Diogo Filipe Pereira Fontes Fernandes Silva

Bachelor of Science in Biomedical Engineering

Combining Synthesis of Cardiorespiratory Signals and Artifacts with Deep Learning for Robust Vital Sign Estimation

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

Adviser: Christoph Hoog Antink, Assistant Professor, RWTH Aachen University

Co-adviser: Pedro Vieira, Assistant Professor, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Examination Committee

Chair: Susana Sério
Rapporteurs: André Wemans
Members: Pedro Vieira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

December, 2019

Combining Synthesis of Cardiorespiratory Signals and Artifacts with Deep Learning for Robust Vital Sign Estimation

This document was created using the (pdf)LaTeX processor, based on the "novathesis" template[1], developed at the Dep. Informática of FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

Acknowledgements

There could be no way to start without expressing my sincere gratitude for my adviser, Christoph Hoog Antink, who was, on all accounts, a mentor as well. He ever so kindly received and took me under his wing throughout this campaign, with genuine concern and interest for both my work and welfare. His self-assured demeanor and way of tackling scientific problems so clearly and elegantly are the true signs of talent and maturity. Still, he always carries himself with humility and shows nothing but empathy and respect to the next. Such a rare set of qualities left an indelible mark which will always shape me as an engineer and a person.

I was equally well received at MedIT in the Helmholtz Institute by its team, of which I would like to especially thank Professor Steffen Leonhardt for approving my stay and swiftly taking care of its inherent bureaucratic matters, despite his incredibly busy schedule. Thank you to Theo Thomas and Christoph Weiss for their kindness and technical support with servers and installations. Naturally, I would also like to address a special thanks to my colleagues Paul Rohwein, Arnold Sterle, Stephan Eisenbrand, Nadine Buczak, Alissa Wenzel, Arnhold Lo and Onno Linschmann for all the conversations, fun and companionship inside and outside of work.

From my home University, I feel extremely lucky to have developed a close academic bond with my co-adviser, Pedro Vieira, for the past three years, whose work, product of his knowledge and versatility, I have always admired and was inspired by. Despite the distance and the amount of mentees, he remained available and has given me precious orientation without fail and, for that, I am truly grateful. I am also much obliged for the Erasmus Traineeship Grant, whose process was so professionally handled by FCT's Academic Division team.

Just as importantly, I cannot thank my family enough. A heartfelt thanks goes out to my father, Jorge Silva, for the wisdom, guidance and togetherness, in spite of the two-thousand miles between us - I hope I did you proud and that someday I can be half the man you are. Thank you to my grandparents, Mário and Cândida Silva, for the unconditional love and support (emotional and financial), ever-present regardless of me winning or losing. My sincere appreciation for all the concern and attention shown by my stepmother and little brother, Ana Caetano and Santiago Caetano e Silva. Last but not least, I am taking the opportunity to thank my friends Gonçalo Martins, for a lifelong friendship, Henrique Avelar, with whom there will never be a dull moment, and, quite especially, Maria Inês Roxo, to whom I owe whatever sanity I have left. All three went out of their way to visit me in Aachen, not unlike how they have always stuck by my side, no matter what.

Abstract

Healthcare has been remarkably morphing on the account of Big Data. As Machine Learning (ML) consolidates its place in simpler clinical chores, more complex Deep Learning (DL) algorithms have struggled to keep up, despite their superior capabilities. This is mainly attributed to the need for large amounts of data for training, which the scientific community is unable to satisfy.

The number of promising DL algorithms is considerable, although solutions directly targeting the shortage of data lack. Currently, dynamical generative models are the best bet, but focus on single, classical modalities and tend to complicate significantly with the amount of physiological effects they can simulate.

This thesis aims at providing and validating a framework, specifically addressing the data deficit in the scope of cardiorespiratory signals. Firstly, a multimodal statistical synthesizer was designed to generate large, annotated artificial signals. By expressing data through coefficients of pre-defined, fitted functions and describing their dependence with Gaussian copulas, inter- and intra-modality associations were learned. Thereafter, new coefficients are sampled to generate artificial, multimodal signals with the original physiological dynamics. Moreover, normal and pathological beats along with artifacts were included by employing Markov models. Secondly, a convolutional neural network (CNN) was conceived with a novel sensor-fusion architecture and trained with synthesized data under real-world experimental conditions to evaluate how its performance is affected.

Both the synthesizer and the CNN not only performed at state of the art level but also innovated with multiple types of generated data and detection error improvements, respectively. Cardiorespiratory data augmentation corrected performance drops when not enough data is available, enhanced the CNN's ability to perform on noisy signals and to carry out new tasks when introduced to, otherwise unavailable, types of data. Ultimately, the framework was successfully validated showing potential to leverage future DL research on Cardiology into clinical standards.

Keywords: Multimodal Cardiorespiratory Signals, Sensor Fusion, Gaussian Copula, Deep Learning, Convolutional Neural Network, Data Augmentation

Resumo

A indústria médica encontra-se em mutação decorrente do fenómeno *Big Data*. Enquanto o *Machine Learning* (ML) já assume tarefas clínicas simples, algoritmos de *Deep Learning* (DL) não tiveram o mesmo tipo de aderência, não obstante serem mais capazes. Tal deve-se à necessidade massiva de dados para treino, à qual a comunidade científica não consegue atender.

Apesar dos muitos algoritmos de DL promissores, poucas são as soluções que endereçam especificamente a escassez de dados. Atualmente, modelos generativos dinâmicos constituem a melhor opção, mas focam-se apenas numa determinada modalidade clássica e tornam-se tão mais complexos quanto o número de efeitos fisiológicos que modelam.

Esta tese visa a fornecer e validar uma *framework* que aborda diretamente o défice de dados nas modalidades cardiorrespiratórias. Primeiramente, um sintetizador estatístico multimodal foi projetado para gerar sinais artificiais, longos e anotados. Ao expressar os dados através de coeficientes de funções predefinidas que lhes são ajustadas e descrevendo a sua dependência com cópulas Gaussianas, associações inter e intramodais são capturadas. Assim, novos coeficientes são amostrados para gerar sinais multimodais artificiais com as dinâmicas fisiológicas originais. Modelos de Markov permitiram a inclusão de batimentos normais e patológicos bem como artefactos. Segundo, uma rede neuronal convolucional (CNN) foi arquitetada para fusão de sensores de forma inovadora e treinada com sinais sintetizados, sob condições experimentais realistas, sendo avaliado o seu desempenho.

Tanto o sintetizador como a CNN atingem o nível do estado da arte e ainda introduzem inovações com a geração de múltiplas categorias de sinal e melhoramentos no erro da tarefa de deteção, respectivamente. Com *data augmentation* rectificaram-se quedas de desempenho derivadas da insuficiência de dados, potenciou-se o desempenho das CNNs em sinais ruidosos e permitiu-se a execução de novas tarefas mediante a apresentação de categorias de sinal, de outra forma, indisponíveis. Arrematando, a *framework* foi validada com sucesso demonstrando potencial para elevar futuros desenvolvimentos do DL em Cardiologia até padrões clínicos.

Palavras-chave: Sinais Cardiorespiratórios Multimodais, Fusão de Sensores, Cópula Gaussiana, *Deep Learning*, Rede Neuronal Convolucional, *Data Augmentation*

Contents

List of Figures

List of Tables

Acronyms

ABP	arterial blood pressure
Adam	adaptive moment estimation
AI	Artificial Intelligence
ANN	artificial neural network
ApEn	approximate entropy
ARMA	autoregressive moving average
AV	atrioventricular node
BP	blood pressure
BPG	ballistocardiography
CC	characteristic component
CE	cross entropy
cECG	capacitively coupled electrocardiography
CNN	convolutional neural network
CO	cardiac output
CorrDim	correlation dimension
CSF	cerebrospinal fluid
CX	left circumflex branch
DBP	diastolic blood pressure
DL	Deep Learning
ECG	electrocardiography
FCN	fully-connected network
FDA	U.S. Food and Drug Administration
FN	false negative
FP	false positive
FS	Fourier series
GAP	global pooling
GMM	Gaussian mixture model

HAR	human activity recognition
HF	high-frequency component
HR	heart rate
HRV	heart rate variability
ICU	intensive care unit
ITS	inverse transform sampling
LAD	anterior descending branch
LAP	lead asymmetrical pooling
LF	low-frequency component
MAD	median absolute deviation
MAP	mean arterial pressure
ML	Machine Learning
NLLSM	non-linear least squares method
NSTEMI	non-ST elevation myocardial infarction
OSA	obstructive sleep apnea
PAC	premature auricular contraction
PCA	principal component analysis
PIT	probability integral transform
PPG	photoplethysmography
PPV	positive predictive value
PTT	pulse transit time
PVC	premature ventricular contraction
ReLU	rectified linear unit
RESP	respiratory signal
RF	respiratory frequency
RIP	respiratory inductance plethysmography
RMSE	root mean squared error
RSA	respiratory sinus arrhythmia
SA	sinoatrial node
SBP	systolic blood pressure
SE	sensitivity

SEM	standard error of the mean
SGD	stochastic gradient descent
SNR	signal-to-noise ratio
SoS	sum of sines
STEMI	ST elevation myocardial infarction
SV	stroke volume
SVR	systemic vascular resistance
TN	true negative
TP	true positive

Chapter 1

Introduction

1.1 Context & Relevance

Recent years have seen an exponential increase in digital data volume: a phenomenon which coined its own buzzword - Big Data. With progressively higher amounts produced daily per capita, the last two years, alone, account for the creation of about 90% of all the data in the world [1]. Artificial Intelligence (AI) started, thus, turning heads for its ability to extract actual information out of this data flood. In 2017, McKinsey reported Healthcare to be the industry which suffered the most core business practice changes due to Big Data, just after High Tech and Energy [2], thanks to patient record digitalization and widespread of continuous monitoring devices both inside and outside clinical infrastructures.

Among AI techniques, ML shone the brightest, consolidating its place in low-risk clinical tasks oriented to both patients and practitioners, from operating room scheduling to gene sequencing. Despite no lack of quality research in higher-risk, point-of-care applications, like diagnostics through signal processing, seldom do these ML algorithms pass the strict clinical regulations [3]. Algorithms stemming from its subdomain, DL, have, however, shown higher potential for advanced clinical tasks given their capability of perceiving extremely complex associations in data, allowing for more robust and accurate analysis. In fact, a Food and Drug Administration (FDA)-approved algorithm for diagnosing diabetic retinopathy through image analysis is already in use, but it was still an isolated case [4].

Naylor highlights that one of the reasons DL has trouble mainstreaming its clinical application is the lack of data available to meet algorithms' training demands in order for them to perform satisfactorily and be scalable. The source of the problem, he points out, is while big corporations detain the monopoly of Healthcare data, coming from an endless array of streams, independent or institute researchers position themselves in the opposite end of the spectrum, topped by scarce data sharing and centralization initiatives between both parts [5]. In truth, data acquisition is typically a cumbersome process requiring trained personnel, stable experimental conditions and compliance with legal and ethical concerns [6]. Such restrictions hinder the development of novel DL methods with clinical potential [4–6]. In other industries, computer vision algorithms have

benefited from projects on artificial image generation in case of data shortage, but, while this is usually an effective solution, alerts have been raised as complexity varies in different applications, requiring a proportional amount of previous validation [7, 8].

This thesis endeavors to deliver a novel, centralized framework to synthesize large, artificial recordings with annotated, multimodal cardiorespiratory signals, whose complexity resembles the ones acquired from humans, to cover the existing discrepancies of data availability. Furthermore, a CNN architecture optimized for sensor-fusion is proposed and used to validate data augmentation promoted by the synthesizer's output, showcasing the benefits that involving artificial data in training has on the development, enhancement and testing of DL algorithms.

1.2 Objectives

To this effect, a set of milestones were defined to ensure that the project's individual components function along the lines of the end goal's requirements and can individually be of some relevance to the state of the art of their respective research fields. Goals for the synthesizer framework include:

- Versatile generation of a variety of cardiorespiratory multimodal data. At least, electrocardiography (ECG), blood pressure (BP), photoplethysmography (PPG), ballistocardiography (BPG) and capacitively coupled electrocardiography (cECG);
- Generation of not only normal, but also pathological heart cycles and artifacts with adequate annotations;
- Integration of multiple heart cycle types and noise into full recordings of arbitrary length;
- Realistically reproduce the physiological dynamics of real, subject-specific data recordings;
- Obtain generated data comparable to their real counterparts in time, frequency and phase domains.

For the DL algorithm:

- Design a generalizable CNN architecture adaptable to a variety of cardiorespiratory multimodal data;
- Capability to robustly obtain vital parameter estimations by means of beat detection;
- Characterize in detail the network's capabilities and optimal hyperparameters;
- Obtain successful performance metrics from tests on real signals, in accordance and compared to state of the art standards.

Both pieces are designed with the possibility of being independently used in the future for signal synthesis, beat detection or classification purposes and further developed. Finally, their conjunction should set the foundation to encourage, through validation, the use of synthetic data to develop, enhance and test cardiorespiratory algorithms, especially from the realm of DL, contributing to unveil the full clinical potential of these tools.

1.3 Dissertation Overview

The remainder of this thesis is arranged in such a way to clearly detail the means by which the synthesizer and the CNN were designed and what they are able to accomplish, culminating on how they support the importance of data augmentation in the light of the introduced DL limitation. A structure that privileges a converging explanation of both components was chosen, as follows:

- **Chapter 2** contains a set of theoretical concepts for the reader to make their acquaintance. These are pertinent to the project, necessary for its comprehension and will be sufficiently deepened to optimize understanding of what would, often, be unfamiliar notions to most;
- **Chapter 3** provides a thorough review of the research output most relevant to the thesis. Contemplated works are either necessary to explain how the suggested framework contributes to push the state of the art forward or have in some way inspired or been integrated into the project. It is also subdivided into research related to signal modeling and DL;
- **Chapter 4** separately details the tools, design and test procedures used in the development of the synthesizer and the CNN to reach the proposed objectives in chronological order;
- **Chapter 5** starts by presenting validation and test results obtained from the generative algorithm, followed by validation of the DL algorithm and a showcase of key results from the joint use of both. All results are adequately supported by brief discussions;
- **Chapter 6** concludes the thesis by abridging the findings and connecting them to possible applications in real-world scenarios, indicating its limitations and contributions. Moreover, it contains reference to research output stemming from this campaign and what could be in store for future developments.

Chapter 2

Theoretical Concepts

2.1 Concepts in Medicine

This section introduces the Physiology and Electrophysiology concepts rooting the dynamics behind the biosignal modalities this thesis targets.

2.1.1 Anatomy and Physiology of the Cardiorespiratory System

Composed by the cooperating cardiovascular and respiratory systems, the cardiorespiratory system endeavors to maintain homeostasis by gate-keeping gas exchanges between carbon dioxide (CO_2) and oxygen (O_2) and providing anatomical infrastructures and mechanisms so that, while the former is removed along with other waste products, the latter is supplied along with nutrients according to cellular demands [9]. To understand how these goals are met, an anatomical and physiological foundation on each of the supra-referred subsystems is laid with Sections 2.1.1.1 and 2.1.1.2.

2.1.1.1 Cardiovascular System

The cardiovascular system is anatomically comprised of the heart and blood vessels. It provides a means of transport and imposes the fluid mechanics to successfully circulate CO_2 and O_2 throughout the organism via the blood and plays an important role in a multitude of physiological processes - always aiming at maintaining homeostasis through transport. Besides gases, it transports heat - managing body temperature by migrating blood flow from peripheral areas, such as limbs, to central areas, such as the chest and head, in cold conditions -, metabolites - byproducts of physiological processes, such as toxins filtered out in the kidneys -, hormones - whose joint-action is responsible for endocrine cellular signaling allowing long-distance communication between different areas in the organism - and blood cells - such as leukocytes, which specialize in fighting pathogens, and thrombocytes, which contribute to repair damaged blood vessels [10].

Circulation of blood and, therefore, all the components it carries, happens quasi-periodically as a result of the heart's mechanical movement. This muscular organ is tasked with cyclically receiving and pumping blood from and to the entire organism, in what can be considered a closed

loop system. It's anatomy can be seen in Figure 2.1. The heart is comprised of muscular chambers, which accommodate and eject blood volumes, and valves, which mediate blood circulation inside and out of the latter. A normal heart has four chambers, dividable into right or left (from a posteroanterior point-of-view) and atria or ventricles. Ventricles are bigger than the atria and are separated by a thick muscle wall called interventricular septum. This implies higher pressures are transmitted to the ejected blood volume upon contraction. Likewise, there are four valves - one at each exit of the chambers. The valves mediating flow from atria to ventricles are called atrioventricular or, more specifically, mitral and tricuspid, on the left and right, respectively. The remaining two are semilunar valves, more specifically, named after the arteries they are connected to - aortic and pulmonary. It is important to understand that the pump mechanism not only ensures ejection, but also refill, always relying on the pressure gradients created by contraction and relaxation of the chambers with a single opening.

The mechanical behavior of the heart results from an underlying, sequential electrical behavior occurring in the heart muscle and specialized fibers enjoying properties of automatism, excitability and conductivity [11]. Among these fibers, lie the cells forming the sinoatrial (SA) node, which independently produce action potentials, motivated by an influx of sodium ions, depolarizing the cell once a threshold value is crossed. These electrical signals are then propagated throughout adjacent cardiac muscle cells and a network of fibers, via permanently open gap junctions, reaching and, consequently, contracting different heart regions, dictating the sequential nature of the cardiac cycle [10].

A cycle begins with the entrance of an incoming blood volume to the relaxed atria from the pulmonary veins and venae cavae, while the atrioventricular valves are closed, holding the volume in the chamber (atrial diastole). Afterwards, the atria contract, as a result of the SA node firing an action potential, and the last-mentioned valves open, forcing the blood into the relaxed ventricles due to the originated pressure gradient, while the semilunar valves are closed (atrial systole and ventricular diastole). Meanwhile, the electrical signal propagates to a secondary node - the atrioventricular (AV) node - which is the only route allowing the signal to reach the ventricles, due to high electrical impedance between atria and the former. Atrioventricular valves are now closed and the ventricles contract when the action potential reaches them after a brief conduction delay at the AV node to maximize chamber filling, as the aortic and pulmonary valves open, ejecting the blood volume into the aorta and the pulmonary artery (ventricular systole), perpetuating blood circulation in both trajectories.

Circulation is performed according to two main trajectories, pulmonary and systemic. While the former guides venous blood from the heart's right ventricle through the lungs to the left atrium, the latter guides arterial blood out of the heart's left ventricle through the remaining organs and tissues (including the heart itself) to the right atrium, closing the circuit. Note how circulation inside, represented by different colors in Figure 2.1, and outside the heart structure ensures separation of both blood types and a one-way flow direction.

Mechanical manifestations of the cardiac cycle are not limited to the heart's anatomical domain as the ejected, circulating blood interacts with vessels, to which energy is transferred along its way. This transfer is not always performed on the same terms, as evidenced by the arterial blood pressure (ABP) along circulation pathways. Blood vessels can be classified further - arteries are subdivided from higher to lower caliber into large, small arteries and arterioles, while, analogously, veins are subdivided into large, small veins and venules. Additionally, a lesser caliber vessel subtype, the capillaries, is transversal to both arteries and veins, being the transition structure

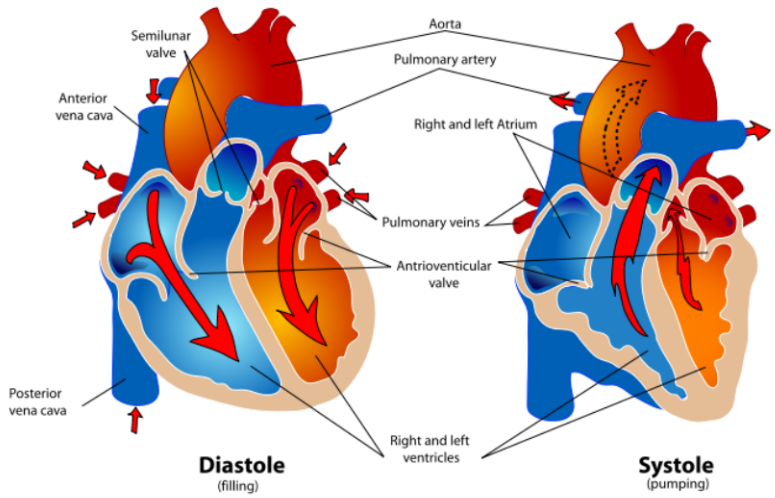

Figure 2.1: Anatomy and circulation mechanism of the heart. Blue colored structures correspond to venous blood circulation, while red colored structures arterial blood circulation. The left image illustrates ventricular relaxation and consequent filling, while the inverse process is shown on the right with blood ejection. Authored by Mariana Ruiz Villarreal.

where blood flows from arteries to veins, after undergoing capillary exchange with surrounding tissues. Typically, the vascular system is organized as a tree centered on the heart, where most large caliber vessels are located, which ramify along the body towards organs and tissues, where smaller caliber vessels and capillaries are lodged. Arteries have higher elasticity and compliance than veins [11]. Therefore, they are more resistant to deformation under pressure, which explains how ABP is maintained until the O_2 and nutrients in the blood reach diffusion zones, thanks to the additional pressure returned by arteries' walls, shown in Figure 2.2. On the other hand, circulation in veins experience lower pressures, despite accumulating higher blood volumes. Decreasing radii of vessels restricts blood flow in small caliber structures and pressure is increased, allowing fluid and other blood components to leak into neighboring tissues due to hydrostatic pressure [12].

2.1.1.2 Respiratory System

The respiratory system is predominantly represented by the respiratory tract, starting at the nose cavity and ending at the alveoli at the airways' extremities, and is supported by a set of additional structures - respiratory muscles and the rib cage -, as represented in Figure 2.3. It aims at performing pulmonary ventilation in order to allow CO_2 and O_2 transportation and alveolar diffusion, turning venous blood into arterial blood by removing the former and supplying the latter [11].

Air flows in and out of the lungs due to pressure gradients between the atmosphere and the inside of the lungs. These pressure gradients change by an expanding and contracting rib cage: during inspiration, the main muscular effort is performed by the diaphragm which, supported by auxiliary inspiratory muscles (e.g. external intercostal muscles), contract, causing rib cage volume to increase and, subsequently, inner pressure to drop in comparison to outside air pressure. Such originates an inward airflow, filling the lungs. The inverse process occurs when the

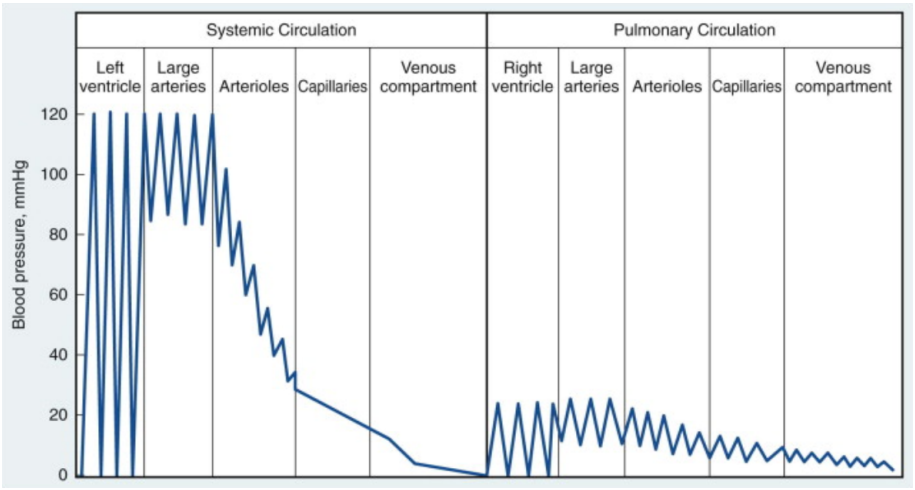

Figure 2.2: Blood pressure along circulation pathways. The leftmost section of the graph represents an overall higher pressure and longer circuit - systemic circulation -, as the right section shows opposing features - pulmonary circulation. Adapted from [13].

rib cage volume decreases motivated by the lungs' own elastic, retractile properties and, in case of active breathing, contraction of auxiliary expiratory muscles (e.g. internal intercostal muscles). The more effort is put into ventilation, the more auxiliary muscles are used in what would, otherwise, be a process sufficiently supported by diaphragmatic activity and lung and rib cage elastic properties [11]. An overview of the process can be seen in Figure 2.3.

Air flows in and out of fine alveolar ducts and alveolar sacks, thus, fully adapted to facilitate gas diffusion, through more conductive respiratory tract structures [11]. At this stage of the tract, the air filled alveoli have higher O_2 and lower CO_2 partial concentrations than the dense capillary network surrounding them, which in a healthy lung is translated as a normal ventilation and perfusion relationship. This leads to O_2 molecules to move out of the alveoli to the capillary lumen and the opposite phenomenon to occur for CO_2.

Already in the bloodstream, most oxygen molecules are transported in the erythrocytes by

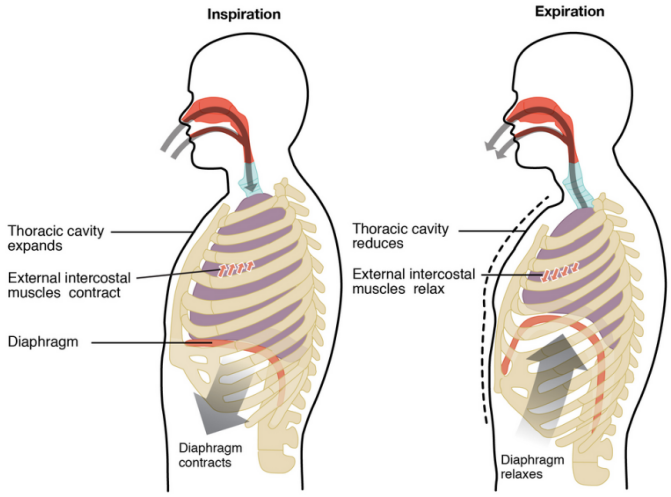

Figure 2.3: Schematic of quiet respiration as a function of rib cage, diaphragm and auxiliary muscle movement. Adapted from [14].

connecting to iron in the hemo groups their hemoglobin proteins [10]. Thus, they then follow the circulation path, making their way into every cell in the organism, following the logic explored earlier in Section 2.1.1.1.

2.1.2 Cardiorespiratory Regulation & Coupling

As far as periodicity is concerned, both the cardiovascular and respiratory system function in analogous ways. In truth, they share a regulatory entity, the autonomous nervous system. Autonomous control is, in fact, a balance between the tone of two opposite systems, the sympathetic and parasympathetic, loosely related to biological activity increase and decrease, respectively. This translates into, for example, a heart rate (HR) increase or decrease, while in the respiratory system, the same occurs with the respiratory frequency (RF). There is, however, the exception of, less often, ventilation being voluntarily controlled.

2.1.2.1 Cardiovascular Regulation

With regards to cardiovascular control, HR relates to sodium channels, previously mentioned in Section 2.1.1.1. Increasing or decreasing the rate at which of the sodium ion influxes causes action potentials to be deployed at different rates. A normal HR of about 60 to 100 BPM can, therefore, rise above that range in situations which champion sympathetic activation, such as physical activity, or fall in others where sympathetic activity is decreased or vagal tone is increased, such as during sleep. Ultimately, autonomous control directly or indirectly affects components with respect to ABP, influencing the latter. These components include stroke volume (SV), which is the volume of blood ejected from the left ventricle per beat, and systemic vascular resistance (SVR), referring to the resistance imposed by the circulatory system to blood flow. These three components relate to ABP as follows:

$$ABP = CO \times SVR$$

where $CO = SV \times HR$, being CO is the cardiac output, or the blood volume ejected from the left ventricle per unit time. SV regulation is essentially done in two ways: inotropy and through Frank-Starling's mechanism. The first one concerns how contractility of the myocardium increase directly correlates to sympathetic activity by direct innervation of the hearts atrial and ventricular wall. The second describes how muscle contractility increases by optimizing the length of muscular fibers (positional relation between actine and myosin filaments), originated by stretching the latter when a higher blood volume lies within the ventricles right before systole occurs (preload) [11]. The autonomous nervous system influences this mechanism by varying the preload via a higher or lower venous return to the heart, which is in turn related to factors directly dependent on nervous control such as variation of the radius of blood vessels. SVR control is more directly influenced by sympathetic or parasympathetic activity as lean muscle in blood vessel walls are directly signaled to vasoconstrict or vasodilate, respectively.

ABP changes at stretch-receptors sited at the carotid sinus and the aortic arch are communicated to the central nervous system, specifically the cardiovascular center in the medulla oblongata, via the afferent nerve fibers of the glossopharyngeal and vagus nerve. After information integration, an adequate response in the form of nervous stimuli is propagated along efferent nerves, either the sympathetic nerve for a sympathetic response or the efferent vagus nerve for a parasympathetic response - baroreflex control. These can directly connect back to the heart,

e.g. atrial and ventricular walls or the SA nerve, or connect to adrenal glands which release norepinephrine, for sympathetic response, and epinephrine, for parasympathetic response, into the blood stream for endocrine signaling. While, until now, only short-term regulation processes were mentioned, long-term baroreflex mechanisms may take place starting by endocrine signaling the kidneys to differ its water retention capability, in order to increase or decrease venous return, through volemia [11]. A baroreflex scheme is presented below in Figure II.1. Additionally, a low-frequency (LF), systematic oscillation of the blood pressure can be discerned, to which the name of Mayer waves was given in honor of one of the German physiologist who first described it on the second half of the nineteenth century. Their origin is not fully known since despite coherence with sympathetic nervous activity, connections with oscillations of the baroreflex or an independent endogenous oscillator were hypothesized. Furthermore, strong associations with the vasomotor tone of the blood vessels were found, meaning it influences ABP by means of altering SVR [6, 15].

2.1.2.2 Respiratory Regulation

The involuntary domain of the respiratory system is supported by a similar nervous structure, relying mostly on peripheral and central, muscle and joint chemoreceptors and stretch-receptors located in the lungs, afferently reporting back to the pons and the medulla oblongata, the respiratory centers for stimuli integration. Central chemoreceptors focus on the pH level of the cerebrospinal fluid (CSF) through hydrogen ion and partial CO_2 concentration monitoring while peripheral receptors monitor the blood with an additional concern for partial O_2 pressure. The latter are also part of the afferent paths which include the glossopharyngeal and vagus nerves [11]. In order to maintain the right partial concentrations of CO_2 and O_2 to allow diffusion, ventilation can be increased or decreased by stimulating accessory respiration muscles and the diaphragm through, for example, the phrenic nerve. Hence, while a normal RF averages 12 ventilations per minute and a normal tidal volume of about 0.5 l, corresponding to a respiratory minute volume of 6 l/min., these values can change according to sympathetic or parasympathetic feedback. The above mechanisms are represented in Figure II.1.

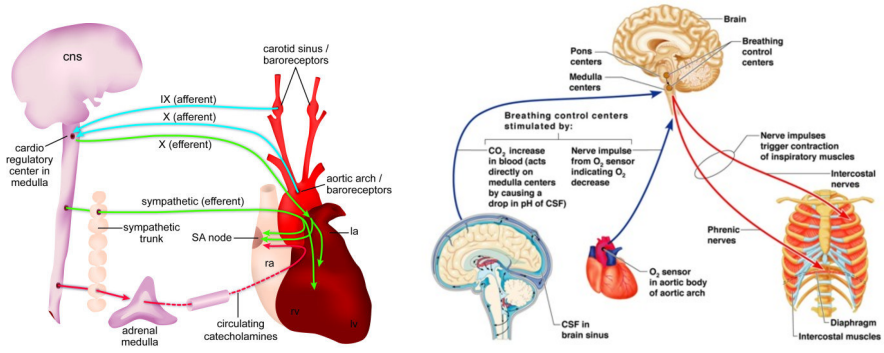

Figure 2.4: Representation of baroreflex and respiratory regulation structure and mechanism. On the left, afferent nerves are represented in blue, while efferent nerves in green. IX pairs code the glossopharyngeal, while X code the vagus nerve pairs. Adapted from [16]; On the right, blue encodes afferent nerves, while red encodes efferent nerves. Authored by Benjamin Cummings.

2.1.2.3 Cardiorespiratory Coupling

Notwithstanding the separation with which both subsystems have been explained until now, their intertwined functioning is equally important. The most prominent example of such is the respiratory sinus arrhythmia (RSA). It describes how HR increases during inspiration and decreases during exhalation, functioning as a high-frequency (HF) regulatory oscillator. Further, ABP is also shown to oscillate during normal respiration and even more so during deep respiration, hinting at the baroreflex being behind the phenomenon. Its origin is, nonetheless, disputed, as the Bainbridge reflex, may also be associated [15].

2.1.3 Cardiorespiratory Signal Modalities

Widespread use of medical sensors to obtain electrical signals or even images provided, not only new, objective insight is obtained about how human physiological processes function, but also pathophysiological processes - completely changing the medical diagnostics' paradigm [17]. Even though this work focuses more on software, hardware also bears importance to how cardiorespiratory signals are obtained. Both traditional and novel medical signal acquisition devices, focusing on ECG and cECG, to detect electrical activity, and BP, PPG and BCG to detect mechanical activity are explored in the next subsections.

2.1.3.1 Electrical Activity Medical Sensors

As explained in Section 2.1.1.1, cell depolarization causes cells' intracellular area to become positiviely charged by a sodium influx to the intracellular space. While in resting state, where the intra and extracellular polarities are shifted. As the action potential propagates to neighboring cells by transferring positively charged particles through gap junctions, causing them to depolarize the cell they enter, while the originally depolarized cell returns to its resting state instants later. On a larger scale electric fields are created as a result of the voltage existing between polarized and depolarized regions' extracellular spaces.

Conductive electrodes placed on the body's surface pick up on these electric current vectors. Common ECG electrodes placement can be divided into two configuration types: bipolar and unipolar leads as shown in Figure 2.5. The former uses a single positive lead and a single negative,

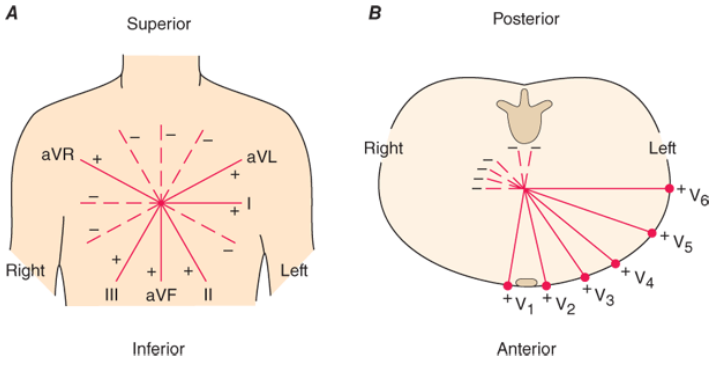

Figure 2.5: On the left, representation of the bipolar limb leads I, II and III (forming Einthoven's triangle) and enhanced unipolar leads aVL, aVR and aVF (forming the vertical plane). On the right, placement of the remaining bipolar ECG electrodes is shown (forming the horizontal plane). Together, they form a standard 12-lead ECG setup. Adapted from [11].

each placed on different vertices of the Einthoven triangle, while the latter resorts to single positive leads placed around the left side of the chest surface, along the fourth and fifth intercostal spaces. Measured voltages are typically low, around 1 mV, which is the reason why the equipment must consider higher voltage sources, such as skeletal muscle activity or nearby power lines, and apply frequency filters and high common mode rejection ratio amplifiers. Different configurations deliver different perspectives of the change in the electric field during a cardiac cycle.

Meanwhile, cECG is able to acquire the same biosignals without directly contacting with the skin, by plate capacitors formed by metallic surfaces placed in close proximity with the body. This method has the advantage of being unobtrusive, allowing integration into everyday life activities, requiring less preparation than a standard acquisition with ECG (where the skin is cleaned and a conductive gel is applied). Successful applications have been reported, for example, in integration with beds [18–20], chairs [21, 22], toilet seats [23, 24] and car seats [25–27]. However, the obtained signal belongs to an even lower voltage range, easily shadowed by artifacts, due to a smaller signal-to-noise ratio (SNR).

Both systems capture projections of the depolarization and repolarization vectors along the axes established by the set up leads. As a result and given the periodicity of the cardiac cycle, characteristic waveforms appear in a electrocardiographic recordings respective to different cycle stages explained in Section 2.1.1.1, as shown in Figure 2.6. It is worth noting that mechanical activity occurs with a slight delay to the electrical event which primes it. Atrial systole is represented by the P-wave, which corresponds to the depolarization vector created by the action potential fired by the SA node and propagated through both atria along the Bachmann's bundle. The PR segment follows, where no activity is recorded due to the conduction delay in the AV node. The QRS-complex is the most energetic portion of the signal and is composed of three individual waves, each representing a different part of the ventricular systole: the Q-wave is respective to the septal depolarization; the R-wave corresponds to most of the ventricular wall's depolarization along the right and left bundle, from inside to outside; the S-wave represents the end phase of ventricular depolarization along the terminal portion of the Purkinje fibers. Finally, the T-wave corresponds to ventricular diastole, when the ventricular walls repolarize [11].

A progression of the HR throughout time - HR variability (HRV) - is also a commonly used signal modality which is directly related to RR-tachograms, which graph the progression of duration between consecutive R peaks. These are useful assets in diagnosing some arrhythmia types and exploring regular physiological phenomena such as RSA, Mayer waves and autonomous control through balancing between sympathetic and parasympathetic tone. Taking the Fourier transform of an HRV signal is a common way to conduct such studies - in the frequency domain, LF and HF wide peaks can be discerned, respectively corresponding to the effects of the Mayer waves and RSA. Moreover dividing the spectrum, as seen in Figure 2.7, into LF and HF segments, provides insight on autonomous control through the integration of each segment and establishing a ratio LF/HF.

2.1.3.2 Mechanical Activity Medical Sensors

Mechanical manifestations of the cardiac cycle and respiration can also be measured as electrical signals from transducers detecting organ motion, deformation or fluid displacement.

Cardiac contraction and relaxation as well as ejection of blood into arteries is a turbulent enough process to cause involuntary, slight body movements which can be related to different

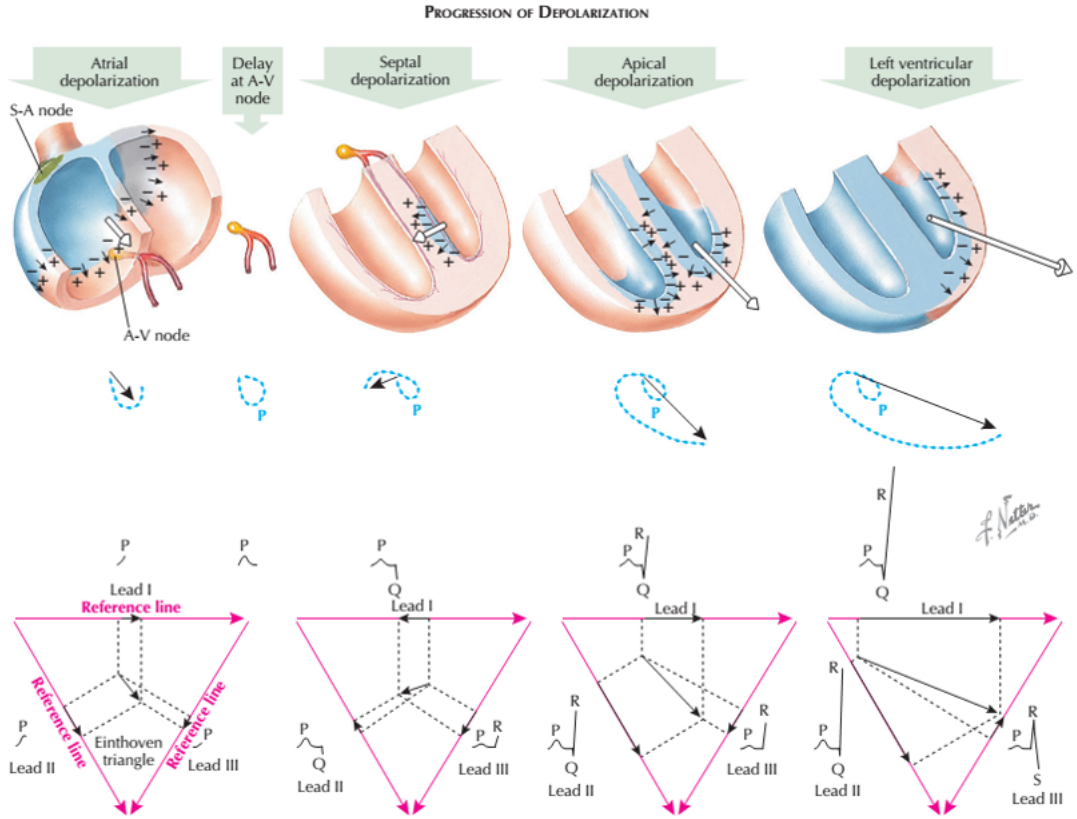

Figure 2.6: Evolution of the depolarization vector until the R-wave to exemplify how current vectors draw ECG waveforms using bipolar leads. Projections of the current vectors produced between depolarized regions, with negative extracellular domain, and polarized regions, with positive extracellular domain, vary during different stages of the cardiac cycle, yielding characteristic waves a each one and ultimately forming the PQRST-complex. Adapted from [28].

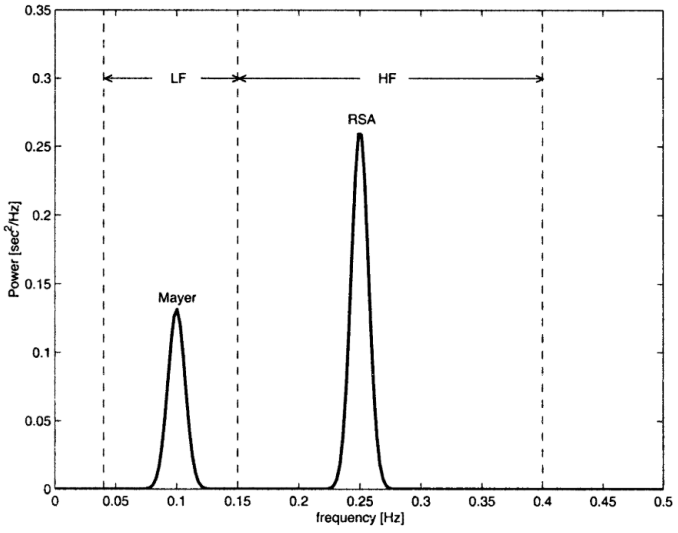

Figure 2.7: Power spectrum of an RR-interval with Mayer and RSA peaks and division into LF and HF segments. Adapted from [29].

stages of the cardiac cycle. BCG and SCG are two of the methods working with these movements, although using different approaches, as the former relies on body movement measurements along different axes on a flat surface while the latter directly measures chest displacement with an accelerometer placed on the sternum area. Similarly to the cECG, these are also unobtrusive modalities, sharing the same advantage of being possible to integrate in daily activities, as evidenced by extensive work: revolving around beds through the use of framing as sensors [30], pressure pads [31], electromechanical film or piezopolymers [32, 33], measuring voltage across moving layers underneath the mattress [34], infrared light scattering inside the mattress; floor mat instrumentation [35]; and wheel-chairs [36, 37]. In these modalities too, characteristic waveforms can be observed, whose segment names remained since the original nomenclature created by Isaac Starr, who implemented the first ballistocardiograph [38, 39], despite other BCG systems showing those same waves with shape and timing differences. A representation of these waves obtained with Starr's system compared with the typical ECG waveform is shown in Figure 2.8. They were labeled taking into consideration variations among different systems and are typically grouped in pre-systolic waves (F and G) representing pre-systolic events, systolic waves, which include the most prominent H-wave, preceding or simultaneous to the ventricular ejection start and I, J and K waves, which correspond to the ventricular systole process from early stages to its end, and, finally, diastolic waves (L through O) [39–41]. Since respiratory displacements are more significant than mechanical displacements of the cardiovascular system, as evidenced by the differences of stroke and tidal volumes, these waveforms highly strongly to respiration and overlapping frequency spectra.

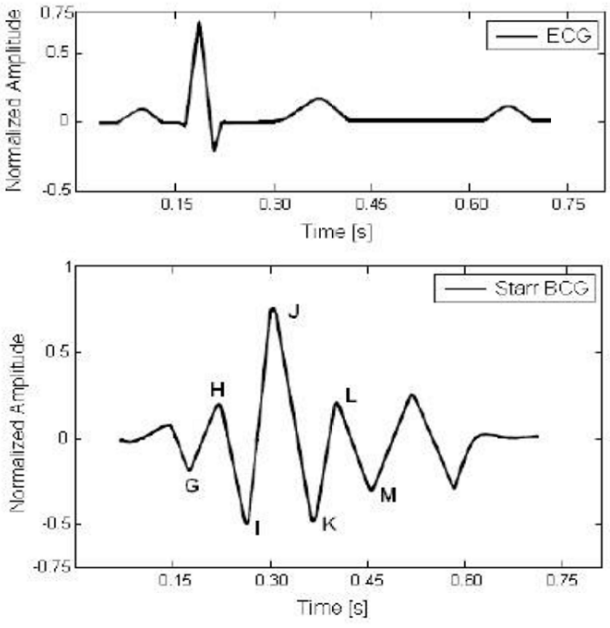

Figure 2.8: Introduction and comparison of BCG waveform from Starr's system with typical ECG cycle signal. Accounting with a delay between electrical activation and blood ejection mechanical impulse propagation, one can see notice a correlation with pre, post and systolic events in ECG and the corresponding BCG waveforms. Adapted from [41].

With respect to circulation, ABP is tracked with two commonly used techniques: invasively, using an intra-arterial catheter, and, non-invasively, with PPG. The first is usually employed in

intensive care units (ICU), where an arterial line is set up with a saline filled tube connected to a pressure transducer, which translates the communicated pressure wave into an electrical signal. The typical waveform obtained from this method is depicted in Figure 2.9, from which a number of measurements can be extracted. Firstly, the peak pressure corresponds to the systolic blood pressure (SBP), achieved right after blood is ejected out of the left ventricle to the aorta. Conversely, the minimum pressure corresponds to the diastolic blood pressure (DBP) and is achieved before ventricular systole, as the atrioventricular valves close. After the peak is reached, a dicrotic notch is noticeable, which represents the aorta's elastic retraction after the aortic valve closes, as explained in Section 2.1.1.1. Pulse pressure is defined as the difference between SBP and DBP, while mean arterial pressure (MAP) is calculated as

$$MAP = \frac{2}{3}DBP + \frac{1}{3}SBP,$$

where the multiplication factors relate to how long each phase takes relative to the whole cardiac cycle. The waveform pulse, naturally, occurs with a slight delay of about 200 ms in relation to the R-wave of the ECG, frequently addressed as pulse transit time (PTT) [11], which is inversely proportional to pulse velocity.

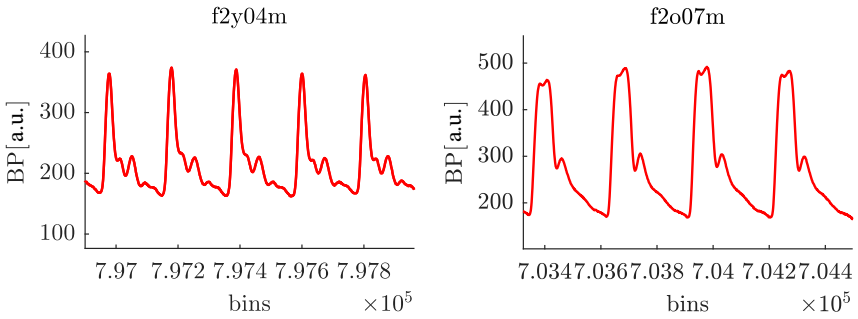

Figure 2.9: Typical BP waveforms in both young (left) and elderly (right) subjects. These waveforms were obtained from the Fantasia [42] database subjects, whose identification labels stand above the graphs.

While the accuracy of BP through a catheter is dependent on a series of factors transversal to most oscillating systems such as oscillation, damping and resonance [43], it is still considered to be more accurate than existing non-invasive alternatives, which overestimate SBP during hypotension [15] although coherence with invasive methods is shown to increase with BP. Typically, PPG makes use of optical properties of blood cells to extract a similar waveform, by radiating a near-infrared or visible light with a known intensity, I_i, to the skin and, after loosing some absorbed intensity, I_a, and reflection with intensity, I_r, obtaining a transmitted frequency, I_f. Thus, after wavelength and location calibration, a similar waveform to that of catheter-based measurements is inferred from blood oxygenation since oxyhemoglobin and deoxyhemoglobin absorb light frequencies differently, as do different blood volumes [44]. Being unobtrusive, this technique mitigates the risk of hospital infection, ischemic disease and vascular damage [43]and further uses have been tested, such as over-clothing applications in chairs [45] and mattresses [46] and image plethysmography at a distance to infer the underlying signal from an exposed skin region [47].

Similarly to what happens with different ECG leads, multiple waveforms can be obtained from different catheter or photoplethysmograph positioning. Peripheral circulation differs from central circulation since, for example and as explained in Section 2.1.1.1, RAP increases and the vascular

tree's impedance builds up, as shown in Figure 2.10. Additionally, differences between young and elderly subjects can be verified as a result of, for example, reduced arterial distensibility, which increases pulse pressure, delays the systolic peak pressure and attenuates the diastolic pressure wave [48, 49], as shown in Figure 2.9.

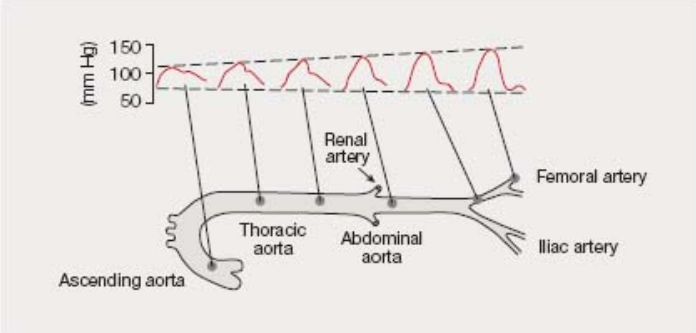

Figure 2.10: Evolution of the arterial waveform along the arterial tree. PP becomes wider while MAP increases only slightly, agreeing with what seen in Figure 2.2. Adapted from [49].

Concerning the respiratory system, various ways of obtaining the respiratory signal (RESP), are available, making use of the displacement of the rib cage or the air displacement during inhalation and exhalation. The most common type of pulmonary function test is spirometry. As lungs are filled with air because of a drop in interior air pressure caused by the expansion of the thorax (inspiration) and emptied by air pressure build up caused by the inverse process, a mouth piece used by a patient is able to conduct the air flow in and out of a chamber where a weight is displaced, drawing the signal corresponding to variations of air volumes breath in and out of the spirometer. After calibration, it is capable of measuring air volumes inside the lungs at different stages of respiration, e.g. tidal volume, total lung capacity, residual volume, yielding graphs as the one shown in Figure 2.11. In polysomnography, a nasal thermistor can also be used

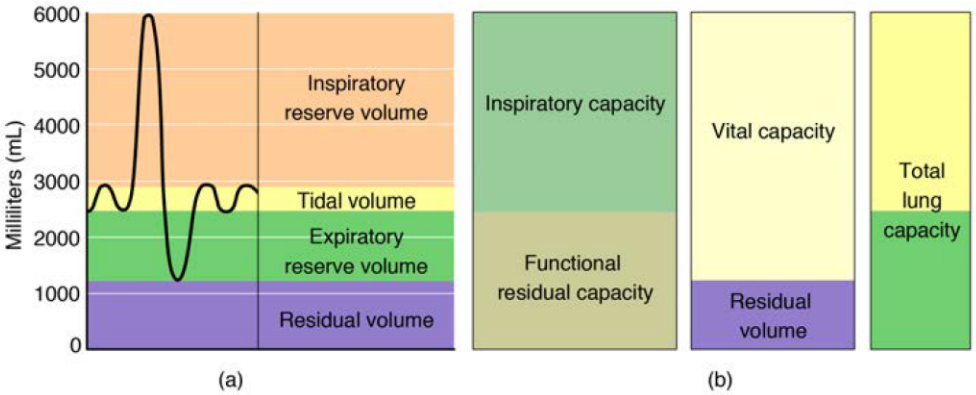

Figure 2.11: Typical waveform obtained from spirometry tests. Respiratory volumes (a) can be directly interpreted from graph values and can be combined to yield respiratory capacity values (b). Both these metric types are useful for diagnosing a number of respiratory diseases such as asthma and chronic obtrusive pulmonary disease. Adapted from [14].

as a less obtrusive way of recording RESP during sleep stages, using temperature as an air flow proxy. As an alternative, cannula pressure transducers can also be used with a higher degree of precision [50, 51]. Rib cage displacement is also commonly used as a surrogate for respiratory

effort - the RIP (respiratory inductance plethysmography) is a technique which relies on changing magnetic fields, due to breathing, that are originated by an alternating current traveling along adhesive bands placed around the chest and abdominal area. The changing magnetic fields change the self-inductance of the coils inside the bands and the oscillation frequency, making RESP measurable with frequency demodulation [52]. Novel methods of measuring RESP continue to be tested, such as the device described in [51], which is comprised of a sensor with a microphone and a transducer placed on the suprasternal notch, capturing respiratory effort using suprasternal pressure as surrogate, snoring and respiratory flow with the microphone (appearing in different frequency ranges).

2.1.4 Cardiorespiratory Pathology & Clinical Presentation

This section introduces common pathology mechanisms where analysis of said signal modalities play a role in diagnostics. Frequently, these diseases culminate in cardiac insufficiency or, in other words, the inability of the heart to adapt to the organism's metabolic needs. Typical causes for cardiac insufficiency include coronary disease, valvular disease, arrhythmia, cardiomyopathy, arterial hypertension and diabetes, but only the first three will be covered since their gold standard means of diagnostic involves the use of ECG and BP recordings.

2.1.4.1 Coronary Disease

Myocardial ischemia, characterized as an imbalanced between O_2 supply and consumption by the myocardium, is a cardiopathy which can lead to muscle damage or even necrosis. It is mainly caused by atherosclerosis, which is an umbrella term for plaque build up inside arteries. The most relevant type for the case is coronary atherosclerosis, which occurs when the plaque build up is located in the coronary tree, mainly divided in right (RCA) and left coronary arteries. The latter bifurcates into the left circumflex branch (CX) and anterior descending branch (LAD) and each arterial division is responsible for the irrigation of different regions of the myocardium. A schematic of coronary irrigation is shown in Figure 2.12.

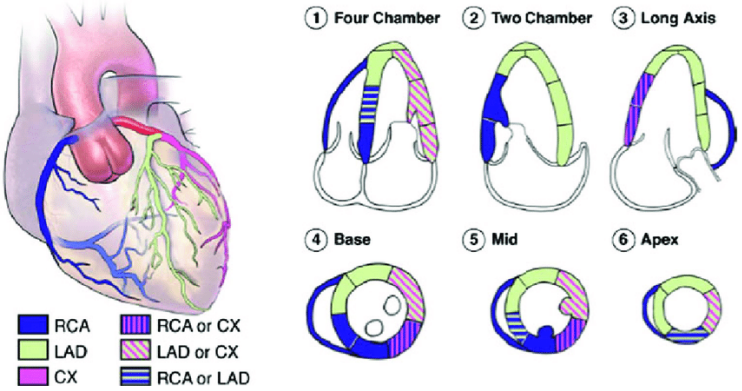

Figure 2.12: Color scheme of myocardial irrigation by coronary arteries. Different cross Sections of the heart are shown along the sagital, coronary and transversal axes to explain the irrigation domains of each of the three spoken artery branches - RCA, LAD and CX. Adapted from [53].

Rupture of the lipid core of the coronary arteries, motivated by a set of risk factors, forming thrombi and/or emboli, which can partially or totally occlude the lumen of the artery or smaller

ones downstream. Depending on the affected ramification, damage to the respective myocardial region alters the depolarization vectors, which form the ECG as seen in Figure 2.6, leading up to abnormalities such as non-ST elevation myocardial infarction (NSTEMI) or ST elevation myocardial infarction (STEMI) [54].

2.1.4.2 Valvular Disease

On the other hand since valvular damage boils down to a mechanical defect in the functioning of the heart instead of an electrical anomaly, direct consequences of valve damage found in BP or sonograms due to alterations on the flow type of the blood in between chambers and into the arteries, although compensating pathophysiological adaptations of the cardiac muscle can later alter the typical ECG waveform). Valvular diseases are broadly grouped into stenosis or regurgitation, meaning a tightening or an alleviation of the valve's opening [54]. Aortic stenosis induces a slow upward stroke, delaying the systolic peak, which ultimately is lower than normal, and may cause the dicrotic notch to not be discernible. Contrariwise, aortic regurgitation allows blood to flow back to the ventricle from the aorta after ejection, countering a rapid upstroke and increased SBP and PP with a low DBP. Mitral stenosis and regurgitation follow the same mechanisms as the previous two conditions but do not majorly affect BP recordings. Instead, pressure changes occur predominantly on the atrioventricular transition: in mitral stenosis, atrial pressure increases as opposed to ventricular pressure, increasing the pressure gradient among both chambers; in mitral regurgitation, pressure builds up in both chambers during ventricular systole, due to blood leakage back into the atria. Figure 2.13 shows typical altered waveforms of aortic stenosis and regurgitation.

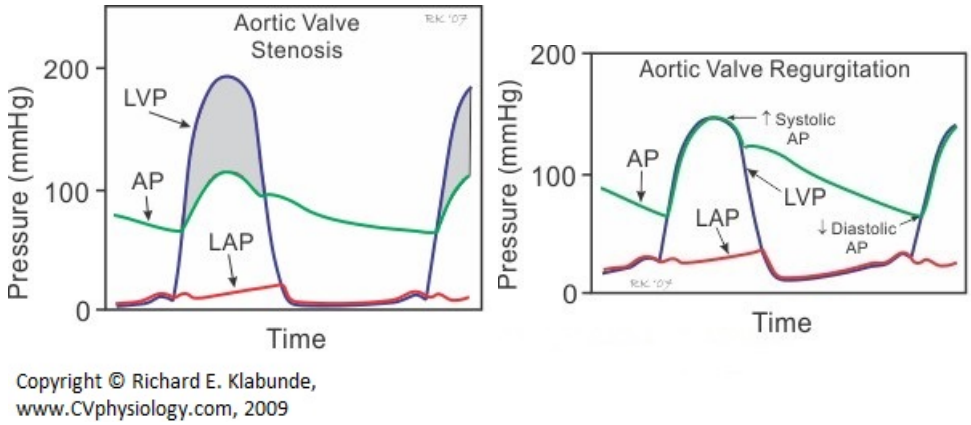

Figure 2.13: Evolution of left ventricular pressure (LVP), ABP (AP) and left arterial blood pressure (LAP) during a heart cycle with valvular damage. Authored by Richard E. Klabunde.

2.1.4.3 Arrhythmias

In this section, ectopic arrhythmias used in this project and representative of the diagnostic capability and importance of ECG tests will be presented. Arrhythmias are HR changes due to physiological regulation mechanisms, as explained in Section 2.1.2, or pathological changes in the way action potentials are formed, conducted or both. A decrease in HR below 60 BPM is called bradycardia, whereas an increase above 100 BPM is called tachycardia and a normal rhythm is called sinus rhythm.

Action potential formation can be changed in a number of ways. Altered automaticity can be caused by changes in the autonomous nervous system balance or overshadowing of the SA node by ectopic pacemaker cells. Differently, damage to the SA node, can induce non-specialized cells to play the role of pacemakers - anormal automaticity. An impulse may be triggered after a first depolarization - triggered activity. Conduction may be affected by a conduction block caused by, for instance, tissue damage or drugs, or reentry, usually caused by an accessory pathway.

Auricular extrasystole, or premature auricular contraction (PAC), is a type of supraventricular tachycardia and is characterized by early P-waves with a different morphology, despite the QRS complex remaining identical. Its occurrence may not be inserted in a pathology setting. Ventricular extrasystole, or premature ventricular contraction (PVC), presents a different QRS morphology with a duration of more than 120 ms without prior P-wave, followed by a compensatory pause. Sequences of PVC and sinus beats can be classified as bigeminy if one PVC occurs every other sinus beat, trigeminy if the ratio is 2:1, etc.. Furthermore, the number of consecutive PVC can also be labeled as pairs, triplets, etc. [54]. Both are considered ectopic beats, represented in Figure 2.14, and will be an implementation focus developed multimodal synthesizer.

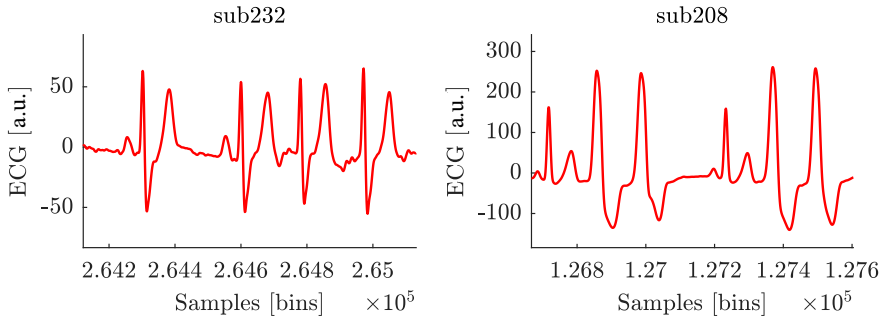

Figure 2.14: Typical ECG waveforms for both PAC (left) and PVC (right) arrhythmic events. These waveforms were obtained from the MIT-BIH Arrhythmia database subjects, whose identification labels stand above the graphs.

2.2 Concepts in Probability & Statistics

This section provides an introduction of the main mathematical tools which were leveraged to design the multimodal synthesizer.

2.2.1 Multivariate Distributions with Copulas

Copulas have been extensively used in Finance as a tool for Integrated Risk Management. Specifically, the Gaussian or normal copula was used under the assumption that prices of many financial products follow a normal distribution, making credit-sensitive asset (financial product with a high likelihood of not complying with the conditions of a loan [55]) modeling simple and elegant. Correlation between these could, therefore, be integrated in the model (which also works for non-correlated assets, where the final distribution would simply be the multiplication of individual assets probability density functions) and any behavioral fluctuation past this normal logic would be considered worth investigating.

2.2.1.1 Mathematically Definition of Copula

Mathematically, an n-copula is defined as a multivariate distribution function, typically denoted by C, with an n-box domain of $[0,1]^n$ and codomain of $[0,1]$ such that it is grounded, n-increasing and has its marginals C_k satisfy $C_k(u) = u$ for $u \in [0,1]$.

Following the theoretical framework used in [56], the concepts of grounded, n-increasing and marginals will be introduced. Let us consider a function H with n variables and $\mathrm{Dom}H = S_1 \times \cdots \times S_n$, where S_k is nonempty, has a smallest element a_k and a biggest element b_k, for $k = 1,\ldots,n$. It is considered to be grounded if $H(\mathbf{t}) = 0$, for all $\mathbf{t} \in \mathrm{Dom}H$ such that $t_k = a_k$ for at least one valid value of k. Moreover, the marginals are the functions $H_k(x) = H(b_1,\ldots,b_{k-1},x,b_{k+1},\ldots,b_n)$ with $x \in S_k$. H is considered to be n-increasing if the H-volume of the n-box $B = [\mathbf{a},\mathbf{b}] = [a_1,b_1] \times \cdots \times [a_n,b_n]$, where $\mathbf{a},\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{a} \leq \mathbf{b}$, $V_H(B) \geq 0$. In turn, this volume is defined by

$$V_H(B) = \Delta_{\mathbf{a}}^{\mathbf{b}} H(\mathbf{t}) = \Delta_{a_n}^{b_n} \ldots \Delta_{a_1}^{b_1} H(\mathbf{t}),$$

where

$$\Delta_{a_k}^{b_k} H(\mathbf{t}) = H(t_1,\ldots,t_{k-1},b_k,t_{k+1},\ldots,t_n) - H(t_1,\ldots,t_{k-1},a_k,t_{k+1},\ldots,t_n).$$

Thus, the previous conditions adapt to the copula as its margins are defined in the same way as H, $C(\mathbf{u}) = 0$ if at least one element of $\mathbf{u}$ is 0 (grounded), $C(\mathbf{u}) = u_k$ if every element of $\mathbf{u}$ is 1 except for u_k (marginals) and, ultimately, $V_C([0,u_1] \times \cdots \times [0,u_n]) \geq 0$($n$-increasing). It follows that C is increasing in each argument as two points $(u_1,\ldots,u_{k-1},x,u_{k+1},\ldots,u_k)$ and $(u_1,\ldots,u_{k-1},y,u_{k+1},\ldots,u_k)$, such that $x \leq y$, result in $C(u_1,\ldots,u_{k-1},x,u_{k+1},\ldots,u_k) \leq C(u_1,\ldots,u_{k-1},y,u_{k+1},\ldots,u_k)$, culminating in $C(\infty,\ldots,\infty) = 1$. Since its margins C_k, with $k = 1,\ldots,n$, satisfy $C_k(u) = u$ (being CDFs of uniform distributions), for $u \in [0,1]$, and $C(\mathbf{u}) = u_k$ if all elements of $\mathbf{u}$ but one are 1, follows

$$C(u_1,\ldots,u_n) = V_C([0,u_1] \times \cdots \times [0,u_n]).$$

The theorem which sponsors the use of copulas to model multivariate distributions of random variables following arbitrary univariate distributions is Sklar's theorem. It states that if H is the n-dimensional multivariate CDF and $F_1,F_2,\ldots,F_n$ are the marginal CDFs for the real random variables, $X_1,X_2,\ldots,X_n$, respectively, then an n-copula C exists such that, for all $\mathbf{x} \in \mathbb{R}^n$,

$$H(x_1,x_2,\ldots,x_d) = C(F_1(x_1),F_2(x_2),\ldots,F_n(x_n)),$$

being C unique if $F_1,F_2,\ldots,F_n$ are continuous, as is the case in this project.

This theorem successfully proves that the dependence structure can be defined based on marginal distributions instead of directly on the random variables. Such a result is extremely useful for modeling efforts, since the former can be often easily obtained, conversely to the latter. Moreover, any arbitrary distribution can be used provided that it can be transformed into a uniform distribution. For effect, it is only required for the CDF F_X of the original, arbitrary distribution X to be monotone. If so, $Y = F_X(x) = \mathrm{P}(X \leq x)$, results in $Y \sim \mathcal{U}(0,1)$ as stated by the Probability Integral Transform (PIT) and proven as follows,

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(F_X(x) \leq y) = \mathrm{P}(X \leq F_X^{-1}(y)) = \mathrm{P}(X \leq F_X^{-1}(F_X(y))) = \mathrm{P}(X \leq x) = F_X(x) = y.$$

Finally, by taking the generalized inverse distribution function of the marginals $F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}$, follows the corollary of Sklar's theorem which concludes the linkage of the margins stemming from arbitrary distributions to a copula dependence structure,

$$C(u_1,\ldots,u_n) = H(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)).$$

For the particular case of the Gaussian copula, the corresponding copula function C_ρ^G of an n-dimensional standard normal distribution with correlation matrix ρ is denoted as

$$C_\rho^G(u_1,\ldots,u_n) = \Phi_\rho^G(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)),$$

where Φ stands for the standard univariate normal distribution function and Φ_ρ^G its multivariate equivalent. Since the obtained distribution is still essentially a Gaussian structure, just as in standard Gaussian distributions, but accepts arbitrary distributions as marginals, it is referred to by some authors as meta-Gaussian [57].

2.2.1.2 Conditional Sampling

There are several ways to sample values from a known copula, although most of these methods are based on the same principle, the inverse transform sampling (ITS). This pseudo-random number sampling method supports data generation from an arbitrary distribution, despite its dimension, using the respective CDF. In fact, the reasoning behind this methods expands from the previously used PIT, by using the knowledge of CDFs being uniformly distributed and plugging into its inverse function F^{-1} a single, pseudo-randomly generated probability $u \in [0,1]$ to obtain the corresponding sample $x = F^{-1}(u)$. Copula sampling methods mostly differ in which CDFs the ITS is applied to. The simplest, quickest and most used way to do so is the random sampling method, which easily takes the inverse of the given copula's CDF and directly proceeds with ITS. In the scope of this work, a more complex and computationally expensive sampling method was used - conditional sampling or conditional draw.

Firstly, note how by giving value 1 to k parameters of an n-dimensional copula C^n, an $(n-k)$-dimensional copula C^{n-k} is obtained. Next, the general way to use a given copula $(U_1,\ldots,U_n) \sim C$ to obtain sample u_k from U_k conditional on previously obtained values of $u_1,\ldots,u_{k-1}$ is by first defining the conditional distribution

$$\begin{aligned} G_k(u_k|u_1,\ldots,u_{k-1}) &= \mathrm{P}(U_k \le u_k | U_1 = u_1,\ldots,U_{k-1} = u_{k-1}) \\ &= \frac{\partial_{u_1,\ldots,u_{k-1}} C(u_1,\ldots,u_k,1,\ldots,1)}{\partial_{u_1,\ldots,u_{k-1}} C(u_1,\ldots,u_{k-1},1,\ldots,1)}, \end{aligned}$$

which, given the first observation, can be rewritten as

$$G_k(u_k|u_1,\ldots,u_{k-1}) = \frac{\partial_{u_1,\ldots,u_{k-1}} C(u_1,\ldots,u_k)}{\partial_{u_1,\ldots,u_{k-1}} C(u_1,\ldots,u_{k-1})}$$

and, finally, following ITS and sample according to $u_k = G_k^{-1}(u_k'|u_1,\ldots,u_{k-1})$, with u_k being a realization of U_k and u_k' a pseudo-randomly generated probability value [58]. Indeed, this same procedure can be used to iteratively obtain estimations for every variable in the simulation space. Following the example of full sea-state dynamics simulation in [58], to model four variables (height H, storm duration D, time between successive storm phases I and storm wave direction A) with a 4-dimensional copula C_{HDIA}, the first variable H must be given, the next variable can be simulated with the value for H and with the 2-dimensional copula C_{HD}, succeeded by the simulation of variable I with obtained H and D and copula C_{HDI} and, finally, simulation of the last variable A with obtained H, D and I and copula C_{HDIA}. However, for substantial simulation spaces, this becomes impractical in the computational and notational point of view. Thankfully, simplifications are possible and extensively used in computational simulations. The same authors

of the example above, suggest one simplification where only 2-dimensional copulas are needed to perform CS on any multivariate distribution. Following the same example, multivariate copulas can be decomposed into bivariate copulas, resuming the required previous knowledge to C_{HD}, C_{HI}, C_{HA}, C_{DI}, C_{DA} and C_{IA}. To demonstrate this, the authors go through the iterative process of CS conventionally until the simulation of I with obtained H and D variables and copula C_{HDI},

$$G_3(u_3|u_1,u_2) = \frac{\partial_{u_1,u_2} C(u_1,u_2,u_3)}{\partial_{u_1,u_2} C(u_1,u_2)},$$

at which point they simplify the numerator as

$$\partial_{u_1} C_{H,I}(D_2 C_{HD}(u_1,u_2), D_1 C_{DI}(u_2,u_3)),$$

solving to

$$D_1 C_{HI}(D_2 C_{HD}(u_1,u_2), D_1 C_{DI}(u_2,u_3)).D_{12} C_{HD}(u_1,u_2).$$

Using the same decomposition by partial differentiation, further simulation of the remaining variables will always come down to a point where only 2-dimensional copulas exist in the expression, as shown.

2.2.1.3 Simulation Space Dependence

Calculating the correlation matrix can differ based on which correlation criteria is used. In some scenarios, linear correlation is a more effective metric to depict associations among data than rank correlation, albeit the opposite case might also be true. One must, therefore, heed the type of data being used and to understand how commensurable dependence and correlation are.

Pearson's ρ_{p} is a unitless scalar in between -1 and 1. Taking as an example two random variables X and Y, it can be derived by the covariance $\mathrm{cov}(X,Y)$ divided by the product between both distributions' standard deviations σ_X and σ_Y, as follows

$$\rho_{\mathrm{P}} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

The unitless property makes the correlation measurement more robust and intuitive than purely using covariance (given different units or value ranges can be taken by the variables) and is guaranteed by the property $\mathrm{cov}(X,Y) \leq \sigma_X \sigma_Y$. Since $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ and $\mathrm{cov}(X,X) = \mathrm{cov}(Y,Y) = 1$ correlation matrix is both symmetric and equal to the unit diagonally.

Kendall's τ and Spearman's ρ_{s} are rank coefficients. To calculate these coefficients, variables are transformed into ranks, where each value takes an integer value according to how high they rank in its variable value range. Spearman's ρ_{s} is simply a modified formula of Pearson's ρ_{p} with rank values instead of the real variable values, while Kendall's τ calculates a ration between concordant and discordant pairs (concordant when variable pairs score equal ranks) between variables. For example, with the same two variables X and Y with n samples each, Kendall's τ equates to

$$\tau = \frac{\#(\text{concordant pairs}) - \#(\text{discordant pairs})}{\binom{n}{2}}.$$

Rank correlation coefficients are calculated based on monotonic relationships independently of the marginal distributions, which makes them more appropriate when a nonlinear transformation is applied since linear correlation is not preserved [59]. For that same reason, while dependence

between variables described by elliptical distributions are sufficiently well represented by Pearson's ρ_P, nonelliptical distributions are better represented by a linear rank coefficient [60, 61]. One important aspect to bear in mind is that while independence implies uncorrelation, uncorrelation does not imply independence. This is a result of a particular correlation metric not fully depicting how dependent two variables might be, but rather how dependent they are with regards to what specific association the metric is designed to measure.

This project has started by using Pearson's ρ_p for the sake of simplicity and since the Gaussian copula is an elliptical distribution, although different metrics were tested to evaluate which one works better given the data framework. On Section 5.1.1.2, Pearson's ρ is compared to Spearman's ρ_s and Kendall's τ in order to understand whether or not linear correlation is enough to define parameter dependence.

2.2.1.4 Limitations and Usefulness

Despite being so widely used, Gaussian copulas are considered to be too simplistic for some simulation scenarios. As an example, they were accused to play a major role in the 2007-2008 financial crisis since market data returns were soon after discovered to be heavy-tailed, skewed and to have clustered extremes, among other nuances which surpassed the copula model's ability to describe dependence at a time when most professionals used it as a black-box, one-size-fits-all model in the absence of a systematic way to choose between existing copula models [62]. Since then, not only new copula models with different specialties have been made available, such as the t-student copula, Gumbel copula and the Archimedean copula, an emphasis was put on Extreme Value Theory, which measures the risks of extreme deviations from the distribution median (such as Lisbon's 1755 earthquake) and models were extended with stochastic correlations and time-varying copulas to enhance the models' scenario coverage [63].

Notwithstanding, Gaussian copulas remain widely in use, being extremely useful when adequate. In fact, the studies in [62, 64] show evidence supporting the theory that a good parameter estimation when fitting a copula model to data improves model accuracy more than switching copula type altogether, possibly evicting the Gaussian copula from blame related to the crisis. In the scope of this project, it is thought that Gaussian copulas are, indeed, recommended since most of marginal distributions assume Gaussian shapes themselves, avoiding some additional concerns about possible heavy tail dependence, and are computationally lightweights with small storage requirements, according to [57].

2.2.2 Markov Chains

A process is said to be Markovian, when its future can be predicted based on its present. Despite having first been introduced in the beginning of the last century, Markov processes are still frequently used in statistical modeling, control engineering, predictive algorithms and state estimators. In fact, it was repeatedly proven that commonly used concepts such as Brownian motion and random walks are also based on Markov processes in continuous and discrete-time, respectively. In particular, Markov chains allow describing a multi-state system and how it evolves, according to state-transition probabilities [65].

2.2.2.1 Mathematical Definition of Markov Chains

Markov chains are described by a state-space $Q = q_1, \ldots, q_N$, where N is the number of states the system can adopt, a transition matrix,

$$A = \begin{bmatrix} a_{1,1} & \ldots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \ldots & a_{N,N} \end{bmatrix},$$

where $a_{i,j}$, with $\forall i, j \in \{1, \ldots, N\}$, is the probability of transitioning from state q_i to state q_j with the restriction of $\sum_{j=1}^{N} a_{i,j} = 1 \forall i \in \{1, \ldots, N\}$ and an initial probability distribution $\pi = \pi_1, \ldots, \pi_N$, where π_i, with $\forall i \in \{1, \ldots, N\}$ is the probability of the system to start at state q_i, given that $\sum_{i=1}^{N} \pi_i = 1$.

This structure illustrates the previously mentioned Markov assumption about the future state only depends on the present, not the past. In mathematical terms

$$p(q_i = a | q_1, \ldots, q_{i-1}) = p(q_i = a | q_{i-1}).$$

Under this premise, after T transitions, the probability vector S_T with respect to each of the states can be calculated starting with the initial probability distribution π by

$$S_T = \pi A^T.$$

Thus, keeping track of the probabilities in $S_t, \forall t \in \mathbb{Z}^+$ makes it possible for states to be sampled after t transitions, building a state-evolution train.

When

$$\lim_{t \to \infty} S_t - S_{t-1} = \mathbf{0}_{1,N},$$

where $\mathbf{0}_{1,N}$ is the zero matrix with dimensions $1 \times N$, it is said that the system has reached a stationary state and the vector S which obeys the equation $S.A = S$ is called the stationary matrix. It is the final probability configuration to which the system's states converge to. This property is not a characteristic of every Markov chain, but if $A^t, \forall t \in \mathbb{Z}^+$ has only positive elements, the Markov chain is called regular and complies with this property, always converging to a steady-state with a defined stationary matrix. In this case a limiting matrix of the form

$$\lim_{t \to \infty} A^t - A^{t-1} = \mathbf{0}_{N,N}$$

is also reached. Another type of converging Markov chain behavior can be observed in absorbing Markov chains, where there is at least one absorbing state in the state-space, out of which the system cannot exit after it settles, and where it is possible to go from a non-absorbing state to an absorbing state in a finite number of steps. In this case, taking increasingly high powers of the transition matrix A will also result in a limiting matrix.

2.2.2.2 Limitations and Usefulness

These stochastic representations of systems can sometimes not be adequate, when future states depend on a more complex association between past and present states. Moreover, popular Markov chains were shown to take exponentially long to converge to a steady-state in many real-world applications, when convergence is in fact intended. However, it is a very elegant, computationally efficient and useful way of describing simple multi-state systems. This lightweight property also

makes it efficient for state sampling and state evolution calculation, which, in most use scenarios, are not related to systems converging to steady-states, avoiding the computational time problem stated above. Furthermore, it provides a solid description of a system which can then be pivoted into an enhanced version of the Markov chain - the hidden Markov chain - capable of analyzing an observation vector (whose observations are tied to a hidden state according to specific probabilities) and calculating the probability of underlying state vectors.

2.3 Concepts in Artificial Intelligence

A brief overview of DL and the methods and concepts used to understand and design the CNN is presented in this Section.

2.3.1 Deep Learning

The umbrella term AI designates a number of methods subdivided according to the nature of the problems it tackles, from finance and economics to medicine and healthcare. Some of the AI methods applied to one subdomain are not transferable to others, but most of them share the algorithmic paradigm of ML. This area of study builds mathematical models capable of learning how to perform a specific task from experience in detriment of explicitly programming an algorithm to do it. As such, ML provides the tools for an algorithm to train with sample data in order to improve its performance at the given task when faced with new, unseen data [66]. A multitude of frameworks have been developed to this extent and choosing one depends on the nature of the task at hand. As this thesis works towards heart cycle detection and classification, only binary and multi-class classification tasks will be addressed. Artificial neural networks (ANN), which have been extensively researched for its proficiency at understanding high-level abstractions based on a strong pattern recognition ability, will be the ML method we will be covering in this Section.

The building blocks of an ANN are artificial neurons. Each outputs the dot product of an input vector $\mathbf{X}^{[1]}$ (or $\mathbf{H}^{[i]}$ as we will see later on) and its respective weight vector $\mathbf{W}^{[i]}$, taking inspiration from the biology of the neuron. Although one simple unit is only capable of modeling a first-order linear function, applying a nonlinear activation function to its output and stacking multiple artificial neurons, now referred to as nodes, into successive layers allows highly nonlinear functions to be simulated. In fact, the more layers the ANN has, the higher the complexity and abstraction it is able to attain. This was thoroughly explored for image recognition tasks, ultimately founding DL as a subdomain of ML in allusion to the high number of hidden layers used in such ANNs [67]. Figure 2.15 shows an example of a binary classification ANN captioned with the standard mathematical notation. With it, one can easily express the process of obtaining the network's output as follows,

$$\hat{Y} = \sigma \circ \left(\mathbf{W}^{[2]} \times R \circ \left(\mathbf{W}^{[1]} \times \mathbf{X}\right)\right),$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ and $R(x) = \max(0, x)$ are the sigmoid and rectified linear unit (ReLU) activation functions (with an arbitrary input x), respectively, which are also used in the proposed architecture, and the weights of the matrices $\mathbf{W}^{[1]}$ and $\mathbf{W}^{[2]}$. This calculation process is called forward propagation and, for binary classification, its result $\hat{Y}$ is the probability of a data sample $\mathbf{X}$ being positive, or classified with 1. For multi-class classification, the ANN would have multiple outputs

- each is capable of binary classification and, through their combinations, the network is capable of classifying $\mathbf{X}$ in 2^M classes. As is the case of the current project, in supervised learning, the correct classification Y of each sample is known and used to evaluate $\hat{Y}$ by means of an error or loss function, in this case, the cross entropy (CE) function

$$CE = -\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij} \ln \hat{Y}_{ij},$$

where N is the number of data samples and M the number of outputs of the ANN.

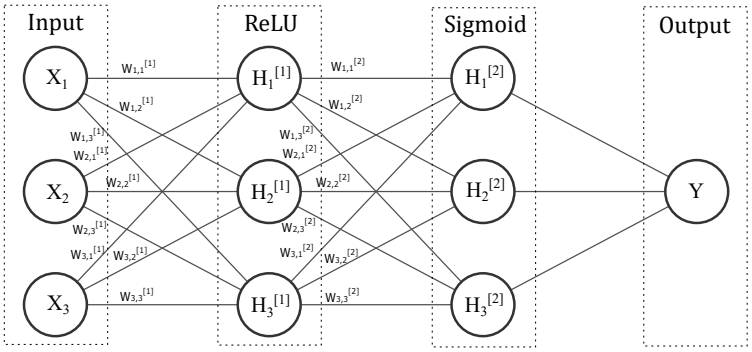

Figure 2.15: Example ANN with 3 input nodes in the input layer, 2 hidden layers, with 3 hidden units each, and 1 output node. Inputs are denoted by the matrix $\mathbf{X}$, weights by matrix $\mathbf{W}$, hidden units outputs with $\mathbf{H}$ and the output with Y. Superscripts indicate the layer and subscripts units in that specific layer.

ANNs are trained by iteratively classifying data samples and updating the weights in order to minimize the error function by means of backward propagation. Most commonly, the total available dataset is divided, without overlapping, into a training set, a validation set and a test set, within which samples are grouped in batches. The training set is used to update the weights via optimization. Typically, the same training set is iterated through several times, batch by batch- each time being called an epoch. After forward propagation is performed on a batch, the chain rule of derivation is used to recursively calculate partial derivatives of the CE $\frac{\partial CE}{\partial W_{ij}^{[k]}}$, where $i, j = \{1, 2, 3\}$ and $k = \{1, 2\}$ with respect to each weight of the matrices, starting at the last layer $\mathbf{W}^{[2]}$ and moving backwards to the first $\mathbf{W}^{[1]}$. Depending on the chosen optimization algorithm, these derivatives are used differently: in stochastic gradient descent (SGD), CE is obtained from a matrix $\mathbf{X}$ containing every sample in a batch and weights are updated according to

$$W_{ij}^{k} = W_{ij}^{[k]} - \alpha \frac{\partial CE}{\partial W_{ij}^{[k]}}; \tag{2.1}$$

in adaptive moment estimation (Adam), calculations are, too, performed per batch, but weights are updated as

$$W_{ij}^{[k]} = W_{ij}^{[k]} - \alpha \frac{V}{\sqrt{S}}, \tag{2.2}$$

where V and S are running-average terms calculated as

$$V \propto \beta_1 V - (1-\beta_1)\frac{\partial CE}{\partial W_{ij}^{[k]}}$$

$$S \propto \beta_2 S - (1-\beta_2)\left(\frac{\partial CE}{\partial W_{ij}^{[k]}}\right)^2,$$

so as to represent a momentum- and RMS-approaches to the error function minimization problem with the first and second equations, respectively. The terms α and β are constants, being the former referred to as learning rate. Both algorithms were tested in this project given their pros and cons: SGD is a stabler but slower method, while Adam allows convergence to occur faster by finding more accurate minimization directions with the partial derivative running-average approach, provided the constants are not so high the updates overshoot the CE's minima [68]. The optimization step is repeated each epoch and the validation set solely undergoes forward propagation to produce a validation error so the user understands of how well training is progressing. Finally, after training is complete, forward propagation is also performed on the test set and results are used to describe the model's performance. To obtain a better estimate of the model's true performance, it is common to divide the initial dataset into sub-datasets and group them together to form the training, validation and test sets, going through all the steps above described. The process is repeated a number of times, differently grouping the sub-datasets every iteration, yielding the same number of performance metrics. This method is known as cross-validation.

2.3.2 Convolutional Neural Networks

Largely employed in image classification, ANNs would either receive each individual pixel as inputs or manually crafted features. With the introduction of CNNs, seen in Figure 2.16, feature design became automatic and part of the learning process, thanks to a series of layers added to the input of an ANN, now called fully-connected network (FCN) in the context of CNNs. Notwithstanding, these layers are differently connected. Instead of connecting each pair of neurons in consecutive layers, local connections assign different regions of the input to different regions (features) of the CNN, similarly to how visual cortex neurons respond to a specific region of the visual field. The area of the input affecting any given feature of a CNN is designated by its receptive field. The best features are extracted as filters (new term for weight matrices in CNNs) convolute over the previous layer, as Figure 2.17 shows, and are updated to identify the most useful patterns (feature maps) to minimize the error function. Each layer can produce a number of distinct feature maps by applying different filters to its input [69]. Typically, filters have higher dimensions near the CNN's input layer, covering larger regions of its input and producing a smaller number of feature maps, where lower-level patterns are learned. As the FCN is approached, filters become smaller and feature maps more numerous, meaning learned patterns are higher-level abstractions of the original input.

After ANNs had been established as gold standard for their high accuracy (around 95%) in classification of the popular MNIST database [71], CNN architectures started being used with clear improvements, some even attaining a perfect score [72]. Besides, since each feature map is designed by a single filter, it adds up to $D \times D \times F$ weights to be learned per layer with F D-dimensional filters, which, comparatively to an ANN, in which each layer is fully connected to the prior layer, constitutes a significant complexity reduction. Moreover, given the flexibility of CNNs

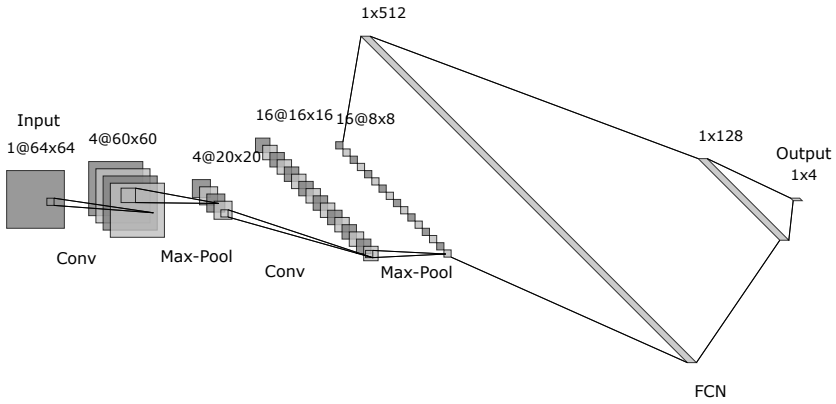

Figure 2.16: Example of a CNN used in computer vision applied to images. The input is a grayscale image with dimensions 64×64. Following convolutional layers abide to the same notation, where, for example, $4@60 \times 60$ means a layer originated by 4 filters with 60 samples of width and height performing convolution over the image to produce 4 feature maps. An also common notation is to describe layers as 3D tensors, where the sizes of the filter are the number of rows and columns, by this order, and the number of feature maps the tensor's depth, translating into $60 \times 60 \times 4$. At the end, a regular FCN, represented by vectors denoting the artificial neurons, as seen in Figure 2.15.

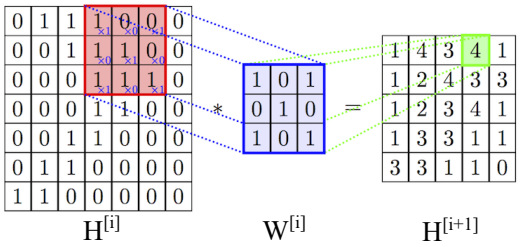

Figure 2.17: Example of a convolution operation of a filter $\mathbf{W}^{[i]}$ on a hidden feature map $\mathbf{H}^{[i]}$, resulting on the feature map of the following hidden layer $\mathbf{H}^{[i+1]}$. Adapted from [70].

in identifying patterns and representing them abstractly, they can sometimes be pre-trained for one known task and learn useful attributes transferable to a new one, in which case the network, or parts of it, can be tuned in order to produce good results.

2.3.3 Model optimization

The model's overall performance can benefit with changes ranging from training, to the optimization algorithm or the network architecture. One of the firstly tackled performance aspects was the bias and variance paradigm - a low bias, high variance model is capable of learning training data to an extremely profound extent, such that it lacks the ability to extrapolate and perform reasonably in unseen data, also called overfit, while a high bias, low variance model does not suffer from this problem, because it lacks the ability to learn training or unseen data to a reasonable extent. Striking the right balance between both optimizes results and, with that in mind, a set of techniques, designated by regularization, are now frequently used to elevate CNN to performant-level architectures [67]. Dropout, described as the stochastic disconnection of a number of layer connections during the training phase by nullifying their weights, demotivates co-dependence among consecutive nodes, forcing the network to not overly rely on any one feature. Additionally,

L1 and L2 regularization add a term to the error function, which is

$$\lambda_1 \sum_{l=1}^{L} \|\mathbf{W}^{[l]}\|_1 \text{or}$$
$$\lambda_2 \sum_{l=1}^{L} \|\mathbf{W}^{[l]}\|_2^2,$$

respectively, where L is the number of layers in the network, $\|\cdot\|_1$ is the element-wise sum of the elements' norm in a matrix, $\|\cdot\|_2^2$ the element-wise sum of the squared elements in the matrix and the λ terms are associated constants. While the former motivates weights to be more sparse, the latter tends to make them more similar. In general, both these dropout and L1 and L2 regularization bring weights closer to zero, leading to better information distribution inside the network and avoiding certain regions from becoming too influential, countering overfitting.

Another performance aspect thoroughly studied is the concept of invariance. A model which respects invariance, is able to recognize a pattern regardless of how it is presented in the input, performing much better in real-world scenarios, where seldom do patterns present without unwanted distortions. Batch normalization is a technique to improve invariance by means of reducing the amount by which nodes vary (covariance shift) [67]. When using a batch of training samples, inputs of every layer are assumed to follow a normal distribution and their node values are standardized accordingly, by subtracting the mean and dividing by the standard deviation, ultimately, following a similar distribution. It also adds two trainable parameters to the model, which correspond to yet another shift and scaling operation to mitigate the effects of the first one, in case it worsens classification. By reducing covariance shift, batch normalization accelerates learning and facilitates different presentations of the same pattern to be identified in the data. Moreover, having versatile training data to teach the model the most scenarios possible pertaining to the desired task is imperative for invariant performance. Such operations include data multiplication, class balancing, noise addition, shifting and scaling, among others.

Before being used, a set of parameters related to the optimization algorithm and the architecture must be initialized. These are, respectively, optimizer and model specific hyperparameters and choosing the right combination bears significance to the algorithm's success. In this project, two optimization routines were used: grid search and Bayesian optimization. Firstly, despite the chosen routine, evaluation metrics must be defined in order to select which hyperparameter set works best with the given model. Metrics can range from validation error, to accuracy, sensitivity (SE) and positive predictive value (PPV), even running time, including combinations between them. In grid search, a P-dimensional grid is defined by R^P discrete points, where P is the number of hyperparameters and R the number of discrete values (numerical or categorical) it is allowed to take, making the points of the grid all the possible value combinations, or search-space. The algorithm is ran with values from every possible combination and the one yielding the best results is chosen for further applications. Despite being very simple to implement, grid search takes a long time when P is high, without any improvements being made as new combinations are tested since they are independent of one another. On the other hand, Bayesian optimization is better suited for tuning numerous hyperparameters since it uses a surrogate probabilistic model $\hat{f} = \mathrm{P}(s|p)$ mapping hyperparameter configurations p to a metric score s and minimizes the objective function f. Each iteration, the configuration which performs better on $\hat{f}$, taking into account previous results, is used in f. To each hyperparameter, a value range is set, continuous or discrete,

with an assumed probability distribution. In this project, a tree-structured Parzen estimator was used to define the surrogate function. From previous iterations, results are used to build $P(p|s)$ which is divided into two sub-distributions $l(p)$ for $s < t$ and $g(p)$ for $s \geq t$, where t is a score threshold. From Bayes' theorem, an expected improvement function (EI) is defined [73] as

$$\mathrm{EI}(p) = \int_{-\infty}^{t} (t-s) \cdot \hat{f}\, ds,$$

which simplifies to

$$\mathrm{EI}(p) \propto \frac{g(p)}{l(p)}.$$

In a nutshell, choosing a configuration p, which maximizes the ratio $g(p)/l(p)$, has a higher chance of yielding a higher score s in the next iteration since picking among configurations in distribution $g(p)$ is most likely to result in higher scores than $l(p)$, which would probably yield lower scores than threshold t. This method is much faster as more emphasis is put on finding the most likely candidate, which takes less time than, effectively, testing it [74].

In supervised learning, a lot of significant metrics stem from the confusion matrix, where the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are counted by comparison between predicted classifications and the real labels, as Table 2.1 exemplifies for a simple detection task. In this thesis, SE and PPV are used. SE, is the fraction of TP out of all the actual positive samples,

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

informing about the capability of the algorithm not to miss an actual positive sample. PPV is the fraction of TP out of all positively predicted samples,

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$

denoting how reliable a positive result is. Finally, accuracy is a well-known, but broader measurement which expresses the number of correctly predicted samples over the total number of samples by

$$\mathrm{accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}.$$

Although it works well for balanced classes and when performing well on one or more specific classes is not relevant, which is rarely the case in medical data.

Table 2.1: Example of a confusion matrix for a binary classification, or detection problem. Each result adds up to the count of the corresponding box - if the prediction and real label match, a class 1 is considered a TP, while a class 0 is a TN. Otherwise, a class 1 is a FP, while a class 0 is a FN.

		Real	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Chapter 3

State of the Art

3.1 Cardiorespiratory Signal Modeling

The range of available generative algorithms for unimodal signal modeling is much greater than for multimodal modeling and most cited works cardiorespiratory simulation specialize in a single modality. Bearing this in mind, in this section, both types of approach will be presented.

3.1.1 Unimodal Models

One of the most influential and cited generative models for ECG signal simulation was first introduced in [29]. A great deal of the currently available dynamical generative algorithms aim at building upon it, by adapting it to other signal modalities and updating it to take into account more physiological effects. Research on generative models has focused mainly on deterministic, dynamical ways to describe physiological phenomena, which lead to a considerable amount of useful algorithms for clinical modalities such as ECG, BP and RESP, quite especially for the first two. Despite non-dynamical models not having been a major research focus in the field, they have been gaining attention due to how they can produce convincing results for emerging, non-clinical modalities such as cECG, remote or reflective PPG and BCG. This Section, therefore, segregates both model types related to each group of modalities.

3.1.1.1 Dynamical Models

In [29], a set of dynamical equations in the form of three ordinary differential equations in a 3-dimensional Cartesian system is used. The equations are enunciated as follows

$$\begin{cases} \dot{x} = \alpha x - \omega y \\ \dot{y} = \alpha y - \omega x \\ \dot{z} = \sum_{i \in \{\mathrm{P,Q,R,S,T}\}} a_i \Delta\theta_i \left(-\frac{\Delta\theta_i^2}{2b_i^2} \right) - (z - z_0), \end{cases} \tag{3.1}$$

where $\alpha = 1 - \sqrt{x^2 + y^2}$, $\Delta\theta_i = (\theta - \theta_i) \mathrm{mod} 2\pi$ and $\theta = \mathrm{atan2}(y, x)$. The first two equations introduce a circumference inserted in the (xy)-plane rotating at angular velocity ω suffering deviations on

the z-axis described by the Gaussian mixture model (GMM) - a summation of individual Gaussian functions - corresponding to the third equation and a baseline wonder term z_0. Such a system builds a typical ECG waveform cycle by cycle, whose length defined by the varying angular velocity ω, emulating HR, and the same standard PQRST-complex, which is determined by the GMM trajectory and manually set parameters, shown in Table 3.1. These parameters are affected by a baseline wonder z_0, represented by a sinusoidal wave with empirically determined amplitude and variable RF in order to mimic respiratory amplitude modulation. Additionally, a data-driven component was included to the model to involve arrhythmogenic effects from RSA and Mayer waves by taking the inverse Fourier transform of the power spectrum of a real RR-tachogram, as introduced in Section 2.1.2.3, to obtain a numerical time series of RR-interval lengths $T(t)$, defining the angular velocity as

$$\omega(t) = \frac{2\pi}{T(t)}.$$

Table 3.1: Set of parameters for Equation 3.1 used to obtain synthetic signals shown in Figure 3.1. Adapted from [29].

Index (i)	P	Q	R	S	T
Time (secs)	-0.2	-0.05	0	0.05	0.3
θ_i (radians)	$-\frac{1}{3}\pi$	$-\frac{1}{12}\pi$	0	$\frac{1}{12}\pi$	$\frac{1}{2}\pi$
a_i	1.2	-5.0	30.0	-7.5	0.75
b_i	0.25	0.1	0.1	0.1	0.4

Results are shown below in Figures 3.1 and 3.2, where a considerable resemblance to a typical ECG waveform can be observed besides a realistic modulation of cardiorespiratory coupling. Furthermore, a positive linear correlation between the QT-interval and the RR-interval is present in the synthetic ECG, as is also the case in real waveforms. In conclusion, the authors emphasize the possibility of the model being used to develop and test biomedical signal processing algorithms, including noise removal techniques and employed in QT-interval calculation correction. Changing the GMM parameters in Table 3.1 also successfully generated non-standard waveforms. This technique was the signal generator around which one of the follow-up papers, where a benchmark for T-wave alternans detection [75], was developed.

Meanwhile, the model has undergone several updates. A two-piece installment, described in [76, 77], included the possibility of obtaining BP waveforms using the same GMM introduced in Equation 3.1, RESP and a new method of generating an RR-interval sequence taking into consideration short- and long-term oscillations in order to produce 24 hour synthetic recordings of ECG, BP and RESP. Having already considered short-term effects in RR-interval variation in the previous version, long-term oscillations - circadian activity and sleep cycles - were added to the array of stationary states. States, their duration and transitions among them were modeled with empirically determined distributions [78] each with its own model function and ultimately assuring transition to other states by also modeling the amount of time spent in each one [78, 79]. As for the BP waveform, the same dynamical equations for ECG were used, but with different parameters, as shown in Table 3.2, with an additional offset term $\delta\theta = \left(1 - \left(\sqrt{\frac{60}{HR}}\right)/2\right)/2\pi$ to the PTT (calculated by $\theta_Q^{BP} - \theta_R^{ECG}$) to model changes in PTT motivated by HR as reported by [80]. Finally, the RESP signal is obtained by once again using the inverse Fourier transform, similarly

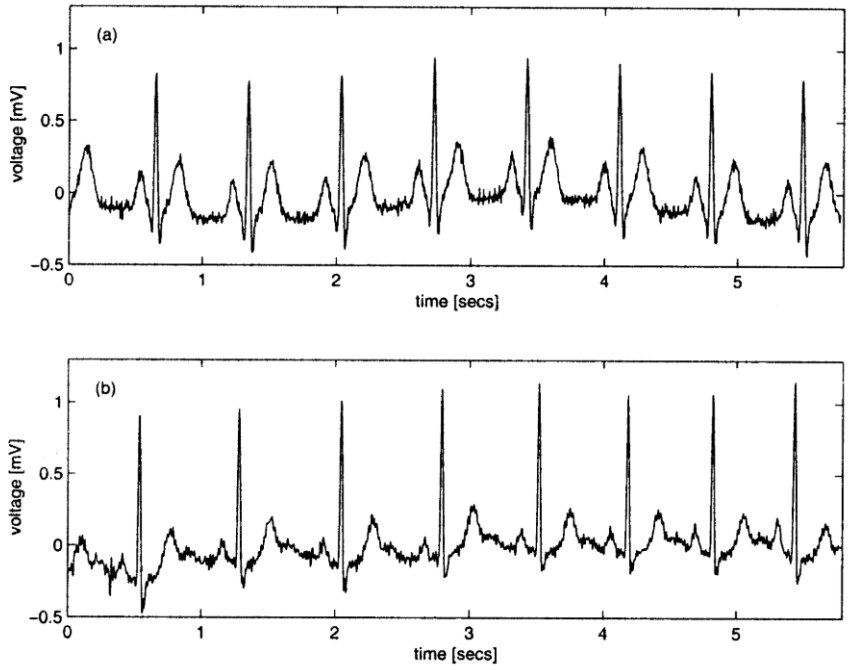

Figure 3.1: Comparison between synthetic (above) and real (below) signals. Strong morphological resemblance can be observed despite the real counterpart showing higher complexity, as its baseline and individual waveforms are more variant. Adapted from [29].

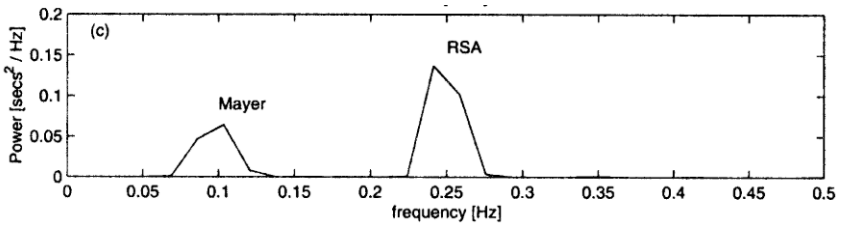

Figure 3.2: RR-tachogram frequency spectrum of the synthetic signal. Results show large peaks corresponding to RSA and Mayer waves, as expected. Adapted from [29].

to what was used to obtain the first iteration of synthetic RR-interval time series, except only the HF component was considered.

Table 3.2: Set of parameters for Equation 3.1 used to obtain BP synthetic signals shown in Figure 3.3. Only the components corresponding to R, S and T waves are used, while P and Q waves are set to null amplitude. Adapted from [76].

Index (i)	P	Q	R	S	T
Time (secs)	-0.21	-0.01	0	0.03	0.22
θ_i (radians $\times \frac{\sqrt{HR}}{60}$)	$-\frac{5\pi}{12}$	$-\frac{\pi}{36}$	0	$\frac{\pi}{18}$	$\frac{4\pi}{9}$
a_i	0	0	0.45	0.25	0.45
b_i ($\frac{\sqrt{HR}}{60}$)	0.25	0.1	0.3	0.5	0.3

Although waveforms resulted in realistic simulations as seen in Figure 3.3 and, especially, BP presents signal complexity within the expected range, as implied by Figure 3.4 and reported by [81], the obtained RR-interval time series showed that there are yet improvements to be made

on a very long- and very short-term level. Moreover, an erroneous decorrelation between HR and LF/HF was found between deep and light stages. These results show the introduced changes were still not thorough enough to produce a convincing enough extensive simulation.

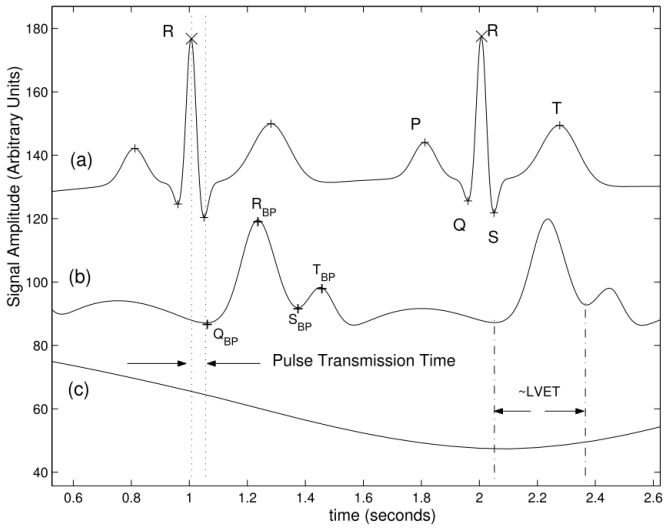

Figure 3.3: Plots of ECG (a), BP (b) and simulated RESP (c). Annotated fiducial points on the BP wave are subscripted according to the modality in order to express how different components of the GMM are used in ECG and BP. Adapted from [76].

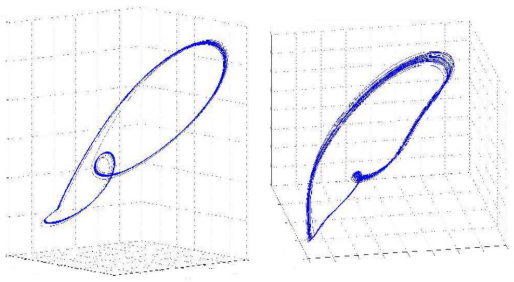

Figure 3.4: Ruelle-Takens reconstructions of the real (right) and synthetic (left) BP waveforms. An estimated optimal step delay of $\frac{5}{256}$ seconds and embedding dimension of 3 were used, yielding Correlation Dimension (CorrDim) around 1.1 and Approximate Entropy (ApEn) around 2.3. Adapted from [77].

Further updates to the model mainly focused on increasing its coverage of different signal modalities while still including minor effects. In [82] the initial model was used to form a single cardiac dipole vector of the form $\mathbf{d}(t) = x(t)\hat{\mathbf{a}_x} + y(t)\hat{\mathbf{a}_y} + z(t)\hat{\mathbf{a}_z}$ with modified dynamic equations

$$\begin{cases} \dot{\theta} = \omega \\ \dot{x} = \sum_{i\in\{\mathrm{P,Q,R,S,T}\}} \frac{a_i^x \omega}{(b_i^x)^2} (\Delta\theta_i^x) 2 \left(-\frac{(\Delta\theta_i^x)^2}{2(b_i^x)^2} \right) \\ \dot{y} = \sum_{i\in\{\mathrm{P,Q,R,S,T}\}} \frac{a_i^y \omega}{(b_i^y)^2} (\Delta\theta_i^y) 2 \left(-\frac{(\Delta\theta_i^y)^2}{2(b_i^y)^2} \right) \\ \dot{z} = \sum_{i\in\{\mathrm{P,Q,R,S,T}\}} \frac{a_i^z \omega}{(b_i^z)^2} (\Delta\theta_i^z) 2 \left(-\frac{(\Delta\theta_i^z)^2}{2(b_i^z)^2} \right), \end{cases}$$

where the variables stand for the same values as in Equation 3.1. However, the first equation is now capable of defining the circular trajectory with angular velocity ω alone, while the remaining

three equations define the baseline deviations respective to the ECG modality waveform in 3 directions, corresponding to the 3 coordinates of the dipole vector $\mathbf{d}(t)$. This representation allows mimicking of orthogonal lead vectorcardiogram (VCG) signals. Multi-channel data can be generated by the function

$$\mathbf{ECG}(t) = \mathbf{H} \cdot \mathbf{R} \cdot \mathbf{\Lambda} \cdot \mathbf{s}(t) + \mathbf{w}(t),$$

where $\mathbf{ECG} \in \mathbb{R}^N$ is a N-channel vector, $\mathbf{s}(t) = [x(t), y(t), z(t)]^T \in \mathbb{R}^3$ the vector containing the three dipole components, $\mathbf{H} \in \mathbb{R}^{N\times 3}$ a body volume conductor model studied in [83], $\mathbf{\Lambda}$ a rank 3 diagonal matrix corresponding to the signal scaling for each direction, $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is the rotational matrix for the dipole and $\mathbf{w}(t) \in \mathbb{R}^N$ is the noise respective to each channel. As for different beat types, [82] and [84] have explored the option of defining different states mapping to different waveforms, such as pathological ones. In this sense, the basic waveform functions are still used with different parameters and each parameter vector is assigned to a stationary state. A Markov chain, explained in Section 2.2.2, is ultimately used to control state evolution and, thus, the occurrence of the different waveforms in the final signal. In [82], new components were inserted into the transition matrix to simulate the dependence of HR and a T-wave alternan state

$$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \quad , p(t) = \frac{\tanh[\vartheta(h(t) - h_{\mathrm{TWA}})] + 1}{2},$$

where h(t) as the instantaneous HR, $h_{\mathrm{TWA}} = 95 \pm 5$, which coincides with the usual HR at which these waveforms occur and $\vartheta = 0.2$ BPM^{-1} controls the slope of transition. Thus, instead of giving the algorithm a predefined state evolution vector in addition to the RR-interval time series vector, only the latter must be inputted, as the algorithm can decide which state comes next based on the HR.

Results of these expansions are shown in Figures 3.5 and 3.6, with convincing waveforms for different leads and states of the Markov chain. Notwithstanding, the authors suggest further addition of HR components to emulate the influence of the HR turbulence phenomenon, which describes a rapid acceleration and deceleration after an ectopic beat [85], coupled with the occurrence of ectopic states. Besides, they also consider the state transition matrix to be incomplete, as autonomic tone, sleep state and other relevant inputs should be added in order for the Markov chain module to keep up with the changes presented in previous installments of the model.

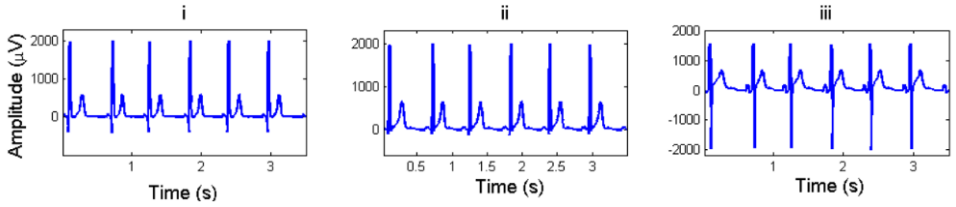

Figure 3.5: Synthetic signal from 3 of the 12-lead ECG channels. Adapted from [82].

Although being a sole example, this model was also able to inspire the simulation of a non-traditional modality, the BCG. In [86], an ensemble model of RESP and ECG generators was used to test an algorithm for non-linear separation of respiratory and cardiac components of the BCG. The authors chose not to pursue modeling of RESP resorting to the inverse Fourier transform, but rather designing a set of dynamical equations which describe the RSA and the actual RESP as oscillating signals whose frequency is time dependent and amplitudes were extracted from

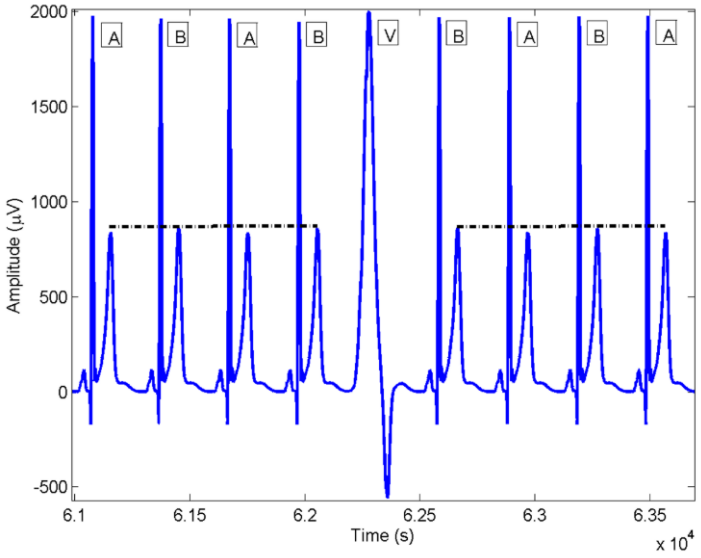

Figure 3.6: Representation of waveforms from different Markov states. Labels A and B code different T-wave alternan morphology while V codes a PVC, as presented in 2.1.4.3. Adapted from [82].

the most energetic harmonic components' coefficients of the Fourier transformation of a real respiration belt recording. The differential equations are as follows

$$\begin{cases} \dot{\theta} = \omega(t) \\ \dot{r_1} = -\omega(t)(c_{r,k_1}\sin(k_1\theta) + c_{i,k_1}\cos(k_1\theta)) \\ \dot{r_2} = -\frac{2\omega(t)a(t)}{N}\sum_{j=1}^{M_r} k_j(c_{r,k_1}\sin(k_1\theta) + c_{i,k_1}\cos(k_1\theta)), \end{cases}$$

where $\omega(t)$ is the varying RF with a fixed mean frequency, maximum deviation and time scale of variation, c_{r,k_j} and c_{i,k_j} with $j = 1,\ldots,M_r$ are the largest M_r coefficients containing 99% of the Fourier transform's energy, $a(t)$ is the varying RESP amplitude with a fixed phase offset between amplitude and frequency modulation, maximum deviation and time scale of variation and N is the length of the RESP obtained from the respiratory belt. The heartbeat equations are similar to those in Table 3.1 with an added $\omega_c^2(t)$ term multiplied by the GMM, where the RSA effect is coupled with cardiac frequency by defining $\omega_c(t) = w_{c,0}(1 + \Delta\omega_c r_1(t))$, in which $w_{c,0}$ and $\Delta\omega_c$ are the mean cardiac frequency and its maximum deviation, respectively. The ensemble signal is obtained by the summation of the RESP r_2, the ECG signal and noise. According to the authors, this was the first published generative model for BCG signals.

Resulting waveforms' morphology resembles its real counterparts, although with apparently lower amplitude LF components as one can observe in Figure 3.7. Frequency-wise, the power spectral density showed considerable overlapping between ECG and RESP components, as expected along the lines of what was explained in Section 2.1.3.2.

3.1.1.2 Non-Dynamical Models

Clinical modalities have also been a target for non-dynamical generative models. In [87], an adaptive algorithm to model a given ECG recording using only Fourier series was developed aiming at compressing real data for online processing. Using a root mean squared error (RMSE)

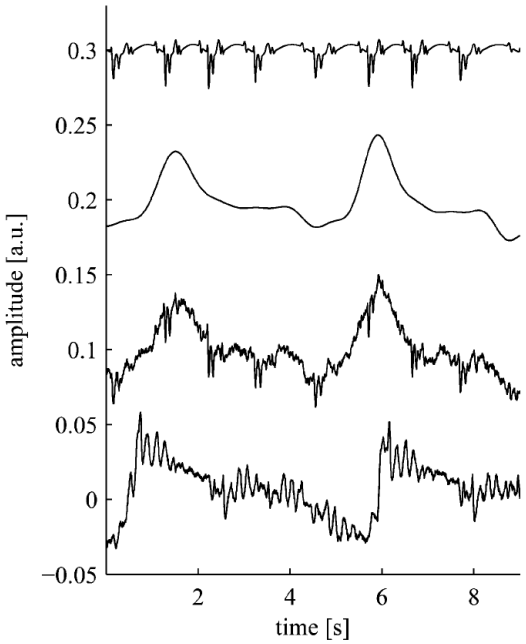

Figure 3.7: Synthesized waveforms obtained from ECG and RESP components, its merging, final BCG waveform and the real BCG (from top to bottom). Adapted from [86].

feedback, coefficients of sine and cosine components are updated every epoch proportionally to a learning rate analogous step size

$$\begin{cases} a_i(n+1) = a_i(n) + 2\mu_a \cos(i\omega n) \\ b_i(n+1) = b_i(n) + 2\mu_b \sin(i\omega n), \end{cases}$$

where a and b terms are the amplitudes and the μ terms are the step sizes for the cosine and sine components, respectively. To avoid the fundamental frequency identification step, a difference equation with initial conditions was used to replace both sinusoidal terms in the update equations

$$\begin{cases} a_i(n+1) = a_i(n) + 2\mu_a \left[2\cos(\frac{2\pi i}{P}) y_a(n-1) - y_a(n-2) \right] \\ b_i(n+1) = b_i(n) + 2\mu_b \left[2\cos(\frac{2\pi i}{P}) y_b(n-1) - y_b(n-2) \right], \end{cases}$$

where the expression inside square brackets equates to $y(n)$ and $y(-1)$ and $y(-2)$ are cosine components with a fundamental frequency empirically deemed to produce reasonable results for multiple ECG recordings. After 10000 epochs and a step size of 10^{-3}, final coefficients are saved and uploaded to the online algorithm which uses them to recalculate the recording signal using the above written difference equation to replace the usual $\cos(i\omega n)$ and $\sin(i\omega n)$ terms in the Fourier series.

A realistic waveform was obtained for both healthy and aortic valve stenosis patients, as Figure 3.8 represents, and an average percent RMSE of 4.29% after testing on 100 ECG recordings. Testing on unhealthy subject revealed a slightly higher percent RMSE of 6.64%. Ultimately, the achieved compression ratio was of 24.87.

Similarly, in [88], despite the usage of regular Fourier fitting, ECG recordings were divided into 500-sample segments and expressed in terms of Fourier series. Frequency scale of the determined coefficients was calculated, components corresponding to 80 Hz or higher were suppressed as a noise removal method and the final signal was recalculated with the remaining coefficients.

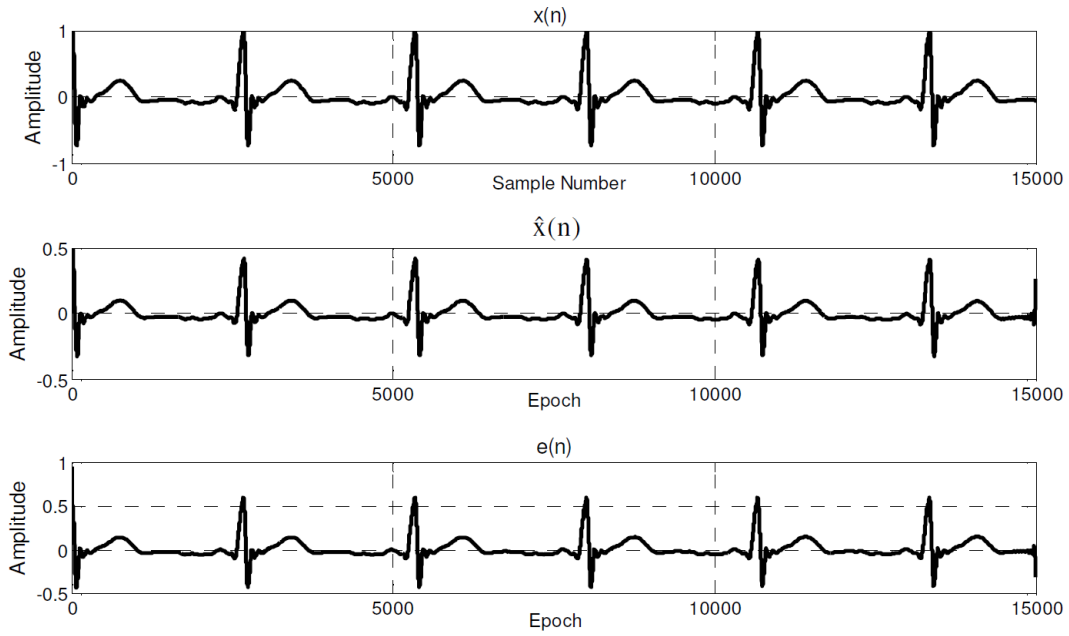

Figure 3.8: From top to bottom: original ECG, reconstructed and error signals. Despite the authors reporting a low average percent RMSE, the error for this particular recording is in the same order of magnitude of the reconstructed signal, being the latter significantly attenuated. Adapted from [87].

This approach was tested for different ECG waveform shapes and noise types (powerline, electromyographic, Gaussian and composite noise) still yielding convincing recalculated signals and significant improvements in SNR, as seen in Figure 3.9.

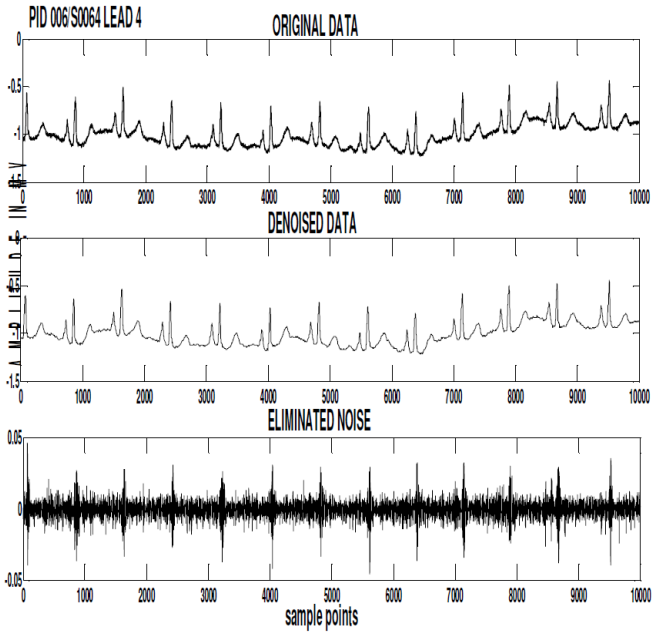

Figure 3.9: Comparison of waveforms before and after denoising of a Physikalisch-Technische Bundesanstalt (PTB) recording. Adapted from [88].

Non-clinical modalities can also be combined to generate an estimation of its clinical counterparts. As an example, in [89], cECG is addressed with a deconvolutional approach to estimate an underlying ECG source signal. The problem is tackled with a multiple-input-single-output deconvolution technique under the assumption that cardiac activity is comprised of a train of impulses subjected to delays and filters, originating the different modalities' signals. Consequently, a one-lead ECG signal is obtained from the application of a delay, filter and addition of noise to a three-channel cECG recording. In the Fourier space, the ECG Fourier transform $Y(\omega)$ is described in terms of the cECG channels' $X_i(\omega), i = 1, 2, 3$ as

$$Y(\omega) = \sum_{i=1}^{3} \vec{E}(\omega) X_i(\omega) \cdot \vec{A}(\omega) + N(\omega),$$

where $\vec{E}(\omega)$ and $\vec{A}(\omega)$ are the delay and filter vectors, respectively, and $N(\omega)$ the noise Fourier transform. A minimization problem was then formulated as follows

$$\underset{\vec{A}}{\operatorname{argmin}} \left\| \mathbf{EX} \cdot \vec{A} - \vec{Y} \right\|,$$

which ultimately allows solving the following equation for the final ECG Fourier transform

$$\vec{Y}^* = \mathbf{EX} \cdot \vec{A},$$

out of which the time domain signal $y^*(t)$ is obtained through the inverse Fourier transform.

From this approach, an estimated ECG signal was achieved with very realistic waveforms, as Figure 3.10 suggests. However, it was observed, by comparing to the real underlying ECG, that some nuances might be abusive, such as an enlarged Q-wave and attenuated R-peak, culminating in a correlation among real and estimated counterparts approximately between 0.7 and 0.8. These results are enough to deem this approach possible, but further development and validation is advisable.

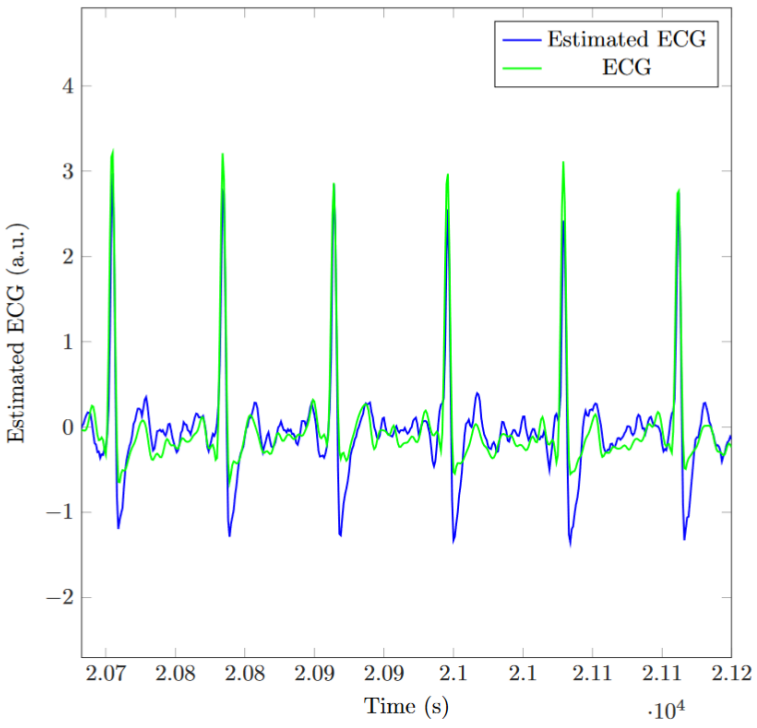

Figure 3.10: Comparison between the real ECG and its estimated waveform. Adapted from [89].

Another non-parametric algorithm is presented in [90]. This work had the goal of producing BCG artificial signals independent of the calibration configurations of the available real BCG recordings, which are used as ground-truth. Since signals from these modalities strongly depend on calibration and experimental conditions, e.g. position of the patient, a data-driven feature such as this in a generative model is of significant importance to allow modeling the most waveform types as possible. In order to achieve this goal, a combination of an inverted Gaussian renewal and a Gaussian process is designed. The first one is used to generate an artificial array of beat times, assuming their corresponding durations follow an inverse Gaussian probability distribution $\Delta t \sim \mathrm{IG}(\mu, \lambda)$,

$$\log(p(\{t_b\}|\mu,\lambda)) = \sum_{b=1}^{B-1} \log(\mathrm{IG}(t_{b+1}-t_b|\mu,\lambda)) + \log(1-\mathrm{IG}'(t_1-t_{\max}|\mu,\lambda)) + \log(1-IG'(t_{\max}-t_B|\mu,\lambda)),$$

where IG′ denote the CDF of the IG distribution and is used in the last two terms to deal with boundary conditions. Distribution parameters μ and λ can be learned by maximizing this log-likelihood for a known recording. The Gaussian process models the actual artificial signal using the standard log-marginal function

$$\log(p(\mathbf{y}|\{t_b\})) = -\frac{1}{2}\mathbf{y}^T(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}+\sigma_n^2\mathbf{I}| - \frac{N}{2}\log(\frac{2}{\pi}),$$

where $\mathbf{y}$ is the artificial signal vector, N the number of samples, $\mathbf{I}$ is the unit matrix of size N, σ_n is the level of noise in measurements and $\mathbf{K}$ is the covariance matrix, customized for this problem with time-varying covariance functions describing the time scales at which the signal is correlated, inter and intra beat correlations, with learnable parameters from a specific recording and optimized by maximizing the joint probability of the signal vector $\mathbf{y}$, and the series of beat duration values $\{\mathbf{t_b}\}$,

$$\log(p(\mathbf{y},\{t_b\})) = \log(p(\mathbf{y}|\{t_b\})) + \log(p(\{t_b\})).$$

Analysis of the final artificial signals showed potential for testing signal processing methods specific to BCG signals and data augmentation in the sense of missing segment estimation, represented in Figure 3.11. Algorithms capable of mimicking multiple types of waveforms, resulting from specific experimental configurations, and making few assumptions greatly potentiate novel algorithm development by supplying extensive datasets of a particular configuration. Moreover, this makes the algorithm heuristic enough to be applied to any other modality.

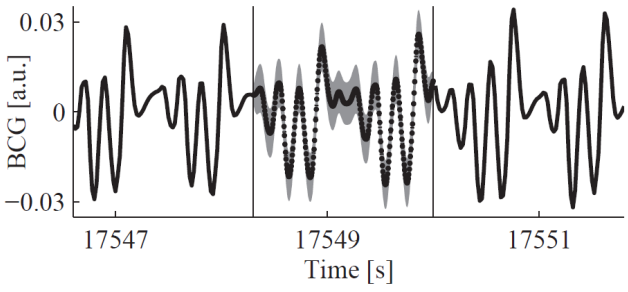

Figure 3.11: Showcase of the capability to fill missing signal segments and realistic waveform generation. Shaded areas represent point-wise standard deviation with respect to the dotted, artificial segment. Adapted from [90].

In an effort to monitor LVET from PPG signals, [91] designed a feature extractor which, under the assumption that the PPG waveform is composed by the summation of four waves representing either the systolic or diastolic portions, fitted a GMM with four Gaussian functions to a collection of pulses. From this model, LVET calculation was based on the first, second and third derivatives of the first two GMM components - considered to be part of the systolic phase. Comparison between the fitted GMM and the real waveform is shown in Figure 3.12.

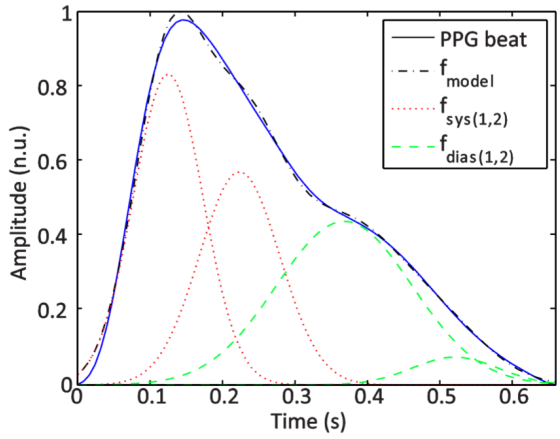

Figure 3.12: Fitted GMM to a PPG waveform and its components for systolic and diastolic segment representation. Adapted from [91].

Healthy subject processing originated state of the art results, but inherent flaws when using the algorithm in subjects with arterial stiffening were detected. The authors attribute these weaknesses to the difficulty the trust region algorithm has in discerning two waveforms when the dicrotic notch is attenuated and, therefore, calculating the onset and end of the systolic portion. Notwithstanding, this problem does not translate into a worsened morphological representation of the waveform, making the algorithm used for fitting, the non-linear least mean squares method (NLLSM), is still valid.

With the same rationale, [92] also adopted the GMM representation of PPG pulses. However, only two Gaussian functions were utilized per pulse, whose parameters were stored in a parameter vector $\mathbf{W}$ and the collection of pulses in recordings were found to follow a Gaussian distribution and to have significant correlation among each other, being modeled by a multivariate Gaussian random variable $\mathbf{W} \sim \mathcal{N}_{10}(\eta_{\mathbf{W}}, \Sigma_{\mathbf{W}})$. Principal component analysis (PCA) was then applied to the elements of the parameter vector and each one gave rise to a single channel, uncorrelated with the remaining ones. Each of these channels were, in turn, modeled by two autoregressive moving average (ARMA) models instead of following a direct probabilistic modeling approach after obtaining $\mathbf{W}$.

Authors emphasized the potential of the model to estimate missing segments of recordings, as shown in Figure 3.13, statistical characterization of clinical parameters, numeric representation of patient-specific physiology through extracted parameters and, finally, in diagnostic support by allowing safety intervals, regarding what should be expected from a patient, to be set according to model variability.

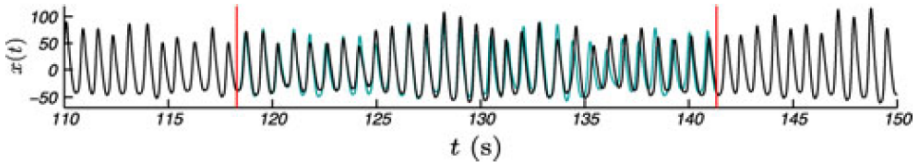

Figure 3.13: Filling missing segment (bounded by two red vertical lines) with artificial signal (in blue) and comparison with its real counterpart (in black). Adapted from [92].

3.1.2 Multimodal Models

One of the most extensive models available is detailed in [93]. Compiling previous studies of specific physiological cardiovascular and respiratory processes dating back to the first half of the past century, the model aimed at emulating sleep-disorders, namely obstructive sleep apnea (OSA), by means of dynamically describing phenomena such as cardiovascular autonomic control, chemoreflex and respiratory control through respiration and airway mechanics, circadian and sleep-wake regulations. It can simulate a number of conditions, besides only OSA, by generating physiological signals, for instance, tidal volume, HR, BP, RF and pleural pressure.

Although it is also one of the most self-contained models available, not needing any real recordings to set up its simulations, it tallies up to 80 states and 472 parameters. Produced signals managed to positively predict how the cardiorespiratory system behaves under various conditions along with presenting realistic waveforms. Despite authors claiming such an in-depth model is very useful in condensing the current knowledge on the physiology behind this system, the amount of states, parameters and underlying models describing subsystems can become overwhelming for users who were not involved in the development of the model as a whole.

In [6], a hybrid multimodal model is introduced with a dynamical, modality-independent component and a data-driven, modality-dependent one, contrasting greatly from the previously presented model. It consists of a framework which simulates the heart as a harmonic oscillator whose frequency (f_{HR}) is modulated by the coupling with a LF (with $f_{\mathrm{LF}} = 0.1$ Hz) and a HF oscillators (with $f_{\mathrm{HF}} = 0.25$ Hz), representative of the Mayer wave and RSA, respectively, as the presence of a lung is integrated. LF and HF peaks' widening are modeled by adding, band-limited, time dependent, random modulation components, $\gamma_{\mathrm{LF}}(t)$ and $\gamma_{\mathrm{HF}}(t)$. Such relations are expressed by a set of coupled nonlinear ordinary differential equations describing the time-dependent states of the heart, lung and Mayer wave oscillators in the phase domain $\phi_{\mathrm{card}}(t)$, $\phi_{\mathrm{resp}}(t)$ and $\phi_{\mathrm{mayer}}(t)$, respectively. The first oscillator is coupled to the last two by adding modulation components of the form $k_i \cdot f(s_i)$, where $i = \{\mathrm{LF}, \mathrm{HF}\}$, f was chosen to be the sinusoidal function, k_i empirically determined coefficients and s_i are the LF and HF oscillators' phase states. So far, phase signals are produced independently of signal modality. Modality-based template calculation is achieved using the previous phase information and based on a respiratory modulation and additive effect on every modality:

$$\hat{x} = T_{\mathrm{sensor}}\left(\phi_{\mathrm{card}}(t)\right) \cdot \Big(1 + T_{\mathrm{resp,mod}}(\phi_{\mathrm{resp}}(t))\Big) + T_{\mathrm{resp,add}}(\phi_{\mathrm{resp}}(t)),$$

where T_{sensor}, $T_{\mathrm{resp,mod}}$ and $T_{\mathrm{resp,add}}$ represent the template signals for the sensor of interest's signal, the respiratory modulation and additive components, respectively. While the first template is obtained by performing segmentation on a real recording and averaging over the obtained

segments, the respiratory templates are obtained by the minimization

$$\underset{T_{\text{resp,mod}}, T_{\text{resp,add}}}{\operatorname{argmin}} \quad \left(x_{\text{resid}} - \tilde{x} \cdot T_{\text{resp,mod}} - T_{\text{resp,add}} \right),$$

where x_{resid} is the signal obtained by the difference between the original signal and a signal $\tilde{x}$ composed only of a train of sensor templates T_{sensor}. This approach also allows modeling modalities where $T_{\text{sensor}} \equiv 0$, such as is the case for belt or transducer respiration measurement devices. Moreover, HRV signals can also be processed and used to create an RSA template T_{RSA} by segmentation according to $\phi_{\text{resp}}(t)$.

In pursuance of validation, multimodal signals (ECG, BP, PPG, airflow, respiratory effort from thorax and abdomen and cECG) from different databases were used to produce modality-based templates. These were then compared to their artificial counterparts in terms of time and frequency domain parameters and visual resemblance, yielding convincing results, as shown in Figure 3.14. The framework generates cardiorespiratory signals with considerable parameter flexibility for template creation, given that any patient recording of any cardiorespiratory modality can be used, while reproducing RSA and Mayer wave influence, shown to correctly capture differences among old, young and differently physically active patients. However, it shares some of the disadvantages of dynamical algorithms such as: fiducial points' relative position depending linearly on the RR-interval, which does not invariably translate to real life scenarios; different signal modalities depending differently on the considered respiratory effects and Mayer wave; no noise and/or artifacts are added nor pathological signals can be recreated. Moreover, averaging over segmented cycles may cause the templates to be smoothed, causing respiratory modulation components to stray too much from what is expected while trying to compensate.

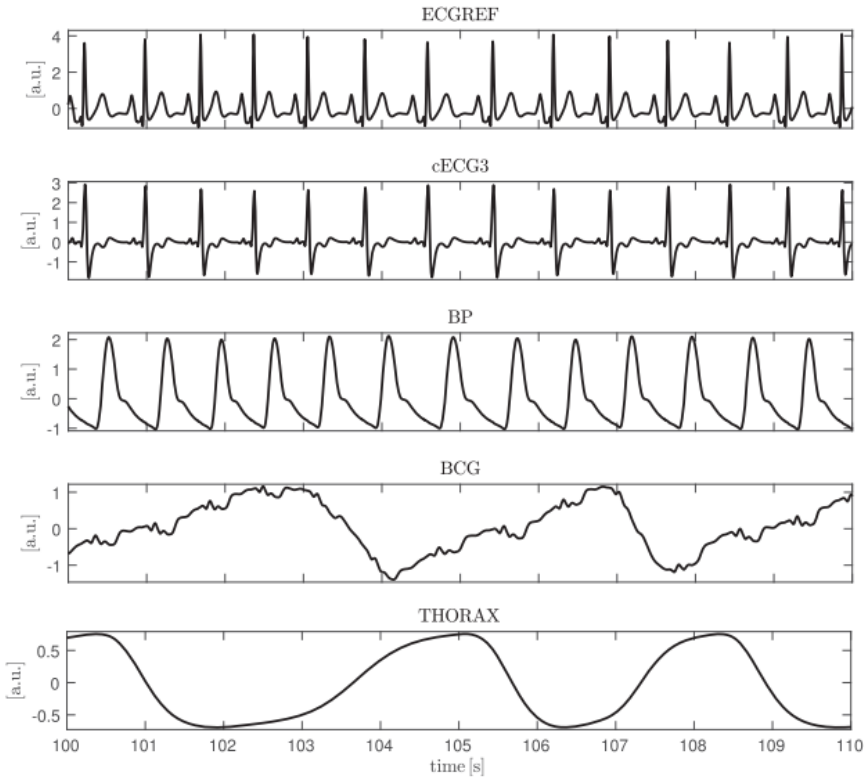

Figure 3.14: Synthesized multimodal signal. Adapted from [6].

Research output on artifact generation is very limited. Notwithstanding, [94] defines a statistical model to synthesize multimodal motion artifacts using data from cECG, PPG and electromechanical BCG sensors. Based on all artifact events in data, univariate analysis was performed by assigning the distribution type which fitted best to voltage amplitude, duration and power histograms to the respective measurement. Due to low correlation with the remaining measurements, raw voltage is modeled independently through a multivariate approach while duration and power were modeled as a multivariate distribution. To that end, a Gaussian copula was fitted to data. Waveform generation was based on the observation that artifacts manifested as either saturation or low-pass signals with noise. While the former was modeled as a state with minimum, maximum and transition time distributions, the latter was modeled as a state governed by an ARMA model. State transition was assured with thresholding – the starting state was defined by the power output of a random copula sample, which, if higher than the threshold, would mean the starting state is saturated, otherwise, it would be the ARMA state; when an ARMA state threshold of a certain duration takes place, it is scaled until its power reaches the randomly sampled power value, but may transition to a saturated state if it reaches a voltage higher than a voltage threshold.

Generated artifacts showed a successful representation of motion artifacts, as seen in Figure 3.15, with considerable variety, displaying potential for noise removal or algorithm testing, including sensor-fusion. The authors find improvements could have been made by applying the same statistical logic to introduce different types of noise and obtain generated artifacts for different signal modalities.

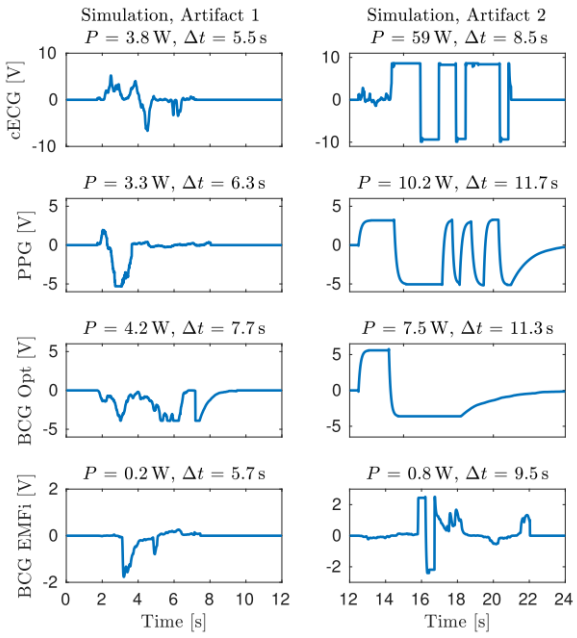

Figure 3.15: Simulated modalities' artifacts with different copula-yielded values. Both low-pass and saturated behavior was shown to comply with real measurements. Adapted from [94].

3.2 Deep Learning Methods for Sensor Fusion

Opposite to the previous Section 3.1 on multimodal modeling, a lot of research has been conducted concerning DL approaches for sensor-fusion, especially focusing on medical data [66, 95] but also

on different cases, such as human activity recognition (HAR). This section covers successful applications which, to some extent, provided the bearings along which the suggested CNN was designed.

3.2.1 Sensor Fusion in Cardiac Signals

The introduced DL research on cardiac signals provides solid practical architectural guidelines on how sensor-fusion can be maximized given cardiovascular waveforms and comparison standards to gauge the proposed network's performance in detection and classification tasks when similar metrics are used.

A CNN intended for simple beat detection in multimodal signals sampled at 250 Hz was developed in [96]. The authors opted for a single-convolutional layer, branched architecture. In other words, each k modality's signal comprised an input branch to the network, which underwent convolution with P filters, yielding the same number of feature maps. Signals corresponding to the feature maps of the network are denoted by $y_{k,p}$ where $k \in 1,\ldots,K$ and $p \in 1,\ldots,P$. Sensor-fusion occurs right afterwards, where corresponding feature maps from each modality are summed, resulting in a final, single set of P feature maps, with $y_{fusion,p} = \sum_{k=1}^{K} y_{k,p}$. After concatenation, these are connected to the inputs of a FCN. After application on the PhysioNet 2014 Challenge database, it was found that a top score of 94% was achieved with two, 20-sample-length filters and no hidden layers in the FCN, as the latter increased computational time with no significant improvements. The used score was the mean value between gross and average SE and PPV.

Training of the algorithm was based on assigning 251-sample signal windows (snippets) labels 1 or 0 for beat and not beat, respectively. Attending to ANSI/AAMI EC38 and EC57 standards [97], every snippet whose central sample was located inside 150 ms distance of a beat annotation (area referred to as location pulse) was labeled 1, while the remaining where labeled 0. Data augmentation was performed in order to produce snippets where only one modality has signal and the remaining are null, as well as data balancing since the number of zero-labeled snippets is much greater with this approach. In this sense, the outputs of the network are location pulse estimations.

As shown in Figure 3.16, robust detection was achieved when either one of the modalities was corrupted with some amount of noise, to which extent data augmentation was beneficial, as some beats are detected even in the absence of one modalities' signal.

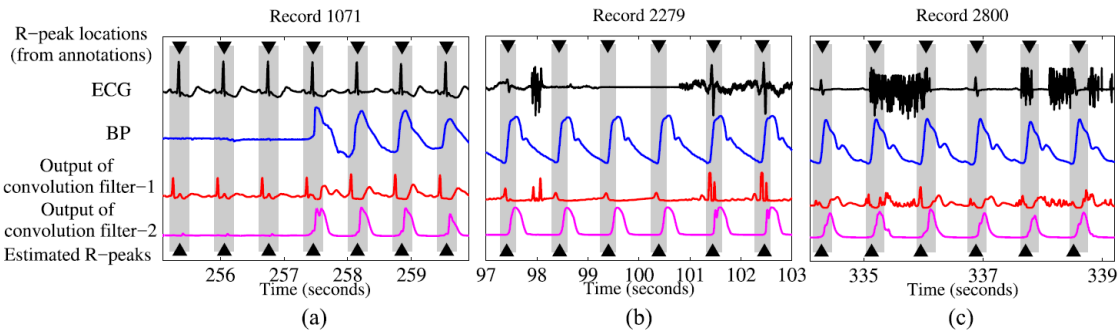

Figure 3.16: Beat detection in recordings from the PhysioNet 2014 Challenge database a) in the absence of BP signal, b) ECG signal and c) in noisy, non-clinical conditions. Adapted from [96].

The previous algorithm tended to primarily be controlled by the ECG modality, using BP information as a complement to the estimation, which was hypothesized by realizing ECG filters' time response resembled the Haar wavelet, frequently used to identify QRS-complexes, while BP filters did not convey any useful conclusions. In [98], however, a CNN was idealized to perform late-fusion at the level of the FCN, allowing individual modality branches to hone their respective features, at which point feature redundancy and, therefore, robustness should be higher. A similar, branched architecture was used. After fine-tuning with the PhysioNet 2014 Challenge database, the final structure, as seen in Figure 3.17, included five layers: a convolutional layer, followed by a max-pooling layer with a 2-sample-width filter and three other convolutional layers. The number of filters of convolutional layers increased: 10, 20, 25 and 30; while the corresponding width decreased: 20, 15, 10, 10. Finally, feature maps from the last layer are concatenated and inputted to a FCN with 3 layers of 100, 40 (both with dropout probability of 20%) and 2 nodes, given the last layer has a softmax activation function outputting the probability of being one of two classes, 1 or 0. ReLU activation functions were used for neuronal layers. A top score of 95% with data from the Fantasia database was achieved with an RMSE concerning beat localization compared to reference annotations of 43 ms. The used score was an average between SE and PPV.

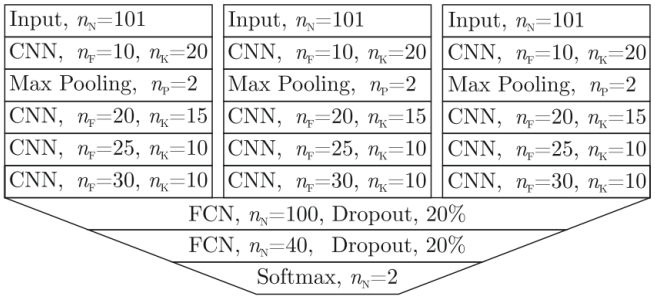

Figure 3.17: Architecture of the precursor CNN upon which this project's network was designed. Adapted from [98].

For training, 101-sample snippets were used at 250 Hz sampling frequency. Snippets whose central sample matched the beat annotation were labeled 1 and same snippets with a random shift of at least 20 samples were labeled 0, ensuring class balance. Such approach allows a more efficient maximization of the available datasets than the previous model. The algorithm was optimized using Adam and CE throughout 8 epochs of training, where one recording was used for testing and the remaining for training - leave-one-subject-out. The output signal of the network was, therefore, a probability signal corresponding to the probability of the current snippet being exactly centered at the location of the beat. Post-processing was conducted with an adaptive peak detector, yielding final estimations of the beat locations.

It was concluded that CNN architectures perform better than ANNs for the task of beat localization, achieving overall higher scores and a decrease in RMSE of 40% in unimodal data. Further results were observed, as all sensor-fusion approaches yielded better scores, approximately from 90% to 95%, at the expense of a slight deterioration in RMSE.

Concerning multi-class beat classification, the branch structure was also adopted in [99]. The network was employed in classifying the heart regions affected by myocardial infarction, along the rationale of what was introduced in Section 2.1.4.1, from 12-lead ECG recordings of specific patients by pre-training and localizing the occurrences in data (detection). Features from different

branches are developed independently through 3 convolutional layers, with 1, 2 and 4 filters and respective filter sizes of 61, 31 and 9, as the first two are followed by a 3-sample max-pooling layer. Each convolutional layer used ReLU. Late sensor-fusion is preferred as final feature maps were concatenated and used as input for a 1-hidden layer FCN, with 2 nodes in case of detection and 6 in case of classification. Softmax was used to produce predicted class probabilities. Opposite to what was done in previously introduced works, testing was conducted using segmented snippets instead of running the network through entire recordings, by testing snippets centered in each sample of the signal. Moreover, the algorithm was firstly pre-trained and validated with global datasets and, subsequently, subject-specific datasets were used for fine-tuning and final validation, being accuracy the chosen metric. Results above 99% were achieved in global testing for detection and for all classified classes, while patient-specific testing yielded slight decreases as detection achieved an approximate score of 99% and classification 95%.

Each branch received a 600-sample segmented snippet, sampled at 1kHz from the PTB diagnostic ECG database. Training was further conducted with SGD, throughout 30 epochs with a batch size of 16, using a learning rate of 0.05 in the first 20 and decreasing to 0.02 in the last 10. The training sets were not balanced as the relative presence of a specific class approximated its real-life relative frequency.

The network proposed by the authors achieved better scores among a number of published high-scoring networks for the task of detection and classification with the advantage of not requiring feature hand-crafting and being capable of distinguishing between 6 classes. However, performing sensor-fusion on 12 leads is very likely to ease the complexity of the proposed tasks, especially for myocardial infarctions, which are typically diagnosed with a 12-lead ECG clinical test, reinforcing how the use of a branched architecture helps maximizing sensor-fusion efforts. In fact, the authors argue such a design allows leveraging two aspects of multiple ECG leads, which could also be transferable to a multimodal setting: integrity, which refers to how different signals report the status of the same heart and, thus, intra-modality local events are correlated across modalities to some extent, making a joint analysis relevant; diversity, which supports multi-scale features of the heart are obtained through signals stemming from multiple angles and distances.

Perpetuating the same philosophy based on integrity and diversity, another network was published [100] for the same purpose but for real-time use with only four leads - the most relevant for the most types of myocardial infarction. Notwithstanding, this work presents a different approach to multi-lead sensor-fusion by dropping the branched architecture for a matrix built by vertical stacking of different leads' signals and proceeding with 2D operations, as for images. The authors assert that since regular 2D filters and single-factor pooling layers would not adequately respect integrity and diversity in the information, sub-2D filters and lead asymmetrical pooling (LAP) are used to ensure both aspects, respectively. While sub-2D filters only diverge from regular filters by having shared coefficients among rows, LAPs use different, manually defined pooling lengths for each lead, performing 0-padding where necessary. This methodology might make sense when working with the same modality on different leads where significant similar waveforms are presented, favoring training of shared coefficients to detect vertical lines at, for example, the QRS-complexes. Nonetheless, when multimodal signals are used, with significantly different and out-of-phase events (e.g. ECG, BP and BCG), a branched architecture makes more sense, still respecting integrity, by independently developing unimodal features, and diversity, by eventually joining feature maps, trained to emphasize features in a modality-tailored scale.

Training and testing were conducted by leaving-k-subjects-out, instead of 5-fold-cross validation, used in the previous model, with recordings from the PTB database resampled at 250 Hz, which is theoretically a least optimistic evaluation of the algorithm. The optimization algorithm was SGD, through 30 epochs with a learning rate of 0.17. A processing time of 17.10 ms per beat was achieve with Intel Core i7-3770 @ 3.4 GHz CPU and 12 GB RAM using MATLAB R2011a in 64-bit. This was deemed fit for real-time application, as posterior migration to an ARM Cortex-A9 microprocessor @ 925 MHz and 1GB RAM controlled by C scripts running on Linux yielded 26.75 ms of processing time per beat.

3.2.2 Fundamental Approaches to Sensor Fusion

On the other hand, given the wider range of waveforms being worked with, research in this section places a stronger emphasis on when should fusion occur in the network and what architectural mechanisms should be used, despite the shape and type of input data. In this case, specific results are not mentioned since no direct comparison to the proposed CNN can be made.

Another CNN with vertically stacked channels as input was also designed for HAR based on motion sensor-fusion to classify movements in [101]. A point was made that, for weakly correlated low-level features, including input signals, a late-fusion approach is preferable in order for unimodal features to independently contribute to high-level abstractions with a higher probability of being correlated since the same underlying process is at the origin. For the case, this rationale was developed to overcome data acquisition problems such as noisy configurations, flawed calibration, use of low-cost sensors or even unstable sampling rates and how these vary among modalities. The architecture included 3 convolutional layers, with 48, 96 and 96 feature maps obtained with 15-sample filter sizes. While the first two used 15 samples wide 1D filters followed by ReLU and 2-sample max-pooling, the last layer used 3 samples high and 15 samples wide, regular 2D filters followed by a ReLU activation. Subsequently, instead of resorting to a FCN to summarize feature maps' activations, a global pooling layer (GAP) was inserted, condensing each feature map into one single activation by taking its average, performing a transformation from an inter- and intra-pattern localization role to a pattern identification one [102], followed by 10 output units (for the desired classes) processed with softmax. In this model, using a 2D filter seems more reasonable than in the previously mentioned one as high-level features might be similar enough, in any given class, to be jointly recognized by one single filter through multiple modalities on one feature map.

Training also encompassed leave-3-subjects-out validation, using the adelta optimizer with an initial learning rate of 1 throughout a minimum of 2000 epochs, with a cutoff at 100 if accuracy did not improve, and dropout probability of 20%. This optimizer is able to adapt the learning rate over iterations and is reported to be more robust to changes in architecture, hyperparameters and noisy gradient information [103]. Since it was not used in the project it will not be further mentioned. Two public HAR datasets were used: UCL for hyperparameter optimization and, in conjunction with PAMAP2, for applicability and comparative evaluation with state of the art algorithms using handcrafted features, CNNs focused on early-fusion and convolutional long short-term memory methods. All used comparison metrics showed the model outperformed the previously mentioned versions in both datasets.

A commitment between early and late-fusion representations was presented in [104] in the form of a lightweight addition to independent modality networks, morphing them into a single

branched architecture, named CentralNet. Sensor-fusion in the ML paradigm is not restricted to signal modalities, but also video [105], image [106], audio [107] and text [108], to the point a taxonomy for different types of algorithms was established in [109]. It highlights two types of approaches the current model attempts to conciliate: joint, where different modalities are projected onto the same multimodal space, for example, via element-wise multiplication, and coordinated representations, where the projection into the multimodal space is constrained, forcing some modalities to play a more complementary role as the learning process dictates. In this case, apart from the unimodal branches, a multimodal branch is defined as the joint multimodal representation space, whose feature maps at each layer are obtained through a linear combination, with learnable weights, of themselves with the unimodal's feature maps from corresponding layers, describing the representation constraints. As Figure 3.18 depicts, and using the same notation, feature maps of multimodal layers are recursively calculated as

$$h_{C_{i+1}} = \alpha_{C_i} h_{C_i} + \sum_{k=1}^{n} \alpha_{M_i^k} h_{M_i^k},$$

where n is the number of modalities and α terms are the weighted sum coefficients and, as the multimodal branch does not receive inputs, its first layer is simply the linear combination of the unimodal branches' first feature maps. This adaptation can easily be extended to an arbitrary number of input modalities, layers and to either CNNs or ANNs, as shown by Tables 3.3 and 3.4.

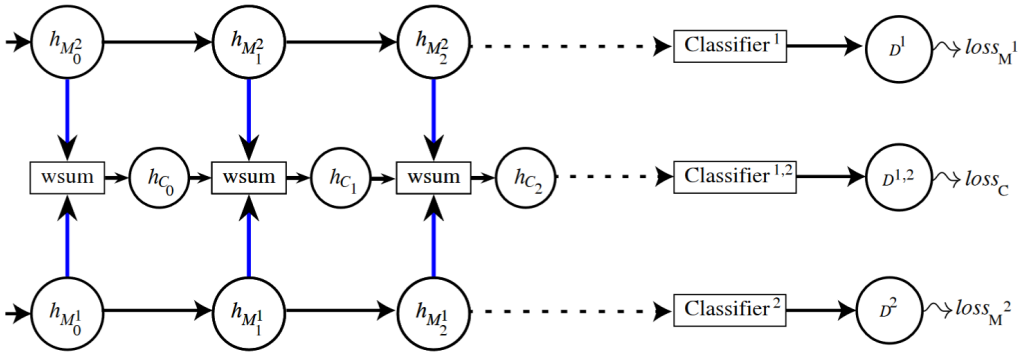

Figure 3.18: Simplified schematic of a CentralNet. Feature maps are represented by h, while the multimodal branch is represented with C and the unimodal branches with M. Respective modality's numbers are denoted as superscripts and subscripts indicate the layer number. Although not represented, convolutional and neuronal operations are performed before each feature map at the unimodal level and before each weighted sum (wsum) at the multimodal level. FCN layers also undergo weighted sum fusion, as the multimodal classifier is superscripted with numericals referring to both modalities. Outputs are captioned as D, yielding the respective losses. Adapted from [104].

The only adaptation required for training is the addition of all the individual branches' losses as total loss since the authors suggest that better overall performance is achieved when individual performance is maintained. After application of individual loss functions to the outputs, the final loss is computed as

$$loss = loss_C + \sum_{k=1}^{n} loss_{M^k}.$$

Training was also conducted using the Adam optimizer and the framework was applied in a variety of datasets for comparison with the previously mentioned models and others featured

Table 3.3: Structure for the MM-MNIST database, both inputs to unimodal branches are images. Adapted from [104].

Unimodal 1		Multimodal		Unimodal 2	
Type	Size	Type	Size	Type	Size
Conv	14x14x32	Conv	14x14x32	Conv	14x14x32
Conv	7x7x64	Conv	7x7x64	Conv	7x7x64
Dense	1024	Dense	1024	Dense	1024
Pred	10	Pred	10	Pred	10

Table 3.4: Structure for the MM-IMDb database. Inputs to unimodal branch 1 are text, while to unimodal branch 2 are the posters. Adapted from [104].

Unimodal 1		Multimodal		Unimodal 2	
Type	Size	Type	Size	Type	Size
Dense	2048	Dense	2048	Dense	2048
Dense	512	Dense	512	Dense	512
Pred	23	Pred	23	Pred	23

in the cataloging in [109] - MM-MNIST - visual dataset derived from MNIST where the original image is decomposed into two or more less energetic images resulting from PCA -, avMNIST - audiovisual dataset derived from MNIST and the Free Spoken Digits Database where images with a quarter of the energy of the originals are taken from the first and spectrograms from the corresponding voice recordings from the second -, Montalbano - an RGB and depth video and joint tracking map database of volunteers performing Italian sign gestures - and MM-IMDb - movie database with compiled information including plot, poster and fifty other properties intended to be used in order to predict the movie's genre. Adapted networks have consistently produced better results across all databases with negligible increase in processing time. Moreover, it was shown with testing on MM-MNIST that adapted networks show higher robustness in classifying lower energy pictures and pictures with mixed features.

It was concluded that the framework produces architectures which successfully enforce complementary aspects of modalities by learning how they are combined better. After training, an estimation of the contribution of every individual branch of the network at each layer can be obtained, resulting in the representations on Figure 3.19, which may provide interesting insight on the types of data being worked on, such as at what level of abstraction each modality better represents a class.

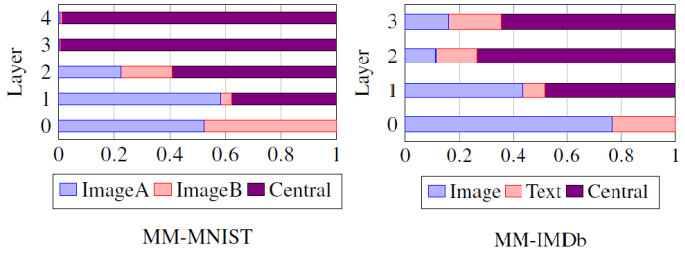

Figure 3.19: Relative importance of each branch's feature map at every given layer, calculated by expressing the linear combinations coefficients relative to one another. Adapted from [104].

Chapter 4

Materials & Methods

4.1 Datasets

In this section, a brief description of the datasets used in the thesis is presented. Multiple types of data were required to help designing the synthesizer so as to endow it with multimodal capabilities and were used to evaluate the resulting signals in Section 5.1. Most were also used to debug, optimize and test the CNN in order to prove the usefulness of cardiorespiratory data augmentation in DL.

4.1.1 Fantasia

This database contains 120-minute, ECG and RESP, recordings from twenty young (21 - 34 years old, with the suffix 'y') and twenty elderly (68 - 85 years old, with the suffix 'o') healthy subjects in awake, continuous supine rest. Only half of each group was used since non-calibrated, non-invasive BP measurements were conducted in half of the subjects (labelled with the prefix 'f2'). Samples were digitized at 250 Hz and annotations were generated by a detection algorithm, but visually confirmed later.

4.1.2 MIT-BIH Polysomnographic

Eighteen recordings were acquired from sixteen male subjects aged 32 to 56 during sleep in Boston's Beth Israel Hospital Sleep Laboratory. Given, the purpose of the database is to evaluate OSA, multiple modalities were recorded at 250 Hz. Of which, only recordings from the 5 subjects with ECG, invasive BP, BCG and RESP were used, along with the beat annotations [110].

4.1.3 MIT-BIH Arrhythmia

Acquired in the same hospital as the previous database, forty-eight half-hour recordings of two-lead ambulatory ECG channels from forty-seven subjects are available. Signals were digitized at 360 Hz and annotations were done manually. From a mixed population of patients, twenty-five

of the recordings include less common types of clinically significant arrhythmias. From available recordings, only arrhythmic events related to ectopic beats were used [111].

4.1.4 UnoVis

In order to foster studies on unobtrusive modalities, this database was created by the MedIT institute at RWTH with one-hundred and forty-five records, mostly unobtrusively acquired, for a total of 16.2 hours. A subsection of the database was used, containing thrity-one records from three cECG leads and one reference ECG, measured while subjects were driving in a city, highway or proving ground settings. The signals are sampled at 200 Hz and both automatic and manual beat annotations are provided [112].

4.2 Multimodal Synthesizer

The current section provides an in-depth description of the synthesizer algorithm and its components, schematized in Figure 4.1 and with the pseudo-code Listing 4.1. The goal of the model is to perceive physiological dynamics represented in a specific multimodal recording by decomposing it into numerically representable portions, as shown in Figure 4.2, from which these dynamics can be probabilistically modeled and reproduced into artificial recordings.

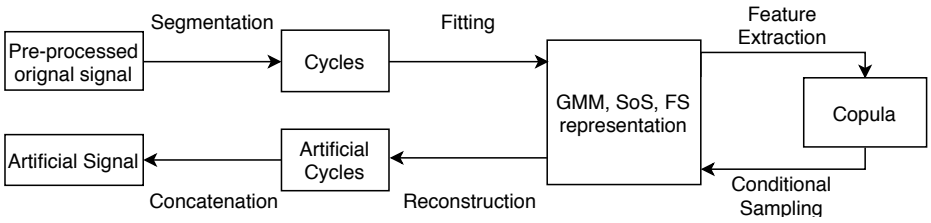

Figure 4.1: Synthesizer outline - flowchart associating the following subsections' components.

Listing 4.1: Algorithm Overview

```
for s in S       % state s in state space S
  B = cycle_segmentation(s)  % segmentation of signal of cycle-state s into B cycles
  for b in B      % beat b in segmented beats B
    for m in M     % modality m in modalities M
      x = delineation(m)  % division of each modality m's segment into portions x
      p = fitting(x)   % parameters p extraction from portions
    end
  end
  c = copula(p)     % copula c definition
  signal = signal_generation(c) % artificial signal generation
end
```

4.2.1 Markov Model Initialization

A Markov model was designed based on the manual classification annotations provided by the original datasets. Simply having a state evolution vector, i.e. a time series of classifications of the cycles present in the dataset, sufficed to extract the N_S-state vector vocabulary $\boldsymbol{S} = s_1, \ldots, s_{N_S}$ (which types of beats are featured in the data) and the state-transition probability matrix $\boldsymbol{A} =$

$a_{1,1},\ldots,a_{1,N_S},\ldots,a_{N_S,1},\ldots,a_{N_S,N_S}$ (by checking how often a given type follows a first type). The initial probability distribution was manually set to $\pi = \pi_1,\ldots,\pi_{N_S}$, where each probability is $\frac{1}{N_S}$ since in most datasets a user only has one recording, hence, one starting cycle, making automatic estimation statistically irrelevant.

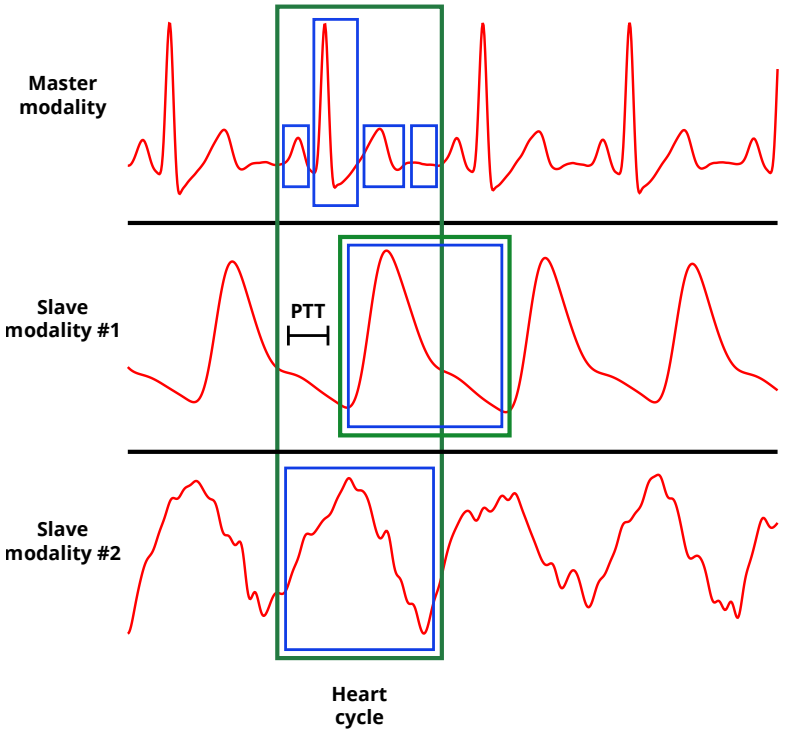

Figure 4.2: Segmentation overview. A post-processed original recording excerpt from subject slp59m of the MIT-BIH Polysomnographic database, with multiple heart cycles. The segmented cycle highlighted in green is further divided into unimodal segments, separated by horizontal lines, which are in turn divided into portions, highlighted in blue.

4.2.2 Pre-Processing

In this project, pre-processing was mostly independently performed on each modality, as shown in Table 4.1, because of the different physiological frequency ranges. The only transversal pre-processing step was sample standardization which was only used to ease the assessment of generated signals when debugging the algorithm and providing a normalization method for the input data of the CNN. As output, the generative algorithm keeps the original scales of the input signals. For this, mean removal sufficed.

Building up to the signal generation step, following Sections of the algorithm were run iteratively for each one of the N_S states. For simplification, mathematical notation in said Sections will omit Markov state terms despite their variables and results being associated with a specific state s. An overview of these iterations is shown in the Listing 4.1.

Table 4.1: Filtering scheme for pre-processing of original signals. Butterworth and zero-phase moving average filters with window sizes varying as a function of the sampling frequency (FS) to account for different morphology were used. In both ECG and cECG, the window size was set for 30% of the longest expected cycle and 10% for BCG due to finer variations, as seen in Section 2.1.3.2. Less variable waveforms' window sizes were set to a tenth of a second.

Modality	Low-pass		High-pass		Average (samples)
	cutoff (Hz)	order	cutoff (Hz)	order	
ECG	50	4th	0.3	4th	$0.3 \times 0.12\text{s} \times FS$
cECG	30	4th	0.3	4th	$0.3 \times 0.12\text{s} \times FS$
BP	30	4th	0.3	4th	$0.1\text{s} \times FS$
PPG	50	4th	0.3	4th	$0.1\text{s} \times FS$
BCG	20	4th	1	4th	$0.1 \times 0.12\text{s} \times FS$
RESP	0.3	4th	0.1	4th	$0.1\text{s} \times FS$

4.2.3 Cycle Segmentation

To integrate the generative algorithm, original multimodal signals were segmented into groups of synchronous, multimodal heart cycles [1] henceforth referred to as cycle-states in compliance with the Markov notation introduced above. A cycle-state represents an underlying common physiological and/or electronic manifestation. As such, every cycle-state from a particular generated signal should share the cardiovascular dynamics of a particular subject as presented by a particular measurement system configuration [113]. Therefore, if a specific subject presents both healthy and pathological heart cycles, not only were these two types of beats grouped as two separate cycle-states but also different sub-states (e.g. ectopic beats in the form of PVC and PAC cycle-states) in order for intra- and inter-state dependence to be correctly modeled.

4.2.3.1 Waveforms

In this project, multimodal signal processing is divided into two processes: 1) temporal delineation of cycle candidates and 2) posterior validation. Furthermore, one modality was selected as master modality, while the remaining are considered slaves. In this case, ECG was always chosen as master since it is the most common biosignal type in public cardiorespiratory datasets and also the best documented. For the process 1), depicted in Figure 4.3, only the master was used but in 2), depicted in Figure 4.4, every modality was analyzed and a cycle candidate was only saved if every modality's segment was validated.

For regular and PAC heart cycles, both cycle-states were considered to start at the onset of the P-wave, when early atrial depolarization occurs, while PVC were considered to start at the onset of the altered QRS-complex shape, when early ventricular contraction occurs. Hence, viewing windows were established around the given dataset QRS annotations, with a predefined number of samples before and after each, to bootstrap simple peak detection. In normal beats, the P-wave was identified as the most prominent peak taking place before the R-peak; in PVC beats, the onset of the altered QRS-complex was considered to be the onset of the first most prominent peak. Prominence was used as an absolute measurement, to improve robustness when dealing with

[1] The segmentation problem can be tackled by the user of the algorithm in their own way besides the one suggested in this work.

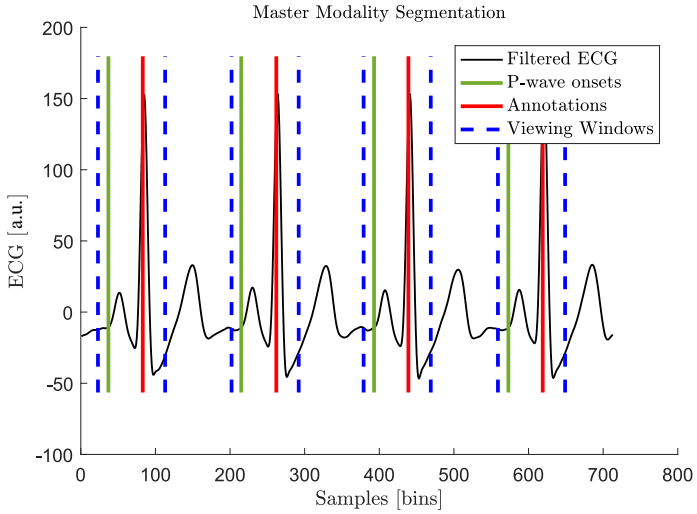

Figure 4.3: Segmentation of master modality based on viewing windows (enclosed by blue dashed lines), around the QRS annotations (in red). Cycle candidates are the signal (in black) portion enclosed by each pair of consecutive P-wave onsets (in green).

different ECG lead signals. A candidate cycle was assumed to be enclosed by two consecutive detected P-wave onsets.

A predefined set of thresholds and scalars were empirically set to accept or reject a waveform based on its duration, RMS and cross-correlation with the average waveform of cycle candidates $\bar{y}$. Valid waveforms $y_{i,\text{candidate}}$ were considered to have a duration between 0.5 and 1.2 seconds (HR between 50 to 120 BPM since most subjects were at rest), an RMS value inside an interval $[0.1 \times \tilde{\text{RMS}}, 10 \times \tilde{\text{RMS}}]$ set from fractions of the median RMS value of every annotation-delimited snippet and a correlation peak $\text{argmax}(\bar{y} \circledast y_{i,\text{candidate}})$ between itself and $\bar{y}$ (scaled to the the same length) located inside a lag of $[-15, 15]$. The duration test was meant to remove segments containing more or less than exactly one heart cycle and RMS and cross-correlation tests to filter out anomalous waveforms.

4.2.3.2 Additionally Extracted Parameters

Given the PTT observed in BP and PPG modalities, time delineation in the molds of the remaining modalities would result in the clipping of the diastolic runoff at the end of the cycle window, while the remaining waveform would only appear in the chronologically following cycle. Bearing this in mind, an additional delay parameter allusive to PTT, was added to the model in the form of an integer, in samples, extracted per segmented cycle, as shown in Figure 4.2. While segments of modalities like ECG, cECG and BCG have null delay since they can be approximated to be synchronous pulses [15], BP will have a positive delay value between 1 and the length of the cycle. BP segment windows were, thus, adjusted by delaying the windowed signal by using the difference between the P-wave onset and the location of the minimum before the upward systolic slope as PTT proxy.

Segment lengths (RR-intervals) and respiratory phase ϕ_{RESP} were saved as vectors independent of modality with one element per segmented cycle. The process of obtaining $\phi_{\text{RESP}} : [1, \infty] \rightarrow [-\pi, \pi]$ from the RESP signal, as shown in Figure 4.5 is detailed in Appendix A.

A mean waveform of the extracted beats of every modality was also calculated for later use in

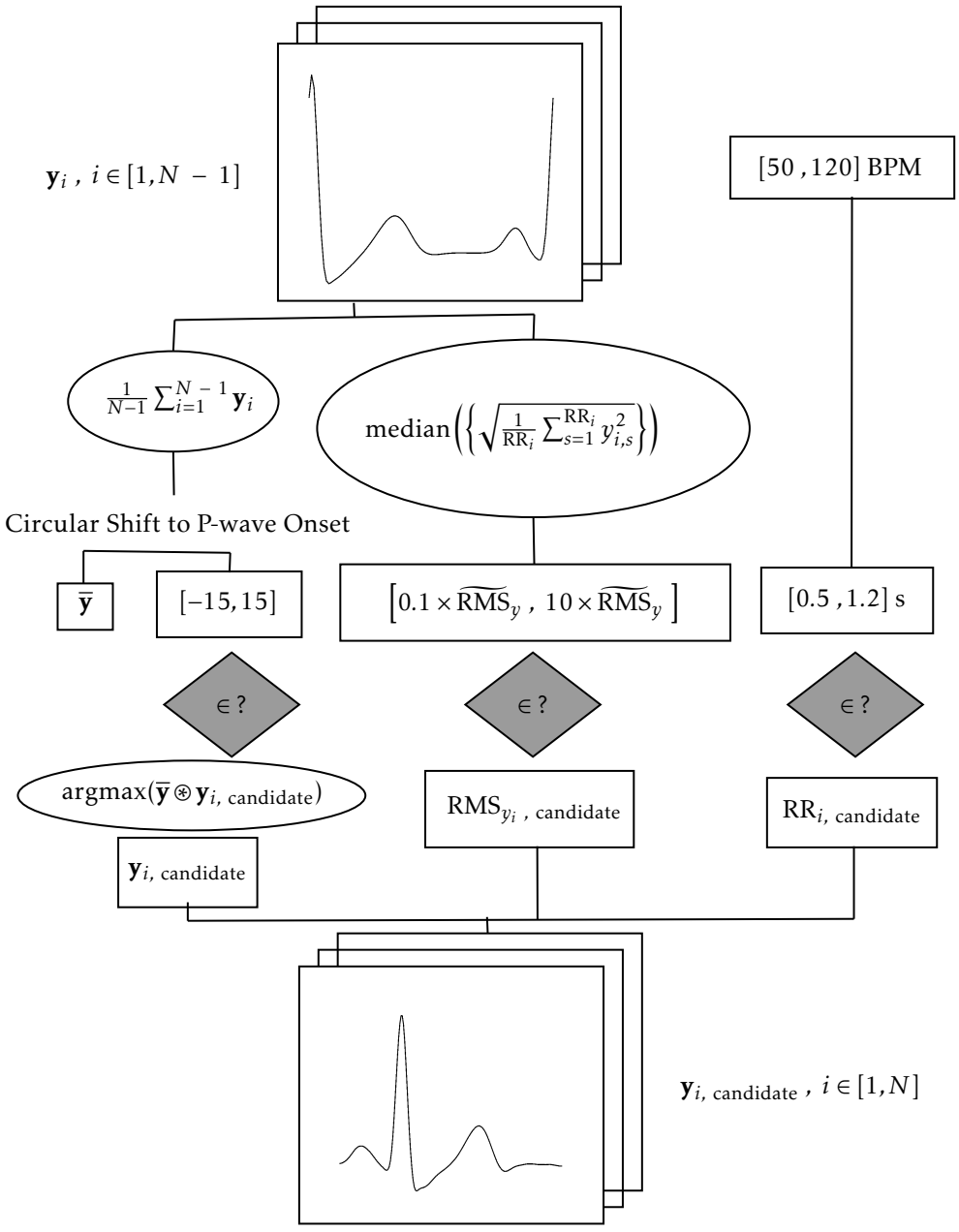

Figure 4.4: Heart cycle validation flowchart, used in every modality. Bottom and top arrays of snippets represent the N_a (number of annotations) previously segmented cycle candidates and annotation-based segmented beats, respectively. The latter are used to calculate the average waveform and a reference RMS value range. Each cycle candidate is cross-correlated with the average waveform, tested for alignment and checked for presence of its RMS and RR-interval values inside the validation range.

signal portion identification and correlation control.

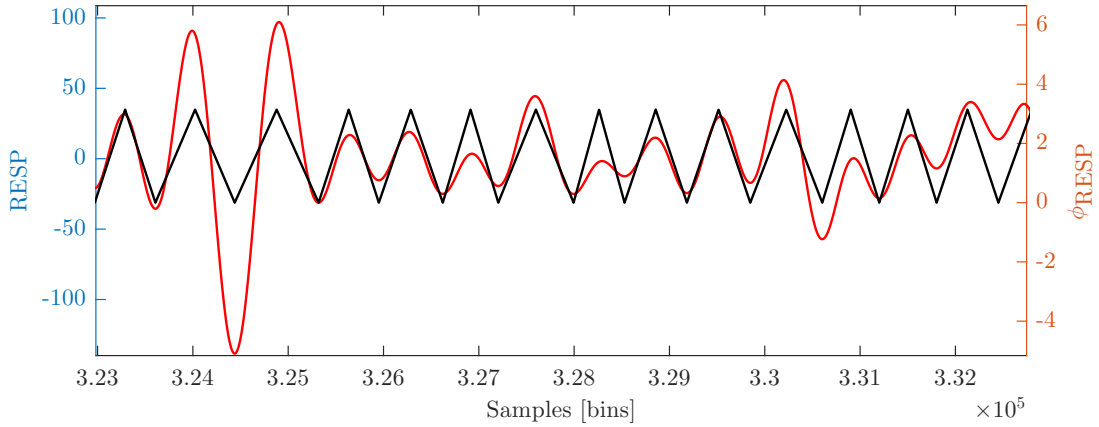

Figure 4.5: Overlapped plots of ϕ_{RESP} (in black) and its respective RESP signal (in red).

4.2.4 Feature Extraction

In pursuance of modeling inter- and intra-modality dependencies based on segmented heart cycles, the latter were represented by a common feature representation capable of being translated back into heart cycle signals. Fitting predefined functions to these segments allowed the extraction of comparable features - coefficients of the fitted functions - which could later be used as inputs of those same functions for signal generation.

Essentially, multiple ways to fully represent the cycles through a combination of GMM, Sum of Sines (SoS) and/or Fourier Series (FS), validated in [29, 76, 77, 82, 84, 87, 88, 91, 92], are allowed by dividing the segmented cycles into different, independently fitted portions (e.g. P, Q, R, S and T waves for the ECG modality), or fitting only one type of function to the entirety of the cycle. The fitting process is, however, not trivial. Several iterations of the algorithm were intended to maximize the number of portion samples that partake in the function fitting process while minimizing segmentation and fitting errors. This approach was applied to every multimodal segmented cycle of a particular state, one modality at a time - the output of this module was, therefore, a matrix of coefficients where each row corresponded to a multimodal coefficient train of concatenated unimodal coefficient vectors, as shown in Appendix C.

Taking into consideration respiration, portion delineation was based on characteristic wave morphology, rather than a common baseline which seldom follows a linear trend, just as explained in Sections 2.1.3 and 2.1.2.3. ECG beats were divided into portions, while BP segments were modeled by fitting the whole pulse with a GMM. On the other hand, low-SNR modalities are characterized by displaying highly volatile waveforms which do not allow a stable mathematical representation of portions with GMMs. Therefore, high volatility portions of these modalities (portions around the QRS-complex in cECG and the whole undivided segment in BCG) and noisy portions of other modalities' segments were modeled using either SoS or FS, which are capable of accommodating much higher variance than a GMM using fewer coefficients. Different portions can therefore be mathematically described as

$$\text{GMM}[k] = c_{\text{GMM}} + \sum_{i=1}^{N_F} a_{\text{GMM},i} \cdot e^{\frac{x[k]-\mu_i}{\sigma_i}^2}, \tag{4.1}$$

$$\mathrm{SoS}[k] = c_{\mathrm{SoS}} + \sum_{i=1}^{N_F} a_{\mathrm{SoS},i} \cdot \sin(\omega_i \cdot k + \varphi_i), \tag{4.2}$$

$$\mathrm{FS}[k] = c_{\mathrm{FS}} + \sum_{i=1}^{N_F} a_{\cos,i} \cdot \cos(\omega \cdot k) + a_{\sin,i} \cdot \sin(\omega \cdot k), \tag{4.3}$$

where every a terms are scalars corresponding to a Gaussian, sine or cosine component's amplitudes, μ and σ are the mean values and standard deviations with respect to a Gaussian, φ are the phase of each sine component, ω are the angular frequencies, c terms are the DC components and N_F is the degree of the fitting function. Note that the segment portion's signal, $x[k], k \in [k_{\mathrm{start}}, k_{\mathrm{end}}]$, is considered to be discrete and is fitted in the form of

$$x[k] = \begin{cases} \mathrm{GMM}[k], & k \in [k_{\mathrm{start}}, k_{\mathrm{end}}] \\ 0, & k \notin [k_{\mathrm{start}}, k_{\mathrm{end}}] \end{cases}.$$

While different GMM portions can be summed together to obtain a segment, FS and SoS functions were joined with neighboring portions through 10-sample cubic interpolations. This implementation is necessary since only GMM models respect Dirichlet and Neumann's boundary conditions

$$\lim_{k\to\infty} \mathrm{GMM}[k] = \lim_{k\to-\infty} \mathrm{GMM}[k] = 0$$
$$\lim_{k\to\infty} \frac{\partial \mathrm{GMM}[k]}{\partial k} = \lim_{k\to-\infty} \frac{\partial \mathrm{GMM}[k]}{\partial k} = 0.$$

Following subsections of the feature extraction algorithm were run for each modality $m = 1, \ldots, M$ of the M modalities featured in the current cycle-state and each beat $b = 1, \ldots, B_{\mathrm{seg}}$ of the B_{seg} validated segmented cycles, as schematized in Listing 4.1.

4.2.4.1 Portion Delineation Methods

The segment division algorithm introduces baseline-independent portion identification with minimal user-input. The first step of the algorithm was delineating the portions of a given segment. A range of possible configurations is provided to cope with different segment waveforms:

1. Positional Segmentation - intended for use in low-complexity waveforms with no energetically dominant portions, such as PVC in the ECG modality. This method relies on a user inputted index vector delineating different portions of the previously calculated average cycle corresponding to the modality currently being processed. Afterwards, zero-crossing indices are identified in the segment and the ones which are closest to the user inputted indices after conversion to the current segment's scale are selected as final portion boundaries. An example of positional segmentation follows where user input is provided in Listing 4.2 and delineation is represented in Figure 4.6;

Listing 4.2: Positional Segmentation Input

```
% Typical setup for PVC beats in ECG
template_locs = [];
segment_locs = [23];
fit_types = {'gauss2', 'gauss3'};
```

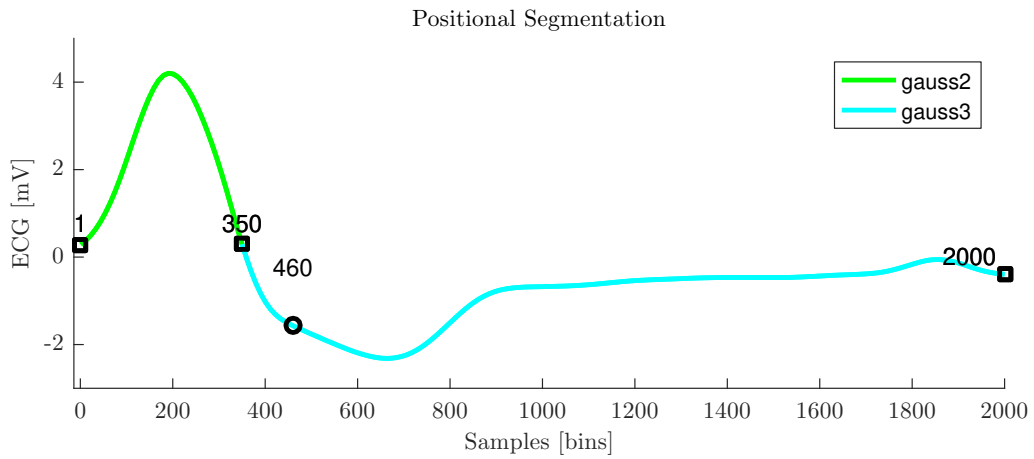

Figure 4.6: Representation of positional segmentation. A PVC segment is divided into portions according to template_locs and segment_locs in Listing 4.2 converted to a 2000-sample scale as reference, highlighted with black circles. Final portion boundaries are highlighted with black squares.

2. Anchor Segmentation - designed for complex waveforms where the number of portions varies but there are energetically distinctive portions. Normal cycles in the ECG and cECG are, thus, indicated for such, given that the QRS-complex is the most energetic portion of the wave, being distinguishable even in low-SNR setups. As such, segmentation becomes a function of the location of this characteristic component (CC). Using the average cycle, the user inputs a 2-element index vector which encloses the desired CC. Using cross-correlation after interpolating the average cycle signal $\bar{y}$ of length $N_{\bar{y}}$ to the scale of the current segment signal y, the highest correlation lag $x_{\text{CC, start}}$ value robustly identifies the onset of the CC signal y_{CC} in the segment:

$$x_{\text{CC, start}} = \text{argmax}_{k\in\mathbb{Z}}(\bar{y}\cdot y)[k], \text{ where,}$$

$$y_{\text{CC}} = y[x_{\text{CC, start}}, \ldots, x_{\text{CC, start}} + N_{\bar{y}}].$$

Subsequently, the zero-crossing indices of the newfound CC, using a mean of its boundary values as baseline, are considered delineation boundaries for inner-CC portions (e.g., Q and S-Waves of the ECG modality). Finally, portions before and after the CC are identified with positional segmentation delineation boundaries appended to a final portion boundaries vector. An example of anchor segmentation follows where user input is provided in Listing 4.3 and delineation is represented in Figure 4.7;

Listing 4.3: Anchor Segmentation Input

```
% Typical setup for normal beats in ECG
template_locs = [14, 46];
segment_locs = [21, 33, 80];
fit_types = {'gauss2', 'gauss2', 'gauss1', 'gauss2', 'gauss2', NaN};

% Typical setup for normal beats in cECG
template_locs = [21, 61];
segment_locs = [36];
fit_types = {'sin8', 'gauss2', 'gauss2', 'sin8'};
```

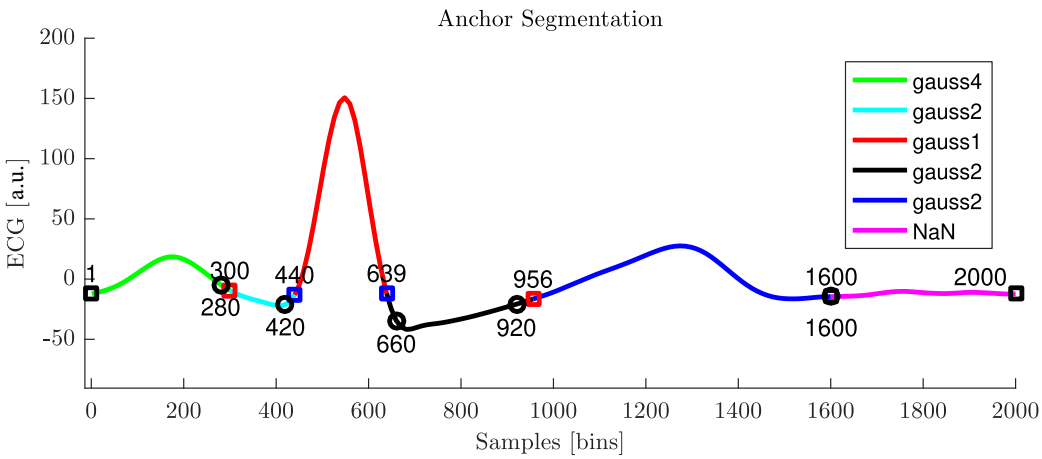

Figure 4.7: Representation of anchor segmentation. A normal ECG segment is divided into six portions (P, Q, R, S, T and transition) duly delineated based on reference, user-inputted template_locs and segment_locs in Listing 4.3, converted to a 2000-sample scale and highlighted with black circles. After processing, final portion boundaries were calculated, which are highlighted with squares: CC boundaries in red; inner-CC division in blue; remaining segment boundaries in black.

3. No Segmentation - the simplest form of delineation, preferably used in modalities where a single function type is able to accommodate the segment's variability. BP, PPG, and BCG are good examples of such. In this case, the segment is considered to have no intermediate bounded portions and is fitted as a whole. The final portion boundaries vector is composed of the first and last sample of the segment. An example of no segmentation follows where user input is provided in Listing 4.4 and delineation is represented in Figure 4.8;

Listing 4.4: No Segmentation Input

```
% Typical setup for normal beats in BP
template_locs = [];
segment_locs = NaN;
fit_types = {'gauss4'};
```

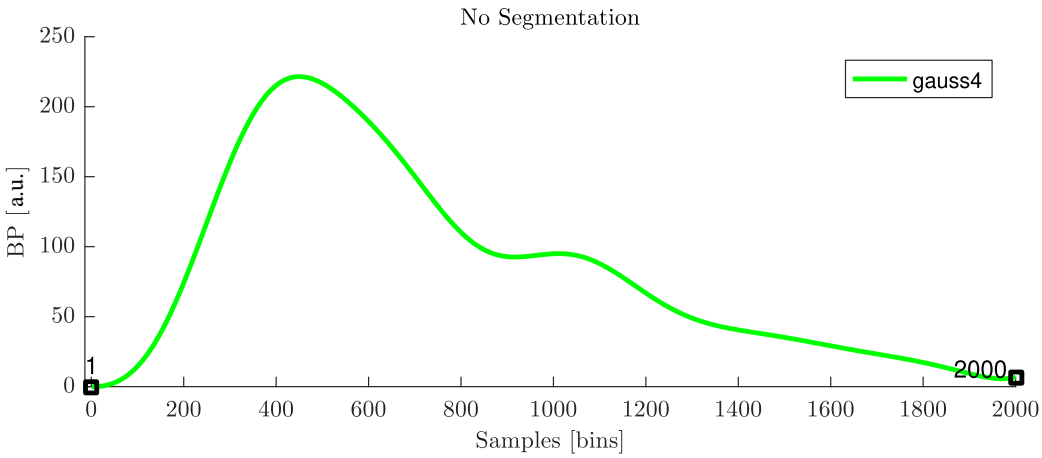

Figure 4.8: Representation of no segmentation. A BP segment left undivided as Listing 4.4 dictates.

Appendix C illustrates the data structure constructed through this process using the same examples previously presented. Additionally, a conceptual overview is provided with the flowchart in Figure 4.9.

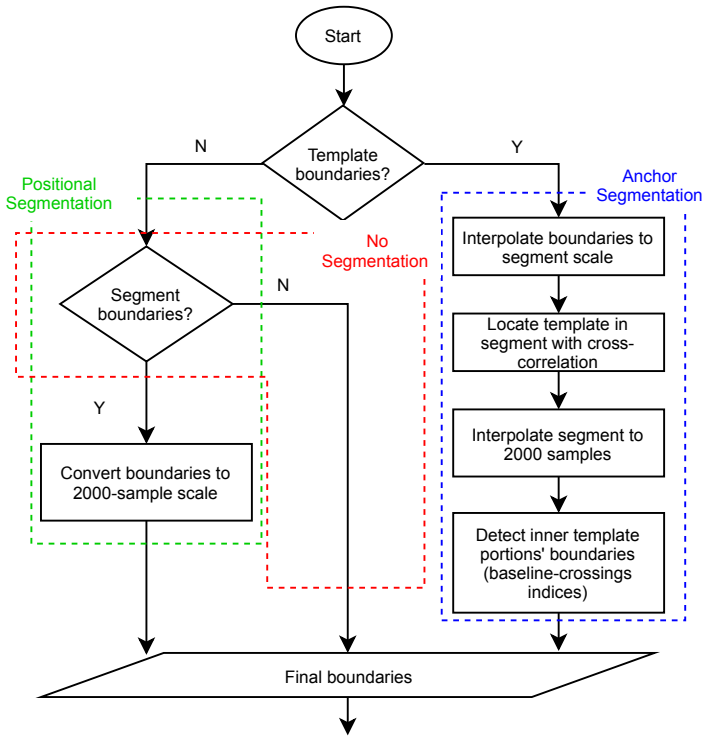

Figure 4.9: Flowchart of the beginning of feature extraction, with the different configurations of portion delineation available to the user, bounded by colored dashed lines - anchor, positional and no segmentation. Routines depicted in this figure are inserted in a loop and repeated for every single segment obtained from segmentation in Section 4.2.3. Inputs to this loop include sample vectors of reference portion boundaries with respect to the average waveform.

Aiming at increasing delineation accuracy, robustness and statistical significance of fit, segmented signals were linearly interpolated to a larger scale of 2000 samples (empirically determined). Previous to this step, segments resulting from cycle segmentation in Section 4.2.3, have a limited length range of 125 to 300 samples as imposed by the RR-interval validation filter, which after partitioning was reduced to mere dozens of samples per portion.

4.2.4.2 Portion Fitting

An iteration through the bounded portions followed, guided by the user-inputted vector fit_types denoting the function type and order intended to attempt fitting to a particular portion, as represented in Figure 4.10.

A fit_type vector element can also be flagged as "not a number"(NaN), in order to be ignored by the algorithm, in which case a 3-element, zero vector is appended to the final vector of fitted function coefficients. Noisy transitions were flagged as such so as to save computational time, memory and stochastic instability when defining a copula. Otherwise, three types of functions, GMM, SoS and FS were fitted to segments as previously explained [2], although with different

[2] Matlab® restricts fit functions to the 8th order, hindering the algorithm.

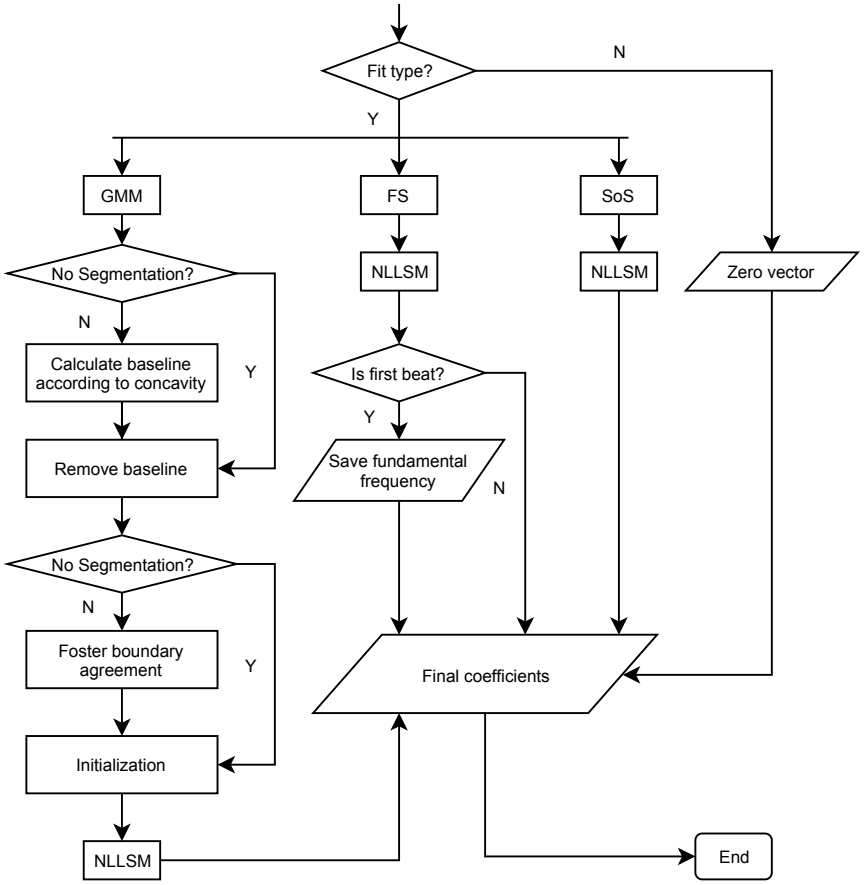

Figure 4.10: Final steps of the feature extraction process as a continuation of the previous flowchart in Figure 4.9, receiving as input its final boundaries vector and iterating over each pair. Further input to this routine includes fit_types. The output of each iteration is a row vector containing the final feature coefficients.

processing requirements:

1. GMM - if anchor or positional segmentation were chosen for delineation, two pre-processing steps were needed before the fitting process since a single portion can have both positive and negative samples. Firstly, the portion's concavity was calculated, according to the algorithm in Appendix B, to identify if its baseline should be its minimum or maximum value.

   The baseline was then removed from the signal, and a boundary agreement process was used to promote similar amplitudes to exist in both extremities of the portion, mitigating delineation variance among different cycles. To do so, the left bound was used as reference to admit new samples located beyond the preliminary right bound, whose signal amplitude is less than 5% of the segment's peak and does not trespass its bounds on more than 25% of the portion's length. Non-segmented modalities simply had their minimum value used as baseline.

   Finally, following the calculation of the best fit coefficients a_{GMM}, μ and σ, these were appended to the final coefficients vector along with the calculated portion's baseline value c_{GMM}, with the physiological reasoning being an additive and modulation effect introduced to the signal by the respiration;

2. SoS and FS - both fit types proceeded similarly, without any need for pre-processing with the exception of the first cycle ran by the algorithm. This serves to probe a fundamental frequency ω, which was saved for use in other cycles, relieving the computational weight and improving probabilistic stability of the copula by making one of its marginal distributions constant.

   Following the calculation of the best fit coefficients a, φ, ω and c coefficients forming a vector for a single fit type, were appended to final coefficients vector.

Suitable fitting coefficients are found with a nonlinear least-squares method simplified by using the trust region approach in [114], also used in [91]. Depending on the fitting function, $\hat{y} = \{\text{GMM}, \text{SoS}, \text{FS}\}$, the set of parameters $\boldsymbol{p} \in \mathbb{R}^{1\times n}$, which minimize the least-squares function $f[\boldsymbol{p}]$ between $\hat{y}[k]$ and the original set of points $x[k]$ is found as follows: a trust region is defined as the neighborhood A (spherical or ellipsoidal shape) around $\boldsymbol{p}$ in which function f is approximated to a linear function q. The trust region sub-problem is then

$$\min_{s} q[\boldsymbol{s}], \boldsymbol{s} \in A;$$

after all the vectors of A are computed, the trial step $\boldsymbol{s} \in \mathbb{R}^{1\times n}$ is implemented $f[\boldsymbol{p}+\boldsymbol{s}] < f[\boldsymbol{p}]$, updating the parameters. If the previous condition is not met, the neighborhood is changed. The approximated function q is described as the first two terms of the Taylor expansion series around $\boldsymbol{p}$ since by assuming $\delta\boldsymbol{p} = \boldsymbol{s}$ is small, higher order terms are negligible, thus making the function quadratic

$$f[\boldsymbol{p}+\boldsymbol{s}] = f[\boldsymbol{p}] + \sum_{i=1}^{n} \frac{\partial f[\boldsymbol{s}]}{\partial \boldsymbol{s}_i}\boldsymbol{s}_i + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial f[\boldsymbol{s}]}{\partial \boldsymbol{s}_i \partial \boldsymbol{s}_j}\boldsymbol{s}_i\boldsymbol{s}_j;$$

the last two terms can be rewritten in vector form as $\boldsymbol{s}^T\boldsymbol{g}$ and $\frac{1}{2}\boldsymbol{s}^T\boldsymbol{H}\boldsymbol{s}$, respectively, being $\boldsymbol{H}$ the Hessian matrix and $\boldsymbol{g}$ the gradient at $\boldsymbol{p}$. Therefore, the previous minimization problem can be rewritten as

$$\min_{s} \boldsymbol{s}^T\boldsymbol{g} + \frac{1}{2}\boldsymbol{s}^T\boldsymbol{H}\boldsymbol{s} \text{ such that } \|\boldsymbol{I}\cdot\boldsymbol{s}\| < \Delta$$

where $\boldsymbol{I}$ is a diagonal matrix and Δ a positive scalar which is reduced if $f[\boldsymbol{p}+\boldsymbol{s}] \geq f[\boldsymbol{p}]$, changing the neighborhood A.

To optimize the parameter estimation process for Gaussians, $\boldsymbol{p}_{\text{GMM}}$ was initialized. For an N_F-th degree model, the starting amplitudes $\boldsymbol{a}$ were set to the N_F-th fraction of the maximum value of the segment, starting mean values $\boldsymbol{\mu}$ were set to spread evenly across the segment's length and starting standard deviations $\boldsymbol{\sigma}$ were defined as the N_F-th fraction of the segment's length. Furthermore, for positive segments, the above-mentioned parameters were restricted to $[0, \max(x[k]) + 10\%]$, $[-\infty, +\infty]$, $[0, \sigma_i + 10\%]$ and, for negative segments, $[\min(x[k]) - 10\%, 0]$, $[-\infty, +\infty]$, $[0, \sigma_i + 10\%]$, respectively, with $i = 1, \cdots, N_F$.

4.2.4.3 Segment Rejection and Data Structure

A filter based on the RMSE calculated between the original and artificial segments resulting from the fitting process is used to minimize imperfect fitting and segmentation. Gaussian distributions $\mathcal{N}(\mu_m, \sigma_m)$ are assumed for the RMSE values of the fitting process of each modality m and those which yield a total RMSE in the range of $[\mu_m + \frac{3}{2}\sigma_m, \infty]$, are discarded - percentile-based outlier removal. The RMSE distribution is iteratively updated, with μ_m and σ_m being recalculated after

fitting one modality's segment. Until the 51st cycle is reached, a constant RMSE rejection threshold of 15 (empirically determined) is used. Once again, an artificial cycle is only kept if every modality's segment is validated.

Processing of each cycle-state yields B_{fit} horizontal coefficient vectors, denoted as $B = B_{\text{fit}}$ for simplification, which were stacked into a matrix $\boldsymbol{D} = (d_{bp}) \in \mathbb{R}^{B \times P}$, where P is the number of coefficients describing each of the B cycles. A row contains an ordered vector of all portions' parameters, while a column is a vector of the different values taken by the same parameter respective to the same portion in different cycles. This data structure captures same-parameter variation including RR-interval, respiratory phase and delays for each of the $M-1$ slave-modalities, featured as the first elements of each row.

To further guarantee the reliability of the obtained coefficients vectors, cluster-based, state-specific outlier removal was performed in the feature space $\boldsymbol{d}_b$ with B data points. Clustering over a feature space containing coefficients from every modality forced both intra- and inter-modality dependence to bias the calculation of the center $\bar{\boldsymbol{d}}_b$, i.e., the point in the feature space less distant from all the other data points, according to the squared Euclidean distance ($d^2(\boldsymbol{d}_{b,i}, \boldsymbol{d}_{b,j})$). This metric yields higher values for outliers when compared to regular Euclidean distance, making it easier to filter these out. The used algorithm follows:

$$\bar{\boldsymbol{d}}_b = \boldsymbol{d}_b, \text{where,}$$

$$i = \text{argmin}_{i \in 1,\dots,B} \left(\sum_{j=1}^{B} (\boldsymbol{d}_{b,i} - \boldsymbol{d}_{b,j})^2 \right).$$

Outliers were then removed from the state coefficient matrices using a median absolute deviation (MAD) approach and the center point and resulting distance threshold were saved for artificial parameter rejection through the data pipeline B after conditional draw, as Table 4.2 explains.

4.2.5 Copula Approach

A Gaussian copula allowed a probabilistic description of parameter dependencies with a single, simple CDF. For each identified cycle-state, one copula was defined.

4.2.5.1 Conversion to Uniform Distributions

Each of the columns $\boldsymbol{d}_p$ of the data structure matrix D follow unknown distributions X_p of the corresponding parameter values. To transform these into marginal distributions for the Gaussian copula, they need to be represented as uniform probability distributions. Hence, the empirical CDF was calculated for each column vectors $\boldsymbol{d}_p$. However, if the cycle-state has B segmented beats, the empirical CDF is a step function with $\frac{1}{B}$ discrete jumps, meaning few beats or close parameter values will result in a highly discrete function with limited ability to extrapolate the CDF of a different set of parameter values. Moreover, $\boldsymbol{d}_p$ might follow any arbitrary distribution, requiring a flexible non-parametric function. This was overcome with a lightweight piecewise linear approximation approach which linearly connects the midpoints of each discrete step of the empirical CDF, as in Figure 4.11. As showcased in Sections 2.2.1.1 and 2.2.1.2, any distribution with a monotone CDF can be transformed into a uniform distribution $\mathcal{U}(0,1)$ by means of the PIT with the inverse of its CDF $F_{X_p}^{-1}$ which was, at this point, straightforward, given the valid CDF approximation F_{X_p}.

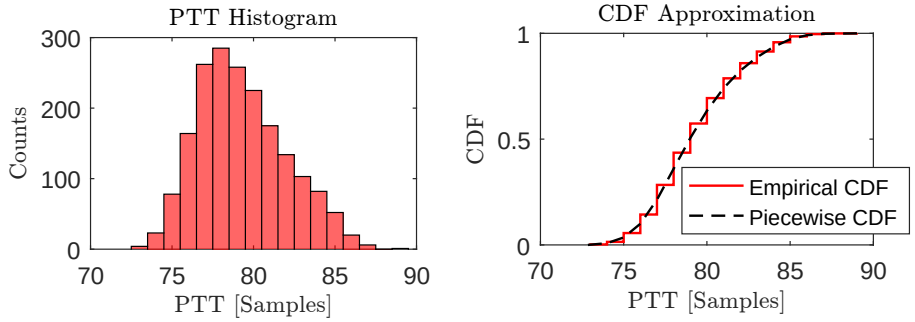

Figure 4.11: Histogram of PTT-allusive delay parameter for the BP modality of subject slp59m of the MIT-BIH Polysomnographic database (on the left) and the respective CDF (on the right). It can be seen that the piecewise approximation provides a better estimate of the real underlying CDF.

4.2.5.2 Gaussian Copula

A P-dimensional copula per cycle-state $C : [0,1]^P \rightarrow [0,1]$ could then be defined with the previously calculated uniform distributions $F_{X_p}^{-1} \sim \mathcal{U}(0,1)$, for each estimated parameter vector $\boldsymbol{d}_p$ as follows:

$$C(\boldsymbol{f}) = \mathrm{P}[F_{X_1}^{-1} \leq f_1, \ldots, F_{X_P}^{-1} \leq f_P],$$

where $\boldsymbol{f}$ denotes the vector of uniformly distributed values $f_1, \ldots, f_P$, of equal dimension to that of the number of parameters and, therefore, number columns of the matrix $\boldsymbol{D}$. The final component to define C, is a positive and semi-definite correlation matrix, which, as explained in Section 2.2.1.3, was calculated by means of Pearson's ρ_{p}, modeling linear correlations between the segmented beats' multimodal parameters.

4.2.5.3 Conditional Draw from Copula

The same process used so far to define the copula function could be reversed in full, using the mathematical tools explained in Section 2.2.1.2 to obtain new, artificial parameters according to their initial unknown distributions in a similar fashion to that of [29]. Moreover, assuming correct dependence modulation between the copula parameters, one or more of them could be used as control variables, leveraging conditional probability. This process was done by means of conditional draw, utilizing two variables as control - RR-interval and respiratory phase ϕ_{RESP} - to generate a conditional CDF

$$G_{f_3,\ldots,f_P}(f_3,\ldots,f_P|\mathrm{RR},\phi_{\mathrm{RESP}}) = \frac{\partial_{\mathrm{RR},\phi_{\mathrm{RESP}}} C(\mathrm{RR},\phi_{\mathrm{RESP}},f_3,\ldots,f_P)}{\partial_{\mathrm{RR},\phi_{\mathrm{RESP}}} C_{\mathrm{RR},\phi_{\mathrm{RESP}}}(\mathrm{RR},\phi_{\mathrm{RESP}})}.$$

Doing so allowed accounting for respiratory modulation and arrhythmogenic effects explained in Sections 2.1.2 and 2.1.4.3 brought by a synthetically calculated RR-tachogram, respiratory phase series and pseudo-randomly generated primers for every conditional parameter $f_3, \ldots, f_P = G^{-1}_{f_3,\ldots,f_P}(f_3', \ldots, f_P'|\mathrm{RR},\phi_{\mathrm{RESP}})$.

A copula C outputs a vector of uniformly distributed variables $\boldsymbol{f}_{\mathrm{artifial}} = f_{\mathrm{artifial},1}, \ldots, f_{\mathrm{artifial},P}$, encoding one multimodal cycle. Subjecting each variable to ITS yielded new coefficients, $\boldsymbol{d}_{b,\mathrm{artifial}}$ following their initial probability distributions X_p. In fact, an artificial matrix $\boldsymbol{D}_{\mathrm{artifial}} = (d_{bp,\mathrm{artifial}}) \in \mathbb{R}^{B_{\mathrm{artificial}} \times P}$ was calculated, whose row vectors were reused with the initial user-specified functions

(variable fit_types as initialized in Listings 4.2, 4.3 and 4.4) to produce artificial, conditional beats of a cycle-state.

4.2.6 Signal Generation

Finally, fitted copulas $C_s, s = 1, \ldots, s_{N_S}$ for the N_S cycle-states and the defined Markov model are jointly used to produce full recording-specific cycles dependent on a given set of RR-interval $RR_{\text{artificial}}$, respiratory phase values $\phi_{\text{RESP, artificial}}$ and arbitrary number of cycles $B_{\text{artificial}}$. A cycle-state evolution vector with $B_{\text{artificial}}$ elements is sampled from the Markov model and a copula C_s is assigned to its corresponding $B_{\text{artificial, s}}$ cycles.

Initial copula definition was executed in R by setting up a common working directory, where .mat data files containing the parameter matrices $\mathbf{D}_s$ specific to a cycle-state s were saved, loaded and processed, resulting in an R object file containing the cycle-state copula C_s being worked with. However, two pipelines were used to produce the final parameters and only Sections 4.2.5.2 and 4.2.5.3 were handled with R scripts. Additional processes were conducted with Matlab®-two data flow structures, represented in Table 4.2, were implemented, relying on .txt files with a single integer, allusive to the Markov state of the current cycle, passed onto R informing which one copula to use at a time. These are described as follows:

A. R and Matlab® only communicate once per cycle-state. For this, a matrix of $B_{\text{art, s}}$ rows of $RR_{\text{artificial}}$ and $\phi_{\text{RESP, artificial}}$ values are passed onto an R script along with the cluster center $\mathbf{d}_{b,s}$ and euclidean distance threshold used to perform cluster-based removal in Section 4.2.4.3. An R script initializes $B_{\text{art, s}}$ rows of $P_s - 2$ pseudo-randomly generated uniform values to perform conditional draw and repeats the process until all the artificially sampled coefficients are validated and ultimately stores them in a .mat file in the directory. From here, all sampled cycles, previously cluster-based validated, are processed with ITS in Matlab® as $\boldsymbol{D}_{\text{artifial}} = (d_{bp,s,\text{artifial}}) \in \mathbb{R}^{B_{\text{art, s}}, P_s}$ and reconstructed into a full recording;

B. For conditional draw, .mat data files containing a pair of $RR_{\text{artificial}}$ and $\phi_{\text{RESP, artificial}}$ followed by $P_s - 2$ pseudo-randomly generated probability values are passed onto an R script, which performs conditional draw, ultimately saving a new .mat file with $\boldsymbol{f}_{\text{artifial,s}} = f_{\text{artifial,s,1}}, \ldots, f_{\text{artifial,s},P_s}$ to the directory. Matlab® then loads it to produce a single cycle $\boldsymbol{D}_{\text{artifial}} = (d_{bp,s,\text{artifial}}) \in \mathbb{R}^{1,P_s}$ with ITS and validate it through an RMSE filter with the respective mean waveform $\bar{y}_s$ (similar to the one implemented in Section 4.2.4.3). If a cycle is rejected, the process is repeated, starting with a different, randomly generated uniform vector until the cycle is accepted. This routine is performed until $B_{\text{art, s}}$ are produced and validated.

While pipeline B allows for more rigorous filtering of the artificial cycles, A allows for reasonable results to be achieved much faster. Thus, in cases where the original data does not result in a stable probabilistic representation, or for which a Gaussian copula is not suited B is preferable, while A is better when long artificial multimodal signals are needed. In both, the final output signal is, ultimately, obtained by concatenating the generated cycles according to the cycle-state evolution vector.

Table 4.2: Scheme of data pipelines, represented by both rows where data flows from left to right columns. While pipeline A only needed to be run once (since all the control parameters and random primers are passed to R and used in conditional draw in a joint fashion) pipeline B was externally run the same number of times as required beats (since only a set of parameters and primers is sent to R at a time).

	Matlab® (pre-processing)	R (copula sampling)	Matlab®(cycle reconstruction)
A	$B_{\text{art, s}}$ sets of $RR_{\text{art}}, \phi_{\text{RESP, art}}$ Cluster center $\bar{\mathbf{d}}_b$ Euclidean distance threshold	$B_{\text{art, s}}$ sets of $f_3', \ldots, f_P'$ Conditional sampling Cluster validation loop	$B_{\text{art, s}}$ sets of $RR_{\text{art}}, \phi_{\text{RESP, art}}, f_3, \ldots, f_P$
B	set of $RR_{\text{art}}, \phi_{\text{RESP, art}}, f_3', \ldots, f_P'$	Conditional sampling	set of $RR_{\text{art}}, \phi_{\text{RESP, art}}, f_3, \ldots, f_P$ RMSE validation loop with $\bar{y}_s$

4.3 Deep Learning Algorithm

The current section details the DL approach used to handle cardiorespiratory data from the architecture to training, testing and additional signal processing steps. Besides the methodology, the decision-making and reasoning behind it is also addressed.

4.3.1 Convolutional Neural Network Architecture

As presented in Section 3.2, there is currently no consensus on the best way to perform sensor-fusion. In fact, the presented networks exploited both early and late-fusion compellingly, achieving high scores regardless of the approach. The CentralNet rationale, introduced in Section 3.2.2 is transferable to structures applied to cardiorespiratory data and makes for more flexible representations, reaping the benefits from both early and late-fusion. Given the type of multimodal data intended to be handled by the designed CNN, lower-level features are frequently uncorrelated, meaning the use of 2D filters for sensor-fusion of stacked multimodal channels does not make as much sense as when applied to similar inputs such as a 12-lead ECG, as previously discussed in Section 3.2.1. For this reason, the use of the multi-branch architecture with 1D inputs was adopted as, with the CentralNet adaptation, the best strategy for fusion can be learned. These include strategies used to support 2D filters, like favoring early-fusion when inputs are similar (such as ECG and cECG), favoring late-fusion when low-level features are scarcely correlated (such as ECG, BP and BCG, or when there were acquisition defects in the channels) or even a combination of both. Furthermore, the principles of multimodal integrity and diversity introduced in the lastly mentioned section can still be respected.

The previously presented CNN architecture in [98], represented in Figure 3.17, was chosen as a starting point for featuring a multi-branch structure, having dealt with the most types of modalities tackled in this thesis and yielding relevant results in a similar set of tasks. The convolutional branch structure was kept the same, but with M architectural twin branches (one per modality), as shown in Figure 4.12. Alterations to the original structure are:

1. Inclusion of an $(M+1)$-th fusion branch, labeled with subscript 0, whose feature maps $\mathbf{H}$ are the result of a linear combination of all the branches' m feature maps from the corresponding layer i, including themselves

$$\mathbf{H}^{[0,i]} = \alpha^{[0,i]}\mathbf{H}^{[0,i]} + \sum_{m=1}^{M} \alpha^{[m,i]}\mathbf{H}^{[m,i]}.$$

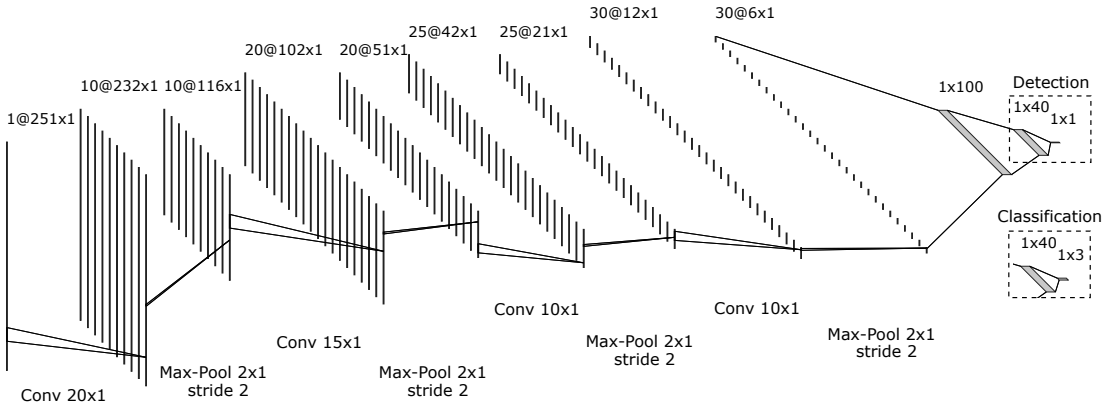

Figure 4.12: Simplified representation of the proposed architecture (only one of the branches). The same notation as in Figure 2.16 is used, although every height parameter was reduced to 1 given both the input and convolutional layers are 1D. The prediction layer changes from 1 to 3 output nodes depending on the intended task.

Linear combination coefficients α were uniformly initialized. This notation is an adaptation of the one used to describe CentralNet, more adequate to the context and in line with what is frequently used in the DL community;

2. Each branch, except the fusion branch, receives a 251-sample snippet, synchronous with the remaining ones. This increase in signal length is motivated by the possibility of providing a wider viewing window capable of accommodating bigger portions of the snippet, causing characteristic waveforms to completely be included. Such a change was expected to elevate the number of TPs at the cost of an increased absolute error (between the predicted peak location and the QRS-annotation), a case that was analyzed with the obtained results. Having been successfully utilized with ECG and BP in [96] for early-fusion, it was thought that incorporating a bigger portion of the BP pulse wave could both elevate activations in the respective branch and also potentiate lower-level correlation with the ECG's T-wave. Such could rehabilitate early-fusion, which would, otherwise, be harder with a smaller snippet;

3. Separation of the FCN into $M + 1$ replicas, each summarizing their respective branch, as schematized in Figure 4.12. The FCN structure was not reduced since a similar number of features to the original was obtained in the last convolutional layer;

4. Replacement of the softmax activation function at the output layer by a sigmoid for the detection task, when the possible classes are 1 or 0, as a simplification to avoid an extra neuron, parameters and computational speed with negligible changes to the outcome. Equivalence is supported by conversion of a binary classification softmax with inputs x_0 and x_1 into a sigmoid with only one x_1

$$\begin{aligned} \text{softmax}(x_1) &= \frac{e^{x_1}}{e^{x_0} + e^{x_1}} \\ &= \frac{e^{x_1 - x_0}}{1 + e^{x_1 - x_0}} \\ \underset{(x_0=0)}{=} \sigma(x_1) &= \frac{e^{x_1}}{1 + e^{x_1}}, \end{aligned}$$

assuming one of the two inputs in the softmax function always yields zero as a result of training;

5. Insertion of 2-sample pooling layers with stride 2 between each consecutive convolutional layer in order to increase abstraction of the enlarged snippet's features;

6. Batch normalization was applied to every layer, except the output, to increase training speed, shrinking computational time required for optimizations, and pattern recognition invariance. Theoretically, it should ease beat identification despite varying waveforms stemming from experimental configurations and conditions or changes with regards to subject-specific nuances, such as variations in elderly and young patients in the BP modality, explained in Section 2.1.3.2. Consequently, the chances of overfitting should decrease. This operation can be mathematically described by

$$\mathbf{H}^{[m,i+1]} = \mathbf{J}^{[m,i]} \left(\frac{\mathbf{H}'^{[m,i]} - \boldsymbol{\mu}^{[m,i]}}{\boldsymbol{\sigma}^{[m,i]}} \right) - \boldsymbol{\beta}^{[m,i]},$$

   where $\mathbf{H}'$ indicates the feature maps after the application of ReLU, $\boldsymbol{\mu}^{[m,i]}$ and $\boldsymbol{\sigma}^{[m,i]}$ are the mean value and standard deviation matrices corresponding to the elements of $\mathbf{H}'^{[m,i]}$, containing feature maps from one batch of samples. $\mathbf{J}^{[m,i]}$ and $\boldsymbol{\beta}^{[m,i]}$ are added learnable corrections to the distribution since, for example, null mean elements with slight deviations can yield very small values after applying certain activation functions, such as the sigmoid. The terms $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are calculated before each batch of an epoch is processed and are used for the remaining samples of that batch;

7. As a consequence of the previous alteration, a simplification was conducted by removing the bias of each layer. By adding learnable parameters $\boldsymbol{\beta}$ to the feature maps, keeping the learnable bias terms B becomes unnecessary as it would force the network to learn how a sum could amount to a single parameter $\boldsymbol{\beta}' = \mathrm{B} + \boldsymbol{\beta}$, which can be directly learned;

4.3.2 Dataset Management

Datasets referred in Section 4.1 were divided in order to perform the tasks explained in the following ones. In general, each task counted with a training, validation and test set, except for hyperparameter optimization, where no test set was used. Similarly to what was done in previous, introduced works, no recordings nor subjects were shared among the sets. While training sets were comprised of a mixture or exclusively real or artificial recordings, validation and test sets were solely composed of real recordings, in an attempt to create a realistic DL research scenario.

For the optimization, only recordings from the MIT-BIH Polysomnographic database were used since the process would become too cumbersome otherwise and it covers most of the modalities targeted in this thesis. From a pool of real recordings from subjects slp59m, slp60m, slp61m, slp66m and slp67xm, leave-one-subject-out cross-validation was performed in order to obtain the most robust hyperparameters. These hyperparameters and respective optimization search-space are summarized in Table 4.3.

After parameter optimization, a series of experiments were conducted, which included training, validation and testing, with different proportions of real and artificial data in the training set. From Fantasia, recordings of subjects f2o05m, f2o07m, f2o08m, f2o09m, f2y03m, f2y07m and f2y09m were used for training, from f2y04m for validation and from f2y05m for testing. From MIT-BIH Polysomnographic, recordings of subjects slp59m, slp60m and slp67xm were used for training, from slp61m for validation and from slp66m for testing. From MIT-BIH Arrhythmia,

Table 4.3: Optimization search-space.

Optimization	Parameter	Search-space	Type
Grid search	Optimizer	Adam or SGD	Categorical
	Initialization	Uniform or Gaussian	Categorical
Bayesian	Epochs	$\{40, 45, \ldots, 95, 100\}$	Discrete
	Batch Size	$\{64, 125, 256\}$	Discrete
	Dropout Probability	$[0.2, 0.6]$	Continuous
	L1	$[0.01, 0.40]$	Continuous
	L2	$[0.01, 0.40]$	Continuous
	Learning Rate	$[1.0 \times 10^{-4}, 0.1]$	Continuous

recordings of subjects sub208, sub113, sub115 and sub230 were used for training, from sub119 for validation and from sub221 for testing. This selection was based on which subjects yielded better artificial signals from the synthesizer - subjects who produced the worse quality artificial signals were included in the validation or test sets, so as to only use their real recordings. Moreover since only slight performance deviations between folds were registered during optimization using the dataset with most variability among recordings, no further cross-validation was used in order to spare computational time, by assuming the use of a single fold produced trustworthy metrics.

Pre-processing was necessary to produce the input snippets required by the model. Regardless of recordings being real or artificial, each was composed by a multimodal signal sampled at 250 Hz with one numerical annotation per sample - 0 if not a beat, 1 if the sample is the QRS-annotation of a normal beat or 2 if a PVC beat (in the case of recordings from MIT-BIH Arrhythmia). The same procedure used in [98] was adapted, where a non-zero label snippet is defined as being centered at the QRS-annotation and a zero-label snippet is produced with a smaller shift of a minimum of 5 samples from the current QRS-annotation and such that it is centered further than 5 samples from adjacent beat QRS-annotations. This method ensures data balance, which, in this case, is necessary given that, both for detection or classification, it is important that the algorithm performs well in every class [115]. The empirically determined reduction in the number of samples shifted to obtain zero-label snippets was assumed to potentiate the algorithm's temporal accuracy of detection without making it too confusing to discern from non-zero-label snippets. Furthermore, in compliance with the inclusion of batch normalization layers, the signals, too, were standardized.

4.3.3 Artificial Data Experiments

This section synthesizes the steps to validate the designed network, the quality of the produced artificial signals and, most importantly, evaluate the effects of real dataset data augmentation in the performance of DL methods. For the first and second objectives, control results were obtained in order to be measured against the ones yielded by previously presented state of the art architectures. The third objective explored scenarios where data augmentation is performed even though enough real data is available, when real data is insufficient, corrupted by noise and when a classification task is intended but not enough beat types are present in the real data, always using the control results as reference. The MIT-BIH Arrhythmia database was used for the classification experiment, while the Fantasia and MIT-BIH Polysomnographic databases were used in the remaining with the exception of the artifact experiment, where only the latter was

used.

Supervised training with the above-mentioned snippets was conducted using the Adam optimizer with a learning rate of 3.149×10^{-4}, as obtained through hyperparameter optimization, and a β_1 of 0.900 and β_2 of 0.999. With the exception of the last layer, dropout with 0.318 probability was used in every layer of the FCN. No dropout was used for CNN layers since, in that case, the process works in a different, less intuitive way. In [116], it is shown that dropping out filter weights by making them null, does not necessarily mean they are not being trained and, in fact, introduces Bernoulli noise to the feature maps. Additionally, L1 and L2 regularization terms were included to the CE loss function, using linear combination weights $\boldsymbol{\alpha}$ with a λ_1 of 0.013 and remaining weights $\mathbf{w}$ with a λ_2 of 0.3432, respectively. This decision was based on allowing modality contributions to the same fusion layer to be as sparse as possible if need be, while regular weights should only be marginally different and softly updated, allowing the network to adapt faster, similarly to what happens in batch normalization. After hyperparameter optimization, the maximum number of epochs was set to 85 and the batch size to 64 samples. To prevent overfit and unnecessary computational time, an early stopping approach was used in order to halt the training process when a hand-crafted stoppage metric presents a non-decreasing tendency for 5 consecutive epochs (patience factor). The mentioned stoppage metric was set to strike a balance between score and error, giving the first 85% importance and the second 15%. After empirically verifying detection rarely resulted in errors above 30 ms, a cutoff was introduced to the error portion of the metric, where its contribution would be null past that threshold. The obtained expression was

$$0.85 \cdot \text{score} + 0.15\left(1 - \frac{\text{error}}{30}\right).$$

4.3.3.1 Artificial/Real Proportion

Control results were obtained from using exclusively snippets from real recordings for training. As a way to test the autonomy of the synthesized multimodal signals, only artificial recordings, approximately with the same length as the real recordings, were used in training in a posterior experiment. Finally, after empirically verifying the amount of real training data in the first experiment is sufficient to produce reasonable results, training sets with different proportions of artificial and real signals were used for training so as to evaluate if data augmentation results in any improvements or degradation. Specifically proportions tested were 0.25, 0.50, 0.75, 1, 1.25, 1.50, 1.75 and 2.

4.3.3.2 Insufficient Real Data

To simulate the type of scenario where data augmentation is most likely to be used in, real recordings part of the training set were shrunk by a number of factors until the length of data was not able to train a performant network. Specifically, the factors were 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 0.005, 0.004, 0.003, 0.002 and 0.001. Afterwards, the real recordings were kept shrunk and a similar process as in Section 4.3.3.1 was used to integrate artificial recordings in different proportions of their original counterparts length: 0.25, 0.50, 0.75 and 1.

4.3.3.3 Robustness to Artifacts

It was noticed that in noisy portions of the signal, especially when SNR was very low as in artifacts, the output probability of positive classification increased, reducing the CNN's performance. As such, in an attempt to work around this issue using data augmentation, an experiment was devised with MIT-BIH Polysomnographic recordings, due to them featuring the most modalities: in a first phase, a training set with only high-SNR real recordings was used to train a CNN to be tested on real recordings with incorporated corrupted segments where one, two or all three modalities were corrupted by abstracts; in a second phase, the training set was also incorporated with artificial recordings with artifacts, as exemplified in Figure 4.13; and trained against the same validation and test sets. Both results were compared and checked for improvements in beat detection.

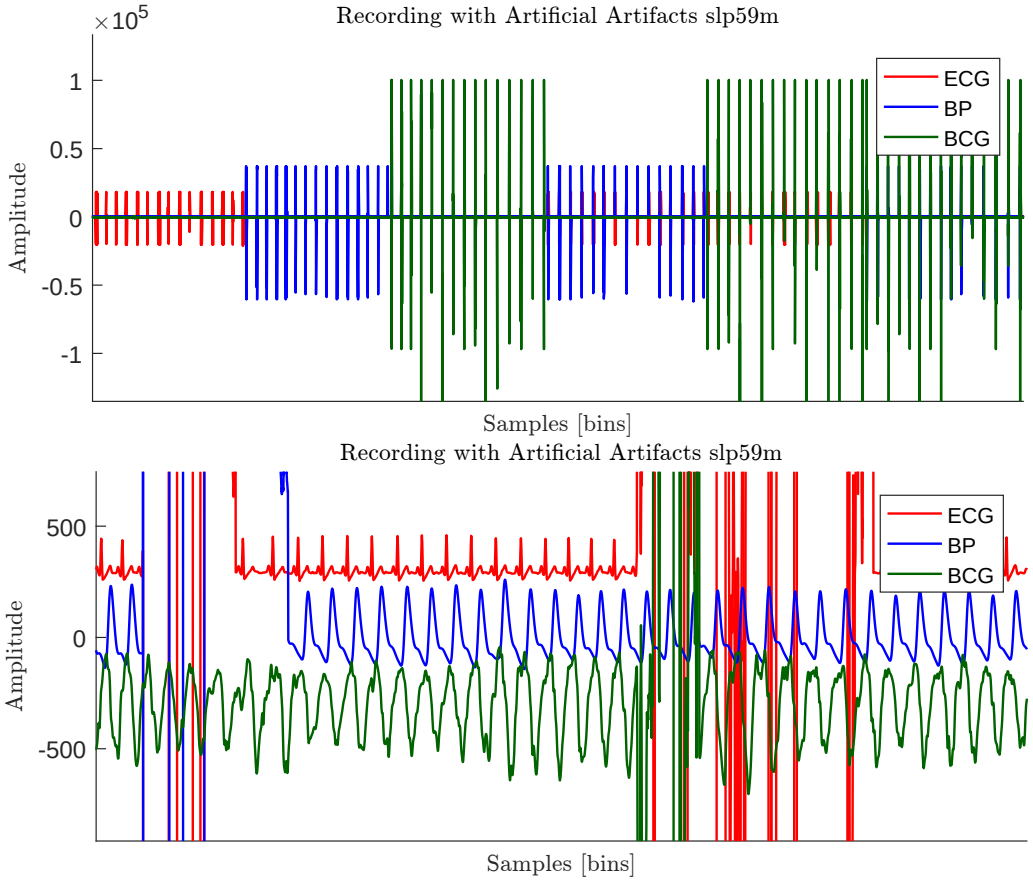

Figure 4.13: Excerpt of the artificial recording produced from subject slp59m of the MIT-BIH Polysomnographic database with artifacts. At the top, an overview of the signal with portions of modality-independent artifacts, covering every combination of artifacts in modalities without contaminating them all simultaneously. At the bottom, a closer view of the same recording.

4.3.3.4 Classification Tasks

In order to check the capability of the CNN to also perform classification tasks with the help of data augmentation, real recordings showing no signs of pathological beats, whose snippets were labeled as 0 or 1 integrated the training set along with an artificial recording synthesized based on a PVC recording, whose beats are labeled with 2. Validation and training sets both included real recordings where all three types of beats existed. Before training, the CNN was adapted to its

3-output softmax form and metrics were employed to measure the performance of the algorithm at both classifying and delineating label 1 and 2 beats.

4.3.4 Test Metrics & Post-Processing

It was important to use the CNN to process every snippet in a full data recording chronologically. Such allowed testing it in what would be a real-world scenario for detection algorithms, with the original data balance, evaluate the error compared to annotations and its feasibility in real-time. The raw output of the CNN detection variation would be a single channel with the probabilities of each signal snippet being a beat, which translated into three channels with the probabilities of the snippets not being a beat, being a normal beat or a PVC beat for the CNN classification variation, in which case only the last two channels were used. Post-processing was performed resorting to a prominence-adaptive peak detection routine to identify probability peaks, which were assumed to be the CNN's best guess, and their respective locations produced a train of labels, the same length as the original signal, comparable to the one used to supervise the training phase. By cross-checking with the real annotations, the score and error metrics were calculated in a similar fashion as in [98]. The former was the mean value between SE and PPV, while the latter was the average time distance, in ms, between a predicted non-zero label and the nearest real occurrence of the same label. Within the location pulse around a given real non-zero label, as stipulated in the standards [97]: a prediction of the same label as the annotated was considered a TP, a FN if no predictions were made or as FP if multiple non-zero label predictions were counted. Further, outside those windows, any non-zero label was considered a FP.

Chapter 5

Results & Discussion

5.1 Multimodal Synthesizer

In order to evaluate the quality of synthesized multimodal signals, these were compared to their real counterparts. All four datasets involved - Fantasia, MIT-BIH Polysomnographic and Arrhyhtmia and UnoVis - were used so as to establish a common base for a three-fold analysis based on the time, frequency and phase domains, as previous works in Section 3.1 have done. Additionally, the output of different components of the algorithm were analyzed so as to describe and validate the basic functioning of the presented framework to provide a solid basis of documentation for future development.

5.1.1 Framework Validation

In this subsection, intermediate results of the synthesizer's components are shown. The adopted multimodal strategy was characterized in terms of its effectiveness at utilizing the original data and quality of segmentation. It was checked whether obtained Markov Models learned physiological transition probabilities. Finally, copula objects were investigated in order to compare how realistic the sampled data is and if there would be differences among the usage of linear and rank correlation coefficients.

5.1.1.1 Cycle Segmentation

The cycle segmentation and feature extraction processes described in Sections 4.2.3, which included data rejection routines, guaranteed the automaticity of the algorithm (allowing input of full, unprocessed datasets) by making sure that used segments were, for one, physiologically valid and uncorrupted and, for another, adequately representable by the mathematical model. Invalid data leakage would prove to be harmful to a generative probabilistic algorithm's ability to learn associations in the original data by introducing residual wrongful probabilities to the dependence structure. As such, the number of output segments from these two filtering steps were compared to the corresponding original number of annotated QRS locations in Table 5.1. Final segments, exemplified in Figure 5.1, were checked for segmentation quality and alignment.

Table 5.1: Coverage statistics of the algorithm for every recording of each dataset.

	Coverage (%)			
Database	Mean	Standard Deviation	Maximum	Minimum
Fantasia	74.0	16.4	97.1	51.4
Polysomnographic	87.1	15.8	97.9	59.2
Arrhythmia	72.3	15.0	97.1	53.8
UnoVis	44.6	26.8	97.0	5.4

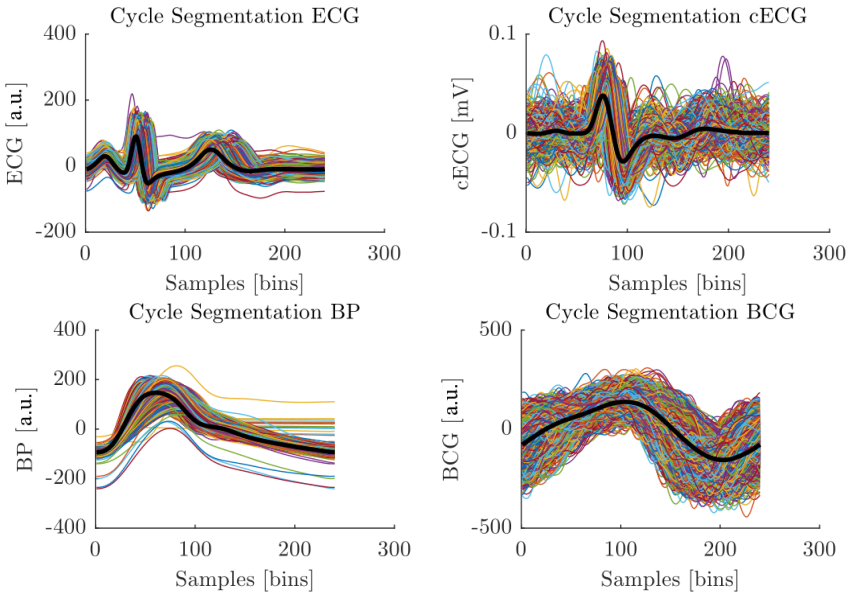

Figure 5.1: Stacked segmented cycles and respective average waveforms (in black) for ECG, BP (subject f2o04m of Fantasia), cECG (first recording of subject 1 of UnoVis) and BCG (subject slp59m from MIT-BIH Polysomnographic) modalities.

Reasonable coverage values can be observed with the exception of the UnoVis database. This was expected since, in the latter, three out of the four channels in each recording were unobtrusively set up, being more prone to the occurrence of artifacts and noisy segments due to a low-SNR. Moreover, some of the shorter recordings (some less than 5 minutes long) yielded very little information after segmentation and, ultimately, respective generated signals were not used in the training of the CNN given their low quality. Such was the case of the recording of which only 5.4% was used, as seen in Table 5.1, and the reason why coverage standard deviation increases in UnoVis.

Figure 5.1 shows that, despite coverage values far from 100%, segmentation was rigorous to the point of aligning segments very consistently, as can be verified by the P-wave alignment in the ECG modality, setting up a trustworthy baseline for the next algorithm components which allows intra- and inter-modality comparison. In cECG and BCG modalities, volatile signal portions can be seen (except for the QRS-complex of the former), where segment signals appear to diverge at almost every point resulting in a dampened average waveform, as seen in the flattened P-wave for cECG average signal. The opposite happens with the remaining, clinical modalities, where portions' signals are highly correlated. This also validates the principle upon which the mathematical representation was chosen - using GMM for high-SNR portions and FS or SoS for low-SNR ones

to accommodate higher variance. In fact, the flexibility of the model is further demonstrated in Figure 5.2.

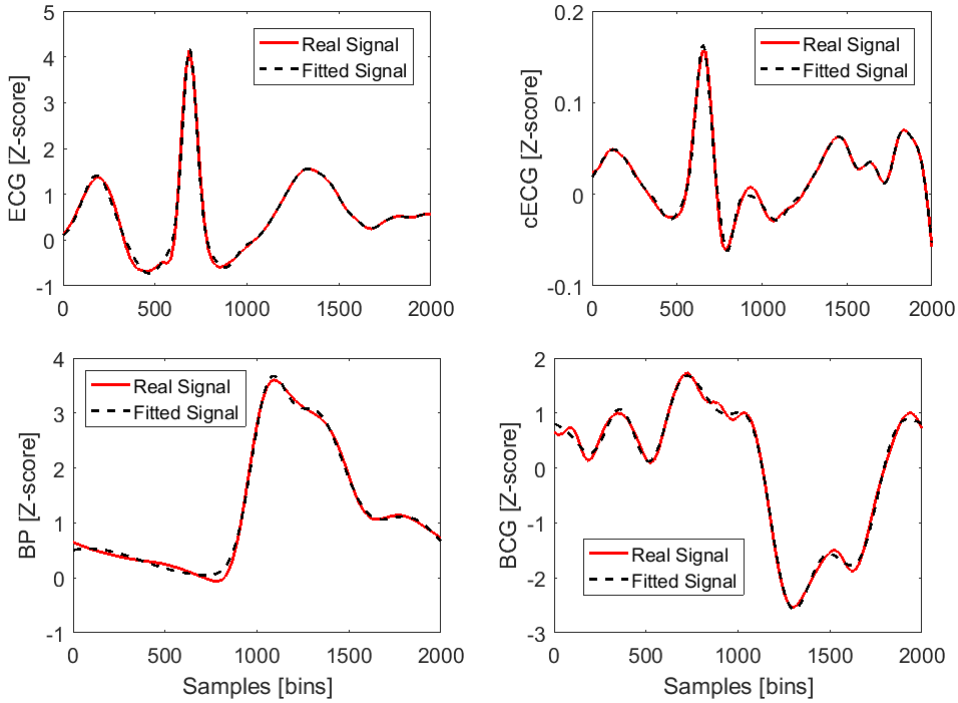

Figure 5.2: Fitting process for a segment of each modality respective to the recordings shown in Figure 5.1.

Some segments, such as the BCG segment represented in the lastly mentioned figure, present a significant high frequency component which cannot be exactly modeled by the current fitting algorithm. Division of the segments into portions alleviates this problem, when portion delineation is possible, resulting in well defined physiological components (such as P, Q, R, S and T waves), including noisy transition portions, and smooth transitions among them.

5.1.1.2 Copula Framework

Some examples of the fitted copulas' correlation matrices are shown in Figure 5.3. The higher order functions and more portions are used to define segments, the more coefficients are needed to describe them, resulting in larger matrices. MIT-BIH Polysomnographic's correlation matrices have more coefficients than Fantasia's due to the extra modality, while UnoVis recordings yield much larger matrices given three of the four modalities are unobtrusive and require modeling with a high order of SoS. From experimentation: ECG modalities are usually well represented with 2^{nd} order GMM for every portion except the Q-wave, for which 1^{st} order suffices; BP with 3^{rd} to 5^{th} order GMM; BCG and cECG's low-SNR portions with 8^{th} order FS or SoS, respectively; noisy transitions with 8^{th} order SoS. In GMM, 2^{nd} order is used to account for portions' varying skewness, as previously pointed out in [29, 76, 77, 79, 82].

Through analysis of the correlation matrices' values, it was shown that correlation between RR-intervals and PTT takes values between -0.20 and -0.30. Although this result is in agreement with what was obtained in [80], the same paper reported a correlation of 0.69 can be obtained when delaying the PTT by 3.17 ± 0.76 beats. Such is suggestive of the proposed model's inability

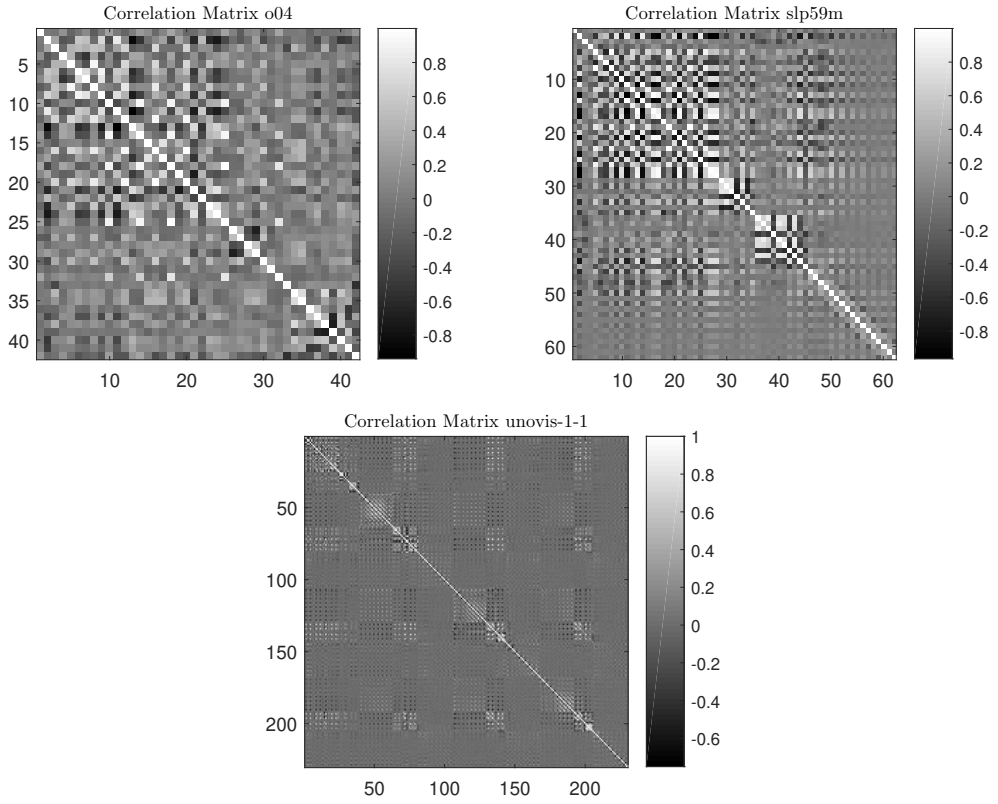

Figure 5.3: Linear correlation matrices for subject f2o04m of Fantasia, subject slp59m from MIT-BIH Polysomnographic and first recording of subject 1 of UnoVis, from top to bottom. The matrices are symmetrical, with unitary diagonals and have $P \times P$, where P is the number of parameters used to describe one recording.

to model out-of-phase dependence, as is the case for PTT changes due to past RR-interval values. However, in-phase correlations can be satisfactorily modeled like additive and multiplicative respiratory modulation used in [6], illustrated by high correlation values between respiratory phase and amplitudes of GMM and FS, as evidenced by Figure 5.8, corresponding to, for example, the R-wave and BCG amplitude. Moreover, as studied in [117], PTT is negatively correlated with BP amplitude since when arterial tension drops, pulse velocity does as well, ultimately increasing PTT, as correlation matrices usually depict with correlation coefficients between PTT and GMM amplitudes of BP -0.5 and -0.7. In fact, significant positive correlation among PTT and location of GMM corresponding to diastolic portions of the BP wave exists, suggesting the model is trying to anticipate diastolic components to overlap systolic ones and yield a higher SBP. At the same time, no degradation of the typical waveform is observed, as the model probably compensates with an increased width of said GMM components.

Another important aspect about this correlation-based modeling is the dependence between ϕ_{RESP} and RR-intervals. Theoretically, correlation between both these variables should be meaningful, due to the RSA phenomenon, which is not verified in the obtained correlation matrices. Nonetheless, both play their roles independently in respiratory modulation and preservation of subject-specific HRV effects, including RSA and Mayer waves, respectively. A problem arises when ϕ_{RESP} and RR-interval values are fed to the model without the user manually correlating them, resulting in uncorrelated ϕ_{RESP}- and RR-interval-derived effects such as the above mentioned

ones. When perpetuating their correlations matter, the ϕ_{RESP} and RR-interval values used for conditional sampling should be the same as the ones extracted from the original recordings or, if a longer artificial signal is desired, could be obtained by dynamical coupling, similar to what was used to couple HR, RSA and Mayer waves in [6]. It is possible that adding this deterministic component to the model even potentiates the remaining correlation coefficients.

Ideally, sampling from the copula should provide artificial coefficients which follow the same distributions as the original distributions that originated them. Moreover, the linear correlation matrix of the artificial coefficients should also remain the same. In order to keep track of the change in distributions and linear correlations from the input to the output of the copulas, calculations of the histogram intersection metric for the former and the difference per element of the matrices for the latter were performed and are shown respectively in Tables 5.2 and 5.3. For this, full input recordings of the dataset were used and artificial recordings with 1000 heart cycles were generated. Histogram intersection is a histogram similarity metric mostly used in histogram-based image classification algorithms [118, 119]. It is mathematically defined between a reference histogram R and another histogram I as

$$\frac{\sum_{j=1}^{n} \min\left(I_j, R_j\right)}{\sum_{j=1}^{n} R_j},$$

where j represents each of the histogram values respective to I_j and R_j counts. Due to the denominator, the metric takes values between 0 and 1 and the higher, the more similarity exists among histograms.

Table 5.2: Statistics of the histogram intersection metric between input and output marginal distributions for every recording of each dataset.

	Histogram Intersection			
Database	Mean	Standard Deviation	Maximum	Minimum
Fantasia	0.78	0.19	1.00	0.61
Polysomnographic	0.86	0.14	1.00	0.76
Arrhythmia	0.90	0.16	1.00	0.79
UnoVis	0.84	0.17	1.00	0.79

Table 5.3: Statistics of the matrix error per element between input and output correlation matrices for every recording of each dataset.

	Matrix Error per Element			
Database	Mean	Standard Deviation	Maximum	Minimum
Fantasia	6.43e-2	8.14e-2	2.65e-1	4.87e-02
Polysomnographic	9.67e-4	3.62e-2	5.80e-2	3.05e-04
Arrhythmia	1.15e-3	4.92e-2	6.14e-2	5.42e-04
UnoVis	1.33e-2	4.31e-2	8.22e-2	8.30e-02

The content in the tables reports accurate sampling of features using the copula and the conditional draw method described in Section 2.2.1.2. In general, high similarity was found between original and artificial marginal distributions as well as low correlation errors (which in only one instance was higher than 0.1 in a scale from 0 to 1) even in the worst cases, highlighting the

reliability of the framework at handling this particular type of coefficient data. In particular, the Fantasia database seems to achieve the worst results probably because of the more variable portion shapes in clinical modalities in relation to the remaining databases, which might make it hard for a simple Gaussian shaped copula to learn the complexity of coefficient variation. However, it is reasonable to infer that also this component of the algorithm performs properly and satisfactorily, validating it for handling this type of data representation.

Additionally, a study of the adequacy of linear correlation for the projects goals was evaluated by comparing correlation matrices calculated with Pearson's ρ_p, Spearman's ρ_s and Kendall's τ for input data from all datasets. Results of the total difference among these matrices are shown in Table 5.4.

Table 5.4: Comparative use of linear and rank correlation matrices.

	Total Matrix Absolute Error			
Database	Mean	Standard Deviation	Maximum	Minimum
Spearman's ρ_s	2.22e-2	1.04e-2	4.23e-2	7.30e-3
Kendall's τ	2.23e-1	9.80e-2	4.09e-1	9.51e-2

Since insignificant differences were observed, it was assumed that the use of a rank correlation metric, capable of depicting non-linear correlations as explained in Section 2.2.1.3 would not improve the model's performance and linear correlation continued to be assumed fit to describe the adopted data representation. This also means that the types of dependence this model is capable of learning are predominantly linear.

5.1.2 Waveform Generation

Resemblance between generated and real waveforms has been explored through qualitative and quantitative analysis in the time, frequency and phase domain.

5.1.2.1 Time Domain

Signal morphology and respective variations are probably the most important characteristics to be properly mimicked by artificial recordings since they form the patterns which a CNN can learn to perform accurate predictions on real data. For the ultimate purpose of classification, making sure generated signals resemble original signals in the time domain was the main focus of the generative algorithm.

Visual analysis of the signals presented in Figure 5.4 is enough to conclude the general ECG waveform is well modeled and that the algorithm is flexible enough to accommodate variations in the sense of different leads. Closer inspection shows the morphology of P, Q, R, S and T waves is also realistically displayed, as well as relative amplitudes, widths and locations. Since, in this case, and as is the case for the remaining full artificial recordings which will be further discussed, noise was ignored by the algorithm. Transitions between cycles are flattened from linear interpolation. Moreover, while subject slp59m's amplitude range is maintained, f2o05m's artificial recording presents what seems to be an attenuated signal. This occurs as a result of the signal generation method explained in Section 4.2.6, where, for the case, GMMs corresponding to different portions are summed together to form a segment. When the segment is divided

Figure 5.4: Two examples of ECG artificial recordings with different waveforms due to different measurement leads. Subject slp59m is part of the MIT-BIH Polysomnographic database, while f2o05m of Fantasia. The time axis was omitted because it bears no relevance for the purpose of this waveform comparison.

into portions with overlapping delineations, consecutive, symmetrical portions originate a signal which is partially canceled out. Typically, such does not happen often in P-waves, but in Q and T-waves, whose boundaries with R and S-waves are hard to discern.

Figure 5.5 shows generated waveforms which are more variable than the ones presented in the respective real recording. More precisely, only the first, second and third cycles could be considered to bear realistic resemblance to their real counterparts. However, this does not necessarily mean the algorithm is not working correctly since the real recording does, in fact, present BP waveforms which resemble all of the generated ones displayed above. Since full recordings were fed into the generative model and some particular recordings contain highly variable, valid waveforms resulting from changes in the measurement conditions, the algorithm considers them as different, possible versions of the same cycle-state. Notwithstanding, it lacks the ability to group these cycle variations together as they appear in reality given they have the same likelihood of being presented at any given set of RR-interval and ϕ_{RESP}. Therefore, it is possible for multiple waveforms to alternate in a small time span, including amplitude differences with no physiological reason. Furthermore, dicrotic notches are modeled smoother than expected. No evidence of smoothing exists in the ECG modality, which might point to a higher likelihood of

this type of occurrence in undivided segments with high-order GMMs, suggesting improvements can be made in the nonlinear fitting method in use. Notwithstanding, the generated BP pulses generally have realistic waveforms, where a systolic phase can be discerned from a diastolic and the amplitude range is respected.

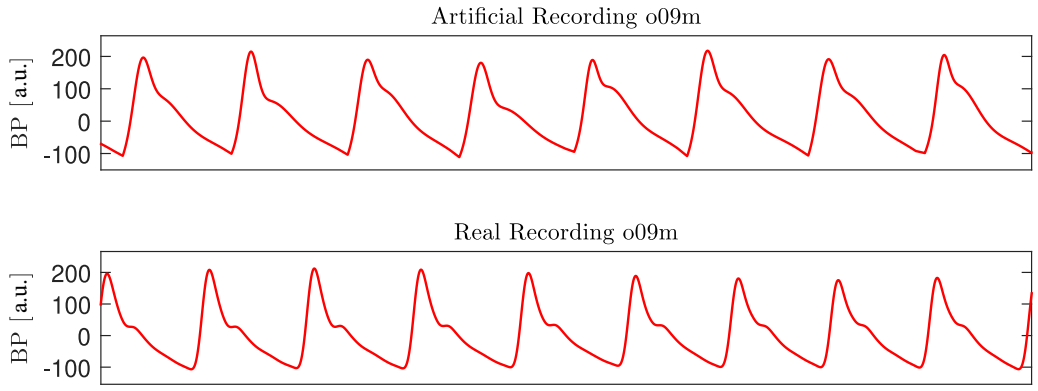

Figure 5.5: Comparison of BP artificial and real recordings corresponding to the subject f2o09m from the Fantasia database. The time axis was omitted because it bears no relevance for the purpose of this waveform comparison.

Unobtrusive modality modeling is shown in Figure 5.6. General morphology can be verified to be realistic and amplitudes ranges are correctly modeled. Concerning the BCG modality, it seems that 8^{th} FS is not enough to fully depict high-frequency components in the signal, causing the real recording to show much higher variation from the most noticeable low frequency waveform. The same does not occur in cECG, as segments modeled by SoS are shorter and 8^{th} is high enough to accommodate most of the variability. Additionally, QRS-complexes are also convincingly modeled. Calculation of the average waveform from generated recordings of previous modalities yielded almost identical results to their real counterparts, shown in Figure 5.1, including cECG, which also features flattened peaks corresponding to the P and T-waves, suggesting SoS functions are capable of integrating these portions into a random, noisy representation.

The Markov model's ability to alternate between normal and ectopic beats is showcased in Figure 5.7. In particular, existing PVC cycles in the referenced subject's recording were recreated and the estimated transition probabilities have produced realistic cycle-state transitions. Moreover, the mean stationary state was calculated and recordings were checked for the existence of absorbing states. No absorbing states were found, which is physiologically expected, and a mean stationary state of $[0.615 \pm 0.12, 0.385 \pm 0.11]$ for normal and ectopic beats, respectively, was reached after 7 ± 2 iterations. PVC waveforms are realistically modeled, too, and the recording respects the original amplitude range. Despite being modeled with only three portions - P, R and T-wave -, normal beats can also be convincingly generated and destructive overlap among portions is prevented.

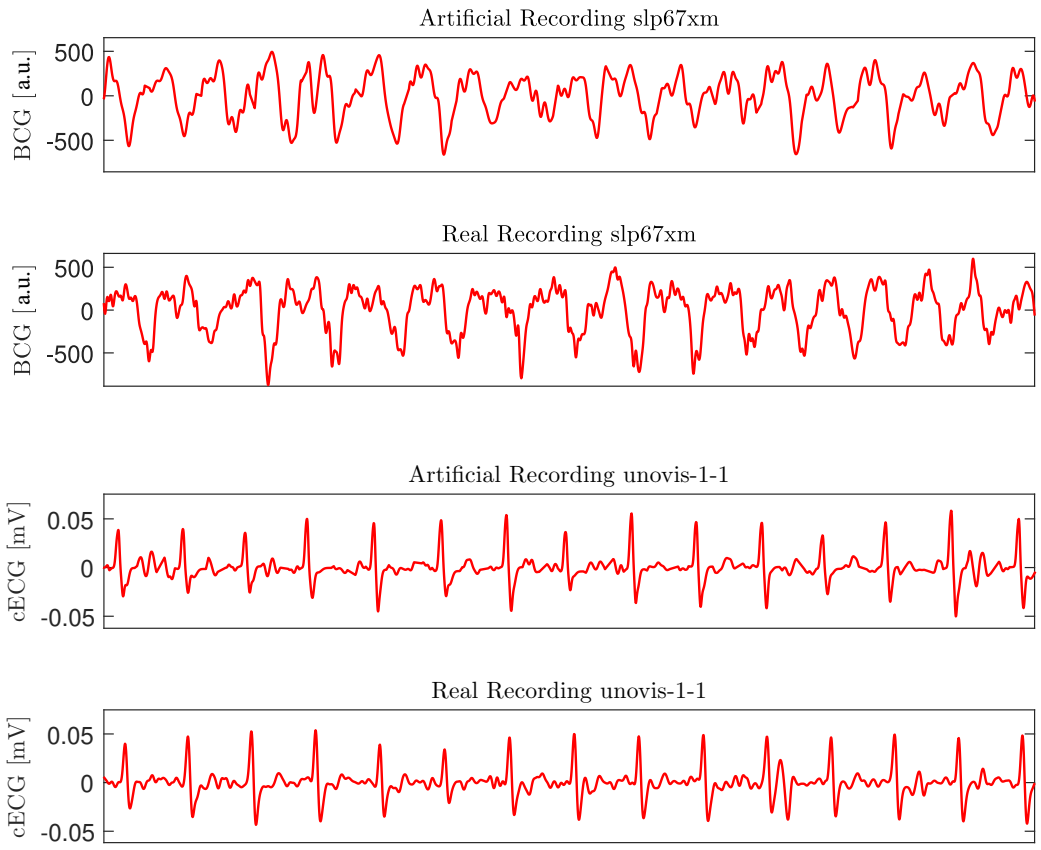

Figure 5.6: Comparison of artificial and real recordings in unobtrusive modalities for slp67xm subject of MIT-BIH Polysomnographic database and recording 1 of subject 1 of UnoVis. The time axis was omitted because it bears no relevance for the purpose of this waveform comparison.

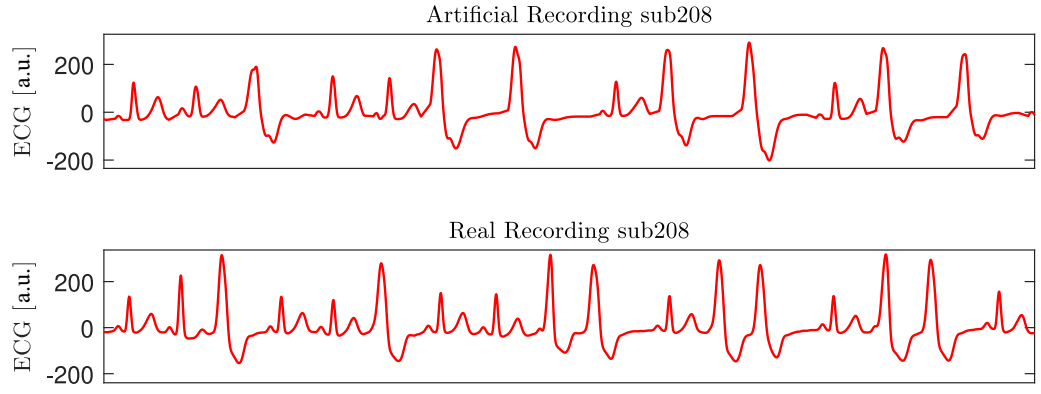

Figure 5.7: Comparison of ECG artificial and real recordings corresponding to the subject 208 from the MIT-BIH Arrhythmia database with a Markov model to insert PVC cycles. The time axis was omitted because it bears no relevance for the purpose of this waveform comparison.

In Figure 5.8, an artificial recording excerpt is plotted against the respective ϕ_{RESP} values to showcase the respiration modulation capability of the algorithm. This modulation is motivated

by high correlation values among not only ϕ_{RESP} and GMM amplitudes but also baseline values - multiplicative and additive modulation. These correlations are also verified in BCG and cECG modalities. Given how BP generation is significantly affected by heterogeneous input data as discussed above, linear correlations allusive to respiratory modulation are not consistent in this modality.

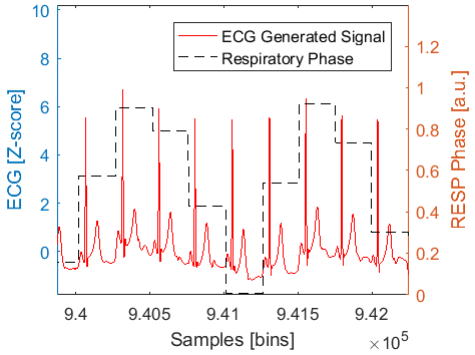

Figure 5.8: Respiratory modulation exemplified in ECG modality in generated recording respective to subject slp60m from the MIT-BIH Polysomnographic database. The signal is plotted in red while the ϕ_{RESP} is plotted with a black dashed line.

Finally, an example of an excerpt of a multimodal recording can be seen in Figure 5.9. The observations above regarding ECG, BP and BCG can also be verified in this representation. Artifact inclusion is not shown in this section since it was used to train the CNN, in which section example artificial recordings exist.

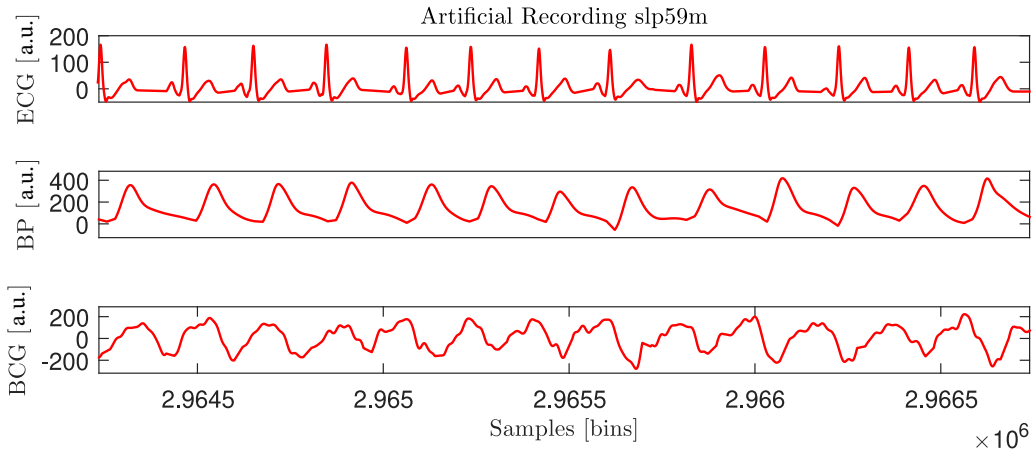

Figure 5.9: Artificial multimodal recording respective to subject slp59m from the MIT-BIH Polysomnographic database.

5.1.2.2 Frequency Domain

Fourier analysis was performed on the obtained artificial recordings, by comparative evaluation of their frequency spectra with those of their real equivalents. Plots were obtained by taking the Fourier transform of entire recordings, restricting them to the modality-dependent frequency ranges used for pre-processing in Table 4.1 and the logarithmic scale was used in the frequency

axis to better distinguish peaks and trends. Also, amplitudes were featured in decibels normalized with the highest value and a sliding average window was used to create an average amplitude signal in order to facilitate spectral shape comparison.

Clinical modality modeling yielded artificial signals which, also in the frequency domain, resembled real recordings by adapting to different spectral shapes and relative amplitudes among its peaks, as can be seen by Figure 5.10. In fact, real frequency peaks are accurately mimicked in terms of location, although artificial peaks tend to be better defined, with smaller widths. Artificial signals tend to have small higher frequency components which are not featured in the original recordings, with the exception of the BP modality, where the opposite occurs. This might be due to the fact that, in some cases, the order of the GMM used to model BP pulses is not enough to fully describe it also causing the time domain representation to display smoothed dicrotic notches, as previously discussed in Section 5.1.1.2 with Figure 5.5. As a matter of fact, increasing the order of the GMM function provides a better frequency reconstruction of higher frequency components (as seen in the previously mentioned figure), but never perfect since orders above 6 become incompatible with probabilistic generation, yielding incorrect waveforms.

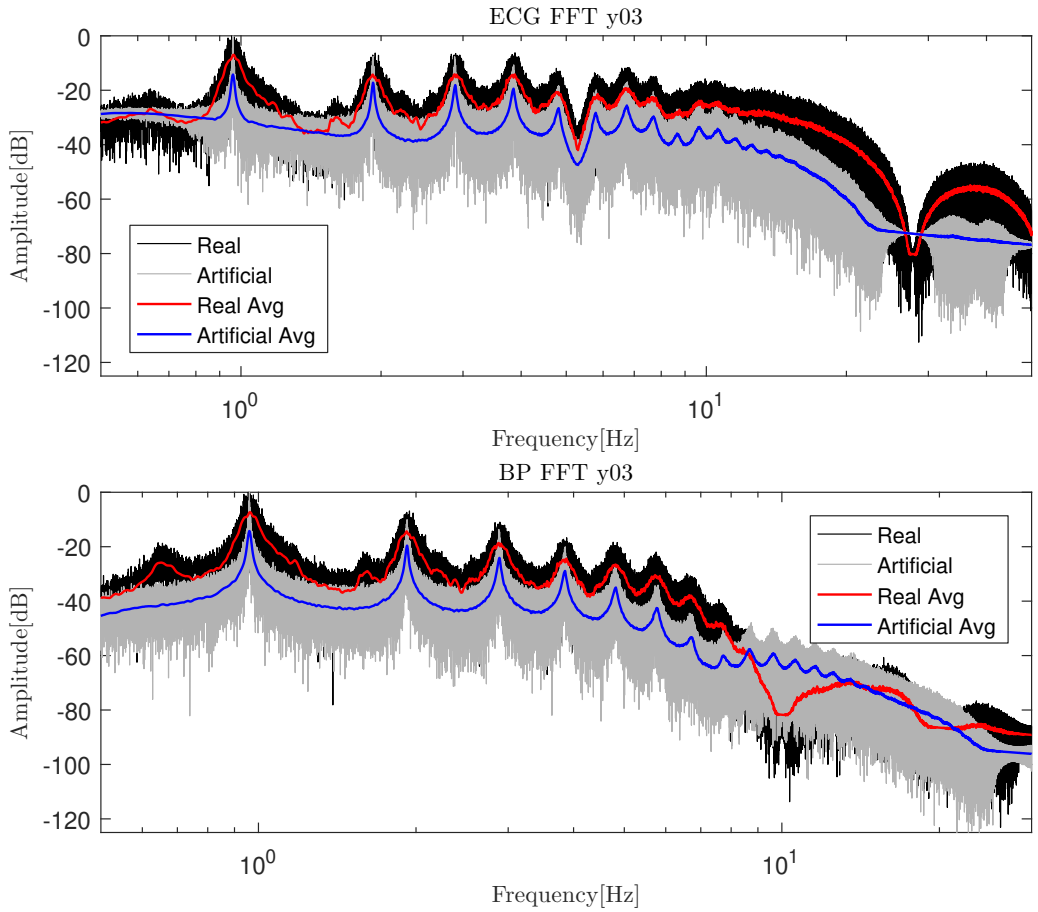

Figure 5.10: Fourier transform of the generated and real ECG and BP recordings relative to subject f2y03 from the Fantasia database.

Non-clinical modalities present frequency signals with less prominent peaks, likely due to the increased amplitude of white noise. Notwithstanding, some wide peaks can be discerned in some recordings, especially in BCG, where some low frequency peaks exist. As seen in Figure 5.11, while

artificial BCG recordings are able to recreate said low frequency peaks, although thinner, cECG artificial recordings display slightly discernible peaks on what should be a more homogeneous baseline, which can be compared to peaks observed in ECG in Figure 5.10. This supports what was discussed earlier in Section 5.1.1.2 with Figure 5.6, about BCG modeling not rigorously representing the incorporated noise and adds that cECG modeling has the same problem in the frequency domain despite it not being evident in the time domain.

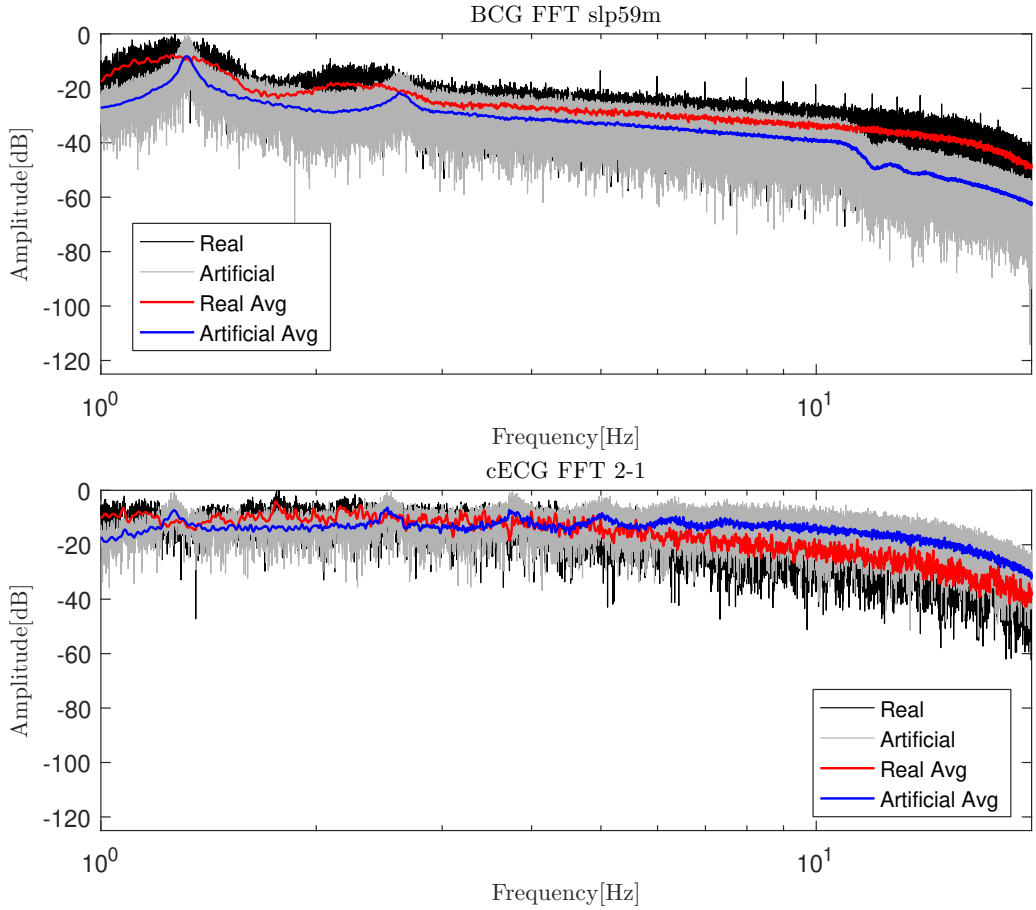

Figure 5.11: Fourier transform of the generated and real BCG recordings relative to subject slp59m from the MIT-BIH Polysomnographic database and cECG recordings relative to recording 1 of subject 2 from the UnoVis database.

Finally, ECG recordings with ectopic cycles present low amplitude peaks which map to better defined peaks, although with similar widths. Figure 5.12 serves as example.

5.1.2.3 Phase Domain

Phase analysis of artificial signals was performed so as to characterize their underlying nonlinear dynamics, given that the data-driven modeling approach was first idealized to mimic the complexity of original data as realistically as possible. The phase fractals respective to each modality of each recording were built and a quantitative comparison between real and artificial recordings based on CorrDim and ApEn was conducted. The first metric characterizes a fractal object by measuring how complex of a topological dimension is needed to contain it, while the second measures the probability of the phase trajectories which comprise the fractal object to follow the

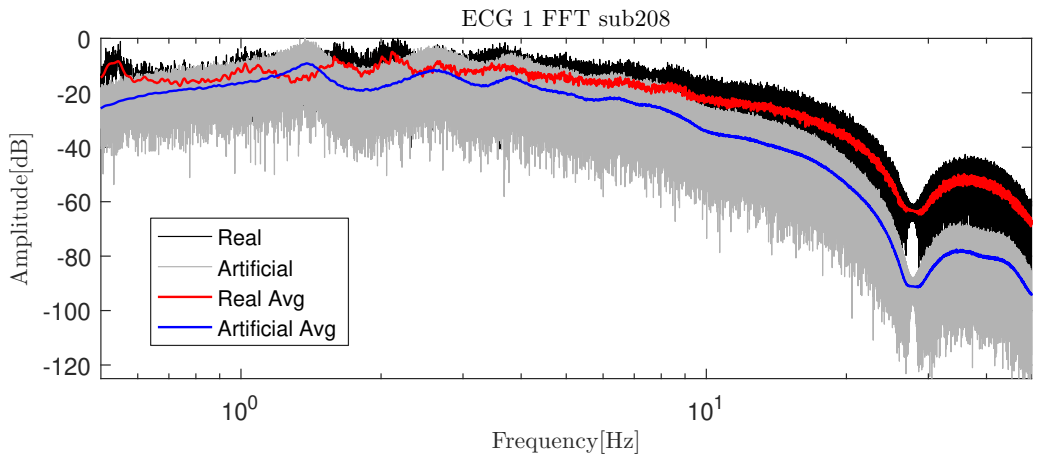

Figure 5.12: Fourier transform of the generated and real ECG recordings with PVC relative to subject sub208 from the MIT-BIH Arrhythmia database.

same trajectory with only slight variations. To build the fractal objects, the optimal time delays τ were found by choosing the first lag value which yielded null auto-correlation, minimizing linear correlation, and the optimal embedding dimensions was determined by the false nearest neighbor method, as described in [120].

Table 5.5 shows how the complexity of the generated signals convincingly resembles the non-linear dynamics from which real recordings stem, with a slightly higher error in unobtrusive modalities. Accordingly, phase reconstruction with a fixed embedding dimension of 3 yielded fractal objects with similar shapes, as shown in Figure 5.13 and reported in [77], given the low absolute error values observed, although a tendency for artificial trajectories to punctually stray away from the expected shape, evidencing a slight excess of entropy of the generated signals, as shown in Figure 5.14. Unobtrusive show the opposite tendency, as real objects follow show more trajectory variance, supporting what is verified from the table and was discussed in the previous two Sections.

Table 5.5: Comparative phase characterization of the artificial signals and their real counterparts in the light of CorrDim and ApEn.

		Absolute error ± standard deviation			
Database	Metrics	ECG	BP	BCG	cECG
Fantasia	CorrDim	0.34±0.12	0.34±0.21	-	-
	ApEn	0.14±0.08	0.19±0.13	-	-
Polysomnographic	CorrDim	0.30±0.24	0.37±0.04	0.97±0.05	-
	ApEn	0.14±0.04	0.30±0.14	0.87±0.16	-
Arrhythmia	CorrDim	0.19±0.08	-	-	-
	ApEn	0.31±0.05	-	-	-
UnoVis	CorrDim	0.35±0.19	-	-	0.86±0.75
	ApEn	0.14±0.12	-	-	0.84±0.57

Furthermore, the waveform generation method discussed above for causing destructive summation of consecutive portions modeled with GMM also hindered some recordings in the phase domain by altering the fractal shape range as seen in Figure 5.14. In the referred illustration,

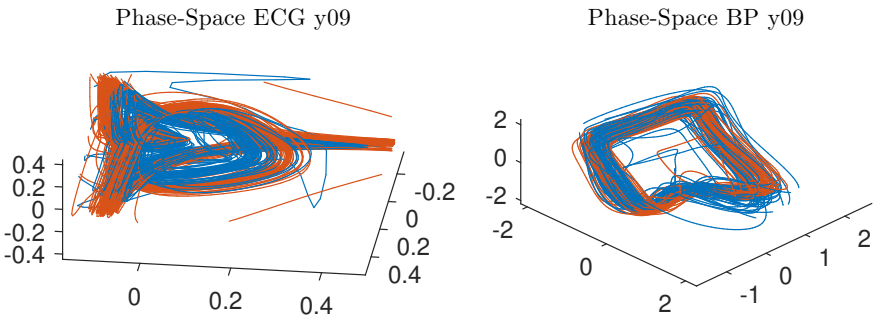

Figure 5.13: Phase reconstruction of real and artificial ECG and BP recordings respective to subject f2y09 from Fantasia. Artificial objects are plotted in blue, while real objects are plotted in orange.

lower, artificial R-peaks shrink the fractal shape in relation to real one. Similarly, some real fractal objects have been altered, especially in the BP modality due to change of measurement conditions in long recordings, also previously discussed in Section 5.1.1.2.

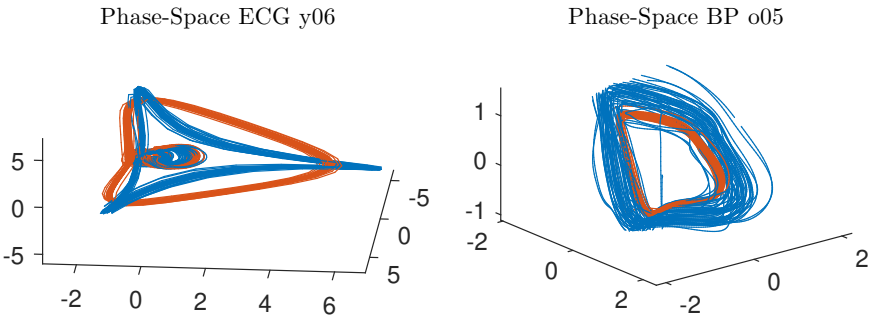

Figure 5.14: Phase reconstruction of real and artificial ECG and BP recordings respective to subjects f2o09 and f2y06 from Fantasia. Artificial objects are plotted in blue while real objects are plotted in orange.

5.2 Deep Learning Algorithm

In this section, results from hyperparameter optimization and the experiments mentioned in Section 4.3 are presented. These will also serve as the basis for further addressing the assumptions previously made about the effectiveness of the chosen architecture and its adequacy given the type of data it handles. Ultimately, after validation of the CNN, the contribution of the combined synthesizer and network to the enhancement of DL methods will be realized.

5.2.1 Hyperparameter Optimization

Grid search optimization was performed first in order to determine a suitable combination of weight initialization and optimization algorithm for the designed network. Both these categorical aspects were kept segregated from the main Bayesian hyperparameter optimization to avoid clogging its search-space on tasks which do not need a great deal of iterations for conclusions to be drawn and which can dramatically vary the time until convergence, unnecessarily slowing down the process. The remaining hyperparameters were kept as PyTorch's default. Obtained training and test loss graphs are shown in Figure 5.15. Prior debugging with a single batch as training set and no regularization had already guaranteed the basic functioning of the algorithm, with both

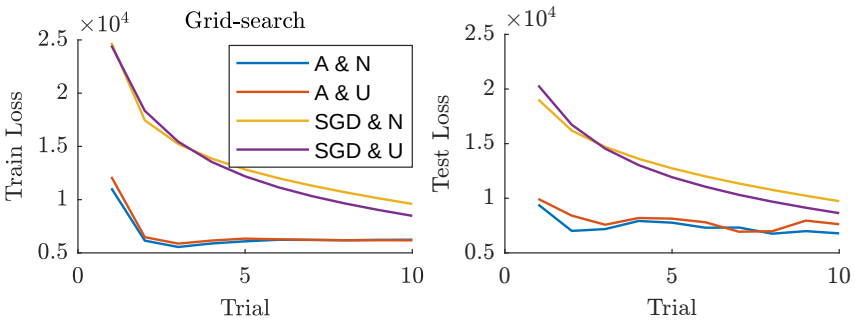

Figure 5.15: Grid search results for training and testing after 10 iterations. A - Adam optimizer; SGD - SGD optimizer; N - normal distribution; U - uniform distribution.

SGD and Adam converging and crossing the Goldilocks point, where test loss starts increasing while training loss keeps decreasing. As expected, the SGD optimizer provides a much slower, but steadier convergence than Adam. The addition of the adopted regularization methods, kept the algorithm from overfitting past that point but still allowed convergence. However, given there was no major instability with the latter, still managing to reduce the training and test loss, this optimization algorithm was preferred. Moreover, normally distributed weights (excluding linear combination coefficients) provided a slight, early improvement in decreasing the loss function and although this initialization type was chosen, opting for uniform initialization would not result in serious changes. The initial loss values are not shown in the graph, but, since fixed pseudo-randomization seeds were used, each initialization type provided a common starting loss value regardless of the used optimizer, leading to the conclusion that a much bigger leap is taken during the first iteration of Adam.

Bayesian optimization followed and the algorithm was ran for 100 iterations for a 6-dimensional search-space to allow the tree-structured Parzen estimator explained in Section 2.3.3 to find the best solution. Results are shown in Figure 5.16. Given the number of iterations, it was chosen to present cross-validation SEM in the current graph, to show the variance in performance for different subjects of a database. The SEM values were assumed to be negligible for further experiments with the network given that, especially when hyperparameter configurations yield lower test losses, small variations are observed. No clear convergence of the algorithm is shown, as it should be intended from Bayesian optimization. It is possible, given the high-dimensional search-space, that more iterations were necessary. Notwithstanding, while the need for optimization is not questionable, the utilization of regularization methods like batch normalization, L1 and L2, tend to expand the range of hyperparameter values which work well with the network, causing merely slight performance changes across different configurations. Furthermore, the inclusion of the maximum number of epochs in the search-space, while being a relevant, unknown hyperparameter, might have not played a part in the probabilistic forecast because of the early stopping approach, contributing with some noise to the choice of the most performant configuration. Therefore, it was concluded that, despite achieving small loss values, Bayesian optimization tended to act more randomly, probably performing with marginal improvements over algorithms such as random grid search. Nonetheless, the algorithm did show preference for some parameters, such as dropout probability and batch size, to remain within tight ranges - $[0.25, 0.35]$ and 64, respectively. The chosen configuration, enunciated in Section 4.3.3, was the one attempted at the 73^{rd} iteration, which yielded the lowest combination of processing time, test and validation error.

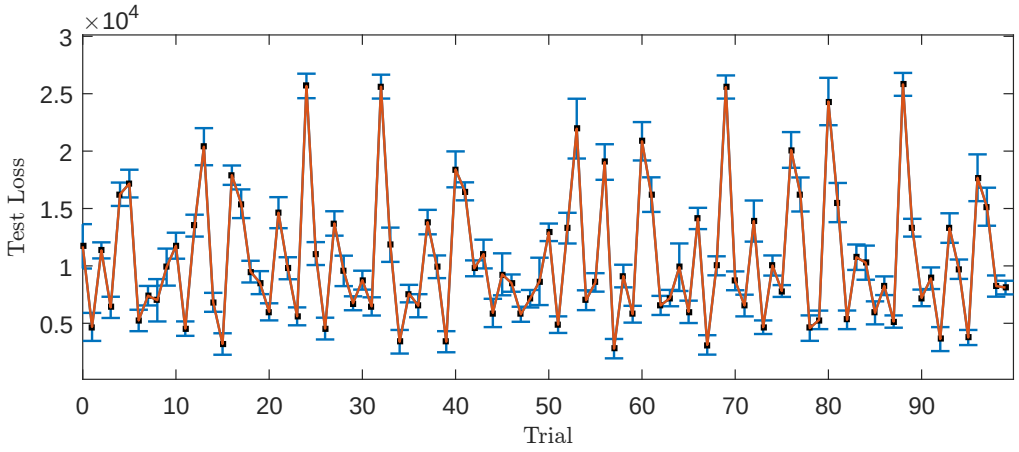

Figure 5.16: Bayesiam optimization test loss after 100 iterations with standard error of the mean (SEM) respective to different subjects in the leave-one-subject-out cross-correlation. X-axis points correspond to different hyperparameter combinations.

5.2.2 Artificial Data Experiments

The optimized network underwent additional tests to check for common problems affecting CNNs, such as exploding, vanishing gradients (although those are typically problems in deeper CNN) or dead ReLUs. Weights kept being updated with loss decrease and no activations were found to become null, while following normal distributions centered near 0, taking small non-null values as intended with regularization. After making sure proper functioning was in place, pre-processing was conducted followed by experimentation, as Sections 4.3.2 and 4.3.3 detailed.

5.2.2.1 Artificial/Real Proportion

Through the various iterations of this experiment, shown in Figure 5.17, it matters to point out the control score and error, obtained when no artificial data was added to the existing real recordings (null artificial/real ratio). The scores above 98% rival with the state of the art results previously referenced. In particular, the score of approximately 97% in the Fantasia database might suggest improvements over the network [98], upon which changes were introduced. Moreover, detection errors below 13 ms already consist of a significant improvement over the errors inside [37, 43] ms obtained by the same network. In general, no significant changes are observable as more artificial data is inserted in the training set, suggesting there should be no concerns regarding how results might worsen by doing so. Since the proposed CNN receives relatively long snippets of data as input, it could be argued that it would have more difficulty pinpointing a specific location given the larger scale, therefore originating higher detection errors than its precursor. However, the fact that it was trained more rigorously, in the sense that zero-label snippets used in training were centered closer to a non-zero label, motivated the network to learn that distinction. Furthermore, by analyzing the receptive field values of both networks' features in Table 5.6, the suggested, deeper structure achieves higher levels of abstraction as features in the last convolutional layer are influenced by more than half of the input snippet's samples - more than double the area influencing the same features in [98]. In fact, to fully seize the potential of the CentralNet adaptation, the suggested CNN was designed in order to stretch the range of representation levels presented at each layer, leaving it up to the linear combinations to promote or demote specific levels to achieve

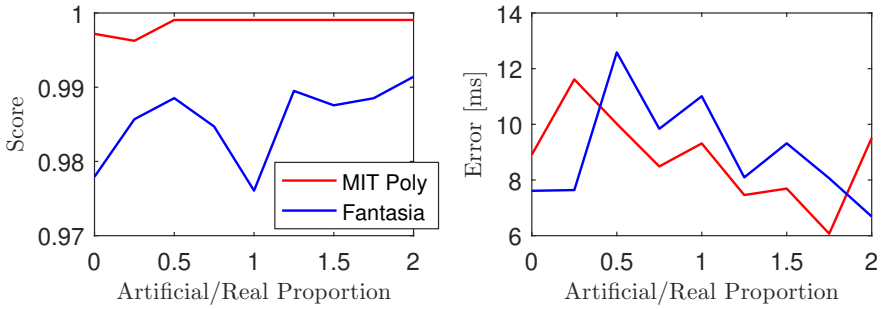

Figure 5.17: Score and error values resulting from training the CNN with an increasingly higher percentage of artificial signals in the training set. The first points relate to a training set solely comprised of real recordings with their original lengths (control), while the last correspond to training sets with exclusively artificial signals.

better results. Additional adaptations would entail filter size manipulation in order to not rely as much on pooling layers to create abstraction and expand the receptive field. This stretch is evidenced in Table 5.6 as the number of samples in the input contributing to the last convolutional layer in [98] is achieved in the sixth layer of the proposed network. Closer inspection showed that the receptive field of the first two layers of the network are small enough to extract information from smaller portions of the network, such as the P-wave in ECG, the systolic upstroke in BP or the positive slope in the beginning of the BCG pulse, while intermediate layers are influenced by combinations of these smaller consecutive portions and the last layers already receive information from most of the beat. The inclusion of a greater portion of the BP pulse in non-zero labels could have also helped the improvements in this sense, as hypothesized.

Table 5.6: Comparative analysis of receptive fields between the proposed CNN and its precursor in [98]. FtD - distance between features; FL - field length; FC - field center (projected onto the input snippet).

Precursor			Proposed		
FtD	FL	FC	FtD	FL	FC
1	1	0.5	1	1	0.5
1	20	10	1	20	10
2	21	10.5	2	21	10.5
2	49	24.5	2	49	24.5
2	67	33.5	4	51	25.5
-	-	-	4	87	43.5
-	-	-	8	91	45.5
-	-	-	8	163	81.5

Figure 5.18 shows an overlapping excerpt of an ECG recording and the respective detection probabilities from the CNN. Despite sporadic probability increases in the transition phase, the use of a prominence-adaptive peak detector is a simple, yet robust addition. One can further observe how the difference in probabilities between zero-label and non-zero-label snippets average around 0.6, which is a safe margin for uncorrupted signals. Top probabilities of 0.8 might be evidence of the amount of regularization integrated into the model to avoid overfit since a deep CNN might become too nonlinear for simple 1D detection tasks such as this. Fully artificial training sets were also used, having yielded only marginally lower scores of 98% for MIT-BIH Polysomnographic

and 97% for Fantasia recordings, with a decreased error of 9.5 and 10.2, respectively, also noticed as artificial/real ratio increases in Figure 5.17. An increase in detection probabilities in peaks was also verified, which can be explained by noisy cycle transitions simply being modeled with linear interpolation and the overall less complex nature of an artificial signal making the QRS-complex more evident due to its high energy. The decrease in error can be attributed to the higher consistency in annotating artificial data while the corresponding beats are generated as opposed to regular, manual annotation in real data.

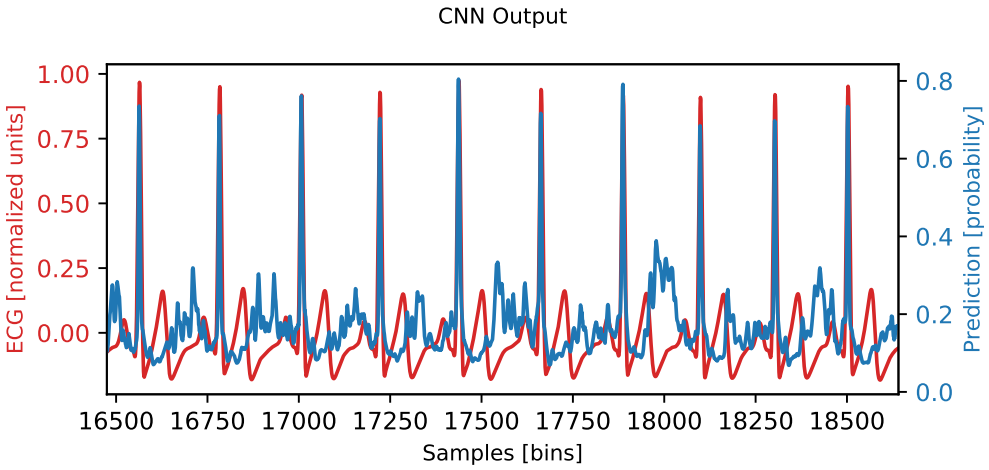

Figure 5.18: Typical CNN output (in blue) plotted against the real recording's ECG channel (in red) it is evaluating.

5.2.2.2 Insufficient Real Data

Results of the real recording shrinking phase are shown in Figure 5.19. Both training sets were reduced to a point where the CNN did not have enough data to be properly trained, with about 150 beats left available. Recordings had to be significantly shrunk to reach scores inside 50 and 60%, which emphasizes the simplicity of a 1D detection task for a DL algorithm but also shows how well the network resists overfitting and efficiently learns from data, even in a scarcity setting. The maintenance of low error values supports the former conclusion.

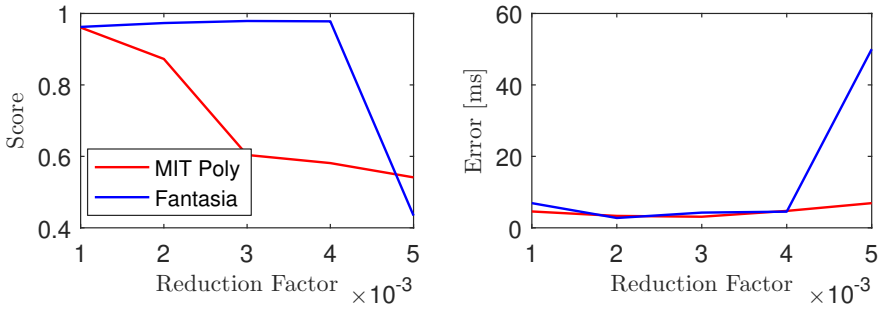

Figure 5.19: Score and error values resulting from training the CNN with an increasingly smaller real training set.

After reaching the target score, the shortened training sets were augmented with artificial data and results are shown in Figure 5.20. Addition of a single artificial recording with one fourth of the length of an original, full-sized one sufficed to bring the score back to the control level. Given the low variations after that point, it was assumed the behavior would tend to follow what occurred on the first experiment.

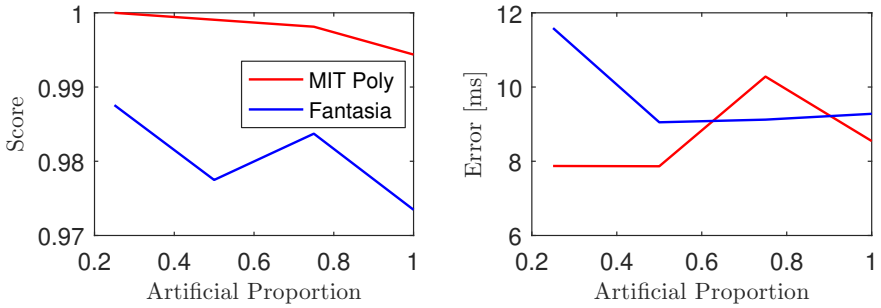

Figure 5.20: Score and error values resulting from training the CNN with the shrunk, real training set with increasingly higher percentage of artificial data.

5.2.2.3 Robustness to Artifacts

After contaminating the real recording in the test and validation set with modality-independent, generated artifacts, the results of using regular, real training set to train the CNN are shown in Figure 5.21. As predicted, activations during artifacts became too high and overshadowed detection of non-zero, uncorrupted samples, whose probabilities drastically decreased to the same level at which non-zero-labels were in the control experiment. This resulted in a decreased score of 93% but a still reasonable error of 6.4 ms. The score worsened due to an increase in FPs when artifacts occurred as in high SNR excerpts of the signal, the peak detector is still able to identify peaks based on prominence, because probabilities of zero-labels decreased proportionally as well.

Improvements can be seen in Figure 5.22, after artificial recordings with artifacts were included in the training set. The score returned to control level, above 99% but the error was increased to 11.1 ms. The number of FPs decreased mainly due to a reversal of the above described saturation process, as probabilities returned close to the 0.8 value in clean excerpts of data and the opposite occurred for corrupted portions. This suggests that data augmentation concerning noise can improve the network's perception of it and improve overall robustness of the model. Particular to signal-fusion, different noise scenarios can be shown to the CNN with some degree of success, as has also been demonstrated in [96], by omitting portions of the signal of a particular modality to train the algorithm to perform even in the absence of one or more channels.

Both in the first and second cases, no significant changes were observed in corruption of different modality combinations, likely because the majority of the artifacts saturate the signal, having similar patterns and yielding similar effects.

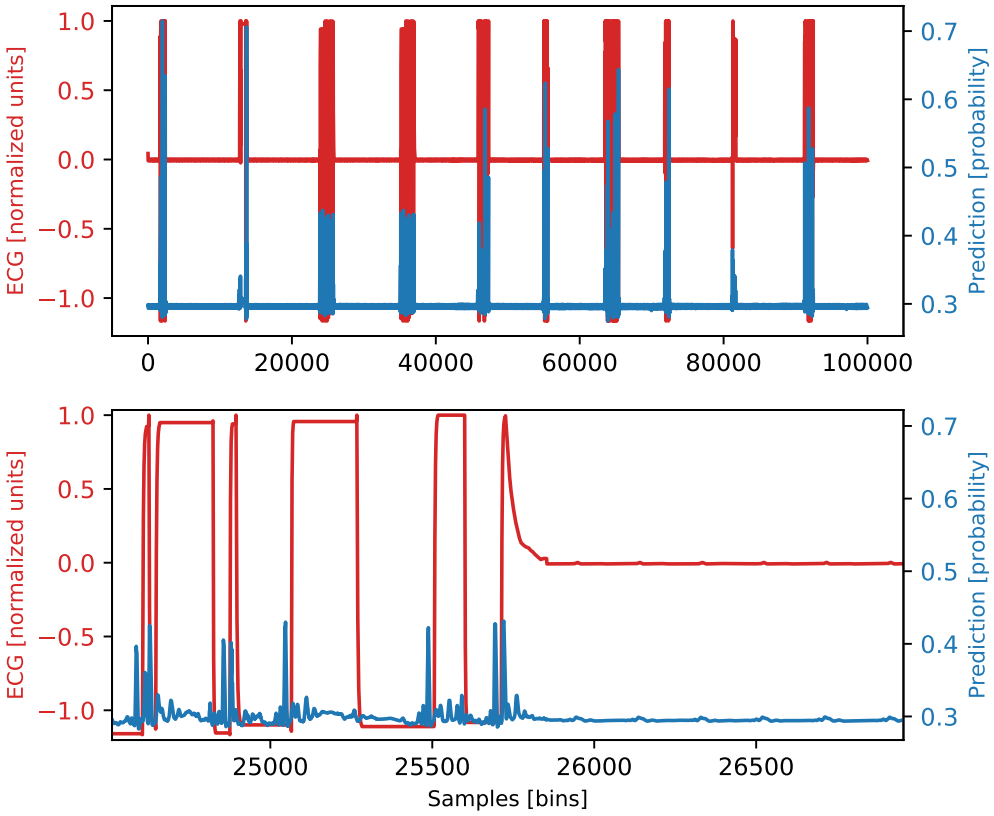

Figure 5.21: ECG signal and predictions resulting from training the network with real recordings with no artifacts.

5.2.2.4 Classification Task

In order to perform classification, three outputs to the network were set but only the two respective to non-zero-labels were used for peak detection post-processing. An excerpt of a 2-lead ECG recording and corresponding outputs of both those channels are represented in Figure 5.23. It can be seen that channel 2, responsible for normal beat detection, can also yield some prediction probability when a PVC occurs. However, due to the high specialization of channel 3 in detecting these, it almost solely outputs positive probabilities in such circumstances, causing its peaks to be much more prominent and detectable by the peak detection algorithm. Withal, in some cases, channel 3's output is not prominent enough and PVCs can be taken as normal beats, as shown in the confusion matrix in Table 5.7. A score of 92% associated with PVC detection and an error of 6.0 ms was achieved, as well as 99% and 9.0 ms for normal beats. Results evidence not only the capability of data augmentation in enabling a network to perform tasks requiring different types of data not available to the user but also the possibility of performing them satisfactorily.

5.2.3 Additional Findings

It is also interesting to observe how modalities are recruited at different layers by the fusion network. A schematic of the relative absolute linear coefficients of the CentralNet adaptation for

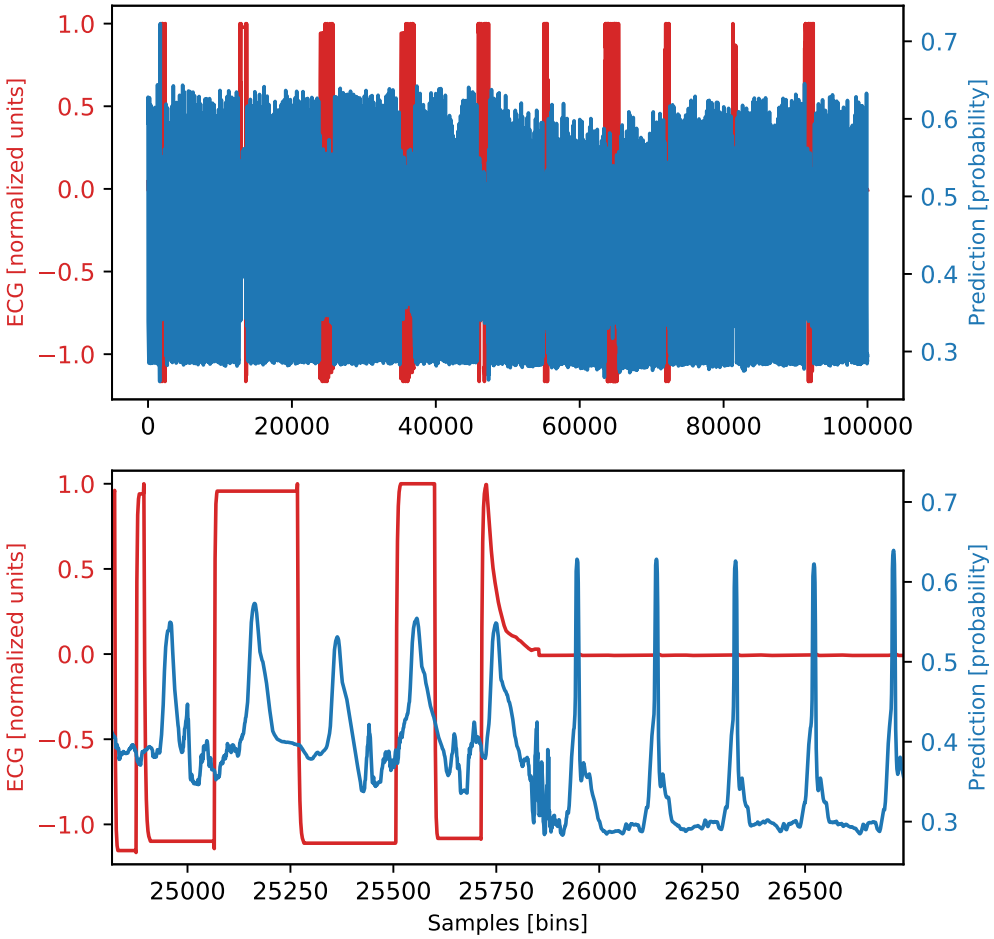

Figure 5.22: ECG signal and predictions resulting from training the network with real recordings with artificial recordings containing artifacts.

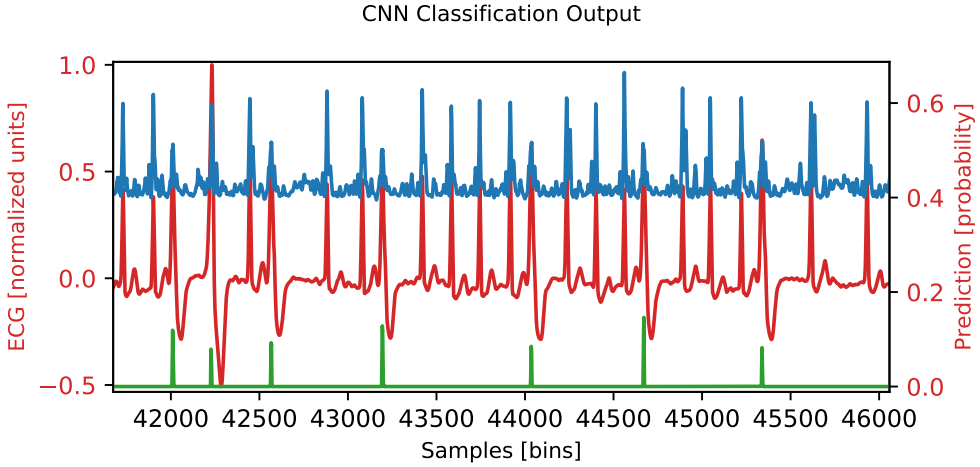

Figure 5.23: Typical CNN output for classification. In blue, channel 2's output is represented, while channel 3's predictions are highlighted in green and the original signal in red.

Table 5.7: Confusion matrix due to the classification experiment.

		Real	
		Normal (%)	PVC (%)
Predicted	Normal (%)	100	4.57
	PVC (%)	0	95.43

particular tests is shown in Figure 5.24, which can be interpreted as a proxy of modality importance and relevance of information at different stages of the CNN. For example, the ECG modality is fully predominant in one layer in the Fantasia and MIT-BIH Polysomnographic datasets, in accordance with what is expected given it is the most informative modality. Moreover, it can be verified that, in the first layer, the lowest level features of each branch are equally important across datasets and the prediction layer also tends to stem from similarly informative branch predictions. Despite the PTT, BP has a high importance in an early and a deep layer in the Fantasia dataset, which can imply the widened snippets are, in fact, enabling more useful information to be extracted from this modality.

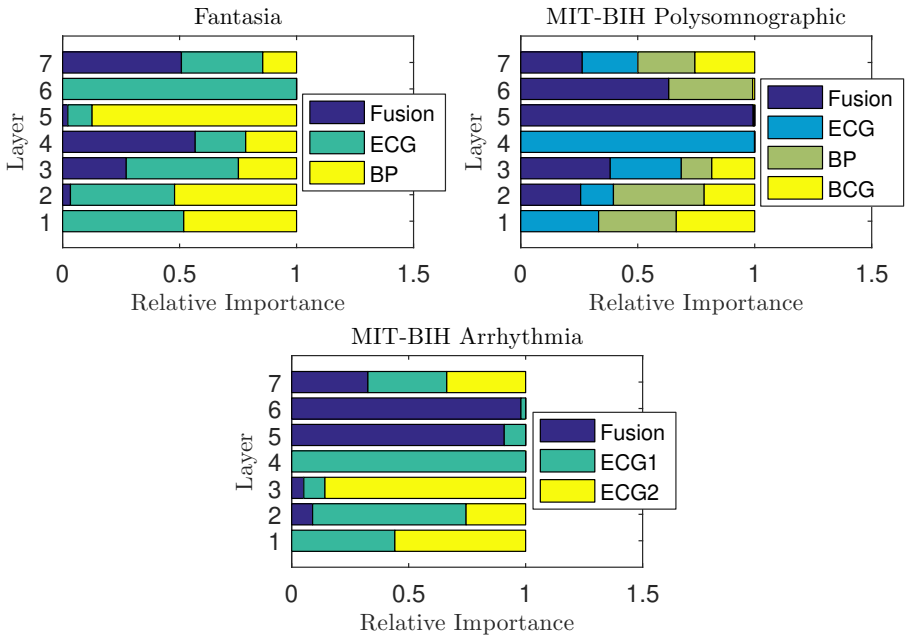

Figure 5.24: Relative absolute coefficients of linear combinations due to every layer for each processed dataset. Each plot concerns a single, trained network.

If we were to calculate the instantaneous HR HR_{inst} using the proposed algorithm as beat detector, according to an average absolute error σ_{CNN} of 8.65 ms and error propagation, we could proceed as follows to obtain the final error. Given two beat temporal location estimations $beat_1$ and $beat_2$ with respective uncertainties of $\sigma_{beat_1} = \sigma_{beat_2} = \sigma_{CNN}$, the RR-interval would be $RR = beat_2 - beat_1$ with an associated uncertainty of $\sqrt{2}\sigma_{CNN}$. Since $HR_{inst} = \frac{1}{RR}$ the associated

uncertainty would be

$$\sigma_{\mathrm{HR}_{\mathrm{inst}}}(\mathrm{RR}) = \frac{\sqrt{2}\sigma_{\mathrm{CNN}}}{\mathrm{RR}^2}.$$

For example, given a normal HR of 60 BPM, or 1 Hz, the uncertainty of measurement would equate to 1.22×10^{-2} Hz, which would mean vital sign estimation could also be accurate enough to integrate clinical monitoring devices.

One may conclude the proposed DL algorithm achieved the versatility it was designed to have given its successful adaptation to multiple types of datasets in the performance of different tasks. It was developed with PyTorch, and intended to be built with simple input from the user: how many modalities, convolutional layers to have, the size and depth of the filters and the structure of the FCN. To form the CNN's layers, default blocks of a convolutional operation, ReLU activation, batch normalization and 2-sample pooling layers with stride 2 are joined together, not needing any additional input. Remaining parameters such as L1 and L2 regularization constants are kept from the optimization step. Therefore, to adapt the architecture to a dataset with different modalities or to perform a different task, no changes need to be made to the code having rendered it to be highly generalizable.

Using the parallel computing platform CUDA® GPU training was conducted on an NVIDIA Quadro P4000 GPU @ 1202 MHz and 8 GB and Intel® Xeon® Gold CPU @ 3.40 GHz and 90 GB RAM. The average beat processing time did not vary much in the number of modalities, but mostly in the type of task. Processing of a snippet took on average 0.112 ± 0.024 ms for detection, while classification took 0.091 ± 0.028 ms, which will be referred to as reference duration Δt_r. The time it would take for a full beat to be processed would depend on the number of samples on that beat and, therefore, on the HR and sampling rate, while real-time applications would require it to be less than the duration of the beat itself $\Delta t_r \times \frac{250}{\mathrm{HR}} \leq \frac{1}{\mathrm{HR}}$. Thus, since the inequality resolves to $\Delta t_r \leq 4$ ms, which is true, the DL model has potential for real-time application having performed faster (in the given setup) than the real-time algorithm developed in [100], which took 17.10 ms per beat. The type of scenario and the hardware used to run it would dictate the feasibility of online applications.

Chapter 6

Conclusions

6.1 Closing Considerations

This thesis proposes a novel statistical synthesizer framework to model and sample artificial, multimodal cardiorespiratory signals. Its data-driven approach, allowed recording-specific, intra- and inter-modality physiological dynamics to be reproduced by means of copulas. Additionally, a CNN was proposed, which included adaptations from the latest research on sensor-fusion networks, to detect and classify beats in cardiorespiratory signals. Ultimately, joining both components, an array of experiments was conducted to validate the usage of data augmentation to overcome the shortage of data hindering the development of DL diagnosis algorithms.

Apart from the final results, some aspects deserve being stressed for the contribution each component represents to their respective research areas.

6.1.1 Multimodal Synthesizer

The synthesizer only requires a small set of simple parameters to capture a wide range of physiological effects, such as respiratory modulation and PTT variations. Besides, the framework uses lightweight statistical methods, allowing fast sampling of long signals. Overall, it can be hypothesized that the amount of effects that can be covered by adding a single parameter to the synthesizer is greater than when a full module is added to a dynamical model.

As the feature extractor is aligned with a data-driven mentality, it can be used to explore associations within data which could, otherwise, be very hard to conceive from a purely medical standpoint, e.g. by analyzing the correlation matrices.

In fact, this data representation is compatible with other types of works, for example, on biometric recognition [121]. It is also compatible with the most cited cardiorespiratory dynamical model [29], meaning it shares the ability to produce waveforms not featured in data by scaling and shifting parameter distributions. Furthermore, in case of sampling instability, parameter variation can be used to calculate a noise component intended to be directly summed to an average parameter vector, generating beats in a similar way to dynamical models.

As far as research on the state of the art was able to tell, this framework is the most complete to date, in the sense that quality normal, pathological and artifact waveforms can be modeled in one single recording, including realistic transitional dynamics between beat types. These signals are, therefore, suitable for DL algorithms tackling the time, frequency and phase domains of cardiorespiratory signals, as proven by the realistically obtained data in all three. Withal, while dynamical models seem to suffer from a few recurrent limitations [6], such as the stretching of artificial beats proportional to their length, the proposed synthesizer generates signals with less restrictions in general given their non-deterministic nature.

6.1.2 Deep Learning Algorithm

Thanks to the adopted CNN architecture, and opposite to what has been previously done, benefits from early-fusion, such as when input signals are highly correlated, and from late-fusion, like when the contrary occurs due to disparate modalities or even faulty signal acquisition, are both leveraged with only a marginal increase in parameters.

However, this does not overshadow the inherent generalizable nature of the architecture and the programming itself. The multi-branch structure and regularization techniques included in the model contribute to faster learning and immunity to overfitting, regardless of the modality combination in the input. Moreover, the regularization effect on the CNN's learning efficiency was evidenced by how heavily training recordings had to be shrunk to yield low-performance networks.

The network yields comparable results to those of the state of the art, in both detection and classification, and even surpasses its precursor's when common metrics are equated. Although the model was ran in a fast setup, its running times show promise in regards to real-time use. Simultaneously, low detection errors, respecting ANSI/AAMI EC38 and EC57 standards, are also encouraging for increased DL use in clinical settings.

6.1.3 Data Augmentation

It was found that, data augmentation for, at least, detection and classification of multimodal cardiorespiratory data has clear benefits. Moreover, no side effects were detected as neither the score nor error worsen significantly from varying the presence of artificial data in a sufficient training set. In scenarios where training data is insufficient, significant improvements, where performance was elevated to state of the art levels, were invariably observed. Such was true for when not enough training data samples were available, no noisy training data was provided in the face of test data of the same nature and if there was a complete absence of a type of data leading up to a classification task. In fact, it was noted a tendency of detection error to decrease with higher amounts of artificial data as annotation generation is more systematic than in manually annotated data.

These cases resonate with typical scenarios in research, emphasizing this framework as a step towards validating such an approach to overcome data scarcity and enhance a network's capability. The quality of the synthesized signals, naturally, bears significance to such results. Thus, the synthesizer could provide a centralized source for artificial cardiorespiratory signals in future research.

6.2 Limitations & Improvements

As per usual given the time constraints, one can only do so much in this type of thesis. In this particular case, the ambitious project of developing and studying two standalone components, which could be individually considered research projects on their own, could not have been concluded without prioritizing as some research directions were foregone for the end goals to be met in time. Inevitably, some limitations, therefore, stand out.

6.2.1 Multimodal Synthesizer

Because the synthesizer was designed to directly use lightly-processed recordings to generate its own signals, it falls victim to the variability of the original data as distorted waveforms produce residual probabilities learned by the copula. At the moment, the model is incapable of mimicking, for example, changes in experimental conditions as they appear in the original data. However, with a bigger care for pre-processing, these situations can be avoided.

Still in terms of signal abnormalities, although elegant, the waveform generation mechanism for GMMs causes some cycle segments with overlapping borders to destructively interfere with one another. Concatenating consecutive portions with interpolating overlap regions, finishing with low-pass filtering to remove high-frequency artifacts could, notwithstanding, pose a viable solution.

With regards to physiological effects, adding impactful parameters to the copula framework could not only provide a better control of the conditional sampling, but also contribute to model additional dynamics. An example would be the addition of a parameter from a delayed cycle, which correlates better to the PTT than in-phase parameters. Such a temporal component could cause a whole new set of out-of-phase physiological dependencies to be found. On the same train of thought, the Markov state transition matrix might also be coupled to a parameter, for example, HR.

Another question which needs to be addressed is the fact that ϕ_{RESP} does not always yield high correlations with other parameters, such as amplitudes, as expected. Different variations were tried, where instead of assuming the ϕ_{RESP} values to be synchronous with the R-peak, the mean or median over a window around it is calculated, but no consistent improvements were observed. Hence, there would be a need for defining a better proxy for RESP.

6.2.2 Deep Learning Algorithm

Since the network was partly built upon a pre-existing structure, there is no way to be sure if the amount of layers and filter sizes are optimal without further optimization at this level. Increased size differences among consecutive layers could provide more disparate abstraction levels from which the learnable linear combinations could benefit. Besides, targeted tests could shed a brighter light on the individual impact of each design choice, allowing better characterization of the model.

This thesis didn't cover the use of DL beyond estimation of vital parameters. Nonetheless, arrhythmia detection, or tasks involving classification of bigger excerpts of data are commonly researched. It is not unreasonable to assume the synthesized signals would also be useful in such scenarios, but the employment of an adapted CNN would further answer the question of how cardiorespiratory data augmentation affects DL applications in general.

6.3 Contributions

Bearing in mind the achieved milestones and interesting findings, it was decided that there is enough relevant and quality content for publishing. In this sense, some results have already been presented during this campaign and further publications are in progress:

- **Russian-German Conference on Biomedical Engineering 2019**: held in Saint-Petersburg, where preliminary results obtained from a first iteration of the synthesizer were featured in a conference paper and presented at the event, in July. Annex I contains the respective abstract submitted to the scientific committee;
- **AIP Conference Proceedings**: in the wake of the previously mentioned conference, the paper was published in August and can be found in [122];
- **Biomedical Engineering Summit 2019**: held in October in Lisbon where a poster will be displayed and a presentation will be delivered with the final results already included. Annex II encloses the respective abstract approved by the event's scientific committee;
- **Scientific Journals**: having used previous presentations and conferences to build a solid prelude to the extensive work developed during the thesis, at least one paper is planned to be submitted to high-profile journals in the field. At the moment, the writing of one specific to the multimodal synthesizer is in progress.

Bibliography

[1] Bernard Marr. *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*. 2018. URL: https://www.forbes.com/sites/bernardmarr/2018/05/21/{\%}0Ahow-much-data-do-we-create-every-day-the-mind-blowing-stats-everyoneshould-read/ (visited on 09/18/2019).

[2] Josh Gottlieb and Khaled Rifai. *Fueling growth through data monetization*. 2017. URL: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/fueling-growth-through-data-monetization (visited on 09/18/2019).

[3] S. A. Bini. "Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care?" In: *The Journal of Arthroplasty* 33.8 (2018), pp. 2358–2361. ISSN: 08835403. DOI: 10.1016/j.arth.2018.02.067. URL: https://linkinghub.elsevier.com/retrieve/pii/S0883540318302158.

[4] W. W. Stead. "Clinical implications and challenges of artificial intelligence and deep learning." In: *JAMA* 320.11 (2018), pp. 1107–1108. ISSN: 0098-7484. URL: http://dx.doi.org/10.1001/jama.2018.11029.

[5] C. D. Naylor. "On the Prospects for a (Deep) Learning Health Care System." In: *Journal of American Medical Association* 320.11 (2018), pp. 1099–1100. DOI: 10.1001/jama.

[6] C. Hoog Antink, S. Leonhardt, and M. Walter. "A synthesizer framework for multimodal cardiorespiratory signals." In: *Biomedical Physics & Engineering Express* 3.3 (2017), p. 035028. ISSN: 2057-1976. DOI: 10.1088/2057-1976/aa76ee. URL: http://stacks.iop.org/2057-1976/3/i=3/a=035028?key=crossref.f4f011d662f119a1c626ea8102202837.

[7] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic. "Deep learning applications and challenges in big data analytics." In: *Journal of Big Data* 2.1 (2015), p. 1. ISSN: 2196-1115. DOI: 10.1186/s40537-014-0007-7. URL: http://www.journalofbigdata.com/content/2/1/1.

[8] A. Triastcyn and B. Faltings. "Generating Artificial Data for Private Deep Learning." In: (2018). arXiv: 1803.03148. URL: http://arxiv.org/abs/1803.03148.

[9] S. J. Mcphee and W. F. Ganong. *Pathophysiology of Disease: An Introduction to Clinical Medicine*. 2007. ISBN: 0071806008.

[10] J. E. Hall. *Guyton and Hall: Textbook of Medical Physiology*. 13th. Saunders, 2015. ISBN: 9781455770052. URL: http://surgicalneurologyint.com/surgicalint-articles/guyton-and-hall-textbook-of-medical-physiology/.

[11] K. Barrett, H. Brooks, S. Boitano, and S. Barman. *Ganong's review of medical physiology*. 2010, p. 593. ISBN: 9780071605687.

[12] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford, D. R. Hose, and E. R. Ritenour. "Medical Physics and Biomedical Engineering." In: *Medical Physics* 28.5 (2001), pp. 861–861. ISSN: 00942405. DOI: 10.1118/1.1369117. URL: http://doi.wiley.com/10.1118/1.1369117.

[13] G. J. Crystal and P. M. Heerdt. "Cardiovascular Physiology." In: *Pharmacology and Physiology for Anesthesia*. Elsevier, 2013, pp. 366–389. DOI: 10.1016/B978-1-4377-1679-5.00021-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9781437716795000211.

[14] K. A. Y. J. Gordon Betts, P. Desaix, E. Johnson, J. E. Johnson, O. Korol, D. Kruse, B. Poe, J. A.Wise, M. Womble. *Anatomy and Physiology*. 2017, pp. 1033–1085. ISBN: 978-1-947172-043.

[15] C. Antink. "On Sensor Fusion for Multimodal Cardiorespiratory Signals." Doctoral dissertation. RWTH Aachen, 2018.

[16] E. M. McNeill, K. P. Roos, D. Moechars, and M. Clagett-Dame. "Nav2 is necessary for cranial nerve development and blood pressure regulation." In: *Neural Development* 5.1 (2010), p. 6. ISSN: 1749-8104. DOI: 10.1186/1749-8104-5-6. URL: http://neuraldevelopment.biomedcentral.com/articles/10.1186/1749-8104-5-6.

[17] J. Dunn, R. Runge, and M. Snyder. "Wearables and the medical revolution." In: *Personalized Medicine* 15.5 (2018), pp. 429–448. ISSN: 1741-0541. DOI: 10.2217/pme-2018-0044. URL: https://www.futuremedicine.com/doi/10.2217/pme-2018-0044.

[18] B. Chamadiya, K. Mankodiya, M. Wagner, and U. G. Hofmann. "Textile-based, contactless ECG monitoring for non-ICU clinical settings." In: *Journal of Ambient Intelligence and Humanized Computing* 4.6 (2013), pp. 791–800. ISSN: 1868-5137. DOI: 10.1007/s12652-012-0153-8. URL: http://link.springer.com/10.1007/s12652-012-0153-8.

[19] K.-f. Wu and Y.-t. Zhang. "Contactless and continuous monitoring of heart electric activities through clothes on a sleeping bed." In: *2008 International Conference on Technology and Applications in Biomedicine*. IEEE, 2008, pp. 282–285. ISBN: 978-1-4244-2254-8. DOI: 10.1109/ITAB.2008.4570586. URL: http://ieeexplore.ieee.org/document/4570586/.

[20] Y. G. Lim, K. H. Hong, K. K. Kim, J. H. Shin, S. M. Lee, G. S. Chung, H. J. Baek, D.-U. Jeong, and K. S. Park. "Monitoring physiological signals using nonintrusive sensors installed in daily life equipment." In: *Biomedical Engineering Letters* 1.1 (2011), pp. 11–20. ISSN: 2093-9868. DOI: 10.1007/s13534-011-0012-0. URL: http://link.springer.com/10.1007/s13534-011-0012-0.

[21] Hyun Jae Baek, Gih Sung Chung, Ko Keun Kim, and Kwang Suk Park. "A Smart Health Monitoring Chair for Nonintrusive Measurement of Biological Signals." In: *IEEE Transactions on Information Technology in Biomedicine* 16.1 (2012), pp. 150–158. ISSN: 1089-7771. DOI: 10.1109/TITB.2011.2175742. URL: http://ieeexplore.ieee.org/document/6078433/.

[22] Y. G. Lim, K. K. Kim, and K. S. Park. "ECG measurement on a chair without conductive contact." In: *IEEE Transactions on Biomedical Engineering* 53.5 (2006), pp. 956–959. ISSN: 00189294. DOI: 10.1109/TBME.2006.872823.

[23] Ko Keun Kim, Yong Kyu Lim, and Kwang Suk Park. "The electrically noncontacting ECG measurement on the toilet seat using the capacitively-coupled insulated electrodes." In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 3. IEEE, pp. 2375–2378. ISBN: 0-7803-8439-3. DOI: 10.1109/IEMBS.2004.1403688. URL: http://ieeexplore.ieee.org/document/1403688/.

[24] H. J. Baek, J. S. Kim, K. K. Kim, and K. S. Park. "System for unconstrained ECG measurement on a toilet seat using capacitive coupled electrodes : The efficacy and practicality." In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 2326–2328. ISBN: 978-1-4244-1814-5. DOI: 10.1109/IEMBS.2008.4649664. URL: http://ieeexplore.ieee.org/document/4649664/.

[25] T. Matsuda and M. Makikawa. "ECG monitoring of a car driver using capacitively-coupled electrodes." In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 1315–1318. ISBN: 978-1-4244-1814-5. DOI: 10.1109/IEMBS.2008.4649406. URL: http://ieeexplore.ieee.org/document/4649406/.

[26] S. Leonhardt and A. Aleksandrowicz. "Non-contact ECG monitoring for automotive application." In: *2008 5th International Summer School and Symposium on Medical Devices and Biosensors*. IEEE, 2008, pp. 183–185. ISBN: 978-1-4244-2252-4. DOI: 10.1109/ISSMDBS.2008.4575048. URL: http://ieeexplore.ieee.org/document/4575048/.

[27] M. Walter, B. Eilebrecht, T. Wartzek, and S. Leonhardt. "The smart car seat: personalized monitoring of vital signs in automotive applications." In: *Personal and Ubiquitous Computing* 15.7 (2011), pp. 707–715. ISSN: 1617-4909. DOI: 10.1007/s00779-010-0350-4. URL: http://link.springer.com/10.1007/s00779-010-0350-4.

[28] F. Netter. *Atlas of human anatomy*. 6th ed. Philadelphia, PA: Saunders/Elsevier, 2006. DOI: 10.1002/ca.20682.

[29] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith. "A dynamical model for generating synthetic electrocardiogram signals." In: *IEEE Transactions on Biomedical Engineering* 50.3 (2003), pp. 289–294. ISSN: 0018-9294. DOI: 10.1109/TBME.2003.808805. URL: http://ieeexplore.ieee.org/document/1186732/.

[30] M. Brink, C. H. Müller, and C. Schierz. "Contact-free measurement of heart rate, respiration rate, and body movements during sleep." In: *Behavior Research Methods* 38.3 (2006), pp. 511–521. ISSN: 1554-351X. DOI: 10.3758/BF03192806. URL: http://www.springerlink.com/index/10.3758/BF03192806.

[31] D. Mack, D. Mack, J. Patrie, P. Suratt, R. Felder, and M. Alwan. "Development and Preliminary Validation of Heart Rate and Breathing Rate Detection Using a Passive, Ballistocardiography-Based Sleep Monitoring System." In: *IEEE Transactions on Information Technology in Biomedicine* 13.1 (2009), pp. 111–120. ISSN: 1089-7771. DOI: 10.1109/TITB.2008.2007194. URL: http://ieeexplore.ieee.org/document/4663856/.

[32] F. Wang, M. Tanaka, and S. Chonan. "Development of a PVDF Piezopolymer Sensor for Unconstrained In-Sleep Cardiorespiratory Monitoring." In: *Journal of Intelligent Material Systems and Structures* 14.3 (2003), pp. 185–190. ISSN: 1045-389X. DOI: 10.1177/1045389X03014003006. URL: http://journals.sagepub.com/doi/10.1177/1045389X03014003006.

[33] K. Niizeki, I. Nishidate, K. Uchida, and M. Kuwahara. "Unconstrained cardiorespiratory and body movement monitoring system for home care." In: *Medical & Biological Engineering & Computing* 43.6 (2005), pp. 716–724. ISSN: 0140-0118. DOI: 10.1007/BF02430948. URL: http://link.springer.com/10.1007/BF02430948.

[34] J. Alihanka, K. Vaahtoranta, and I. Saarikivi. "A new method for long-term monitoring of the ballistocardiogram, heart rate, and respiration." In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 240.5 (1981), R384–R392. ISSN: 0363-6119. DOI: 10.1152/ajpregu.1981.240.5.R384. URL: http://www.physiology.org/doi/10.1152/ajpregu.1981.240.5.R384.

[35] K. Watanabe, Y. Kurihara, and H. Tanaka. "Ubiquitous Health Monitoring at Home – Sensing of Human Biosignals on Flooring, on Tatami Mat, in the Bathtub, and in the Lavatory." In: *IEEE Sensors Journal* 9.12 (2009), pp. 1847–1855. ISSN: 1530-437X. DOI: 10.1109/JSEN.2009.2030987. URL: http://ieeexplore.ieee.org/document/5291944/.

[36] J. Alametsä, A. Palomäki, and J. Viik. "Short and longer term repeatability of ballistocardiography in a sitting position with EMFi sensor." In: *Medical & Biological Engineering & Computing* 49.8 (2011), pp. 881–889. ISSN: 0140-0118. DOI: 10.1007/s11517-011-0746-y. URL: http://link.springer.com/10.1007/s11517-011-0746-y.

[37] E. Pinheiro, O. Postolache, and P. Girão. "Study on Ballistocardiogram Acquisition in a Moving Wheelchair with Embedded Sensors." In: *Metrology and Measurement Systems* 19.4 (2012), pp. 739–750. ISSN: 0860-8229. DOI: 10.2478/v10178-012-0065-0. URL: http://journals.pan.pl/dlibra/publication/104023/edition/90029/content.

[38] S. A. Franzblau. "Ballistocardiography in Cardiovascular Research: Physical Aspects of the Circulation in Health and Disease." In: *JAMA: The Journal of the American Medical Association* 203.2 (1968), p. 157. ISSN: 0098-7484. DOI: 10.1001/jama.1968.03140020085031. URL: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1968.03140020085031.

[39] I. Starr. "Prognostic Value of Ballistocardiograms." In: *JAMA* 187.7 (1964). ISSN: 0098-7484. DOI: 10.1001/jama.1964.03060200043008. URL: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1964.03060200043008.

[40] I. Starr. "Progress Towards a Physiological Cardiology." In: *Annals of Internal Medicine* 63.6 (1965), p. 1079. ISSN: 0003-4819. DOI: 10.7326/0003-4819-63-6-1079. URL: http://annals.org/article.aspx?doi=10.7326/0003-4819-63-6-1079.

[41] E. Pinheiro, O. Postolache, and P. Girão. "Theory and Developments in an Unobtrusive Cardiovascular System Representation: Ballistocardiography." In: *The Open Biomedical Engineering Journal* 4.1 (2010), pp. 201–216. ISSN: 1874-1207. DOI: 10.2174/1874120701004010201. URL: https://openbiomedicalengineeringjournal.com/VOLUME/4/PAGE/201/.

[42] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz. "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics." In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 271.4 (1996), R1078–R1084. ISSN: 0363-6119. DOI: 10.1152/ajpregu.1996.271.4.R1078. URL: https://www.physiology.org/doi/10.1152/ajpregu.1996.271.4.R1078.

[43] B. Gupta. "Monitoring in the ICU." In: *Update in Anaesthesia - ECG, pulse oximetry and capnograph* 28 (2012), pp. 32–37.

[44] A. B. Hertzman. "The Blood Supply of the Various Skin Areas as Estimated by the Photoelectric Plethysmograph." In: *American Journal of Physiology-Legacy Content* 124.2 (1938), pp. 328–340. ISSN: 0002-9513. DOI: 10.1152/ajplegacy.1938.124.2.328. URL: http://www.physiology.org/doi/10.1152/ajplegacy.1938.124.2.328.

[45] Hyun Jae Baek, Gih Sung Chung, Ko Keun Kim, Jung Soo Kim, and Kwang Suk Park. "Photoplethysmogram Measurement Without Direct Skin-to-Sensor Contact Using an Adaptive Light Source Intensity Control." In: *IEEE Transactions on Information Technology in Biomedicine* 13.6 (2009), pp. 1085–1088. ISSN: 1089-7771. DOI: 10.1109/TITB.2009.2031108. URL: http://ieeexplore.ieee.org/document/5256146/.

[46] M. Y. M. Wong, E Pickwell-MacPherson, and Y. T. Zhang. "Contactless and continuous monitoring of heart rate based on photoplethysmography on a mattress." In: *Physiological Measurement* 31.7 (2010), pp. 1065–1074. ISSN: 0967-3334. DOI: 10.1088/0967-3334/31/7/014. URL: http://stacks.iop.org/0967-3334/31/i=7/a=014?key=crossref.db145bebf34c4ac6c1f589729b5a767b.

[47] M.-Z. Poh, D. J. McDuff, and R. W. Picard. "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." In: *Optics Express* 18.10 (2010), p. 10762. ISSN: 1094-4087. DOI: 10.1364/OE.18.010762. URL: https://www.osapublishing.org/abstract.cfm?URI=oe-18-10-10762.

[48] I. Moxham. "Understanding Arterial Pressure Waveforms." In: *Southern African Journal of Anaesthesia and Analgesia* 9.1 (2003), pp. 40–42. ISSN: 2220-1181. DOI: 10.1080/22201173.2003.10872991.

[49] M. Safar. "Antihypertensive efficacy and destiffening strategy." In: *Medicographia* 32 (2010), pp. 234–240.

[50] R. Budhiraja, J. L. Goodwin, S. Parthasarathy, and S. F. Quan. "Comparison of Nasal Pressure Transducer and Thermistor for Detection of Respiratory Events During Polysomnography in Children." In: *Sleep* 28.9 (2005), pp. 1117–1121. ISSN: 1550-9109. DOI: 10.1093/sleep/28.9.1117. URL: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/28.9.1117.

[51] A. Amaddeo, M. Fernandez-Bolanos, J. Olmo Arroyo, S. Khirani, G. Baffet, and B. Fauroux. "Validation of a Suprasternal Pressure Sensor for Sleep Apnea Classification in Children." In: *Journal of Clinical Sleep Medicine* 12.12 (2016), pp. 1641–1647. ISSN: 1550-9389. DOI: 10.5664/jcsm.6350. URL: http://jcsm.aasm.org/ViewAbstract.aspx?pid=30894.

[52] K. Konno and J. Mead. "Measurement of the separate volume changes of rib cage and abdomen during breathing." In: *Journal of Applied Physiology* 22.3 (1967), pp. 407–422. ISSN: 8750-7587. DOI: 10.1152/jappl.1967.22.3.407. URL: http://www.physiology.org/doi/10.1152/jappl.1967.22.3.407.

[53] A. Shalbaf, H. Behnam, Z. Alizade-Sani, and M. Shojaifard. "Left ventricle wall motion quantification from echocardiographic images by non-rigid image registration." In: *International Journal of Computer Assisted Radiology and Surgery* 7.5 (2012), pp. 769–783. ISSN: 1861-6410. DOI: 10.1007/s11548-012-0786-2. URL: http://link.springer.com/10.1007/s11548-012-0786-2.

[54] I. Damjanov. *Pathophysiology*. 1st. Saunders, 2008. ISBN: 9780323314053.

[55] G. Courtadon. "The Pricing of Options on Default-Free Bonds." In: *The Journal of Financial and Quantitative Analysis* 17.1 (1982), p. 75. ISSN: 00221090. DOI: 10.2307/2330930.

[56] P. Embrechts, F. Lindskog, and A. Mcneil. "Modelling Dependence with Copulas and Applications to Risk Management." In: *Handbook of Heavy Tailed Distributions in Finance* (2003), pp. 329–384. DOI: 10.1016/b978-044450896-6.50010-8.

[57] S. Hazarika, S. Dutta, H. W. Shen, and J. P. Chen. "CoDDA: A Flexible Copula-based Distribution Driven Analysis Framework for Large-Scale Multivariate Data." In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 1214–1224. ISSN: 19410506. DOI: 10.1109/TVCG.2018.2864801.

[58] G. Salvadori, C. de Michele, N. T. Kottegoda, and R. Rosso. *Extremes in Nature: An Approach Using Copulas*. 56th ed. map C. Dordrecht: Springer, 2007. ISBN: 978-0-306-48065-2. DOI: 10.1007/0-306-48065-4.

[59] J. D. Gibbons. *Nonparametric Measures of Association*. Sage Publications, 1993, p. 104.

[60] A. Stuart. "Rank Correlation Methods. By M. G. Kendall, 2nd edition." In: *British Journal of Statistical Psychology* 9.1 (1956), pp. 68–68. ISSN: 0950561X. DOI: 10.1111/j.2044-8317.1956.tb00172.x. URL: http://doi.wiley.com/10.1111/j.2044-8317.1956.tb00172.x.

[61] P. Y. Chen and P. N. Popovich. *Correlation: Parametric and Nonparametric Measures*. 137th ed. Sage Publications, 2002.

[62] D. MacKenzie and T. Spears. "'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking." In: *Social Studies of Science* 44.3 (2014), pp. 393–417. ISSN: 0306-3127. DOI: 10.1177/0306312713517157. URL: http://journals.sagepub.com/doi/10.1177/0306312713517157.

[63] L. Makkonen. "Problems in the extreme value analysis." In: *Structural Safety* 30.5 (2008), pp. 405–419. ISSN: 01674730. DOI: 10.1016/j.strusafe.2006.12.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167473007000045.

[64] C. Donnelly and P. Embrechts. "The Devil is in the Tails: Actuarial Mathematics and the Subprime Mortgage Crisis." In: *ASTIN Bulletin* 40.1 (2010), pp. 1–33. ISSN: 0515-0361. DOI: 10.2143/AST.40.1.2049222. URL: https://www.cambridge.org/core/product/identifier/S0515036100000350/type/journal{_}article.

[65] P. A. Gagniuc. *Markov Chains*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017. ISBN: 9781119387596. DOI: 10.1002/9781119387596. URL: http://doi.wiley.com/10.1002/9781119387596.

[66] K. W. Johnson, J. Torres Soto, B. S. Glicksberg, K. Shameer, R. Miotto, M. Ali, E. Ashley, and J. T. Dudley. "Artificial Intelligence in Cardiology." In: *Journal of the American College of Cardiology* 71.23 (2018), pp. 2668–2679. ISSN: 07351097. DOI: 10.1016/j.jacc.2018.03.521. URL: https://linkinghub.elsevier.com/retrieve/pii/S0735109718344085.

[67] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." In: *Nature* 521.7553 (2015), pp. 436–444. ISSN: 0028-0836. DOI: 10.1038/nature14539. URL: http://www.nature.com/articles/nature14539.

[68] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: (2014). arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980.

[69] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going Deeper with Convolutions." In: (2014). arXiv: 1409.4842. URL: http://arxiv.org/abs/1409.4842.

[70] I. S. Mohamed, A. Capitanelli, F. Mastrogiovanni, S. Rovetta, and R. Zaccaria. "Detection, localisation and tracking of pallets using machine learning techniques and 2D range data." In: (2018). arXiv: 1803.11254. URL: http://arxiv.org/abs/1803.11254.

[71] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." In: *Proceedings of the IEEE* 86 (1998), pp. 2278–2324.

[72] Kaggle. *Digit Recognizer Competition Leaderboard*. URL: https://www.kaggle.com/c/digit-recognizer/leaderboard (visited on 09/18/2019).

[73] Bergstra, J. S., R. Bardenet, Y. Bengio, and B. Kégl. "Algorithms for hyper-parameter optimization." In: *Advances in neural information processing systems* (2011), pp. 2546–2554.

[74] J. Bergstra, D. Yamins, and D. D. Cox. "Making a Science of Model Search." In: (2012). arXiv: 1209.5111. URL: http://arxiv.org/abs/1209.5111.

[75] A. Khaustov, S. Nemati, and G. Clifford. "An open-source standard T-Wave alternans detector for benchmarking." In: *2008 Computers in Cardiology*. IEEE, 2008, pp. 509–512. ISBN: 978-1-4244-3706-1. DOI: 10.1109/CIC.2008.4749090. URL: http://ieeexplore.ieee.org/document/4749090/.

[76] G. D. Clifford and P. E. Mcsharry. "Generating 24-Hour ECG , BP and Respiratory Signals with Realistic Linear and Nonlinear Clinical Characteristics Using a Nonlinear Model." In: *Computers in Cardiology* 31 (2004), pp. 709–712.

[77] G. D. Clifford and P. E. McSharry. "A realistic coupled nonlinear artificial ECG, BP, and respiratory signal generator for assessing noise performance of biomedical signal processing algorithms." In: *Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II* 5467 (2004), p. 290. DOI: 10.1117/12.544525.

[78] P. Bernaola-Galván, P. C. Ivanov, L. A. Nunes Amaral, and H. E. Stanley. "Scale Invariance in the Nonstationarity of Human Heart Rate." In: *Physical Review Letters* 87.16 (2001), p. 168105. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.87.168105. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.168105.

[79] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith. "Method for generating an artificial RR tachogram of a typical healthy human over 24-hours." In: *Computers in Cardiology*. IEEE, pp. 225–228. ISBN: 0-7803-7735-4. DOI: 10.1109/CIC.2002.1166748. URL: http://ieeexplore.ieee.org/document/1166748/.

[80] M. J. Drinnan, J. Allen, and A. Murray. "Relation between heart rate and pulse transit time during paced respiration." In: *Physiological Measurement* 22.3 (2001), pp. 425–432. ISSN: 0967-3334. DOI: 10.1088/0967-3334/22/3/301. URL: http://stacks.iop.org/0967-3334/22/i=3/a=301?key=crossref.b9ef46e59c7fa12dbef9913c8de96774.

[81] D. Kaplan, M. Furman, and S. Pincus. "Techniques for analyzing complexity in heart rate and beat-to-beat blood pressure signals." In: *Proceedings Computers in Cardiology*. IEEE Comput. Soc. Press, pp. 243–246. ISBN: 0-8186-2225-3. DOI: 10.1109/CIC.1990.144206. URL: http://ieeexplore.ieee.org/document/144206/.

[82] G. D. Clifford, S. Nemati, and R. Sameni. "An Artificial Vector Model for Generating Abnormal Electrocardiographic Rhythms." In: *Physiological Measurement* 31.5 (2010), pp. 595–609. DOI: 10.1088/0967-3334/31/5/001.An.

[83] G. E. Dower, H. B. Machado, and J. A. Osborne. "On Deriving the Electrocardiogram from Vectorcardiographic Leads." In: *Clinical Cardiology* 3.2 (1980), pp. 87–95. ISSN: 01609289. DOI: 10.1002/clc.1980.3.2.87. URL: http://doi.wiley.com/10.1002/clc.1980.3.2.87.

[84] O. Sayadi and M. B. Shamsollahi. "A combined dynamical sequential network for generating coupled cardiovascular signals with different beat types." In: *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010)*. 1. IEEE, 2010, pp. 1–5. ISBN: 978-1-4244-8131-6. DOI: 10.1109/ISABEL.2010.5702821. URL: http://ieeexplore.ieee.org/document/5702821/.

[85] A. Bauer, M. Malik, P. Barthel, R. Schneider, M. A. Watanabe, A. J. Camm, A. Schömig, and G. Schmidt. "Turbulence dynamics: An independent predictor of late mortality after acute myocardial infarction." In: *International Journal of Cardiology* 107.1 (2006), pp. 42–47. ISSN: 01675273. DOI: 10.1016/j.ijcard.2005.02.037. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167527305005061.

[86] Yu Yao, C. Bruser, U. Pietrzyk, S. Leonhardt, S. van Waasen, and M. Schiek. "Model-Based Verification of a Non-Linear Separation Scheme for Ballistocardiography." In: *IEEE Journal of Biomedical and Health Informatics* 18.1 (2014), pp. 174–182. ISSN: 2168-2194. DOI: 10.1109/JBHI.2013.2261820. URL: http://ieeexplore.ieee.org/document/6514106/.

[87] P. Chaiyakum and P. Wardkein. "A new ECG compression by using adaptive cosine Fourier Transform." In: *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*. IEEE, 2012, pp. 1–4. ISBN: 978-1-4673-2025-2. DOI: 10.1109/ECTICon.2012.6254260. URL: http://ieeexplore.ieee.org/document/6254260/.

[88] D. Sadhukhan and M. Mitra. "ECG noise reduction using Fourier coefficient suppression." In: *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*. IEEE, 2014, pp. 142–146. ISBN: 978-1-4799-2044-0. DOI: 10.1109/CIEC.2014.6959066. URL: http://ieeexplore.ieee.org/document/6959066/.

[89] A. Bohm, C. H. Antink, S. Leonhardt, and D. Teichmann. "Determining the connection between capacitively coupled electrocardiography data and the ground truth." In: *2015 Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 677–680. ISBN: 978-1-5090-0685-4. DOI: 10.1109/CIC.2015.7411001. URL: http://ieeexplore.ieee.org/document/7411001/.

[90] Y. Yao, J. Schiefer, S. van Waasen, and M. Schiek. "A non-parametric model for Ballistocardiography." In: *2014 IEEE Workshop on Statistical Signal Processing (SSP)*. IEEE, 2014, pp. 69–72. ISBN: 978-1-4799-4975-5. DOI: 10.1109/SSP.2014.6884577. URL: http://ieeexplore.ieee.org/document/6884577/.

[91] R. Couceiro, P. Carvalho, R. P. Paiva, J. Henriques, M. Antunes, I. Quintal, and J. Muehlsteff. "Multi-Gaussian fitting for the assessment of left ventricular ejection time from the Photoplethysmogram." In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 3951–3954. ISBN: 978-1-4577-1787-1. DOI: 10.1109/EMBC.2012.6346831. URL: http://ieeexplore.ieee.org/document/6346831/.

[92] D. Martin-Martinez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, and C. Alberola-Lopez. "Stochastic Modeling of the PPG Signal: A Synthesis-by-Analysis Approach With Applications." In: *IEEE Transactions on Biomedical Engineering* 60.9 (2013), pp. 2432–2441. ISSN: 0018-9294. DOI: 10.1109/TBME.2013.2257770. URL: http://ieeexplore.ieee.org/document/6497551/.

[93] L. Cheng, O. Ivanova, H.-H. Fan, and M. C. Khoo. "An integrative model of respiratory and cardiovascular control in sleep-disordered breathing." In: *Respiratory Physiology & Neurobiology* 174.1-2 (2010), pp. 4–28. ISSN: 15699048. DOI: 10.1016/j.resp.2010.06.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1569904810002193.

[94] F. Berief, S. Leonhardt, and C. H. Antink. "Modelling and Synthesizing Motion Artifacts in Unobtrusive Multimodal Sensing using Copulas." In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Vol. Accepted. IEEE, 2018, pp. 6006–6009. ISBN: 978-1-5386-3646-6. DOI: 10.1109/EMBC.2018.8513690. URL: https://ieeexplore.ieee.org/document/8513690/.

[95] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya. "Deep learning for healthcare applications based on physiological signals: A review." In: *Computer Methods and Programs in Biomedicine* 161 (2018), pp. 1–13. ISSN: 01692607. DOI: 10.1016/j.cmpb.2018.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169260718301226.

[96] B. S. Chandra, C. S. Sastry, and S. Jana. "Robust Heartbeat Detection From Multimodal Data via CNN-Based Generalizable Information Fusion." In: *IEEE Transactions on Biomedical Engineering* 66.3 (2019), pp. 710–717. ISSN: 0018-9294. DOI: 10.1109/TBME.2018.2854899. URL: https://ieeexplore.ieee.org/document/8410035/.

[97] ECAR and AAMI. *Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms*. 1987.

[98] C. Hoog Antink, E. Breuer, D. U. Uguz, and S. Leonhardt. "Signal-Level Fusion With Convolutional Neural Networks for Capacitively Coupled ECG in the Car." In: *Computing in Cardiology* 45 (2018).

[99] W. Liu, Q. Huang, S. Chang, H. Wang, and J. He. "Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram." In: *Biomedical Signal Processing and Control* 45 (2018), pp. 22–32. ISSN: 17468094. DOI: 10.1016/j.bspc.2018.05.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S1746809418301150.

[100] W. Liu, M. Zhang, Y. Zhang, Y. Liao, Q. Huang, S. Chang, H. Wang, and J. He. "Real-Time Multilead Convolutional Neural Network for Myocardial Infarction Detection." In: *IEEE Journal of Biomedical and Health Informatics* 22.5 (2018), pp. 1434–1444. ISSN: 2168-2194. DOI: 10.1109/JBHI.2017.2771768. URL: https://ieeexplore.ieee.org/document/8103330/.

[101] P. Kasnesis, C. Z. Patrikakis, and I. S. Venieris. "PerceptionNet: A Deep Convolutional Neural Network for Late Sensor Fusion." In: (2018). arXiv: 1811.00170. URL: http://arxiv.org/abs/1811.00170.

[102] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. "Learning Deep Features for Discriminative Localization." In: (2015). arXiv: 1512.04150. URL: http://arxiv.org/abs/1512.04150.

[103] M. D. Zeiler. "ADADELTA: An Adaptive Learning Rate Method." In: (2012). arXiv: 1212.5701. URL: http://arxiv.org/abs/1212.5701.

[104] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie. "CentralNet: a Multilayer Approach for Multimodal Fusion." In: (2018). arXiv: 1808.07275. URL: http://arxiv.org/abs/1808.07275.

[105] X. Yang, P. Molchanov, and J. Kautz. "Multilayer and Multimodal Fusion of Deep Neural Networks for Video Classification." In: *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*. New York, New York, USA: ACM Press, 2016, pp. 978–987. ISBN: 9781450336031. DOI: 10.1145/2964284.2964297. URL: http://dl.acm.org/citation.cfm?doid=2964284.2964297.

[106] M. Kang, K. F. Li, X. Leng, and Z. Lin. "Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection." In: *Remote Sensing* 9.8 (2017), p. 860. ISSN: 2072-4292. DOI: 10.3390/rs9080860. URL: http://www.mdpi.com/2072-4292/9/8/860.

[107] C. Cangea, P. Veličković, and P. Liò. "XFlow: Cross-modal Deep Neural Networks for Audiovisual Classification." In: (2017). arXiv: 1709.00572. URL: http://arxiv.org/abs/1709.00572.

[108] Z. Gu, B. Lang, T. Yue, and L. Huang. "Learning Joint Multimodal Representation Based on Multi-fusion Deep Neural Networks." In: 2017, pp. 276–285. DOI: 10.1007/978-3-319-70096-0_29. URL: http://link.springer.com/10.1007/978-3-319-70096-0{_}29.

[109] T. Baltrusaitis, C. Ahuja, and L.-P. Morency. "Multimodal Machine Learning: A Survey and Taxonomy." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (2019), pp. 423–443. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2018.2798607. URL: https://ieeexplore.ieee.org/document/8269806/.

[110] I. Y and M. GB. "Development of the polysomnographic database on CD-ROM. Psychiatry and Clinical Neurosciences 53:175-177 (April 1999)." In: *Psychiatry and Clinical Neurosciences* 53 (1999), pp. 175–177.

[111] G. B. Moody and R. G. Mark. "The impact of the MIT-BIH arrhythmia database." In: *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society* 20.3 (), pp. 45–50. ISSN: 0739-5175. URL: http://www.ncbi.nlm.nih.gov/pubmed/11446209.

[112] T. Wartzek, M. Czaplik, C. H. Antink, B. Eilebrecht, R. Walocha, and S. Leonhardt. "UnoViS: the MedIT public unobtrusive vital signs database." In: *Health Information Science and Systems* 3.1 (2015), p. 2. ISSN: 2047-2501. DOI: 10.1186/s13755-015-0010-1. URL: http://link.springer.com/10.1186/s13755-015-0010-1.

[113] S. Saechia, J. Koseeyaporn, and P. Wardkein. "Human Identification System Based ECG Signal." In: *TENCON 2005 - 2005 IEEE Region 10 Conference*. IEEE, 2005, pp. 1–4. ISBN: 0-7803-9312-0. DOI: 10.1109/TENCON.2005.300986. URL: http://ieeexplore.ieee.org/document/4085231/.

[114] J. J. Moré and D. C. Sorensen. "Computing a Trust Region Step." In: *SIAM Journal on Scientific and Statistical Computing* 4.3 (1983), pp. 553–572. ISSN: 0196-5204. DOI: 10.1137/0904038. URL: http://epubs.siam.org/doi/10.1137/0904038.

[115] P. Shukla and K. Bhowmick. "To improve classification of imbalanced datasets." In: *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. IEEE, 2017, pp. 1–5. ISBN: 978-1-5090-3294-5. DOI: 10.1109/ICIIECS.2017.8276044. URL: http://ieeexplore.ieee.org/document/8276044/.

[116] H. Noh, T. You, J. Mun, and B. Han. "Regularizing Deep Neural Networks by Noise: Its Interpretation and Optimization." In: (2017). arXiv: 1710.05179. URL: http://arxiv.org/abs/1710.05179.

[117] J. R. Stradling. "Which aspects of breathing during sleep influence the overnight fall of blood pressure in a community population?" In: *Thorax* 55.5 (2000), pp. 393–398. ISSN: 00406376. DOI: 10.1136/thorax.55.5.393. URL: http://thorax.bmj.com/cgi/doi/10.1136/thorax.55.5.393.

[118] H. Xi and T. Chang. "Image Classification Based on Histogram Intersection Kernel." In: *Journal of Computer and Communications* 03.11 (2015), pp. 158–163. ISSN: 2327-5219. DOI: 10.4236/jcc.2015.311025. URL: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jcc.2015.311025.

[119] S. Boughorbel, J.-P. Tarel, and N. Boujemaa. "Generalized histogram intersection kernel for image recognition." In: *IEEE International Conference on Image Processing 2005*. IEEE, 2005, pp. III–161. ISBN: 0-7803-9134-9. DOI: 10.1109/ICIP.2005.1530353. URL: http://ieeexplore.ieee.org/document/1530353/.

[120] H. Yang, Y. Chen, and F. Leonelli. "Characterization and Monitoring of Nonlinear Dynamics and Chaos in Complex Physiological Systems." In: *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2016, pp. 59–93. DOI: 10.1002/9781118919408.ch3. URL: http://doi.wiley.com/10.1002/9781118919408.ch3.

[121] S. Saechia, J. Koseeyaporn, and P. Wardkein. "Human Identification System Based ECG Signal." In: *TENCON 2005 - 2005 IEEE Region 10 Conference*. IEEE, 2005, pp. 1–4. ISBN: 0-7803-9312-0. DOI: 10.1109/TENCON.2005.300986. URL: http://ieeexplore.ieee.org/document/4085231/.

[122] D. F. Silva, S. Leonhardt, and C. H. Antink. "Synthesis of cardiac signals using a Copula-approach." In: *AIP Conference Proceedings* 2140.1 (2019), p. 020070. DOI: 10.1063/1.5121995. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.5121995. URL: https://aip.scitation.org/doi/abs/10.1063/1.5121995.

Appendix

Respiratory Phase Algorithm

This appendix details the algorithm used in Section 4.2.3.2 to obtain ϕ_{RESP} from the pre-processed RESP based on simple peak detection.

1. Define the minimum peak prominence constraint as 5% of the difference between the average value of the positive RESP and the average value of the negative RESP. This is a more robust method than using the full range of the signal, which is susceptible to fluctuations due to changes in experimental conditions;
2. Peak detection to identify time locations of respiratory peaks with a prominence greater than the constraint;
3. Calculation of the RF with adjacent peak locations and consequent outlier detection with a MAD approach. Outlier peaks were thus rejected and a minimum peak distance constraint was set as the minimum, non-outlier peak distance;
4. Peak detection with the previous constraints, but with the negative value of range, to locate respiratory valleys. Steps thus far are shown in Figure A.1;

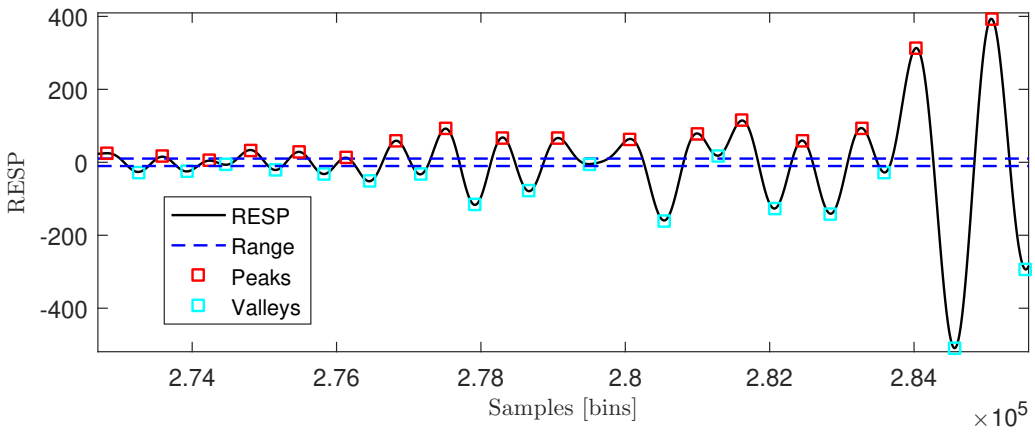

Figure A.1: Representation of the prominence constraint for the RESP.

5. Linear interpolation between each pair of valleys, considering the first has a value of $-\pi$ and the second of π. A saw-tooth signal was obtained, being $-\pi$ where a respiratory cycle starts and π when it ends. This is the typical ϕ_{RESP} waveform, but not the one used in this thesis;

6. Application of the following operation to the previously obtained waveform

$$\phi_{\text{RESP}} = \pi - |\phi_{\text{RESP}}|, \tag{A.1}$$

where $|\cdot|$ is the absolute value operator. The final signal $\phi_{\text{RESP}} : [1, \infty] \rightarrow [0, \pi]$, indeed used in this project, is obtained.

APPENDIX B

Concavity Determination Algorithm

This appendix details the algorithm used in Section 4.2.4.2 to determine whether a segment's portion has positive or negative concavity using derivatives.

1. Calculation of the first derivative of the signal;
2. Transformation of the first derivative into its respective sign signal, ranging from -1 to 1, and calculation of its absolute, cumulative sum;
3. Location of positive or negative transitions in the sign signal;
4. Election of the transition which is temporally closer to the cumulative signal's maximum. If it is a positive transition, determine the concavity to be faced up, otherwise, determine the concavity to be faced down. A representation of the process can be seen in Figure B.1. Although it is not the case in the representation, this algorithm is effective at this task when the portion's signal has energetic fluctuations.

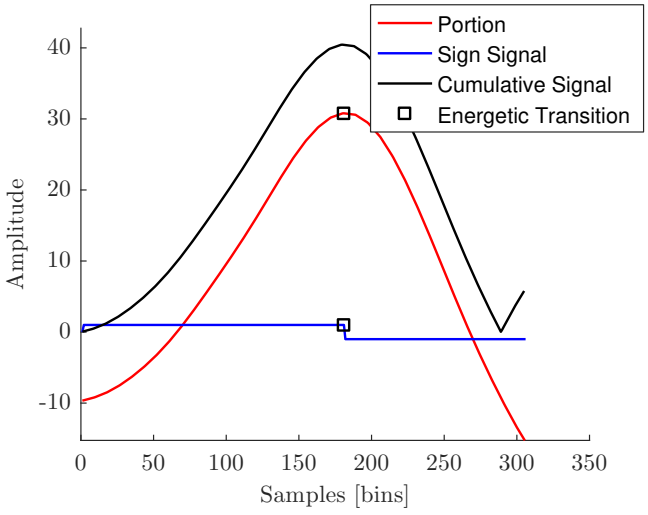

Figure B.1: Representation of the process to determine the concavity of a segment's portion.

Appendix C

Detailed Multimodal Synthesizer Data Structure

This appendix provides a visual representation of the data structure used in Section 4.2.4, to ease its understanding. For effect, Table C.1 is presented ahead using the case of a fictitious MIT-BIH Polysomnographic subject's recording with B segmented beats as an example.

Table C.1: Representation of a matrix $\boldsymbol{D}$ obtained from B segmented beats in the ECG, BP and BCG modalities. The color coding and nomenclature is the same as the one used for the segment's portions exemplified in Section 4.2.4, specifically, Figures 4.7 and 4.8. As such, coefficients corresponding to each portion, fitted according to the respective user-inputted fit_types (e.g. 'gauss3', NaN or 'fourier8'), are placed in the cells where the portion label (P-wave, Q-wave, etc.) is located. All the coefficients from the same segment are placed in the same row by the same order with which portions appear in the segment, regardless of the modality they result from. Cells labeled as "Additional", contain the modality-independent coefficients respective to RR-interval, ϕ_{RESP} and the modality delays. This table is a simplification since a cell containing coefficients from a portion fitted with a GMM 'gauss2' would, for example, enclose seven coefficients, as explained in the same Section.

	Modality-independent				ECG						BP	BCG
	RR-interval	ϕ_{RESP}	Delay BP	Delay BCG	'gauss2'	'gauss2'	'gauss1'	'gauss2'	'gauss2'	NaN	'gauss3'	'fourier8'
Beat 1	Additional				P-wave	Q-wave	R-wave	S-wave	T-wave	Transition	BP pulse	BCG waveform
...	...	...	...	...	...	...	...	...	...	...	...	...
Beat B	Additional				P-wave	Q-wave	R-wave	S-wave	T-wave	Transition	BP pulse	BCG waveform

Annex I

Russian German Conference on Biomedical Engineering 2019

This annex contains the abstract of the paper presented in this year's edition of the Russian-German Conference on Biomedical Engineering, which was held at the Saint-Petersburg Electrotechnical University from the July 3rd to July 6th. It was later published in the American Institute of Physics (AIP) Conference Proceedings on August 12th [122].

Synthesis of Cardiac Signals using a Copula-Approach

Diogo Filipe Pereira Fontes Fernandes Silva, Steffen Leonhardt and Christoph Hoog Antink [a)]

Medical Information Technology, Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Germany

[a)]Corresponding author: hoog.antink@hia.rwth-aachen.de

Abstract. The increasing need for comprehensive medical signal data has been scarcely tackled in the scope of non-classical modalities and multimodal signal. To overcome this, a data-driven, statistical approach to model cardiorespiratory signals combined with copulas to emulate inter and intramodality dependence is suggested. Gaussian Mixture Models, Fourier Series and Sum of Sines are fitted to different physiological portions of segmented heart cycles. Marginal distributions of the yielded parameters and their linear correlation are used to define a Gaussian Copula from which random sets of parameters are generated to simulate new beats. The model is applied to regular and capacitively coupled Electrocardiography (ECG-cECG) as well as a combination of ECG with blood pressure and ballistocardiography (ECG-BP-BCG) to demonstrate generation of multimodal recordings. Results show realistic wave morphologies and relevant inter and intramodality correlations with a wide range of potential applications such as testing, training and integrating signal processing algorithms.

ANNEX II

Biomedical Engineering Summit 2019

This annex contains the extended abstract submitted to this year's edition of the Biomedical Engineering Summit, which will be held at the Order of Engineers in Lisbon on October 11th and 12th, with a focus on Machine Learning in Healthcare and innovation in biological devices. This work was accepted by the scientific committee for participation in the poster display and a presentation in the opening session of the second day.

Combining Synthesis of Cardiorespiratory Signals and Artifacts with Deep Learning for Robust Vital Sign Estimation

Diogo Filipe Silva, Steffen Leonhardt and Christoph Hoog Antink

Medical Information Technology, Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Germany

Abstract. Machine Learning methods have already consolidated their place in highly specific, low-risk clinical tasks and keep advancing towards higher-risk applications such as diagnostics. In particular, cardiorespiratory signals have been the target of a great deal of research by the artificial intelligence community. Modern Deep Learning methods, namely convolutional neural networks, often outperform traditional Machine Learning ones, showing very convincing results, even for clinical standards. However, a set of obstacles, such as logistical problems and legal concerns related to data acquisition, along with the limited sharing of data among healthcare corporations and researchers, have hindered the development of clinically ready Deep Learning algorithms, as these require a substantial amount of annotated datasets to be trained. In the current paper, a multimodal cardiorespiratory signal synthesizer is presented, given the very restricted research output in the area, along with a convolutional neural network taking advantage of signal fusion to perform tasks of detection and classification in cardiorespiratory data. Primarily, the combination of both these tools aims at showing how the generation of large, synthetic datasets can boost Deep Learning's performance in situations where the amount of data available isn't enough and, hopefully, increase the clinical significance of artificial intelligence in cardiorespiratory diagnostics.

A synthesizer was, therefore, designed, aiming at producing artificial cardiorespiratory recordings (electrocardiography, pulse plethysmography, blood pressure, ballistocardiography, capacitively coupled electrocardiography and respiration) of arbitrary length, whose physiological effects and electrical properties mimic those of a particular recording obtained from a specific subject with specific acquisition configurations. Additionally, both pathological beats and artifacts presented in the original signal should be able to be included. Real multimodal recordings from the Fantasia, MIT-BIH Polysomnographic, MIT-BIH Arrhythmia and UnoVis databases were segmented, in order to segregate multimodal heart cycles. With minimal user input, a fitting algorithm was designed to decompose said cycles into Gaussian mixture models, Fourier series, sums of sines, or adequate combinations, ultimately, describing every segmented cycle by means of the coefficients of these fitted, pre-defined functions. These coefficients were then used to describe how each particular feature of the segment is probabilistically distributed and dependent on the remaining inter- and intra-modality coefficients by means of a Gaussian copula – a multivariate distribution highly used in finance to model credit-sensitive assets based on the correlation among uniformly distributed marginals. After defining the copula by transforming each arbitrary coefficient distribution into uniform distributions, a conditional sampling approach was defined to draw new, artificial coefficients dependent on a set of given RR-interval and respiratory phase values as inputs to the copula. Finally, a copula for each type of beat and artifact was designed and integrated into a Markov chain, also estimated from real data, to realistically model how likely one type of beat transitions to another, or to an artifact and vice-versa.

A real-time capable convolutional neural network was designed with the goal of detecting or classifying heartbeats from multimodal data, by efficiently using sensor fusion. While most works choose to fuse information from different sensors earlier or later in the network, a novel architecture was used, which creates as many twin networks as modalities in the signal, plus a fusion network, whose feature maps are

learnable, linear combinations of the feature maps of the modalities' twin networks at every layer. Grid search and Bayesian optimization were used to define the best hyperparameters for the architecture and training was done by considering a signal snippet with a centered annotation to be labelled accordingly and considering label zero the same snippet with a random delay, ensuring class balance.

The copula framework was validated by verifying artificial coefficients were similarly distributed to their original counterparts, which was only achieved by a rigorous segmentation and a successful fitting process, as evidenced by convincing goodness of fit metrics. Time, frequency and phase domain analysis were performed to validate the obtained artificial recordings and it was found that, in every domain, every modality's signal was realistically modeled, achieving convincing visual and numerical results, such as waveform morphology, fiducial point relative amplitudes, Fourier transform peaks, fractal object shape, voltage and frequency range, approximate entropy and fractal dimension. Linear correlation was deemed enough to model physiological effects given the chosen data representation, as the effect of respiratory sinus arrhythmia, Mayer waves, respiratory modulation, among others, can be verified in the signals and checked for high correlation values in the copulas' correlation matrices. These results show how the current framework can be a useful tool for realistic data augmentation. Furthermore, compatibility with state of the art, deterministic models, namely because of the use of the same data representation with coefficients from Gaussian mixture models and Fourier series, provides the opportunity for further, joint development and the possibility of converting the current model into a deterministic one, simply by manually changing the obtained coefficients to yield new, unseen waveforms.

A series of tests were conducted with the designed network and analysed though a score, based on the mean of the sensitivity with positive predictive value, and absolute detection error. Firstly, beat detection using only artificial signals in the training phase yielded an average score of 97,6% and error of 9,3 ms compared to 98,8% and 9,5 ms from real signals, always using real data for testing. Then, real signals in the training set were reduced to about 150 beats, yielding an average score of 53,4% with an average error of 25,0 ms, and posteriorly enhanced with artificially modelled data, which improved the scores to an average of 98,3% and reduced the error to an average of 8,7 ms. Additionally, given a test set with modality-independent artifacts, it was shown that a training set with real data without artifacts performed at an average score of 93,2% with an average error of 9,8 ms, while using an artificial training set with modality-independent artifacts improved the score to an average of 99,1% and 11,4 ms error. Finally, a real, healthy recording used for training was enhanced with an artificially generated ectopic beat recording and successfully used to train the network to classify normal and ectopic beats in real test data with an average score of 92,8% and error of 5,9 ms. The designed architecture was validated by comparable scores and smaller absolute error values than previous, related works. It was also proven the efficacy of data augmentation in not only improving detection tasks with Machine Learning when not enough real data is available for training, and also how including artificial data of a type not contemplated in the original dataset allows algorithms to be more robust in the face of overwhelming noise and capable of new tasks, such as beat classification.

Combining Synthesis of Cardiorespiratory Signals and Artifacts with Deep Learning for Robust Vital Sign Estimation

Diogo Filipe Silva [1], Steffen Leonhardt [2], Christoph Hoog Antink [2]

[1] Faculty of Sciences and Technology, NOVA University of Lisbon. Corresponding author: dfp.silva@campus.fct.unl.pt
[2] Medical Information Technology, Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University

Introduction

- Healthcare is the 3rd industry to change its core business practice the most due to Big Data [1]
- Traditional Machine Learning has been integrated into everyday low-risk clinical practice
- Deep Learning outperforms traditional approaches in higher-risk tasks, from diagnostics to guided surgery, but require larger, annotated datasets for training [2]
- Scientific community is not currently able to attend to these algorithms' needs, namely in the Cardiology domain, hindering their development

Potential Solution: centralized source of large, annotated, synthetic cardiorespiratory datasets

Materials & Methods

Multimodal Synthesizer

Coefficient Extraction

- Multimodal segmentation of an existing dataset's signal using the P-wave onset of the ECG as reference
- Division of the segmented waveforms into physiological portions (e.g. P-, Q-, R-, S- and T-wave)
- Fit Gaussian mixture models, Fourier series and sum of sines functions to portions, extracting resulting coefficients
- Extraction of RR-interval, respiratory phase and pulse delays from other modalities relative to the ECG as additional coefficients

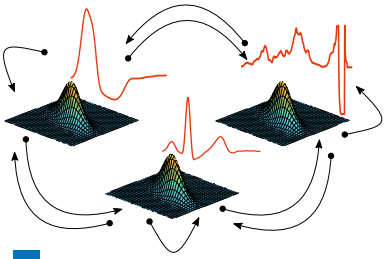

Fig.1 Markov model for normal, PVC beats and artifacts. Arrows are transition probabilities between copulas.

Modeling Physiological Dynamics

- Obtain same-coefficient probability distributions for each portion and calculation of the respective correlation matrix
- Model coefficient dependence with a Gaussian copula, capturing intra- and inter-modality dynamics
- Repeat process to design more copulas characterizing different waveforms
- Design a Markov model to describe how often each waveform type transitions to another

Signal Reconstruction

- Provide an RR-tachogram to obtain artificial coefficients with conditional sampling from copulas mediated by the Markov model
- Reconstruct original waveforms with fitted functions and new coefficients

Convolutional Neural Network- CNN

- Multi-branch architecture: each receives input from one modality's signal + additional fusion branch
- Sensor-fusion at every layer with learnable linear combinations of branches' feature maps
- Hyperparameters chosen with grid search and a tree-structured Parzen estimator

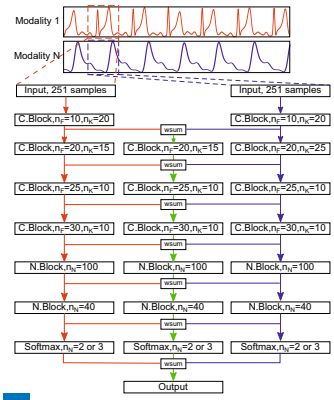

Fig.2 Designed sensor-fusion network. Fusion branch receives results of feature map linear combination.

- Cross-entropy as loss function and mean of sensitivity and positive predictive value as score
- Inclusion of L1 regularization for linear combination coefficients and L2 for remaining weights
- Early-stopping and leave-one-subject-out training with 251-sample signal snippets
- Testing always conducted on full, real datasets

Results

Multimodal Synthesizer

- Synthetic signals resembled their original counterparts in the time, frequency and phase domains
- Several original physiological effects were simulated, such as RSA, Mayer waves, respiratory modulation and pulse transit time
- Most complete cardiorespiratory model to date; each added parameter can simulate more effects than a new module in a dynamical model

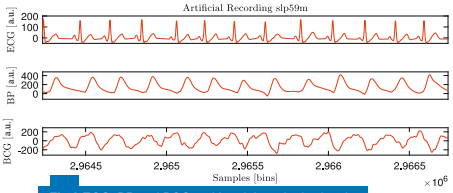

Fig.3 ECG, BP and BCG multimodal synthetic signals.

Data augmentation experiments with CNN

- Training set with variable proportions of artificial data
- Insufficient real training set with increasing artificial data

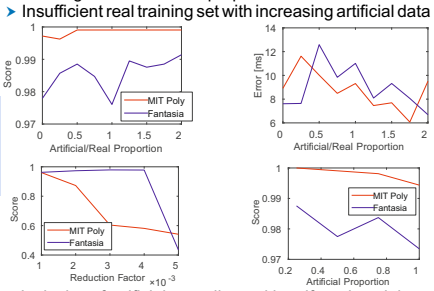

Fig.4 Artificial proportion experiment (top); Reduced training set and improvement (bottom).

- Inclusion of artificial recordings with artifacts in training set
- Classification task with synthetic PVC cycles not featured in the original training set. Classification score of 95%
- In general, data augmentation enhanced scores to state of the art level. Average score of 97.32% and error of 8.65 ms

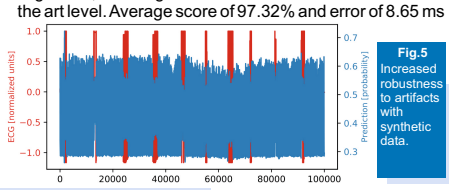

Fig.5 Increased robustness to artifacts with synthetic data.

References

[1] Bernard Marr. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. 2018.
[2] W. W. Stead. "Clinical implications and challenges of artificial intelligence and deep learning." In: JAMA 320.11 (2018), pp. 1107–1108.

Univ.-Prof. Dr.-Ing. Dr. med. Dr. h.c. Steffen Leonhardt
Tel: +49 241 / 80 - 23211
Fax: +49 241 / 80 - 623211
Email: medit@hia.rwth-aachen.de
http://www.medit.hia.rwth-aachen.de

Figure II.1: Poster that was showcased and presented at the conference.