Noisy\ samples\ classified\ correctly\ )}{Total\ noisy\ samples} \tag{4.15}$$

In [P5] the Utility Score $U(s,t)$ is defined as a function of patient s and time interval t.

$$U(s,t) = \begin{cases} U_{TP}(s,t), & positive\ prediction\ at\ time\ t\ for\ sepsis\ patient\ s \\ U_{FN}(s,t), & negative\ prediction\ at\ time\ t\ for\ sepsis\ patient\ s \\ U_{FP}(s,t), & positive\ prediction\ at\ time\ t\ for\ non-sepsis\ patient\ s \\ U_{TN}(s,t), & negative\ prediction\ at\ time\ t\ for\ non-sepsis\ patient\ s \end{cases} \quad (4.16)$$

The utility function rewards or penalized the predictions. To be more specific, the utility function rewards if the model predicts sepsis between 12 hours before and 3 hours after the sepsis occurs. Moreover, the utility score penalizes the model if it does not predict sepsis or predict sepsis more than 12 hours before or late (more than 3 hours after sepsis occurs) prediction. The weights of such penalties and rewards vary between patients with and without sepsis. For more detailed information, refer to [55].

In addition, the area under the receiver operation characteristic (AUROC) is used in [P5]. AUROC is calculated as a trade-off between $Sensitivity$ and false positive rate. False positive rate is defined as follows,

$$false\ positive\ rate = \ 1 - specificity \quad (4.17)$$

Contrary to $Accuracy$, the AUROC provides a more informative measure for imbalanced problems as it includes both $Sensitivity$ and $specificity$. However, if the numbers of negative samples are excessively higher than the positive samples, AUROC doesn't reflect the FP in $Specificity$ (see Eq. 4.2).

The area under the $Precision$ -recall curve (AUPRC) is another evaluation metric used in [P5]. Recall is equal to $Sensitivity$ and $Precision$ defined as follows,

$$Precision = \ \frac{TP}{TP+FP} \quad (4.18)$$

This measure elucidates the performance of a model that has a relatively high sensitivity but low precision. This means that the majority of samples are classified as positive samples. Moreover, In AUPRC TN is not considered. Often in biomedical problems, number of negative observations compared to positive samples are high. Therefore, AUPRC is not sensitive to such unbalanced classes and can provide a more clear interpretation.

4.2 Nonlinear Dynamics in Electroencephalogram Analysis

Motivation

Epilepsy is one of the most common disorders affecting almost 50 million people world-wide. Epilepsy is defined as a neurological condition where a person has an intrinsic risk of having more than one seizure. It is noteworthy to highlight that seizures and epilepsy are different. Seizures are temporary conditions within the brain in which the brain function is disturbed or altered. Seizures can be epileptic or non-epileptic. Non-epileptic seizures are triggered because of a head injury, central nervous system infection, brain tumor, lack of oxygen, or chemical imbalance. On the other hand, epileptic seizures are not usually provoked. Epileptic seizures can be associated with one or multiple physical causes, such as structural brain problems, metabolic disorders, or inherited genes.

From the neurobiology perspective, the occurrence of seizures is associated with a communication distraction between brain neurons. Seizures can originate at different cellular levels, such as alterations in ions, cell membranes, synapses, and microstructural changes [56]. The underlying reason for such variations in the nervous system is caused by multiple (potentially unknown) intertwined mechanisms. Due to the complexity of these mechanisms, it is difficult to determine the underlying cause. To narrow down the causes and facilitate the diagnosis, epileptic seizures are categorized into different types based on the involved location in the brain and the symptoms. For example, focal seizures are defined where only a part of the brain is affected and can cause a strange taste in the mouth, fear, or stiffness in the patient. Another example is generalized seizures, where a focal seizure spreads out to both sides of the brain and causes loss of consciousness and convulsion.

The other challenge of epilepsy is that seizure mechanisms can evolve through time [57]. Thus, the severity and frequency of seizures between and within epileptic patients vary drastically, making epilepsy a tremendously diverse condition. The most common medical test for epilepsy is Electroencephalogram (EEG). EEG acquisition during seizures provides valuable information for neurologists to determine the type of seizure. However, due to the nonlinearity involved in the mechanism of epilepsy, seizure detection is not a straightforward task, and as a result, a high disagreement can occur among epileptologists [58] [59].

Several well-established Signal Processing and ML methods have been developed for automatic seizure detection [60] [61] [62] [63]. Notably, it has been shown that time-frequency features can capture discriminative characteristics between seizure events

and background EEG activities [64] [65]. The basis of these spectral methods is the decomposition of the signal into different frequencies and time scales. However, these methods cannot reveal nonlinear EEG dynamics. EEG signals are generated by nonlinear interactions between numerous neurons and reflect the dynamic of a large number of interrelated variables in the brain. Therefore, the EEG signal is generated from a high dimensional model. If the high-dimensional model of the original dynamic, or its shadow version, is accessible, then more accurate properties of the system can be achieved [66]. These characteristics cannot be achieved by focusing on the frequency and time properties of one-dimensional time series. An extensive description of prior work can be found in [P1] and [P2].

In publications [P1] and [P2], we proposed empirical approaches to extract a set of discriminative descriptors based on well-known techniques of nonlinear dynamics. These features were used to measure some hidden properties of the EEG time series. Afterward, a patient-specific classification framework was proposed for epileptic seizure detection based on the extracted properties. Extensive comparative analysis shows that the extracted features achieved state-of-the-art performance in differentiating seizures from non-seizure events.

Datasets and Experimental Protocols

Datasets

In these studies, we used the CHB-MIT benchmark dataset [67] [68]. The scalp EEG dataset was collected from 23 pediatric patients (males, ages 3–22; and females, ages 1.5–19) at the Boston children's hospital. Between 9 to 24 EEG recordings were collected from each patient with a sampling frequency of 256 Hz. We used 23 common EEG channels for our analysis: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8 (see Figure 4.1).

Experimental Protocol

Most research works in this domain use non-invasive scalp EEG to monitor seizures activities [69]. Often, EEG electrodes are located on the scalp based on the international 10-20 system (see Figure 4.1). Due to the high variability of seizures' types and emergence of epileptogenesis process in different brain regions, signal processing and ML models depend on the utilization of all EEG electrodes (e.g., [70]).

Commonly, there are two main approaches to design seizure detection models. The first approach is patient-independent, which means that the model is trained and tested independent of the patients. Such models are usually developed for detecting a specific type of seizure [71]. The second approach is patient-specific [72]. This means that the model is trained and validated on the recordings of the same patient.

Each of these approaches has its own advantages and shortcomings. In a patient-independent approach, the models are more convenient to be utilized in medical environments. To be more specific, it is more convenient if the same model can be used for all patients. However, this is not realistic since the brain activities profoundly vary between patients. On the other hand, the patient-specific models are tuned to detect seizures specifically for the patient of interest. Although this approach can generate more accurate detection, utilizing such models need some manual labeling for each patient, which can be a cumbersome procedure for medical experts. In this thesis, a patient-specific approach is used for seizure detection. To minimize the manual labeling in real-world scenario, only 25% of the EEGs are used for training.

For evaluating the ML models, the k-fold cross-validation procedure is used. However, this should be done with caution for time series where the temporal transient of data conveys relevant information. In a standard cross-validation scheme, the EEG segments are randomly partitioned into train and test without considering their occurrence in time. As a result, it is possible that a model is trained with $t - 1$ and $t + 1$ segments and tested with segment t. In this way, not only the causality of the model is not considered (specifically for predictive models), but also, a naïve classifier can interpolate the label for segment t. Thus, such a validation process does not provide a fair evaluation. To address this, in this thesis, the earlier (25%) data is used for training, and the rest (75%) are used for validation (test set).

4.2.1 Epileptic Seizure Detection Using Poincaré Section ([P1])

Methodology

First, for each 1s EEG segment, the phase space was reconstructed with embedding dimension 5 and time lag 6 (see Section 2.3). Then, using Principal Component Analysis (PCA), the first and second linearly uncorrelated principal components (PCs) of the 5-dimensional phase space were extracted. Afterward, a line (polynomial with the degree of one) using the least square method was fitted to the space formed by the first two PCs. The intersection points between the projected attractor and the fitted line were used to extract features. The features are listed in Table 4.2.

Once the features were extracted, they were fed into a two-layer classifier. In the first layer, the features of each channel were classified with the corresponding LDA. Then, in the second layer, the outputs (i.e., labels) of LDAs were then integrated using a naïve Bayes with the multivariate multinomial distribution. The obtained classification labels are then filtered using morphological filters based on continuity and neighborhood similarity properties.
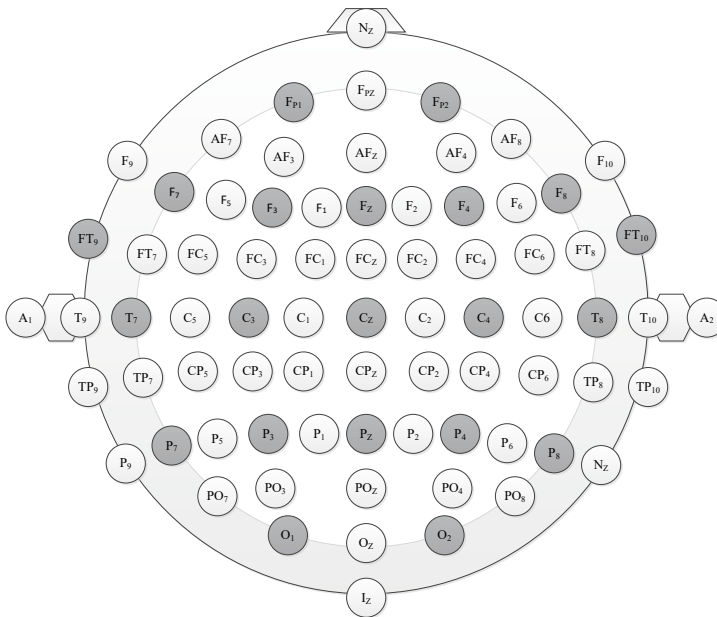
Figure 4.1 Location of electrodes in the international 10-20 system [P1].

Table 4.2 Features extracted from the intersection points of the first two principal components and the fitted line ([P1])

	Features extracted from the intersection points
1	Range
2 & 3	The 0.13 quantile and interquartile range
4	Shannon entropy
5	Root Mean Squared Amplitude
6	Coefficient of Variation
7	Energy

Results and Discussion

The proposed approach was trained with two training rates: 25% and 50%. Using the 25% training rate, we achieved an average 88.27% sensitivity and 93.21% specificity on the test set. As expected, providing more data for training (i.e., 50% training rate) improved both the sensitivity and specificity of the test set by almost 1%.

Comparative evaluations were carried out at three levels. The comparative analysis is performed against different feature sets, different classifiers, and state-of-the-art methods in this domain. The average results over the dataset for different sets of features and classifiers are shown in Table 4.3.

As can be seen, the proposed approach achieved the best results compared to different sets of features. Additionally, the proposed classification approach achieved the most robust performance among all classifiers in the experiment. Moreover, the proposed approach provides more balance between sensitivity and specificity compared to the other methods. For example, although the Support Vector Machine (SVM) classifier achieved slightly higher average sensitivity, its sensitivity was less than 60% over the EEG records of two patients. Such a low sensitivity indicates that more than 40% of the seizure frames were missed and thus makes this method unreliable for clinical usage.

The proposed feature extraction is inspired by the Poincaré section concept (discussed in Section 2.4). Although the proposed features are shown to be effective in detecting seizure events from the background EEG, they do not indicate which property of the

Table 4.3 The classification performance using different feature sets and classifiers. These results are achieved with a 50% training rate. DWT, ApEn, and MDA denote discrete wavelet transform, approximate entropy, and Mahalanobis discriminant analysis, respectively.

		Sensitivity (%)	Specificity (%)
Features	The energy of DWT Coeff.	82.80	85.18
	DWT-based APEn.	83.46	87.78
	Nonlinear features	84.74	82.57
	Proposed	**89.10**	**94.80**
Classifiers	QDA	84.66	83.77
	MDA	87.26	92.02
	Naïve Bayes	81.80	93.48
	SVM	**90.11**	94.41
	LDA (proposed)	89.10	**94.80**

system they are characterizing. Moreover, the proposed method loses (potential) information while applying PCA and fitting a line to reduce the dimension of the phase space.

In this study, we followed a real-world scenario for training and testing. The system is trained with the earlier EEG records and then tested using the later part of the data. In this case, after the system is trained using the labeled data, the system can be used to help the neurologist for the same patient over and over [38]. Another reason for such an evaluation scheme compared to the conventional cross-validation is that in cross-validation, the temporal dependencies in time series are not considered. For example, the classifier may be tested with an observation at time t, while it has been trained with observations at time $t - 1$ and $t + 1$. This is an unfair evaluation because the causality of the model is not considered. Therefore, for an accurate and fair simulation of a real scenario, we should consider standing in the present (training phase) and try to detect the unseen data (testing phase). For this, the testing data should come "*chronologically*" after the training data. One solution for this case would be using forward-chaining, i.e., successively assigning the previous (earlier) data into the training and testing on the rest (later). However, because of the highly unbalanced dataset and a very limited number of seizure events in each patient's record, this solution is not applicable to this study.

The intention of using only 25% of early data in the training phase is to prevent a high variance. Moreover, applying the proposed approach to 23 patients (which introduces randomness - that is also the primary purpose of techniques such as bootstrapping), can provide a fair evaluation for its generalization performance. On the contrary, using only 25% for training can lead to high bias error (underfitting). However, the state-of-the-art performance level still achieved in the publication reveals the discriminant power of the proposed feature over such a limited training set.

Although there are numerous EEG classification methods proposed in the literature, only a few of them consider the real-world scenario, where the limited patient history is used to train the classifier. For example, in many recent studies, the majority (i.e., >75%) of the EEG record is used for training, which is not a feasible option in a real clinical case. Therefore, we follow the same strategy for training and testing as set in [P2].

To explain the intuition behind the two-layer classification approach, we need to address two main points:

- By classifying the EEG channels separately, we do not require a *priori* information about the relevant channels. This is crucial in the seizure detection problem where the anatomic location of seizures' onset varies within and between patients.

- Labels (i.e., outputs) of the LDA classifiers (i.e., base models) can be considered as meta-features. These meta-features can provide more expressive space for learning in contrast to the base-level features. It has been shown in many applications that learning the meta-features using a second-layer classifier outperforms each of the individual base models (e.g., [73] [74] [75]). In our case, the classification of each channel individually makes it possible to use such a two-layer classification approach and achieves a higher accuracy.

4.2.2 Epileptic Seizure Detection Using Nullclines ([P2])

Methodology

In this publication, the EEG signal was filtered between 1 to 60 Hz using a band-pass second-order Butterworth filter. This frequency range covers almost all seizure events. Although there are some high-frequency oscillations associated with epilepsy, they are not visible using scalp EEG due to the conductance characteristics of the skull. Afterward, the phase space of each 1s windowed signal was reconstructed with embedding dimension 3 and time lag 31 (almost 121 ms). Then, the nullclines (discussed in Section 2.5) of the three-dimensional space were obtained. For this purpose, the analysis was performed according to the following steps:

1. Obtaining the numerical gradient of each dimension, i.e., $\frac{dX_1}{dt}$, $\frac{dX_2}{dt}$, and $\frac{dX_3}{dt}$, in the phase space.
2. Finding the zero-crossing points of $\frac{dX_1}{dt}$, $\frac{dX_2}{dt}$, and $\frac{dX_3}{dt}$.

The zero-crossing points from the second step give an estimate of the point on the nullclines. Zero-crossing was used because it satisfies the two conditions of nullclines: 1) $\frac{dX}{dt} = 0$, and 2) $\frac{dX}{dt}$ has an opposite sign on different sides of the nullcline (see Section 2.5).

In the next step, the Euclidean distances of the nullcline points from the origin were calculated. Then, the medians of these distances in each dimension were obtained as features. This results in three features (one feature for each coordinate) in total. For classification, we used the same approach as in [P1]. However, an ANN in the second layer was used to learn the relevant EEG channels and maximize the final classification performance.

Moreover, to show the feasibility of the designed method, we evaluated the nullcline analysis over five chaotic nonlinear systems; Lorenz, Rabinovich–Fabrikant, Rössler, Chua's circuit, and a modified version of Genesio system.

Results and Discussion

The proposed method using only three features achieved an average sensitivity, specificity, and AUC of 91.15%, 95.16%, 95.11%, respectively. This shows an improvement compared to the obtained results from [P1], and achieved the state-of-the-art compared to other methods applied on the same dataset.

The primary and novel contributions of this study are the fact that it was the first time (to the best of our knowledge) that the nullcline points were estimated from the reconstructed phase space without the presence of the differential equation. Moreover, this was the first time that nullcline points were used as discriminative features to detect seizure events in EEG recordings.

The confusion matrix is provided in [P2] where it can be seen that sensitivity and specificity alone are not providing a deep insight into the performance of the method. For example, consider the EEG record of patient 10 in the dataset (*Table II*, in [P2]), out of 37506 non-seizure segments 96.66% of them are correctly detected as non-seizure, and only 3.44% are misclassified. However, 3.44% here means 1250 segments (seconds). This is a high false-positive rate and requires further research. Yet, to the best of our knowledge, this is the lowest false positive rate achieved compared with the reported results over the same dataset in the literature (even with the same or even a much higher training rate).

As opposed to several other earlier studies that used a large set of features (temporal, spectral, morphological), publications [P1] and [P2] have demonstrated that the proposed feature extraction schemes based on nonlinear dynamics can single-handedly achieve a superior discrimination capability compared to all the other conventional features combined [38] (see Figure 4.2).
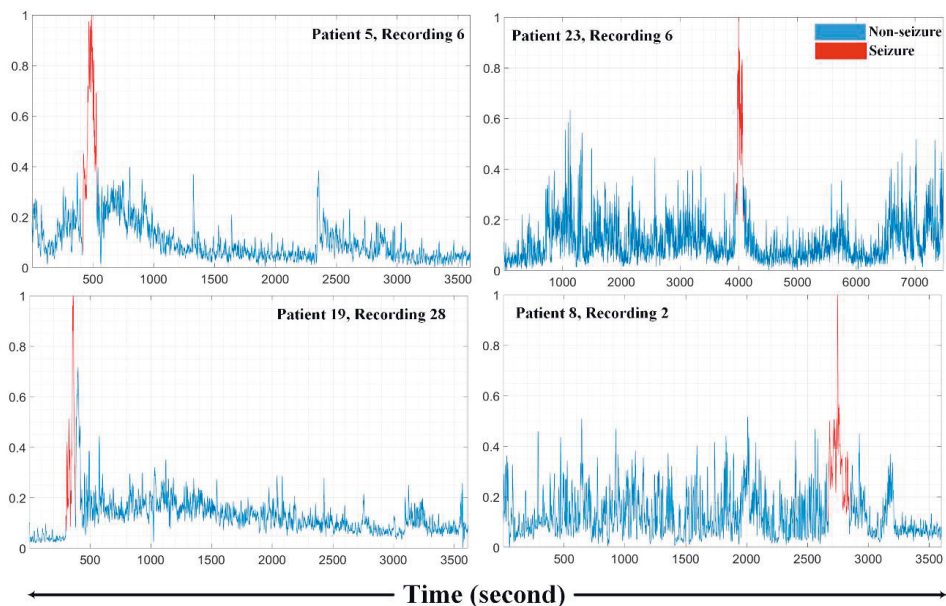
Figure 4.2 Illustration of the nullcline-based features and their discriminative power in four different patients (CHB-MIT dataset). The vertical coordinate indicates the median values of the Euclidean distances of the x-nullcline points from the origin in the phase space.

4.3 Assistive Medical Diagnostic Methods

In this section, three scientific challenges involving classification and prediction methods for different medical assistive diagnostic tasks are presented. Physionet hosts a series of challenges by providing a global platform for researchers to address clinically relevant problems, where current solutions either do not exist or fail to solve the problems adequately. The major issue with such problems is that the developed ML and Signal Processing models in the literature have been hindered by the lack of standardized and validated open databases. The main advantage of such scientific competitions is that researchers are able to evaluate their solutions under the same fair conditions. In the sequel, the developed solutions for the detection of heart anomalies using PCG, the detection of atrial fibrillation using ECG, and the early prediction of sepsis using clinical data are summarized. In this section, for each problem, we shall provide a brief overview of the state-of-the-art methods in the domain including their advantages and limitations. Then, we provide a description of the dataset, and the developed solution.

4.3.1 Heart Anomalies Detection Using PCG Analysis ([P3])

The work summarized below comes from publications [P3][1].

Motivation

Heart sound or phonocardiogram (PCG) signal is produced by contracting the heart muscles and vibrations of heart valves. PCG consists of two fundamental components: S1 and S2. S1 or the first heart sound is produced by the closure of the atrioventricular valves after blood is pumped from the atria to the ventricles. The S2 or the second heart sound is caused by the closing of aortic and pulmonary valves, just after the blood is pumped from the ventricles. The time interval between S1 and S2 shows the systole. This is the time when the ventricles contract. Conversely, the time interval between S2 and the next S1 demonstrates the diastole, which is the time that the ventricles are filling with blood.

Therefore, any anomalies in the heart sound can be seen as a sign of a pathological condition. Auscultation of heart sound can be used to detect abnormalities such as mitral or tricuspid regurgitation, aortic and pulmonic stenosis, ventricular septal defect, and prolapse of the mitral valve [76]. PCG is considered to be the primary, accessible, and cost-effective diagnostic test for cardiovascular monitoring. However, several factors make the interpretation of heart sounds a challenging task.

The PCG signals are recorded from different cardiac listening posts, which results in diverse characteristics, such as different intensity levels in S1 and S2. Besides, the accuracy of PCG analysis highly depends on the cognitive skills, the audible frequency range, and the expertise of the examiner. Moreover, body posture, respiration, and environmental noise affect heart sounds features.

In the past few decades, several studies have proposed different automated PCG analysis [77] [78] [79]. Although such models have achieved relatively high accuracy, the majority of such solutions suffer from two main limitations. Commonly, the localization (segmentation) of S1 and S2 is considered as a prerequisite of PCG analysis. Different methods, including the enveloped-based method [80], the feature-based [81], and ML

---

[1] The method proposed in publication [P3] ranked 2nd out of 47 teams in the Physionet Challenge 2016

methods [82], have been developed for this purpose. These methods can provide relevant medical information about the systole and diastole. However, any miss-segmentation in the PCG signal negatively affects the classification of pathological events.

The second limitation is the evaluation of the developed methods using a limited amount of data and the lack of PCG dataset collected from different institutes. In [83], the PCG recordings of 120 patients have been studied. In [84] and [85], 64 and 107 patients have been used, respectively. Using such limited datasets can lead to miss interpretation of the results and does not show the generalization power of the developed methods.

The main objective of this study is to overcome the mentioned shortcomings by developing a robust classification model that can detect normal and abnormal heart sounds (see Figure 4.3). The main contributions of this thesis are the detailed investigation of discriminative features and the design of an ANN-based ensemble classifier for this task without segmentation.
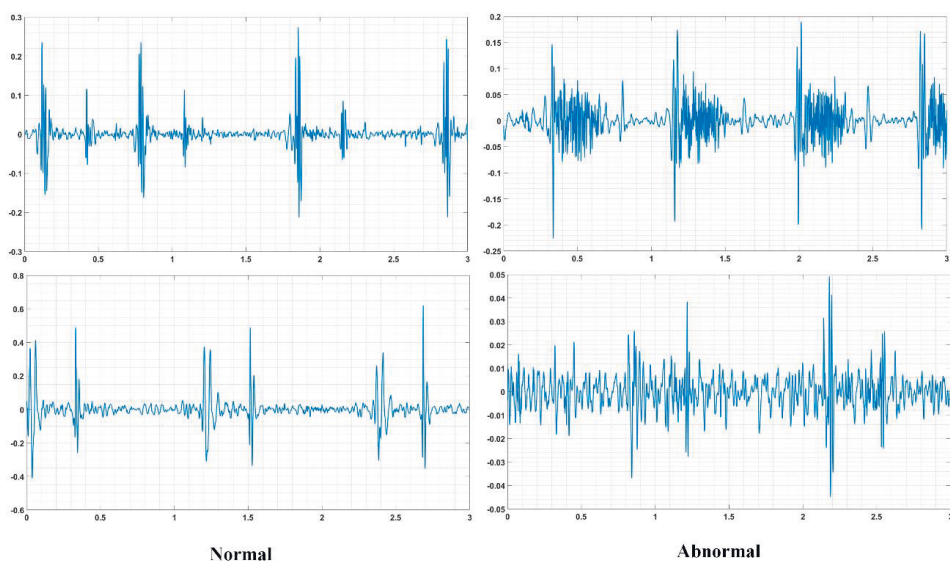

Figure 4.3 Normal and abnormal heart sound recordings. The left and the right column show normal and abnormal samples, respectively. The horizontal axis indicates the time in seconds.

Datasets and Methods

The dataset used in this study was provided by the Physionet challenge 2016, which is the largest open-access heart sound database [68] [86]. The dataset includes 4430 PCG

signals recorded with a sampling frequency of 2000 Hz from clinical and nonclinical environments. The length of the recordings varies from 5s to 121s. Each patient may have more than one PCG signal in the dataset. However, the training and test sets are mutually exclusive populations. Moreover, the locations of the recordings on the body vary between PCG signals. The training set includes 3153 records from 764 subjects, and the hidden test set consists of 1277 PCG from 308 subjects. Each recording is characterized by two types of labels: *normal or abnormal*, and *good or poor quality*.

In this study, we extracted 40 features from each PCG signal. Features were extracted from linear predictive coefficients, Natural and Tsallis entropy, Mel Frequency Cepstral Coefficients (MFCC), Wavelet-based features, and power spectral density. Then a subset of 18 features was selected using a wrapper-based feature selection scheme [87] The list of the selected features is shown in Table 4.4.

Once the feature selection was carried out, a given signal is initially classified as either *good* or *poor* or *quality*. Then, *good quality* signals were fed into the second classifier to decide whether they belong to the *normal* or *abnormal* class. For this purpose, we used an ensemble of ANNs. The classifier consists of 20 feedforward ANNs where each ANN has two hidden layers and 25 neurons in each layer.

For the imbalance problem, we used bootstrap resampling for each ANN. Finally, the results of 20 classifiers are integrated using a combination rule, which is learned in 10-fold cross-validation. A signal is classified as a *poor quality* signal if at least 17 out of 20 ANN classify it as *poor quality*. Otherwise, the signals are detected as *good quality*. Additionally, if at least seven out of 20 classifiers recognize the signal as *abnormal*, the combination rule detects the signal as *abnormal*. Otherwise, the signal is classified as *normal*.

Results and Discussion

For evaluation, the overall score was designed as the average of the modified sensitivity and specificity. The sensitivity and specificity were modified based on the percentages of *good* and *poor quality* in both *normal* and *abnormal* recordings. In Table 4.5, the performance of the proposed method using 10-fold cross-validation of the train set and the corresponding test set is shown.

As opposed to state-of-the-art methods in this domain, the proposed method is independent of S1 and S2 segmentations. To be more specific, the method is independent of S1 and S2 heart sounds detection. Although segmentation in PCG signals has shown

to be beneficial, any errors in the segmentation will propagate through the whole classi-
fication pipeline.

Table 4.4 Selected features ([P3])

Type	Features
Linear Predictive Coefficients	The 1st, 3rd, 6th, 8th, 9th, and 10th coefficients of 10th-order linear predictor
Entropy	Natural and Tsallis entropy
Mel Frequency Cepstral Coefficients (MFCCs)	$$\frac{1}{N}\sum_{j=1}^{N}\min_{i\in I}\mathcal{C}_{i,j}$$ $$E(\max_{i\in I}\mathcal{C}_{i,j}-\mu)^2$$ $$E\left(\underset{i\in I}{Skew}\,\mathcal{C}_{i,j}-\mu\right)^2,$$ where $\mathcal{C}_{i,j}$, is the MFCC coefficient at i^{th} features and j^{th} frames. E is the second-order central moment.
Wavelet transform (Daubechies 4)	$$-\sum_{i}\mathcal{P}(d_{4_i})\ln\mathcal{P}(d_{4_i})$$ $$-\sum_{i}\mathcal{P}(a_{5_i})\ln\mathcal{P}(a_{5_i})$$ $$\ln\left(\sum_{i}d_{5_i}{}^2\right)$$ $$\log_2(\sigma^2(d_3))$$ where d_3, d_4, and d_5 are the details coefficients of the 3rd, 4th, and 5th levels, respectively. a_5 is the approximation coefficient of the 5th level.
Power spectral density	$$PSD_{Centroid}=\frac{\int f PSD(f)^2 df}{\int PSD(f)^2\,df}$$ $$AUC_{PSD}(0.7-0.8)=\int_{0.7}^{0.8}PSD(f)df$$ $$AUC_{PSD}(0.9-1)=\int_{0.9}^{1}PSD(f)df$$ where $PSD(f)$ and f denotes the power spectral density and frequency, respectively.

Table 4.5 Performance of the proposed method ([P3]).

	Modified Sensitivity	Modified Specificity	Overall Score
Training set Average (standard deviation)	94.23 (2.22)	88.76 (1.96)	91.50 (1.35)
Testing set	86.91	84.90	85.90

In Table 4.6, the performance of the top 5 teams are listed. As can be seen, although the first team achieved a higher overall score ($< 1\%$) compared to [P3], they obtained a relatively low specificity of 77.81%. In [88], 59 selected time and frequency features were extracted after S1 and S2 segmentation and fed into an Adaboost classifier. In addition, each PCG cycle is decomposed into four different frequency bands (i.e. 25-45, 45-80, 80-200, and 200-400 Hz) and then fed into a 1-dimentional CNN. Finally, the outputs of two classifiers were aggregated and the final prediction was produced. Moreover in [89], a total of 131 features were extracted after heart sounds segmentation and fed into a nested set of ensemble classifiers including Random Forest and LogitBoost. Compared to the top 5 teams, the proposed method in [P3] achieved the highest specificity while obtaining a relatively high sensitivity. Such a balance between sensitivity and specificity is desired in this classification task. In addition to the relatively low computational complexity of the classifier, in [P3], only 18 features were extracted. This not only further decreases the run-time but also demonstrates the crucial role of the well-chosen features.

Table 4.6 Performance of the top 5 teams on the test (unseen) set.

Refs	Rank	Modified Sensitivity	Modified Specificity	Overall Score	Method
[88]	1	94.24	77.81	86.02	Adaboost/CNN
[P3]	2	86.91	84.90	85.90	Ensemble of ANNs
*	3	87.43	82.97	85.20	Regularized Neural Network
*	4	86.39	82.69	84.54	MFCCs, Wavelets, Tensors & KNN
[89]	5	88.48	80.48	84.48	Random Forest + LogitBoost

*no publications from competing teams have been found to date.

4.3.2 Atrial Fibrillation Detection Using Electrocardiogram Analysis ([P4])

The work summarized below comes from publications [P4][2].

Motivation

Atrial fibrillation (AF) is the most common heart arrhythmia. In a healthy heart, the sino-atrial (SA) node initiates all the electric impulses in the atria while during AF, these impulses originate randomly from many different regions called ectopic sites. These regions are commonly located near the roots of the pulmonary veins. Random impulses lead the atria to beat greater than 300 beats per minute. Often, the majority of these pulses are filtered through the atrioventricular node (due to the refectory period). Therefore, only a

---

[2] The method proposed in this publication was tied for first place out of 67 teams in the Physionet Challenge 2017.

few irregular pulses can pass to the ventricle. Consequently, this makes the heart beat irregularly and cannot pump blood effectively [90].

One serious complication of AF is the formation of blood clots. Since the amount of pumped blood into body organs is reduced, there is a chance that the blood will pool in the heart and form a blood clot, and therefore, increases the risk of a stroke [91]. AF can have symptoms such as weakness, shortness of breath, or skipped heartbeats. However, in some cases, such as silent or asymptomatic AF, the patient does not experience any symptoms at all [92]. This can be life-threatening because severe complications of atrial fibrillation can develop without any warnings.

The main and primary tool for AF diagnosis is ECG. In ECG, AF is identified as the absence of P wave because the latter corresponds to the depolarization of the atrium and atrial contraction. Additionally, the other indicator of AF is the presence of narrow and irregular QRS complexes. QRS complexes indicate the ventricular depolarization. AF can be missed and remain undiagnosed even during a regular 12-lead ECG monitoring period. Therefore, portable and hand-held devices that can monitor the ECG daily enhance the diagnosis of missed AF. However, only a few studies have exploited signals obtained from hand-held ECG devices [93] [94].

In [95], the analysis of 196 patients showed that with one-lead ECG recording, the same sensitivity for AF detection was achieved compared to using two leads ECG recordings. However, using only one-lead ECG resulted in a significantly lower (almost 10%) specificity. This demonstrates that even though one-lead ECG recordings produce more false alarms, they can still be considered effective for AF detection in non-medical environments. In [94], a total of 313 patients (including 109 and 2014 patients in the training and validation sets, respectively) were studied. The results of [94] showed that automated algorithms using a single one-lead hand-held ECG devices have acceptable accuracy for AF monitoring. A critical point in this study was the Kappa coefficient, which is a statistical measure to assess the interobservers' agreement. The 1% difference in the Kappa coefficient in the training set between two cardiologists in such a limited dataset proves the necessity of having more than one expert's labeling.

In this thesis, we developed a multi-class classification method to classify ECG signals into four classes: *AF arrhythmia, healthy rhythm*, *other arrhythmias*, and *too noisy* to analyze (see Figure 4.4). Our objective is to address the mentioned shortcomings by designing an accurate AF detection algorithm based on a comprehensive study of 500 features and evaluate their performance based on the largest publicly available dataset in this domain.
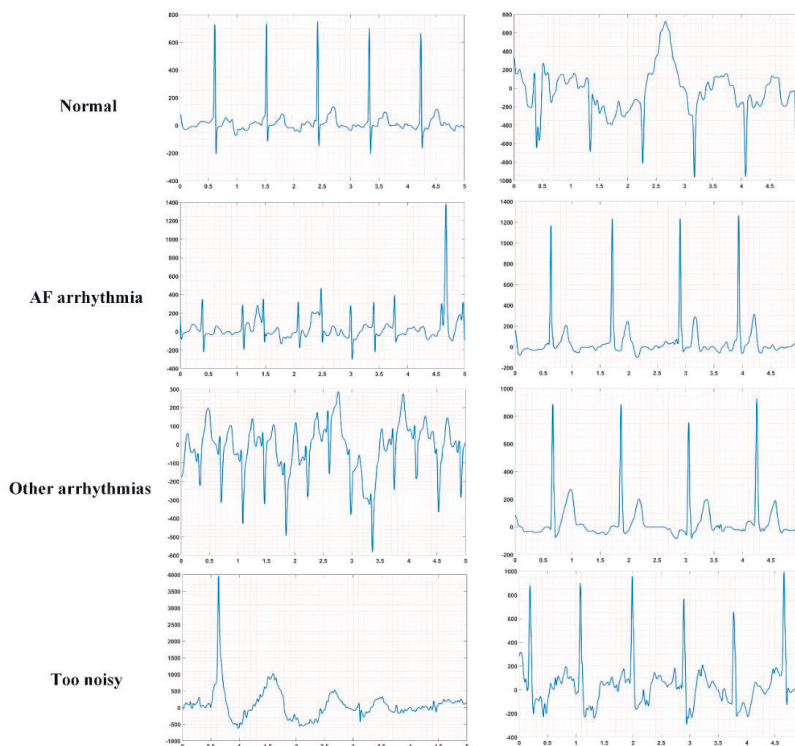
Figure 4.4 Illustration of four signal classes. Each row shows two samples from the same class. The second column shows the challenging samples in each class. The horizontal axis indicates the time in seconds.

Datasets and Methods

The ECG dataset is provided by the Physionet challenge 2017 [96] [68]. The dataset includes ECG signals from 8528 patients for training, while ECG signals from 3658 patients were kept hidden for testing. Each recording is a single-lead ECG signal collected using the AliveCor device with the sampling frequency of 300 Hz. The duration of each ECG signal varies between 9 to 61 seconds.

In this study, first, the baseline wander of each signal was removed. Then, 150 features were selected from a set of 491 features using a random forest classifier. The selected feature set consists of two types of features: base-level and meta-level features. The base-level features were extracted from time, frequency, time-frequency, and nonlinear

domains (see Table 4.7). The meta-level features were obtained based on the predictions of LDA, QDA, and random forest classifiers (base learners). The meta-level features were extracted using the following steps:

1. Splitting randomly 20% of the whole training data,
2. Windowing the signals with 5s and 4s overlap,
3. Extracting the nine features listed in Table 4.8 from each 5s segment,
4. Training the three base learners using the features obtained from the previous step,
5. The mean and standard deviation of the posterior probabilities of each 5s segment was used as meta-level features.

Once the base-level and meta-level features were extracted, they were fed into an external random forest with 500 decision trees. The external random forest was trained using the remaining 80% of the training data to avoid overfitting.

Results and Discussion

The performance of the proposed method was evaluated using F-measure. The F-measure was calculated for the training set in 10-fold cross-validation. Additionally, the F-measure is obtained for the hidden test set, as well. As can be seen in Table 4.9, the achieved results in both train and test sets are similar, which shows the robustness of the proposed method.

In this study, the primary source of misclassification is *noise* in the signals and their resemblance to the *other arrhythmia* class. This problem appears explicitly while the signals are recorded using hand-held devices, which are more prone to noise and artifact compared to standard 12-lead ECG devices. The other reason for such a low accuracy is that the total number of noisy samples provided in the training data is 46.

Table 4.9 Performance of the proposed method ([P4])

	F-measure (%)				
	Normal	**AF**	**Other**	**Noisy**	**Overall**
Training set (Average ± standard deviation)	90.49 ± 0.96	79.43 ± 4.52	75.64 ± 3.11	61.11 ± 7.53	**81.85 ±2.57**
Test set	80.87	83.51	73.41	50.42	**83**

Table 4.10 Performance of the top 4 teams on test (unseen) set.

Refs	Rank	F-measure (%)	Method
[97]	1	83.1	Rhythm/morphological features XGBoost +LSTM
[98]	2	82.9	Rhythm/morphological features AdaBoost
[P4]	3	82.6	LDA, QDA, and random forest
[99]	4	82.5	Deep features + Rhythm/morphological features LSTM + XGboost

Table 4.**10** lists the work of the teams who were tied for the first place. In [97], for each ECG signal, a description in multiple abstraction levels was provided using the *Consture* algorithm. Then, two types of features are extracted from the abductive interpretation of the signal. The first set of features is composed of 79 global features including rhythm, morphological, and signal quality features, while the second set is local features focusing on each heartbeat of ECG. Once the features were extracted, the first set of features was fed into an XGBoost classifier and the second set into a LSTM classifier. Finally, the predictions of XGBoost and LSTM were combined by a LDA to generate the final prediction using the stacking technique.

In [98], a two-layer cascade classification approach was employed, where the first classification layer classified each ECG recording into two intermediate classes (i.e., normal vs. AF). In the second layer, the actual classification was carried out. For this purpose, morphological, rhythm, frequency, and statistical features were extracted. For all classificaition levels AdaBoost was used. In [99], first, statistical, frequency, time-frequency, and rhythm-based features were extracted. Then using a 1-dimensional CNN, deep features were extracted. The combination of features was then fed into an XGBoost classifier for final classification. As can be seen, all the wining methods used more sophisticated ensemble learning methods and Deep Learning models such as XGBoost and LSTM as opposed to [P4], where a standard Random Forest was utilized. This shows the descriptive power of the extracted features in the proposed method.

4.3.3 Early Sepsis Prediction Using Clinical Data ([P5])

The work summarized below comes from publications [P5][3]

Motivation

Sepsis occurs due to a dysregulated host response to lung, urinary tract, skin, or gut infection. It is highly related to clinical changes such as a decrease in blood pressure, an increase in heart rate, body temperature, respiratory rate, elevated lactate level, and

---

[3] The method presented in this publication was ranked 3rd place out of 78 teams in the Physionet Challenge 2019.

alteration in mental status. However, sepsis is a dynamic syndrome, and all of its associated symptoms may not show up all the time. The diagnosis can get even more challenging for patients with critical conditions, such as Intensive Care Units (ICU) patients, due to the overlap between sepsis symptoms and other diseases [100].

Sepsis demands urgent attention and is considered as a life-threatening medical emergency. Although very young or old patients or those with a weakened immune system are more likely to get an infection, sepsis can impact people of all ages and levels of health. The mortality rate associated with sepsis is the cause of one out of three hospital deaths, and every year sepsis causes around six million deaths worldwide [101] [102].

Additionally, the definition of sepsis is not clear and has been evolving over the last years to reflect the latest knowledge in this domain. Therefore, due to the rapid and dynamic sepsis development and lack of certain clinical criteria, there is no standard monitoring protocol for sepsis detection. Although new clinical criteria have been developed in recent years [103], there is insufficient evidence for the efficiency of their diagnostic performance [104].

Several studies have focused on the detection and prediction of sepsis using ML to promote the performance of current clinical tools [105] [106] [107]. In [105], Artificial Intelligence Sepsis Expert (AISE) algorithm was developed for early prediction of sepsis. For this purpose, a total of 65 features, including clinical, laboratory, and demographics features were extracted from roughly 69000 ICU patients. For prediction, a modified Weibull-Cox proportional hazards model was designed to predict the onset of sepsis. The AUROC, specificity, and accuracy of 85%, 67%, and 67% were achieved for predicting sepsis 6 hours in advance, respectively.

In [106], multiple covariates, including vital signs, peripheral capillary oxygen saturation, Glasgow Coma Scale, and age, were used. Moreover in [106], the performance of the proposed model was compared against four clinical scoring systems, systemic inflammatory response syndrome (SIRS), quick sequential organ failure assessment (qSOFA), modified early warning score (MEWS), simplified acute physiology score (SAPS) II, and sequential organ failure assessment (SOFA) for comparative analysis. They achieved almost an AUROC of 75% and AUPRC of 30% for prediction of sepsis four hours in advance. These results demonstrate the superiority of ML models compared to standard medical scoring systems.

In [107], the covariates of age, diastolic and systolic blood pressures, heart rate, temperature, respiration rate, peripheral oxygen saturation, Glasgow Coma Scale, white blood cell count, and platelet count were used for sepsis prediction 4 hours in advance. The

proposed model using XGBoost obtained an average AUROC of 71.8%, a sensitivity of 75%, and a specificity of 70%.

All the mentioned studies promise improvement in sepsis prediction using ML models in this domain. However, a direct comparison between these studies is not possible due to the use of different clinical variables, datasets, and evaluation metrics. Moreover, the employed evaluation metrics in the above studies do not differentiate between too early, too late predictions, and false alarms. The Physionet Challenge 2019 dataset [55] [68], provide the possibility to address these limitations.

Our study [P5] provides a systematic approach to perform early sepsis prediction using clinical data. The main contributions of this study are:

- Proposing a comprehensive set of features to predict sepsis at an early stage,
- Introducing a feature extraction scheme for missing data values,
- Developing an ensemble of five extreme gradient boost (XGBoost) classifiers for the early detection of sepsis.

Datasets and Methods

In this study, the clinical dataset was provided by Physionet Challenge 2019 [68] [55]. The dataset was collected from 63097 ICU patients. The training and test sets include 40336 and 22761 patients, respectively. The clinical data for each patient contains eight vital signs, six demographic variables, and 26 laboratory values for every hour.

To have a predictive model, the labels were shifted ahead by six hours. Also, the data for each patient were fed to the model in a causal manner. To be more specific, at each time-stamp, only past and present observations were presented to the model. This was done to ensure that the predictive model does not have access to future data.

In this publication, two sets of features were designed. The first type of features was extracted from 13 covariates, which have less than 70% of missing values. The second type of features was modeling the pattern of absence in the clinical data. We first defined the sequence abstraction as a set of consecutive measurements where the values are only either absent (missing) or present. Once the clinical data were represented as a sequence for each patient, we extracted the second type of features. The features are listed in Table 4.11.

After feature extraction, we randomly split the training data into two disjoint sets. We used 10% of the training data to perform feature selection and hyper-parameter tuning.

For this purpose, we selected five sets of best-performing features and hyper-parameters in a 5-fold cross-validation scheme. Afterward, we trained five distinct extreme gradient boosting (XGBoost) classifiers using the five sets of selected features and hyper-parameters over 90% of the training data. Finally, during the test phase, a new observation was classified using each classifier, and the geometric mean of their outputs is used as the final decision.

The feature selection was performed using a wrapper feature selection algorithm based on XGboost. Also, for class imbalance, we used the random undersampling technique to balance the data for each classifier

Table 4.11 Features used for Early Sepsis predication ([P5])

Type	Feature
13 covariates	Mean, minimum, maximum, median, variance, 95%, 99%, 5%, and 1% quantiles from the last 5 and 11 hours
	Energy, Shannon entropy, mean of the first differences, and the lengths of observations
	last observation values of the 13 covariates
Missingness pattern	Mean and variance of the lengths of sequences along with each covariate
	Summation and variance of the lengths of sequences with only valid values (without the missing ones) along with each covariate
	Mean and variance of the lengths of sequences along with each observation in the last 5 hours.

Results and Discussion

The proposed method was evaluated using utility scores, the Area Under Receiver operating characteristic Curve (AUROC), and the Area Under Precision-Recall Curve (AUPRC). The results are shown in Table 4.12. The challenge organizers designed the utility score in a way that it rewards the prediction if the model can predict Sepsis from 12 hours before until 3 hours after the Sepsis occurrence. Also, the utility score penalizes the model if it cannot predict or predict Sepsis more than 12 hours before the actual sepsis occurrence.

As can be seen, the proposed model achieved the best results when it is tested over the data gathered from Hospitals A and B. On the other hand, most errors occurred when it is tested over the data of Hospital C. The main reason is that the training set includes data from Hospital A and B only. This shows the dynamic nature of Sepsis and the necessity of having datasets from different institutes for training a robust predictive model.

The other reason that potentially explains the performance drop is the missing patterns. The missing features represent human behavior in data recording. Although these features are able to show the correlations between the continuity of the recorded data and the sepsis occurrence, they are not medically interpretable. Therefore, the variation in missing patterns can negatively affect the robustness of the model. However, it is worth mentioning that the other contestants' performances dropped drastically on Hospital C even if they had not utilized the missing data patterns in their methods.

We observed that the *hours between hospital admit and ICU admit*, the missing patterns of *temperature*, *age*, *the administrative identifier for ICU unit*, the variance of *heart rate*, and *temperature* in the last 11 hours are ranked among the top 10 features. The feature ranking was performed by XGBoost. The importance of each feature indicates how the feature split point improves the performance. Again, it should be noted that the hours *between hospital-admit* and *ICU-admit*, and the *administrative identifier for the ICU unit* are among the non-medical features.

One feature of the proposed approach is its ability to transform the varying length data into a proper feature space so that various discriminative methods can classify it. This also enables us to perform different sampling methods. For example, we cannot merely use the resampling method if the time series data is directly used because we should consider their time order as well.

Table 4.12 The performance of the proposed method ([P5])

		Utility Score	AUROC	AUPRC	Accuracy
	A	0.422	0.814	0.102	0.803
Hospital	B	0.395	0.844	0.110	0.882
	C	-0.146	0.793	0.058	0.765

Table 4.13 Performance of the top 4 teams on test (unseen) set.

Refs	Rank	Final Score	Hospital A	Hospital B	Hospital C
[108]	1	0.36	0.433	0.434	-0.123
[109]	2	0.345	0.409	0.396	-0.042
[P5]	3	0.339	0.422	0.395	-0.146
[110]	4	0.337	0.42	0.401	-0.156
[111]	5	0.337	0.401	0.407	-0.094

The performance of the top 5 teams are listed in Table 4.13. As can be seen, the achieved final scores are almost similar between the teams. In addition, the obtained scores between Hospital A and B are robust. However, the performance of the developed models were dropped significantly when evaluated using the dataset from Hospital C. For training, no patients' data from Hospital C was provided. Such a generalization issue is more evident in the proposed solutions by [110] and [P5].

[108] achieved the best performance. They have two major contributions. The first contribution is representing the covariates as the signature of paths. The signature is a transformation that maps the given covariates into a d-dimensional space where each coordinate is a k-fold iterated integral of the covariates. The main feature of the signature is that it provides the hidden cross correlated information of the sequential data [112]. The second contribution is redefining the Sepsis labels. By modification of the labels, they changed the classification problem into a regression task, which leads to higher performance.

Interestingly, all the top teams employed XGBoost for classification. This shows the discrimination power of ensemble approaches. In addition, it proves that Deep Learning models are not always the best candidates to tackle the problem of interest. However, a recurrent neural network was used in [110] to capture the long-term time dependencies. In [110], three models were utilized based on the stay time in ICU. In the early, middle,

and long-term stages of ICU stay, XGBoost, a tree-based model, and an RNN model were used, respectively.

In this study, the missing values were treated differently by each team. In [108], a forward-fill method was used. In [109], the missing values were imputed with the most recent valid value. In [110], for the first two models, a forward-fill method was utilized. In [111], no imputation method was employed, and the NaN values were directly fed into the model. They demonstrated that imputation can lead to loss of information. As mentioned earlier, one possible explanation of the relatively low performance on Hospital C was the different pattern of missing values compared to the provided training set. As can be seen in Table 4.13, [109] and [111] achieved the highest scores in this dataset, while using either simple or no imputation method. The source of missingness in this study reveals the behavior of medical staff in recording the data. This can be a subjective recording pattern and mainly depends on the medical environment. Thus, treating missing values should be done by taking the nature of the problem into account.

5. Conclusions

Conventional clinical decision-making models are facing an abundant amount of challenges. Some of these challenges include the dynamic nature of medical conditions, the correlations between numerous clinical covariates, the high biosignal pattern variation within- and between-patients, real-time monitoring, and the exponential growth of clinical data, to name just a few. Accordingly, traditional clinical decision-making methods need to be revised. Biomedical signal analysis attempts to address these challenges by providing assistive tools upon which clinicians may gain additional insights on health status and diagnosis. However, there is no global approach for the diagnosis of all medical conditions. In this thesis, we focused on developing Signal Processing and ML approaches for four prevalent medical conditions.

The first objective of the thesis is to answer two questions: a) can we develop a set of discriminative nonlin-ear descriptors for Electroencephalogram (EEG) signals? and b) can we build a framework for epileptic seizure detection based on these nonlinear features? In publications [P1] and [P2], novel sets of discriminative features were introduced for seizure detection, which are inspired by nonlinear dynamics. The proposed patient-specific system achieved a relatively high performance compared to competing methods in this area. Also, the low computational complexity of the proposed method makes it independent of cloud computing implementations. Furthermore, due to the generic nature of the proposed features, they can be utilized in other applications, with minor adjustments. The findings in these publications adequately answer the two aforementioned research questions. Although the proposed methods yield high specificity, the false alarm rate is still high and may not be a practical solution for epilepsy monitoring units. This demonstrates that the usual evaluation metrics, i.e., sensitivity and specificity, in this domain, may not be sufficient to assess the performance and more adequate measures are needed, e.g. reporting the full confusion matrix for the proposed methods.

The second objective is to develop Signal Processing and ML techniques for three challenging diagnosis-focused applications. In [P3], [P4], and [P5], we aim to answer the following research questions, respectively: which classification approaches are suitable for (a) detecting heart anomalies using short-term phonocardiogram (PCG) signals, (b) detecting atrial fibrillation using one-lead electrocardiogram (ECG) signals, and (c) predicting sepsis for intensive care unit patients early enough?

In publications [P3], a novel heart sound classification method was proposed to identify heart anomalies. The method was developed to classify both heart anomalies and the quality of the heart sound. The detection of signal quality is crucial for heart sounds as

they are often contaminated with noise and artifacts. The main advantage of the developed method over state-of-the-art methods was that it is segmentation-agnostic. That is, most methods for PCG analysis and anomaly detection rely on segmentation to find the fundamental components of the heart sound. However, segmentation increases the computational burden, and errors in segmentation propagate to the classification pipeline and, hence, increase the misclassification rate.

Publication [P4] presents a multiclass classification system for AF detection using handheld ECG devices. The model distinguishes AF from other arrhythmias, healthy, and noisy signals using short-term ECG signals. A rich set of features from time, frequency, time-frequency, and nonlinear domains were extracted and classified with a random forest. Although ECG signals recorded by portable devices are prone to noise and artifacts, the proposed method still achieved top performance (tied for first place out of 67 teams in the Physionet Challenge 2017) for automatic AF monitoring using such devices.

In publication [P5], a novel predictive model was designed for the early detection of sepsis from 6 to 12 hours before clinical recognition. The proposed feature extraction scheme addresses two main challenges in the field, missing data, and varying length data. Clinical data are not collected consistently, and therefore, many data samples are missing. Additionally, the number of observations at each timestamp is different and this requires special treatment, particularly in discriminative classifiers. The experimental results show that the proposed method can predict sepsis with significant accuracy.

In summary, the contributions in publications [P3-P5] answer the aforementioned research questions. Apart from the success achieved in the thesis, the proposed Signal Processing and ML methods still fall short from meeting all the challenges in the biomedical field. We next highlight the main challenges and propose possible future research to improve the current state-of-the-art.

First, in all the studies in this thesis, the largest known open-access datasets were employed for each medical application. This is crucial for empirical evaluations of the developed methods. Additionally, this provides a benchmark for fair comparative analysis against the state-of-the-art methods in the domain of interest. However, as was stated in the thesis, some of the datasets remain still small and available ground truth remains scarce, even if one discards issues related to conflicting annotations done by different experts or those which are not solely based on the signals. For example, the standardized annotations for epileptic seizures rely on the symptoms and patient history besides the EEG signals. This issue has also been mentioned by the Physionet organizers about the heart sound labeling [86]. Therefore, larger scale datasets with labels agreed upon

by multiple experts will certainly open new horizons for more advanced ML and Signal Processing algorithms to reach higher performance.

Second, proper experimental protocols and adequate performance metrics should be designed specifically for each task through a close collaboration between experts in both disciplines. International challenges, such as Physionet, is a good example of such an endeavor.

Third, our experiments over the physiological datasets show that well-chosen features can outperform or perform as well as more sophisticated approaches, such as Deep Learning (DL) based methods. This has been shown especially in publication [P5], where the top three winning teams' classification pipelines did not include any DL based methods. One explanation for the limited performance of DL methods is the insufficient labeled data in the medical field, see earlier comment. One potential extension is configuring time series augmentation methods similar to image augmentation for DL methods.

Forth, ML methods using ensemble classifiers have been among the top winners of data science competitions. Our experiments support the hypothesis that using multiple classifiers and ensemble models outperform a single classifier. However, ensembling techniques should be chosen based on the given task and data.

Finally, some recent developments in nonlinear dynamics ([108] and [109]) may improve the performance of the feature extraction scheme used in the first two publications. Different time lags in the phase space reconstruction rather than using a constant value for successive elements of a delay vector may improve the discrimination power of the extracted features. Moreover, using EEG segments with a duration of more than a few seconds may improve the quality of the phase space reconstruction. Furthermore, using the combination of handcrafted features and DL approaches such as recurrent ANNs may boost the prediction and classification of physiological time series. These will be topics of future research.

References

[1]   J. Jeong, J.H. Chae, S.Y. Kim, S.H. Han, "Nonlinear Dynamic Analysis of the EEG in Patients With Alzheimer's Disease and Vascular Dementia.," *Journal of Clinical Neurophysiology,* vol. 18, no. 1, pp. 58-67, 2001.

[2]   T.H. Mäkikallio, J.M. Tapanainen, M.P. Tulppo, H.V. Huikuri, "Clinical Applicability of Heart Rate Variability Analysis by Methods Based on Nonlinear Dynamics," *Cardiac Electrophysiology Review,* vol. 6, no. 3, pp. 250-255, 2002.

[3]   L.S. Liebovitch, and A.T. Todorov, "Using Fractals and Nonlinear Dynamics to Determine the Physical Properties of Ion Channel Proteins," *Critical Reviews™ in Neurobiology,* vol. 10, no. 2, 1996.

[4]   D. Parthimos, K. Osterloh, A.R. Pries, T.M. Griffith, "Deterministic Nonlinear Characteristics of in Vivo Blood Flow Velocity and Arteriolar Diameter Fluctuations," *Physics in Medicine & Biology,* vol. 49, no. 9, p. 1789, 2004.

[5]   M. Le Van Quyen, M. Chavez, D. Rudrauf, J. Martinerie, "Exploring the Nonlinear Dynamics of the Brain," *Journal of Physiology-Paris,* vol. 97, no. 4-6, pp. 629-639, 2003.

[6]   J. Fell, A. Kaplan, B. Darkhovsky, and J. Röschke, "EEG Analysis With Nonlinear Deterministic and Stochastic Methods: a Combined Strategy," *Acta Neurobiologiae Experimentalis,* vol. 60, no. 1, pp. 87-108, 2000.

[7]   H. Kantz, and T. Schreiber, Nonlinear Time Series Analysis, vol. 7, Cambridge University Press, 2004.

[8]   R.M. May, "Simple Mathematical Models With Very Complicated Dynamics," *Nature,* vol. 261, no. 5560, pp. 459-467, 1976.

[9]   H.S. Karagueuzian, H. Stepanyan, W.J. Mandel, "Bifurcation Theory and Cardiac Arrhythmias," *American Journal of Cardiovascular Disease,* vol. 3, no. 1, pp. 1-16, 2013.

[10] FL. Da Silva, W. Blanes, SN. Kalitzin, J. Parra, P. Suffczynski, D. N. Velis, "Epilepsies as Dynamical Diseases of Brain Systems: Basic Models of the Transition Between Normal and Epileptic Activity," *Epilepsia,* vol. 44, p. 72–83, 2003.

[11] D. Kugiumtzis, Surrogate Data Test on Time Series, vol. 2, Boston, MA: Springer, 2002.

[12] M. Akay, Nonlinear Biomedical Signal Processing, Wiley Online Library, 2000.

[13] D.W. Jordan, and P. Smith, Nonlinear Ordinary Differential Equations: An Introduction to Dynamical Systems, vol. 2, Oxford University Press, 1999.

[14] R. C. Hilborn, Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers, Oxford University Press on Demand, 2000.

[15] F. Takens, "Detecting Strange Attractors in Turbulence," *Dynamical systems and turbulence,* pp. 366-381, 1981.

[16] H. Whitney, J. Eells, and D. Toledo, Collected Papers of Hassler Whitney, Nelson Thornes, 1992.

[17] M. Casdagli, S. Eubank, J.D. Farmer, J. Gibson, "State Space Reconstruction in the Presence of Noise," *Physica D: Nonlinear Phenomena,* vol. 51, no. 1, pp. 52-98, 1991.

[18] D. Aeyels, "Generic Observability of Differentiable Systems," *SIAM Journal on Control and Optimization,* vol. 19, no. 5, pp. 595-603, 1981.

[19] C. Rhodes and M. Morari, "The false nearest neighbors algorithm: An overview," *Computers & Chemical Engineering,* osa/vuosik. 21, pp. 1149-1154, 1997.

[20] H.D.I. Abarbanel, Analysis of Observed Chaotic Data, New York: Springer-Verlag, 1996.

[21] MB. Kennel and H.D.I. Abarbanel, "False Neighbors and False Strands: a Reliable Minimum Embedding Dimension Algorithm," *Physical Review E,* vol. 66, no. 2, 2002.

[22] R. Barrio, F. Blesa, and S. Serrano, "Qualitative Analysis of the Rössler Equations: Bifurcations of Limit Cycles and Chaotic Attractors," *Physica D: Nonlinear Phenomena,* vol. 238, no. 13, pp. 1087-1100, 2009.

[23] EA Coddington and N. Levinson, Theory of ordinary differential equations, Tata McGraw-Hill Education, 1955.

[24] J. Llibre and A.E. Teruel, Introduction to the Qualitative Theory of Differential Systems, Basel: Springer, 2014.

[25] S. Yang, "Nonlinear Signal Classification Using Geometric Statistical Features in State Space," *Electronics Letters,* vol. 40, no. 12, pp. 780-781, 2004.

[26] R. Yaghoobi Karimui, S. Azadi, "Cardiac Arrhythmia Classification Using the Phase Space Sorted by Poincare Sections," *Biocybernetics and Biomedical Engineering,* vol. 37, no. 4, pp. 690-700, 2017.

[27] B. Sharif, and A.H. Jafari, "Prediction of Epileptic Seizures From Eeg Using Analysis of Ictal Rules on Poincaré Plane," *Computer Methods and Programs in Biomedicine,* vol. 145, pp. 11-22, 2017.

[28] M. Amiri, E. Davoodi-Bojd, F. Bahrami, M. Raza, "Bifurcation Analysis of the Poincaré Map Function of Intracranial EEG Signals in Temporal Lobe Epilepsy Patients," *Mathematics and Computers in Simulation,* vol. 81, no. 11, pp. 2471-2491, 2011.

[29] A. Kubičková, J. Kozumplík, Z. Nováková, M. Plachý, P. Jurák, J. Lipoldová, "Heart Rate Variability Analysed by Poincaré Plot in Patients With Metabolic Syndrome," *Journal of Electrocardiology,* vol. 49, no. 1, pp. 23-28, 2016.

[30] E. Simonyi, "The Dynamics of the Polymerization Processes," *Periodica Polytechnica Electrical Engineering,* vol. 11, no. 4, pp. 309-316, 1967.

[31] D.H. Wolpert, "The Lack of a Priori Distinctions Between Learning Algorithms," *Neural computation,* vol. 8, no. 7, pp. 1341-1390, 1996.

[32] K. M. Al-Aidaroos, A. A. Bakar, and Z. Othman, "Medical Data Classification with Naive Bayes Approach," *Information Technology Journal,* vol. 11, no. 9, pp. 1166-1174, 2012.

[33] M. M. Casas, R. L.Avitia, F. F. Gonzalez-Navarro, J. A. Cardenas-Haro, M. A. Reyna, "Bayesian Classification Models for Premature Ventricular Contraction Detection on ECG Traces," *Journal of healthcare engineering,* 2018.

[34] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision Trees: An Overview and Their Use in Medicine," *Journal of medical systems,* vol. 26, no. 5, pp. 445-463, 2002.

[35] G. Quellec, M. Lamard, L. Bekri, G. Cazuguel, C. Roux, B. Cochener, "Medical Case Retrieval From a Committee of Decision Trees," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 5, pp. 1227-1235, 2010.

[36] J. Zhang, F. Lin, P. Xiong, H. Du, H. Zhang, M. Liu, Z. Hou, X. Liu, "Automated Detection and Localization of Myocardial Infarction With Staked Sparse Autoencoder and TreeBagger," *IEEE Access,* vol. 7, pp. 70634-70642, 2019.

[37] M. Heyat, D. Lai, FI. Khan, Y. Zhang, "Sleep Bruxism Detection Using Decision Tree Method by the Combination of C4-P4 and C4-A1 Channels of Scalp EEG," *IEEE Access,* vol. 7, pp. 102542-102553, 2019.

[38] S. Kiranyaz, T. Ince, M. Zabihi, D. Ince, "Automated Patient-specific Classification of Long-term Electroencephalography," *Journal of Biomedical Informatics,* vol. 49, pp. 16-31, 2014.

[39] M. Zabihi, AB. Bahrami, S. Särkkä, S. Kiranyaz, A. K. Katsaggelos, M. Gabbouj, "Automatic Sleep Arousal Detection Using State Distance Analysis in Phase Space," in *Computing in Cardiology Conference (CinC)*, Maastricht, Netherlands, 2018.

[40] S. Kiranyaz, T. Ince, and M. Gabbouj, "Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias," *Scientific Reports volume,* vol. 7, no. 1, p. 9270, 2017.

[41] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, John Wiley & Sons, 2012.

[42] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[43] Y. Guo, T. Hastie, and R. Tibshirani, "Regularized Linear Discriminant Analysis and Its Application in Microarrays," *Biostatistics,* vol. 8, no. 1, pp. 86-100, 2006.

[44] T. M. Mitchell, "The Basic Decision Tree Learning Algorithm," tekijä: *Machine Learning*, McGraw-Hill, 1997, pp. 55-58.

[45] J. R. Quinlan, C4.5: Programs for Machine Learning, Elsevier, 2014.

[46] L. Breiman, J. H. Friedman, R.A. Olshen, C. J. Stone, Classification and Regression Trees, Chapman & Hall, 1984.

[47] J. R. Quinlan, "Induction of decision trees," *Machine Learning,* osa/vuosik. 1, p. 81–106, 1986.

[48] J. Friedman, "Stochastic Gradient Boosting," *Computational Statistics & Data Analysis,* vol. 38, no. 4, pp. 367-378, 2002.

[49] "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences,* osa/vuosik. 55, nro 1, pp. 119-139, 1997.

[50] "XGBoost: a Scalable Tree Boosting System," tekijä: *Proceedings of the 22nd ACM SIGkDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[51] B. C. Csáji, "Approximation With Artificial Neural Networks," *M.SC thesis, Faculty of Sciences, Etvs Lornd University, Hungary,* p. 48, 2001.

[52] P. McCullagh, and J. A. Nelder, Generalized Linear Models, Chapman & Hall/CRC, 1990.

[53] D. J. C. MacKay, "Bayesian Interpolation," *Neural Computation,* vol. 4, no. 3, pp. 415-447, 1992.

[54] MT. Hagan and MB. Menhaj, "Training Feedforward Networks With the Marquardt Algorithm," *IEEE transactions on Neural Networks,* vol. 5, no. 6, pp. 989-993, 1994.

[55] MA. Reyna, C. Josef, R. Jeter, SP. Shashikumar, MB. M. Brandon Westover, S. Nemati, GD. Clifford, A. Sharma, "Early Prediction of Sepsis From Clinical Data: the Physionet/Computing in Cardiology Challenge 2019," *Critical Care Medicine,* 2019.

[56] HE. Scharfman, "The Neurobiology of Epilepsy," *Current neurology and neuroscience reports,* vol. 7, no. 4, pp. 348-354, 2007.

[57] J. Milton, and P. Jung, Epilepsy as a Dynamic Disease, Springer-Verlag Berlin Heidelberg, 2013.

[58] SB. Wilson, ML. Scheuer, C. Plummer, B. Young, S. Pacia, "Seizure Detection: Correlation of Human Experts," *Clinical Neurophysiology,* vol. 114, no. 11, pp. 2156-2164, 2013.

[59] N. J. Stevenson, R. R. Clancy, S. Vanhatalo, I. Rosen, J. M. Rennie, G. B. Boylan, "Interobserver Agreement for Neonatal Seizure Detection Using Multichannel EEG," *Annals of clinical and translational neurology,* vol. 2, no. 11, pp. 1002-1011, 2015.

[60] "Adaptive Seizure Onset Detection Framework Using a Hybrid PCA-CSP Approach," *IEEE Journal of Biomedical and Health Informatics,* vol. 22, no. 1, pp. 154-160, 2017.

[61] T. Guan, X. Zeng, L. Huang, T. Guan, and M. Seok, "Neural Network Based Seizure Detection System Using Raw EEG Data," in *International SoC Design Conference (ISOCC)*, Jeju, South Korea, 2016.

[62] M. Zhou, C. Tian, R. Cao, B. Wang, Y.Niu, Ting Hu, H. Guo, and J. Xiang, "Epileptic Seizure Detection Based on EEG Signals and CNN," *Frontiers in Neuroinformatics,* vol. 12, p. 95, 2018.

[63] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A Multi-View Deep Learning Framework for EEG Seizure Detection," *IEEE Journal of Biomedical and Health Informatics,* vol. 23, no. 1, pp. 83-94, 2019.

[64] A. Bhattacharyya, and R. B. Pachori, "A Multivariate Approach for Patient-Specific EEG Seizure Detection Using Empirical Wavelet Transform," *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 9, pp. 2003-2015, 2017.

[65] M. Kaleem, A. Guergachi, and S. Krishnan, "Patient-specific Seizure Detection in Long-term EEG Using Wavelet Decomposition," *Biomedical Signal Processing and Control,* vol. 46, pp. 157-165, 2018.

[66] E. Bradley, and H. Kantz, "Nonlinear Time-series Analysis Revisited," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 25, no. 9, p. 097610, 2015.

[67] A. Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," *Ph.D. Thesis, Massachusetts Institute of Technology,* 2009.

[68] A. L. Goldberger, L. 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, p. e215–e220, 2000.

[69] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, F. E. Abd El-Samie , "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing volume,* osa/vuosik. 183, 2014.

[70] B. Hunyadi, M. Signoretto, W. V. Paesschen, J. A. Suykens, SV. Huffel, M. D. Vos, "Incorporating structural information from the multichannel EEG improves patient-specific seizure detection," *Clinical Neurophysiology,* osa/vuosik. 123, nro 12, pp. 2352-2361, 2012.

[71] V. Sridevi, M. R. Reddy, K. Srinivasan, K. Radhakrishnan, C. Rathore, D. S. Nayak, "Improved Patient-Independent System for Detection of Electrical Onset of Seizures," *Journal of Clinical Neurophysiology,* osa/vuosik. 36, nro 1, pp. 14-24, 2019.

[72] M. Kaleem, A. Guergachi, S. Krishnan, "Patient-specific seizure detection in long-term EEG using wavelet decomposition," *Biomedical Signal Processing and Control,* osa/vuosik. 46, pp. 157-165, 2018.

[73] K. Matlock, C. De Niz, R. Rahman, S. Ghosh, and R. Pal, "Investigation of Model Stacking for Drug Sensitivity Prediction," *BMC Bioinformatics,* vol. 19, no. 3, 2018.

[74] M. Gjoreski, M. Simjanoska, A. Gradisek, A. Peterlin, M. Gams, G. Poglajen, "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers," in *International Conference on Intelligent Environments*, Seoul, 2017.

[75] W. Li and A. Drygajlo, "Global and Local Feature Based Multi-classifier a-stack Model for Aging Face Identification," in *IEEE International Conference on Image Processing (ICIP)*, Hong Kong, 2010.

[76] A. A. Luisada, The Sounds of the Diseased Heart, St. Louis, Missouri: Waren H. Green, Inc., 2013.

[77] S. Patidar, R. B. Pachori, N. Garg , "Automatic diagnosis of septal defects based on tunable-Q wavelet transform of cardiac sound signals," *Expert Syst. Appl. ,* osa/vuosik. 42, p. 3315–3326, 2015.

[78] A. Gharehbaghi, I. Ekman, P. Ask, E. Nylander, B. Janerot-Sjoberg, "Assessment of aortic valve stenosis severity using intelligent phonocardiography," *International Journal of Cardiology,* osa/vuosik. 198, pp. 58-60, 2015.

[79] R. Saracoglu, "Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction," *Engineering Applications of Artificial Intelligence,* osa/vuosik. 25, nro 7, pp. 1523-1528, 2012.

[80] A. Moukadem, A. Dieterlena, N. Hueberb, C. Brandtc, "A robust heart sounds segmentation module based on S-transform," *Biomed. Signal Process. Control,* osa/vuosik. 8, pp. 273-281, 2013.

[81] J. Pedrosa, A Castro, T. T. Vinhoza, "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms," tekijä: *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Chicago, 2014.

[82] H. Tang, T. Li, T. S. Qiu, Y. Park, "Segmentation of heart sounds based on dynamic clustering," *Biomed. Signal Process. Control,* osa/vuosik. 7, pp. 509-516, 2012.

[83] H. Uguz, "Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy," *Neural Comput. Appl.,* osa/vuosik. 21, pp. 1617-1628, 2012.

[84] S. Ari, K. Hembram, G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classier," *Expert Syst. Appl.,* osa/vuosik. 37, pp. 8019-8026, 2010.

[85] Y. N. Zheng, X. M. Guo, X. R. Ding, " A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Syst. Appl.,* osa/vuosik. 42, pp. 2710-2721, 2015.

[86] GD. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, RG. Mark, "Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016," in *Computing in Cardiology Conference*, Vancouver, BC, Canada, 2016.

[87] I. Guyon, and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research ,* osa/vuosik. 3, pp. 1157-1182, 2003.

[88] C. Potes, S. Parvaneh, A. Rahman, B. Conroy, , "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," tekijä: *2016 Computing in Cardiology Conference (CinC)*, Vancouver, BC, Canada, 2016.

[89] M. N. Homsi, N. Medina, M. A. H. Rodriguez, N. A. Quintero, G. Perpiñan, A. Quintana, P. A. Warrick, "Automatic Heart Sound Recording Classification using a Nested Set of Ensemble Algorithms," tekijä: *2016 Computing in Cardiology Conference (CinC)*, Vancouver, BC, Canada, 2016.

[90] Developed with the special contribution of the European Heart Rhythm Association (EHRA), et al., "Guidelines for the Management of Atrial Fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)," *European heart journal,* vol. 31, no. 19, pp. 2369-2429, 2010.

[91] P.A. Wolf, R.D. Abbott, and W.B. Kannel, "Atrial Fibrillation as an Independent Risk Factor for Stroke: the Framingham Study," *Stroke,* vol. 22, no. 8, pp. 983-988, 1991.

[92] W. Bishop, and P. Edman, Atrial Fibrillation Explained: Understanding the Next Cardiac Epidemic, Dr. Warrick Bishop, 2019.

[93] Z. C. Haberman, R. T. Jahn, R. Bose, H. Tun, J. S. Shinbane, R. N. Doshi, P. M. Chang, L. A. Saxon, "Wireless smartphone ECG enables large-scale screening in diverse populations," *Journal of Cardiovascular Electrophysiology,* osa/vuosik. 26, nro 5, pp. 520-526, 2015.

[94] J. Lau, N. Lowres, L. Neubeck, D. B. Brieger, R. W. Sy, C. D. Galloway, D. E. Albert, S. B. Freedman, "iPhone ECG application for community screening to detect silent atrial fibrillation: A novel technology to prevent stroke," *International Journal of Cardiology,* osa/vuosik. 165, nro 1, pp. 193-194, 2013.

[95] K. M. Griffiths, E. N. Clark, B. Devine, and P. W. Macfarlane, "Assessing the accuracy of limited lead recordings for the detection of Atrial Fibrillation," tekijä: *Computing in Cardiology Conference (CinC)*, Cambridge, MA, USA, 2014.

[96] G. Clifford, C. Liu, B. Moody, LH. Lehman, I. Silva, Q. Li, A. Johnson, RG. Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017," in *Computing in Cardiology*, Rennes, 2017.

[97] T. Teijeiro, C. A. García, D. Castro, P. Félix, "Arrhythmia classification from the abductive interpretation of short single-lead ECG records," tekijä: *2017 Computing in Cardiology (CinC)*, Rennes, France, 2017.

[98] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. D. Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, S. Khandelwal, "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," tekijä: *2017 Computing in Cardiology (CinC)*, Rennes, France, 2017 .

[99] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, J. Xie, "ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks," tekijä: *2017 Computing in Cardiology (CinC)*, Rennes, France, 2017.

[100] J. L. Vincent, "The Clinical Challenge of Sepsis Identification and Monitoring," *PLoS Medicine,* vol. 13, no. 5, 2016.

[101] J. Hajj, N. Blaine, J. Salavaci, D. Jacoby, "The Centrality of Sepsis: a Review on Incidence, Mortality, and Cost of Care," *Healthcare,* vol. 6, no. 3, p. 90, 2018.

[102] C. Fleischmann, A. Scherag, N.K. Adhikari, C.S. Hartog, T. Tsaganos, P. Schlattmann, D.C. Angus, K. Reinhart, "Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations," *American journal of respiratory and critical care medicine,* vol. 193, no. 3, pp. 259-272, 2016.

[103] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.D. Chiche, C.M. Coopersmith, R.S.Hotchkiss, "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *Jama,* vol. 315, no. 8, pp. 801-810, 2016.

[104] Z. Zhang, N.J. Smischney, H. Zhang, S. Van Poucke, P. Tsirigotis, J. Rello, P.M. Honore, W.S. Kuan, J.J. Ray, J. Zhou, Y. Shang, "Ame Evidence Series 001— the Society for Translational Medicine: Clinical Practice Guidelines for Diagnosis and Early Identification of Sepsis in the Hospital," *Journal of thoracic disease,* vol. 8, no. 9, p. 2654, 2016.

[105] S. Nemati, A. Holder, F. Razmi, M.D. Stanley, G.D. Clifford, T.G. Buchman, "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU," *Critical care medicine,* vol. 46, no. 4, pp. 547-553, 2018.

[106] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M.D. Feldman, C. Barton, D.J. Wales, "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: a Machine Learning Approach," *JMIR medical informatics,* vol. 4, no. 3, p. e28, 2016.

[107] S. Le, J. Hoffman, C. Barton, J.C. Fitzgerald, A. Allen, E. Pellegrini, J. Calvert, R. Das, "Pediatric Severe Sepsis Prediction Using Machine Learning," *Frontiers in Pediatrics,* vol. 7, p. 413, 2019.

[108] L. Pecora, L. Moniz, J. Nichols, and T. Carroll, "A Unified Approach to Attractor Reconstruction," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 17, no. 1, p. 013110, 2007.

[109] D. Holstein and H. Kantz, "Optimal Markov Approximations and Generalized Embeddings," *Physical Review E,* vol. 79, no. 5, p. 056202, 2009.

[110] "physionet," MIT Laboratory for Computational Physiology, [Online]. Available: https://physionet.org/challenge/. [Accessed 6 11 2019].

PUBLICATION 1

Analysis of High-dimensional Phase Space via Poincaré Section for Patient-specific Seizure Detection

M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M. Gabbouj, and T. Ince

Analysis of high-dimensional phase space via Poincaré section for patient-specific seizure detection

Morteza Zabihi[*], *Student Member, IEEE*, Serkan Kiranyaz, *Senior Member, IEEE*,
Ali Bahrami Rad, *Student Member, IEEE*, Aggelos K. Katsaggelos, *Fellow*, *IEEE*,
Moncef Gabbouj, *Fellow*, *IEEE*, and Turker Ince, *Member, IEEE*

Abstract— **In this paper, the performance of the phase space representation in interpreting the underlying dynamics of epileptic seizures is investigated and a novel patient-specific seizure detection approach is proposed based on the dynamics of EEG signals. To accomplish this, the trajectories of seizure and non-seizure segments are reconstructed in a high dimensional space using time-delay embedding method. Afterwards, Principal Component Analysis (PCA) was used in order to reduce the dimension of the reconstructed phase spaces. The geometry of the trajectories in the lower dimensions is then characterized using Poincaré section and seven features were extracted from the obtained intersection sequence. Once the features are formed, they are fed into a two-layer classification scheme, comprising the Linear Discriminant Analysis (LDA) and naïve Bayesian classifiers. The performance of the proposed method is then evaluated over the CHB-MIT benchmark database and the proposed approach achieved an 88.27% sensitivity and 93.21% specificity on average with 25% training data. Finally, we perform comparative performance evaluations against the state-of-the-art methods in this domain which demonstrate the superiority of the proposed method.**

Index Terms— **Dynamics, EEG, phase Space, Poincaré section, seizure detection, two-layer classifier topology.**

I. INTRODUCTION

EPILEPTIC seizures are transient excessive neuronal discharges originated from cortical gray matter and considered as the main definition of epilepsy. Indeed the concept of epilepsy covers a wide range of disorders, which can be classified according to the variety in types of seizures.

M. Zabihi[*], and M. Gabbouj are with the Department of Signal Processing, Tampere University of Technology, Tampere, Finland (e-mail: morteza.zabihi@tut.fi, moncef.gabbouj@tut.fi).

S. Kiranyaz is with the Electrical Engineering Department, Qatar University, Doha, Qatar (e-mail: mkiranyaz@qu.edu.qa)

A. B. Rad is with the Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway (e-mail: ali.bahrami.rad@uis.no).

A. K. Katsaggelos is with the Electrical Engineering Department, Northwestern University, Evanston, IL, USA (e-mail: aggk@eecs.northwestern.edu).

T. Ince is with the Department of Electrical and Electronics Engineering, Izmir University of Economics, Izmir, Turkey (e-mail: turker.ince@ieu.edu.tr).

Therefore, the epileptologist in the first place, should diagnose if an epileptic seizure occurred and then determine the seizure type. For this purpose, electroencephalography (EEG) is commonly used due to its unique properties, such as cost-effectiveness and high temporal resolution, which make it an influential and compulsory tool for exploring the brain functioning of patients with epilepsy.

The estimated number of people suffering from epilepsy in the world is around 50 million [1]. In addition, an increasing need for recording EEG signals in the long term, and the contamination of these signals with physiological and non-physiological artefacts renders their interpretation through visual inspection only a daunting and challenging task. These factors add impetus to the need of an automatic seizure detection system to ease the neurologist's burden of inspecting such long-term EEG data [2]. Thus, in the recent years, several techniques have been developed in order to detect patterns of interest from background patterns, including time [3], frequency [4], time-frequency [5], and nonlinear methods [6] - [8].

Despite conventional time series analysis, nonlinear dynamics addresses nonlinear relationships among the variables of a system by investigating only the variables (i.e., states) in phase space whilst discarding time or spectral components. The main power of this approach is that it provides information regarding the underlying dynamics of the system without knowing all the factors in the system evolution. Hence, nonlinear time series analysis, unlike differential equations, is a top-down approach where information about the states of the system or the relationship among the states is not available. Therefore, the approach is to reconstruct the system dynamics in phase space and then quantify the reconstructed attractors (e.g., [9] and [10]).

Numerous measures originated from nonlinear dynamics have been introduced and used for the analysis of EEG signals. Correlation dimension [11], Lyapunov exponents [12], phase synchronization [13], and mutual dimension [14] can be named among the traditional and novel measures. Nevertheless, most of these measures do not have straightforward interpretations and can only be used as tentative indices. This may instigate a false impression of chaos, hence surrogate data tests are needed in order to check the validity of the analysis. In other words, surrogate data

analysis is required in this case to reject the least null hypotheses of linear stochastic and white noise [15] [16]. On the contrary, phase space reconstruction can be considered as a tool to demonstrate the evolution of a dynamic system through time, while the dynamic system can fall into several categories such as conservative or dissipative, linear or nonlinear, deterministic or stochastic. Therefore, using geometrical features, based only on the phase space and without asserting any assumption about the type of the underlying dynamic, made the studies independent from chaotic hypothesis and consequently independent from the surrogate data analysis [6].

However, despite the emphasized properties of the phase space, characterizing the reconstructed trajectories of EEG signals, purely based on its geometry, is quite rare and only few works exist up to date. In [17], the phase space of an EEG signal was reconstructed after decomposing the signal using empirical mode decomposition with application to seizure detection. Two measures, namely the 95% confidence ellipse area and the interquartile range of the Euclidian distance, were extracted from two- and three-dimensional phase spaces, respectively. M. Chang et al. compared the efficiency of features obtained by amplitude-frequency analysis [18] and autoregressive (AR) model [10] using phase space and raw data in a Brain Computer Interface (BCI) task. In their study, the phase space of two EEG channels was reconstructed and then the AR parameters, peak and mean values of the absolute value of amplitude samples in two frequency bands 8-13 Hz and 14-25 Hz were calculated as features. It was shown that these features improve the classification results in contrast to the same features extracted from the raw data.

In [19], the wavelet (Daubechies 4) coefficients of an EEG signal at 5 levels were used to plot a two dimensional phase space. Then, the Euclidean distances between the origin and every point in the phase space were calculated. The mean, median, average power and standard deviation of these distances in each sub-band were used as features in order to detect seizure events. The classification results on real EEG data showed the significance of the extracted features. Furthermore, the phase space of EEG signals has been studied in a behavioral neuroscience research. In [20], the slope of a regression line in a two-dimensional phase space was obtained as a function of different time lags and considered as a feature in classification of sleep-wake states. The results showed the features based on phase space achieved higher performance than the power spectral approach.

Accurate reconstruction of the phase space has a great impact on the characterization of its trajectory properties. In the time-delay embedding method this accuracy depends on the proper selection of the time lag and the embedding dimension. According to the embedding theorem, any time lag will be acceptable; however, it should be noted that choosing a too small or a too large time lag value prompts completely dependent and independent coordinates [6]. Thus, the necessary and major task in keeping the physical properties of attractors is to determine a large enough embedding dimension. In the aforementioned studies, the maximum

embedding dimension of three (in [17]), or two (in [18], [10], [19], and [20]) were used, this is generally insufficient and hence inadequate for mathematical modeling of such complex signals. Many published works in EEG signal processing (e.g., [21] - [24]) propose high dimensional phase spaces, i.e., d>3. However, using a higher dimensional phase space makes the interpretation and visualization of the trajectories a challenging task.

In this study, in order to address the aforementioned deficiencies such as the false impression of chaos and the necessity of applying surrogate data analysis, a novel phase space method is proposed. The proposed method aims to capture the underlying dynamics of the epileptic seizures and hence to discriminate them from the non-seizure segments in an efficient way. Therefore, our primary goal is to create a new set of nonlinear features for seizure detection which decreases the computational complexity while at the same time increases the seizure detection accuracy. This method describes the characterization of the geometry of a high dimensional phase space in such a way that it keeps the reconstructed trajectories unfolded. More specifically, this paper describes a novel feature extraction method based on a high dimensional phase space along with a classification scheme, where the main objective is to maximize the seizure detection accuracy with a minimal feedback from a human expert. Besides the discrimination ability of the extracted features, there are two main factors potentially affecting the performance of the automatic seizure detection methods: variation in seizure types and the brain regions where seizures have originated. Epileptic seizures develop as the results of different disorders, and as such, they cause significant variations of seizure types among patients (inter-patient variability). Therefore, in this work patient-specific setting is employed as a convenient candidate for such classification problems since this approach has more potential to learn the patterns of seizures in each individual specifically. In order to address the second factor, the signals of all channels are utilized. Besides providing the information of seizures in different brain areas, this approach gives the flexibility on the proposed framework to be compatible with various EEG recording montages.

To accomplish these objectives, first we reconstruct the trajectories of each 1-second EEG segment with fixed values of time lag and embedding dimension (Section II.B). Then, a Poincaré map of the reconstructed trajectories is obtained using Poincaré section, which is mainly chosen according to the first and second Principal Components (PCs) of the phase space coordinates (Section II.C). Afterwards, 7 discriminative features (Section II.D) are extracted from the obtained Poincaré map and fed into a classifier topology with two-layer architecture. The first layer consists of 23 Linear Discriminant Analysis (LDA) classifiers and a Naïve Bayes classifier formed in the second layer (Section II.E) in order to fuse the decisions of the first layer classifiers and hence perform the final classification. Finally, the proposed approach is tested on a benchmark dataset with EEG recordings of pediatric patients with intractable seizure and compared with the state-of-the-art

methods in this domain. The paper is concluded in Section IV.

II. EEG DATA PROCESSING

A. EEG Dataset

The EEG recordings were collected from 23 pediatric patients (males, ages 3-22; and females, ages 1.5-19) at the Children's Hospital, Boston, to assess their candidacy for surgical operation [25] [26]. Nine to twenty four EEG recordings were recorded for each individual. All the recordings were labeled as seizure or non-seizure with one second resolution. The sampling frequency was 256 Hz with 16-bit resolution. There are 23 common channels (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8) for each recording, which are based on the international 10-20 system of EEG electrode positions (see Fig. 1). In this study, only those records which contain at least one seizure event were used. In TABLE I, the lengths of each EEG recording (used in this study) are shown. Recordings 21 and 1 were obtained from the same female patient with 1.5 year apart, which were considered as two extra patients in this study. In addition, patient 15 was excluded from our analysis because we failed to read the EEG data of this patient.
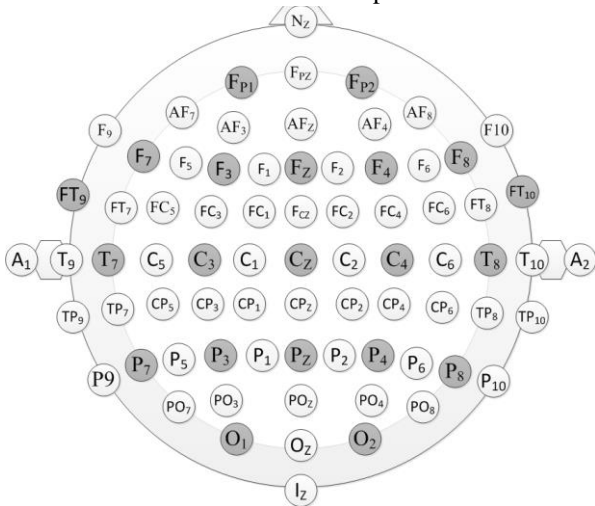

Fig. 1. Location of electrodes in international 10-20 system; the used channels are shown with gray color.

B. Phase Space Reconstruction

The phase space of a system represents how the states' dynamics evolve over time. The aim of phase space reconstruction is to obtain state vectors using the sequence of observed measurements. In this study, time-delay embedding method [27] is used for reconstruction of the EEG phase space. Takens' theorem expresses that the topological features of any higher dimensional system with coupled variables are reconstructable from a single time series of observations [28]. This theorem proves the independence of our study from the surrogate data analysis, which is mentioned in Section I.

The main idea is to create a series of time-shifted samples in d dimensions so that d coordinates would be provided using the map:

$$X[n] \rightarrow Y[n] = (X[n], X[n + T], \dots, X[n + (d - 1)T]) \quad (1)$$

where T is the time lag. In order to determine the convenient dimension and time lag the two commonly known methods of correlation dimension and the mutual information are employed [6]. The embedding dimension 5 and time lag 6 (about 23 milliseconds at a sampling frequency of 256 Hz) were achieved and used for constructing an EEG attractor. The achieved values were validated empirically, where different time lags and embedding dimensions were used and their classification accuracy compared. In Fig. 2, the reconstructed phase spaces of sample seizure and non-seizure segments from the 1st, 17th and 21st patients are shown.
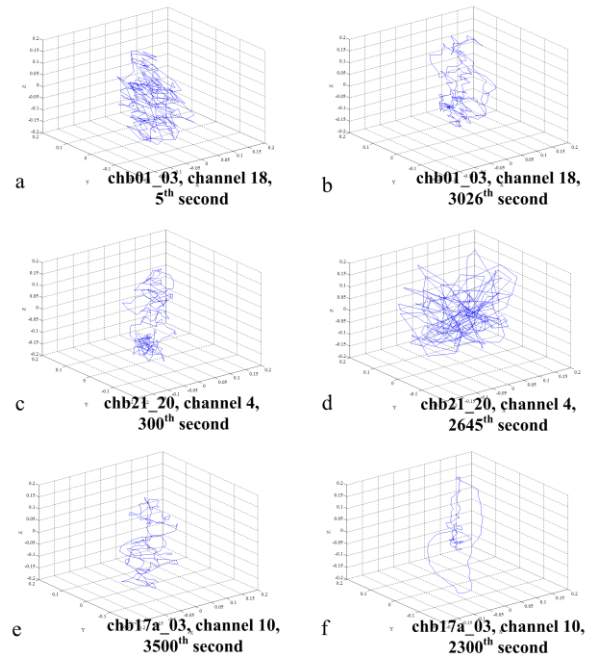
Fig. 2. The phase space plots of 1-s non-seizure segments in the left column (a, c, and e), and seizure segments in the right column (b, d, and f) reconstructed from the 1st, 17th and 21st patients (each row) in the CHB-MIT benchmark database. 3-D phase spaces are plotted for visualization.

C. Poincaré Section Delineation

Poincaré section, which was named in Henry Poincaré honor, is a well-known method for analyzing the type of attractors. In this method, a line (or plane) cut the attractor and then the intersection points are investigated. In fact, Poincaré section provides a geometric view of a trajectory's behavior through those intersection points. One application of Poincaré section, for instance, is to study the stability of limit cycles. This method also can be considered as a sampling method, which converts the continuous trajectories of a phase space to a discrete sequence of intersection points. The investigation of

TABLE I. CHB-MIT benchmark. The patients with longest and shortest duration of recordings are shown in bold.

Patient	Gender	Age	Number of seizure events (Tmax-Tmin in seconds)	Total duration of seizures (sec)	Total duration of non-seizures (sec)	Total duration (sec)
1	F	11	7 (28-102)	449	23476	23925
2	**M**	**11**	**3 (10-83)**	**175**	**7984**	**8159**
3	F	14	7 (48-70)	409	24791	25200
4	M	22	4 (50-117)	382	37977	38359
5	F	7	5 (97-121)	563	17437	18000
6	**F**	**1.5**	**10 (13-21)**	**163**	**93053**	**93216**
7	F	14.5	3 (87-144)	328	32209	32537
8	M	3.5	5 (135-265)	924	17076	18000
9	F	10	4 (63-80)	280	34219	34499
10	M	3	7 (36-90)	454	50010	50464
11	F	12	3 (23-753)	809	9250	10059
12	F	2	27 (14-98)	1016	33844	34860
13	F	3	10 (18-71)	450	24750	25200
14	F	9	8 (15-42)	177	25023	25200
16	F	7	8 (7-15)	77	17923	18000
17	F	12	3 (89-116)	296	10528	10824
18	F	18	6 (31-69)	323	19951	20274
19	F	19	3 (78-82)	239	10307	10546
20	F	6	8 (30-50)	302	19734	20036
21	F	13	4 (13-82)	203	13587	13790
22	F	9	3 (59-75)	207	10593	10800
23	F	6	7 (21-114)	431	31823	32254
24	Unknown	Unknown	16 (17-71)	539	42661	43200

how the trajectories pass through the Poincaré section reveals information about the dynamics of a system that is not obtainable otherwise [29].

The most common approach in this context is applying the Poincaré section on two- or three-dimensional attractors. The reason is that Poincaré section aims to generate a plot stroboscopically so that it samples the motion of the observed trajectories. In order to construct such a view, the Poincaré section must have a dimension less than the corresponding trajectories. Therefore, for instance, if the embedding dimension of a phase space is three, then a plane is used as a Poincaré section. In our case, where the number of embedding dimension is five, the Poincaré section will be a surface with four dimensions. In order to avoid using such complex geometric shapes and at the same time keep the trajectories unfolded without ambiguities, the following solution is proposed: first, Principal Component Analysis (PCA) is employed in order to convert the 5-dimensional embedding coordinates into a set of values of linearly uncorrelated principle components (PCs). Then, a polynomial with the degree of one (i.e., Poincaré section) is fitted to the space formed by the 1st and 2nd PCs. The intersection points of the fitted line and trajectories in the 2-dimensional space are then obtained. In Fig. 3, the whole process of Poincaré mapping is shown. The reason for deploying PCA is to project the entire phase space onto a different space (the space originated from 1st and 2nd PCs) where the reconstructed trajectories are more spread. In this way, the chosen space contains more information about the dynamics of the states. It is worth mentioning that the first 2 PCs are just linear transformations of the original variables and do not necessarily contain more information than the other PCs. Therefore, we empirically checked all the possible combinations of PCs i.e., 1st and 3rd

PCs, 2nd and 3rd PCs, etc., and the best results was achieved by the first and seconds PCs for feature extraction. In order to find the intersection points, the trajectories and the Poincaré section lines were considered as polylines and then the Bézier clipping method [30] was applied.

D. Feature Extraction

In the next step, seven features in total were extracted from the first PC of the intersection sequence obtained in Section II.C. These features are as follows:

Range:

$$Range = \max(X) - \min(X), \tag{2}$$

Quantile and interquantile range:

The 0.13 quantile and the interquartile range (the difference between the first -0.25- and the third -0.75- quartile).

Shannon entropy:

$$H_s(X) = -\sum P(X)\log_2\big(P(X)\big), \tag{3}$$

Root Mean Squared Amplitude (RMS Amp):

$$RMS_{Amp}(X) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} X^2(k)}, \tag{4}$$

Coefficient of Variation:

$$COV(X) = \frac{\sqrt{\frac{\sum(X-\overline{X})}{N}}}{\overline{X}}, \tag{5}$$

and energy:

$$En(X) = \sum_{k=1}^{N} |X(k)|^2 \tag{6}$$

where X is the sequence of the intersection points, and $\overline{X}$ is its mean value. N is the number of intersection points and P(X) is the probability distribution function.

E. Classification and Post Processing

Once the feature vectors of each patient were formed, they are fed into a two-layer classifier network. In the first layer, a
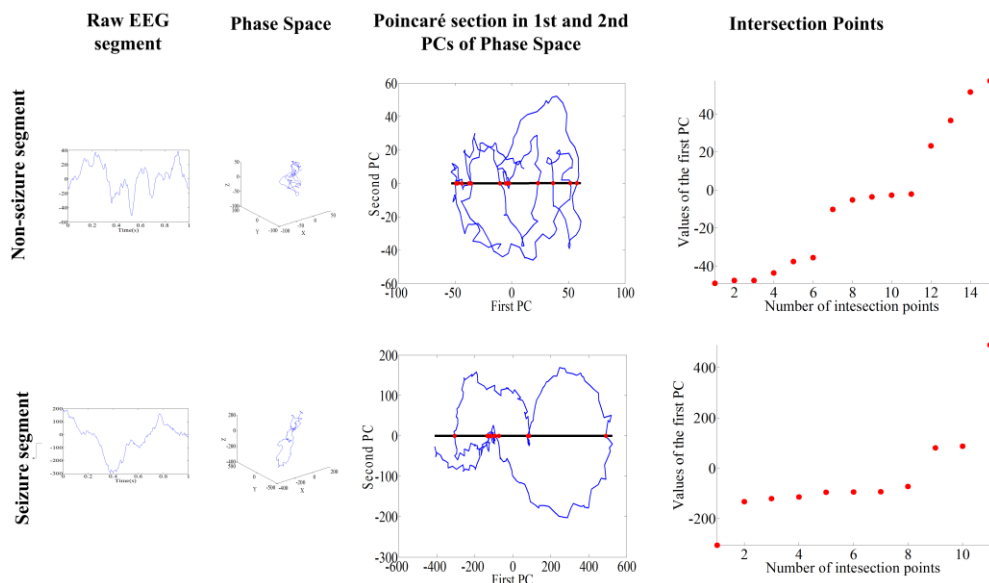
Fig. 3. Poincaré mapping procedure. The phase space (column 2) is obtained from the raw signal (column 1) in a non-seizure (top row) and a seizure segment (bottom row). In order to draw the Poincaré section, a polynomial curve with degree of one is fitted to the 1st and 2nd PCs of the phase space coordinates (column 3). Once the intersection points were determined, their values on the first PC were used for feature extraction. For visualization purpose, only the first three coordinates of phase space are shown.

single LDA classifier trained over the features of each 1-second EEG segment of each EEG channel. Once the first layer classifiers are trained, then their outputs (class vectors) are fed into a naïve Bayes (NB) classifier as a feature matrix, which makes the final classification for the input 1-second EEG segment. In the naïve Bayes classifier the "*multivariate multinomial distribution*" is used to model the outputs of the LDA classifiers because the input features of the second layer are discrete (i.e., binary). The schematic diagram of the proposed classification framework is shown in Fig. 4.

The main scenario is that once the system is trained for a specific patient it can then be used over and over for the same patient. In this case, after the system is trained using the labeled data, the system can be used and help the neurologist for the same patient over and over. Therefore, we divided the benchmark database is divided into a training and test datasets, both of which contain seizure and non-seizure frames, where

the training set contains seizure and non-seizure segments which occurred earlier in time and the remaining segments constitute the test set. The classifier network was trained over the EEG recording of each patient's training set that is formed using two different training sizes: 25% and 50% of the available data.

In the post processing step, the fuzzy rule-based morphological filter proposed in our previous work [31] was applied to the outputs of each classifier in both layers (i.e., 23 LDA and 1 NB). The principal aim of the morphological filter is to filter out the classification outliers based on some global properties such as continuity and neighborhood similarity.

III. EXPERIMENTAL RESULTS

In this section, first the overall results of the proposed patient-specific approach are presented. For comparative
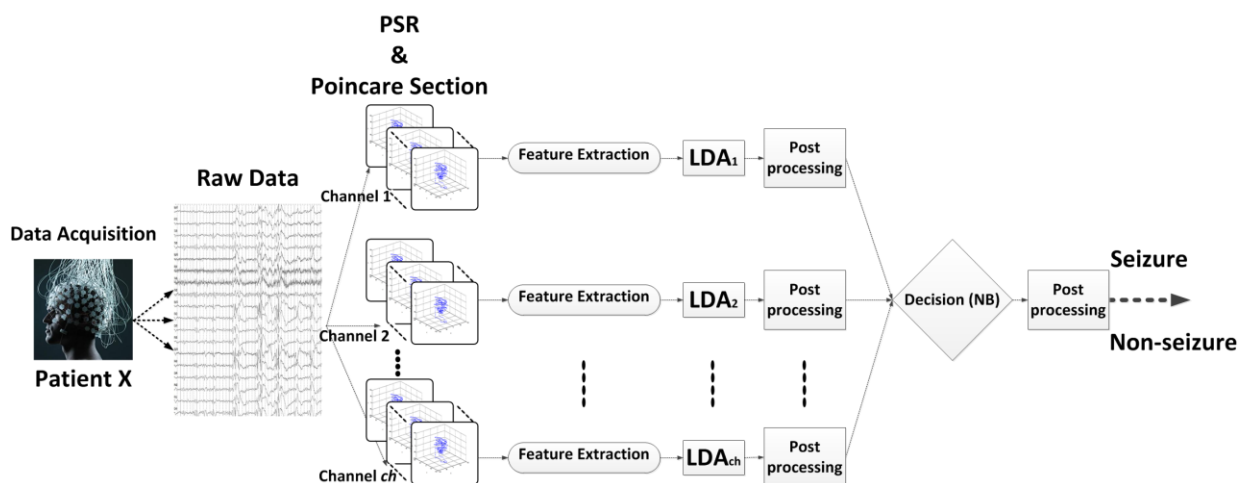

Fig. 4. The proposed classification framework (PSR is the phase-space reconstruction, LDA and NB are the linear discriminant analysis and naïve Bayes classifier, respectively).

evaluations, in section III.B, the seizure detection systems using the CHB-MIT benchmark are briefly presented. In section III.C, the proposed feature extraction approach is compared against the three state-of-the-art methods. In section III.D, four different classifiers within the proposed classification topology are evaluated against the proposed classifier. Finally, the computational complexity analysis is presented in section 0.

A. Classification Performance Evaluations

In this work, the standard performance measures of sensitivity (*Sen*), specificity (*Spe*) and accuracy (*Acc*) are used. They are defined as follows,

$$Sen = \frac{TP}{TP+FN}, \quad (7)$$
$$Spe = \frac{TN}{TN+FP}, \quad (8)$$
$$Acc = \frac{TP+TN}{TP+TN+FP+FN}, \quad (9)$$

where TP (True Positive) is the number of segments correctly detected as seizure, FN (False Negative) is the number of segments incorrectly detected as non-seizure, TN (True Negative) is the number of segments correctly detected as non-seizure, and FP (False Positive) is the number of segments incorrectly detected as seizure. The confusion matrix and the overall classification performance measures are shown in TABLE II and TABLE III. In addition, the Region of Convergence (ROC) plots are presented in Fig. 5 and Fig. 6 for better visualization of the performance of the proposed method.

As TABLE III shows, the best classification performance is achieved using a training rate of 50% with an average sensitivity and specificity of 89.10% and 94.80%, respectively. However, even with 25% training rate only an

insignificant performance loss is encountered, i.e., 88.27% and 93.21% are the average sensitivity and specificity rate, respectively. This demonstrates a delicate generalization capability of the proposed approach and effectiveness of the proposed feature extraction on the discrimination of the seizures segments.

However, the results given in TABLE III indicate that a relatively low classification accuracy is obtained on a few patients i.e., 6, 12 and 24. The reason is that in the recordings of these patients, there are many similarities between seizure and non-seizure segments as well as the high variations within each type. This is visible in Fig. 7 where few segments of non-seizure and seizure recordings from patient 6 are displayed. As can be seen in the figure, there is a high variability between patterns of non-seizure segments c, *e* and *g*. In addition, this difference is evident between seizure segments *d* and *h*. Furthermore, both segments in *a* and *b* have high frequency and low amplitude signals while the former is a non-seizure and the latter is a seizure segment.

Furthermore we calculated the Average Detection Sensitivity Rate (ADSR), Average False Alarm (AFA) per hour and Average Alarm Delay (AAD) as expressed below in order to evaluate the "seizure event detection" performance of the proposed method. In order to calculate these metrics, we defined a seizure event if at least 7 consecutive seizure segments (with resolution of 1 second) are detected.

$$ADSR = \frac{\sum_{S=1}^{NS} number\ of\ detected\ seizure\ events\ in\ subject\ "S"}{Total\ number\ of\ seizure\ events\ in\ all\ subjects}, \quad (10)$$

TABLE II. True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP) achieved for test set using 50% and 25% training rate.

patients	50% Training rate				25% Training rate			
	Seizure detected as seizure (TP)	Seizure detected as non-seizure (FN)	Non-seizure detected as non-seizure (TN)	Non-seizure detected as seizure (FP)	Seizure detected as seizure (TP)	Seizure detected as non-seizure (FN)	Non-seizure detected as non-seizure (TN)	Non-seizure detected as seizure (FP)
1	206	18	11693	45	327	9	17371	236
2	80	7	3955	37	131	0	5645	343
3	203	1	12087	308	306	0	17084	1509
4	169	22	18409	579	276	10	23663	4819
5	219	62	8677	41	350	72	13011	66
6	58	23	41672	4854	99	23	47232	22557
7	124	40	16069	35	211	35	23988	168
8	373	89	8522	16	574	119	12404	403
9	133	7	17084	25	206	4	24844	820
10	212	15	24750	255	293	47	37129	378
11	383	21	4521	104	597	9	6556	381
12	407	101	11344	5578	560	202	20491	4892
13	219	6	11751	624	325	12	17455	1107
14	83	5	11814	697	129	3	17258	1509
16	33	5	6681	2280	42	15	11859	1583
17	148	0	4862	402	191	31	7618	278
18	161	0	8460	1515	242	0	11953	3010
19	101	18	5048	105	135	44	7647	83
20	142	9	9790	77	213	13	14525	275
21	101	0	6727	66	151	1	10040	150
22	102	1	5278	18	151	4	7836	108
23	183	32	15757	154	193	130	23716	151
24	179	90	21142	188	272	132	31195	800

TABLE III. The classification results using the proposed method. Patients with measures less than 70% are highlighted. Sen, Spe, and Acc are the sensitivity, specificity and accuracy obtained over the test data.

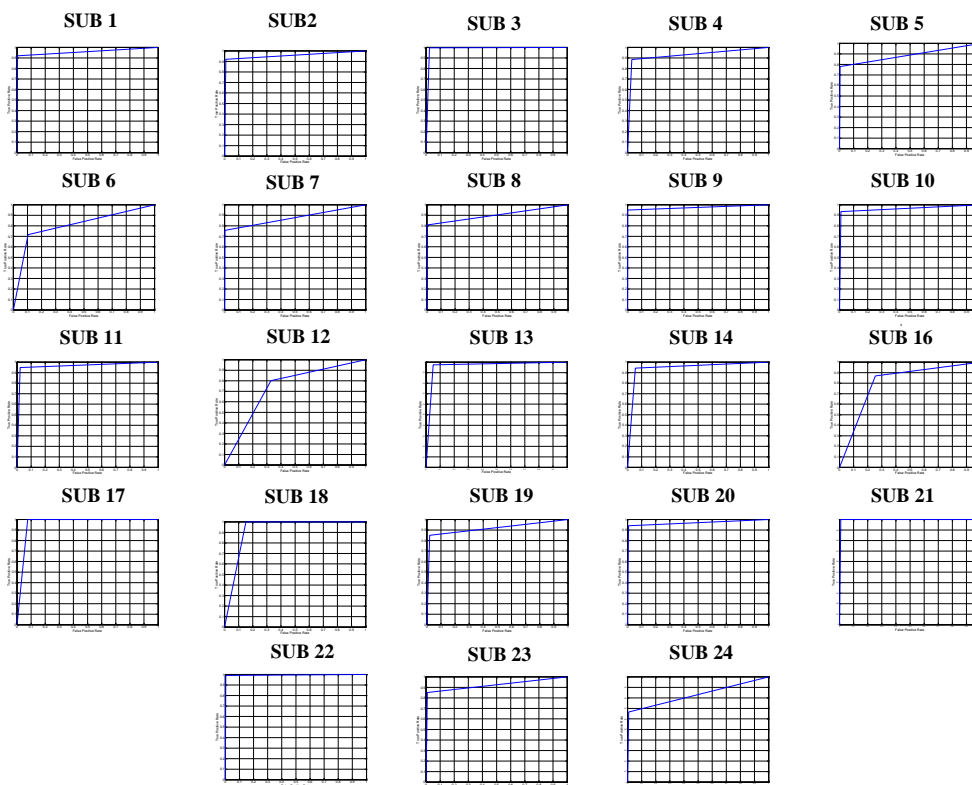| Patient	50% Training (%)			25% Training (%)		
	Sen.	Spe.	Acc.	Sen.	Spe.	Acc.
1	91.96	99.62	99.47	97.32	98.66	98.63
2	91.95	99.07	98.92	100	94.27	94.39
3	99.51	97.52	97.55	100	91.88	92.02
4	88.48	96.95	96.87	96.50	83.08	83.21
5	77.94	99.53	98.86	82.94	99.50	98.98
6	71.60	89.57	89.54	81.15	67.68	67.70
7	75.61	99.78	99.54	85.77	99.30	99.17
8	80.74	99.81	98.83	82.83	96.85	96.13
9	95.00	99.85	99.81	98.10	96.80	96.82
10	93.39	98.98	98.93	86.18	98.99	98.88
11	94.80	97.75	97.51	98.51	94.51	94.83
12	80.12	67.04	67.42	73.49	80.73	80.52
13	97.33	94.96	95.00	96.44	94.04	94.08
14	94.32	94.43	94.43	97.73	91.96	92.00
16	86.84	74.56	74.61	73.68	88.22	88.16
17	100	92.36	92.57	86.04	96.48	96.19
18	100	84.81	85.05	100	79.88	80.20
19	84.87	97.96	97.67	75.42	98.93	98.39
20	94.04	99.22	99.14	94.25	98.14	98.08
21	100	99.03	99.04	99.34	98.53	98.54
22	99.03	99.66	99.65	97.42	98.64	98.62
23	85.12	99.03	98.85	59.75	99.37	98.84
24	66.54	99.12	98.71	67.33	97.50	97.12
Average	89.10	94.80	94.69	88.27	93.21	93.11

Fig. 5. ROC plots for 50% training rate per patient. The x- and y axis represent the false positive rate and true positive rate, respectively.
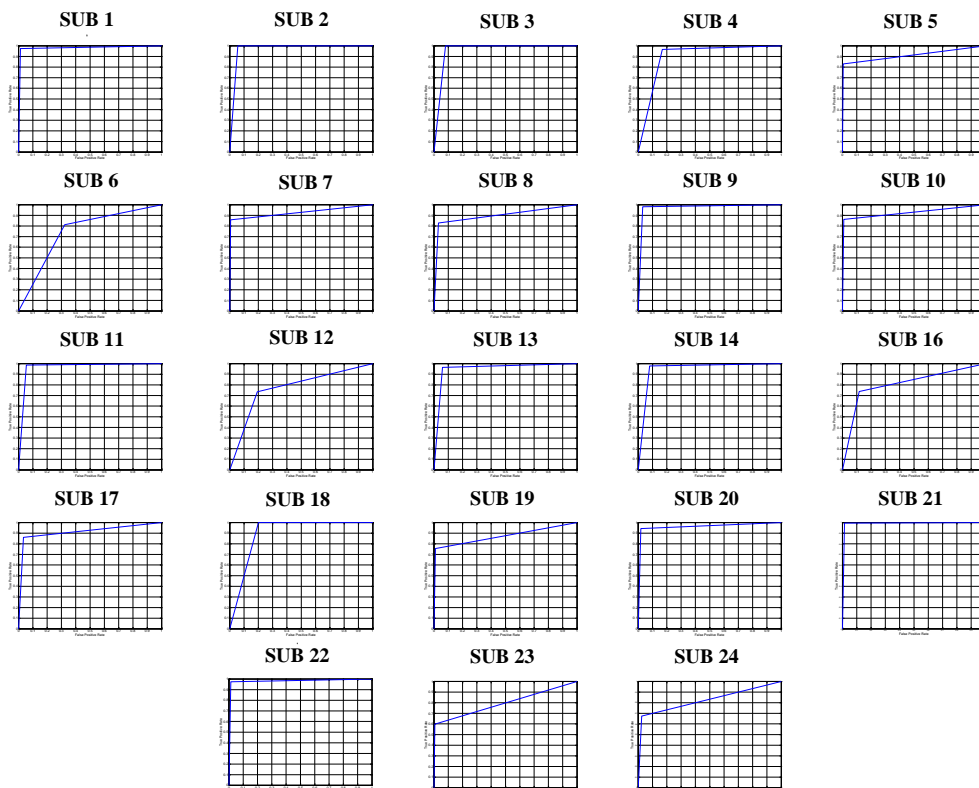
Fig. 6. ROC plots for 25% training rate per patient. The x- and y axis represent the false positive rate and true positive rate, respectively.

AFA per hour $=$

$$\frac{\sum_{S=1}^{NS} number\ of\ false\ alarms\ in\ subject\ "S"}{Total\ duration\ of\ EEG\ recordings\ (in\ hours)\ in\ all\ subjects}, \quad (11)$$

$$AAD = \frac{\sum_{S=1}^{NS} length\ of\ delays\ in\ subject\ "S"}{Total\ number\ of\ detected\ seizure\ events\ in\ all\ subjects}, \quad (12)$$

where NS = 23 is the total number of patients. The ADSR, AFA, and AAD metrics are reported for the test set in TABLE IV.

TABLE IV. Average Detection Sensitivity Rate (ADSR), Average False Alarm (AFA) and Average Alarm Delay (AAD) achieved on test set for seizure event detection

	Using 50% training rate	Using 25% training rate
ADSR (%)	96.29	91.34
AFA per hour	3.04	4.86
AAD (second)	4.65	5.03

As can be seen in TABLE IV, the best results have been achieved using 50% training rate as expected. Because of the existence of noise and artifacts in the EEG records of the patients 6 and 12, the average of false alarms increased significantly. To be more specific, using 50% training rate and not taking into account patients 6 and 12, we achieved 97.14%, 1.85, and 5.63 for ADSR, AFA and AAD, respectively. However, the delay is increased by about 1 second on average while the number of false alarms per hour is approximately reduced by half. Similarly, with 25% training rate, 94.93%, 2.75 and 5.65 were obtained for ADSR, AFA and AAD, respectively.

B. Comparative Evaluations of the Classification Performance

Few recent studies have used this benchmark for evaluation. There are three main reasons for this: 1) high seizure variations both within- and among patients, 2) only bipolar longitudinal montage information provided (lack of full montage information), and 3) long-term EEG recordings in this dataset contain other patterns such as sleep and physiological artefacts which reduce the performance of seizure detection (in particular in patients 6 and 12).

In TABLE V, we summarized the seizure detection methods (i.e., methods which detect the entire duration of seizure events and not only the onset of seizure) applied on CHB-MIT dataset. In order to perform a fair comparative evaluation we compared our approach only with those studies which used a training rate higher than or equal to ours and also used complete data from the benchmark database for evaluation.

Note that no comparisons can be drawn between our results and the ones from [32], [33] and [34], since the patients used in these studies were not specified. In [35], only accuracy is reported which is not a proper metric for such highly unbalanced dataset. Still in the proposed approach with only 25% training rate we achieved around 13% higher accuracy level than the method in [35] although they used 80% training rate.
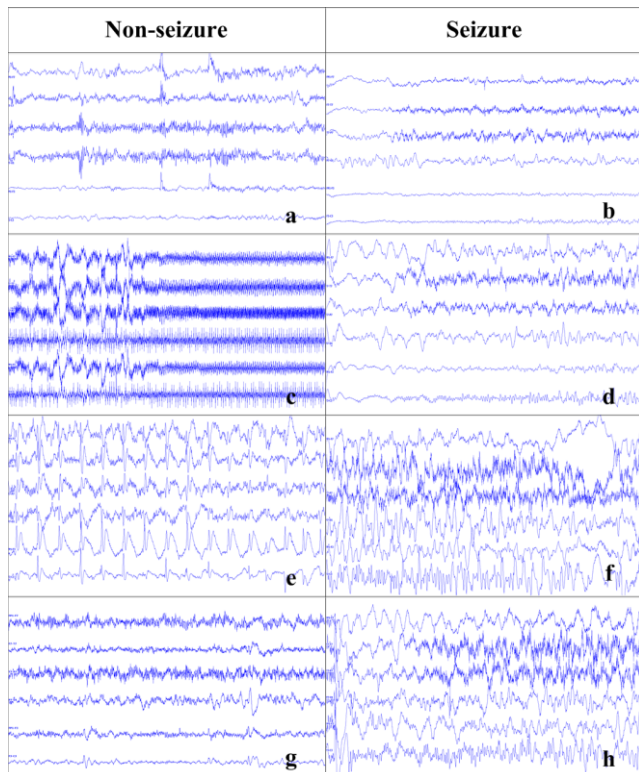
Non-seizure	Seizure

Fig. 7. The non-seizure (first column) and seizure (second column) segments of patient 6 illustrate a high variability within the same class and similarity between distinct classes (length of each segment=10 seconds, channels: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, and F3-C3).

Direct comparison with [36] is not feasible either because there is an ambiguity in how the training and test sets are constructed. Regardless of this ambiguity, the best achieved detection sensitivity rate of 83% was reported in the fifth experiment which corresponds to the 80% training rate. Regardless, we still achieve a significantly higher sensitivity rate with only 25% training rate.

Additionally, in this work an average sensitivity and specificity rate of 89.31% and 95.03% were achieved using only the *early* 25% of each EEG record as training (excluding the patients 6 and 12), which further shows approximately 1% improvement in contrast to our previous work [31] with the same training rate but using a large set of features in a significantly sophisticated classifier network topology. Therefore, our previous method needs a cloud computing implementation with massive parallelization for any on-line processing because we used a large network of classifiers in each Network of Binary Classifier (NBC) [37]. While in the proposed method, only 24 and significantly faster classifiers (23 LDAs and 1 Naïve Bayes) are used. In addition, the proposed method needs only about 3 ms for feature extraction from a 1-s EEG segment while in our previous method it took 280 ms. As a result, the proposed method can easily be ported on a tablet or pocket size computer which makes the proposed EEG classification approach feasible on a mobile application.

C. Comparative Evaluations of the Feature Extraction Approach

In order to perform a fair comparison between the discriminative powers of the proposed feature extraction technique against the other state-of-the-art methods, three different feature sets are extracted from the CHB-MIT benchmark and classified using the same classification scheme explained in section II.E. The first feature set consists of six features used in [38], which are the energy of details and approximation coefficients (d_1, d_2, d_3, d_4, d_5, and a_5). The second feature set was proposed by Kumar et al. [39], where approximate entropy (ApEn) of details and approximation coefficients (d_1-d_5, and a_1-a_5) were proposed for epileptic seizure detection. In these two feature sets, wavelet Daubechies 4 were used in order to decompose the EEG segments into five levels. The third feature set is a combination of eight nonlinear features, including ApEn [40], correlation dimension [41], and recurrence quantification analysis (RQA). The RQA based features are recurrence rate, determinism, averaged diagonal line length, entropy, laminarity, and trapping time [28]. TABLE VI presents the obtained results against the three competitor feature sets over the test data.

The results clearly indicate that the proposed feature extraction approach yields the highest average sensitivity and specificity rates. Based on TABLE VI, the proposed features obtained relatively low results for only two patients, i.e., less than 70% in either sensitivity or specificity while the competing feature sets totally fail in some patients; for instance, the first, second and third feature sets obtained less than 50% specificity for patients 4, 13, and 18. This makes them entirely unreliable in practice for a medical diagnosis application using the proposed classification scheme.

D. Comparative Evaluations of the Classifiers

To evaluate the selection of the classifiers used in the first layer of the proposed approach (i.e., LDA), the performances of different classifiers are also presented while the NB classifier in the second layer is used. In TABLE VII, we compared the sensitivity and specificity of 4 state-of-the-art classifiers including Quadratic Discriminant Analysis (QDA), Mahalanobis Discriminant Analysis (MDA), Naïve Bayes (NB), and SVM (with a linear kernel). The NB classifier used in the first layer employed *kernel smoothing density estimation* in order to model the input features. It is clear that the proposed LDA classifiers can reach much higher sensitivity compared to QDA and NB classifiers. In addition, they can achieve approximately 2% higher average sensitivity and specificity compared to MDA classifier. Even though SVM classifier with linear kernel achieved 1% higher sensitivity than LDA, it fails to classify the EEG data of patients 4 and 6. Moreover, LDA achieved 0.4% higher specificity on average, which is a noteworthy improvement in such unbalanced data, i.e., recall that the average duration of non-seizure segments over 23 patients is around 26444 seconds, 0.4% of which are about 106 seconds. So on average, using SVM would lead to an extra of 106 false alarms. Furthermore, the computational complexity of the SVM classifier is much higher than that of LDA, which may cause problems in practice for real-time processing. This is further investigated in detail in the next sub-section.

TABLE V. The proposed seizure detection methods using CHB-MIT dataset

Authors	Features	Patients	Channels	Training Rate (%)	Av. Sen. (%)	Av. Spe. (%)	Av. Acc. (%)
Rafiuddin et. al. [35]	Interquartile range, median absolute deviation of raw data, energy and coefficient of variation extracted from the Daubiches (db4) wavelet coefficients	23	23	80	-	-	80.16
Uzzaman Khan et. al. [32]	Relative values of normalized coefficient of variation (NCOV) based measure	5 (not specified)	Not specified	80	83.6	100	91.8
Hunyadi et. al. [36]	16 features extracted from time and frequency domain	23	Not specified	≈ 80	83	-	-
Supratak et. al. [33]	Uses stacked autoencoders as unsupervised feature learner	6 (randomly selected)	Not specified-channels were selected manually	Totally 30 epochs used	100	-	-
Fürbass et al. [34]	EpiScan (automatic seizure detection method)	23	-	-	67	-	-
Kiranyaz et. al. [31]	342 features including time, frequency, and time-frequency features	21(excluding patients 6, 12, and 15)	18	25	89.01	94.71	-
Proposed Method	7 features extracted from intersection points of Poincaré section and phase space	23 (excluding patient 15)	23	25	88.27	93.21	93.11
				50	89.10	94.80	94.69

TABLE VI. The classification results of three different feature sets using the proposed classification scheme. Performance metrics less than 70% are highlighted.

	Length of One EEG Segment = 1 second, Training Rate = 50%							
	Proposed feature set		1st Feature Set Energy of DWT Coeff. [38]		2nd Feature Set DWT-based ApEn. [39]		3rd Feature Set Nonlinear [40] [41] [28]	
Patient	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.
1	91.96	99.62	94.64	98.11	94.64	99.48	93.75	90.45
2	91.95	99.07	98.85	16.51	96.55	36.62	78.16	95.57
3	99.51	97.52	96.08	97.01	95.59	97.52	94.61	97.04
4	88.48	96.95	24.61	26.01	21.99	34.78	67.02	83.96
5	77.94	99.53	79.36	98.65	78.29	99.60	98.93	85.66
6	71.60	89.57	65.43	67.99	64.20	73.44	64.20	70.28
7	75.61	99.78	73.17	99.79	69.51	99.96	50.00	93.45
8	80.74	99.81	74.89	99.23	74.46	99.40	77.71	96.78
9	95.00	99.85	82.14	100	87.14	100	68.57	66.46
10	93.39	98.98	84.58	99.86	83.26	99.64	94.27	91.89
11	94.80	97.75	95.30	97.69	94.31	94.14	99.01	74.68
12	80.12	67.04	76.97	79.66	87.80	87.24	77.76	65.88
13	97.33	94.96	100	28.75	98.22	32.21	99.11	42.97
14	94.32	94.43	64.77	98.63	76.14	98.63	85.23	99.22
16	86.84	74.56	81.58	91.75	84.21	97.51	97.37	76.97
17	100	92.36	92.57	93.43	97.30	93.79	91.22	74.58
18	100	84.81	97.52	78.42	96.27	82.52	100	43.19
19	84.87	97.96	84.87	95.85	85.71	99.28	77.31	91.97
20	94.04	99.22	92.05	97.67	94.04	97.14	94.04	95.46
21	100	99.03	90.10	98.45	90.10	98.45	90.10	84.31
22	99.03	99.66	96.12	99.30	97.09	98.51	83.50	91.77
23	85.12	99.03	93.02	99.39	88.84	99.08	90.23	92.76
24	66.54	99.12	65.80	97.05	63.94	99.99	76.95	93.84
Average	89.10	94.80	82.80	85.18	83.46	87.78	84.74	82.57

TABLE VII. The classification results of the proposed features using four different classifiers. Patients with performance metrics of less than 70% are highlighted.

Length of One EEG Segment = 1 second, Training Rate = 50%										
	LDA		QDA		MDA		NB		SVM	
Patient	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.
1	91.96	99.62	95.98	99.45	100	91.00	95.87	98.93	95.98	99.80
2	91.95	99.07	98.85	13.45	90.80	99.97	93.98	98.76	97.70	98.95
3	99.51	97.52	99.51	95.35	100	97.05	100	96.51	99.51	97.10
4	88.48	96.95	23.04	40.85	18.32	97.70	11.11	97.52	32.46	96.00
5	77.94	99.53	86.83	93.90	100	67.80	96.59	93.62	99.64	94.23
6	71.60	89.57	58.02	29.72	18.75	98.10	16.05	77.51	55.56	91.24
7	75.61	99.78	82.32	96.35	91.46	94.98	87.90	96.41	89.02	97.44
8	80.74	99.81	88.74	99.33	93.51	97.21	83.63	99.30	89.83	99.06
9	95.00	99.85	97.14	96.05	97.14	95.69	97.84	76.14	96.43	97.21
10	93.39	98.98	96.92	96.70	96.04	94.55	95.96	98.20	95.59	99.08
11	94.80	97.75	96.53	95.35	97.03	87.87	96.47	98.67	96.04	99.11
12	80.12	67.04	59.76	84.85	86.22	68.38	53.66	77.17	87.40	74.74
13	97.33	94.96	56.89	94.56	96.41	95.31	72.85	97.50	96.44	95.01
14	94.32	94.43	88.64	91.99	88.64	96.92	91.67	95.38	94.32	95.72
16	86.84	74.56	84.21	41.95	76.32	97.07	63.64	79.22	92.11	72.05
17	100	92.36	100	92.00	100	87.08	100	92.97	100	90.81
18	100	84.81	100	77.30	100	75.26	100	85.81	100	82.09
19	84.87	97.96	88.14	93.13	97.48	87.70	82.18	94.83	95.80	96.35
20	94.04	99.22	90.73	99.04	93.38	96.73	87.50	99.13	94.04	98.62
21	100	99.03	85.15	99.85	86.14	99.87	87.75	99.63	100	99.44
22	99.03	99.66	100	97.90	100	96.85	100	98.33	100	98.92
23	85.12	99.03	97.67	99.13	98.60	97.94	98.07	99.06	98.14	99.06
24	66.54	99.12	72.12	98.48	80.67	95.49	68.83	99.53	66.54	99.50
Average	**89.10**	**94.80**	**84.66**	**83.77**	**87.26**	**92.02**	**81.81**	**93.48**	**90.11**	**94.41**

E. Computational Complexity Analysis

In this work, we implemented the proposed method using MATLAB version R2014b and the computations were performed on a standard desktop computer with a 3.4 GHz processor and 16 GB ram. Two metrics are used for evaluating the computational complexity: The first metric is the time taken to extract each feature set from every 1-s long segment of one EEG channel. This value was repeatedly calculated across one hour (3600 segments) to ensure the repeatability of the measured time and is reported in TABLE VIII for the proposed and competing feature extraction methods.

TABLE VIII. Run time (in millisecond) using three different feature extraction methods in the proposed classification scheme

	Proposed Feature Set	1st Feature Set	2nd Feature Set	3rd Feature Set
Average elapsed time for feature extraction of 1s-long EEG segment	2.6	2.8	5.3	146.2

The second metric is the elapsed times for classification (including training and test) over the proposed features for one hour recording with different classifiers, which are shown in TABLE IX. The results clearly demonstrate the superiority of

the proposed approach with LDA classifiers over the proposed features in terms of computational complexity.

Clearly, the proposed method has the advantages of having relatively high sensitivity and specificity, and low computational complexity. The proposed method on the other hand lacks preprocessing steps for removing noise and artifacts. Although, this decreases the computational complexity, it can increase the false alarms due to the contaminated EEG signals especially in patients 6 and 12. Additionally, the proposed method is not able to detect the

TABLE IX. Classification run times (in seconds) using four different classifiers over the proposed features

	LDA (Proposed)	QDA	MDA	NB	SVM
Average elapsed time for classification of 1h EEG recording	0.25	0.27	0.26	58.29	24.28

pre-ictal states.

IV. CONCLUSIONS

In this study, a new multi-channel EEG seizure detection method is presented based on the dynamics of the trajectories in phase space. The proposed Poincaré mapping procedure enables us to study the difference between the dynamics of the seizure and non-seizure segments in a high dimensional phase space. The proposed approach keeps the reconstructed

trajectories unfolded and the computational complexity low.

The proposed seizure detection approach was performed over CHB-MIT database in order to achieve the aforementioned objectives. The results indicate an improved classification performance over competing techniques without any pre-processing. The proposed approach achieves the highest accuracy and minimum false alarm rate among the three state-of-the-art feature extraction methods and four different classifiers, and offers the best trade-off between the anomaly detection accuracy and computational burden. Furthermore, the achieved run time shows the potential application of the proposed approach in Epilepsy Monitoring Units (EMUs).

Extracting better features directly from the phase space rather than from its principal components is the subject of our future study. To accomplish this, noise and artifacts in EEG signals will need to be significantly suppressed prior to subsequent analysis.

REFERENCES

[1] W. H. Organization, "Fact sheet N°999," October 2012. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/.
[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-Based Systems,* vol. 45, pp. 147-165, 2013.
[3] J. Zhang, J. Zou, M. Wang, L. Chen, C. Wang, and G. Wang, "Automatic detection of interictal epileptiform discharges based on time-series sequence merging method," *Neurocomputing,* vol. 110, pp. 35-43, 2013.
[4] O. Blanke, G. Lantz, M. Seeck, L. Spinelli, R. Grave de Peralta, G. Thut, T. Landis, and C.M. Michel, "Temporal and spatial determination of EEG-seizure onset in the frequency domain," *Clinical Neurophysiology,* vol. 111, no. 5, pp. 763-772, 2000.
[5] B. Boashash, and GH. Azemi , "A review of time–frequency matched filter design with application to seizure detection in multichannel newborn EEG," *Digital Signal Processing,* vol. 28, pp. 28-38, 2014.
[6] C. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology,* vol. 116, no. 10, p. 2266–2301, 2005.
[7] V. Bajaj and R. B. Pachori, "Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals," *Biomedical Engineering Letters,* vol. 3, no. 1, pp. 17-21, 2013.
[8] R. Pachori, "Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition," *Research Letters in Signal Processing,* vol. 2008, 2008.
[9] P. Gifani, H.R. Rabiee, M.R. Hashemi, and M. Ghanbari,, "Dimensional Characterization of anesthesia dynamic in reconstructed embedding space," in *Engineering in Medicine and Biology Society (EMBS 2007).*, Lyon, 2007.
[10] Y. Fang, M. Chenb, and X. Zhengc, "Extracting features from phase space of EEG signals in brain-computer interfaces," *Neurocomputing,* p. In Press, 2014.
[11] N. Sriraam, "Correlation dimension based lossless compression of EEG signals," *Biomedical Signal Processing and Control,* vol. 7, no. 4, pp. 379-388, 2012.
[12] F. Shayegh, S. Sadri, R. Amirfattahi, and K. Ansari-Asl, "A model-based method for computation of correlation dimension, Lyapunov exponents and synchronization from depth-EEG signals," Computer Methods and Programs in Biomedicine, vol. 113, no. 1, pp. 323-337, 2014.
[13] C. Allefeld, and J. Kurths, "An approach to multivariate phase synchronization analysis and its application to event-related potentials," International Journal of Bifurcation Chaos, vol. 14, pp. 417-426, 2004.
[14] X. Meng, J. Xu, and F. Gu, "Generalized dimension of the intersection between EEGs," Biological Cybernetics, vol. 85, no. 4, p. 313, 2001.
[15] C. J. Stam, Nonlinear Brain Dynamics, Nova Publishers, 2006.
[16] D. Kugiumtzis, "Surrogate Data Test on Time Series," in *Modelling and Forecasting Financial Data*, Springer Science & Business Media, 2002, pp. 267-282.
[17] R. Sharma , and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Systems with Applications,* vol. 42, no. 3, pp. 1106-1117, 2015.
[18] M. Chen, Y. Fang, and X. Zheng, "Phase space reconstruction for improving the classification of single trial EEG," *Biomedical Signal Processing and Control,* vol. 11, pp. 10-16, 2014.
[19] S. Lee, J. S. Lim, J. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance," *Computer Methods and Programs in Biomedicine,* vol. 116, no. 1, pp. 10-25, 2014.
[20] A. Brignol, T. Al-ani b, and X. Drouot, "Phase space and power spectral approaches for EEG-based automatic sleep–wake classification in humans: A comparative study using short and standard epoch lengths," *Computer Methods and Programs in Biomedicine,* vol. 109, no. 3, pp. 227-238, 2013.
[21] J. Jeong, D. Kim, J. Chae, S. Y. Kim, H. Ko, and I. Paik, "Nonlinear analysis of the EEG of schizophrenics with optimal embedding dimension," *Medical Engineering & Physics,* vol. 20, no. 9, pp. 669-679, 1998.
[22] N. Kannathal, U. R. Acharya, C.M. Lim, and P.K. Sadasivan, "Characterization of EEG—A comparative study," *Computer Methods and Programs in Biomedicine,* vol. 80, no. 1, pp. 17-23, 2005.
[23] Y. Yuan,Y. Li, and DP. Mandic, "A comparison analysis of embedding dimensions between normal and epileptic EEG time series.," *Journal of Physiological Sciences,* vol. 58, no. 4, pp. 239-247, 2008.
[24] C. Anderson, E. Forney, D. Hains, and A. Natarajan, "Reliable identification of mental tasks using time-embedded EEG and sequential evidence accumulation," *Journal of Neural Engineering,* vol. 8, no. 2, 2011.
[25] A. L. Goldberger, L. 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, p. e215–e220, 2000.
[26] A. Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," Ph.D. dissertation, Massachusetts Institute of Technology, September 2009.
[27] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics,* vol. 898, pp. 366-381, 1981.
[28] C. L. Webber, J. P. Zbilut, "Recurrence Quantification Analysis of Nonlinear Dynamical Systems," in *Tutorials in Contemporary Nonlinear Methods for Behavioral Sciences,* Retrieved March 6th from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp, 2005, pp. 27-94.
[29] A. Lyon, and M. Colyvan , "The Explanatory Power of Phase Spaces," *Philosophia Mathematica,* vol. 2, no. 16, pp. 227-243, 2007.
[30] T.W. Sederberg, and T. Nishita, "Curve intersection using Bézier clipping," *Computer-Aided Design,* vol. 22, no. 9, pp. 538-549, 1990.
[31] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term Electroencephalography," *Journal of Biomedical Informatics,* vol. 49, pp. 16-31, 2014.
[32] Y. Uzzaman Khan, N. Rafiuddin, and O. Farooq, "Automated seizure detection in scalp EEG using multiple wavelet scales," in *Signal Processing, Computing and Control (ISPCC), 2012 IEEE International Conference on*, Waknaghat Solan, 2012.
[33] A. Supratak, L. Li, and Y. Guo, "Feature Extraction with Stacked Autoencoders for Epileptic Seizure Detection," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Chicago, IL, 2014 .

[34] F. Fürbass, P. Ossenblok, M. Hartmann, H. Perko, A.M. Skupch, G. Lindinger, L. Elezi, E. Pataraia, A.J. Colon, C. Baumgartner, and T. Kluge, "Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units," *Clinical Neurophysiology*, p. In Press, 2014.

[35] N. Rafiuddin, Y. Uzzaman Khan, and O. Farooq, "Feature Extraction and Classification of EEG for Automatic Seizure Detection," in *International Conference on Multimedia, Signal Processing and Communication Technologies*, Aligarh, 2011.

[36] B. Hunyadi, M. Signoretto, W. Van Paesschen, J. A.K. Suykens, S. Van Huffel, and M. De Vos, "Incorporating structural information from the multichannel EEG improves patient-specific seizure detection," *Clinical Neurophysiology*, vol. 123, no. 12, p. 2352–2361, 2012.

[37] S. Kiranyaz, T. Ince, and M. Gabbouj, "The Classifier Framework: Collective Network of Binary Classifiers," in Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition, Springer Science & Business Media, 2013, pp. 265 - 270.

[38] N. Sadati, H. R. Mohseni, and A. Maghsoudi , "Epileptic Seizure Detection Using Neural Fuzzy Networks," in *IEEE International Conference on Fuzzy Systems*, Vancouver, 2006.

[39] *Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," Signal, Image and Video Processing, vol. 8, no. 7, pp. 1323-1334, 2014.*

[40] N. Kannathal, M. L. Choo, U. R. Acharya, and P.K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187-194, 2005.

[41] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189-208, 1983.

PUBLICATION 2

Patient-specific Seizure Detection Using Nonlinear Dynamics and Nullclines

M. Zabihi, S. Kiranyaz, V. Jäntti, T. Lipping, and M. Gabbouj

IEEE Journal of Biomedical and Health Informatics,

Patient-Specific Seizure Detection Using Nonlinear Dynamics and Nullclines

Morteza Zabihi, Student Member, IEEE, Serkan Kiranyaz, Senior Member, IEEE, Ville Jäntti,
Tarmo Lipping, Senior Member, IEEE, and Moncef Gabbouj, Fellow, IEEE

Abstract— Nonlinear dynamics has recently been extensively used to study epilepsy due to the complex nature of the neuronal systems. This study presents a novel method that characterizes the dynamic behavior of pediatric seizure events and introduces a systematic approach to locate the nullclines on the phase space when the governing differential equations are unknown. Nullclines represent the locus of points in the solution space where the components of the velocity vectors are zero. A simulation study over 5 benchmark nonlinear systems with well-known differential equations in 3D exhibits the characterization efficiency and accuracy of the proposed approach that is solely based on the reconstructed solution trajectory. Due to their unique characteristics in the nonlinear dynamics of epilepsy, discriminative features can be extracted based on the nullclines concept. Using a limited training data (only 25% of each EEG record) in order to mimic the real-world clinical practice, the proposed approach achieves 91.15% average sensitivity and 95.16% average specificity over the benchmark CHB-MIT dataset. Together with an elegant computational efficiency, the proposed approach can, therefore, be an automatic and reliable solution for patient-specific seizure detection in long EEG recordings.

Index Terms— EEG, seizure detection, nonlinear dynamics, phase space, nullcline, LDA, ANN.

I. INTRODUCTION

EPILEPTIC seizures occur when the normal neuronal system in the brain turns into a hyperexcitable network. This hyperexcitability is the result of a disruption in the balance of the neuronal network [1]. The disruption factors, for instance, can change the neural excitation and inhibition balance, or the ion concentration of the cells [1]. In a healthy nervous system, this balance is maintained using different mechanisms, while in an epileptic brain this balance is altered [2].

The series of mechanisms that affect the epileptogenesis, such as the neurons firing rates or the permeability of their membrane, prove that the nature of the neuronal system has a predominantly wide dynamic and nonlinear behavior. Therefore, theories of nonlinear dynamic systems are suitable mathematical characterizations for such biological systems. To be more precise, it has been shown that nonlinear dynamics can truly provide an explanatory framework for complex neuronal systems and bioelectric data. This applies to both small-scale neuronal circuits [3] and large-scale systems such as brains [4] [5] [6].

Often, dynamic systems are formulated by a set of coupled differential equations. Differential equations express the evolution of the system's variables with respect to each other and for a specific initial condition. The evolutions are shown as trajectories in an N-dimensional space, known as the "Phase Space". The phase space represents the correlation between the variables and their dynamic evolution from one state to the next. Therefore, to have a rigorous characterization of a dynamic system, all variables should be known in advance. However, in large-scale systems, high-dimensional nonlinear differential equations may make it infeasibly hard to derive an analytic solution. In such cases, numerical solutions and graphical representations of the equations are usually pursued to assess the dynamics of the differential equations. Using numerical methods, for instance, one can approximate the system behavior in a neighborhood of equilibrium points (steady states) based on the linearized stability principle [7]. In graph-based methods, phase space [8] plays a major role. More specifically, the dynamic properties of the trajectories in the phase space can be revealed without solving the underlying differential equations themselves. For instance, while studying a neural circuit, where the system dynamics are modeled by a set of differential equations, one can explore bifurcation events by changing the parameters of the differential equation [9] or observe the strange attractors using the phase space. However, this approach is usually followed at a cellular level or in small-scale systems.

So far, such applications of phase space are possible only when the governing differential equations of the system are available. However, these applications are extremely restricted in the study of large-scale cortical systems. The main reason is that finding these differential equations, which describe such complex systems (e.g., brain) is a challenging problem. Examples of such difficulties may include the uniqueness of the coefficients of the differential equation, the complexity of the system, and the chaotic behavior in the measurement [10] while the underlying differential equations can have a simple form (e.g., Rössler system [11]). Due to these limitations, the phase space is directly reconstructed from the time series measured from the system [12] [13] based on the Takens' delay embedding theorem [8]. This

M. Zabihi is with the Department of Computing Sciences, Tampere University, Tampere, Finland (correspondence e-mail: morteza.zabihi@tuni.fi).

S. Kiranyaz is with the Electrical Engineering Department, Qatar University, Doha, Qatar

T. Lipping, and M. Gabbouj are with the Department of Computing Sciences, Tampere University, Tampere, Finland

V. Jäntti is with the Clinical Neurophysiology Department, Seinäjoki Central Hospital, Seinäjoki, Finland

theorem states that given the lack of knowledge about the evolution of a system states, the system information can be retained. In several studies focusing on seizure occurrence using Electroencephalograph (EEG) recordings, phase space reconstruction has been used. The common hypothesis is that multiple states (e.g., seizure and non-seizure) exist in an epileptic brain and their differences can be described by the parameters and the initial conditions of the unknown differential equations. Consequently, these differences can be qualitatively visualized in the reconstructed phase space [14].

The studies mentioned above can be categorized into two major groups. The first group is mainly application-based where the phase space features are extracted and tuned to discriminate between seizure and non-seizure states [15] [16] [17]. Although the extracted geometrical features are effective for a particular application, they do not specifically describe a relevant property of the system. In other words, it is not clear what characteristics of the system are reflected by these features. For instance, in [16] a seizure detection method based on the Poincaré Section was proposed, where the reconstructed trajectory of EEGs was projected on a plane formed by the first two principal components (PCs). A least square line was then fitted to the projected trajectory, and several statistical descriptors of the intersection points (between the line and the trajectory) were used as features. As mentioned earlier, despite the high accuracy achieved for seizure detection, the proposed features do not reflect an accurate insight for characterizing the underlying system's behavior. The main drawback of this feature extraction method is that when the reconstructed trajectory is projected on a 2-dimensional space using PCs, the crucial information about the time order of the states is lost.

In the second group, the extracted features describe properties such as the sensitive dependence of the system evolution with respect to the initial values (e.g., Lyapunov exponents [18] [19]), and some properties of the attractors' geometry (correlation dimension [19]). However, any practical use of such measures is challenging, and many considerations should be taken into account. Lyapunov exponent, for example, measures how fast nearby trajectories separate over time. This definition can precisely be determined on the "real phase space". However, in the absence of the main differential equation, it should be estimated with caution. The main reason is that distances in the real phase space may shrink (instead of growing) in the reconstructed phase space due to the angle of the observation (i.e., the physical measurement) [10]. Moreover, Lyapunov exponent expresses the long-term dynamics of the system, while in seizure detection the features are supposed to be extracted from a relatively short time window (e.g., one second) for a reasonable time resolution.

This study proposes a novel characterization of complex systems based on nonlinear dynamics. The proposed method is designed to study the nullclines of 3D systems. Nullclines denote the locus of points in the phase space where the components of the velocity vectors (i.e., $\frac{dx}{dt}$, $\frac{dy}{dt}$, and $\frac{dz}{dt}$) are zero. Such a unique set of points provides us qualitative information about the dynamics of the states without having access to the solutions of the governing differential equation.

Nullclines are used to discover the direction of the trajectories in the phase space (solution space), equilibrium points, and their stability status without solving the differential equation analytically. The proposed method is then used for the characterization and detection of epileptic seizures in long-term EEG recordings.

The main objective of the paper is to propose novel descriptors that can discriminate seizures from normal brain activity. In particular, we aim to accomplish: 1) a novel discriminative feature extraction method based on the nullcline analysis; 2) a "Proof of Concept" by direct evaluation over the five well-known chaotic systems, and 3) a practical and systematic approach with a significantly low computational complexity. The overall systematic approach has been tested over the benchmark CHB-MIT EEG dataset to validate these objectives. Although there are numerous EEG classification methods proposed in the literature, only a few of them (e.g., [16] and [20]) consider the real-world scenario where only the past patient history can be used to train the classifier. For example, the study in [21] used the majority (i.e., >75%) of the EEG record for training which is not a feasible option in a real clinical case. This is why we evaluated the proposed approach against the recent state-of-the-art methods [16] and [20] that used only a small portion (i.e., the early 25%) of each EEG record for training while the larger portion (i.e., the remaining 75%) is used for testing.

The rest of the paper is organized as follows. Section II presents the theory of the proposed feature characterization using the nullcline analysis. To further show the effectiveness of the proposed approach, the Appendix includes an application to 5 differential equations, each describing a well-known system in Chaos theory. In Section III, the proposed approach is evaluated over CHB-MIT benchmark EEG dataset. The results of classification and the effectiveness of the extracted features are presented in Section IV. Section V concludes the paper and suggests topics for future research.

II. THEORETICAL FRAMEWORK

Nullclines [22] present a more comprehensive understanding of the solution of a differential equation by providing information about the direction of the trajectories and the equilibrium points in the phase space. The concept of nullclines plays a central role, especially when the given differential equation is not analytically solvable or a qualitative interpretation only based on the phase space is not sufficient. In this section, nullclines and the proposed approach for visualizing their geometric shapes in the reconstructed phase space are described. The evaluation of the proposed approach over 5 well-known nonlinear systems is provided in the Appendix.

As a motivation for the concept, let us consider the following differential equation in 2D:

$$\begin{cases} \frac{dx}{dt} = f(x,y) \\ \frac{dy}{dt} = g(x,y) \end{cases} \tag{1}$$

Then x-nullcline is defined as a set of points in the phase space where $\frac{dx}{dt} = 0$. Likewise, the y-nullcline is a set of points in the

phase space where $\frac{dy}{dt} = 0$. The x-nullcline geometrically represents the set of points that remain unchanged along the x coordinate through time. Such points evolve only vertically (up or down) with time. Similarly, the y-nullcline evolve only horizontally (left or right).

The intersection of "all" nullclines yields the location of the equilibrium points derived by solving $\frac{dx}{dt} = \frac{dy}{dt} = 0$. Additionally, nullclines divide the phase space into different regions in terms of the relative signs of $\frac{dx}{dt}$ and $\frac{dy}{dt}$. In other words, it shows the flow direction (direction of the velocity vectors) of the variables in the solution space. Thus, depicting the flow direction in a neighborhood of the equilibrium point reveals the stability status (i.e., stable, non-stable, or saddle) of the steady states. This concept can be generalized to higher dimensions.

This stability test is used for qualitative analysis of differential equations. In this approach, first, the nullclines are sketched in the phase space formed based on the given differential equations. Then, the evolution of the system states is determined inside the regions formed by the nullclines. Once the aforementioned information is depicted in the phase space, the dynamics of the system including the equilibrium points and their stability status can be explored. However, if the differential equations are unknown and only physical measurements from the system exist, then another approach is needed to sketch the nullclines. In this study, an approach is proposed to extract information about these "directivity curves" using the nullcline concept in the absence of the governing differential equation. The proposed systematic approach can be described in the following three steps:

Step 1: the phase space is reconstructed using the embedding dimension m and time lag τ based on the delay-embedding theorem [8]. The i'th state vector (i.e., a point) in the reconstructed phase space of the time series $s(n), n = 1, 2, \ldots, N_s$, is defined as:

$$S_i = [s(i), s(i+\tau), s(i+2\tau), \ldots, s(i+(m-1)\tau)]^T \quad (2)$$

Step 2: The flow of each state in the reconstructed phase space (i.e., the time evolution of a projectile in an m-dimensional space) is obtained using numerical gradients. This provides the estimated gradient values, i.e. $\frac{dx}{dt}$ and $\frac{dy}{dt}$.

Step 3: As mentioned earlier, the x- and y-nullclines are defined as the points where $\frac{dx}{dt} = 0$ and $\frac{dy}{dt} = 0$, respectively. Moreover, nullclines divide the phase space into different regions in terms of the relative signs. Therefore, the zero-crossing points of the $\frac{dx}{dt}$ and $\frac{dy}{dt}$, which are achieved in Step 2, are calculated in the phase space. The zero-crossing points, in fact, reflect the position of the nullclines on the reconstructed trajectories.

In Fig. 1, for instance, the evolution of the Lorenz system is shown with small arrows in the reconstructed phase space. In addition, the nullcline points of the system are shown with red, green, and black filled circles which correspond to X-, Y-, and Z-nullclines, respectively. It can be clearly seen that the three types of nullcline points are getting close to each other when

they approach the centers of the two wings and intersect in the center of the reconstructed phase space. This is the location where the states switch their spirals from one wing to the other (left part of the figure). This means that the nullclines show the signs of convergence in the center of the wings, where the steady states of the system are located. The same arguments also apply for all evaluated systems in the Appendix.

The proposed approach uses the phase space reconstruction and benefits from the main properties of the delay-embedding theorem [8], which preserves the structure of the original attractor. It uses the nullcline to reflect information about those states of the system that have a zero rate of change (i.e., zero velocity). However, some practical limitations should be addressed. In the shown Lorenz system (and all the systems covered in the Appendix), the x variable is directly used as a measurement of the system to show the concept of the proposed nullcline analysis. However, a realistic physical measurement is a projection of the system's variables, and it is also contaminated with noise. In addition, although the reconstructed phase space preserves the structure of the underlying dynamic, it is a sparse representation of the original attractor. This limits our access to the entire solution space. For instance, if enough information about the evolution of the trajectory is provided, then the convergence of the nullcline points near the equilibrium points is more visible. Thus, as the proposed approach is based on the phase space reconstruction, its limitations will reside too.

III. PROPOSED METHODS

A. Benchmark Dataset

CHB-MIT dataset [23] [24] was collected at the Boston Children's Hospital from pediatric patients with intractable seizures. In this study, we used the following 23 channels (based on the international 10-20 system) FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8 and T8-P8. The sampling frequency is 256 Hz with 16-bit resolution. All seizure events were annotated with one-second (1-sec) resolution. We have used only those records, which contain at least one seizure event. As records 1 and 21 were obtained from the same female patient with 1.5 years apart, we consider them as two distinct patients. In addition, patient 15 was excluded from our analysis because we failed to read the EEG data of this patient (the same issue was reported in earlier studies using this dataset). Therefore, 23 subjects with more than 171 hours EEG recordings are analyzed in this study. Further details about the used EEG segments and the duration of the seizure events for each patient can be found in [16].

B. Nullcline Feature Design

In this study, first, the EEG recordings are partitioned into non-overlapping 1-sec segments. The reasons for choosing 1-sec windowing are motivated by the following: (1) the time resolution of the ground truth provided by the experts is one second; (2) the same segment length was used by previous works, e.g. [16] [20]. Thus, using the same time frame length
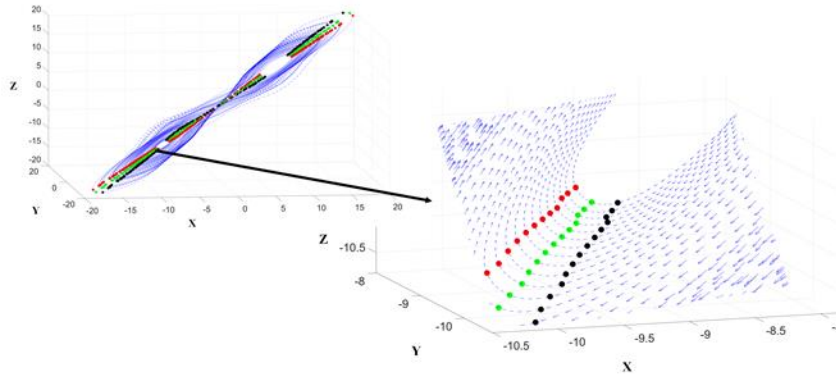
Fig. 1. The velocity plot and nullcline points of the Lorenz system on the reconstructed phase space

makes the comparison possible. Because EEG recordings are prone to noise and artifact contamination, each segment is passed through a band-pass second-order Butterworth filter (1 - 60 Hz). The filtered segments are further smoothed using a 3-point moving average filter. After filtering, three features are extracted based on the location of the nullclines as follows:

First, the phase space of each segment is reconstructed using an embedding dimension $m = 3$ and a time lag $\tau = 31$ [8] (these values are chosen empirically). Although there are some approaches for selecting the proper values of m and τ, visual inspection is still needed to insure a reasonable choice. For instance, too small values of τ make the states strongly correlated. On the other hand, too large values of τ may hide informative structures in the phase space.

Second, the trajectory points located on the nullclines of the three dimensions are determined using the approach presented in Section II. Afterward, the Euclidean distances of the x-, y-, and z-nullcline points from the origin $(x_0, y_0, z_0) = (0,0,0)$ are calculated. The medians of these distances in each dimension are used as features. In other words,

$$F_j = Median\left(d_j^1, \dots, d_j^{N_j}\right) \qquad (3)$$

where F_j is the j-nullcline feature and d_j^i is the distance of the i'th state vector (point) on the j-nullcline from the origin $(0,0,0)$. N_j is the total number of j-nullcline points. In Fig 2, two sample feature plots of patients 1 and 11 are shown. The red and blue samples represent the features of the seizure and non-seizure segments, respectively. Visual inspection of both plots reveals the discrimination power of the proposed features. A further statistical analysis of nullclines on the relevant channels (for patients 1 and 11) indicates that during a seizure attack, the nullclines get far away from the origin of the reconstructed phase space. Moreover, in the absence of seizure, the nullclines tend to be near the origin. In addition, it can be seen that seizure features have more variation compared to non-seizure features. In other words, for these 2 patients, during a seizure the irregularities of nullclines locations increase as opposed to the background EEG of the normal brain function.

C. Classification, Training Strategy and Postprocessing

Classification: Epileptiform activities arise from different regions of the cerebral cortex. In some seizures, such as bilateral seizures, the focus can also vary with respect to time. Moreover, an epileptic patient can experience different seizure types.

Therefore, in the diagnosis phase, when no prior information is available, analyzing all the channels of each EEG record is necessary. As this presents a large amount of data from several independent channels, we follow a "Divide & Conquer" strategy that is experimentally verified to yield the highest classification performance. To accomplish this, an ensemble of classifiers network is constituted where all available channels are used in a two-layer classification scheme. In the first layer, 23 Linear Discriminant Analysis (LDA) classifiers are used, each of which corresponds to one of the 23 EEG channels. Each LDA is a binary classifier that uses the proposed features extracted from the one-second long segments as inputs to learn the significance of that channel. The 23 classification results form 23 class vectors for each segment of the EEG record. The main advantages of the LDA classifier are their closed-form solution and their low computational complexity. These properties make LDA an excellent choice to tackle long-term EEG data.

At this stage, for each 1-sec segment of EEG data, 23 labels ($C_{seizure}$ or $C_{nonseizure}$) are generated, while a single classification decision is required for each segment. Therefore, a fusion rule is needed to combine the generated labels. For the fuser classifier, the feedforward Artificial Neural Network (ANN) with one hidden layer and 23 hidden neurons yielded the highest overall classification accuracy. With proper training by the Bayesian regularization backpropagation algorithm [25], the proposed ANN is able to learn and discriminate the relevant channels for seizure detection from the corresponding LDA outputs and discard the others in order to maximize the final classification accuracy. Therefore, the ANN is trained to fuse the 23 class vectors to generate the final classification output for each 1-sec segment.

Moreover, in order to train the ANN classifier appropriately for such highly imbalanced problem (in total, the ratio of seizure to non-seizure is ~1:66 in the training set), we balance the training data. For this purpose, we compare the performance of random undersampling, random oversampling, and synthetic minority oversampling technique [26]. The best performance is achieved by random undersampling. Therefore, the dominant class (non-seizure) is randomly sampled so that the number of seizure and non-seizure samples are kept equal.

Training Strategy: In order to mimic a real-life personalized EEG monitoring scenario, each EEG recording is split into two time segments, where the first time segment (considered as
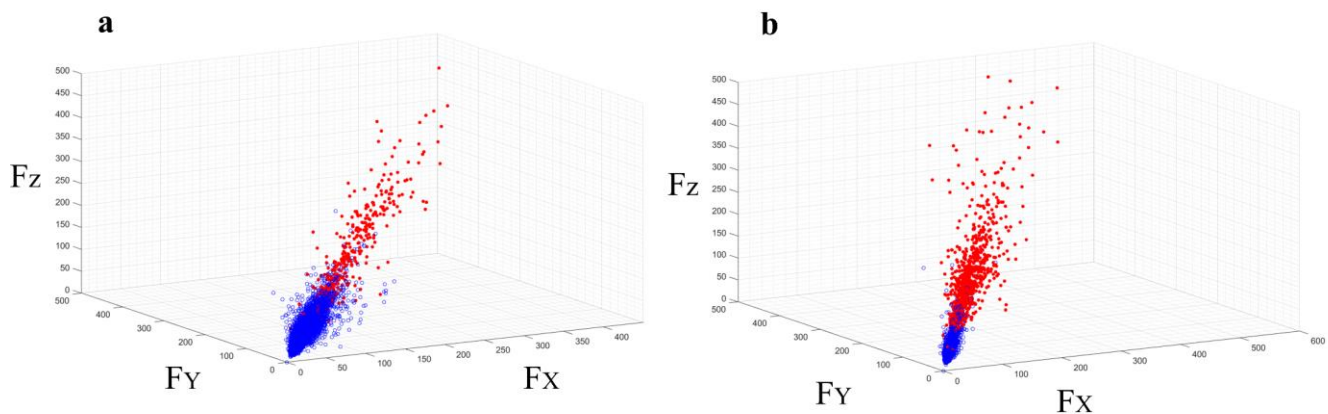
Fig 2. The sample scatter plots of the feature space (the axes represent the proposed features). The features of the seizure and non-seizure samples are shown with red and blue, respectively. a) Feature space for patient 1-channel F7-T7. b) Feature space for patient 11-channel F3-C3.

"early" or prior segment) is used for training and the second time segment (considered as the future record over which the seizure detection takes place) is used for testing. The first time segment containing 25% of the early seizure and non-seizure events in each EEG record is used for training (this segment include 2307 seconds of seizure and 152052 seconds of nonseizure events). The remaining 75% of the record (the second time segment) is used for testing to evaluate the classification performance. Such a partitioning also mimic the simulated real-world patient-specific scenario, where a medical expert would only annotate a small portion of the seizure/non-seizure events, and then the trained system is expected to detect accurately all the seizure events that occur in the future (see Fig. 3).

Postprocessing: Segment labels are filtered to refine the classification output. This procedure can increase the classification performance at a low cost by removing the occasional classification noise (misclassifications) within a seizure or non-seizure event. The purpose of such a postprocessing step is to ensure the desired classification properties such as continuity and temporal similarity (via morphological filtering), and smoothness (via moving average filtering). The morphological filter is designed based on the proposed fuzzy rule-based filtering technique in [20].

IV. EXPERIMENTAL RESULTS

A. Classification Performance Evaluation

To evaluate the classification performance, we report standard performance metrics in Table I: Sensitivity (Sens), Specificity (Spec), Accuracy (Acc), and the Area Under the ROC Curve (AUC). The respective definitions of the first three metrics using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are as follows: Accuracy is the ratio of the number of correctly classified frames to the total number of frames classified, i.e., $Acc = (TP+TN)/(TP+TN+FP+FN)$; Sensitivity is the rate of correctly classified seizure frames among all seizure frames, $Sen = TP/(TP+FN)$; and Specificity is the rate of correctly classified non-seizures among all non-seizures, $Spe = TN/(TN+FP)$. In addition, the ROC graph is obtained by plotting the *Sensitivity* versus *(1 − Specificity),* and it is shown in Fig. 4.

Even though accuracy reflects the overall classification performance, due to the highly unbalanced numbers of seizure and non-seizure frames, the accuracy is not a reliable performance metric. Since the accurate classification of seizure frames is of utmost importance, the sensitivity rate becomes the primary performance metric. On the other hand, a reasonable level of specificity should suffice since the human expert can easily discard or correct false positives. To further clarify the archived performance, the confusion matrix for each patient is also reported in Table II. Therefore, other evaluation metrics, such as Precision and F-measure, can be computed.

For a fair comparative evaluation, the classification performance of the proposed approach along with the Nullcline features is tested against the state-of-the-art EEG classification methods, [20] and [16], which also used the same train/test partitioning. Therefore, we excluded some other works (e.g. [21]), which used 80% and even higher portions of each patient data for training. In practical clinical usage, assuming such a large volume of training data is obviously quite hard, if feasible at
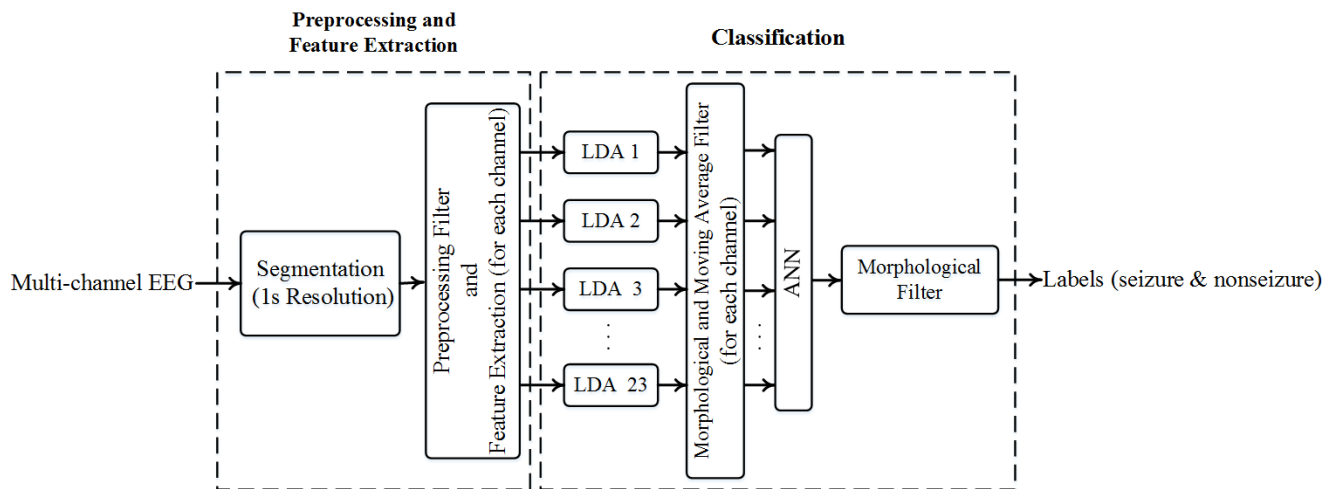
Fig. 3. Block diagram of the proposed seizure detection system for multi-channel EEG recordings

TABLE I. The classification results over the test set using 25% of the data for training. The achieved results are based only on a single feature (the median of the Euclidean distances of x-, y-, and z-nullcline from the origin). The results of our two previous works are reported between parentheses ([20] & [16]). NR means "Not Reported".

Patient	Sens (%)	Spec (%)	Acc (%)	AUC (%)
1	99.40 (94.66 & 97.32)	99.59(99.72 & 98.66)	99.59 (NR & 98.63)	99.50
2	99.24 (98.48 & 100)	93.90(80.17 & 94.27)	94.02 (NR & 94.39)	96.57
3	100 (97.07 & 100)	93.27 (92.74 & 91.88)	93.37 (NR & 92.02)	96.63
4	100 (71.78 & 96.50)	97.65 (52.46 & 83.08)	97.67 (NR & 83.21)	98.83
5	91.47 (96.93 & 82.94)	98.75(99.00 & 99.50)	98.52(NR & 98.98)	95.11
6	0.00 (NR & 81.15)	100(NR & 67.68)	99.83 (NR & 67.70)	50.00
7	93.90 (90.24 & 85.77)	99.36 (99.60 & 99.30)	99.31 (NR & 99.17)	96.63
8	77.20 (95.67 & 82.83)	97.20 (94.51 & 96.85)	96.17 (NR & 96.13)	87.20
9	97.14 (100 & 98.10)	98.16 (75.58 & 96.80)	98.15 (NR & 96.82)	97.65
10	95.59 (92.38 & 86.18)	96.67 (96.55 & 98.99)	96.66 (NR & 98.88)	96.13
11	96.53 (95.88 & 98.51)	99.96 (95.73 & 94.51)	99.68 (NR & 94.83)	98.25
12	92.65 (NR & 73.49)	95.60 (NR & 80.73)	95.51 (NR & 80.52)	94.13
13	97.92(86.13 & 96.44)	91.62 (96.25 & 94.04)	91.73 (NR & 94.08)	94.77
14	98.48 (81.95 & 97.73)	90.06 (99.75 & 91.96)	90.12 (NR & 92.00)	94.27
16	91.23(86.81 & 73.68)	79.69(98.80 & 88.22)	79.74 (NR & 88.16)	85.46
17	99.55 (63.51 & 86.04)	95.57 (98.23 & 96.48)	95.68 (NR & 96.19)	97.56
18	100 (81.07 & 100)	75.36 (93.99 & 79.88)	75.76 (NR & 80.20)	87.68
19	96.09(98.89 & 75.42)	97.79 (97.13 & 98.93)	97.75 (NR & 98.39)	96.94
20	99.56 (93.01 & 94.25)	94.20 (97.79 & 98.14)	94.28 (NR & 98.08)	96.88
21	98.68 (95.42 & 99.34)	98.26 (98.34& 98.53)	98.27 (NR & 98.54)	98.47
22	89.68 (100 & 97.42)	97.67 (93.29 & 98.64)	97.52 (NR & 98.62)	93.67
23	96.28 (67.9 & 59.75)	98.58 (99.05 & 99.37)	99.47 (NR & 98.84)	97.45
24	85.89 (83.08 & 67.33)	99.74 (99.59 & 97.50)	99.57 (NR & 97.12)	92.82
Average	**91.15** **(89.01 & 88.27)**	**95.16** **(94.71 & 93.21)**	**95.11** **(NR & 93.11)**	**93.16**

all, for long-term EEG recordings. In addition, we excluded the studies that did not report the overall classification performance from all the records (e.g. [27], [28]) in the benchmark dataset. Finally, the majority of studies in this domain (e.g., [29]) are mere onset detectors, i.e., they only detect the earliest stage of a seizure event. Therefore, those studies have also been omitted from the comparative evaluations since the proposed approach can detect both the onset and the duration of each seizure event

Table I presents the performance metrics obtained per patient in the database using the proposed approach and the two competing methods. The proposed approach achieved an average sensitivity and specificity of 91.15% and 95.16%, respectively, which exhibit a superior performance level compared to the competing methods. For a fair comparison against the state-of-the-art method [20] if patients 6 and 12 are

excluded the average sensitivity and specificity of the proposed approach are 95.22% and 94.91%, respectively. Now a significant performance gap in sensitivity becomes visible. Furthermore, the results of patients 4, 7, 9, 10, 12, 17, 23 and 24 are significantly improved compared to the competing methods. Specifically, the superiority of the proposed approach over the challenging EEG data for patients 4, 12 and 23 is visible.

Among the 23 patients' data, the proposed approach performed poorly only for two patient records: 6 and 8. To shed more light on these cases, the corresponding confusion matrices can be seen in Table II. For these patients, the seizure segments were misclassified as non-seizure and this makes False Negative-FN counts relatively high. Obviously, for these two patients the unprecedented low training rate, 25% used in this
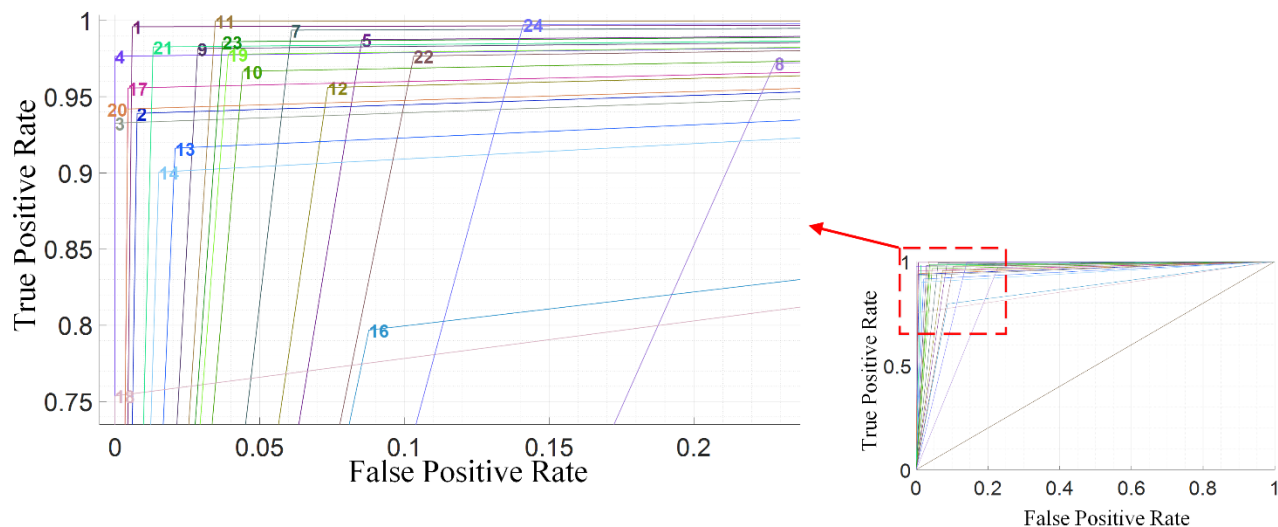
Fig. 4. The ROC curve of the 23 patients. The left-hand side plot shows the magnified area for all patients (excluding patient 6). The corresponding ROC curve of each patient is determined by the patient's number.

TABLE II. True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP) achieved for the test set using 25% of the data for training.

Patient	Seizure detected as seizure (TP)	Seizure detected as non-seizure (FN)	Non-seizure detected as non-seizure (TN)	Non-seizure detected as seizure (FP)
1	334	2	17534	72
2	130	1	5622	365
3	306	0	17340	1252
4	286	0	27813	669
5	386	36	12913	164
6	0	122	69789	0
7	231	15	24002	154
8	535	158	12447	359
9	204	6	25191	472
10	325	15	36256	1250
11	585	21	6933	3
12	706	56	24265	1117
13	330	7	17006	1555
14	130	2	16900	1866
16	52	5	10711	2730
17	221	1	7545	350
18	242	0	11276	3686
19	172	7	7558	171
20	225	1	13940	859
21	150	2	10012	177
22	139	16	7759	185
23	311	12	23534	332
24	347	57	31912	83

study is clearly insufficient for a reasonable generalization, and therefore a larger training dataset is needed. However, the total FP for all patients in [16] was 45626. This large number of FP was reduced by the proposed approach to 17871.

Table III presents the performance of the recent studies that used CHB-MIT dataset. As discussed earlier, although it is unfair to perform a direct comparison with the proposed method that used significantly less amount of data for training in order to mimic an actual clinical case, it still achieves higher performance than the methods proposed in [30] - [35] and [44]. The methods proposed in [29], [36] - [40], [43], [46] and [47] have achieved higher sensitivity and accuracy while using the majority of data for training. Besides, for supplementary reading in epileptic detection algorithms and the hardware

implementations of seizure detection methods, the readers are referred to [48] - [52].

B. Computational Complexity

The proposed seizure classification algorithm was implemented using MATLAB version 9.1.0.441655 (R2016b). The experiments were performed on a workstation with Intel® Xeon® Processor E5-2650 v4 at 2.20 GHz (48-core processor) and 32 GB memory. MATLAB parallel computing toolbox was used to execute the tasks. The average time for preprocessing and feature extraction of a 1-sec EEG segment in 23 channels (23×1-sec segments) is about 10 msec. Training the classifier network for one-hour EEG recording (with 23 channels) takes on average 1.78 sec. In the testing phase, the average time for

TABLE III. The experimental setting and performance results of the existing methods using the CHB-MIT dataset

Ref	Patients	Training and Testing set	Results
[29]	All	- Leave-one-out cross validation (not patient specific and onset detection)	Avg. Sen = 96.00%
[30]	All	- Leave-one-out cross validation (not patient specific)	Avg. Sen = 82% Avg. Spe = 95.5%
[31]	3 patients (3, 8, and 13)	- Randomly use 2/3 of the data for the training step and the 1/2 left for the test set	Avg. Acc = 94.85%
[32]	All	- Each test dataset contains one seizure for each patient, and the signals are augmented	Avg. Sen = 80.6% Avg. Spe = 91.7% Avg. Acc = 85.6%
[33]	17 patients (patients are not specified)	-	Avg. Sen = 89.24% Avg. Spe = 82.98%
[34]	All	- Recordings containing the seizure activity were split such that duration of seizures in both data portions, i.e., those employed for testing and training, was nearly equal (patient-specific) ~50%	Avg. Sen = 87% Avg. Spe = 99%
[35]	7 patients (1, 3, 5, 8, 19, 20 and 24)	-	Avg. Acc = 92.68%
[36]	All	- 10-fold cross-validation (90% training)	Avg. Sen = 97.9% Avg. Spe = 99.6% Avg. Acc = 99.4%
[37]	All	- Two experiments with different window size (30 sec and 70 sec) - 10-fold cross-validation (90% training)	Avg. Sen = 89.73% Avg. Spe = 94.77% Avg. Acc = 92.46% Avg. Sen = 97.12% Avg. Spe = 99.29% Avg. Acc = 98.30%
[38]	All	- Two sets of experiments with convolutional neural network using time and frequency domains signals - 6-fold cross-validation	Avg. Sen = 61.2% Avg. Spe = 63.3% Avg. Acc = 62.3% Avg. Sen = 96.9% Avg. Spe = 98.1% Avg. Acc = 97.5%
[39]	All	- 5-fold cross-validation (not patient specific)	Avg. AUC = 95.72% Avg. Acc = 94.37%
[40]	All	- Leave-one-out cross validation (not patient specific)	Avg. AUC = 96.1%
[41]	9 patients (1, 2, 5, 8, 11, 18, 19, 21 and 22)	- Leave-one-out cross validation (not patient specific)	Avg. Sen = 98.28% Avg. Spe = 87.04% Avg. Acc = 93.16%
[42]	All	- Unsupervised	Avg. Spe = 86.00%
[43]	All	- 5-fold cross-validation (not patient specific)	Avg. Sen = 99.8% Avg. Spe = 99.6%
[44]	All	- Dataset randomly partitioned into 70% as training and 30% for testing set (not patient specific)	Sen = 88% Spe = 88%
[45]	18 patients (1, 3, 5, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22 and 23)	- Dataset is randomly partitioned into 70% as training and 30% for testing set (not patient specific)	Avg. Sen = 87.5% Avg. Spe = 99.9%
[46]	All	- Stratified 10-fold cross-validation (patient specific)	Avg. Sen = 99.84% Avg. Spe = 99.86%
[47]	All	- 5-fold cross-validation (not patient specific)	Avg. Sen = 99.05% Avg. Spe = 99.98% Avg. Acc = 99.52%

each 1-sec segment, including classification and postprocessing, is 0.08 msec. Therefore, the average time spent in the testing phase including preprocessing, feature extraction, classification, and postprocessing is 10+0.08=10.08 msec. However, the postprocessing stage requires 9 seconds before and after each segment. Hence, this will induce a buffering delay and require additional memory space. Moreover, the training which is an offline process will be performed only once for each patient. For one-hour training data, it only takes about 1.78+ (0.01*3600) = 37.78 seconds, which is not a significant time.

V. CONCLUSIONS

Due to the nonlinear nature of the neuronal systems, theories of Nonlinear Dynamics have become an indispensable tool in the analysis of epileptic seizures. Accordingly, in this study, we proposed a novel method that characterizes the dynamic behavior of seizure events by Nullcline analysis in the phase space. To reveal the underlying foundations, we first performed a simulation study to evaluate the proposed method for detecting nullclines in the reconstructed phase space in the absence of the actual differential equation that characterizes the

chaotic system. The results demonstrated that the proposed method can accurately follow the nullcline definition such as the convergence of the detected nullclines near the equilibrium points.

The proposed seizure detection system is then evaluated over the benchmark CHB-MIT Scalp long-term EEG dataset. We experimentally verified that the irregularities of nullclines' locations increase during seizure events. The results over the CHB-MIT database show that the proposed approach achieved all the objectives of the study as stated in Section 1 using a single Nullclines feature. With only 25% training rate, the proposed approach achieved an average classification performance of 91.15% sensitivity and 95.16% specificity over the test data of each EEG record. Comparative evaluations against the competing method exhibit the superiority of the proposed method, which achieved more than 6% higher average sensitivity rate (95.22%) over the same test data. Despite the limited training, such seizure detection performance demonstrates the efficiency and the discrimination capability of the proposed approach. Another observation worth mentioning is that the overall sensitivity rate is above 90% for all patient records except for four patients (patients 6, 8, 22 and 24). While for the last two patients, 22 and 24, the average sensitivity is still reasonably high (i.e., > 85%); for patients 6 and 8, however, the sensitivity rates were too low for practical clinical usage. The detailed investigation revealed that for these patients the seizure EEG signal characteristics exhibit high variations. Therefore, the training set formed only from the early 25% of the EEG record cannot capture the seizure and non-seizure characteristics sufficiently well. It is evident that for those patients, a larger training dataset is needed. Finally, another crucial impact of the proposed approach is the ultimate reliability and robustness levels achieved since it significantly reduced (by more than 2.5 times) the false alarms in seizure detection compared to earlier work. With an elegant computational efficiency, the proposed approach can thus be a practical solution for automatic seizure detection in long EEG recordings.

APPENDIX

To demonstrate the feasibility of the proposed method for estimating the nullcline points in the reconstructed phase space, we evaluated the proposed method on five well-known 3-dimensional differential equations in Chaos theory; Lorenz, Rabinovich–Fabrikant, Rössler, Chua's circuit, and an alternation of Genesio system. These systems have been chosen due to their strong nonlinear nature and relevant chaotic characteristics. The definition and equilibrium points of each system are presented as follows:

1) Lorenz system [53] is a simplified mathematical model of atmospheric convection:

$$\begin{cases} \frac{dx}{dt} = \sigma(y-x) \\ \frac{dy}{dt} = x(\rho-z)-y \\ \frac{dz}{dt} = xy - \beta z \end{cases} \tag{5}$$

where its equilibria are located at

$$\begin{cases} (0,0,0) \\ (\pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho-1) \end{cases} \tag{6}$$

2) Rabinovich–Fabrikant system [54] consists of three coupled ordinary differential equations, which exhibit chaotic behavior for certain parameters.

$$\begin{cases} \frac{dx}{dt} = y(z-1+x^2)+\gamma x \\ \frac{dy}{dt} = x(3z+1-x^2)+\gamma y \\ \frac{dz}{dt} = -2z(\alpha+xy) \end{cases} \tag{7}$$

and its equilibria are located at

$$\begin{cases} (0,0,0) \\ \left(\pm\sqrt{\frac{\alpha R_1+2\alpha}{4\alpha-3\gamma}}, \pm\sqrt{\alpha\frac{4\alpha-3\alpha}{R_1+2}}, \frac{\gamma R_1+R_2}{(4\alpha-3\gamma)R_1+8\alpha-6\gamma}\right) \\ \left(\pm\sqrt{\frac{\alpha R_1-2\alpha}{-4\alpha+3\gamma}}, \pm\sqrt{\alpha\frac{4\alpha-3\alpha}{-R_1+2}}, \frac{\gamma R_1-R_2}{(4\alpha-3\gamma)R_1-8\alpha+6\gamma}\right) \end{cases} \tag{8}$$

where $R_1 = \sqrt{3\gamma^2-4\alpha\gamma+4}$ and $R_2 = 4\gamma\alpha^2-7\alpha\gamma^2+3\gamma^3+2\gamma$

3) Rössler system [11] is described by three non-linear ordinary differential equations defining a continuous-time dynamic system that exhibits chaotic dynamics associated with the fractal properties of the attractor. Note that two of the equations are linear and one is nonlinear.

$$\begin{cases} \frac{dx}{dt} = -y-z \\ \frac{dy}{dt} = x+ay \\ \frac{dz}{dt} = b+z(x-c) \end{cases} \tag{9}$$

where its equilibria are located at

$$\begin{cases} (R_1a, -R_1, R_1) \\ (R_2a, -R_2, R_2) \end{cases} \tag{10}$$

and $R_1 = \frac{c+\sqrt{c^2-4ab}}{2a}$ and $R_2 = \frac{c-\sqrt{c^2-4ab}}{2a}$.

4) Chua's circuit [55] is an electronic circuit that produces an oscillating waveform that never "repeats" and exhibits chaotic behavior.

$$\begin{cases} \frac{dx}{dt} = a(y-x-f(x)) \\ \frac{dy}{dt} = x-y+z \\ \frac{dz}{dt} = -by \end{cases} \tag{11}$$

where $f(x) = cx + 0.5(d-c)(|x+1|-|x-1|)$ and its equilibria are located at

$$\begin{cases} (0,0,0) \\ \left(\frac{c-d}{c+1}, 0, -\frac{c-d}{c+1}\right) \\ \left(-\frac{c-d}{c+1}, 0, \frac{c-d}{c+1}\right) \end{cases} \tag{12}$$

5) Modified Genesio system: the dynamic systems considered next is derived from Genesio system [56]

$$\begin{cases} \frac{dx}{dt} = y \\ \frac{dy}{dt} = z \\ \frac{dz}{dt} = -\rho x - qy - z + y^2 - xy \end{cases} \tag{13}$$

with only one equilibrium point at (0,0,0).

To simulate the case where the differential equation is not available, the values of the x variable are used as measurements

of the system. In this way, the obtained information using the proposed method can be validated with the information derived from the original (known) differential equation. In Table IV the parameters and initial values used in this study are listed.

In Fig. 5-9, the original phase space of the 5 systems with their true equilibrium points (left) and the estimated nullcline points in the reconstructed phase space (right) are shown.

TABLE IV. The simulation parameters and initial values

System	Parameters	Initial value (x_0, y_0, z_0)	Dimension and time lag
Lorenz	$\rho = 28$ $\sigma = 10$ $\beta = 8/3$	$(0.001, 0.001, 0.001)$	$m = 3$ $\tau = 5$
Rabinovich– Fabrikant	$\gamma = 0.1$ $\alpha = 0.288$	$(0.05, -0.05, 0.3)$	$m = 3$ $\tau = 32$
Rössler	$a = 0.2$ $b = 0.2$ $c = 5.7$	$(0.1, 0.2, 0.3)$	$m = 3$ $\tau = 150$
Chua's circuit	$a = 15$ $b = 25.58$ $c = -\dfrac{5}{7}$ $d = -8/7$	$(-1.6, 0, 1.6)$	$m = 3$ $\tau = 12$
Modified Genesio	$p = 0.8$ $q = 0.2$	$(0.1, 0.1, 0.1)$	$m = 3$ $\tau = 30$
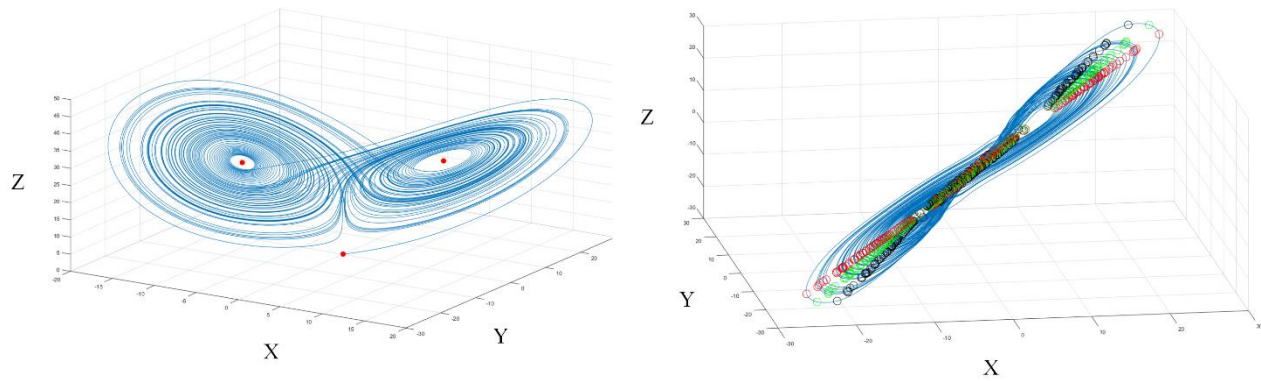
Fig. 5. The nullcline analysis of the Lorenz system. Left: the original attractor and the equilibrium points with red circles, Right: the reconstructed phase space and the estimated nullclines by red (x-nullcline), green (y-nullcline), and black (z-nullcline) circles.
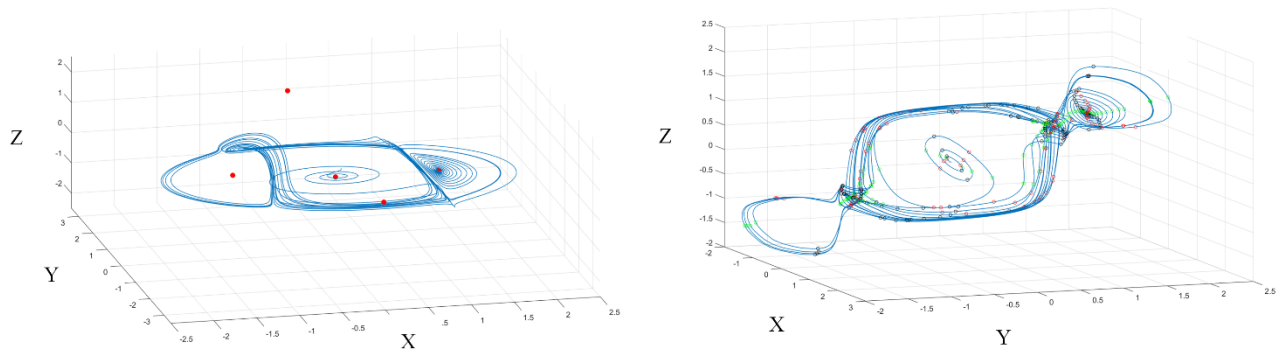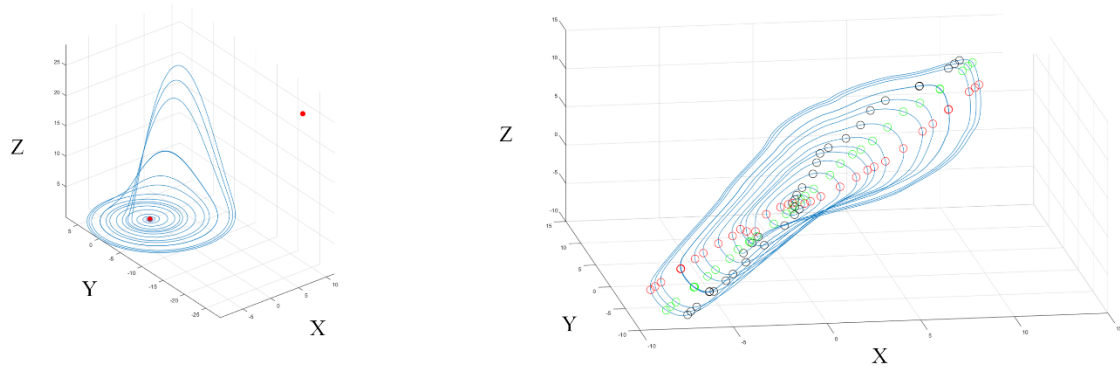

Fig. 6. The nullcline analysis of Rabinovich–Fabrikant system. Left: the original attractor and the equilibrium points with red circles, Right: the reconstructed phase space and the estimated nullclines by red (x-nullcline), green (y-nullcline), and black (z-nullcline) circles.

Fig. 7. The nullcline analysis of the Rössler system. Left: the original attractor and the equilibrium points with red circles, Right: the reconstructed phase space and the estimated nullclines by red (x-nullcline), green (y-nullcline), and black (z-nullcline) circles
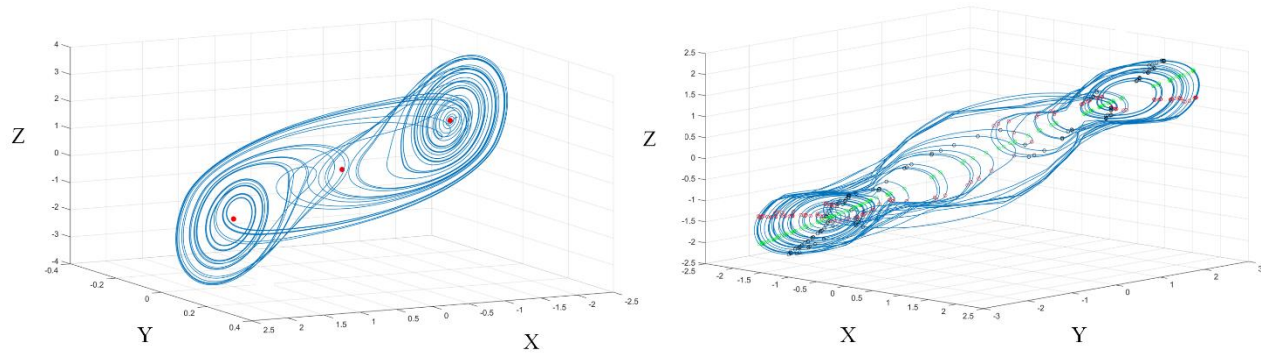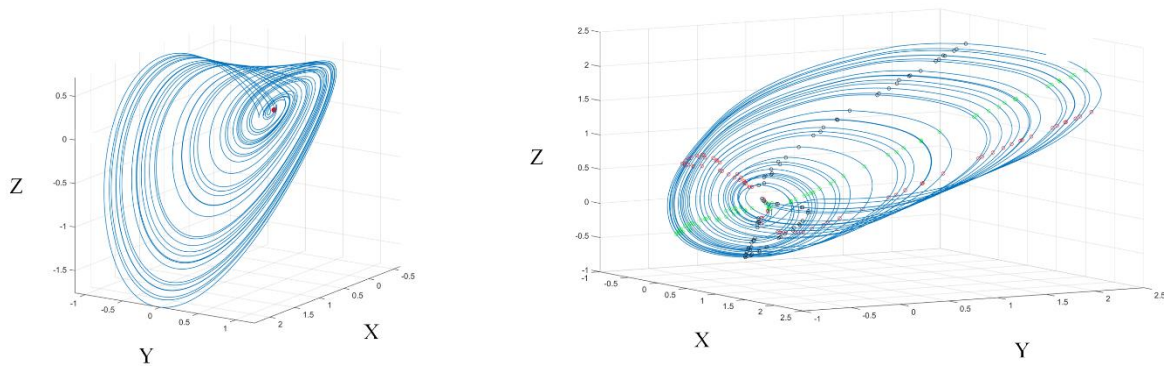

Fig. 8. The nullcline analysis of Chua's circuit. Left: the original attractor and the equilibrium points with red circles, Right: the reconstructed phase space and the estimated nullclines by red (x-nullcline), green (y-nullcline), and black (z-nullcline) circles

Fig. 9. The nullcline analysis of the attractor, which is derived from Genesio system. Left: the original attractor and the equilibrium points with red circles, Right: the reconstructed phase space and the estimated nullclines by red (x-nullcline), green (y-nullcline), and black (z-nullcline) circles

REFERENCES

[1] J. Milton, and P. Jung, Epilepsy as a Dynamic Disease, Springer Science & Business Media, 2013.

[2] N. Chakravarthy, K. Tsakalis, S. Sabesan, and L. Iasemidis, "Homeostasis of Brain Dynamics in Epilepsy: A Feedback Control Systems Perspective of Seizures," *Annals of Biomedical Engineering,* vol. 37, no. 3, p. 565–585, 2009.

[3] J. J. HOPFIELD, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences,* vol. 81, no. 10, pp. 3088-3092, 1984.

[4] D. Sussillo, and L.F. Abbott, "Generating coherent patterns of activity from chaotic neural networks," *Neuron,* vol. 63, no. 4, pp. 544-557, 2009.

[5] R. Laje, D. V. Buonomano, "Robust timing and motor patterns by taming chaos in recurrent neural networks," *Nature Neuroscience,* vol. 19, no. 7, p. 925–933, 2013.

[6] W. J. Freeman, "Mechanism and significance of global coherence in scalp EEG," *Current Opinion in Neurobiology,* vol. 31, pp. 199-205, 2015.

[7] V.I. Arnold, R. A. Silverman, Ordinary Differential Equations, Cambridge: MIT Press, 1973.

[8] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics,* vol. 898, pp. 366-381, 1981.

[9] V. K. Jirsa, W. C. Stacey, P. P. Quilichini, Anton I. Ivanov, and C. Bernard, "On the nature of seizure dynamics," *Brain,* vol. 8, p. 2210–2230, 2014.

[10] H. Kantz, and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, 2010.

[11] R. Barrio, F. Blesa, and S. Serrano, "Qualitative analysis of the Rössler equations: Bifurcations of limit cycles and chaotic attractors," *Physica D: Nonlinear Phenomena,* vol. 238, no. 13, p. 1087–1100, 2009.

[12] J. Cremers, and A. Hübler, "Construction of Differential Equations from Experimental Data," *Zeitschrift für Naturforshung A,* vol. 42, pp. 797-802, 1987.

[13] V. Gorodetskyi, and M. Osadchuk, "Analytic reconstruction of some dynamical systems," *Physics Letters A,* vol. 377, pp. 703-713, 2013.

[14] F. H. Lopes Da Silva, W. Blanes, S N. Kalitzin, J Parra, P Suffczynski, and D. N. Velis, "Epilepsies as dynamical diseases of brain systems: Basic models of the transition between normal and epileptic activity," *Epilepsia,* vol. 44, p. 72–83, 2003.

[15] L.M. Hively, and V.A. Protopopescu, "Channel-consistent forewarning of epileptic events from scalp EEG," *IEEE Transactions on Biomedical Engineering ,* vol. 50, no. 5, pp. 584 - 593, 2003.

[16] M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M Gabbouj, and T. Ince, "Analysis of High-Dimensional Phase Space via Poincaré Section for Patient-Specific Seizure Detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 24, no. 3, pp. 386 - 398, 2016.

[17] S. Ge, Q. Yang, R. Wang, P. Lin, J. Gao, Y. Leng, Y. Yang, H. Wang, "A Brain-Computer Interface Based on a Few-Channel EEG-fNIRS Bimodal System," *IEEE Access,* vol. 5, pp. 208 - 218, 2017.

[18] S. Nesaei, and A. R. Sharafat, "Real-time mining of epileptic seizure precursors via nonlinear mapping and dissimilarity features," *IET Signal Processing,* vol. 9, no. 3, pp. 193 - 200, 2015.

[19] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy," *IEEE transactions on biomedical engineering,* vol. 54, no. 2, pp. 205 - 211, 2007.

[20] S. Kiranyaz, T. Ince, M. Zabihi, D. Ince, "Automated patient-specific classification of long-term Electroencephalography," *Journal of Biomedical Informatics,* vol. 49, pp. 16-31, 2014.

[21] N. Rafiuddin, Y. Uzzaman Khan, and O. Farooq, "Feature Extraction and Classification of EEG for Automatic Seizure Detection," in *International Conference on Multimedia, Signal Processing and Communication Technologies*, Aligarh, 2011.

[22] E. Simonyi, "The Dynamics of the Polymerization Processes," Periodica Polytechnica Electrical Engineering (Archives), Zagreb, 1967.

[23] A. L. Goldberger, L. 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, p. e215–e220, 2000.

[24] A. Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[25] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation,* vol. 4, no. 3, pp. 415-447, 1992.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Artificial Intelligence Research,* vol. 16, no. 1, pp. 321-357, 2002.

[27] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Medicine and Biology Soc. (EMBC)*, Chicago, IL, USA, 2014.

[28] Z. GU, G. Yan, J. Zhang, Y. Li, Z. Yu, "Automatic Epilepsy Detection Based on Wavelets Constructed From Data," *IEEE Access,* vol. 6, pp. 53133 - 53140, 2018.

[29] S. Khanmohammadi and C. A. Chou, "Adaptive Seizure Onset Detection Framework Using a Hybrid PCA-CSP Approach," *IEEE Journal of Biomedical and Health Informatics,* no. 99, 2017.

[30] T. Guan, X. Zeng, L. Huang, T. Guan, and M. Seok, "Neural network based seizure detection system using raw EEG data," in *International SoC Design Conference (ISOCC)*, Jeju, South Korea, 2016.

[31] S. Ammar, and M. Senouci, "Seizure detection with single-channel EEG using Extreme Learning Machine," in *17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Sousse, Tunisia, 2016.

[32] C. Park, G. Choi, J. Kim, S. Kim, T-J. Kim, K. Min, K-Y Jung, and J. Chong, "Epileptic seizure detection for multi-channel EEG with deep convolutional neural network," in *International Conference on Electronics, Information, and Communication (ICEIC)*, Honolulu, 2018.

[33] P. Prathap, and T. A. Devi, "EEG spectral feature based seizure prediction using an efficient sparse classifier," in *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Kannur, India, 2017.

[34] M. Solija, S. Saleem, K. Khurshid, S. A. Hassan, and A. M. Kamboh, "Dynamic Mode Decomposition Based Epileptic Seizure Detection from Scalp EEG," *IEEE Access,* vol. 6, pp. 38683 - 38692, 2018.

[35] S. Lin, Istiqomah, L-C. Wang, C-Y. Lin, and H. Chiueh, "An Ultra-low Power Smart Headband for Real-Time Epileptic Seizure Detection," *IEEE Journal of Translational Engineering in Health and Medicine,* vol. 6, 2018.

[36] A. Bhattacharyya, and R. B. Pachori, "A Multivariate Approach for Patient-Specific EEG Seizure Detection Using Empirical Wavelet Transform," *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 9, pp. 2003 - 2015, 2017.

[37] M. Pinto-Orellana, F. Cerqueira, "Patient-dependent epilepsy seizure detection using random forest classification over one-dimension transformed EEG data," *BioRxiv*, 2016, doi: https://doi.org/10.1101/070300

[38] M. Zhou, C. Tian, R. Cao, B. Wang, Y.Niu, Ting Hu, H. Guo, and J. Xiang, "Epileptic Seizure Detection Based on EEG Signals and CNN," *Frontiers in Neuroinformatics*, vol. 12, 2018, doi: 10.3389/fninf.2018.00095

[39] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A Multi-View Deep Learning Framework for EEG Seizure Detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 83-94, 2019, doi: 10.1109/JBHI.2018.2871678

[40] N.D. Truong, A.D. Nguyen, L. Kuhlmann, M.R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Integer Convolutional Neural Network for Seizure Detection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 4, pp. 849-857, 2018, doi: 10.1109/JETCAS.2018.2842761

[41] T.N. Alotaiby, F. E. Abd El-Samie, S.A. Alshebeili, K.H. Aljibreen, and E. Alkhanen, "Seizure detection with common spatial pattern and Support Vector Machines," *International Conference on Information and Communication Technology Research (ICTRC)*, Abu Dhabi, 2015, doi: 10.1109/ICTRC.2015.7156444

[42] M. Zubair Ahmad, A. Mehmood Kamboh, S. Saleem, and A. Ali Khan, "Mallat's Scattering Transform Based Anomaly Sensing for Detection of Seizures in Scalp EEG," *IEEE Access*, vol. 5, pp. 16919-16929, 207

[43] M. Kaleem, A. Guergachi, and S. Krishnan, "Patient-specific Seizure Detection in Long-term EEG Using Wavelet Decomposition," *Biomedical Signal Processing and Control*, vol. 46, pp. 157-165, 2018

[44] P. Fergus, D. Hignett, A. Jaffar Hussain, and D. Al-Jumeily, "An Advanced Machine Learning Approach to Generalised Epileptic Seizure Detection," *International Conference on Intelligent Computing*, vol. 8590, pp. 112-118, 2014

[45] L. Orosco, A. Garcés Correa, P. Diez, and E. Laciar, "Patient non-specific algorithm for seizures detection in scalp EEG," *Computers in Biology and Medicine*, vol. 71, pp. 128-134, 2016

[46] K. M.Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, D.I. Fotiadis, "A Long Short-Term Memory Deep Learning Network For the Prediction of Epileptic Seizures Using EEG Signals," *Computers in Biology and Medicine*, vol. 99, pp. 24-37, 2018

[47] Y. Tang. D. Chen, L. Wang, A.Y. Zomaya, J. Chen, H. Liu, "Bayesian tensor factorization for multi-way analysis of multi-dimensional EEG," *Neurocomputing*, vol. 318, pp. 162-174, 2018

[48] L. Orosco, A. Garcés Correa, E. Laciar, "Review: A Survey of Performance and Techniques for Automatic Epilepsy Detection," Journal of Medical and Biological Engineering, vol. 33, no. 6, pp. 526-537, 2013.

[49] A. Sharmila, "Epilepsy Detection From EEG Signals: A Review," Journal of Medical Engineering & Technology, vol. 42, no. 5, 2018, doi: 10.1080/03091902.2018.1513576

[50] C. Zhang, MAB Altaf, and J. Yoo, "Design and Implementation of an On-Chip Patient-Specific Closed-Loop Seizure Onset and Termination Detection System, *" IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 996 - 1007, 2016

[51] K. H. Lee, and N. Verma, "A Low-Power Processor With Configurable Embedded Machine-Learning Accelerators for High-Order and Adaptive Analysis of Medical-Sensor Signals," *IEEE Journal of Solid-State Circuits,* vol. 48, no. 7, pp. 1625 – 1637, 2013

[52] H. Wang, W. Shi, CS. Choy, "Hardware Design of Real Time Epileptic Seizure Detection Based on STFT and SVM, " *IEEE Access,* vol. 6, pp. 67277 – 67290, 2018

[53] A. Algaba, M. Merino, and A.J. Rodríguez-Luis, "Superluminal periodic orbits in the Lorenz system," *Communications in Nonlinear Science and Numerical Simulation,* vol. 39, pp. 220–232, 2016.

[54] M. F. Danca, M. Feckan, N. Kuznetsov, and G. Chen, "Looking more closely to the Rabinovich-Fabrikant system," *International Journal of Bifurcation and Chaos,* vol. 26, no. 02, 2015.

[55] P. Deregel, "Chua's oscillator: A zoo of attractors," in *Chua's Circuit: A Paradigm for Chaos*, Salem, MA, World Scientific Publishing Co., 1993, pp. 179-230.

[56] Q. He, and H. Xiong, "Shilnikov chaos and Hopf bifurcation in three-dimensional differential system," *International Journal for Light and Electron Optics,* vol. 127, no. 9, p. 7648–7655, 2016.

PUBLICATION 3

Heart Sound Anomaly and Quality Detection Using Ensemble of Neural Networks without Segmentation

M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos

2016 Computing in Cardiology Conference (CinC),

Heart Sound Anomaly and Quality Detection using Ensemble of Neural Networks without Segmentation

Morteza Zabihi[1,*], Ali Bahrami Rad[2,*],
Serkan Kiranyaz[3], Moncef Gabbouj[1], Aggelos K Katsaggelos[4]

[1]Tampere University of Technology, Tampere, Finland
[2]University of Stavanger, Stavanger, Norway
[3]Qatar University, Doha, Qatar
[4]Northwestern University, Evanston, USA

Abstract

Phonocardiogram (PCG) signal is used as a diagnostic test in ambulatory monitoring in order to evaluate the heart hemodynamic status and to detect a cardiovascular disease. The objective of this study is to develop an automatic classification method for anomaly (normal vs. abnormal) and quality (good vs. bad) detection of PCG recordings without segmentation. For this purpose, a subset of 18 features is selected among 40 features based on a wrapper feature selection scheme. These features are extracted from time, frequency, and time-frequency domains without any segmentation. The selected features are fed into an ensemble of 20 feedforward neural networks for classification task. The proposed algorithm achieved the overall score of 91.50% (94.23% sensitivity and 88.76% specificity) and 85.90% (86.91% sensitivity and 84.90% specificity) on the train and unseen test datasets, respectively. The proposed method got the second best score in the PhysioNet/CinC Challenge 2016.

1. Introduction

Heart auscultation is one of the cursory and cost-effective diagnostic tests. It can provide primary evaluation of hemodynamic status and detect a cardiovascular disease, such as ventricular septal defects, and stenosis in aorta [1]. Heart sound (or phonocardiogram) can also offer additional diagnostic tests for further medical assessments.

However, the practical applications of heart sound highly depend on cognitive skills and expertise of the medical examiner. The limitation of audible frequency range, environmental noise, and variation in recording regions are other major shortcomings of this test. In order to address these shortcomings in a cost-effective diagnostic tests in ambulatory monitoring, several techniques [2-5] have been proposed for automatic analysis of heart sounds.

Phonocardiogram (PCG) signal analysis can fall into two major categories. The first type of approaches is based on temporal segmentation, i.e. identifying the cardiac cycles and localizing the position of the first (S1; beginning of the systole) and second (S2; end of the systole) primary heart sounds. The variation in the duration of S1 and S2, and their intensities are considered as the conclusive signs of cardiac anomalies.

Several studies have been conducted to PCG segmentation using different envelope extraction methods such as Shannon energy [2], Shannon entropy [3], Hilbert-Huang transform [4], and autocorrelation [5]. The envelope of signal attenuates the noise and amplifies the low-intensity components of the signal. Some segmentation approaches use envelope extraction based on wavelet transform to gain the frequency characteristics of S1 and S2 sound [6]. In the second type approaches, abnormal PCG records are detected without segmentation [7,8].

In this study we follow the second type approach towards PCG classification. The main motivation behind this is to remove the dependency on segmentation and reduce the computational burden. The main contributions of this study are the detailed investigation of time-frequency features (Section 2.1) and the design of effective neural network ensembles (Section 2.2). The proposed approach is evaluated on one of the largest public heart sound database [10]. The results are discussed in Section 3 and the Section 4 conclude the paper and suggests topics for future research.

2. Methodology

For this challenge, 3454 PCG labelled records

---

* The first two authors have contributed equally to this paper.

(including training and validation sets) are provided by Physionet/Computing in Cardiology Challenge 2016 [9]. More detailed Information can be found in [10]. The proposed feature extraction and classification approaches will be discussed next.

2.1. Feature extraction

In the initial phase of this work, 40 features in the time, frequency, and time-frequency domains were extracted. Then, a subset containing 18 features were selected using a wrapper-based feature selection scheme [11] in which sequential forward selection search algorithm [12] was used. The selected features can be categorized into 5 types as follows:

(1) Linear Predictive Coefficient (LPC): the first, third, sixth, eight, ninth, and tenth coefficients of 10th-order linear predictor are used as features.

(2) Entropy based features: Natural and Tsallis entropy of PCG signals are calculated as,

$$H(x) = -\sum_i p(x_i) \ln p(x_i) \qquad (1)$$

$$S_q(x) = \frac{k}{q-1}\left(1 - \sum_i p(x_i)^q\right) \qquad (2)$$

where $p(x_i)$ is the probability of ith samples of PCG signal, x. k and q are real parameters equal to 1 and 2, respectively.

(3) Mel Frequency Cepstral Coefficients (MFCCs) based features: The MFCCs of each PCG signal are computed based on the parameter of 14 coefficients for frame duration of 25 ms with 10 ms overlap. The extraction of MFCCs results in 14 coefficients for each frame, $C_{i,j}$, where i and j are the number of features and frames, respectively. Once the $C_{i,j}$ is calculated, three features are extracted as follows:

$$AV_{MFCC} = \frac{1}{N}\sum_{j=1}^N \min_{i \in I} C_{i,j}, \qquad j = 1,2,\ldots,N \qquad (3)$$

$$\mu_r^{max,MFCC} = E(\max_{i \in I} C_{i,j} - \mu_j)^r \qquad (4)$$

$$\mu_r^{skew,MFCC} = E\left(\underset{i \in I}{Skew}\, C_{i,j} - \mu_j\right)^r \qquad (5)$$

where $\min_{i \in I} C_{i,j}$, $\max_{i \in I} C_{i,j}$, and $\underset{i \in I}{Skew}\, C_{i,j}$ are the minimum, maximum and the skewness of each column of matrix C. μ is the average and $r = 2$.

(4) Wavelet transform based features: Discrete wavelet transform (Daubechies 4) is applied to each PCG signal and the approximation coefficients of level 5 (a_5) and the detail coefficients of level 3 to 5 (d_3,

d_4, and d_5) are used for feature extraction as follow:

$$H(a_5) = -\sum_i p(a_{5_i}) \ln p(a_{5_i}) \qquad (6)$$

$$H_q(d_5) = \frac{1}{q-1} ln(\sum_i p(d_{5_i})^q) \qquad (7)$$

$$H(d_4) = -\sum_i p(d_{4_i}) \ln p(d_{4_i}) \qquad (8)$$

$$\lambda(d_3) = log_2(\sigma^2(d_3)) \qquad (9)$$

where σ^2 is the variance. H_q is known as Rényi entropy in which $q = 2$.

(5) Features extracted over power spectral density: The power spectral density of each signal is calculated based on the normalized frequency (i.e. between 0 and 1), and then the following features are extracted as follows:

$$MPSD_{Centroid} = \frac{\int fP(f)^2 df}{\int P(f)^2\, df} \qquad (10)$$

$$AUC_1 = \int_{0.7}^{0.8} P(f)df \qquad (11)$$

$$AUC_2 = \int_{0.9}^{1} P(f)df \qquad (12)$$

where $p(f)$ represents the power spectral density and $MPSD_{Centroid}$ is the modified power spectral density centroid. AUC_1 and AUC_2 show the areas under the curve over the two specified frequency intervals, i.e. 0.7-0.8, and 0.9-1.

2.2. Classification

In this work, the proposed classification algorithm has two main steps. In the first step, the bad quality recordings (class 0) are detected. In the second step, among the good quality signals, the normal (class -1) and abnormal (class 1) PCGs are classified. This process includes two different classifiers: one for good/bad quality recordings, and the other for normal/abnormal recordings (Fig. 1). In the following the structure of these classifiers are described in details.

2.2.1. Ensemble of neural networks

Ensemble based classification systems construct a set of classifiers and then classify new samples by integrating the results of those classifiers to obtain a better classification performance. In this work, we used an ensemble of 20 feedforward Artificial Neural Networks (ANNs) with two hidden layers in each, and 25 hidden neurons at each layer.
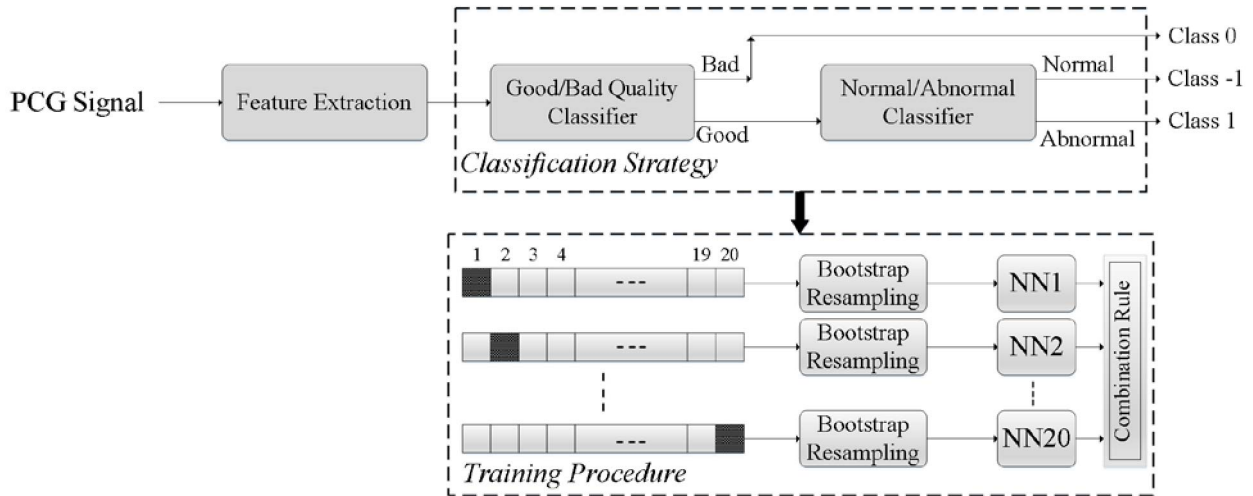
Figure 1. Schematic demonstration of classification strategy and training procedure.

The number of neurons in the output layer is 4 for the purpose of two classification tasks simultaneously, i.e., classification of the signal type (normal vs. abnormal) and quality (good vs. bad). We used the hyperbolic tangent activation function as the transfer function. In addition, we used Levenberg–Marquardt optimization method [13] with Bayesian regularization backpropagation [14] training algorithm.

2.2.2. Training

To construct the training data for the proposed ensemble of ANNs, we used 20-fold cross-validation committee [15]. To do so, we generate 20 replicates from the original training data, and then by removing 5% disjoint random subset from each replica, 20 overlapping training sets were constructed. In addition, due to the data imbalance problem between normal and abnormal signals, in each training sets we used bootstrap resampling method to make the data balanced in the following way: First, we calculate the number of normal signals (which is higher than abnormal). Then, by random sampling with replacement from abnormal signals the size of the selected set becomes equal to the size of the normal set.

Although the data get balanced by using the aforementioned technique, caution is needed. In bootstrap resampling usually the size of the selected samples is equal to or smaller than the size of the original data, but in the proposed method the size of the selected samples is larger than the original data. This would mimic a situation where we have a larger dataset than what we actually had, i.e. we will get higher precision in the bootstrap resampling than what we have in our data, which in our case leads to an overfitting problem for abnormal recordings. Our impression was that to address this drawback we could use the so-called *jackknife* resampling (i.e. random sampling

without replacement) of normal recordings instead of bootstrap resampling of abnormal recordings. However, this technique had inferior performance than the proposed technique. Thus we discarded it. Fig. 1 demonstrates both the classification strategy and the training procedure.

2.2.3 Combination rule

The last key factor for the proposed classification technique is the combination rule to integrate the results of 20 classifiers. In this work we used two approaches: 1) non-trainable rule and 2) trainable rule. In the first approach, we used unweighted average of class-specific outputs [16] of the ANNs. In the second approach, the combination is based on the voting system of the class labels which is learned during a 10-fold cross-validation scheme as follows: if at least 17 out of 20 classifiers recognize a signal as bad quality, our algorithm recognizes it as bad quality and assigns the label 0. For the remaining signals, which recognized as good quality, our algorithm decides whether it is normal or abnormal such that if at least 7 out of 20 classifiers recognize it as abnormal our algorithm detects the signal as abnormal (1) and otherwise as normal (-1).

3. Results and discussion

We have conducted experiments in order to compare the two proposed combination rules (Section 2.2.3) for PCG classification. The performance of each combination rule is evaluated using Sensitivity (Se), Specificity (Sp), and Score (Sc) based on the provided scoring mechanism of the PhysioNet/Computing in Cardiology Challenge 2016 [9,10] by running a 10-fold cross-validation procedure.

In Table 1, these results are shown. Although the

615

Table 1. The average (Ave) and standard deviation (Std) of Sensitivity (Se), Specificity (Sp), and Score (Sc) using 10-fold cross-validation procedure for the combination rules.

	Train (Rule 1)			Train (Rule 2)		
	Se (%)	Sp (%)	Sc (%)	Se (%)	Sp (%)	Sc (%)
Ave.	89.82	92.53	91.17	94.23	88.76	91.50
Std.	2.79	1.23	1.67	2.22	1.96	1.35

performance of the two rules are fairly close (91.17% vs. 91.50%), the second rule is proposed for applying on the unseen test data. The proposed solution achieved the overall score of 85.90% (86.91% Se and 84.90% Sp) on the unseen test dataset, which is the second best score in the competition.

As discussed in Section 2.1, 18 proposed features were selected using a wrapper-based feature selection scheme. In that scheme, an internal "feature selection classifier" (FS classifier) was used to detect only normal/abnormal signals. This means that the features were not selected by considering the quality detection task into account and this was in accordance by the initial scoring strategy. Only during the final stage, the organizers changed the scoring strategy and the remaining time was not sufficient to redesign the proposed system accordingly. Consequently, this decreases the final score. Thus, we decided to adapt our former method to the new strategy. In the future, for further performance improvement, two independent classification scenarios will be designed with such features that will be selected accordingly.

4. Conclusion

This study proposes a solution for anomaly and quality detection of PCG recordings without segmentation. The proposed method got the second best score in the PhysioNet/CinC Challenge 2016. Many previous methods based on PCG analysis are relied on segmentation which potentially increases the computational burden. The achieved sensitivity (86.91%) and specificity (84.90%) on the unseen test dataset demonstrate the potential of improvement in the future. Designing specific features and additional classifier for quality detection may increase the system performance.

References

[1] McConnell ME, Pediatric Heart Sounds, Springer Science & Business Media, 2008.
[2] Balili CC, Sobrepena M, Naval PC, Classification of heart sounds using discrete and continuous wavelet transform and random forests. In 3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, 2015.
[3] Moukadem A, Dieterlen A, Brandt C. Shannon Entropy based on the S-Transform Spectrogram applied on the classification of heart sounds. In IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013.
[4] Zhang D, He J, Jiang Y, Du M. Analysis and classification of heart sounds with mechanical prosthetic heart valves based on Hilbert-Huang transform. International Journal of Cardiology 2011;151(1):126–127.
[5] Kao WC, Wei CC. Automatic phonocardiograph signal analysis for detecting heart valve disorders. Expert Systems with Applications 2011;38(6):6458–68.
[6] Huiying L, Sakari L, Iiro H. A heart sound segmentation algorithm using wavelet. In 19th international Conference IEEE/EMBS, Chicago, IL, 1997.
[7] Yuenyong S, Nishihara A, Kongprawechnon W, Tungpimolrut K. A framework for automatic heart sound analysis without segmentation. In BioMedical Engineering OnLine, 2011; 10-13.
[8] Deng SW, Han JQ. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. Future Generation Computer Systems 2016;60:13–21.
[9] Physionet. Physionet Challenge 2016. https://physionet.org/challenge/2016/. Accessed 8 September 2016.
[10] Liu C, Springer D, Li Q, et. al. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(11).
[11] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell1997;97:273–324.
[12] Whitney AW. A direct method of nonparametric measurement selection. IEEETrans Comput 1971;20:1100–3.
[13] Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE transactions on Neural Networks 1994;5(6):989-93.
[14] MacKay DJC. Bayesian interpolation. Neural Computation 1992;4(3):415–447.
[15] Parmanto B, Munro PW, Doyle HR. Improving committee diagnosis with resampling techniques. Advances in neural information processing systems 1996:882-8.
[16] Polikar R. Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE 2006;6(3):21–45.

Address for correspondence:

Morteza Zabihi, Ali Bahrami Rad
P.O. Box 527, FI-33101 Tampere, Finland.
P.O. Box 8600 Forus, N-4036 Stavanger, Norway.
morteza.zabihi@tut.fi, abahramir@gamil.com.

PUBLICATION 4

Detection of Atrial Fibrillation in ECG Hand-held Devices using a Random Forest Classifier

M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj

2017 Computing in Cardiology Conference (CinC),

Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier

Morteza Zabihi[1*], Ali Bahrami Rad[2*], Aggelos K. Katsaggelos[3],
Serkan Kiranyaz[4], Susanna Narkilahti[2], and Moncef Gabbouj[1]

[1]Tampere University of Technology, Tampere, Finland
[2]University of Tampere, Tampere, Finland
[3]Northwestern University, Evanston, Illinois, USA
[4]Qatar University, Doha, Qatar

Abstract

Atrial Fibrillation (AF) is characterized by chaotic electrical impulses in the atria, which leads to irregular heartbeats and can develop blood clots and stroke. Therefore, early detection of AF is crucial for increasing the success rate of the treatment. This study is focused on detection of AF rhythm using hand-held ECG monitoring devices, in addition to three other classes: normal or sinus rhythm, other rhythms, and too noisy to analyze. The pipeline of the proposed method consists of three major components: preprocessing and feature extraction, feature selection, and classification. In total, 491 hand-crafted features are extracted. Then, 150 features are selected in a feature ranking procedure. The selected features are from time, frequency, time-frequency domains, and phase space reconstruction of the ECG signals. In the final stage, a random forest classifier is used to classify the selected features into one of the four aforementioned ECG classes. Using the scoring mechanism provided by PhysioNet/Computing in Cardiology (CinC) Challenge 2017, the overall score (mean±std) of 81.9±2.6% is achieved over the training dataset in 10-fold cross-validation. The proposed algorithm tied for the first place in the PhysioNet/CinC Challenge 2017 with an overall score of 82.6% (rounded to 83%) on the unseen test dataset.

1. Introduction

Atrial Fibrillation (AF) is associated with too quick or chaotic contraction of atria's muscle fibers. This can cause uncompleted blood transfer from atria to ventricles and decrease the efficiency of heart functioning. The AF global prevalence is estimated as 33.5 million in 2010 [1], and its rate is increasing based on regional studies [2]. This arrhythmia is one of the main public health problems because of not only its prevalence but also its complications and costs. Symptomatic AF patients are more probable to be diagnosed and treated, whereas asymptomatic patients (silent AF) are more prone to serious complications caused by AF such as ischemia, stroke, or early mortality [3]. Therefore, early detection of AF is crucial for effective treatment, improving the clinical outcomes, and decreasing the costs.

Based on the AF management guidelines [4], prompt ECG (at least 30s recording) is a diagnostic and effective method. The absence of significance P wave and irregular distances of QRS complexes are the main signs of AF on the ECG recordings. Therefore, to date, several studies have been conducted to automatically detect AF rhythm using signal processing and machine learning methods [5] [6]. However, only a few of them studied the ECG recorded by single-lead portable devices. Although it is shown that the hand-held devices cannot substitute a conventional ECG devices [7] [8], they can be used for daily usage and improve the accuracy of early AF detection [7].

This work proposes a hybrid classification approach for ECGs recorded by the *AliveCor* hand-held devices [9]. It combines features from multi domains including time, frequency, time-frequency, phase space, and meta-level. It utilizes a feature selection approach based on a random forest classifier. Finally, the selected features are classified by another random forest classifier. The main contributions of this study are:

1) To investigate a comprehensive set of discriminative features, which is independent of the ECG lead positioning (Section 2.1). This is crucial because there are different alternatives for lead placement in hand-held devices, e.g., the measurement between left and right hand, or directly on the chest.

2) To design an effective classification algorithm in order to classify four ECG types including the AF rhythms

---

(Section 2.2).

The proposed algorithm is evaluated over recently released single-lead ECG dataset [9]. In Sections 3 and 4, the results and conclusion are discussed.

2.   Materials and methods

For this challenge, 8528 single-lead ECG recordings with sampling frequency of 300 Hz are provided by Physionet/Computing in Cardiology Challenge 2017 [10]. The objective of the challenge is to classify each ECG recording into one of the following classes: healthy (normal), AF, other rhythms, and noisy. More detailed information can be found in [9]. The proposed feature extraction and classification approach will be presented next.

2.1.   Feature engineering

First as the preprocessing stage, the quality of the ECG recordings are enhanced based on the sparse derivative decomposition and denoising algorithm [11]. Once the ECGs are denoised and the baseline wander is removed, a set of 491 hand-crafted features is extracted. The extracted features are a combination of base-level (i.e., signal-level) features and meta-level features (i.e., the prediction of the base-level classifiers). Then, a random forest classifier ranks the features in decreasing order of importance. The importance of each feature is evaluated based on the reduction of the entropy. A subset of 150 highest-ranked features is then selected. The selected features are listed as follows:

(1)   Base-level time domain and morphological features: The 67 selected features in the time domain are: the average of RR intervals ($\overline{RR}$); the coefficient of variation of RR intervals ($CoefVar(RR)$); variance of the P wave amplitudes; mean of kurtosis values of T waves; eigenvalues of the covariance matrix of beats; the correlation coefficients and Rényi entropy [12] of P waves. Furthermore, mean, standard deviation, range, interquartile range (IQR), percentiles of energy, slope and angles of P-QRS-T waves [13], PR intervals, and R amplitudes are extracted.

(2)   Base-level frequency domain features: First, the power spectral density of each beat is estimated using Burg's method (P). Then, two features are calculated for each beat in different frequency (f) range (Hz) as follows:

$$Pf_1 = \frac{\sum_{f=5}^{15} P_f}{\sum_{f=5}^{40} P_f} \qquad (1)$$

$$Pf_2 = \frac{\sum_{f=1}^{40} P_f}{\sum_{f=0}^{40} P_f} \qquad (2)$$

The average of Pf_1 and Pf_2 in each signal are used as two features in this domain.

(3)   Base-level time-frequency domain features: In total, 46 features are selected from this domain. Shannon, Tsallis, and Rényi entropies [12] of the five levels of detail and one level of approximation coefficients obtained by Symlet 4 wavelet are used. These entropy measures are extracted separately from the whole signal, P, and T waves. In addition, the statistical and morphological features of details and approximation coefficients of seven level decomposition, obtained by Daubechies 4 wavelet, are extracted [14]. Moreover, the homogeneity of the ECG signal is defined using continuous wavelet transform (CWT):

$$Homogenity = \frac{\sum W_{i,j}}{K} \qquad (3)$$

where $W_{i,j} = P_{i,j} + |i - j|$, $P_{i,j}$ is the probability of bin (i, j) in CWT space, and K is the number of all bins.

(4)   Base-level nonlinear (phase space) features: In phase space representation, 1-D time series are embedded into higher dimensional space in order to reveal their dynamical evolution through time. In this study, we have used different embedding methods of RR series in order to characterise the different types of arrhythmia. The first feature is defined as if an ellipsoid can be fitted in the 2-D phase space with lag 1 (reconstructed based on Takens' delay embedding theory [15]). The possibility or impossibility of fitting an ellipse representing the geometrical properties of RR's dynamic. Moreover, the local temporal behaviour of the phase space points is analyzed based on co-occurrence matrix [6].

The next selected feature is defined as:

$$stepping = \frac{\frac{1}{n-2}\sum_{k=1}^{n-2}\sqrt{(I_j - I_{j+1})^2 + (I_{j+1} - I_{j+2})^2}}{\frac{1}{n}\sum_{j=1}^{n} I_j}, \qquad (4)$$

where I_j is jth point in a 2-dimensional space formed by RRs (horizontal coordinate) and $\frac{d\,RRs}{dt}$ (vertical coordinate) [16].

Furthermore, two different phase space reconstruction methods, parabolic [17] and triangle [18] mappings, are used. Parabolic mapping is formed by RR_i and $(\overline{RR} - RR_i)^2$ as the horizontal and vertical coordinates, respectively. The coefficients of a fitted second order polynomial in this space are used as descriptive features. Likewise, the perimeter and area of the triangle phase space, which is constructed by RR_i and $|(\overline{RR} - RR_i)|$, are selected as the key characteristics of this domain. In addition, AFEvidence, ATEvidence, and OrgIndex metrics [5] are used.

(5)   Meta-level features: these features are the statistical descriptors of the prediction of the base-level classifiers. In this work, we use three base-level classifiers: linear and quadratic discriminant analysis (LDA & QDA), and a random forest with 30 decision trees. These classifiers are then trained on the 20% random subset of the training data to generate the meta-level features for the next level classifier (see Fig.1). This process is discussed
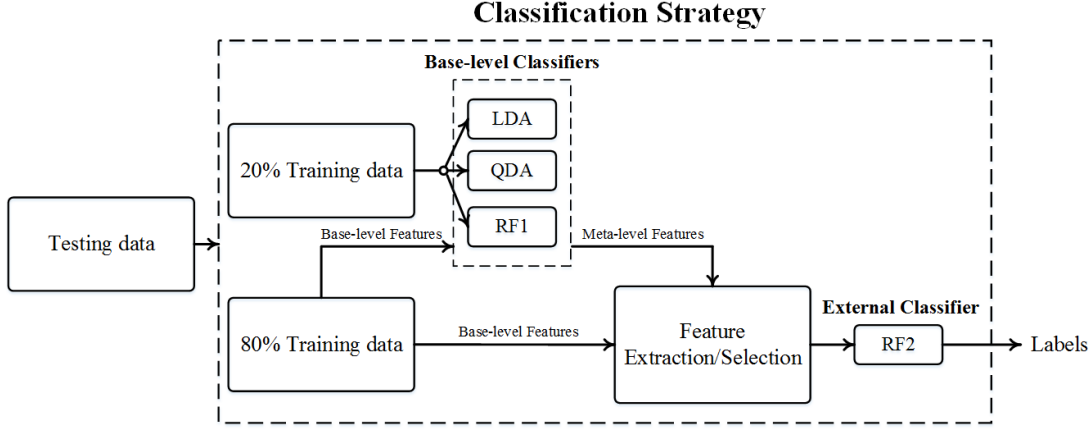
Figure 1. The proposed classification strategy

in the following.

The number of abnormal segments in an ECG signal can signify the irregularities with more resolution. For this purpose, first, the signal is windowed into 5s segments with 4s overlap. Once the R, P and T waves are detected in each segment, the following base-level features are extracted:

$$f_{seq1} = CoefVar\ (RR) \qquad (5)$$

$$f_{seq2} = mean(std(T_{wave})) \qquad (6)$$

$$f_{seq3} = \max(mean(T_{wave})) \qquad (7)$$

$$f_{seq4} = \sum mean(E_{p_{wave}}) \qquad (8)$$

$$f_{seq5} = \sum mean(E_{T_{wave}}) \qquad (9)$$

where $E_{p_{wave}}$ and $E_{T_{wave}}$ are the energy of P and T waves, respectively. In addition, each segment is modelled as an order 4 autoregressive process. The parameters of this model are used as new features (f_{seq6}-f_{seq9}). Then, all nine base-level features are fed into the three aforementioned base-level classifiers. Each classifier is used to generate four posterior probabilities of classes for each 5s segment. The mean and standard deviation of these posterior probabilities are used as meta-level features.

2.2.   Classification

In this work, we used a hybrid classification framework in the sense that we combined the base-level and meta-level features to generate the hybrid feature vectors, and then fed them into a single learning algorithm to classify. For this purpose, an (external) random forest classifier is trained over the remaining 80% of the training data by using 500 decision trees and random selection of features at each node creation. We use bagging, i.e. bootstrapped

replicas of the training data, to train each decision tree, and 30 features are randomly selected for each node. Then the entropy measure is used to decide which feature to split on at each node.

3. Results and discussion

The accuracy of the proposed method is evaluated in 10-fold cross validation manner. Because 20% of data has already been used to train the base-level classifiers, we have used the remaining 80% of the training data for evaluation in order to avoid overfitting. These results are reported in Table 1.

Table 1: Results of the proposed method: the overall score (mean±std) over 80% of the training dataset in 10-fold cross-validation and the overall score on the unseen test dataset.

Evaluation metrics	Training set (%)	Testing set (%)
F1n (Normal)	90.49 ± 0.96	90.87
F1a (AF)	79.43 ± 4.52	83.51
F1o (Other)	75.64 ± 3.11	73.41
F1p (Noisy)	61.11 ± 7.53	50.42
F1	81.85 ± 2.57	83

In hand-held devices, each ECG recording typically includes noise and artifacts, low-quality signals, intermediate rhythms, and transitional states between rhythms. The proposed algorithm in this paper only partially handles these difficulties. The sequential classification algorithms such as hidden Markov models (HMM), conditional random fields (CRFs), and recurrent neural networks (RNN) which analyze consecutive windows can be a possible solution for the aforementioned difficulties. They will be investigated in our future work.

4. Conclusions

In this paper, we have proposed a systematic approach for the detection of AF rhythms in ECG hand-held devices. We have investigated a comprehensive set of hand-crafted 491 features, and ranked them based on their importance. A set of 150 highest-ranked features is selected and fed into a random forest classifier in order to detect AF rhythms in addition to three other ECG rhythms/types. The proposed method tied for the first place in the PhysioNet/CinC Challenge 2017 with an overall score of 82.6%. With this overall performance, the proposed algorithm has a potential for improvement, which is the subject of our future work.

References

[1] S. S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E. J. Benjamin, R. F. Gillum, Y. Kim, J. H. McAnulty, Z. Zheng, M. H. Forouzanfar, M. Naghavi, G. A. Mensah, M. Ezzati, C. J.L. Murray, "Worldwide Epidemiology of Atrial Fibrillation A Global Burden of Disease 2010 Study," *Circulation,* vol. 129, no. 8, 2014.

[2] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, "Epidemiology of atrial fibrillation: European perspective," *Clinical Epidemiology,* no. 6, p. 213–220, 2014.

[3] P. E. Dilaveris, and H. L. Kennedy, "Silent atrial fibrillation: epidemiology, diagnosis, and clinical impact," *Clinical Cardiology,* vol. 40, no. 6, pp. 413-418, 2017.

[4] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H. Diener, H. Heidbuchel, J. Hendriks, G. Hindricks, A. S. Manolis, J. Oldgren, B. A. Popescu, U. Schotten, B. V. Putte, P. Vardas, "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal,* vol. 37, no. 38, pp. 2893-2962, 2016.

[5] S. Sarkar, D. Ritscher, and R. Mehra, "A Detector for a Chronic Implantable Atrial," *IEEE Transactions on Biomedical Eengineering,* vol. 55, no. 3, pp. 1219 - 1224, 2008.

[6] S. Moharreri, S. Parvaneh, N. J. Dabanloo, A. M. Nasrabadi, "Utilizing occurrence sequence of Heart Rate's phase space points in order to discriminate heart Arrhythmia," in *17th Iranian Conference of Biomedical Engineering (ICBME)*, Isfahan, Iran, 2010 .

[7] D. C. Peritz, A. Howard, M. Ciocca, E. and H.Chung,, "Smartphone ECG aids real time diagnosis of palpitations in the competitive college athlete," *Journal of Electrocardiology,* vol. 48, no. 5, pp. 896-899, 2015.

[8] K. M Griffiths, E. N Clark, B. Devine, and P. W. Macfarlane, "Assessing the accuracy of limited lead recordings for the detection of Atrial Fibrillation," in *Computing in Cardiology Conference (CinC)*, Cambridge, MA, USA, 2014.

[9] G. Clifford, C. Liu, B. Moody, L. H. Lehman, I. Silva, Q. Li, A. Johnson, and R. Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017," *Computing in Cardiology (In Press),* vol. 44, 2017.

[10] "AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017," 2017. [Online]. Available: https://physionet.org/challenge/2017/. [Accessed 7 9 2017].

[11] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems,* vol. 139, pp. 156-167, 2014.

[12] M. Zabihi , A. Bahrami Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation," in *Computing in Cardiology Conference (CinC)*, Vancouver, BC, Canada, 2016.

[13] T. Hamdi, A. B. Slimane, and A. B. Khalifa, "A novel feature extraction method in ECG biometrics," in *Image Processing, Applications and Systems Conference (IPAS)*, Sfax, Tunisia, 2014.

[14] A. Bahrami Rad, T. Eftestol, K. Engan, U. Irusta, J. T. Kvaloy, J. Kramer-Johansen, L. Wik, A. K. Katsaggelos, "ECG-based Classification of Resuscitation Cardiac Rhythms for Retrospective Data Analysis," *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 10, pp. 2411 - 2418, 2017.

[15] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics,* vol. 898, pp. 366-381, 1981.

[16] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in Poincare plot," *BioMedical Engineering OnLine,* vol. 38, no. 8, 2009.

[17] S. Moharreri, S. Rezaei, N J. Dabanloo, S. Parvaneh, "Extended Parabolic Phase Space Mapping (EPPSM): Novel quadratic function for representation of Heart Rate Variability signal," in *Computing in Cardiology Conference (CinC)*, Cambridge, MA, USA, 2014.

[18] S. Moharreri, S. Rezaei, and S Salavatian, "Discrimination of heart arrhythmias using novel features in heart rate phase space," in *Computing in Cardiology Conference (CinC)*, Zaragoza, Spain, 2013.

Address for correspondence:

Morteza Zabihi, Ali Bahrami Rad
P.O. Box 527, FI-33101 Tampere, Finland.
P.O. Box 100, FI-33014 Tampere, Finland
morteza.zabihi@tut.fi, ali.bahrami.rad@uta.fi

PUBLICATION 5

Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models

M. Zabihi, S. Kiranyaz, and M. Gabbouj

2019 Computing in Cardiology Conference (CinC),

Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models

Morteza Zabihi[1], Serkan Kiranyaz[2], Moncef Gabbouj[1]

[1]Tampere University, Tampere, Finland
[2]Qatar University, Doha, Qatar

Abstract

Sepsis is caused by the dysregulated host response to infection and potentially is the main cause of 6 million death annually. It is a highly dynamic syndrome and therefore the early prediction of sepsis plays a key role in reducing its high associated mortality. However, this is a challenging task because there is no specific and accurate test or scoring system to perform early prediction. In this paper, we present a systematic approach for sepsis prediction. We also propose a new set of features to model the missingness in clinical data. The pipeline of the proposed method comprises three major components: feature extraction, feature selection, and classification. In total, 407 features are extracted from the clinical data. Then, five different sets of features are selected using a wrapper feature selection algorithm based on XGboost. The selected features are extracted from both valid and missing clinical data. Afterwards, an ensemble model consists of five XGboost models is used for sepsis prediction. The proposed algorithm is ranked officially as third place in the PhysioNet/Computing in Cardiology Challenge 2019 with an overall utility score of 0.339 on the unseen test dataset (our team name: Separatrix).

1. Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection [1] and is often associated with lung, urinary tract, skin, and gut infections. The recent report of Center for Disease Control (CDC) shows that sepsis causes one out of every three hospital deaths [2] [3]. Besides the high mortality rate of sepsis, it imposes immense challenges to healthcare systems. From an economic perspective, sepsis implies high costs of hospital care with almost 17 billion USD annually in the United States [4] and 2.5 billion pounds in the UK [5]. Thus, early prediction of sepsis is a crucial element for appropriate clinical management and improvement of clinical outcomes.

The recent clinical criteria of sepsis [1] in the general hospital ward setting, recommend that quick Sequential (Sepsis-related) Organ Failure Assessment (qSOFA) should be used as a rapid evaluation of sepsis risk. This means that the patient should have at least two of the following clinical criteria to be considered as a patient with suspected infection: respiratory rate of 22 per minute or greater, altered mentation, and systolic blood pressure of 10 mmHg or less. Moreover, in [1], the SOFA$\geq$ 2 score is determined to represent organ dysfunction. SOFA score monitors laboratory values and vital signs such as the fraction of inspired oxygen (FiO2), the partial pressure of oxygen (PaO2), platelets, liver bilirubin, and mean arterial pressure [1]. However, sepsis is a dynamic condition, and such criteria may not meet or present in all the time. This leads to inaccurate results of such approaches [6]. In addition, using clinical criteria for sepsis diagnosis in patients with critical situations (e.g., ICU patients) can be even more challenging due to the misleading symptoms caused by other diseases [7].

Despite the slow changes in sepsis definitions, several studies have focused on the development of Machine Learning models to overcome the aforementioned challenges. In [8], the proposed method achieved significantly higher accuracy compared to the three standard sepsis-related scoring systems (i.e., SOFA, qSOFA, and MEWS). In [9], a variant recurrent neural network model is proposed for sepsis prediction. Their proposed model revealed that ICU length-of-stay, heart rate, white blood cell count, and temperature are the most relevant features for sepsis prediction. In [10], a model based on Weibull-Cox proportional hazards mode is proposed to predict the onset of sepsis in an ICU patient 4 to 12 hours prior to clinical recognition. Their method achieved the area under the receiver operating characteristic (AUROC) between 0.83–0.85. These models have achieved higher accuracy compared to traditional clinical criteria. However, further studies are needed to improve the robustness, false alarm rate, and interpretability of such models.

In this paper, we explore the use of an ensemble learning technique (Figure 1) for sepsis prediction in ICU. The main contributions of this study are:
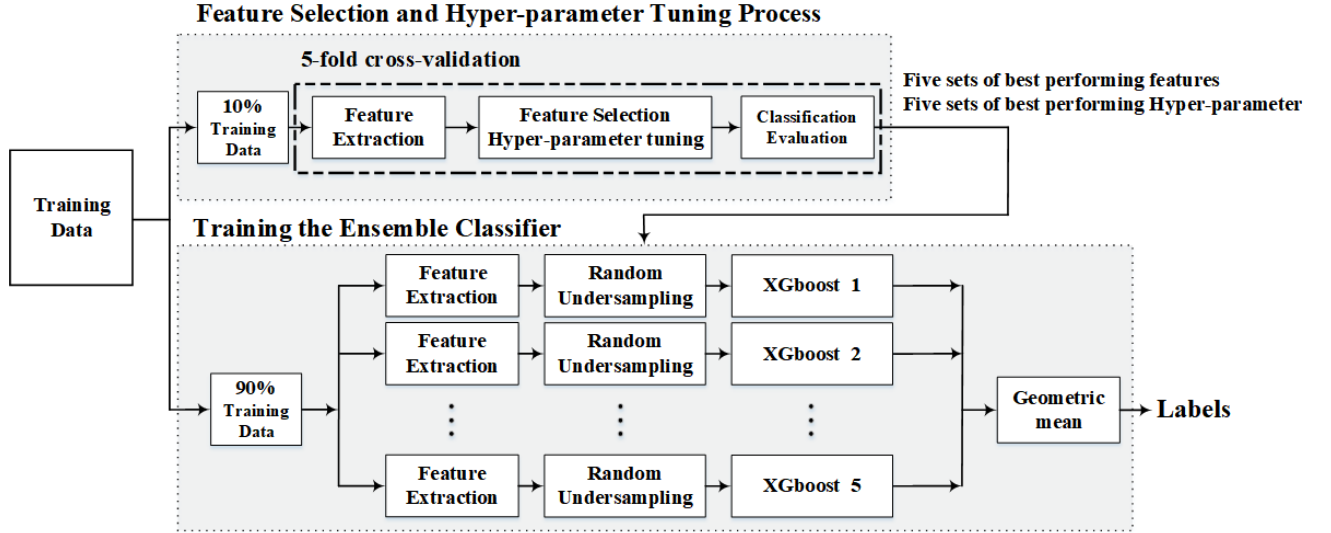
Figure 1. The training strategy of the proposed method

1) Investigating a comprehensive set of features and tracking the top clinically relevant features.
2) Introducing discriminative features for revealing the patterns of missing values in clinical data.
3) Designing a predictive model by ensembling 5 classifiers.

The remainder of this paper is organized as follows: In Section 2, the dataset is briefly described and the proposed method is explained. In Section 3, the evaluation results are presented and discussed. Finally, concluding remarks are outlined in Section 4.

2. Materials and methods

The dataset used in this competition is collected from 63097 ICU patients in three distinct hospitals. The training set includes 40336 records from two hospitals (hospitals *A* and *B*), while the remaining 22761 patient records (from hospitals *A*, *B*, and *C*) are kept hidden to be used for final ranking. For each patient, eight vital signs, six demographics variables, and 26 laboratory values are provided for every hour. More detailed information can be found in [11]. The feature extraction, feature selection, and classification approach are described next.

2.1. Feature engineering

Often the clinical data are not collected consistently. Therefore, it is expected that the majority amount of data for some covariates is missing. It has been shown that the imputation of missing values for such covariates does not significantly improve the prediction performance [12]. On the other hand, the missingness may convey useful information [13]. Therefore, in this work, two different types of features are extracted. The first type of feature targets the covariates with less than 70% of missingness

while the second type of feature focuses on the patterns of missing values in clinical data. The combination of these features forms a set of 407 features in total (see Table 1). Once the features are extracted, they are normalized to a mean of 0 and unit standard deviation. The extracted features are described as follows:

The first type of features are extracted from 13 covariates of heart rate (HR), pulse oximetry (O2Sat), temperature (Temp), systolic blood pressure (SBP), mean arterial pressure (MAP), diastolic blood pressure (DBP), respiration rate (Resp), age, gender, administrative identifier for MICU unit (Unit1), administrative identifier for SICU unit (Unit2), hours between hospital admit and ICU admit (HospAdmTime), and ICU length-of-stay (ICULOS). Before extracting the first type of features, the missing value imputation is carried out by linear interpolation. If the data is less than 3 hours (i.e., less than three observations), then the missing values are replaced with the mean value of the corresponding covariate in the training data. For age and gender, the missing values are replaced by the first valid value. If all the values in the given observations were missing, then they are replaced by mean values. Once the imputation is performed, the following features are extracted:

(1) Sliding-window based features: Mean, minimum, maximum, median, variance, 95%, 99%, 5%, and 1% quantiles are calculated from the last 5 and 11 hours observations. We use two different time windows (i.e., 5 and 11 hours) to capture the short- and long-term temporal evolution of covariates.

(2) Non sliding-window based features: Energy, Shannon entropy, mean of the first differences, and the lengths of observations are calculated from the given observations.

(3) The last observation values of the 13 covariates are also used as 13 separate features.

Table 1. List of the extracted features

Type	Features	#features	
1 *Imputation, 13 covariates*	*Mean*, *minimum*, *maximum*, *median*, *variance*, *95%*, *99%*, *5%*, and *1% quantiles* from the last 5 and 11 hours.	198	245
	Energy, *Shannon entropy*, *mean of the first differences*, and *the lengths of observations*	34	
	last observation values of the 13 covariates	13	
2 *No imputation, 38 covariates*	*Mean* and *variance* of L_C	76	162
	Summation and *variance* of L_{CV}	76	
	Mean and *variance* of L_O	10	

To calculate the second type of features, age and gender are excluded from the given covariates. These two demographic variables are constant for each patient during the monitoring and therefore their absence does not convey any information. To represent the missingness, we define the *sequence* abstraction. Each *sequence* is defined as a set of consecutive measurements where the values are only either missing or present. Therefore, each *sequence* can only have missing or present values. For instance, let's imagine the SBP measurements for 6 hours are $\{nan, 122, 98, nan, nan, 123\}$, then based on the definition, we have 4 *sequences* of $\{nan\}$, $\{122, 98\}$, $\{nan, nan\}$, and $\{123\}$. Using the *sequence* abstraction the following features are calculated (see Figure 2):

(1) Mean and variance of the lengths of *sequences* along each covariate, L_C.

(2) Summation and variance of the lengths of *sequences* with only valid values (without missing) along each covariate, L_{CV}.

(3) Mean and variance of the lengths of *sequences* along each observation, L_O, in the last 5 hours.

It is worth mentioning that the input clinical data have varying lengths, and it is possible that the number of observations is not enough to extract the sliding-window based features. For such cases, the clinical data is padded using the first observation. The amount of padding equals to the difference between the number of observations in the given data and the number of needed ones. This enables us to transform the raw data into a feature space with a fixed length. Thus, discriminative methods, such as XGboost and random forest, can be applied to such dynamic data.

2.2. Feature selection and classification

The proposed classification algorithm consists of two main steps:

(1) In the first step, five sets of best performing features and hyper-parameters are selected. We perform the feature selection and hyper-parameter tuning in a 5-fold cross-validation scheme using 10% of the original training data. For feature selection, we employ a wrapper feature selection algorithm based on XGboost (BoostARoota [14]). The importance metric is the number of times that a particular feature was split on in the XGboost algorithm. In addition, a grid search is used to find the best performing combinations of hyper-parameters.

(2) In the second step, we used an ensemble of five XGboost models. XGboost is a decision tree based ensemble using a gradient boosting framework [15] and its effectiveness has been established in a wide range of applications especially in prediction problems. To train the proposed ensemble, we randomly split the remaining 90% of the original data into five equally disjoint sets. Then, each set is used to train a distinct classifier. Moreover, due to the imbalance problem between sepsis and non-sepsis observations, we separately balance the data for each XGboost using the random undersampling technique. Finally, we use the geometric mean to integrate the outputs of the five classifiers. The training strategy of the proposed method is shown in Figure 1.

3. Results and discussion

We test our predictive model in a 5-fold cross-validation scheme using the training data. The results are reported in Table 2 (for more information about the score and metrics refer to [11]). The obtained utility scores (AUROC, AUPRC, F-measure) on the unseen test set A, B, and C are 0.422 (0.814, 0.102, 0.128), 0.395 (0.844, 0.110, 0.130), and -0.146 (0.793, 0.058, 0.044), respectively. Clearly, the performance of the proposed model on hospitals A and B (which are present in the training set) are robust with respect to our cross-validation. However, the performance drops drastically on the test set C. We believe that the main reason is that the missingness in hospital C has a different

Observation	HR	Temp	SBP	L_O
1	NaN	NAN	NaN	{3}
2	97	NaN	98	{1, 1, 1}
3	89	NaN	122	{1, 1, 1}
	90	NaN	NaN	{1, 2}
	103	NaN	122	{1, 1, 1}
	110	NaN	NaN	{1, 2}
.	108	36.11	123	{3}
.	106	NaN	93	{1 1, 1}
	104	NaN	133	{1, 1, 1}
10	102	NaN	134	{1, 1, 1}
L_C	{1, 9}	{6, 1, 3}	{1, 2, 1, 1, 4}	
L_{CV}	{9}	{1}	{2, 1, 4}	

Figure 2. Sequence abstraction for HR, Temp and SBP covariates

Table 2. The results of the proposed method on the training data in a 5-fold cross-validation scheme and on the hidden test set. AUROC and ACC are area under the receiver operating characteristic and accuracy, respectively.

Fold	AUROC	ACC	Score
0	0.8387	0.8394	0.4366
1	0.8357	0.8418	0.4412
2	0.8436	0.8477	0.4521
3	0.8221	0.8451	0.3899
4	0.8268	0.8464	0.4208
Average	0.8333	0.8440	0.4281
(std)	(0.0078)	(0.0030)	(0.0215)
The hidden test data			0.339

pattern compared to other hospitals. Here, missingness represents human behavior in recording the covariates and does not convey medical information. Therefore, missingness should be used with caution. That said, it should be noted that all the contestants fail to achieve a high score on test set C even if they have not used missingness information in their proposed methods.

Additionally, we observe that among the second type features (missingness) 102 out of 162 features were selected commonly using the BoostARoota algorithm. This shows the significance of the proposed features in sepsis prediction. Moreover, among the selected features, the HospAdmTime, the summation of L_{CV} for TMP, age, Unit1, variance of HR and TMP in the last 11 hours were ranked among the top 10 features.

4.   Conclusions

In this work, we proposed a systematic approach for sepsis prediction in ICU. We investigate a set of features to capture the transitional states of covariates by using two time windows with different lengths. In addition, we introduce a new set of features to represent the missingness in clinical data. We examined the importance of features using the BoostARoota algorithm and found that the missing data convey relevant information for sepsis prediction in two out of three hospitals. The proposed method is officially ranked as the third team with a utility score of 0.339 on the unseen data (our team name: *Separatrix*).

References

[1]   M. Singer, C. S. Deutschman, C. W. Seymour, et al., "The third international consensus definitions for sepsis and septic shock (Sepsis-3)," *JAMA,* vol. 315, no. 8, pp. 801-810, 2016.

[2]   "Making health care safer: Think sepsis," *Center for Disease Control and Prevention (CDC)*, 2016.

[3]   J. Hajj, N. Blaine, J. Salavaci, D. Jacoby, "The centrality of sepsis: A review on incidence, mortality, and cost of care," *Healthcare (Basel)*, vol. 6, no. 3, 2018.

[4]   DC. Angus, WT. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, MR Pinsky, "Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care," *Crit Care Med,* vol. 29, no. 7, pp. 1303-10, 2001.

[5]   M. Sanderson, M. Chikhani, E. Blyth, S. Wood, L. K. Moppett, T. McKeever, M. JR Simmonds, "Predicting 30-day mortality in patients with sepsis: An exploratory analysis of process of care and patient characteristics," *Journal of the Intensive Care Society,* vol. 19, no. 4, p. 299–304, 2018.

[6]   Z. Zhang, NJ. Smischney, H. Zhang, S. Van Poucke, et. al., "AME evidence series 001-The society for translational medicine: Clinical practice guidelines for diagnosis and early identification of sepsis in the hospital," *Journal of Thoracic Disease,* vol. 8, no. 9, pp. 2654-2665, 2016.

[7]   J. L. Vincent, "The clinical challenge of sepsis identification and monitoring," *PLoS Medicine,* vol. 13, no. 5, 2016.

[8]   A. Mitra, K. Ashraf, "Sepsis prediction and vital signs ranking in intensive care unit patients," *arXiv,* 2018.

[9]   S. P. Shashikumar, C. Josef, A. Sharma, S. Nemati, "DeepAISE--An end-to-end development and deployment of a recurrent neural survival model for early prediction of sepsis," *arxiv,* 2019.

[10]   S. Nemati, A. Holder, F. Razmi, M. Stanley, G. Clifford, T. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the ICU," *Critical Care Medicine,* vol. 46, no. 4, pp. 547-553, 2018.

[11]   MA. Reyna, C. Josef, R. Jeter, SP. Shashikumar, MB. M. Brandon Westover, S. Nemati, GD. Clifford, A. Sharma, "Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019," *Critical Care Medicine,* 2019, In press.

[12]   N. Razavian, D. Sontag, "Temporal convolutional neural networks for diagnosis from lab tests," *arXiv,* 2015.

[13]   J. H. Lin, P. J. Haug, "Exploiting missing clinical data in Bayesian network modeling for predicting medical problems," *Journal of Biomedical Informatics,* vol. 41, no. 1, pp. 1-14, 2018.

[14]   C.   DeHan,   "BoostARoota," https://github.com/chasedehan/BoostARoota, 2017.

[15]   T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system," in *International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2016.

Address for correspondence:

Morteza Zabihi
P.O. Box 553, FI-33014, Tampere, Finland
Morteza.zabihi@tuni.fi