A fast signal quality assessment algorithm applied to a contactless electrocardiogram system

by

Lucas TARGINO LINS

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
IN ELECTRICAL ENGINEERING WITH THESIS
M.A.Sc.

MONTREAL, MARCH 24, 2021

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Ghyslain Gagnon, Thesis supervisor
Department of Electrical engineering, École de technologie supérieure

Mr. Carlos Vazquez, President of the board of examiners
Department of Software and IT engineering, École de technologie supérieure

Mr. Jean-Marc Lina, External examiner
Department of Electrical engineering, École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON MARCH 15, 2021

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

FOREWORD

This master's thesis is the fruit of a work collaboration between the microelectronics and communications research lab (LaCIME), at École de technologie supérieure, Concordia University, and SIG.NUM. The collaboration aimed to contribute to the development of a novel automatized contactless electrocardiogram (cECG) system that promises to revolutionize telemetry and patient monitoring.

This work presents an overview of a complex cECG system, developed by SIG.NUM since 2014, some of its critical problems, and proposes a machine learning based signal quality assessment (SQA) algorithm for characterizing, and quantifying the quality of the measured vital signals, using the fast wavelet transform (FWT). The author emphasizes mathematical analysis of the presented systems, and methods for designing and training classifiers, applied to the context of cECG SQA.

Lucas T. Lins, the main author of this work, is an electronic engineer, graduated in 2018 from the Federal University of Pernambuco (UFPE), Brazil. He has 7 years of experience in research and development (R&D), in which he contributed to the development of many medical devices in the fields of electroencephalography (EEG), brain-computer interfaces (BCIs), functional near infrared spectroscopy (fNIRS), wearable oximetry (SpO2), bioimpedance instrumentation and electrocardiography (ECG).

ACKNOWLEDGEMENTS

At first, I would like to express my gratitude to my supervisor, Ghyslain Gagnon. Thank you for believing on my potential, and giving me the opportunity to prove my competences while working in collaboration with SIG.NUM. Thank you for all the financial and moral support. Thank you for your time and guidance. Thank you for the multiple teaching, and feedback. Thank you for believing on my intelligence and capacity to get back on my feet after several moments of failure.

Second, I would like give thanks to my team leader, Mokhtar Liamini. I am extremely thankful for his moral support, his patience, and kindness during my master's. I certainly could not have had a better mentor. Thank you for teaching me about how to work in the industry, how to make presentations and how to communicate myself with people. Thank you for advising me, brain storming for ideas, and for being a real friend.

I give thanks to my colleagues from SIG.NUM for all their support, encouragements and insights. I am very thankful for the incredible team spirit of the company. It is a joy to work with all of you.

I am extremely thankful for my "Conseillère à la vie étudiante", Romy Schlegel, who was always present to help me whenever I needed. Thank you for being an amazing facilitator, and helping me completing my master's at ÉTS.

I am truly blessed with amazing parents. Thank you, Isaias and Kerima (painho e mainha), for supporting my studies, my career choice and my passion for science! Thank you for educating and preparing me to this world! Thank you for teaching me important values, such as: honesty, generosity and hard work.

I give thanks to my girlfriend Kelsey, who has supported me with love, kindness and patience through her letters and her dedication to make our relationship work. I am truly honoured to be with you!

Furthermore, I give thanks to my friends Marco and Eliana, who acted like parents to me during my studies in Montreal. Your love, kindness, support and acts of service were extremely important to me. I am truly blessed to have both of you in my life!

Last but not least, I want to give thanks to the amazing love of my friends and family in my life. I am who I am because of Love.

Un algorithme rapide d'évaluation de la qualité du signal appliqué à un système d'électrocardiogramme sans contact

Lucas TARGINO LINS

RÉSUMÉ

L'évaluation de la qualité du signal (EQS), ou les systèmes de détection de bruit, promettent d'améliorer la qualité de l'électrocardiogramme (ECG) collecté. Ceci est réalisé en permettant aux systèmes automatisés de collecter et de traiter avec précision les données ECG, ou en facilitant la télémétrie grâce à la sélection précise de données acceptables, afin d'être analysées à distance par un cardiologue.

Ce mémoire est une collaboration de recherche & développement avec SIG.NUM, une compagnie spécialisée dans le développement et la commercialisation d'une technologie innovante sans contact et automatisée pour l'acquisition d'ECG à court et à long terme, connu sous le nom de cECG. Ce travail présente les principes fondamentaux des différents systèmes d'instrumentation ECG; une revue de littérature de la EQS, de ses fondements et de ses applications aux systèmes cECG; une revue de littérature des ondelettes et de ses applications concernant la représentation du signal ECG; et enfin, la conception d'un système de classification EQS, ou détecteur de bruit, à l'aide de réseaux de neurones artificiels et de techniques rapide de traitement du signal basées sur scalogramme d'ondelettes à des fins d'extraction de caractéristiques. Le système de classification EQS est destiné à être utilisé comme un sélecteur de canal en temps réel pour le système cECG développé par SIG.NUM.

L'algorithme développé basé sur l'apprentissage automatique pour ECG EQS est formé par une fusion de plusieurs ensembles de données annotés, disponibles sur PhysioNet (Goldberger, Amaral, Glass, Hausdorff, Ivanov, Mark, Mietus, Moody, Peng & Stanley, 2000). Les résultats du système conçu sont principalement comparés à la technique EQS à une seule dérivation proposée par les travaux de Clifford, Behar, Li & Rezek (2012). En conclusion, cette thèse a montré qu'un simple réseaux de neurones artificiels peu profond est suffisant pour bien modéliser un système de classification EQS généralisé. De plus, les résultats ont montré que la méthode d'extraction des caractéristiques de transformée en ondelettes, proposée par ce mémoire, est non seulement très puissante pour représenter la qualité du signal ECG, mais aussi efficace en termes de coût de calcul.

Mots-clés: ECG, évaluation de la qualité du signal, indice de qualité du signal, transformée en ondelettes rapide, apprentissage automatique, réseaux de neurones, extraction de caractéristiques.

A fast signal quality assessment algorithm applied to a contactless electrocardiogram system

Lucas TARGINO LINS

ABSTRACT

Signal quality assessment (SQA), or noise detector systems, promise to improve the quality of the collected electrocardiogram (ECG). This is achieved by enabling automated systems to accurately collect and process ECG data, or facilitating telemetry through the accurate selection of acceptable data, in order to be analyzed remotely by a cardiologist.

This thesis is a research & development (R&D) collaboration with SIG.NUM, a company that is specialized in the development and commercialization of an innovative proprietary contactless and automated technology for both short and long term ECG acquisition, known as cECG. This work presents fundamentals regarding different ECG instrumentation systems; a review about SQA, its fundaments and applications to cECG systems; a review about wavelets and its applications regarding ECG signal representation; and finally, the design of an SQA classifier system, or noise detector, using artificial neural networks (ANN), and wavelet scalogram-based fast signal processing techniques for feature extraction purposes. The SQA classifier system is meant to be used as a real-time channel selector for the cECG system developed by SIG.NUM.

The developed machine learning based algorithm for ECG SQA is trained by a merge of several annotated data-sets, available from PhysioNet (Goldberger *et al.*, 2000). Results of the designed system are mainly compared with the single-lead SQA technique proposed by the work of Clifford *et al.* (2012). In conclusion, this thesis showed that a simple shallow ANN is sufficient for well modelling a generalized SQA classifier system. Moreover, results showed that the wavelet transform feature extraction method, proposed by this thesis, is not only very powerful to represent ECG signal quality, but also computationally efficient.

Keywords: ECG, signal quality assessment, signal quality index, fast wavelet transform, machine learning, neural networks, feature extraction.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

AF	Adaptive filter
AFE	Analog front end
Ag/AgCl	Silver-silver chloride
ANN	Artificial neural network
ART	Adaptive resonance theory
AUC	Area under the ROC curve
BSS	Blind source separation
BW	Baseline wander
cECG	capacitive ECG / contactless ECG
CMOS	Complementary metal-oxide-semiconductor
CNN	Convolutional neural network
CORDIC	Coordinate rotation digital computer
Cplx	Complexity of a signal
CT	Cooley-Tukey
CTB2	Cooley-Tukey base 2
CV	Cross validation
CWT	Continuous wavelet transform
DAQ	Data acquisition
DC	Direct current

DFT	Discrete Fourier transform
DSP	Digital signal processing
DTWS	Discrete time wavelet series
DWT	Discrete wavelet transform
ECG	Electrocardiography / Electrocardiogram
EMD	Empirical mode decomposition
EMG	Electromyography / Electromyogram
FIR	Finite impulse response
FFT	Fast Fourier transform
FP	Feature pruning
FS	Feature selection
FT	Fourier transform
FWT	Fast wavelet transform
ICU	Intensive care unit
LA	Left arm
LC	Linear convolution
LL	Left leg
LTM	Long-term memory
MII	Modified lead II
MA	Motion artifact

MATLAB	Matrix laboratory
MLP	Multilayer perceptron
mRMR	minimum-redundancy-maximum-relevance
MSE	Mean squared error
MV1	modified lead V1
NSTD	Noise stress test database
PCA	Principal component analysis
PMF	Probability mass function
PLI	Power line interference
PSD	Power spectral density
PVC	Premature ventricular contraction
QMF	Quadrature mirrored filter
RA	Right arm
RBF	Radius basis function
R&D	Research & development
RL	Right arm
RLD	Right leg drive
RLS	Recursive least square
RMS	Root mean squared / Root mean square
ROC	Receiver operating characteristic

SDR	Spectral density ratio
SER	Signal to error ratio
SM	Statistical moment
SNR	Signal to noise ratio
SQA	Signal quality assessment
SQI	Signal quality index
STM	Short-term memory
SVM	Support vector machine
SYN	Synthetic lead
TA	Threshold adjustment
TEI	Tissue-electrode impedance
V1	lead V1
V2	lead V2
V3	lead V3
V4	lead V4
V5	lead V5
V6	lead V6
WCT	Wilson’s central terminal
wPMF	Wavelet probability mass function
ZCR	Zero crossing rate

LIST OF IMPORTANT SYMBOLS AND UNITS OF MEASUREMENTS

I	Lead I
Ip	Posterior lead I
II	Lead II
IIp	Posterior lead II
III	Lead III
$IIIp$	Posterior lead III
F_{pos}	Number of false positives
F_{neg}	Number of false negatives
LA	Left arm potential
LL	Left leg potential
LA	Right arm potential
RL	Right leg
T_{pos}	Number of true positives
T_{neg}	Number of true negatives
V_1	Potential in the fourth intercostal space (between ribs 4 and 5) just to the right of the sternum (breastbone)
V_2	Potential in the fourth intercostal space (between ribs 4 and 5) just to the left of the sternum
V_3	Potential between V_2 and V_4

V_4	Potential in the fifth intercostal space (between ribs 5 and 6) in the mid-clavicular line
V_5	Horizontally potential with V_4, in the left anterior axillary line
V_6	Horizontally potential with V_4 and V_5 in the mid-auxiliary line
V_G	Guard
V_o	Amplifier differential output voltage between two different biopotential measurements (Chapter 1)
WCT	Wilson's central terminal potential
aVL	Augmented vector left
$aVLp$	Posterior augmented vector left
aVF	Augmented vector foot
$aVFp$	Posterior augmented vector foot
aVR	Augmented vector right
$aVRp$	Posterior augmented vector right
Hz	Hertz
s	Seconds
ms	Milliseconds (10^{-3}s)
$\mathbf{Q}$	Matrix of the ensemble of M ECG waveforms; The operation $\mathbf{Q}^T\mathbf{Q}$ is analogous to the time-average correlation matrix estimation (Haykin, 2013, p. 410)
$\mathbf{R}$	Correlation matrix
$\mathbf{X}$	Matrix of measured signal vectors, or data-set

$\mathbf{X}^T$	Transpose of a matrix
$\mathbf{x}_i$	i_{th} column of matrix $\mathbf{X}$
$\|\|\mathbf{x}\|\|_2$	Euclidian norm of a vector
$\mathbf{x}_{in,f}$	Frequency analysis feature vector
$\mathbf{x}_{in,tf}$	Time-frequency analysis feature vector
$\mathbf{x}^{FS}_{in,f,Db4}$	Db4 frequency analysis feature vector after FS
$\mathbf{x}^{FS}_{in,f,Db6}$	Db4 frequency analysis feature vector after FS
$\mathbf{q}_i$	i_{th} ECG waveforms
$\mathbf{f}$	Vector form of the wPMF of the signal $s(n)$
$\mathbf{f}^{(i)}$	Vector form of the wPMF of the signal block $s^{(i)}(n)$
$\mathbf{I}_{nf}$	Vector form of the information distribution of a wPMF
$\boldsymbol{\mu}$	Vector form of the SMs of a wPMF
SNR_E	Eigenvalue SNR estimation using matrix $\mathbf{Q}^T\mathbf{Q}$
SNR_{RMS}	Correlation SNR estimation using matrix $\mathbf{Q}^T\mathbf{Q}$
$\chi(\mathbf{R})$	Eigenvalue spread of $\mathbf{R}$
$P(f)$	Power spectral density in Hz
f	Frequency in Hz
Ω	Frequency in rad s^{-1}
f_c	Cutoff frequency in Hz
f_s	Sampling frequency in Hz

ω	Normalized version of frequency Ω, given in rad
ω_c	Normalized version of cutoff frequency Ω_c, given in rad
$x(t)$	Continuous time signal
$x(n)$	Discrete time signal
$\delta(n)$	Unit sample function
$s(t)$	Continuous time signal
$s(n)$	Discrete time signal
$s^{(i)}(n)$	i_{th} non-overlapping signal block, or windowed signal, from $s(n)$
$\bar{s}$	Average value of the discrete time signal $s(n)$
$\psi_{j,k}(t)$	Continuous time dyadic wavelet function
$\phi_{j,k}(t)$	Continuous time dyadic scaling function
p_j	j_{th} relative wavelet energy
$f(j)$	wPMF of the signal $s(n)$
$f^{(i)}(j)$	wPMF of the signal block $s^{(i)}(n)$
I_j	Information of an event j
S_{ent}	Entropy of information distribution
λ_{max}	Maximum matrix eigenvalue
λ_{min}	Minimum matrix eigenvalue
μ	i_{th} statistical moment of a wPMF
$(\downarrow 2)$	Signal decimation by a factor of 2 operator

$(\uparrow 2)$	Signal expansion by a factor of 2 operator
cos	Cosine function
sin	Sine function
exp	Natural exponential
log	Logarithm
$*$	Symbol for convolution
$\circ$	Symbol for Hadamard product
$+$	Superscript for pseudoinverse of a matrix
$\subset$	Symbol for "is a subset of"
$\in$	Symbol for "belongs to"
$=$	Symbol for "is equal to"
$<$	Symbol for "is less than"
$>$	Symbol for "is greater than"
$\leq$	Symbol for "less than or equal to"
$\geq$	Symbol for "greater than or equal to"
$[a, b]$	The close interval $[a, b]$ of a variable x signifies that $a \leq x \leq b$
$\oplus$	Symbol for denoting the union of spaces (like the union of sets)
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integers

INTRODUCTION

Motivation and context

The standard electrode, in electrocardiogram (ECG) patient monitoring applications, is the conventional silver-silver chloride (Ag/AgCl), supported by adhesive backing and gel for skin contact improvement purposes. Despite its high quality signal extraction, wet electrodes are reported to cause skin irritation and discomfort (Sohmyung Ha, Chul Kim, Chi, Akinin, Maier, Ueno & Cauwenberghs, 2014, p. 1523). Moreover, this kind of electrode is cumbersome, inconvenient for mobile use (Chi, Jung & Cauwenberghs, 2010, p. 106) and not ideal for long-term automated monitoring, as the moisture content evaporates (Sohmyung Ha *et al.*, 2014, p. 1523).

Recent advances in alternative biopotential electrode technologies such as dry-contact and non-contact or contactless ECG electrode (cECG) interfaces promise to overcome many of the aforementioned problems. Despite all improvements, there is still a lot to be done in the domains of signal processing, microelectronics, machine learning, etc., until cECG electrodes can be considered sufficient, compared to the quality of standard Ag/AgCl wet electrodes.

Problem statement

cECG systems are much more susceptible to interference - especially motion artifact (MA) interference - in comparison to the standard Ag/AgCl. In summary, any physical displacement between the electrode and the skin will generate interference, sometimes many orders of magnitude larger than the actual signal (Sohmyung Ha *et al.*, 2014, p. 1525). This severely affects the quality of measured data.

Among the factors that influence the performance and reliability of ECG instrumentation systems, it is fundamental to choose effective signal processing techniques capable of properly

filtering without compromising information and assessing the quality of measured data. Many of the current solutions that fully focus on filtering, such as MA compensation methods, are not very reliable in in regards to preserving signal morphology. For instance, according to Oster & Clifford (2015, p. 346), adaptive filtering, a very popular MA mitigation method, can lead to even worse results, depending on the tuning methodology. That being said, an indication of signal quality would often be preferred over an automatic compensation algorithm or system (Such & Muehlsteff, 2006, p. 51-52).

Moreover, many of the ECG SQA algorithms, currently available in the literature, are exclusively designed to be reliable, but not necessarily computationally efficient. It is important for a digital signal processing (DSP) system to be not only effective, but also computationally efficient, in order to be used in real-time applications.

Thesis objective

This thesis is a research & development (R&D) collaboration with SIG.NUM, a company that is specialized in the development and commercialization of an innovative proprietary contactless and automated technology for both short and long term ECG acquisition. This work aims to present a simple and computationally inexpensive, machine learning based, ECG SQA system, designed to be used as a real-time channel selection algorithm and improve both the performance and reliability of SIG.NUM's cECG system (the SIG.ECG system). The results from this thesis are compared to the work of Clifford *et al.* (2012) ("Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms"), a group of consolidated authors in the field of ECG signal processing.

Thesis organization

The subsequent chapters of this thesis are structured as follows:

1. Chapter 1 provides a background and an overview regarding ECG acquisition systems.

2. Chapter 2 defines important concepts such as signal quality assessment (SQA) and signal quality indexes (SQIs). Furthermore, it also provides an overview of previous works related to SQA and cECG systems.

3. Chapter 3 presents the theory behind wavelets, filter banks and scalograms, applied to ECG SQA.

4. Chapter 4 presents the data-sets, their characteristics and how each of them was prepared to be used for training, validating and testing the machine learning SQA classification system.

5. Chapter 5 presents the application itself, its implementation, the experimental procedures, description of results and performance compared to some of the approaches presented in the work of Clifford *et al.* (2012).

Thesis contributions

The major contributions presented by this work are summarized as following:

1. Aiming the R&D collaboration with SIG.NUM, the work of this thesis was dedicated to contribute to the understanding and improvement of both the performance and reliability of the SIG.ECG system. As a result, the characterization and understanding of the SIG.ECG system contributed to the publication of the work of Weeks, Elsaadany, Lessard-Tremblay, Targino, Liamini & Gagnon (2020).

2. Chapter 2 presents a technical literature review, detailing important concepts and metrics (SQIs) related to SQA systems and cECG systems applications. This review is quite different from the current available ones (e.g., Orphanidou (2018); Oster & Clifford (2015)).

3. Chapter 3 presents two efficient and compact methodologies for representing an ECG signal, inspired from the use of wavelet scalograms. It is believed that such methodologies are novel to the application of ECG SQA.

4. Chapter 4 presents a detailed methodology of data preparation and labelling, applied to SQA classification system types. The methodology is inspired and adapted by the work of Clifford *et al.* (2012).

5. Chapter 5 shows that both the *frequency analysis* and the *time-frequency analysis* feature representations, believed to be novel in the application of ECG SQA, are more accurate when assessing the quality of ECG signals, compared to some of features proposed by the work of Clifford *et al.* (2012).

6. Chapter 5 shows that the *frequency analysis* and the *time-frequency analysis* feature representations are computationally more efficient to be extracted, compared to possible implementations of some of the representations proposed by the work of Clifford *et al.* (2012).

CHAPTER 1

AN OVERVIEW OF ECG SYSTEMS

This current chapter introduces a background and an overview about ECG leads, its acquisition, the SIG.ECG system, different types of ECG electrode technologies and their main sources of interference.

1.1 Biopotential measurement fundamentals

The basic principle of an ECG lead acquisition system is illustrated by Figure 1.1. Generally standard Ag/AgCl electrodes use long cables to connect the patient to the analog front-end (AFE) circuit. Such circuit resumes to a differential voltage measurement system. In case of Figure 1.1, the differential system resumes to an instrumentation amplifier, i.e., the INA129. Up to this point, V_o should already measure an amplified ECG signal.

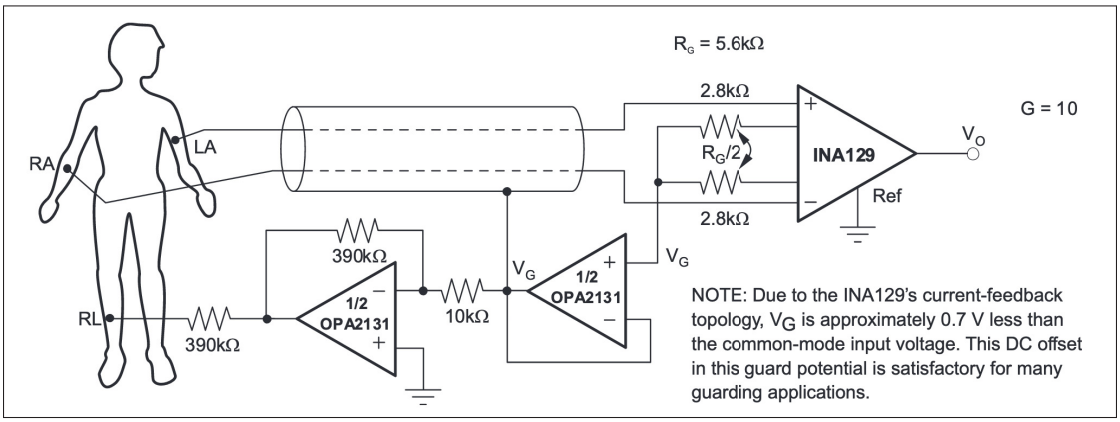


Figure 1.1 Basic principle of ECG acquisition
Taken from Texas Instruments (2020)

Other than the INA129 circuit, the rest of the Figure 1.1 consists on what is known as the *right leg drive* (RLD) circuit (composed by both OPA2131). Basically the signal common to the points RA and LA (V_G) is measured, and sent to the input of another amplifier connected as a buffer, i.e., OPA2131. Finally, the buffer outputs V_G, known as the *guard*, is connected

to both a conductive layer that surrounds the cables (if cables are used), and the input of an inverting amplifier, i.e., another OPA2131. This feedback reduces the amount of common signal interference (e.g., power line interference (PLI) (Friesen, Jannett, Jadallah, Yates, Quint & Nagle, 1990)) measured in the output of the differential system.

Ideally, V_o is an amplified difference between two biopotential measurements, measured at two distinguished points of the body (e.g., RA and LA in Figure 1.1). Therefore, the difference between the outputs of two unipolar amplifiers, that measure two different locations of the body, should behave as a differential system. This is the fundamental principle of a lead measurement.

1.2 The ECG waveform

Figure 1.2 summarizes the structure of a regular ECG waveform, from a healthy patient. P, Q, R, S, T and U, are known as *waves*. Furthermore, each interval and segment are named and described accordingly.

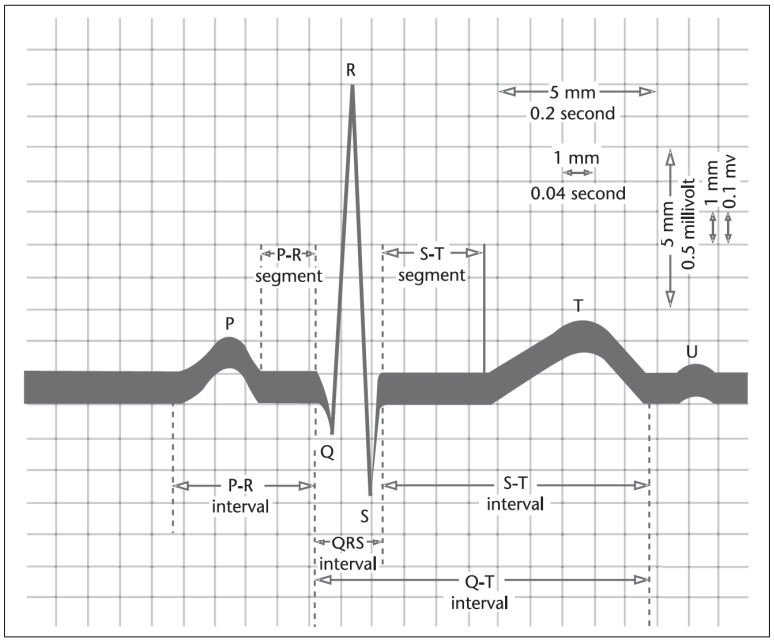
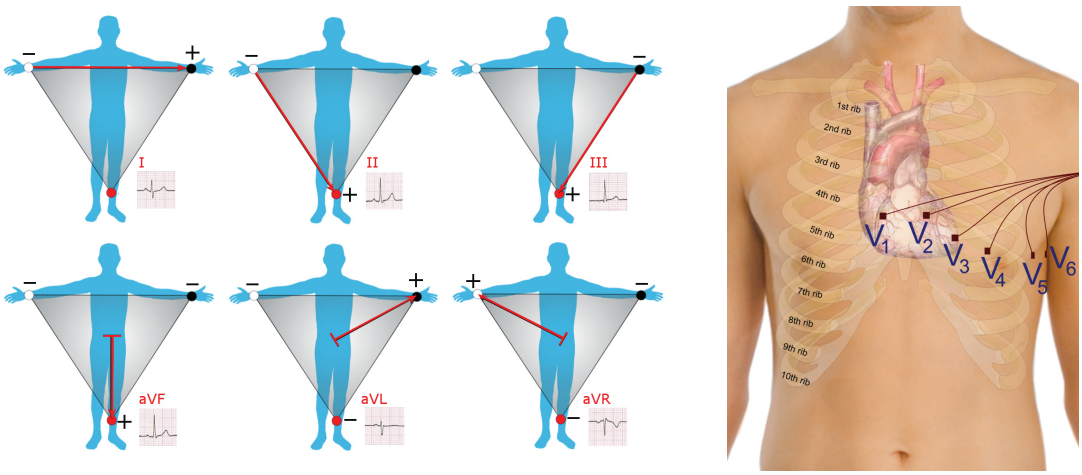


Figure 1.2 Illustration of an ECG waveform, containing all its principal elements along with its respective nomenclatures
Taken from Clifford *et al.* (2006, p. 11)

1.3 The standard 12 leads ECG

The standard 12 leads ECG is composed by 3 limb leads (leads *I*, *II* and *III*), 3 augmented limb leads (leads *aVR*, *aVL* and *aVF*), and 6 precordial leads (leads V_1, V_2, V_3, V_4, V_5 and V_6). Figure 1.3a illustrates both limb leads, augmented limb leads and their respective electrode positions. Figure 1.3b shows the placement of electrodes for measuring the precordial leads.



(a) Limb leads, augmented limb leads positioning and Einthoven's triangle

(b) Precordial leads positioning

Figure 1.3 Standard 12 lead ECG electrode positioning
(a) Taken from Npatchett (2015)
(b) Taken from Häggström (2012)

The potential potential at the left arms is denoted *LA*, the potential measured at the right arm is denoted *RA*, and the potential measured at the left leg is denoted *LL*. The common ground, known as *Wilson's central terminal* (WCT), is given as

$$WCT = \frac{1}{3}(RA + LA + LL). \tag{1.1}$$

The limb leads are calculated according to equations (1.2), (1.3) and (1.4). Where

$$I = LA - RA, \tag{1.2}$$

$$II = LL - RA, \tag{1.3}$$

and

$$III = LL - LA. \tag{1.4}$$

The augmented limb leads are calculated according to equations (1.5), (1.6) and (1.7). Where

$$aVR = RA - \frac{1}{2}(LA + LL) = \frac{3}{2}(RA - WCT), \tag{1.5}$$

$$aVL = LA - \frac{1}{2}(RA + LL) = \frac{3}{2}(LA - WCT), \tag{1.6}$$

and

$$aVF = LL - \frac{1}{2}(RA + LA) = \frac{3}{2}(LL - WCT). \tag{1.7}$$

Finally, each precordial lead is calculated by subtracting V_W from each measured precordial potential.

1.4 Electrode technologies for biopotential sensing

An electrode consists on a technology interface that couples the patient to the AFE. Sohmyung Ha *et al.* (2014) classify biopotential electrodes, mainly, in three classes: wet, dry-contact and

non-contact electrodes. A summary of the fundamentals of biopotential electrode classification, in the vision of Sohmyung Ha *et al.* (2014), is explained in the following subsections. The electrical characteristics of all types of electrodes is shown in Figure 1.4.

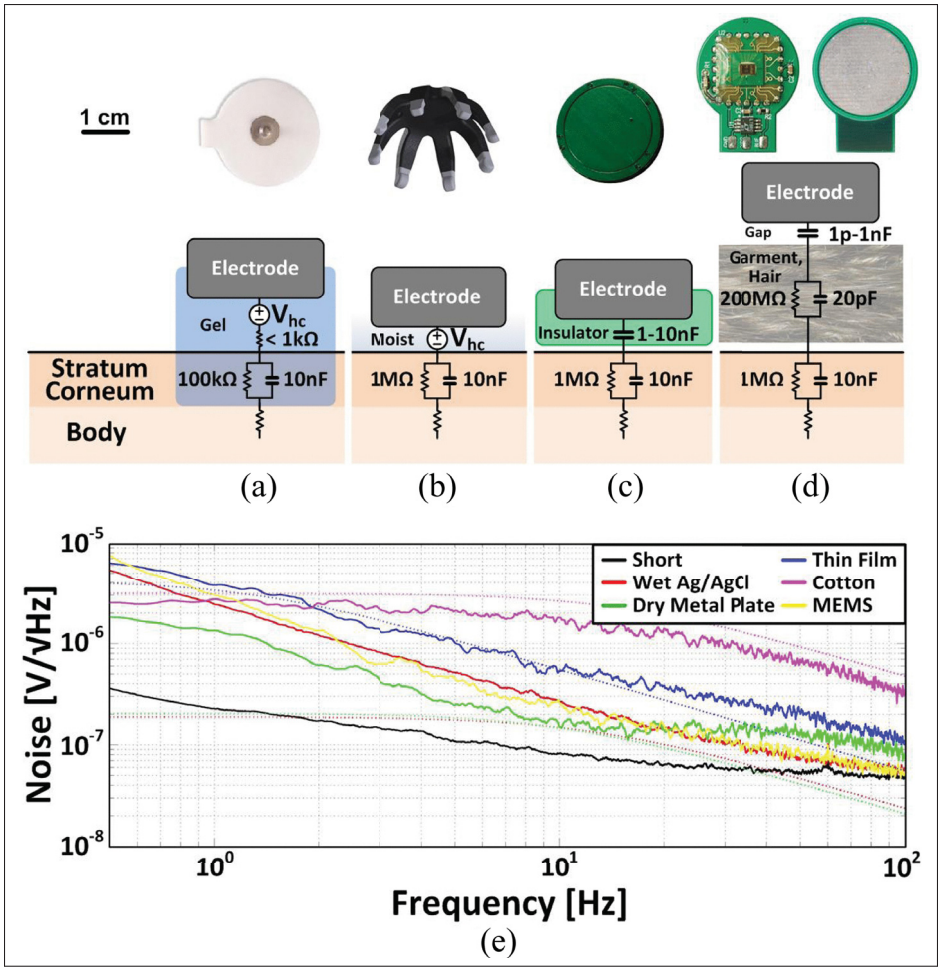


Figure 1.4 Electrical coupling of the skin–electrode interface, or tissue-electrode impedance (TEI), for various electrode topologies, including (a) wet-contact gel-based Ag/AgCl, (b) dry-contact flexible thru-hair sensor (galvanic coupling), (c) thin-film dry-contact insulated capacitive metal plate (capacitive coupling), and (d) non-contact metal plate coupling through hair or clothing such as cotton. (e) Measured noise spectrum of various electrode types, placed at close proximity on forearm at rest, along with predicted (dotted lines) thermal noise limits from measured skin–electrode coupling impedance data, and the instrumentation noise floor of the shorted amplifier, for reference
Taken from Sohmyung Ha *et al.* (2014)

1.4.1 Wet electrodes

Wet electrodes are considered to be the "gold standard" for both clinical and research applications. A typical wet electrode consists of some metal (silver or gold) surrounded by a wet or solid hydrogel containing chloride. A classic example of such electrode is the (Ag/AgCl) (Sohmyung Ha *et al.*, 2014, p. 1523). The performance of wet electrodes degrade as the moisture content evaporates. Its useful lifetime is limited to a few days (Sohmyung Ha *et al.*, 2014, p. 1523). Moreover, since its tissue-electrode impedance (TEI) is characterized to be low (see Figure 1.4), the AFE input impedance is not required to be ultra high.

1.4.2 Dry-contact electrodes

Dry-contact electrodes operate without the use of a wet gel. On bare skin, dry-contact electrodes normally exhibit a TEI by one order of magnitude higher, compared to wet electrodes (Baba & Burke, 2008, p. 95). Unlike wet electrodes, the performance of dry-contact electrodes increase with time, as TEI tends to decrease due to sweat and moisture build-up (Geddes & Valentinuzzi, 1973, p. 356). This property has a huge impact on the noise performance of dry-contact electrodes (see the yellow, and green noise curves, compared to the red noise curve, shown in Figure 1.4).

Finally, dry-contact electrodes can be divided into two main sub-classes: galvanic coupling (Figure 1.4 (b)) and capacitive coupling electrodes (Figure 1.4 (c)). The former simply uses a metal plate that connects the skin to the AFE input. The latter utilizes some thin known dielectric layer to form an insulated contact to the skin. Notice that both topologies make the use of skin-contact for interconnection. Capacitive coupling electrodes are also known as "capacitive electrodes".

1.4.3 Non-contact or contactless electrodes

Contactless electrodes, can be considered a special case of dry-contact capacitive electrodes, where the TEI is not well determined, neither approximately constant. Hence, this type of

electrode operates without gel, through an insulation layer of unknown dielectric properties (e.g., clothing), enabling signal acquisition without direct skin contact. Generally, the TEI of such electrodes is unknown, and can be very high, in the order of tens of picofarads in parallel with hundreds of megaohms (Sohmyung Ha *et al.*, 2014, p. 1524). Thus, the AFE input impedance of such type of electrode is required to be extremely high.

1.5 The SIG.ECG system in a nutshell

Figure 1.5 shows an electronic mattress that containing a matrix of 8x16 AFEs (cECG sensors array), and the central processing unit (CPU), that should connect to the mattress, read and process the analog signals while communicating with other devices.

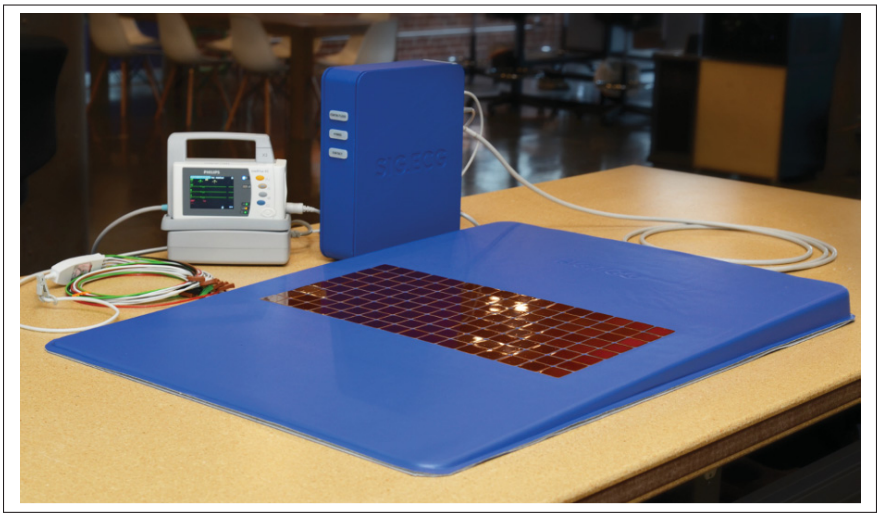


Figure 1.5 The SIG.ECG system
Courtesy of SIG.NUM

1.5.1 The SIG.ECG RLD system

From figure 1.4, it became clear that the coupling between an electrical circuit and the patient skin, can be represented by an impedance, connected in series. A simplified electrical model of a unipolar biopotential acquisition from contactless or capacitive electrodes can be seen in Figure 1.6. The source represents the patient skin, C_2 is modelled as a mixture of air and hair,

and C_1 represents some arbitrary dielectric material (e.g., fabric), R_{in} and C_{in} are both input capacitance and input resitance of the AFE. It is important to mention that, in this simplified model, the skin impedance is considered to be zero (fully conductive). Such approximation is reasonable, since the C_{eq} is the dominant TEI.

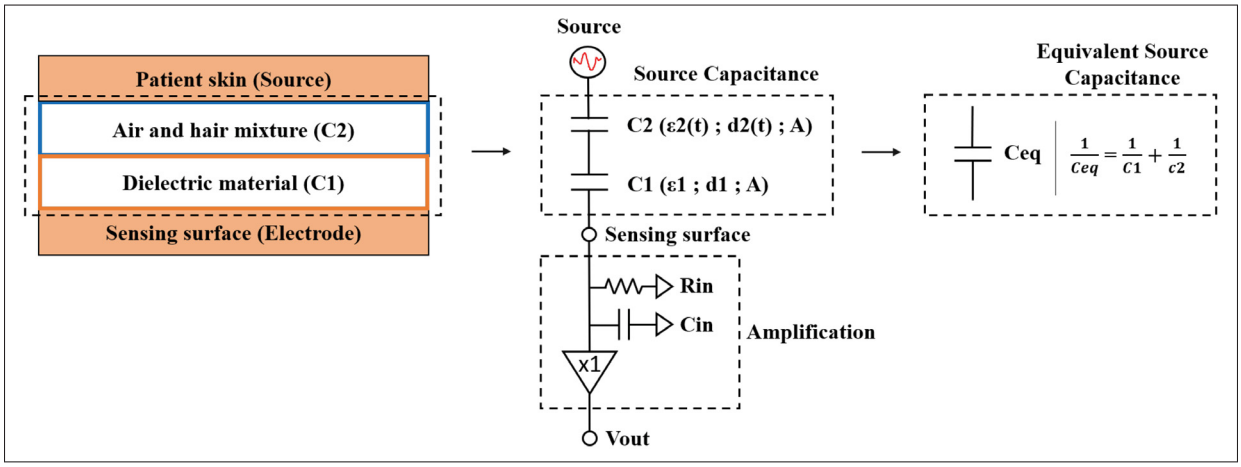


Figure 1.6 Capacitive coupling of skin biopotentials and electrode
Taken from Lessard-Tremblay *et al.* (2020)

The principle of the RLD circuit is illustrated by Figure 1.7. This example illustrates how the SIG.ECG system is used in the application of echocardiography, where an average sum of the 3 best selected sensors is inverted and fed back to the patient body.

In the echocardiography application, a selection algorithm determines the approximately best subset of sensors to automatically provide an accurate ECG acquisition (Weeks *et al.*, 2020, p. 4122). In this case, the subset is equal to 3 sensors. The RL terminal is basically a capacitive coupling interface that injects the inverted common signal back to the patient body. The principle of capacitive coupled feedback is analogous to the principle of capacitive coupling for biopotential acquisition.

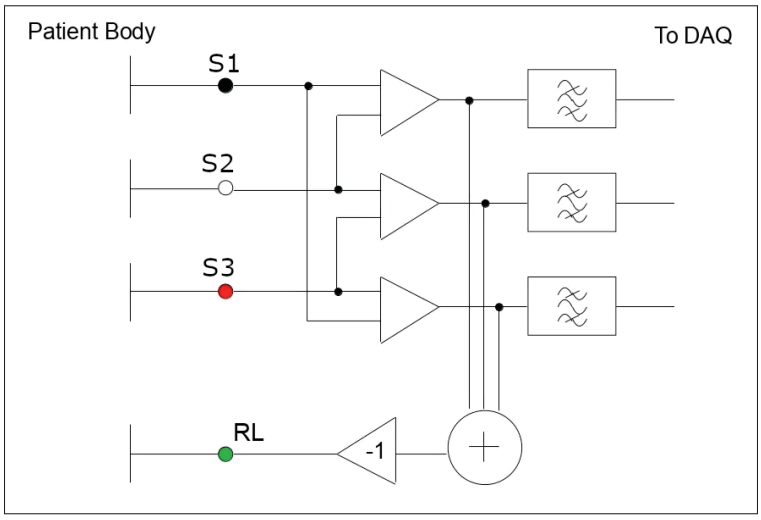


Figure 1.7 SIG.ECG RLD circuit principle
Adapted from the work of Weeks *et al.* (2020)

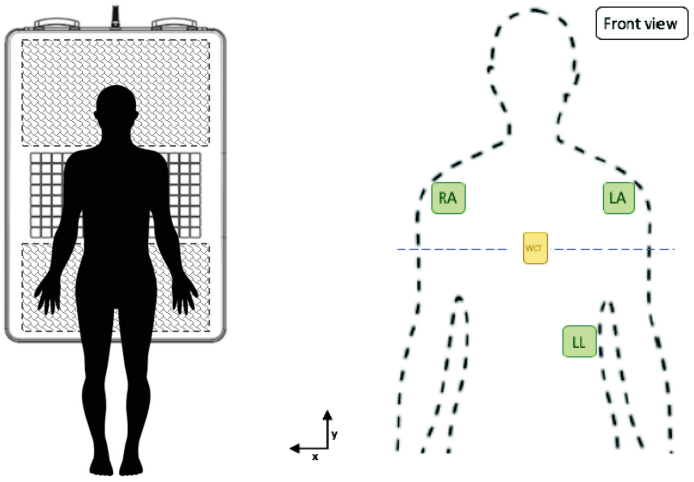
1.6 SIG.NUM custom 6 leads

The custom leads, defined by SIG.NUM are analogous to the standard 12 leads ECG, except that the precordial leads are not taken in consideration. Figure 1.8b redefines the positions of RA, LA and LL to the posterior part of the body, and Figure 1.8a illustrates an example of patient monitoring, where it is possible to see the matrix of sensors (AFEs) and two capacitive coupling interfaces (top and bottom of the array matrix) for the RDL feedback.

The procedures for obtaining the custom limb leads and custom augmented limb leads are the same as described by equations (1.2), (1.3), (1.4), (1.5), (1.6) and (1.7). Furthermore, leads I, II, III, aVR, aVL and aVF are called Ip, IIp, $IIIp$, $aVRp$, $aVLp$ and $aVFp$, when RA, LA and LL are measured from the posterior part of the body.

1.7 Noise vs Interference

Prior to the continuation of the topics presented in this thesis, it is important to clarify the difference between noise and interference. According to Santos (2011), noise sources are generally intrinsic to the instrumentation device electronics (e.g., thermal noise and flicker



(a) Patient placed over the SIG.ECG mattress

(b) SIG.ECG lead positioning

Figure 1.8 Illustration of (a) how a patient interfaces with the SIG.ECG system and (b) its respective posterior limb leads and posterior augmented limb leads positioning, along with the approximate location of the *Wilson's central terminal*
Courtesy of SIG.NUM

noise present in CMOS devices (Lundberg, 2002)), while interference sources are related to external coupling (e.g., PLI). In this thesis, both terms ("interference" and "noise") will be treated equivalently, since many authors do not distinguish them.

1.8 Sources of interference related to ECG instrumentation systems

Based on the work of Oster & Clifford (2015), Table 1.1 shows possible types of interference events that affect the quality of measured signals, and are present in the majority of ECG instrumentation applications.

Table 1.1 Different types of interference that can affect ECG signals
Adapted from Oster & Clifford (2015)

Interference name	**Description**
Power line interference (PLI)	Narrow frequency bandwidth around 50 Hz or 60 Hz depending on the geographic location.
Electrode pop or contact noise	Sharp changes, with possible saturation, due to the contact loss between the electrodes and the skin.
Motion artifact (MA)	Rapid baseline changes, with possible saturation due to the electrode movement. Where movement can produce either rapid changes in the gain of the AFE (lateral movements), or triboelectric effects (horizontal movements) (Ottenbacher & Heuer, 2009; Wartzek, Lammersen, Eilebrecht, Walter & Leonhardt, 2011).
Electromyographic (EMG) noise	High-frequency noise induced by the patient's muscle electrical activity.
Baseline wander (BW)	Low-frequency noise, often induced by the patient's respiration. Both MA and BW are related to movement in electrodes. They differ in terminology due to frequency characteristics.
Data collecting device noise	Saturation induced by the signal processing or communication hardware.
Electronics noise	Instability issues, cut-off frequency alterations, thermal or flicker noise introduced to measurement, etc. Mostly caused by design fault issues.
Electrosurgical noise	High-frequency noise being generated by other medical (or non-medical) devices present in the clinical environment.
Quantization noise and aliasing	Induced by the necessity of digitizing and band-limiting the signal.
Signal Processing artifacts	Distortion induced by signal processing (e.g., ringing or Gibbs oscillations).

ECG recordings can serve as a good indicator of heart activity and are useful for early diagnosis of cardiac diseases, provided that signals are properly acquired and interpreted. When the signal quality is not properly controlled, it can lead to misinterpretation, and eventual false diagnose of patients. MA, BW, and EMG types of interference are most commonly present in ECG recordings, and are very challenging to mitigate due to frequency overlap with the ECG signal, regardless of the electrode technology. Since many of the known artifact mitigation solutions are not always very reliable, signal quality assessment (SQA) systems are considered to be very a attractive solution. Moreover, SQA systems are very helpful when it comes to signal selection applications for automated ECG multi-channel systems. That being said, an elaborated overview ECG SQA systems is presented in chapter 2.

1.9 Conclusion

This chapter presented the main important fundamentals regarding biopotential intrumentation, ECG signals, the SIG.ECG sytem, types of ECG electrodes, and its sources of interference. Moreover, this current chapter also mentions some of the challenges and motivations for assessing the quality of ECG signals, instead of mitigating artifacts.

CHAPTER 2

SIGNAL QUALITY ASSESSMENT APPLIED TO ECG SYSTEMS

As mentioned in Chapter 1, the use of filtering techniques, in many times, can lead to signal distortion. Hence, in many application, such as automated disease diagnosis, the signal morphology can be severely altered by filters or compensation algorithms, leading to false alarm detection of diseases. Therefore, the main reason why signal quality assessment (SQA) systems are attractive, in order to select signals that are useful for different applications, such as diagnosis purposes.

Before continuing with the fundamentals of SQA and signal quality indexes (SQIs), it is important to understand how the performance of a filtering algorithm can be evaluated.

2.1 Performance evaluation of filtering algorithms

In many simulation experiments it is possible to have direct access to the clean signal (ground-truth) and the pure noise (e.g., the works of Chang & Liu (2011); Kuzilek, Kremen, Soucek & Lhotska (2014); Ren, Du, Li, Hu, Yang & Abbas (2017)). In such case, many metrics like mean squared error (MSE) (Ren *et al.*, 2017), signal-to-noise ratio (SNR) (Blanco-Velasco, Weng & Barner, 2008), signal-to-error ratio (SER) (Blanco-Velasco *et al.*, 2008), among others, can be easily calculated. Notice that these indicators have the role of quantifying the signal quality before, and after filtering.

In many real life scenarios, it is not feasible to have access to the clean signal and the pure noise. Hence, it is necessary to use alternative metrics, such as: Counting number of QRS complexes and comparing the estimation of the true QRS complexes, estimated by a specialist, with the results of an automated QRS detector algorithm (Wiklund, Karlsson, Östlund, Berglin, Lindecrantz, Karlsson & Sandsjö, 2007; Xu, Rabotti, Zhang, Harpe, Mischi, Meftah & Ouzounov, 2018); Comparison of the filtered lead with some clean reference lead or signal (Jarchi, Rodgers,

Tarassenko & Clifton, 2018; Wedekind, Kleyko, Osipov, Malberg, Zaunseder & Wiklund, 2018; Xu, Rabotti, Zhang, Ouzounov, Harpe & Mischi, 2019); etc.

Moreover, there are even situations where there is no available clean reference lead nor an accessible specialist to identify useful QRS complexes. Hence, Outram (1998, p. 65) suggests estimating the SNR, by assuming that the noise is purely white. For the understanding of such estimation, it is important to define that, in this thesis, bold uppercase letters are used to refer to a matrix, that can contain several column vectors. Bold lowercase letters, represent column vectors. Examples of a matrix and a vector time-series of length N are shown in equations (2.1) and (2.2). Where

$$\mathbf{X}(n) = \Big[\mathbf{x}_1(n), \mathbf{x}_2(n), \cdots, \mathbf{x}_M(n)\Big], \tag{2.1}$$

and

$$\mathbf{x}(n) = \Big[x(n-N+1), x(n-N+2), \cdots, x(n-1), x(n)\Big]^T. \tag{2.2}$$

For the sake of notation simplicity, many times $\mathbf{x}(n)$ will be simply referred as $\mathbf{x}$. Given such definitions, Outram (1998, p. 65) estimates the SNR according to the following equation:

$$SNR_E = \sqrt{\frac{\lambda_{max}}{M - \lambda_{max}}}. \tag{2.3}$$

Where λ_{max} is the eigenvalue of the matrix $\mathbf{Q}^T\mathbf{Q}$. Given some signal $\mathbf{s}$, its filtered estimation is referred as $\hat{\mathbf{s}}$. Hence, $\mathbf{Q}$ is the matrix of the ensemble of M ECG waveforms (ECG complexes, or PQRST complexes), detected by some QRS detection algorithm applied to the filtered data (e.g., $\hat{\mathbf{s}}$) (Outram, 1998, p. 271) (Wu, Shen, Zhou, Lin, Zeng & Gao, 2013).

Likewise, Outram (1998, p. 66) introduces another method that estimates the SNR through correlation. Thus,

$$SNR_{RMS} = \sqrt{\frac{\eta}{1-\eta}}. \tag{2.4}$$

Where

$$\eta = \frac{2}{M(M-1)} \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} \mathbf{q}_j^T \mathbf{q}_k, \tag{2.5}$$

and $\mathbf{q}_i$ is the i_{th} column of $\mathbf{Q}$ (Outram, 1998, p. 271) (Wu *et al.*, 2013). Both methods of SNR estimation are designed for ECG signal analysis applications. The mathematical demonstration of equations (2.3), (2.4) and (2.5) can be found in the work of Outram (1998).

Both methods were validated by simulating an ECG signal corrupted by some known Gaussian noise source, where the true SNR is compared to SNR_E and SNR_{RMS}. Results showed that, for high SNR, both techniques demonstrated acceptable results. On the other hand, for low SNR, SNR_E showed poor accuracy (Outram, 1998, p. 67).

Furthermore, Haykin (2013, p. 807 - 811) demonstrates a generalist approach for estimating signal quality, compared to white noise. Let the Hermitian matrix $\mathbf{R}$ denote the N-by-N autocorrelation matrix of a wide-sense stationary discrete-time stochastic process $\hat{\mathbf{s}}$, in the same format as described by (2.2). The so called *eigenvalue spread* $\chi(\mathbf{R})$ is bounded according to the following inequality:

$$\chi(\mathbf{R}) = \frac{\lambda_{max}}{\lambda_{min}} \leq \frac{S_{\hat{\mathbf{s}},max}}{S_{\hat{\mathbf{s}},min}}. \tag{2.6}$$

Where $S_{\hat{\mathbf{s}},max}$ and $S_{\hat{\mathbf{s}},min}$ denote, respectively, the maximum and minimum values of the power spectral density (PSD) of $\hat{\mathbf{s}}$. As N approaches infinity, $\chi(\mathbf{R})$ approaches the ratio $S_{\hat{\mathbf{s}},max}/S_{\hat{\mathbf{s}},min}$

(Haykin, 2013, p. 811). Such ratio is interesting when quantifying the difference of power between the desired signal and decorrelated non repetitive noise.

It is important to understand that all the presented concepts already serve as indicators that quantify the quality of ECG signals. This is the core of SQA.

2.2 Signal quality assessment

SQA systems seek to improve the reliability of physiological measurements (Orphanidou, 2018, p. 1), aiming to identify instances of artefact (in addition to *interference / noise*) in a signal segment, such that the signal extracted from that segment can be either *ignored* or *corrected / enhanced* (Orphanidou, 2018, p. 5). Orphanidou (2018, p. 5) also states that the terms *quality assessment*, *quality appraisal* and *artefact / noise detection* are considered equivalent.

Furthermore, it is important to mention that, according to Orphanidou (2018, p. 19-20), SQA is defined in two ways, in terms of rigorousness:

- **Basic quality:** Where R-peaks are clearly identifiable. In such case, a reliable heart rate can be extracted from good quality signals, as well heart rate variability and some types of arrhythmia can be detected and analyzed.
- **Diagnostic quality:** Where P wave (if present), the QRS complex and the T wave are clearly identifiable. In such case, the ECG signal can also be used for clinical diagnosis of more subtle conditions such as: myocardial ischemia and coronary heart disease.

As it will be presented in Section 2.4, many works actually quantify the quality of measured ECG signals by *scoring*, instead of *labelling / classifying* (e.g., *acceptable* or *unacceptable*). Hence, in this thesis, types of SQA, regarding the output response, are understood as:

- **Classification type of quality assessment:** Where signals are classified into categories (e.g., *acceptable* or *unacceptable*).

- **Scoring type of quality assessment:** Where the output of the system is either a continuous value that can be based on several indicators (e.g., SNR_E or SNR_{RMS}), or selection of best leads based on quality ranking (e.g., multichannel ECG signal selection applications). Notice that, given the appropriate threshold, it is possible to separate signals into different classes.

Unlike filtering techniques, that can be easily separated into families of algorithms (e.g., adaptive filters (AFs), blind source separation (BSS), empirical mode decomposition (EMD), etc.), many of the quality assessment techniques are very unique and specific. Hence, the reason for separating SQA systems in these two classes (*classification* or *scoring* types) is simply to assist the reader during the process of finding and organizing what is in the literature in regards to SQA, since many SQA methods are not always explicit in many publications. Many authors implicitly refer to SQA systems as artifact detection algorithms, or methods for evaluating and validating the performance of filtering or acquisition systems, under certain conditions. Thus, the reader should be aware that many SQA systems or metrics might be implicit in many works, such as: MA filtering publications (as discussed in Section 2.1).

2.3 Signal quality indexes

The term SQI is equivalent to *estimation of noise level* (Oster & Clifford, 2015, p. 353). Examples of SQIs, or metrics, presented in the works of Clifford *et al.* (2012, p. 1425-1426), Behar, Oster, Li, Clifford, Qiao Li, Clifford, Li, Clifford, Qiao Li & Clifford (2013, p. 3-4), Zhao & Zhang (2018), Oh (2004), Castro, Varon, Torfs, Van Huffel, Puers & Van Hoof (2018), Koenig, Rehg & Rasshofer (2015) and (Hou, Xiang, Dong, Xue, Xiong & Yang, 2018) are summarized in Table 2.1.

Each SQI has its own particular motivation and justification. For instance, it is expected a good ECG signal to be highly non-gaussian (Behar *et al.*, 2013, p. 3), due to its quasi-stationary properties. Hence, kSQI should be close to zero, for gaussian random processes (low quality ECG signals). Furthermore, Behar *et al.* (2013, p. 3) also expect the ECG signal to be highly skewed (high sSQI), due to the QRS complex.

Table 2.1 Examples of SQIs, or metrics, found in the literature

SQI	**Description**
iSQI	The percentage of beats detected on each lead which were detected on all leads.
bSQI	The percentage of beats detected by the *wqrs* algorithm (Zong, 2010), that were also detected by the *eplimited* algorithm (Hamilton, 2002).
rSQI	The ratio of the number of beats detected by the *eplimited* algorithm (Hamilton, 2002), over the number of beats detected by the *wqrs* algorithm (Zong, 2010).
pSQI	The relative power in the QRS complex: $\int_{5Hz}^{15Hz} P(f)df / \int_{5Hz}^{40Hz} P(f)df$.
sSQI	The third standardized moments of a signal (skewness).
kSQI	The fourth standardized moments of a signal (kurtosis).
fSQI	The percentage of the signal, that appears to be a flat line.
basSQI	The relative power in the baseline: $1 - \int_{0Hz}^{1Hz} P(f)df / \int_{0Hz}^{40Hz} P(f)df$.
pcaSQI	A ratio comprising of the sum of the eigenvalues associated with the five principal components over the sum of all eigenvalues obtained by principal component analysis applied to the time-aligned ECG cycles detected in the window by the *eplimited* algorithm (Hamilton, 2002), segmented at 100 ms either side of the R-peak. Interestingly, such SQI is somehow similar to SNR_E, presented in Section 2.1.
cSQI	The ratio between the standard deviation (STD) estimate and the mean estimate (Hayn, Jammerbund & Schreier, 2012), both from the distribution of R-R intervals within a segment of ECG: $\sigma_{\hat{RR}} / \mu_{\hat{RR}}$.
M_x	The so-called *mobility* (Oh, 2004, p. 41) of signal x is the ratio between the STD estimate of the derivative of x (known as x'), and the STD estimate of x: $\sigma_{x'} / \sigma_x$.
Cplx	The so-called *complexity* (Oh, 2004, p. 40) of signal x is the ratio between the *mobility* of x' and the *mobility* of x: $M_{x'} / M_x$.
ZCR	The so-called *zero crossing rate* (ZCR) (Oh, 2004, p. 39) is the ratio between the total number of times the signal x changes polarity ($x(n)x(n-1) < 1$) and the number of samples.

Table 2.1 Examples of SQIs, or metrics, found in the literature *(cont'd)*

SQI	**Description**
SDR	The so-called *spectral density ratio* (SDR) is a generalization of the basSQI or pSQI. Its value is calulated as: $\int_a^b P(f)df / \int_c^d P(f)df$, for several possible intervals $[a, b]$ and $[c, d]$. Some interesting interval references are proposed by the work of Castro *et al.* (2018, p. 6).
corrSQI	Correlation comparison of a detected QRST complex, with a tamplate obtained by averaging QRST complexes (Castro *et al.*, 2018, p. 7).
varSQI	The variance of the heart rate. In the work of Koenig *et al.* (2015, p. 199), this SQI is 1 if the variance of the heart rate is larger then 40, else the SQI is 0.
aSQI	The absolute value of the heart rate. If the absolute value is in between 30 and 200, the SQI is 0. Else is -5. Such SQI is proposed by Koenig *et al.* (2015, p. 199).
$Rate$	The gridded phase portraits of an ECG signal is constructed, and $Rate = \frac{N_p}{N_t}$. Where N_p is the number of boxes visited by the phase portrait, and N_t is the total number of boxes in the phase space. This SQI is originally proposed by Hou *et al.* (2018, p. 7-8).

It is important to state that SQA and SQI are referred as synonyms by many authors. However, in this thesis, both terms are treated differently. Therefore, in this thesis, SQIs are understood as signal features that *quantify / estimate* some type of noise level, while SQA is understood as a *system* that performs *artifact / noise detection* or *signal evaluation*. For instance, the non-linear combination of SQIs can be used to design very powerful *artifact / noise detectors* (classification SQA systems), as it is shown, or mentioned, in the works of Behar, Oster, Li & Clifford (2012); Behar *et al.* (2013); Clifford *et al.* (2012); Oster & Clifford (2015).

2.4 SQA applied to contactless ECG systems

The following list summarizes different quality appraisal techniques applied to cECG systems. The list is formatted as **authors:** type of SQA - description of SQI or metric.

1. **Seung Min Lee, Ko Keun Kim & Park (2008):** Classification type of quality assessment prior to filtering - artifacts are identified by using the FWT to decompose the ECG signal and comparing the energy level increase. When the energy level increase is lower then a specific threshold (artifact identification), the high frequency components are zeroed (filtering). Subsequently, the inverse FWT is applied to the remaining coefficients.

2. **Ottenbacher, Kirst, Jatoba, Huflejt, Grossmann & Stork (2008):** Classification type of quality assessment - motion artifact is detected by the use of heuristical thresholding on correlated artifact signals from either impedance sensing, or accelerometers.

3. **Heuer, Chiriac, Kirst, Gharbi & Stork (2011):** Classification type of quality assessment - the presence of motion artifact is detected through a capacitance measurement module, integrated to the ECG system, that senses coupling capacitance variations.

4. **Schumm, Arnrich & Troster (2012):** Scoring type of quality assessment - the *eplimited* algorithm (Hamilton, 2002) is used to detect R-peaks, and a *support vector machine* (SVM) network is used to classify if the R-peak detection is correct or not. The SVM network classifies an ECG waveform, contained in a window of 150ms, an interval where R-peaks are counted as correct, according to the American National Standards Institute and the Association for the Advancement of Medical Instrumentation (Schumm *et al.*, 2012, p. 31). The score is $QI = \frac{\text{samples with QL = 1}}{\text{overall amount of samples}}$, where QL is the classification of the window sample.

5. **Schneider, Koellner & Heuer (2012):** Scoring type of quality assessment - each R-peak of the reference signal has its location manually annotated. An algorithm is used to detect the R-peaks location of the cECG signal, subsequently the accuracy of each annotation is compared to the locations of the reference signal.

6. **Koenig *et al.* (2015):** Scoring type of quality assessment - the quality is assessed through the operation $SQI = kSQI + varSQI + aSQI$. In this work, kSQI consists on a flag, where its value is 1 for signals with kurtosis lower or equal to 80, or 0 otherwise. Higher SQI signals are considered to be more reliable. SQI is used as a control feedback for a heart rate estimation KF.

7. **Bohm, Antink, Leonhardt & Teichmann (2015):** Scoring type of quality assessment - a reference signal is estimated by using a linear FIR deconvolution. The Mean Squared Error (MSE) of the estimated signal compared to the reference signal is implicitly defined as an SQI.

8. **Castro, Morariu, Torfs, Van Hoof & Puers (2016):** Classification of acceptability for either beat detection or higher quality monitoring applications - the classification is based on empirical thresholding. bSQI, kSQI (both as described by Li, Mark & Clifford (2008)) and $SDR = \int_{4Hz}^{50Hz} P(f)df / \int_{0.5Hz}^{70Hz} P(f)df$ are used. If $bSQI > 0.6$, then the signal is considered suitable for beat detection. If $SDR > 0.5$ and $kSQI > 0.5$, the signal is considered to be acceptable for higher quality monitoring.

9. **Castro *et al.* (2018):** Classification type of quality assessment - two-class or three-class quality assessment classification based on heuristic threshold of SQIs, such as: kSQI, SDR, bSQI and corrSQI.

10. **Castro *et al.* (2018):** Scoring type of quality assessment - selection of best lead based on non-linear decision rules build and SQI-Fusion.

11. **Wedekind *et al.* (2018):** Scoring type of quality assessment - signals are filtered and selected from a multi-channel cECG system, by applying Blind Source Separation (BSS) based on Independent Component Analysis. A BSS heartbeat-detection-based approach, combined with an altered Kurtosis calculation selects the best quality component.

12. **Hou *et al.* (2018):** Scoring type of quality assessment - the $Rate$ is used to generate a weight matrix, that synthesizes a single-channel ECG signal, from the linear combination of a multi-channel cECG system.

13. **Kido, Tamura, Ono, Altaf-Ul-Amin, Sekine, Kanaya & Huang (2019):** Classification type of quality assessment - a 1D convolutional neural network (CNN) is used to classify signals into 3 categories: clear, blurry and noisy.

14. **Leicht, Skobel, Knackstedt, Mathissen, Sitter, Wartzek, Mohler, Reith, Leonhardt & Teichmann (2019):** Classification of signals into 4 categories - the indexes or categories classified by two cardiologists consisted on the following labels: good, relative good, relative bad, bad.

2.5 Conclusion

It is important to consider the fact that there is no perfect artifact removal filter design solution. In addition, many of the MA mitigation techniques can be computationally expensive, and limited to conditions such as: additional sensing requirements, quasi-stationarity of ECG signals (arrhythmia constrains), and linearly separable artifacts. Furthermore, artifact mitigation techniques can even introduce distortion to filtered signals. With all that being said, SQA systems have become an attractive solution for selecting or segmenting data with acceptable quality.

This chapter presented and described the fundamentals, definitions, and a literature review, all related to SQA and SQI. The main takeaway of this chapter is the understanding of the presented fundamentals, definitions of the terms used in this thesis, and the fact that quality assessment metrics can be very unique and different from each other. Furthermore, it is important to have in mind that SQA systems serve many different purposes, in terms of output and rigorousness, depending on the application.

CHAPTER 3

WAVELETS, FILTER BANKS AND SCALOGRAMS APPLIED TO ECG SIGNALS

The main purpose of this chapter is providing a background regarding wavelets, filter banks and wavelet scalograms, along with explaining and defining how these mathematical tools can be used to represent "acceptable" and "unacceptable" signals. The following equations of this chapter assume that $n \in \mathbb{Z}$ and $t \in \mathbb{R}$. Moreover, all signals are assumed to be real numbers.

3.1 Wavelet transforms

Let $L^2(\mathbb{R})$ denote the space of functions with finite energy $\left(L^2(\mathbb{R}) = \left\{f(t) | \int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty\right\}\right)$. Let $\psi(t) \in L^2(\mathbb{R})$ be a fixed function, said to be a *wavelet* if its FT $\Psi(\omega)$ satisfies

$$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{|\omega|} \, d\omega < \infty. \tag{3.1}$$

The relationship (3.1) represents the *admissibility* condition (Olkkonen, 2011, p. 3). Dilated-translated wavelets are defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right). \tag{3.2}$$

Where $b \in \mathbb{R}$ is the translation parameter and $a \in \mathbb{R}^+$ ($a \neq 0$) is the dilation or scale parameter. The factor $a^{-\frac{1}{2}}$ is a normalization term such that the energy is the same for all scales (Olkkonen, 2011, p. 4). The *admissibility* condition implies that the *wavelet* must have no DC component (Olkkonen, 2011, p. 4). Hence,

$$\int_{-\infty}^{\infty} \psi(t) \, dt = \Psi(0) = 0. \tag{3.3}$$

3.1.1 The continuous wavelet transform

The *continuous wavelet transform* (CWT) of $x(t) \in L^2(\mathbb{R})$ (Olkkonen, 2011, p. 4) is defined as

$$W_x(a,b) = \langle x(t), \psi_{a,b}(t) \rangle = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)\, dt. \tag{3.4}$$

The so-called *resolution of the identity* provides the reconstruction of $x(t)$ (Olkkonen, 2011, p. 5), given as

$$x(t) = \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{W_x(a,b)\psi_{a,b}(t)}{a^2}\, da\, db. \tag{3.5}$$

3.1.2 The discrete wavelet transform

It is important to emphasize that, in the CWT, a and b are both continuous. Many authors, such as Addison (2005); Torrence & Compo (1998a); He (2016, p. 25), define the *discrete wavelet transform* (DWT) by discretizing both a and b. Other authors, such as (CINTRA, 2005, p. 88), call it *wavelet series*.

The most common form of the DWT employs the use of a dyadic grid (integer power of two for a and b) and orthonormal *wavelets* associated to orthonormal bases of $L^2(\mathbb{R})$ (Olkkonen, 2011, p. 5), exhibiting zero redundancy (unlike CWT) (Addison, 2005, p. R163). An example of a dyadic grid wavelet is:

$$\psi_{i,j}(t) = 2^{-i/2}\psi\left(2^{-i}t - j\right). \tag{3.6}$$

Where $i, j \in \mathbb{Z}$. According to Addison (2005, p. R163), discrete dyadic grid wavelets are usually chosen to be orthonormal. Hence, $\langle \psi_{j,k}(t), \psi_{l,m}(t) \rangle = \delta_{jk}\delta_{lm}$, for $j, k, l, m \in \mathbb{Z}$. Where δ_{jk} represents the Kronecker delta.

Therefore, for a real finite energy signal $x(t)$, using the dyadic grid, the DWT is written as:

$$T_{i,j} = \int_{-\infty}^{\infty} x(t)\psi_{i,j}(t)\, dt. \tag{3.7}$$

Where $T_{i,j}$ is known as the wavelet (or *detail*) coefficient (Olkkonen, 2011, p. R164). Hence, $x(t)$ can be represented (Brunton & Kutz, 2019, p. 77) as:

$$x(t) = \sum_{i,j} T_{i,j}\psi_{i,j}(t). \tag{3.8}$$

3.1.3 Example of a basis construction by scaling

Given a signal $x(t)$ that belongs to the space spanned by the basis $\{\phi_k(t)\}$, it is possible to represent $x(t)$ as a linear combination of the basis vectors (signals). Hence,

$$x(t) = \sum_{k} c_k\phi_k(t). \tag{3.9}$$

Where, in equation 3.9, c_k represents the coefficients of the projection of $x(t)$ into the space spanned by $\{\phi(t)\}$. Consider the following continuous function:

$$\theta(t) = \begin{cases} 1 & 0 \leq t < 1; \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

The scaled version of $\theta(t)$ is defined as

$$\theta_n(t) = \theta(2^n t). \tag{3.11}$$

The family $\{\theta_n(t)\}$ is not orthonormal, since $\langle\theta_0(t), \theta_1(t)\rangle \neq 0$. By applying the Gram-Schmidt process, it is possible to obtain a set of orthonormal basis from $\{\theta_n(t)\}$. Hence,

$$\begin{aligned}
\phi_0(t) &= \theta_0(t) = \phi(t) \\
\phi_1(t) &= \theta_1(t) - \frac{\langle\theta_1(t), \theta_0(t)\rangle}{\langle\theta_0(t), \theta_0(t)\rangle}\theta_0(t) \\
&= \begin{cases} \frac{1}{2} & 0 \leq t < \frac{1}{2}; \\ -\frac{1}{2} & \frac{1}{2} \leq t < 1; \\ 0 & \text{otherwise.} \end{cases} \\
&= \frac{\psi(t)}{2}.
\end{aligned} \tag{3.12}$$

The result of the described basis is very popular. $\psi(t)$ is known as the *Haar mother wavelet*, and $\phi(t)$ is its scaling function (or *father wavelet*). $\psi(t)$ is characterized by high frequencies, and $\phi(t)$ by low frequencies. If x(t) is given as

$$x(t) = \begin{cases} 2 & 0 \leq t < \frac{1}{2}; \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

Then,

$$x(t) = a\phi(t) + d\psi(t) = \phi(t) + \psi(t). \tag{3.14}$$

Where, in equation (3.14), $a = 1$ is known as *approximation* coefficient, and $d = 1$ is known as *detail* coefficient. Therefore, this simple example concludes how $x(t)$ is projected in the *Haar wavelet* basis, using the DWT.

3.1.4 Multiresolution analysis and the Mallat algorithm

It is possible to construct an orthonormal basis, by translating and scaling functions (Chun-lin, 2010, p. 12). Thus,

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \tag{3.15}$$

and

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k). \tag{3.16}$$

The parameter j refers to the visibility in frequency, and k to the position in time. In practice, j is referred as *resolution* (Chun-lin, 2010, p. 12) or *level*. Let us define the following subspaces:

$$V_j = \text{Span}\{\phi_{j,k}(t)\}, \tag{3.17}$$

and

$$W_j = \text{Span}\{\psi_{j,k}(t)\}. \tag{3.18}$$

Where its requirements (Gonzalez & Woods, 2006, p. 481-482) are enumerated as follows:

1. The scaling function is orthogonal to its integer translates.
2. $V_j \subset V_{j+1}$.
3. Only the null space is common to all V_j.
4. Any function can be represented with arbitrary precision.

Notice, from requirement 2 and the Haar scaling function, that:

$$\begin{aligned}\phi(t) &= \phi_{0,0}(t) \\ &= \frac{1}{\sqrt{2}}\phi_{1,0}(t) + \frac{1}{\sqrt{2}}\phi_{1,1}(t) \\ &= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t-1).\end{aligned} \tag{3.19}$$

In the context of dyadic grid wavelets, requirement 2 translates to equation (3.20), known as the *refinement equation* (CINTRA, 2005, p. 89). Thus,

$$\phi(t) = \sum_k h_0(k)\sqrt{2}\phi(2t-k). \tag{3.20}$$

Such equation has a unique solution for $h_0(n)$ (CINTRA, 2005, p. 89). For the Haar wavelets, $h_0(n) = \{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\}$ (Chun-lin, 2010, p. 13). Notice that $h_0(n)$ composes the elements of an FIR filter, known as *smoothing filter* (CINTRA, 2005, p. 89).

Analogous to the *refinement equation*, another fundamental recursive equation states that:

$$\psi(t) = \sum_k h_1(k)\sqrt{2}\phi(2t-k). \tag{3.21}$$

Where $h_1(n)$ is known as *detail filter* (CINTRA, 2005, p. 90). For the Haar wavelets, $h_1(n) = \{\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\}$ (Chun-lin, 2010, p. 13). The relation between $h_0(n)$ and $h_1(n)$, known as the quadrature mirrored filter (QMF) relation (Ariño, Morettin & Vidakovic, 2004, p. 5), is given as

$$h_1(n) = (-1)^n h_0(1-n). \tag{3.22}$$

The motivation of equation (3.22) will be discussed in Section 3.2.4. Moving forward, according to Gonzalez & Woods (2008, p. 483),

$$V_{j+1} = V_j \oplus W_j. \tag{3.23}$$

Once the infinite wavelet sets are united (Gonzalez & Woods, 2008, p. 484),

$$\begin{aligned} L^2(\mathbb{R}) &= V_0 \oplus W_0 \oplus W_1 \oplus \ldots \\ &= \ldots W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \ldots \\ &= V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \ldots \end{aligned} \tag{3.24}$$

Where $\oplus$ denotes the union of spaces. This concludes that the $L^2(\mathbb{R})$ space is a linear combination of subspaces formed by the *mother wavelet* and *father wavelet* functions.

3.1.5 The DWT for discrete input signals

Let $l^2(\mathbb{Z})$ denote the space of functions with finite energy $\left(l^2(\mathbb{Z}) = \left\{\sum_{-\infty}^{\infty} |f(n)|^2 < \infty\right\}\right)$. In practical applications, the DWT is implemented to receive discrete signals as inputs. In particular, Jerônimo da Silva Júnior (2008, p. 31) names such implementation as the discrete time wavelet series (DTWS).

As it was demonstrated in Section 3.1.3, a signal $x(t)$ could be decomposed in a *scaling* function summed with a *wavelet* function. Using the dyadic grids shown in equations (3.15) and (3.16), and sampling the *wavelet* and *scaling* functions, according to Chun-lin (2010, p. 12) and Gonzalez & Woods (2008, p. 489), a discrete signal $x(n)$ in $l^2(\mathbb{Z})$ can be decomposed as:

$$x(n) = \frac{1}{\sqrt{M}} \sum_k a_{j_0,k} \phi_{j_0,k}(n) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k d_{j,k} \psi_{j,k}(n). \tag{3.25}$$

Where $k, j \in \mathbb{Z}$; $M = 2^J$; $n = 0, 1, 2, ..., M - 1$. Equation (3.25) is known as the *inverse* DWT (Gonzalez & Woods, 2008, p. 489). The approximation ($a_{j_0,k}$) and detail ($d_{j,k}$) coefficients are obtained by applying the *forward* DWT to $x(n)$ (Gonzalez & Woods, 2008, p. 489), according to

$$a_{j_0,k} = \frac{1}{\sqrt{M}} \sum_n x(n)\phi_{j_0,k}(n), \tag{3.26}$$

and

$$d_{j,k} = \frac{1}{\sqrt{M}} \sum_n x(n)\psi_{j_0,k}(n). \tag{3.27}$$

3.2 Filter banks and FWT

This section aims to present how the DWT is implemented through the use of filter banks. Such implementation is known as fast wavelet transform (FWT).

3.2.1 Intro

Figure 3.1 illustrates an example of a 2 channel filter bank. From Figure 3.1, the block where the input is $u(n)$ and the output is u_d is known as a *compressor*. Its input-output relationship is given as:

$$u_d(n) = u(2n) = (\downarrow 2)u(n). \tag{3.28}$$

The block where the input is $u_d(n)$ and the output is $u_s(n)$ is known as *expander*. Where

$$u_s(n) = (\uparrow 2)u_d(n) = \sum_{k=-\infty}^{\infty} u_d(n)\delta(n - 2k) = \begin{cases} u(n) & \text{if } n \text{ is even;} \\ 0 & \text{otherwise.} \end{cases} \tag{3.29}$$

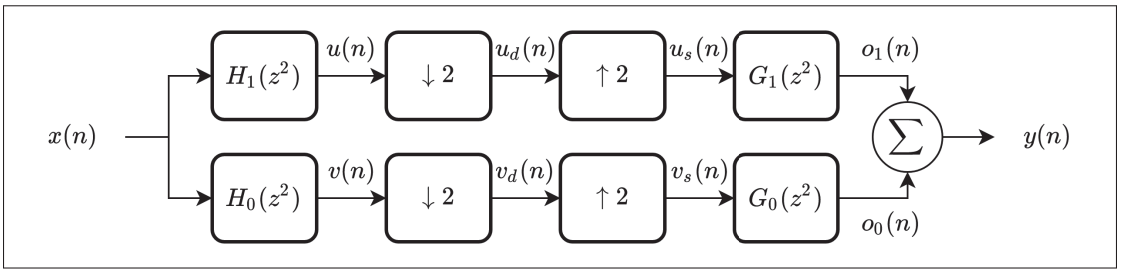


Figure 3.1 Example of a 2 channel filter bank

The Z-transform of $u_d(n)$ and $u_s(n)$, in terms of $u(n)$, are given by:

$$U_d(z) = \frac{U(z^{\frac{1}{2}}) + U(-z^{\frac{1}{2}})}{2}, \tag{3.30}$$

and

$$U_s(z) = \frac{U(z) + U(-z)}{2}. \tag{3.31}$$

The so called the *noble identity*, is known as

$$\begin{aligned} U_d(z) &= \frac{H_1(z^2)X(z) + H_1(-z^2)X(-z)}{2} \\ &= \frac{X(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}})}{2} H_1(z) \\ &= X_d(z)H_1(z). \end{aligned} \tag{3.32}$$

Therefore, the block diagram, illustrated by Figure 3.1, is reformulated to a more efficient implementation, as shown in Figure 3.2. This is very important to avoid unnecessary multiplications.

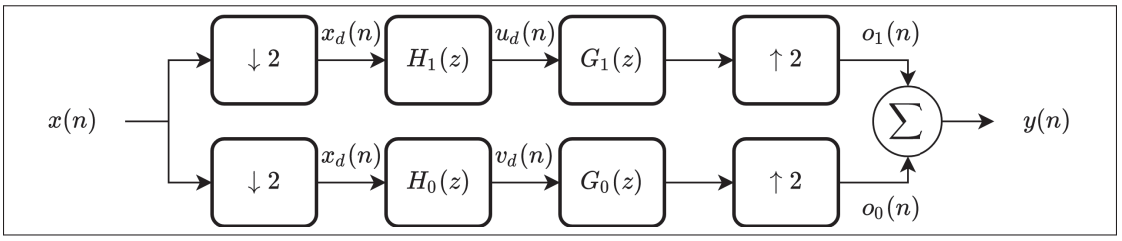


Figure 3.2 Example of a 2 channel filter bank, after applying the *noble identity* to Figure 3.1

3.2.2 Perfect reconstruction

Figure 3.3 illustrates an example of a 2 channel filter bank. The process of decomposing $x(n)$ into the coefficients $d_{1,k}$ and $a_{1,k}$ is called *analysis*. The process of recovering $\hat{x}(n)$, given the coefficients $d_{1,k}$ and $a_{1,k}$, is called *synthesis*.

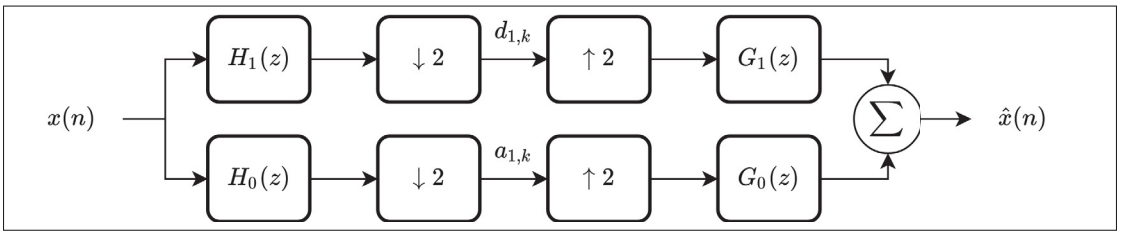


Figure 3.3 Example of a 2 channel filter bank

From Figure 3.3, the Z-transform of $\hat{x}(n)$ is given as:

$$\hat{X}(z) = \frac{1}{2}\left[X(z)H_1(z) + X(-z)H_1(-z)\right]G_1(z) + \frac{1}{2}\left[X(z)H_0(z) + X(-z)H_0(-z)\right]G_0(z) \tag{3.33}$$

$$= \frac{1}{2}X(z)\left[H_0(z)G_0(z) + H_1(z)G_1(z)\right] + \frac{1}{2}X(-z)\left[H_0(-z)G_0(z) + H_1(-z)G_1(z)\right]. \tag{3.34}$$

It is desired that $\hat{x}(n) = x(n-d)$. Hence,

$$\begin{cases} H_0(z)G_0(z) + H_1(z)G_1(z) = 2z^{-d}; \\ H_0(-z)G_0(z) + H_1(-z)G_1(z) = 0. \end{cases} \tag{3.35}$$

Where, in the system of equations (3.35), the former is the *no distortion* condition and the latter is the *no aliasing* condition. Therefore,

$$\mathbf{H_m}(z)\mathbf{G_m}(z) = \begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix} \begin{bmatrix} G_0(z) \\ G_1(-z) \end{bmatrix} = \begin{bmatrix} 2z^{-d} \\ 0 \end{bmatrix}. \tag{3.36}$$

The so-called *modulation matrix* is represented by $\mathbf{H_m}(z)$. Its determinant $\Delta(z) = -\Delta(-z)$. By calculating the inverse of $\mathbf{H_m}(z)$,

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \frac{2z^{-d}}{\Delta(z)} \begin{bmatrix} H_1(z) \\ -H_0(-z) \end{bmatrix}. \tag{3.37}$$

From equation equation (3.37), several wavelets can be derived. From Fig. 3.3, as an example, the *forward* Haar DWT is obtained by using:

$$H_0(z) = 1 + z^{-1}, \tag{3.38}$$

and

$$H_1(z) = 1 - z^{-1}. \tag{3.39}$$

The solution for the *inverse* Haar DWT is obtained by respecting equation (3.37). Therefore,

$$\begin{aligned}\Delta(z) &= (1+z^{-1})(1+z^{-1}) - (1-z^{-1})(1-z^{-1}) \\ &= 4z^{-1}.\end{aligned} \tag{3.40}$$

Thus,

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \frac{z^{-d+1}}{2} \begin{bmatrix} H_1(z) \\ -H_0(-z) \end{bmatrix}. \tag{3.41}$$

If $d = 1$ the solution is causal. Hence,

$$G_0(z) = \frac{1}{2} H_1(-z), \tag{3.42}$$

and

$$G_1(z) = -\frac{1}{2} H_0(-z). \tag{3.43}$$

3.2.3 The analysis tree based FWT filter bank structure

According to Jerônimo da Silva Júnior (2008, p. 31), the DTWS can be implemented using a tree based filter bank structure called *logarithm tree*. For a decomposition of J levels, the lowest frequency signal vector ($a_{J,k}$) consists on the space of lowest temporal resolution. While the other components, when added, form a space of higher resolution (multiresolution) (Jerônimo da Silva Júnior, 2008, p. 31).

According to Jerônimo da Silva Júnior (2008, p. 32), the coefficients in the scale j, for J levels, can be written as:

$$d_{j,k} = d_j(k) = \sum_n x(n) h_1^{(j)}(2^j k - n) = (\downarrow 2^j) \left(x(n) * h_1^{(j)}(n) \right), \tag{3.44}$$

for $j = 1, 2, ..., J$, and

$$a_{J,k} = a_J(k) = \sum_n x(n) h_0^{(J)}(2^J k - n) = (\downarrow 2^J) \left(x(n) * h_0^{(J)}(n) \right). \tag{3.45}$$

In the Z-domain,

$$H_0^{(j)}(z) = \prod_{i=0}^{j-1} H_0(z^{2^i}) = H_0^{(j-1)}(z) H_0(z^{2^{j-1}}), \tag{3.46}$$

and

$$H_1^{(j)}(z) = H_1(z^{2^{j-1}}) \prod_{i=0}^{j-2} H_0(z^{2^i}) = H_0^{(j-1)}(z) H_1(z^{2^{j-1}}). \tag{3.47}$$

An example of a 3 level ($J = 3$) analysis filter bank is illustrated by Figure 3.4.

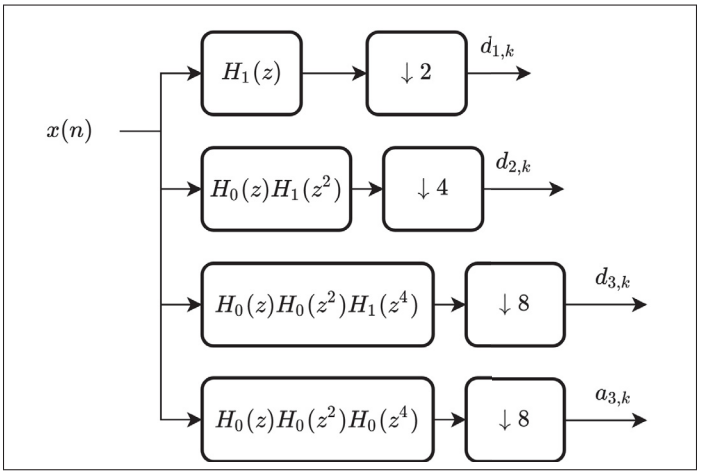


Figure 3.4 Example of a 3 level analysis filter bank

It is important to notice that each level has a different amount of samples. As an example, if $x(n)$ has 512 samples, $d_{1,k}$ has 256 samples, $d_{2,k}$ has 128 samples, $d_{3,k}$ has 64 samples and $a_{3,k}$ has 64 samples. By applying the *noble identity* to the block diagram illustrated by Figure 3.4, the implementation becomes recursive, as shown in Figure 3.5.

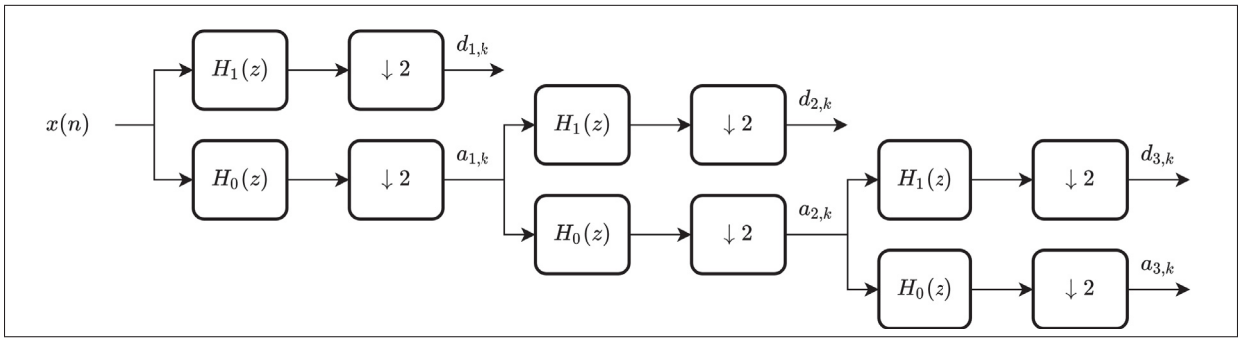


Figure 3.5 Example of a 3 level recursive analysis filter bank

The implementation illustrated by Figure 3.5 is known as fast wavelet transform (FWT), and it is much more efficient then the implementation illustrated by Figure 3.4, since it avoids unnecessary multiplications. By applying the noble identity to Figure 3.5, the computational cost is further reduced by a factor of 2. The result of this is illustrated by Figure 3.6.

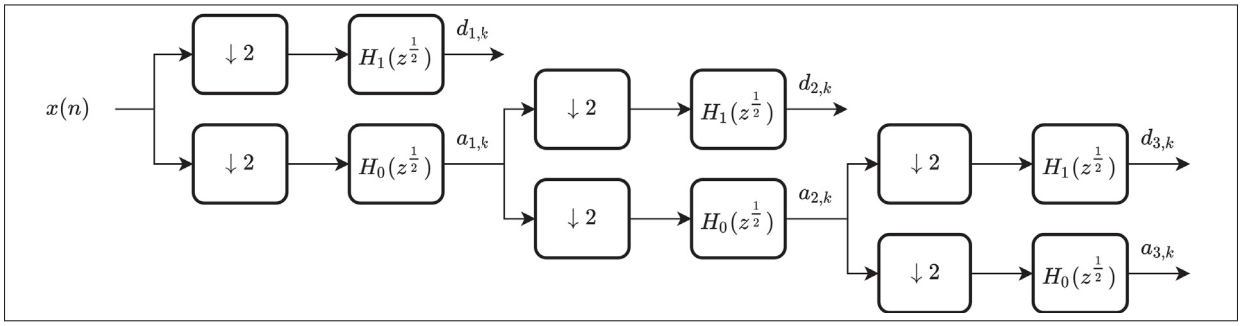


Figure 3.6 Noble identity applied to Figure 3.5

The computational complexity of the FWT is known to be $O(N)$ operations (CINTRA, 2005, p. 91). While many fast Fourier transform (FFT) algorithms, like Cooley–Tukey, have a computational complexity of $O(N \log N)$ operations.

As a conclusion, choosing to use the FWT instead of using an FFT algorithm is a win-win! The FWT provides temporal resolution, reduces the computational cost, is designed to be used with non-stationary signals, and has no windowing leakage issues like the short-time Fourier transform (STFT) or Gabor transform.

3.2.4 Example illustrating the Daubechies wavelets

The so-called *no aliasing* condition, presented by equation (3.35), is satisfied when

$$\begin{cases} G_0(z) = H_1(-z); \\ G_1(z) = -H_0(-z). \end{cases} \tag{3.48}$$

Let us define $P_0(z)$ as

$$P_0(z) = G_0(z)H_0(z). \tag{3.49}$$

The so-called *no distortion* condition can be rewritten as

$$P_0(z) - P_0(-z) = 2z^{-d}. \tag{3.50}$$

Equation (3.50) is known as the *halfband filter*. Since $G_0(z)$ and $H_0(z)$ are both low pass filters, $P_0(z)$ is also a low pass filter. Therefore, the design of the filter bank structure, presented by Figure 3.5, simply consists on the following steps:

1. Design $P_0(z)$ as a low pass filter that satisfies equation (3.50).
2. Factorize $P_0(z)$ in order to obtain $G_0(z)$ and $H_0(z)$.
3. Obtain $G_1(z)$ and $H_1(z)$ by using the condition (3.48).

Say $P_0(z) = o+pz^{-1}+qz^{-2}+rz^{-3}+sz^{-4}$. Then, satisfying equation (3.50), $2pz^{-1}+2rz^{-3} = 2z^{-d}$. This concludes that there are no constrains on even coefficients, while one of the odd coefficients should be equal to 1 and the others should be equal to 0. Let us define $P(z) = G_0(z)H_0(z)z^d$. Hence, equation (3.50) can be rewritten as

$$P(z) + P(-z) = 2. \tag{3.51}$$

The Daubechies solution states that

$$P(z) = (1 + z^{-1})^{2m}Q(z)z^l. \tag{3.52}$$

The Daubechies 4-tap filter solution is found by setting $m = 2$, and $Q(z) = a + bz^{-1} + az^{-2}$. Thus,

$$P(z) = \left[a + (4a + b)z^{-1} + (7a + 4b)z^{-2} + (8a + 6b)z^{-3} + ... + az^{-6}\right] z^l. \tag{3.53}$$

For $l = 3$, and satisfying the condition of the odd coefficients,

$$\begin{cases} 8a + 6b = 1; \\ 4a + b = 0. \end{cases} \tag{3.54}$$

Hence, $a = -1/16$, $b = 1/4$ and $Q(z) = \frac{1}{16}(-1 + 4z^{-1} - z^{-2}) = (1 - c_1z^{-1})(1 - c_2z^{-1})$. The roots of $Q(z)$ are: $c_1 = 2 + \sqrt{3}$, and $c_2 = 2 - \sqrt{3}$.

It is known that there are many ways of factoring $P_0(z)$. One choice is $G_0(z) = (1 + z^{-1})^2$ and $H_0(z) = (1 + z^{-1})^2\frac{1}{16}(-1 + 4z^{-1} - z^{-2})$. Another is $G_0(z) = (1 + z^{-1})^3$ and $H_0(z) = (1 + z^{-1})\frac{1}{16}(-1 + 4z^{-1} - z^{-2})$. Both choices are interesting because both $G_0(z)$ and $H_0(z)$ are

symmetric, meaning they have a linear phase response. If $G_0(z) = (1 + z^{-1})^2(2 - \sqrt{3} - z^{-1})$ and $H_0(z) = (1 + z^{-1})^2(\frac{1}{2-\sqrt{3}} - z^{-1})$, the filter bank has no linear phase response, but it has the property of orthogonality. For deriving the Daubechies 4-tap filter coefficients a solution must be found, such that, the frequency response of $H_0(z)$ and $H_1(z)$ are exactly mirrored around $\pi/2$. Therefore, the filter bank has the property of orthogonality, and both $H_0(z)$ and $H_1(z)$ are known to be QMFs. Such filter is designed by respecting equation (3.22).

As mentioned before, by using the nobel identity, the implementation from Figure 3.5 can be further improved by placing decimators before filters in every recursive filtering block. Same as shown in Figure 3.6, and in the work of Freitas, Inocêncio, Lins, Alves & Benedetti (2019, p. 18). If $H_0(z^{\frac{1}{2}}) = o + pz^{-1} + qz^{-2} + rz^{-3}$ and $H_1(z^{\frac{1}{2}}) = r - qz^{-1} + pz^{-2} - oz^{-3}$, the order 4 Daubechies (Db4) filter coeficients are $o = \frac{1-\sqrt{3}}{4\sqrt{2}}$, $p = \frac{3-\sqrt{3}}{4\sqrt{2}}$, $q = \frac{3+\sqrt{3}}{4\sqrt{2}}$ and $r = \frac{1+\sqrt{3}}{4\sqrt{2}}$ (Freitas *et al.*, 2019, p. 18). Hence, the Db4 analysis filters use 4 taps, instead of 8. Similarly by using the implementation shown in Figure 3.6, the number of taps for an efficient implementation of the Db6 analysis filters can be 6, instead of 12.

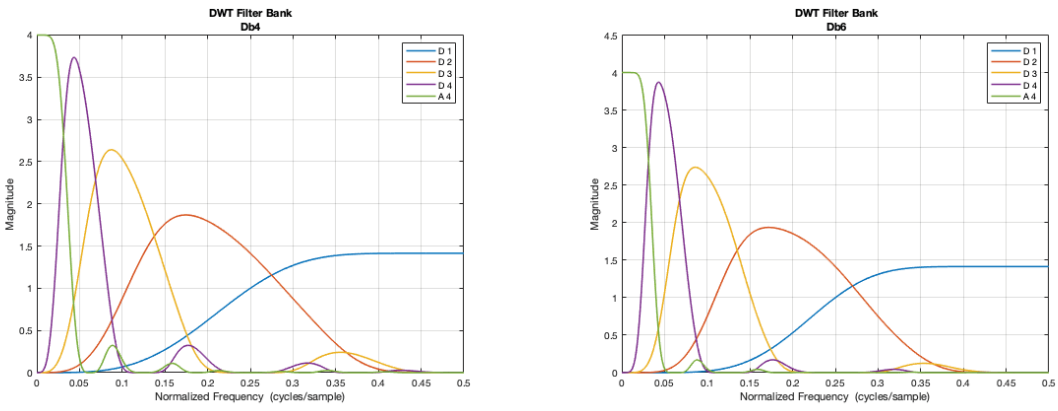
3.2.5 Frequency response of the FWT filter banks

As presented in equations (3.44) and (3.45), for notation simplicity purposes, $d_{j,k} = d_j(k)$ and $a_{j,k} = a_j(k)$. According to He (2016, p. 208), the frequency band ranges contained in the signal components $d_j(k)$, and $a_j(k)$ are described as:

$$\begin{cases} d_j(k) : [2^{-(j+1)} f_s, 2^{-j} f_s], & \text{for } 1 \leq j \leq J; \\ a_j(k) : [0, 2^{-(j+1)} f_s], & \text{for } 1 \leq j \leq J. \end{cases} \tag{3.55}$$

In the application presented by this thesis, f_s =256 Hz, and $J = 9$. Therefore, $a_9(k)$ has its highest frequency limited to 0.25 Hz, and lowest frequency limited to 0 Hz. Here both Db4 and Db6 wavelet transforms are chosen to represent ECG signals, because both wavelets have shown good results regarding the application of ECG beat classification in previous works (Böck, 2015,

p. 36). For instance, Böck (2015, p. 36) mentions the works of Can Ye, Kumar & Coimbra (2012) and Saha & Ghorai (2015) as interesting references of ECG beat classifier systems using, respectively, the Db4 and Db6 wavelet transforms. In many of the experiments performed by Chapter 5, other wavelets such as Db2, Db5 and Db8 were tested, other then Db4 and Db6, but none of them have demonstrated an outstanding performance, in comparison to the use of Db4 and Db6. Moreover, Wang (2018, p. 16) justifies the use of Db6 wavelets to generate features, that describe the quality of ECG signals, due to the fact that previous works have demonstrated that Db6 wavelets have a similar structure compared to the QRS complex of an ECG waveform (Rai, Trivedi & Shukla, 2013; Saritha, Sukanya & Murthy, 2008). Hence, a strong reason why this thesis focuses on the choice of the Db6 wavelet transform, in the application presented by Chapter 5. For illustration purposes, the frequency response of the Db4 and Db6 FWT filter banks, for a 4 level decomposition, are both shown in figures 3.7a and 3.7b.



(a) Frequency response of a 4 level decomposition filter bank that uses the Db4 *mother wavelet*

(b) Frequency response of a 4 level decomposition filter bank that uses the Db6 *mother wavelet*

Figure 3.7 FWT filter banks frequency response

Notice that despite both filter banks having the same cut-off frequency, the Db6 FWT filter bank has less residual lobes compared to the Db4 FWT filter bank. Thus, the space spanned by Db4 wavelet should be different compared to the space spanned by the Db6 wavelet. This will be clarified in Section 3.3.1, where examples of the Db4 and Db6 scalograms are illustrated.

3.3 Wavelet scalograms

According to the work of Ariño *et al.* (2004, p. 37), a wavelet scalogram is the DWT analogue of the periodogram from the spectral analysis of time series. The wavelet scalogram (Ariño *et al.*, 2004, p. 41), for a signal $s(n)$ of length M, decomposed by orthogonal wavelets, as a vector formed by the E_j elements, can be define as

$$E_j = \begin{cases} \sum_{k=0}^{M/2^j-1} |d_j(k)|^2, & \text{for } 1 \leq j \leq J; \\ \sum_{k=0}^{M/2^J-1} |a_J(k)|^2, & \text{for } j = J+1; \\ 0, & \text{otherwise.} \end{cases} \tag{3.56}$$

In the work of Ariño *et al.* (2004, p. 40), $M = 2^J$ and J is the maximum level of decomposition. It is important to adjust M such that is divisible by 2^j, for $j = 1, 2, ..., J$. According to Rosso, Blanco, Yordanova, Kolev, Figliola, Schürmann & Başar (2001, p. 67), the *relative wavelet energy* is defined as

$$p_j = \frac{E_j}{E_{tot}}. \tag{3.57}$$

Where

$$E_{tot} = \sum_{n=0}^{M-1} |s(n)|^2. \tag{3.58}$$

Using the Parseval's theorem, and the fact the wavelets are orthogonal, according to Rosso *et al.* (2001, p. 67):

$$E_{tot} = \sum_{j=1}^{J+1} E_j. \tag{3.59}$$

Hence,

$$\sum_{j=1}^{J+1} p_j = 1. \tag{3.60}$$

Therefore, the *normalized scalogram*, or *wavelet probability mass function* (wPMF) is defined in this thesis as:

$$f(j) = \begin{cases} p_j, & \text{for } 1 \leq j \leq J+1; \\ 0, & \text{otherwise.} \end{cases} \tag{3.61}$$

Notice that $j = 1$ represents the highest frequency, and $j = 10$ represents the lowest frequency. In a vector form, the wPMF from equation (3.61) can be represented as:

$$\mathbf{f} = \Big[f(1), f(2), ..., f(J+1) \Big]^T. \tag{3.62}$$

As a conclusion of this section, it is possible to define two approaches for representing a signal: a *frequency analysis* and a *time-frequency analysis*. Both approaches are explained in the following sections.

3.3.1 Frequency analysis

The *frequency analysis* simply consists on representing a zero mean signal $s(n)$ by its wPMF. In such case M is set to a number that is divisible by 2 and $s(n)$ is decomposed till it reaches the level J, where M is no longer divisible by 2^{J+1}. In the case where $M = 2^J$, the last level of decomposition should be simply J.

In the application presented by this thesis, $M = 2560$ (10 seconds of data, for a sampling frequency $f_s = 256$ Hz). Hence, $J = 9$, because M is not divisible by 2^{10}. Examples of how the

wPMF represent acceptable and unacceptable signals can be visualized in figures 3.8, 3.10, and 3.12, using the Db4 FWT, and figures 3.9, 3.11, and 3.13, using the Db6 FWT.

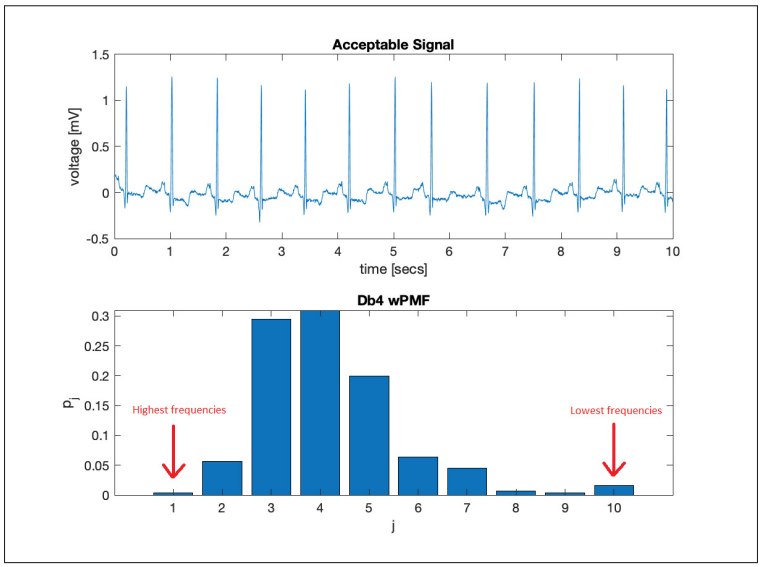


Figure 3.8 Example of an acceptable ECG signal, represented by the wPMF using the Db4 *mother wavelet*

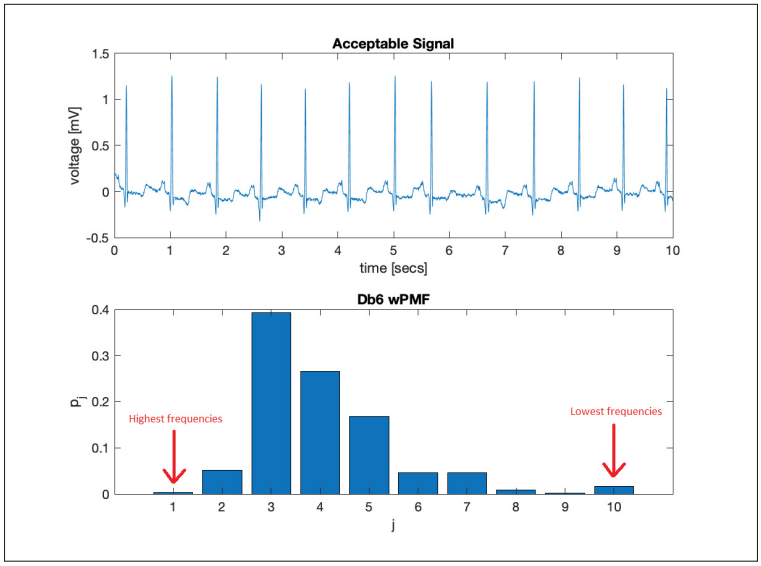


Figure 3.9 Example of an acceptable ECG signal, represented by the wPMF using the Db6 *mother wavelet*

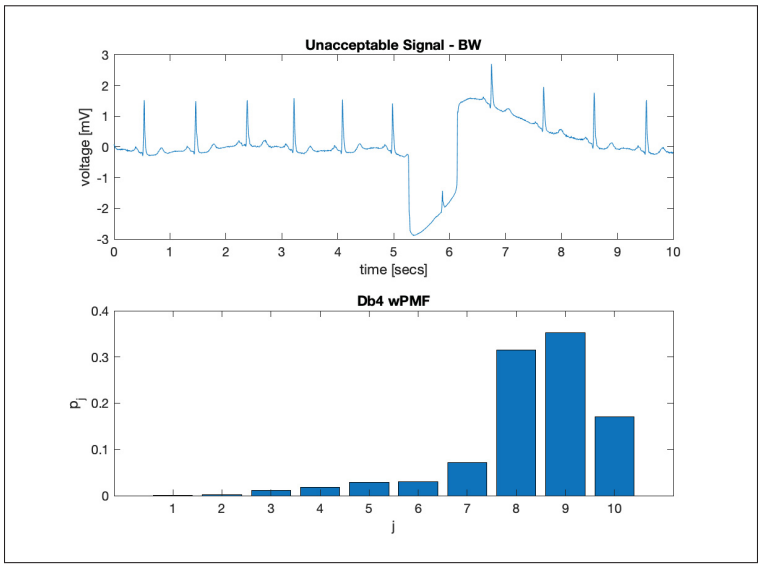


Figure 3.10 Example of an unacceptable ECG signal with BW interference, represented by the wPMF using the Db4 *mother wavelet*

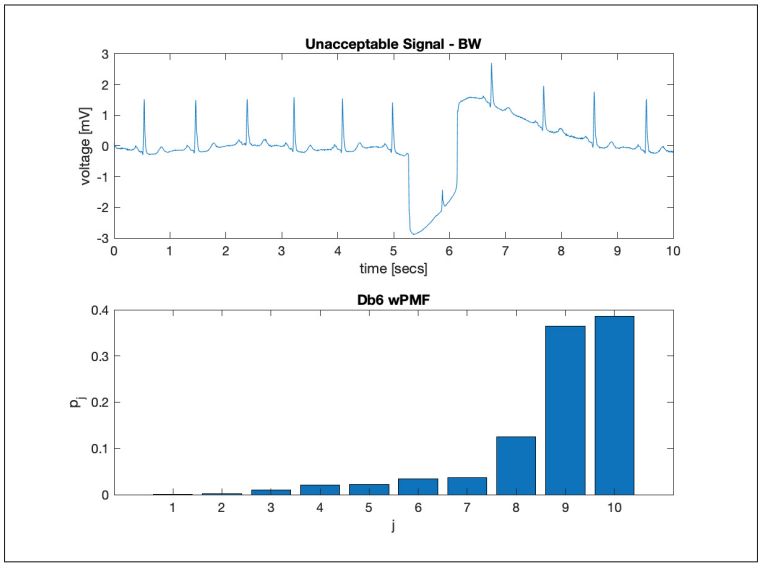


Figure 3.11 Example of an unacceptable ECG signal with BW interference, represented by the wPMF using the Db6 *mother wavelet*

Notice from figures 3.8 and 3.9 that, for the acceptable example, most of the energy is located at the high frequency range (around levels 3, 4, and 5), where the QRS complex is located.

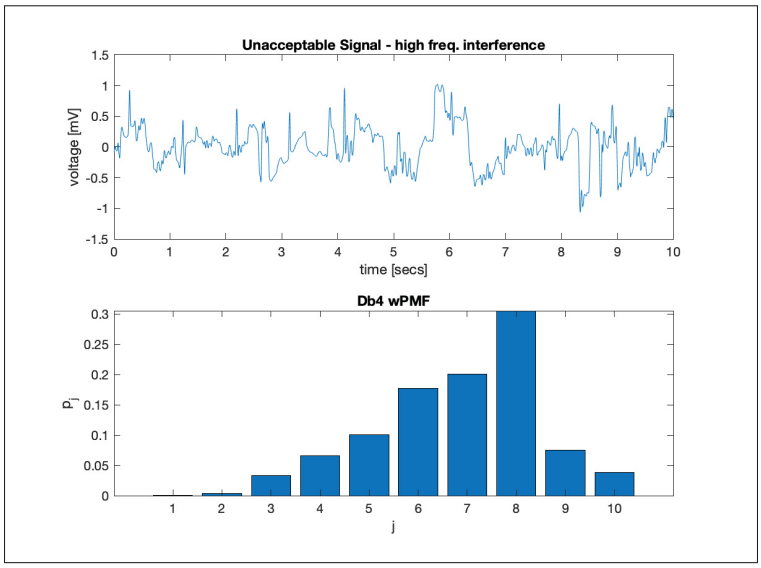


Figure 3.12 Example of an unacceptable ECG signal with high frequency interference, represented by the wPMF using the Db4 *mother wavelet*

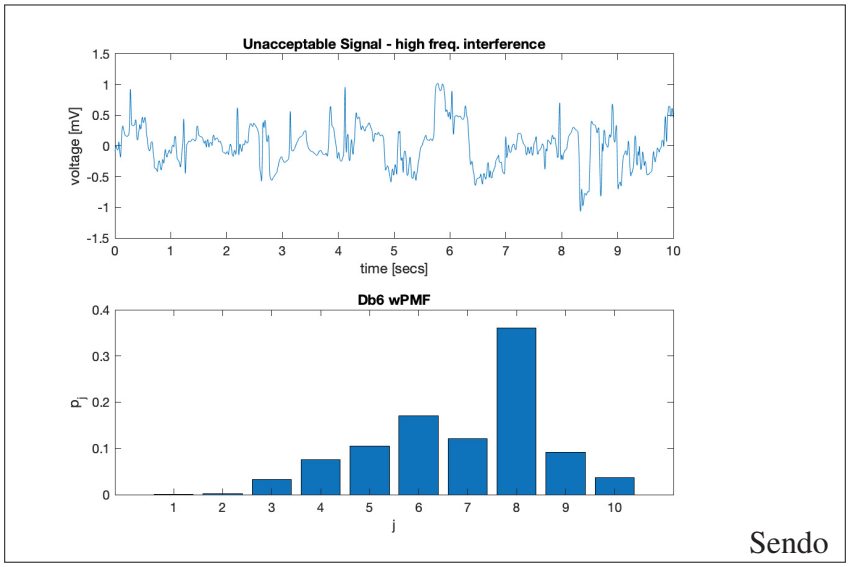


Figure 3.13 Example of an unacceptable ECG signal with high frequency interference, represented by the wPMF using the Db6 *mother wavelet*

In the case of low frequency interference (figures 3.10 and 3.11), most of the energy is concentrated at the lowest frequencies ($d_8(k)$, $d_9(k)$ or $a_9(k)$ coefficients). In other words, the largest probability events are either p_8, p_9 or p_{10}.

In the case of high frequency interference (figures 3.12 and 3.13), the energy is more uniformly distributed in every frequency. Hence, given a signal that has an approximately uniform wPMF, its entropy should be considered high. Such concept of entropy will be clarified in Chapter 5.

It is also important to mention that the basSQI can be calculated using the wPMF. The basSQI characterizes BW interference as a signal of frequency range lower then 1 Hz. The wPMF, in this thesis, performs a 9 level decomposition, using $f_s = 256$ Hz, and 10 seconds of data. It can detect BW interference in the range of [0 Hz, 0.25 Hz], [0.25 Hz, 0.5 Hz], [0.5 Hz, 1 Hz]. Such ranges are represented by the events p_{10}, p_9, and p_8. Therefore, $basSQI \approx p_8 + p_9 + p_{10}$. Notice that the wPMF, compared to the basSQI, is capable of providing more frequency resolution, with the computational cost of $O(N)$.

Finally, another advantage of the wPMF, is that it can be characterized by statistical moments (SMs), analogous to the analysis of Behar *et al.* (2013, p. 3). This concludes that the wPMF is a powerful tool that can be potentially used to discriminate acceptable and unacceptable signals.

3.3.2 Time-frequency analysis

It is important to know that the higher the highest level of decomposition is, the better will be the resolution in frequency of the wavelet *analysis*. The *time-frequency analysis* consists on dividing $s(n)$ into non-overlapping blocks, or signal windows, of size $N = 2^J$. Where J is the largest integer, such that, M is divisible by N. Therefore, providing the highest resolution in frequency. The i_{th} block of $s(n)$, represented by $s^{(i)}(n)$, can be written as:

$$s^{(i)}(n) = \sum_{m=M(i-1)/B}^{Mi/B-1} s(m)\delta(n-m). \tag{3.63}$$

Where $\delta(n)$ is a *unit sample* function, and B is the number of blocks. Subsequently, a grayscale image is formed with the wPMFs of each block. In such case, it is important for M to be divisible

by 2. The i_{th} block of $s(n)$, $(s^{(i)}(n))$, has a wPMF $(f^{(i)}(j))$, that is arranged to a column vector, described as

$$\mathbf{f}^{(i)} = \left[f^{(i)}(1), f^{(i)}(2), ..., f^{(i)}(J+1) \right]^T. \tag{3.64}$$

Where $j = 1, 2, ..., J+1$. Finally, the grayscale image matrix has the following structure:

$$\mathbf{I_m} = \left[\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, ..., \mathbf{f}^{(B)} \right]. \tag{3.65}$$

Hence,

$$E_j^{(i)} = \begin{cases} \sum_{k=M(i-1)/B2^j}^{Mi/(B2^j)-1} |d_j(k)|^2, & \text{for } 1 \leq j \leq J; \\ \sum_{k=M(i-1)/B2^J}^{Mi/(B2^J)-1} |a_J(k)|^2, & \text{for } j = J+1; \\ 0, & \text{otherwise.} \end{cases} \tag{3.66}$$

Furthermore,

$$E_{tot}^{(i)} = \sum_{n=M(i-1)/B}^{Mi/B-1} |s(n)|^2 = \sum_{j=1}^{J+1} E_j^{(i)}, \tag{3.67}$$

$$p_j^{(i)} = \frac{E_j^{(i)}}{E_{tot}^{(i)}}, \tag{3.68}$$

and,

$$f^{(i)}(j) = \begin{cases} p_j^{(i)}, & \text{for } 1 \leq j \leq J+1; \\ 0, & \text{otherwise.} \end{cases} \tag{3.69}$$

A block diagram summarizing how the scalogram image is obtained by using the *time-frequency analysis* procedure, is illustrated by Figure 3.14.

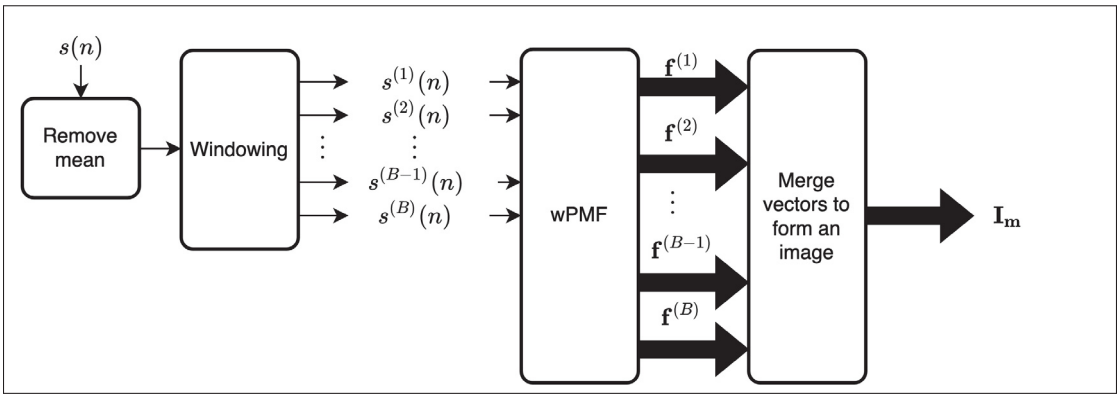


Figure 3.14 Block diagram illustrating how the scalogram image is obtained from a signal $s(n)$

As an example, in the application presented by this thesis, $M = 2560$ (10 s of data, for a sampling frequency f_s =256 Hz), so $B = 5$, because $J = 9$ provides the highest resolution in frequency, and M is not divisible by 2^{10}. In this case, each block $s^{(i)}(n)$ has 512 samples. For illustration purposes, figures 3.15, 3.16 and 3.17 show examples of 10×5 images, formed by using a Db6 *time-frequency analysis*, along with its respective signals.

The advantage of such analysis is the time dependency. Notice the acceptable signal example has the majority of energy concentrated in the high frequencies, the signal interfered by BW has its energy concentrated in the low frequencies, and the high frequency interfered signal has no pattern of energy concentration.

3.4 Limitations

It is clear that, for the *frequency analysis*, the wPMF had no time localization. Hence, the *frequency analysis* is limited to the fact that it cannot detect where exactly the noise is present, for a zero mean signal $s(n)$.

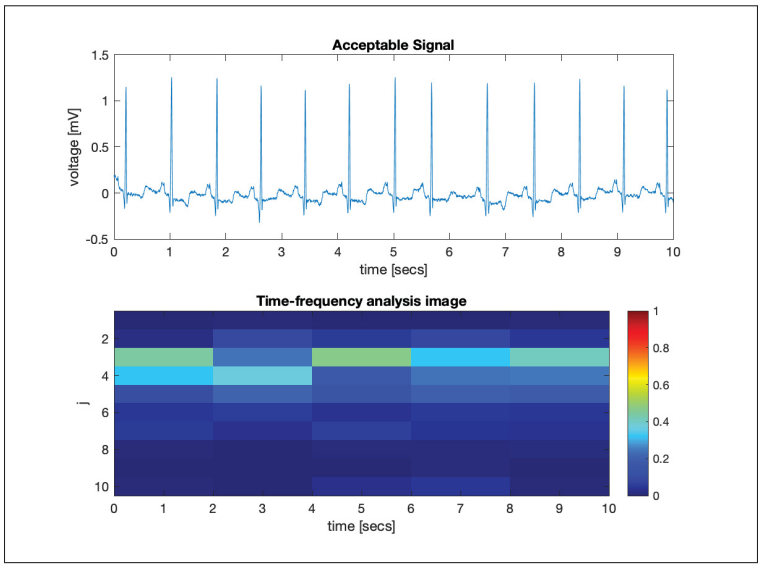


Figure 3.15 Example of an acceptable ECG signal, represented by the *time-frequency analysis* using the Db6 *mother wavelet*

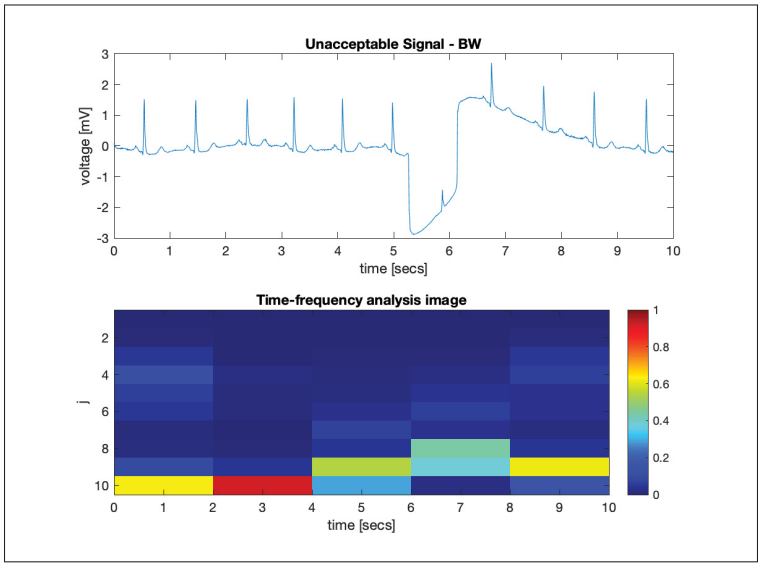


Figure 3.16 Example of an unacceptable ECG signal with BW interference, represented by the *time-frequency analysis* using the Db6 *mother wavelet*

Furthermore, notice that Figure 3.16 actually shows a poor localization of noise. That is evident when inspecting the scalogram during the intervals of 0-4 s, where the peaks of energy should ideally be located around levels 3, 4 and 5, similar to Figure 3.15. In Figure 3.16, it is still

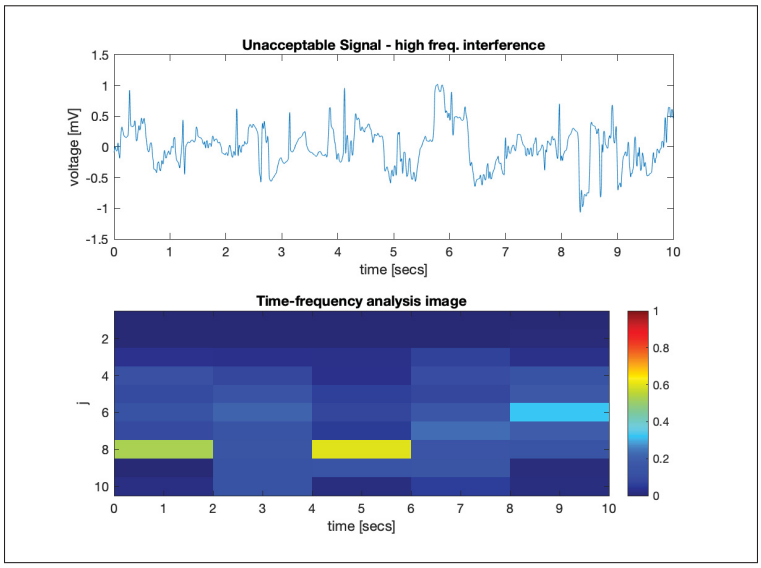


Figure 3.17 Example of an unacceptable ECG signal with high frequency interference, represented by the *time-frequency analysis* using the Db6 *mother wavelet*

possible to see some peaks in level 4, during the intervals of 0-2 s and 2-4 s. However, these peaks are being masked, or are biased, by the influence of the DC component that is present in the windows of intervals 0-2 s and 2-4 s.

According to Figure 3.14, the entire signal $s(n)$ is demeaned. Still, that does not mean that each window of 2 s duration is demeaned. Due to that, any little DC displacement in the window can cause such masking, or biasing, issues. By inspecting figures 3.15, 3.16 and 3.17, it is clear that the noise presence is evident, but time-localization is poor. Hence, the time-localization information is not very much useful when these DC components are present, since they mask, or bias, the time dependency representation.

To resolve this issue, each window is demeaned separately, and the wPDF of each windows is estimated, using the same procedure described in Section 3.3.1. Subsequently, each wPDF is merged to forme an image, same as described in equation (3.65). The result of demeaning each individual signal window, prior to the estimation of each wPDF, is illustrated in figures 3.18, 3.19, 3.20.

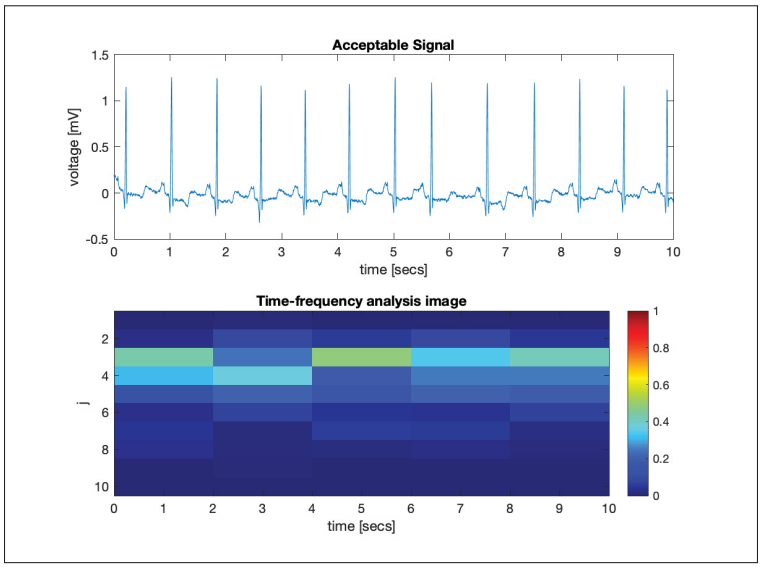


Figure 3.18 Example of an acceptable ECG signal, represented by the *time-frequency analysis*, using the Db6 *mother wavelet* and demeaning each analyzed signal window individually

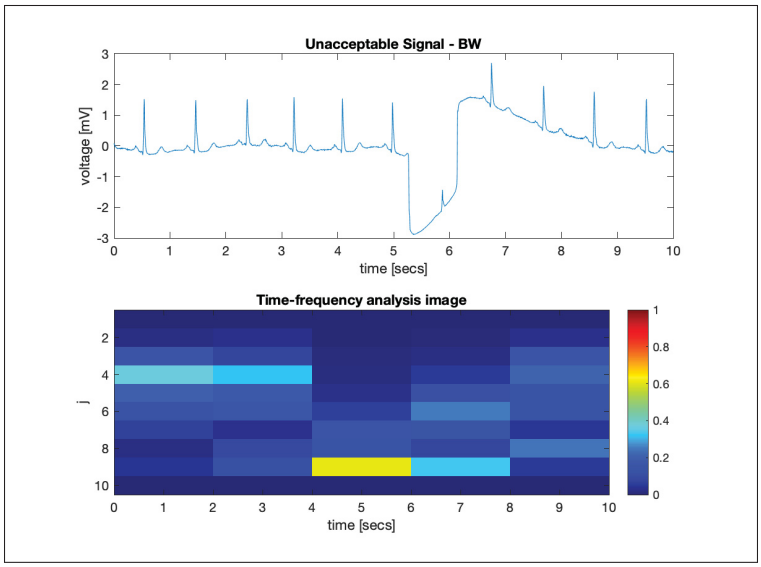


Figure 3.19 Example of an unacceptable ECG signal with BW interference, represented by the *time-frequency analysis*, using the Db6 *mother wavelet* and demeaning each analyzed signal window individually

Notice that, in Figure 3.19, the clean segment located in 0-4 s is now more alike to an expected behaviour from an acceptable ECG segment, where maximum energy peaks are located around

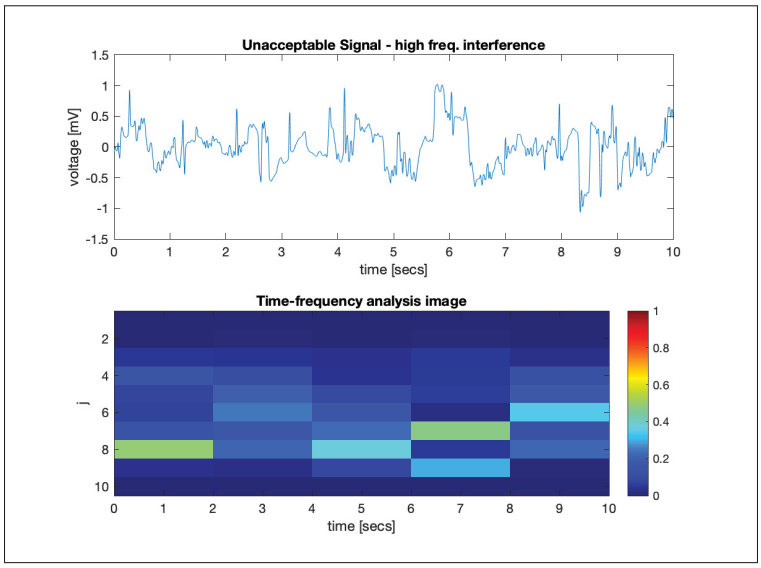


Figure 3.20 Example of an unacceptable ECG signal with high frequency interference, represented by the *time-frequency analysis*, using the Db6 *mother wavelet* and demeaning each analyzed signal window individually

levels 3, 4 and 5. Furthermore, the energy of BW interference, observed at segments located in 4-8 s, matches the expected frequency intervals of such interference (around 0.1 - 1 Hz, or levels 8, 9 and 10).

This whole issue related to residual DC components masking, or biasing, the time localization representation was only noticed in the very end of the research. Since the results were already very satisfying using the *frequency analysis* representation, the classifiers presented in Chapter 5 were not tested using this other approach, where residual DC components are removed from each signal window. Hence, re-training the classifiers using the *time-frequency analysis* without residual DC components in each windowed signal is left for future works.

3.5 Conclusion

This chapter presented the main necessary tools that are used to compactly represent ECG signals. The presented representations are shown to be efective to determine noise presence, but not always effective in determining noise location, due to the presence of residual DC

components. Still, Section 3.4 shows how to overcome such biasing issue. Chapter 5 presents the actual SQA application, using both the *frequency* and the *time-frequency analysis* wavelet scalograms as features to train different ANN classification architectures.

CHAPTER 4

DATA SET PREPARATION

The current and following chapter present the design methodology and implementation of an SQA system, developed to be applied as a *channel selection* or *lead selection* algorithm for the SIG.ECG system. In particular, the current chapter aims to explain the preparation methodology of the data-sets used to train and test the SQA system. The first data-set is prepared by using several different databases from PhysioNet (Goldberger *et al.*, 2000), and is meant to be used for training and validating a classifier that detects if a single-lead ECG signal is acceptable or not. For simplicity, the first data-set is refereed as *PhysioNet data-set*. The second data-set, provided by SIG.NUM, is used as a blind test that validates the performance of the SQA classifier applied to the SIG.ECG system. The second data-set is refereed as *SIG.test data-set*

The *PhysioNet data-set* contains a total of 6519 examples of 10 s labelled as *acceptable* and 6526 examples of 10 s labelled as *unacceptable*, based on several criteria and annotations from specialists. All examples are resampled to 256 Hz, since the accepted range of the diagnostic ECG is often quoted to be from 0.05 Hz (for ST analysis) to 40 Hz or 100 Hz (Clifford *et al.*, 2006, p. 57). If the sampling frequency (f_s) is 256 Hz, the maximum frequency that that can be observed is 128 Hz, according to the Nyquist–Shannon sampling theorem. Hence, a butterworth low-pass filter, of order 10, is chosen to filter the frequencies higher than 100 Hz. Moreover, it is important to mention that the choice of a base 2 sampling frequency impacts the performance of fast DSP algorithms.

The following sections focus on describing the databases used to generate the *PhysioNet data-set*, the preprocessing of signals and the labelling protocol, inspired by the work of Clifford *et al.* (2012).

4.1 Preprocessing of the PhysioNet data-set

Prior to the construction of the PhysioNet data-set, it is necessary to preprocess all signals (ECG and noise) by simply removing the mean, or direct current (DC) component of all 10 s segments. This is very important, especially when generating *unacceptable* signals with the noise stress test database, since most of the signal power is concentrated at the DC component.

4.2 PhysioNet data-set

The following list presents the databases that were used to build the *PhysioNet data-set*. The main important details to be understood from each database are explained in dedicated subsections.

- MIT-BIH Arrhythmia database.
- MIT-BIH Noise Stress Test database.
- PhysioNet / CinC Challenge 2011 database.
- PhysioNet / CinC Challenge 2017 database.

4.2.1 MIT-BIH Arrhythmia database

This database (Moody & Mark, 2001) includes 48 complete two leads ECG records with reference annotations from specialists. The records have a diagnostic bandwidth of 0.1−100 Hz with 12-bit resolution and $f_s = 360$ Hz. Mainly, the first channel is the modified lead II (MII) and the second channel is the modified precordial lead V1 (MV1), occasionally V2 or V5, and in one instance V4. From this database, patients 100 to 106 are analyzed. The possible annotated symbols for the analyzed patients are summarized in Table 4.1.

The analysis of each patient consisted on segmenting the data in pieces of 10 s and labelling each segmented example. The labelling protocol, or list of metrics is described as follows. Where, $SQA = 0$ stands for “unacceptable”, $SQA = 1$ stands for “acceptable”, and $SQA = 2$ stands

Table 4.1 Possible annotated symbols of analyzed segments from patients 100 to 106

Symbol	Annotation description
"+"	Rhythm change
"N"	Normal beat
"A"	Atrial premature beat
"V"	Premature ventricular contraction
"~"	Change in signal quality
"|"	Isolated QRS-like artifact
"Q"	Unclassifiable beat
"/"	Paced beat
"f"	Fusion of paced and normal beat

for "unknown". It is important to keep in mind that only examples with $SQA < 2$ are used for training, testing and validating.

1. If all QRS complexes of both leads are clear, all beats are classified by the annotators and there is no significant existence of noise, despite the presence of "~", $SQA = 1$.

2. If the annotation "Q" is present, meaning at least one of the beats is unclassifiable, $SQA = 0$.

3. If the indication of quality change ("~") is present, and any of the leads have at least one QRS complex that is not clear, even if all beats are classifiable, $SQA = 2$.

4. If both leads have visible QRS complexes, but one of them contain some significant amount of noise (e.g., BW), $SQA = 2$.

For illustration purposes, examples of labelled segments, using metrics 3 and 4, are shown in figures 4.1, and 4.2.

Notice that both figures 4.1, and 4.2 contain 2 ECG signals measured in parallel. Therefore, even if one signal is doubtfully "acceptable" (e.g. lead MII from Figure 4.1), for sake of simplicity, both ECG segments are labelled as "unknown" ($SQA = 2$).

It is important to mention the specialists annotated the data based on the simultaneous visualization of both leads. Hence, the criteria that says if both leads are acceptable is the presence of clear

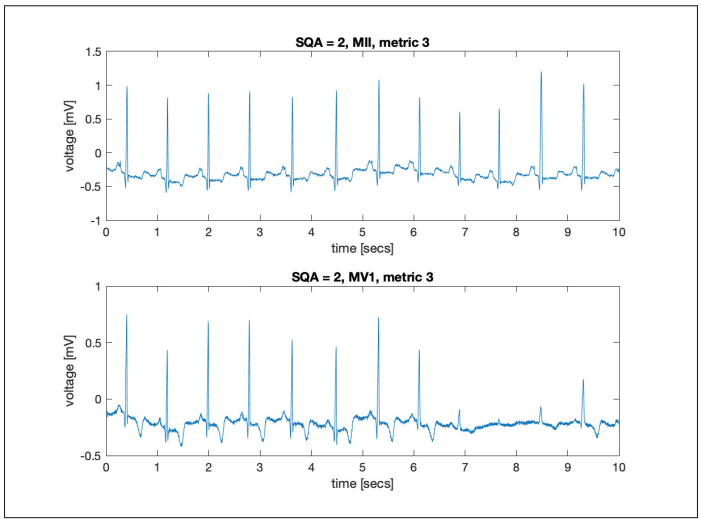


Figure 4.1 Example of signals labelled as “unknown” quality, using metric 3

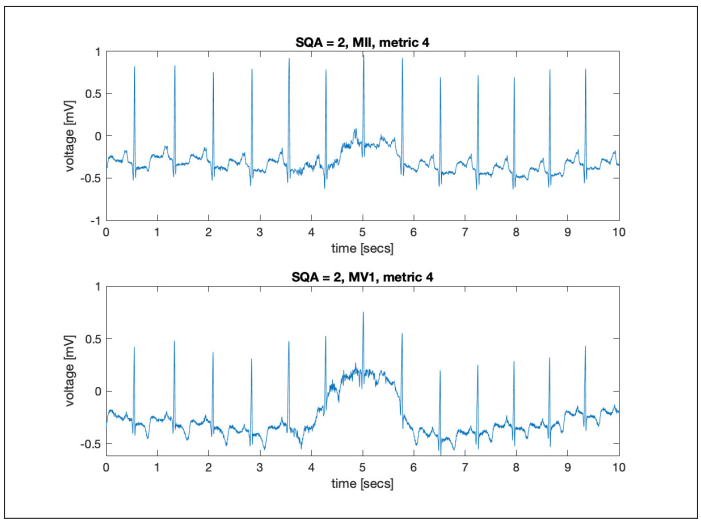


Figure 4.2 Example of signals labelled as “unknown” quality, using metric 4

QRS complexes and no significant noise present in both leads. If one lead seems to be corrupted, and all annotations are present, the signal is labelled as “uknown” quality. The total number of signals labelled as “acceptable” are 1880, and 52 are labelled as “unacceptable”.

4.2.2 PhysioNet / CinC Challenge 2011 database

The training data-set of this challenge consists on 998 sets of 12 lead ECG signals (a total of 11976 single-lead ECG signals examples). Each set is classified, by specialists, as “acceptable” or “unacceptable”. Meaning that the SQA annotation applies only for the set of 12 leads, and not to its individual 12 signals contained in the set. Therefore, another set of metrics was designed in order to verify if some single lead is considered to be “acceptable” or not. Such metrics are listed bellow.

1. If it is possible to clearly visualize all QRS complexes of the example and the signal does not contain any significant amount of noise, $SQA = 1$.

2. If it is not possible to clearly visualize all QRS complexes of the example, but the signal does not contain any significant amount of noise, $SQA = 2$.

3. If it is not possible to clearly visualize all QRS complexes of the example and the signal contains a significant amount of noise, $SQA = 0$.

4.2.3 MIT-BIH Noise Stress Test database

A similar strategy, to the one adopted in the work of Clifford *et al.* (2012), was implemented using this database (Moody, Muldrow & Mark, 1984). The strategy simply consisted on creating a set of “unacceptable” examples, using “acceptable” examples and pure noise data from the MIT-BIH Noise stress test database (NSTD).

The set of clean signals (“acceptable”) come from the PhysioNet / CinC Challenge 2011 database. Furthermore, an artificial signal, generated by using a code from the work of Mcsharry & Cifford (2004), is also used to create “unacceptable” examples. Most of the default parameters, suggested by McSharry, Clifford, Tarassenko & Smith (2003, p. 291), are considered to generate the synthetic lead (SYN). Table 4.2 summarizes the non-default parameters used for generating SYN. The parameter “sfecg” is the ECG sampling frequency, “hrmean” is the mean heart rate, “hrstd” is the heart rate standard deviation and “lfhfratio” is the low frequency / high frequency ratio.

Table 4.2 Parameters used for the generation of a clean signal

Signal name	sfecg	hrmean	hrstd	lfhfratio
Signal 1	250	60	1	0.5

Based on the values of Table 4.2, the synthetic lead is illustrated in Figure 4.3.

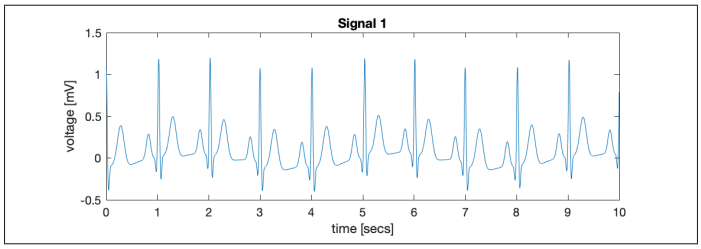


Figure 4.3 Illustration of Signal 1 (lead SYN)

For constructing the "unacceptable" signals, noise is added to "Signal 1" according to:

$$a = \sqrt{\exp\left(-\frac{S \ln 10}{10}\right)\left(\frac{\mathbf{x}^T\mathbf{x}}{\mathbf{v}^T\mathbf{v}}\right)}, \tag{4.1}$$

and

$$\mathbf{y} = \mathbf{x} + a\mathbf{v}. \tag{4.2}$$

Where, $\mathbf{x}$ represents the clean signal vector, $\mathbf{v}$ represents the pure noise. Both signals are demeaned (see Section 4.1), a is the noise gain and S is the SNR parameter that dictates the magnitude of a. In the work of Clifford *et al.* (2012), $S = -6$. The same value is used to generate "unacceptable" examples. The application presented by this thesis only used the *em* and *ma* records from the NSTD. It is important to mention that, in the context of the NSTD, record *em* contains electrode motion artifact (MA) with a significant amount of baseline wander (BW) and muscle noise (EMG), while record *ma* contains mainly muscle noise (EMG).

Finally, in summary, 720 “unacceptable” examples are generated using lead SYN and 1600 “unacceptable” signals are generated by randomly selecting and merging “acceptable” signals from the PhysioNet / CinC Challenge 2011 database and pure noise signals (*em* and *ma* records) from the NSTD, according to equation (4.1) and (4.2). Hence, a total of 2320 “unacceptable” examples.

4.2.4 PhysioNet/CinC Challenge 2017 database

The 2017 PhysioNet / CinC Challenge aimed to encourage the development of algorithms to classify, from a single short lead I ECG recording (between 30 s and 60 s in length), whether it shows normal sinus rhythm, atrial fibrillation, an alternative rhythm, or if it is too noisy to be classified. It is important that the recording is measured on the surface of the thumb fingers (Left arm (+), Rigth arm (-)), collected using the AliveCor device.

The training data-set of the challenge contains a set of 279 recordings labelled as noisy by specialists. It is important to mention that the “noisy” label reefers to the whole recording (30 s to 60 s). If the recording is divided into segments of 10 s, each segment should not necessarily be labelled as “noisy”.

The protocol for labelling a signal as “unacceptable” simply consisted on visualizing if there was any unclear QRS complex ($SQA = 0$), otherwise, $SQA = 2$. Most of the data is completely noisy, and not hard to label. A total of 616 10 s examples are labelled as “unacceptable”.

4.3 Summary of the PhysioNet data-set

In summary, the *PhysioNet data-set*, created by merging several different databases, contains information from a total of 856 patients. 6519 examples are labelled as “acceptable” and 6526 are labelled as “unacceptable”. Hence, a total of 13045 balanced examples.

Figures 4.4, and 4.5, illustrate the histogram of leads used to create the *PhysioNet data-set*. Figure 4.4 shows that the “acceptable” examples from the data-set are not only very diverse, but

also approximately balanced for leads I to V6. For “unacceptable” examples, since each noisy data is very unique, examples do not need to be balanced in terms of leads (see Figure 4.5).

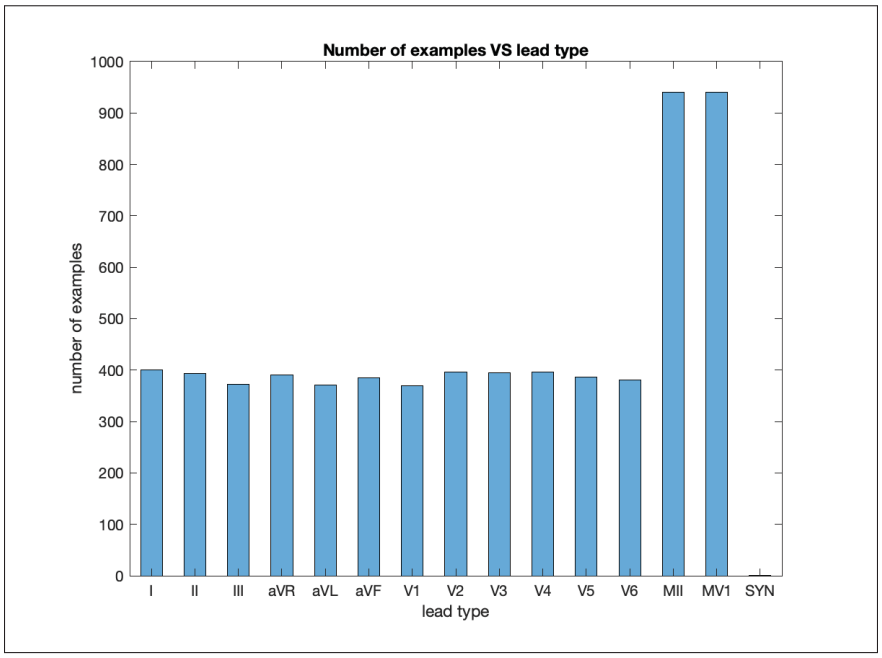


Figure 4.4 *PhysioNet data-set* histogram illustration of leads labelled *acceptable*

4.4 Preprocessing of the SIG.test data-set

Prior to the labelling of the *SIG.test data-set*, it is necessary to preprossess the data, in order to avoid issues with *linearly separable* interference. If the majority of the interference is contained in a deterministic spectrum range, and it is linearly separable from the ECG signal, filtering is not an issue. Therefore, the resulting filtered signal is expected to have high quality.

Types of interference that are generally easy to remove are: PLI and, many times, BW. This is achieved by removing the moving average of the signal (BW filtering) and using notch filters (PLI filtering).

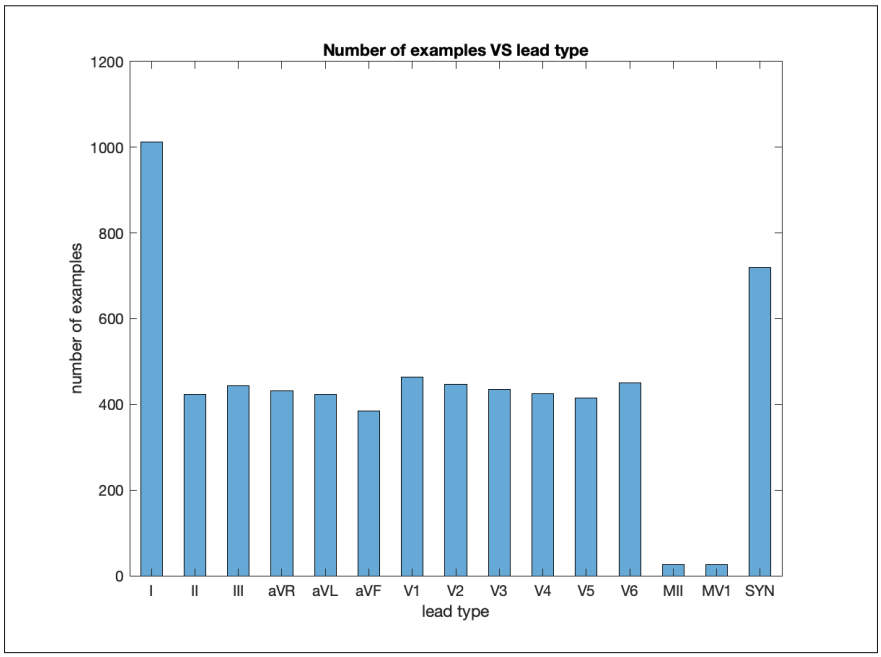


Figure 4.5 *PhysioNet data-set* histogram illustration of leads labelled *unacceptable*

4.4.1 BW removal

The correct way of removing BW is: knowing the specifications of the AFE, and then correctly removing the deterministic spectrum of the interference by using a high-pass filter. Usually this is implemented by using a moving average filter. The cutoff frequency of the moving average filter is dictated by its length.

The cutoff frequency of the SIG.ECG AFE is around 200−250 mHz. If no interference is present, the clean cECG signal should be contained in a range of 250 mHz to 100 Hz. Hence, a well designed moving average filter should remove frequencies below 250 mHz. The approximation of the cutoff frequency f_c, for a long moving average filter of size N, is given as:

$$f_c \approx \frac{0.3362 f_s}{N}. \tag{4.3}$$

The proof of (4.3) is demonstrated in Appendix I. For $N = 800$, the cutoff frequency is approximately 420 mHz. This is reasonable considering the nature of the SIG.ECG AFE (cutoff frequency of approximately 250 mHz). If $x(n)$ is a raw signal ecg signal, and $y(n)$ is the filtered signal, the BW is removed by converting the low-pass moving average filter to a high-pass filter. Hence,

$$y(n) = (\delta(n - d) - h(n)) * x(n). \tag{4.4}$$

Where, "$*$" represents the convolution operation and $\delta(n - d)$ is a *unit sample* function delayed by d samples (the number of samples delayed by the moving average filter).

4.4.2 PLI removal

The PLI can be easily removed by using a comb filter, or cascaded notch filters with the notch frequencies centered in 60 and 120 Hz or 50 and 100 Hz, depending on the local power line. Generally comb filters are more efficient, but require very specific sampling frequency. Cascaded notch filters, with quality factor of 35, were chosen.

4.5 The SIG.test data-set

Given the preprocessing procedures, the *SIG.test data-set* is labelled. The labelling procedure consisted on checking the visibility of all QRS complexes. If all QRS complexes are clear, $SQA = 1$. If any sort of interference is clearly detected, $SQA = 0$. If there is no clear interference, and QRS complexes are also not clear, then $SQA = 2$. $SQA = 2$ signals are discarded.

The *SIG.test data-set*, after labelling, contains information from a total of 99 patients. Details about the diagnosis of these patients can be found in Appendix III. 538 examples were labelled as "acceptable" and 13 were labelled as "unacceptable". Figure 4.6 shows a histogram of the labelled *SIG.test data-set*, according to each lead type, for "acceptable" signals. Similarly, Figure 4.7 illustrates a similar plot, for "unacceptable" signals.

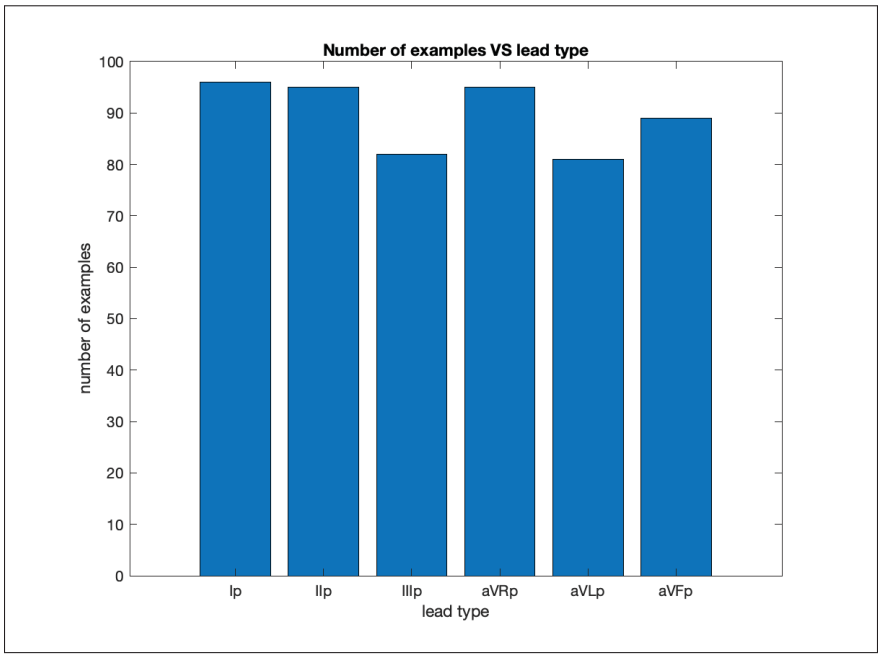


Figure 4.6 *SIG.test data-set* histogram illustration of leads labelled *acceptable*

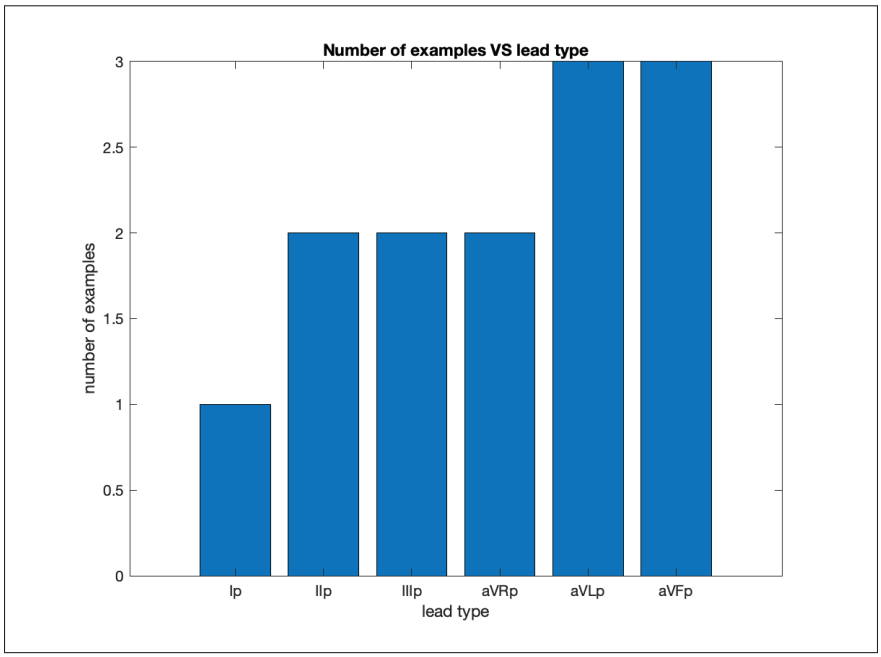


Figure 4.7 *SIG.test data-set* histogram illustration of leads labelled *unacceptable*, prior to the generation of additional *unacceptable* signals

4.5.1 Generation of unacceptable signals

Figure 4.6 shows that the "acceptable" examples from the data-set are very diverse and approximately balanced for all leads. On the other hand, it is not the same case for "unacceptable" examples (see Figure 4.5). Hence, 551 "unacceptable" examples are generated by randomly selecting and merging signals from the labelled *SIG.test data-set* and pure noise signals (*em* and *ma* records) from the NSTD, according to equation (4.1) and (4.2).

4.5.2 Summary of the SIG.test data-set

In summary, the final number of examples contained in the *SIG.test data-set* is 1102. The data-set contains information from a total of 99 patients. 538 examples are labelled as "acceptable", and 564 examples are labelled as "unacceptable".

4.6 Conclusion

This chapter presented and described the formulation of each data-set used for designing a machine learning based SQA classifier system. The main takeaway is that the data preparation presented by this current chapter seems to be reliable for training and testing the SQA system.

CHAPTER 5

A MACHINE LEARNING BASED SQA CLASSIFIER SYSTEM

This chapter aims to present the results and the methodology behind the design of the single-lead acceptability detector or automated SQA classifier system. The used approach consisted on training an ANN classifier that receives a feature vector as an input. The following sections consist on describing feature extraction, feature selection (FS), the separability analysis, present the ANN topologies that were used to design the classifier, the results that evaluate the performance of the automated SQA classifier system, and a computational complexity analysis.

5.1 Experimental procedures framework

The general experimental procedure framework, for the *frequency analysis* methodology, consisted on the following:

1. **Feature extraction:** Features are extracted by using a *frequency analysis* methodology.
2. **Feature selection:** The best features are selected by a wrapper methodology, based on an SVM classifier.
3. **Separability analysis:** The separability of the training and validating data-set is analyzed by using principal component analysis (PCA).
4. **Classifier training and testing experimental procedure:** The classifiers are trained and tested in order to be analyzed.

The feature extraction, features selection and separability analysis procedures are illustrated in the flowchart presented by Figure 5.1. Each procedure of such figure is clarified through sections 5.3 to 5.5.

Figure 5.2 illustrates the classifier training and testing framework. Each procedure of such figure is clarified in Section 5.6.

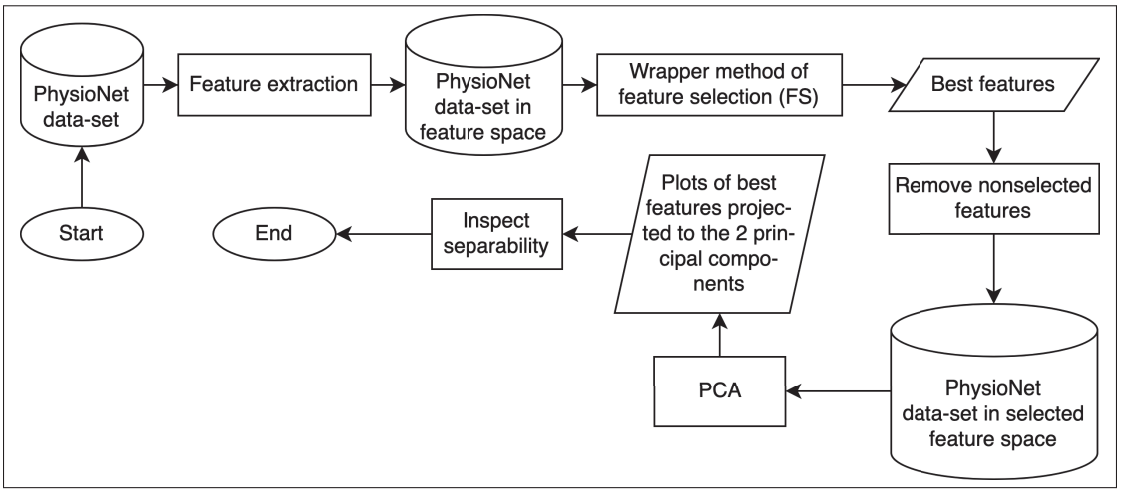


Figure 5.1 Flowchart illustrating the feature extraction, feature selection and separability analysis procedures for the *frequency analysis* methodology

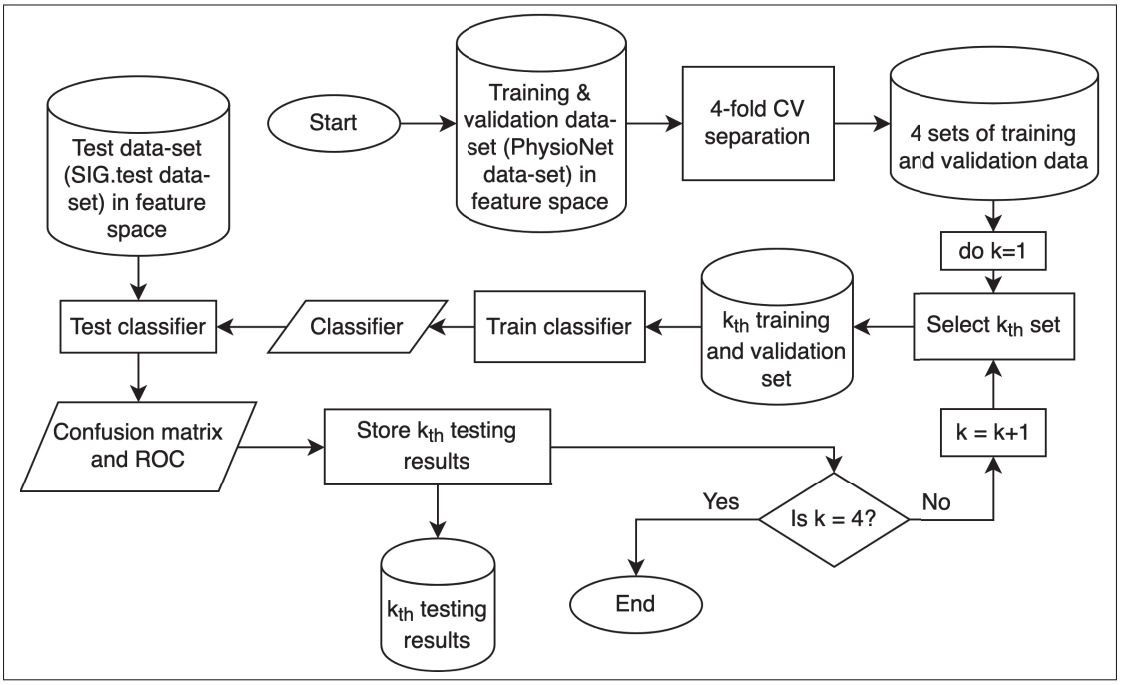


Figure 5.2 Flowchart illustrating the classifier training and testing experimental procedure for both the *frequency analysis* and the *time-frequency analysis* methodologies

For the *time-frequency analysis*, no feature selection is necessary. Furthermore, no separability analysis is performed. Hence, Figure 5.1 is only valid for the *frequency analysis* methodology.

It is important to mention that the framework, illustrated in Figure 5.2, is also used for classifiers with features acquired using the *time-frequency analysis* methodology.

5.2 System framework

After both the classifier and important features are defined, the system framework is very simple. It simply consists on the following:

1. **Necessary feature extraction:** Where all necessary features are extracted in accordance to the classifier architecture.

2. **Classification:** The classifier simply outputs its result.

Figure 5.3 illustrates the system framework for both the *frequency analysis* and the *time-frequency analysis* methodologies.

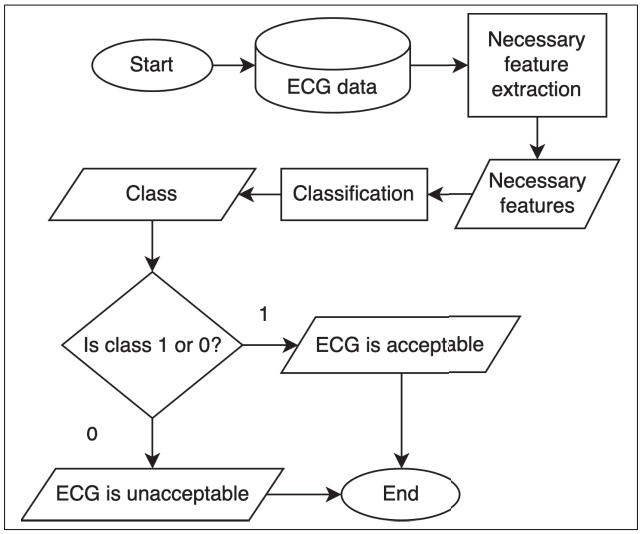


Figure 5.3 Flowchart illustrating the system framework for both the *frequency analysis* and the *time-frequency analysis* methodologies

In specific, for the *frequency analysis*, the necessary features are defined by the results presented by Section 5.7, where features are further pruned, in order to simplify the classifier system. The simplification of the system aims to generalize its performance and avoid overfitting or

underfitting issues due to the *curse of dimensionality* (this issue is further explained in Section 5.4).

5.3 Feature extraction

This current section aims to clarify the feature extraction procedure. If necessary, revisit Chapter 3 for clarification of definitions.

5.3.1 Frequency analysis feature representation

Given the definitions described by Section 3.3.1, a signal $s(n)$ can be represented by a wPMF, using the *frequency analysis*.

5.3.1.1 Statistical Moments

The discrete statistical moments (SMs) are here defined as:

$$\mu_i = \begin{cases} \mathbb{E}\{0.1j\} = \sum_{j=1}^{J+1} 0.1j f(j) = \sum_{j=1}^{J+1} 0.1j p_j, & \text{for } i = 1; \\ \mathbb{E}\left\{(0.1j - \mu_1)^i\right\} = \sum_{j=1}^{J+1} (0.1j - \mu_1)^i f(j) = \sum_{j=1}^{J+1} (0.1j - \mu_1)^i p_j, & \text{for } i > 1. \end{cases} \tag{5.1}$$

Where, $\mathbb{E}\{\cdot\}$ stands for the expected value operator and $j = 1, 2, ..., J + 1$. The factor 0.1 is multiplied to the discrete random variable j in order to reduce the magnitude of the calculated SM values. Hence, the order of magnitude of each SM value, is similar to the order of magnitude of each probability value p_j. In the vector form, the SMs can be written as:

$$\boldsymbol{\mu} = \left[\mu_1, \mu_2, \mu_3, \mu_4\right]^T. \tag{5.2}$$

5.3.1.2 Information distribution

The information distribution is a non-linear transform applied to the probability values that magnifies low probability values and attenuates high probability values. Its interpretation assumes that information is described by low probability events (surprise events). The information of an event j (I_j) is here defined as:

$$I_j = -\log_{10}(f(j)) = -\log_{10}(p_j). \tag{5.3}$$

The $\log_{10}$ was used, instead of $\log_2$, in order to bound the maximum entropy to 1. In the vector form, the information distribution can be written as:

$$\mathbf{I}_{nf} = \left[I_1, I_2, ..., I_{10}\right]^T. \tag{5.4}$$

5.3.1.3 Entropy

The entropy (S_{ent}) consists on the expected value of the information distribution. Hence, it is calculated as:

$$S_{ent} = \mathbb{E}\{I_j\} = \sum_{j=1}^{J+1} f(j) I_j = \sum_{j=1}^{J+1} p_j I_j. \tag{5.5}$$

Its interpretation simply relates to the average information contained at some probability mass function (PMF).

5.3.1.4 Feature vector

Given the definitions about wPMFs, information distribution and entropy, a signal $s(n)$ can be represented by the feature vector $\mathbf{x}_{in,f}$. The elements of $\mathbf{x}_{in,f}$ are known as *features*. $\mathbf{x}_{in,f}$ is given as:

$$\mathbf{x}_{in,f} = \left[f(1), \cdots, f(J+1), \mu_1, \cdots, \mu_4, I_1, \cdots, I_{10}, S_{ent} \right]^T. \tag{5.6}$$

Notice the length of $\mathbf{x}_{in,f}$ is 25. The framework of the feature extraction for the *frequency analysis* is illustrated by Figure 5.4.

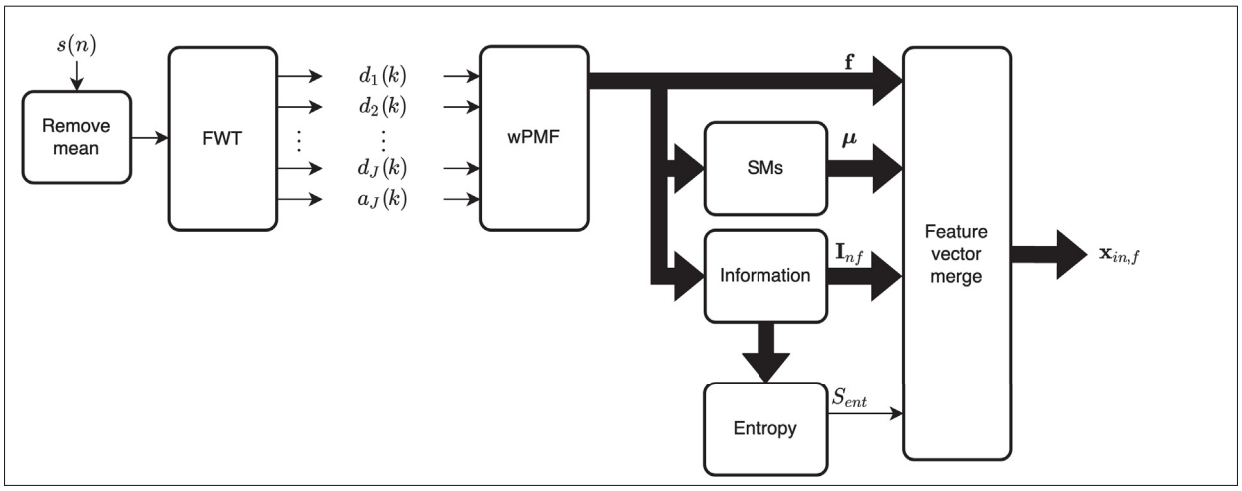


Figure 5.4 Framework of the *frequency analysis* feature extraction

5.3.2 Time-Frequency analysis feature representation

Given the definitions described by Section 3.3.2, a signal $s(n)$ can be represented by a grayscale image matrix, using the *time-frequency analysis*.

5.3.2.1 Feature vector

The feature vector $\mathbf{x}_{in,tf}$ is obtained by reshaping the *time-frequency* grayscale image matrix representation, of the signal $s(n)$, to a vector. The process of reshaping the image matrix to a vector is known as a *vectorizing* operation. Therefore, the feature vector is given as:

$$\mathbf{x}_{in,tf} = \left[f^{(1)}(1), \cdots, f^{(B)}(1), f^{(1)}(2), \cdots, f^{(B)}(2), f^{(1)}(3), \cdots, f^{(B)}(J+1) \right]^T. \quad (5.7)$$

The framework of the feature extraction for the *time-frequency analysis* is illustrated by Figure 5.5.

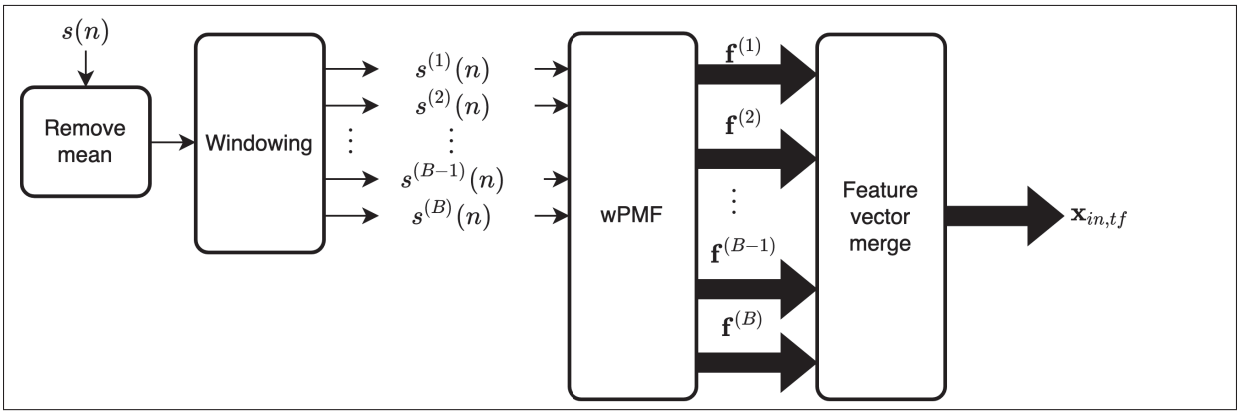


Figure 5.5 Framework of the *time-frequency analysis* feature extraction

5.4 Feature selection

Typically, in machine learning, systems with a few parameters and input features are easier for being both generalized and optimized. This problem is referred, by Richard Bellman (an American applied mathematician, who introduced dynamic programming in 1953), as the *curse of dimensionality*. It relates to the difficulty of using brute force when optimizing a function of too many input variables. The more features are used to represent a class, the more samples, or examples, are necessary to optimize a classifier. As an example, a typical rule of thumb for

training a k-means classifier is: using 10 samples for a data-set represented by 1 feature, 100 samples for a data-set represented by 2 features, 1000 samples for a data-set represented by 3 features, and so on, in order to maintain the same class density in the feature space.

Therefore, if a class is represented by too many features, it is important to select the most important ones, because simpler systems tend to perform better in terms of problem representation generalization. Moreover, feedforward networks, such as SVM and MLP, tend to perform better with decorrelated input data, due to its lack of feedback. Hence, feature selection (FS) is not only important to simplify the class representation, but also important to find an input subspace set that can be mostly decorrelated. Hence, FS can be understood as the process where features, that contribute the most to the class prediction, are automatically or manually selected. FS methods can be divided into *wrapper* or *filter* methods.

Filter methods use a well defined criteria to rank or to score a feature subset, instead of the error rate. These criteria can be based on the variance, information, redundancy of features, among others. Examples of *filter* methods are: *mutual information*, *minimum-redundancy-maximum-relevance* (mRMR), *joint mutual information*, etc.

Wrapper methods use the classification error (e.g., MSE) as a metric to evaluate subsets of the feature vector, unlike *filter* methods. Hence, *Wrapper* methods aim to find the best subset of features that maintain a a tolerable classification error. For a forward sequential feature selection, the first searching step, or round, for finding the best features, consist on training and evaluating a classifier with only one feature, then selecting the best feature. The second round searches for the second best feature, among the remaining ones, combined with the previous selected feature. The iterations continue until the error reaches a stop criteria. This method is known as *forward feature selection*. In summary, the first search round finds the best feature, then the second round finds the second best, combined with the first best feature, then the third round finds the third best feature, combined with the second and first best features, and so on.

In this thesis, features are only selected for the *frequency analysis* feature representation. Features are not selected for the *time-frequency analysis*, due to the fact the each pixel from

the *time-frequency analysis* image is assumed to be equally important for classification. The *wrapper* method used in this thesis uses a support vector machine (SVM) binary learner to quantify the MSE. Figures 5.6a and 5.6b show the FS results. The FS stop criteria is a difference of MSE lower then 10^{-8} in between two rounds of search.

```
Start forward sequential feature selection:
Initial columns included:  none
Columns that can not be included:  none
Step 1, added column 5, criterion value 3.41396e-05
Step 2, added column 3, criterion value 2.51245e-05
Step 3, added column 24, criterion value 2.23282e-05
Step 4, added column 4, criterion value 2.12015e-05
Step 5, added column 8, criterion value 2.03875e-05
Step 6, added column 11, criterion value 1.99493e-05
Step 7, added column 23, criterion value 1.94904e-05
Step 8, added column 15, criterion value 1.83426e-05
Step 9, added column 2, criterion value 1.78625e-05
Step 10, added column 10, criterion value 1.75078e-05
Step 11, added column 21, criterion value 1.74243e-05
Step 12, added column 16, criterion value 1.74034e-05
Final columns included:  2 3 4 5 8 10 11 15 16 21 23 24
```

(a) FS of the data set represented by the Db4 *mother wavelet*

```
Start forward sequential feature selection:
Initial columns included:  none
Columns that can not be included:  none
Step 1, added column 5, criterion value 3.59344e-05
Step 2, added column 4, criterion value 2.62724e-05
Step 3, added column 23, criterion value 2.10347e-05
Step 4, added column 6, criterion value 1.99287e-05
Step 5, added column 15, criterion value 1.89689e-05
Step 6, added column 9, criterion value 1.83428e-05
Step 7, added column 3, criterion value 1.81967e-05
Step 8, added column 2, criterion value 1.78628e-05
Step 9, added column 19, criterion value 1.74664e-05
Step 10, added column 20, criterion value 1.72786e-05
Step 11, added column 11, criterion value 1.70282e-05
Step 12, added column 7, criterion value 1.69864e-05
Step 13, added column 25, criterion value 1.67986e-05
Step 14, added column 17, criterion value 1.66525e-05
Step 15, added column 13, criterion value 1.66108e-05
Final columns included:  2 3 4 5 6 7 9 11 13 15 17 19 20 23 25
```

(b) FS of the data set represented by the Db6 *mother wavelet*

Figure 5.6 FS of the data set represented by a *frequency analysis* feature representation

Notice that FS reduces $\mathbf{x}_{in,f}$ to a vector of length 12, for the Db4 *frequency analysis* feature representation, and a vector of length 15 for the Db6 *frequency analysis* feature representation. Therefore, after FS, the feature vectors are given as:

$$\mathbf{x}^{FS}_{in,f,Db4} = \Big[f(5), f(3), I_9, f(4), f(8), \mu_1, I_9, I_1, f(2), f(10), I_7, I_2\Big]^T, \tag{5.8}$$

and

$$\mathbf{x}^{FS}_{in,f,Db6} = \Big[f(5), f(4), I_9, f(6), I_1, f(9), f(3), f(2), I_5, I_6, \mu_1, f(7), S_{ent}, I_3, \mu_3\Big]^T. \tag{5.9}$$

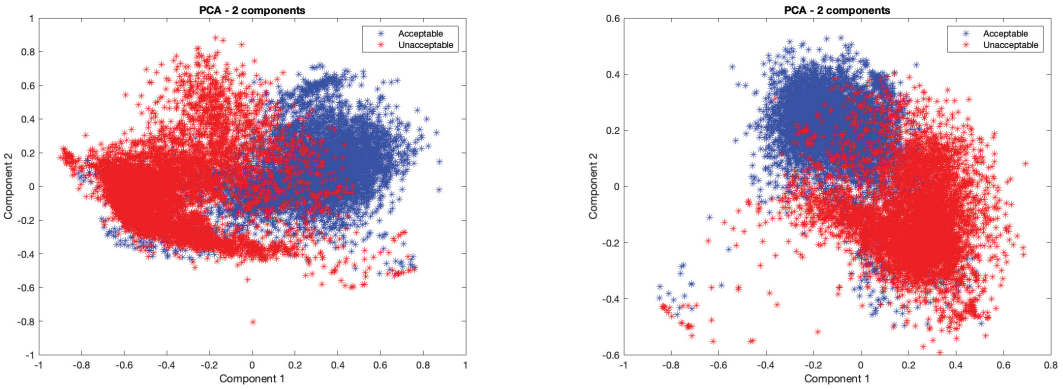
Where $\mathbf{x}^{FS}_{in,f,Db4}$ and $\mathbf{x}^{FS}_{in,f,Db6}$ represent, respectively, *frequency analysis* feature vectors, after FS, using the Db4 and Db6 *mother wavelets*. Each feature vector, given by equations (5.8) and

(5.9), have its elements ordered in degree of importance. Where the first element is the most important, and the last element is the least important feature. The remaining features that are not in the vector are considered completely irrelevant.

5.5 Separability analysis

Principal component analysis (PCA) is here used to analyze the separability of the *frequency analysis* feature representation, after FS, in accordance to figures 5.6a and 5.6b, and equations (5.8) and (5.9). PCA is a dimensionality reduction technique that projects the feature vector into the space of the principal components. The first principal component vector points to the direction of highest variance. The second principal component points to the direction of highest variance, constrained to the condition of being orthonormal to the first principal component. The third principal component, points to the direction of highest variance constrained to the condition of being orthonormal to both the first and second principal components. Subsequent principal components are calculated by following this logic. For details about PCA, the reader can consult the work of Haykin (2009). Figures 5.7a and 5.7b illustrate a the training data-set, represented by the feature vectors from equations (5.8) and (5.9), projected to the 2 principal components, in order to visualize if the data set is nicely separable, and if the examples are decorrelated.

Notice that the features seem to be decorrelated and the data is non-linearly separable for both Db4 and Db6 *mother wavelets*. According to Cybenko (1989), the *universal approximation theorem* concludes that a single hidden layer neural network, with sigmoid activation functions (e.g., logistic function), can represent a wide variety of interesting functions when given appropriate weights, number of neurons and features. Hence, a non-linear sigmoid activated feedforward ANN is sufficient for generalizing a non-linear decision boundary that accurately separates both classes. Furthermore, by inspecting figures 5.7a and 5.7b, the representation using Db6 seems to provide a slightly better separability, compared to Db4. Thus, it is expected for classifiers to perform better when using a Db6 *mother wavelet* for representing samples.



(a) PCA of data set represented by the Db4 *mother wavelet*

(b) PCA of data set represented by the Db6 *mother wavelet*

Figure 5.7 PCA plots of the data set represented by a *frequency analysis* feature representation and 2 principal components

5.6 Classifier architectures, training and test experimental procedure description

This section presents the classifier architectures and the experimental procedure description for training and testing a single lead ECG acceptability detector, based on the data-sets described in Chapter 4, and the features described in this current chapter.

5.6.1 Classifier architectures

For details about the mathematics behind each classifier, read Appendix II.

5.6.1.1 Frequency analysis ANN architectures

In conformance to Section 5.4, features are selected in accordance to figures 5.6a and 5.6b. In order to simplify the architectures of the *frequency analysis* feature representation classifiers, features are further reduced to a maximum of 8 most important features and a minimum of 1 most important feature. In this thesis this is referred as *feature pruning* (FP).

For the *frequency analysis* feature representation, prior to FP, the chosen ANN architectures to be trained were:

- An MLP with one hidden layer, 50 neurons in the hidden layer, a hyperbolic tangent transfer function in the hidden layer, a softmax output layer transfer function, gradient descent with adaptive learning rate, and a value of 0.9 momentum constant.

- An SVM, with Radius Basis Function (RBF) as Kernel Function, and zero mean normalized inputs.

After FP, the number of neurons in the hidden layer is reduced to 16. The reduced MLP architecture is presented by Figure 5.8.

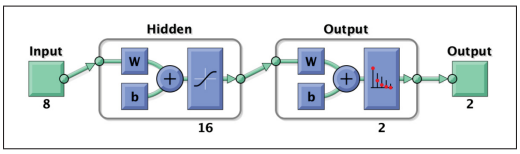


Figure 5.8 Reduced MLP architecture after FP, with 8 features as input

It is important to emphasize that a hyperbolic tangent is chosen in order to provide the ANN an extra degree of freedom, where the output of the neurons from the hidden layer can also be negative values. In addition, experimental results of the ANN training showed that the weights did not tend to infinity. As a conclusion, the use of both regularization and normalization are not necessary, since all input features are approximately bounded by 1. The same design choices are applied to the *time-frequency analysis* feature representation ANN architectures.

5.6.1.2 Time-frequency analysis ANN architectures

For the *time-frequency analysis* feature representation, the chosen ANN architectures to be trained were:

- An MLP with one hidden layer, 100 neurons in the hidden layer, a hyperbolic tangent transfer function in the hidden layer, a softmax output layer transfer function, gradient descent with adaptive learning rate, 0.9 momentum constant. This architecture is refereed here as MLP1 (see Figure 5.9).

- An MLP with one hidden layer, 30 neurons in the hidden layers, a hyperbolic tangent transfer function in the hidden layers, a softmax output layer transfer function, gradient descent with adaptive learning rate, 0.9 momentum constant. Since the columns of the *time-frequency analysis* can be time dependent (linearly correlated), each hidden layer receives an independent input and the previous layer output as an input. Here, the assumption is that the non-linear transform of each layer is responsible for decorrelating inputs from one-another. This architecture is refereed here as MLP2 (see Figure 5.10).

- An ART2 network, with an *outstar Grossberg layer* appended to the output, axactly as described in Appendix II, Section 4.

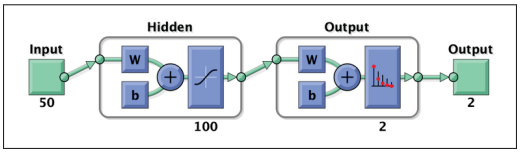


Figure 5.9 MLP1 architecture

The *time-frequency analysis* feature representation classifiers, with the exception of the MLP2, are designed to classify an image that is vectorized, according to Section 5.3.2. Since the input features consist on image pixels, FS and FP are not necessary, because all features are assumed to be equally important.

Additionally, the images have the same format as presented by figures 3.15, 3.16, 3.17. Each row of the image is decorrelated, because both the Db4 and Db6 wavelet transforms are characterized by the property of orthogonality. However, each column of the image has a time dependency. Therefore, the image columns are expected to be correlated. To overcome this problem, the MLP2 architecture is used, with each hidden layer receiving an image column and the previous output layer as inputs, except for Hidden 1 (see Figure 5.10).

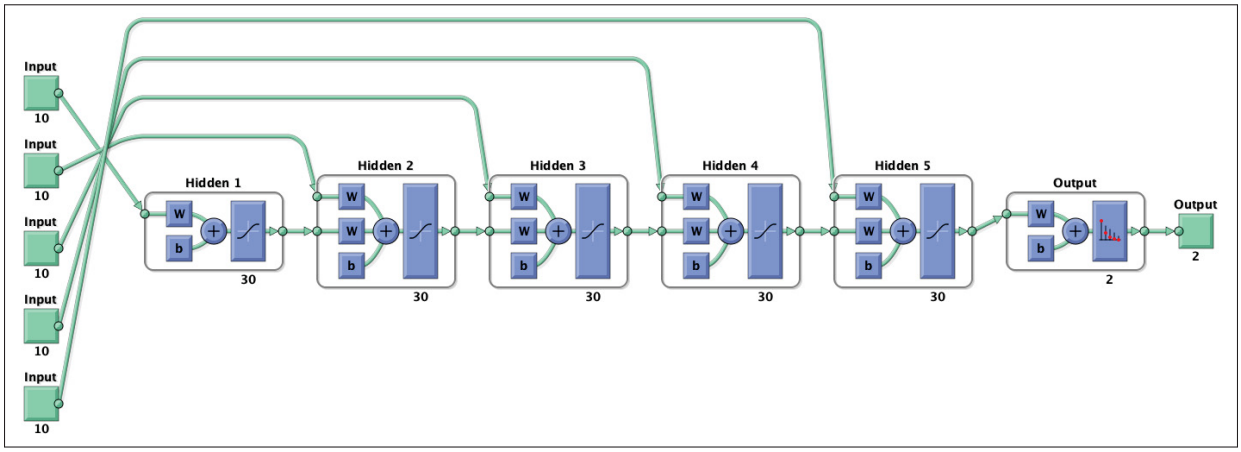


Figure 5.10 MLP2 architecture

5.6.1.3 Standard ANN architectures

The previously described architectures are compared to an MLP and an SVM that use the input features basSQI, kSQI and pSQI, as described by Clifford *et al.* (2012). For sake of simplicity, the bSQI is not considered in the results comparison in this thesis, since Clifford *et al.* (2012, p. 1428) showed that both kSQI and basSQI, alone, have shown to be better features, compared to the bSQI (see Table 2 from Clifford *et al.* (2012, p. 1428)). Moreover, Clifford *et al.* (2012) calculates the process time required for computing each SQI, as a manner of quantifying the computational cost. The bSQI required at least 33.18 ms, while the power estimation routine (necessary for basSQI) required 2.46 ms, kSQI required 0.33 ms, and pSQI required 1.92 ms (see Table 8 from Clifford *et al.* (2012, p. 1431)). It is evident that bSQI is the most computationally expensive SQI to be calculated, compared to other indexes. Hence, not worthy of being used for comparison of results. The standard architectures are structured in the same format as the ones described in Section 5.6.1.1, prior to FP. Here, these architectures are referred as *MLP Cliff* and *SVM Cliff*.

5.6.2 Data separation and training stop criteria

As mentioned before in Chapter 4, there are two data-sets: the *PhysioNet data-set* and the *SIG.test data-set*. Since the training data-set most likely contain some noise, training the classifiers with a random subset of the original data-set will increase its accuracy over the test data-set. Therefore, the *PhysioNet data-set* is divided based on a 4-fold cross validation (CV). Each classifier is trained and tested a total of 4 times, where 3-folds, from the *PhysioNet data-set*, are used for training and the *SIG.test data-set* is entirely used for testing, for all combinations of folds. This is known as *leave-one-out* CV, except that the testing set (remaining fold) is always replaced by the *SIG.test data-set*.

When training the MLP classifiers, within the set of 3-folds (75% of the *PhysioNet data-set*), 75% of the 3-folds is randomly selected for training, and the remaining 25% is used for validation. Again, the test set is always the *SIG.test data-set*, and the training and testing procedures happen for all combinations (total of 4) of 3-folds out of a set of 4-folds. The validation set is used as a stop criteria for overfitting precaution purposes. Hence, the classifier is most likely generalized.

The training stop criteria of the MLP classifiers are: 1) if the maximum number of epochs (5000) is reached; 2) if the training MSE reaches a value of 10^{-5}; 3) if the minimum gradient reaches a value of 10^{-5}; 4) if the MSE of the validation data-set increases 20 consecutive times, while the MSE of the training data-set is decreasing.

For the remaining classifiers (ART2 and SVM), where no validation set is used. Each classifier is trained with 3-folds (75% of the *PhysioNet data-set*) and tested with the *SIG.test data-set*, for all combinations (total of 4) of 3-folds. Hence, the training and testing occur a total of 4 times. The training stop criteria of the ART2 classifier is reaching 300 epochs, while the SVM classifier has no training stop criteria because it is trained based on a global optimization procedure, given the support vectors.

All presented results are based on the best trained classifier, out of 4 training trials. The best trained classifier is chosen based on a parameter called *area under the curve* (AUC). AUC is

simply the area under the receiver operating curve (ROC), calculated based on the ROC of the testing data-set classification (*SIG.test data-set*).

5.7 Frequency analysis feature representation results

Results follow the experimental procedures described in Section 5.6.2. As expected in the section of separability analysis using PCA (Section 5.5), the experimental procedure results showed that the Db6 *frequency analysis* feature representation represents a better separability, compared to the Db4 *frequency analysis* feature representation. Therefore, the following results will concern the Db6 *frequency analysis* feature representation only. Results concerning the Db4 *frequency analysis* feature representation are illustrated in Appendix IV.

5.7.1 Results prior to feature pruning

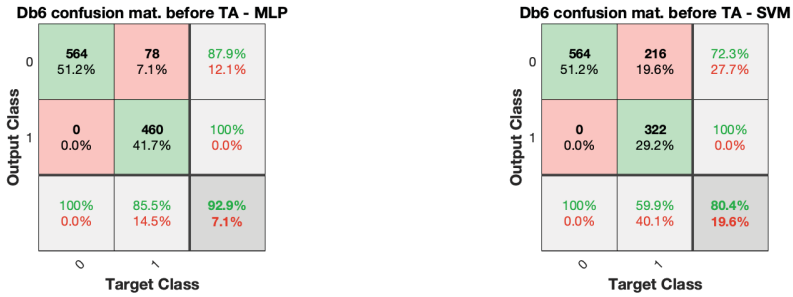
This section concerns results prior to (before) FP. The following sections concern results before a threshold adjustment (TA). TA is explained in Section 5.7.1.2.

5.7.1.1 Confusion matrix prior to threshold adjustment

Figures 5.11a and 5.11b, each, concerns a testing result, before TA, of the best trained classifier, out of 4, for both MLP and SVM tested on the *SIG.test data-set*. As a reminder, the best classifier is chosen based on the highest value of the AUC parameter.

5.7.1.2 Threshold adjustment

The threshold adjustment (TA) simply consists in setting an optimal class discriminant threshold, or decision rule, aiming to enhance the classifier accuracy, with the price of augmenting the probability of false alarm. In the case of MLP classifiers, the first element of the output vector is a probability value (positive semi-definite) that indicates whether the input is an acceptable signal or not. For the SVM classifier, the output value is either positive or negative. Where



(a) MLP confusion matrix (b) SVM confusion matrix

Figure 5.11 Confusion matrix from the best trained MLP or SVM, before FP and TA, using a Db6 *frequency analysis* feature representation, tested on the *SIG.test data-set*

doubtful inputs are less then unity, centred in zero. The default threshold is 0.5 for the MLP and 0 for the SVM.

The decision rule is optimized by using a Neyman-Pearson criteria, where the threshold adjustment is constrained to a limit of at most 3% of the the probability of false alarm estimation, from the testing data-set. The probability of false alarm, probability of detection and the probability of miss are estimated by calculating the sensitivity (detection rate) and specificity, according to:

$$Sen = \frac{T_{pos}}{T_{pos} + F_{neg}} \tag{5.10}$$

$$Spec = \frac{T_{neg}}{T_{neg} + F_{pos}}. \tag{5.11}$$

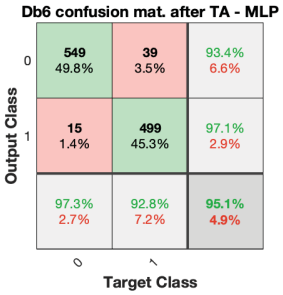
Where T_{pos} is the number of true positives, T_{neg} is the number of true negatives, F_{pos} is the number of false positives, and F_{neg} is the number of false negatives. The probability of false alarm is estimated as $1 - Sen$ (false alarm rate). The probability of miss is estimated as $1 - Spec$. As a result, the TA using a 3% Neyman-Pearson criteria, increases the accuracy (probability of correct decision) of the system. Where

$$Acc = \frac{T_{pos} + T_{neg}}{T_{pos} + T_{neg} + F_{pos} + F_{neg}}. \tag{5.12}$$

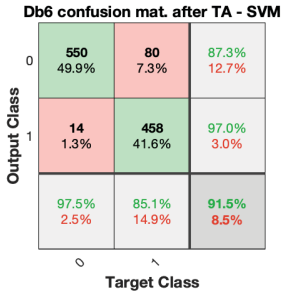
Figures IV-3, IV-4, IV-5 and IV-6, in Appendix IV, illustrate that, by decreasing the threshold, the probability of miss decreases, the accuracy increases, and the probability of false alarm increases up to a maximum.

5.7.1.3 Confusion matrix after threshold adjustment

Figures 5.12a and 5.12b, each, concerns a testing result, after TA, of the best trained classifier, out of 4, for both MLP and SVM, tested on the *SIG.test data-set*.



(a) MLP confusion matrix



(b) SVM confusion matrix

Figure 5.12 Confusion matrix from the best trained MLP or SVM, before FP, after TA, using a Db6 *frequency analysis* feature representation, tested on the *SIG.test data-set*

5.7.1.4 Comparison with the standard architectures

The Db6 *frequency analysis* feature representation classifiers are compared with the standard architectures: MLP Cliff and SVM Cliff. Figure 5.13 presents such comparison, prior to FP.

By inspecting Figure 5.13, it is evident that, for the same types of features, an MLP classifier has a higher *AUC* compared to an SVM classifier. For this reason, further tests will be focused on the use of MLP ANN architectures.

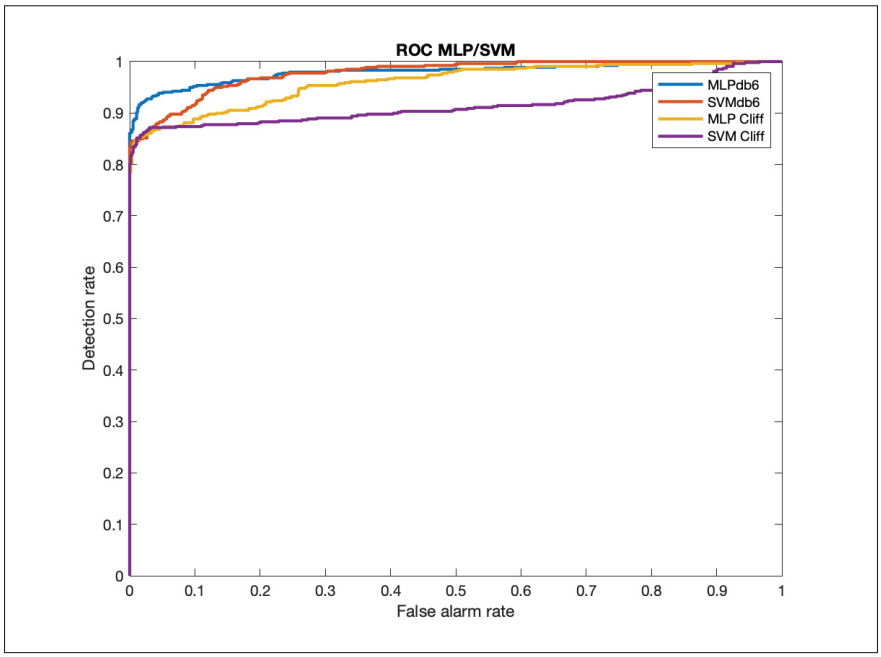


Figure 5.13 ROC from the best trained classifiers, prior to FP, using a Db6 *frequency analysis* feature representation, compared to the standard ANN architectures (MLP Cliff and SVM Cliff), all tested on the *SIG.test data-set*

5.7.2 Results after feature pruning

Table 5.1 show results prior to TA, for a 4-fold CV. Results concern an MLP ANN with a Db6 *frequency analysis* feature representation as inputs prior to FP, an *MLP Cliff* architecture, and a reduced MLP architecture, after FP, with 1 to 8 features as inputs. The reason why 8 maximum features were chosen is the AUC did not augmented more with the use of 9 features or more. When features are pruned, the reduced MLP architecture, has always 16 neurons in the hidden layer, and 2 neurons in the output layer (see Figure 5.8). The table shows values of: mean $Acc \pm 2$ standard deviation (std), mean Sen, mean $Spec$ and mean AUC.

From Table 5.1, it is evident that pruned features classifiers are better generalized (higher AUC). The ROC curves of the best trained classifier (highest AUC), for different feature pruning trials, is shown by Figure 5.14.

Table 5.1 4-fold CV results of comparison between MLP Db6 - no FP, MLP Cliff, and reduced MLP architectures after FP

ANN architecture	Mean Acc ± 2std	Mean Sen	Mean Spec	Mean AUC
MLP Db6 - no FP	92.1 ± 2.1	100.00	86.69	0.97
MLP Cliff	90 ±0.78	99.88	83.76	0.96
1 feat	88.7 ± 0.1	93.71	84.94	0.94
2 feats	91.7 ±0.44	99.23	86.36	0.96
3 feats	90.6 ±0.54	99.71	84.59	0.96
4 feats	90.4 ± 1.1	99.54	84.39	0.96
5 feats	90.3 ±0.46	99.66	84.27	0.96
6 feats	91.7 ±0.68	99.56	86.20	0.97
7 feats	91.1 ±0.31	99.66	85.39	0.97
8 feats	91.4 ±0.94	99.78	85.76	0.98

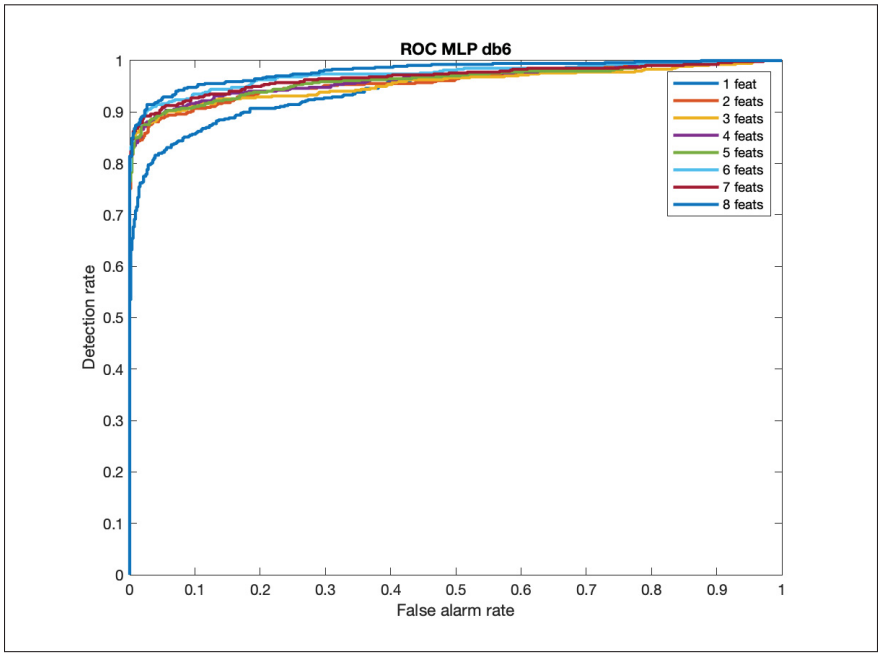
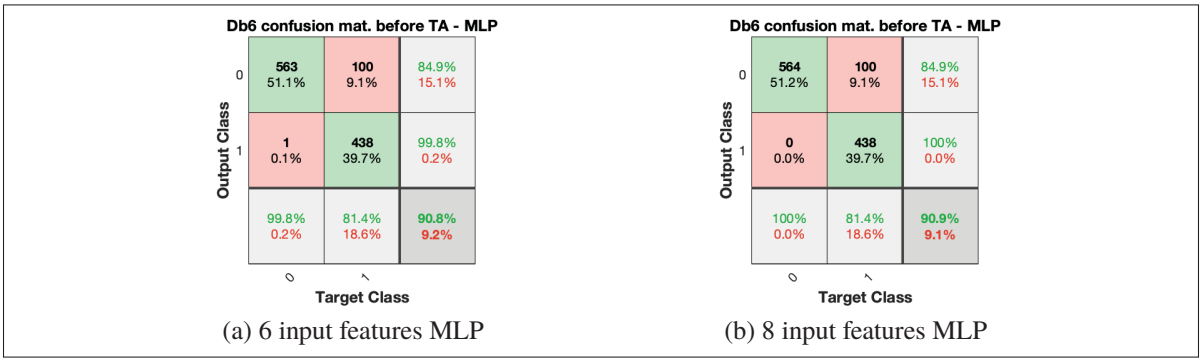


Figure 5.14 ROC from the best trained classifiers, after FP, using a Db6 *frequency analysis* feature representation and tested on the *SIG.test data-set*

5.7.2.1 Confision matrix of best frequency analysis classifiers, prior to threshold adjustment

This current section shows the confusion matrix results (figures 5.15a and 5.15b) of the two best trained *frequency analysis* feature representation classifiers, chosen based on highest *AUC*, with pruned features, before TA.



(a) 6 input features MLP (b) 8 input features MLP

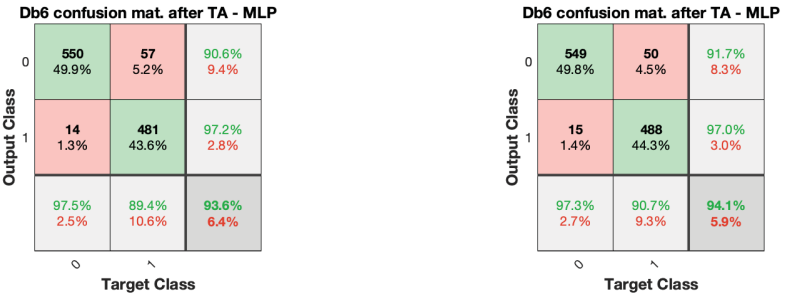
Figure 5.15 Confusion matrix from the two best trained MLPs, after FP, before TA, using a Db6 *frequency analysis* feature representation, tested on the *SIG.test data-set*

5.7.2.2 Confusion matrix of best frequency analysis classifiers, after the threshold adjustment

Similar to Section 5.7.2.1, this current section shows the confusion matrix results (figures 5.16a and 5.16b) of the two best trained *frequency analysis* feature representation classifiers, chosen based on highest *AUC*, with pruned features, after applying TA.

5.8 Time-frequency analysis feature representation results

Analogous to Section 5.7, the results presented in this current section follow the described experimental procedures described in Section 5.6.2. For the same reasons presented in Section 5.7, the following results will concern a *time-frequency analysis* feature representation using a Db6 *mother wavelet* only. Table 5.2 summarizes the 4-fold CV results, prior to TA, of the different *time-frequency analysis* feature representation architectures.



(a) 6 input features MLP (b) 8 input features MLP

Figure 5.16 Confusion matrix from the best trained MLPs, after FP and TA, using a Db6 *frequency analysis* feature representation, tested on the *SIG.test data-set*

Table 5.2 4-fold CV results of comparison between best *time-frequency analysis* feature representation architectures

ANN architecture	Mean Acc ± 2std	Mean Sen	Mean Spec	Mean AUC
MLP1	91.9 ± 0.6	99.56	86.53	0.97
MLP2	91.8 ± 1.3	99.61	86.41	0.96
ART2	91 ±0.68	97.82	86.16	0.96

Furthermore, Figure 5.17 illustrates the ROC curves of the best trained classifier for each architecture. Again, the best classifier is chosen based on the highest *AUC* value.

5.8.1 Confusion matrix of best classifier

This current section presents the confusion matrix of the best trained *time-frequency analysis* feature representation classifier (MLP1). Figures 5.18a and 5.18b illustrate, respectively, the confusion matrix test of MLP1 before, and after TA.

5.9 Time-frequency analysis vs frequency analysis feature representation results

Table 5.3 presents a comparison summary of the best architectures, prior to TA. Even though all performances were quite similar, the best performance occurred for a simpler system architecture (a reduced MLP ANN with 8 features as input), due to its highest value of mean *AUC*.

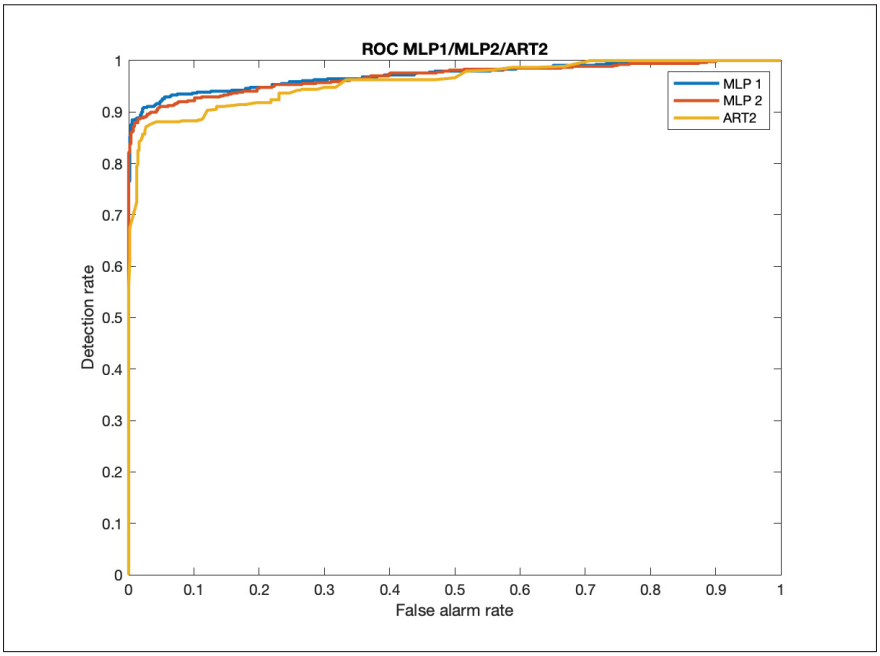
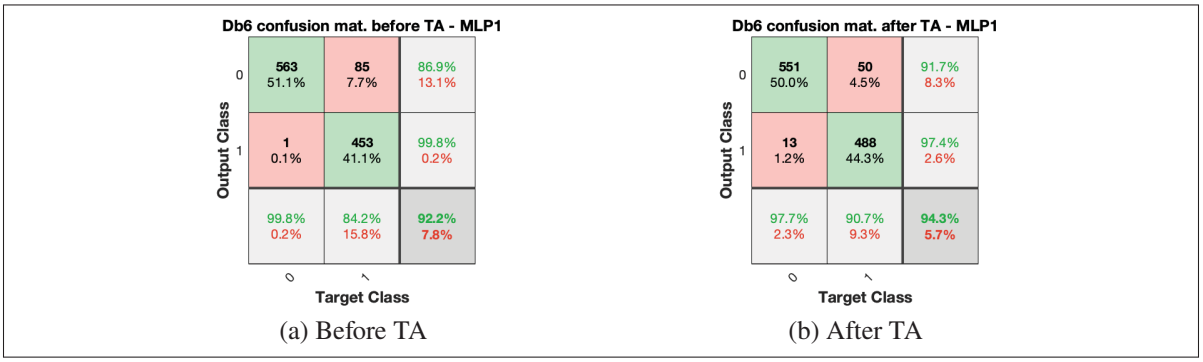


Figure 5.17 ROC from the best trained classifiers, using a Db6 *time-frequency analysis* feature representation and tested on the *SIG.test data-set*



(a) Before TA (b) After TA

Figure 5.18 Confusion matrix from the best trained MLP1, using a Db6 *time-frequency analysis* feature representation, tested on the *SIG.test data-set*

It is recommended to implement the SQA system using the reduced MLP ANN with the best 8 input features architecture, because of its great performance and simplicity. Therefore, promising accurate results, simplicity and low computational cost.

Table 5.3 4-fold CV results of comparison between best architectures

ANN architecture	Mean Acc ± 2std	Mean Sen	Mean Spec	Mean AUC
MLP1	91.9 ± 0.6	99.56	86.53	0.97
MLP Db6 - no FP	92.1 ± 2.1	100.00	86.69	0.97
8 feats	91.4 ±0.94	99.78	85.76	0.98

5.10 Computational complexity analysis of feature extraction

In the analysis presented by this current section, two different approaches will be assumed for the work of Clifford *et al.* (2012) to calculate the basSQI, kSQI and pSQI. Method 1, uses an approach that introduces a very small quantity of distortion to the calculated features, but it is computationally expensive. Method 2 is not computationally expensive, compared to method 1, but it can introduce a significant amount quantity of distortion to the calculated features. The method for obtaining a *frequency analysis* feature representation is here referred as method 3.

It is important to mention that, in the work of Clifford *et al.* (2012), the sampling frequency of the signals is 125 Hz. Hence, for 10 seconds, there are 1250 samples. In the application presented by this current chapter, there are 2560 samples, for 10 s of data.

The following complexity analysis is based on the work of Blahut (2010). Here, we simply take in consideration the number of multiplications to define the computational cost. The computational complexity analysis will assume a sampling frequency of 125 Hz and 10 s of data for methods 1 and 2, according to the work of Clifford *et al.* (2012), and a sampling frequency of 256 Hz and 10 s of data for the method 3, according to the proposal of this thesis.

5.10.1 Method 1

One way of estimating the power spectrum density (PSD), used to calculate the pSQI and basSQI, is computing the FT of the auto-correlation function. This approach is interesting, because it provides a very accurate PSD, with minimum distortion. The multiplication complexity of a linear

convolution (LC) operation, of two vectors of size L and N (LC($L \times N$)), is $M_c(L \times N) = L \times N$, by using a direct method. Where the operator $M_c(\cdot)$ indicates the multiplication complexity.

By using a fast convolution algorithm, such as the method of Cook-Toom, the multiplication complexity (Blahut, 2010, p. 149) of the linear convolution is reduced to:

$$M_c(L \times N) = L + N - 1. \tag{5.13}$$

The following methods will assume the use of Cook-Toom's implementation for every linear convolution operation. Therefore, 2499 multiplications are necessary for estimating the auto-correlation. A zero is appended to the end of the auto-correlation estimation vector, thus introducing a negligible quantity of distortion. As a result, the length of the auto-correlation signal vector should be 2500. For estimating the FT, 62500 multiplications are necessary by using a Cooley–Tukey (CT) tree fast Fourier transform (FFT), and factors of 2 and 5, for a signal of length 2500. The multiplication complexity of the branch of a CT tree decomposition, for a signal of length N, where $N = N'N''$, according to Blahut (2010, p. 71), is given as:

$$M_c(N) = N' M_c(N'') + N'' M_c(N') + N. \tag{5.14}$$

If N is prime, $M_c(N) = N^2$ (Blahut, 2010, p. 68). If $N = 2^m$, for $m \geq 1$, according to Blahut (2010, p. 75), the multiplication complexity is:

$$M_c(N) = \frac{N}{2} \log_2 N. \tag{5.15}$$

This is a special case of the CT algorithm, known as Cooley–Tukey base 2 (CTB2), or simply *radix-two* (Blahut, 2010, p. 72). Hence, the multiplication complexity for an optimal FFT tree decomposition is given as:

$$
\begin{aligned}
M_c(2500) &= 4M_c(625) + 625M_c(4) + 2500 \\
&= 4\Big[125M_c(5) + 5M_c(125) + 625\Big] + 625[\frac{4}{2}\log_2 4] + 2500 \\
&= 4\Big[3125 + 5\big[5M_c(25) + 25M_c(5) + 125\big] + 625\Big] + 5000 \\
&= 4\Big[3750 + 5\Big[5\big[10M_c(5) + 25\big] + 750\Big]\Big] + 5000 \\
&= 62500.
\end{aligned} \tag{5.16}
$$

Both the basSQI and pSQI are estimated by summing a particular interval of the PSD. For estimating the kSQI of a signal $s(n)$, of length $M = 1250$, 5003 multiplications are required. Therefore, a total of 70002 multiplications are required, using method 1. The kSQI is given as:

$$
kSQI = \frac{\frac{1}{M}\sum_{n=0}^{M-1}\left(s(n) - \frac{1}{M}\sum_{n=0}^{M-1} s(n)\right)^4}{\left(\frac{1}{M}\sum_{n=0}^{M-1}\left(s(n) - \frac{1}{M}\sum_{n=0}^{M-1} s(n)\right)^2\right)^2} = \frac{M\sum_{n=0}^{M-1}(s(n) - \bar{s})^4}{\left(\sum_{n=0}^{M-1}(s(n) - \bar{s})^2\right)^2}. \tag{5.17}
$$

5.10.2 Method 2

In this method, the signal must be filtered a total of 3 times by an FIR type 1 filter. The FIR type 1 filter is chosen because of its property of linear phase. Thus, no phase distortion is introduced to the filtered signal.

On the other hand, the transition band is not as sharp as an IIR filter. Thus, amplitude distortion might be introduced to the filtered signal. Such effect can be attenuated by using efficient multistage techniques. For simplicity of the analysis, the filter is implemented as a single LC, using Cook-Toom's fast filtering implementation.

The first FIR filter ($h_1(n)$) is low-pass, and is designed with a passing band edge (f_{pass}) of 39.8 Hz, a stop band edge (f_{stop}) of 40.2 Hz, a passband ripple of (A_{pass}) 0.02 dB and a stop band ripple (A_{stop}) of 25 dB. The second FIR filter ($h_2(n)$) is also low-pass and is designed with the same specifications as the previous filter, except for $f_{pass} = 0.8$ Hz and $f_{stop} = 1.2$ Hz. The third filter ($h_3(n)$) is band-pass, with specifications $f_{stop1} = 4.8$ Hz, $f_{pass1} = 5.2$ Hz, $f_{pass2} = 14.8$ Hz, $f_{stop2} = 15.2$ Hz, $A_{pass} = 0.02$ dB, $A_{stop1} = A_{stop2} = 25$ dB. The number of required coefficients for all these FIR type 1 filters is 656. The fourth FIR filter ($h_4(n)$) is high-pass, with $f_{pass} = 15.2$ Hz, $f_{stop} = 14.8$ Hz, and same A_{pass} and A_{stop} as $h_1(n)$. The number of required coefficients for $h_4(n)$ (FIR type 1) is 616. All coefficients were estimated by using an FIR equiripple filter approximation design method.

The power of signal $s(n)$, in the interval of 0 Hz $\sim$ 40 Hz, can be estimated by computing the power of signal $s_1(n) = s(n) * h_1(n)$. The power of signal $s(n)$, in the interval of 0 Hz $\sim$ 1 Hz, can be estimated by computing the power of signal $s_2(n) = s(n) * h_2(n)$. The power of signal $s(n)$, in the interval of 5 Hz $\sim$ 15 Hz, can be estimated by computing the power of signal $s_3(n) = s(n) * h_3(n)$. The power of signal $s(n)$, in the interval of 5 Hz $\sim$ 40 Hz, can be estimated by computing the power of signal $s_4(n) = s_1(n) * h_4(n)$. Hence, 4 fast filtering computations are required.

The number of multiplications required for an FIR type 1 filter of N taps, to filter a signal of length L, using an LC Cook-Toom implementation, is $L + (\frac{N}{2} + 1) - 1$, due to the symmetry of FIR type 1 filters. The multiplication complexity for filtering is 3×1578, total, for the filters of length 656, and 1558 for the filter of length 616. Thus, a total of 6292 multiplications.

The energy is estimated by summing the squared convoluted signals. Hence, $3 \times (1250 + 656 - 1) + (1250 + 616 - 1) = 7580$. 2 additional divisions are necessary for estimating the basSQI and pSQI power rates. Finally, like method 1, the kSQI requires 5003 multiplications. Therefore, a total of 18877.

5.10.3 Method 3

Assuming a similar filter bank structure as presented by Figure 3.5, using a Db6 *mother wavelet* and a Cook-Toom linear convolution algorithm, 5200 multiplications are necessary to estimate the wavelet coefficients, from a total of 2560 inputs samples. The estimation of the multiplication cost is:

$$M_c(2560) = 2\sum_{i=1}^{9} M_c\left(\frac{2560}{2^i} \times 6\right) = 5200. \tag{5.18}$$

To estimate the wPMF, more 2570 multiplications are necessary. For the SMs, 60 multiplications are necessary (see equation (5.1)). The multiplication complexity for approximating the logarithm function, using the coordinate rotation digital computer (CORDIC), can be 0 (Owen, 2020). However, being very pessimistic, by using a Maclaurin series for the natural logarithm, the base 10 logarithm function, for $|x| < 1$, is here estimated as:

$$\log_{10}(x) \approx -\frac{1}{\ln(10)} \sum_{k=1}^{16} \frac{(1-x)^k}{k}. \tag{5.19}$$

Therefore, a total of 320 multiplications are necessary for estimating the information distribution from the wPMF, and finally, 10 multiplication for estimating the entropy. Hence, the method 3 requires a total of 8160 multiplications. By using only the 8 most important features, this number is reduced to 7834. The 8 most important features are given as:

$$\mathbf{x}_{in,f,Db6}^{FS,8} = \Big[f(5), f(4), I_9, f(6), I_1, f(9), f(3), f(2)\Big]^T. \tag{5.20}$$

The following section presents a summary of the computational cost investigation.

5.10.4 Summary of computational cost

Table 5.4 summarized the computational cost analysis for all 3 methods of feature extraction.

Table 5.4 Summary of computational cost analysis

Authors	**Assumptions**	**Formula**	M_c
Clifford *et al.* (2012) (method 1)	Auto-correlation estimation.	$M_c(L \times L) = L + L - 1$.	2499
	CT & CTB2.	$M_c(N) = N' M_c(N'') + N'' M_c(N') + N$. $M_c(N) = \frac{N}{2} \log_2 N$. $M_c(N) = N^2$. See equation (5.14).	62500
	kSQI.	See equation (5.17).	5003
			Total: 70002.
Clifford *et al.* (2012) (method 2)	LC.	For FIR type 1 filters, using Cook-Toom LC, $M_c(L \times N/2) = L + (N/2 + 1) - 1$ multiplications.	6292
	Energy.	$L + N - 1$ multiplications.	7580
	Additional division.	Additional divisions required to calculate basSQI & pSQI.	2
	kSQI.	See equation (5.17).	5003
			Total: 18877.

Table 5.4 Summary of computational cost analysis *(cont'd)*

Authors	**Assumptions**	**Formula**	M_c
Lucas T. Lins (method 3)	FWT and energy.	$M_c(L \times N) = L + N - 1$; See Section 3.2.3 and equation (5.18).	5200
	wPMF.	See equation (3.61).	2570
	SMs	See equation (5.1).	60
	Info. dist.	See equation (5.3).	320
	Entropy.	See equation (5.5).	10
			Total: 8160. Using 8 most important features, total: 7834.

5.11 Conclusion

In conclusion, the multiplication complexity of the features extraction methodology presented in this thesis is reduced, in comparison to the presented methodologies for representing features based on the work of Clifford *et al.* (2012). Additionally, the *frequency analysis* feature representation proposed by this thesis (method 3) has a better accuracy, compared to the system proposed in the work of Clifford *et al.* (2012). Furthermore, this thesis simplifies the MLP architecture to 16 neurons in the hidden layer, while Clifford *et al.* (2012) proposed an MLP of 50 neurons in the hidden layer.

Notice that the work of Clifford *et al.* (2012) used a lower sampling frequency, compared to the methodology proposed by this thesis. Still, the multiplication complexity of method 2 was superior compared to method 3. If a lower sampling frequency is used for implementing method 3 (e.g. 128 Hz), the multiplication complexity of method 3 could even be further reduced.

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This thesis first presented, in Chapter 1, some fundamentals about ECG instrumentation systems and then, in Chapter 2, a literature review about SQA, in the context of ECG and cECG, and a review about many trust metrics, known as SQIs, that are used in the application of ECG SQA, in order to assure the reliability of monitoring systems. Subsequently, in Chapter 3, two computationally inexpensive and compact techniques for representing an ECG signal, inspired from the use of wavelet scalograms (in the vision of Ariño *et al.* (2004)), are presented. It is believed that these two methods are novel to the application of ECG SQA. Afterwards, Chapter 4 describes, in the context of ECG SQA, the data-sets preparation and labelling criteria used for a merge of databases from PhysioNet (Goldberger *et al.*, 2000) and the *SIG.test* data-set. At last, Chapter 5 presents the application itself, and its performance results. Most of the proposed SQA classification systems, presented in this thesis, showed to be more accurate and computationally more efficient, compared to the system presented by Clifford *et al.* (2012).

In general, the SQA systems, proposed in this thesis, showed to be very efficient in terms of both accuracy and computational complexity. The work of this thesis proves that ECG SQA systems do not require a lot of computation, neither deep ANN techniques for achieving outstanding results. Differently from many other works, this thesis aimed to achieve not only high accuracy, but also simplicity and computational efficiency. The main purpose of the SQA classifier systems, proposed in this work, is to be used as a real-time channel selection algorithm for the SIG.ECG system. Due to the fact that the SIG.ECG system is capable of measuring approximately the same lead by using different combinations of proximal pairs of sensors, it is necessary to have such SQA classifier to only select and analyze the pair of sensors that are measuring acceptable signals. Therefore, both the computational cost and accuracy are very important factors to be considered in such real-time application scenario.

It is evident that any automated clinical data analysis technique should proceed with an estimation of reliability or trust level. The reliability or trust metrics can then be transferred to humans or other processing units in order to assure that only reliable data is being used for important estimates, predictions, evaluation or decision support. Without the appropriate quantification of the quality of signals that are being used for estimating parameters or controlling the response of a system, clinical actions most likely will be based on erroneous data, leading to false alarms or a distrust of measurements. Hence, the main reason why SQA systems are so important.

At last, but not least, it is important to mention that the characterization and understanding of the SIG.ECG system were fruits of investigations that motivated this thesis, and contributed to the publication of the work of Weeks *et al.* (2020).

6.2 Recommendations

The method 3, proposed by this thesis (based on the *frequency analysis* representation), in conjunction with a simplified MLP architecture of 8 inputs and 16 neurons in the hidden layer, has a lower computational complexity and better accuracy performance compared to the system proposed in the work of Clifford *et al.* (2012). Therefore, as concluded by Chapter 5, the *frequency analysis* based fast feature extraction algorithm (method 3), in conjunction with the simplified MLP architecture of 8 input features (8 most important features), and 16 neurons in the hidden layer, is recommended to be used by the SIG.ECG system as a fast *channel selection* algorithm solution.

Even if such method has no time information, it is evident that the FWT filter bank ($\mathcal{O}(N)$) allows the features to be extracted much more rapidly then FFT approaches ($\mathcal{O}(N \log N)$). Furthermore, it also allows ECG signals to be represented in a much more compact way (lower frequency resolution compared to the FFT). In conclusion, a low frequency resolution, with no time information, is enough for extremely shallow ANNs being able to assess the quality of

ECG signals by doing classification. Again, all of that is very much important to consider in real-time application scenarios, where simplicity and computational cost are both crucial.

6.3 Future work suggestions

It is suggested to investigate the use of wavelet scalograms for other applications, such as: bio-signal diagnose, ECG beat classification applications, etc. These applications are related to the proposed task of the PhysioNet - Computing in Cardiology Challenge of 2020 (Perez Alday, Gu, Shah, Robichaux, Wong, Liu, Liu, Rad, Elola, Seyedi, Li, Sharma, Clifford & Reyna, 2020). More information about wavelets and and beat classification applications can be found in the work of Böck (2015).

In such case, a richer scalogram representation of signals might be required. As an example of a recent work, with very good results, Wang, Lu, Sun, Yang, Liu & Ou (2021) uses the CWT and a CNN for ECG diagnosis classification purposes. Examples of a CWT scalogram representation, using the Morse wavelet, are illustrated in figures 6.1 and 6.2.

γ characterizes the symmetry of the Morse wavelet. The (demodulate) skewness of the Morse wavelet is equal to 0 when $\gamma = 3$. The Morse wavelets also have the minimum Heisenberg area when $\gamma = 3$. For this reason, $\gamma = 3$ in figures 6.1 and 6.2. P^2 is the time-bandwidth product. It is the measure of localization in time and frequency domains, simultaneously.

Notice that figures 6.1 and 6.2 show a richer representation of the ECG signal. Meaning, more information is present, in comparison to the compact *frequency analysis* and *time-frequency analysis* representations presented in Chapter 3. The cost is that the CWT scalogram image is likely computationally more expensive to be processed. Furthermore, the CWT has no property of orthogonality, and contains a lot of redundancy. Therefore, it is suggested to try other compact ways to represent and classify ECG signal, rather than using CWT scalogram images.

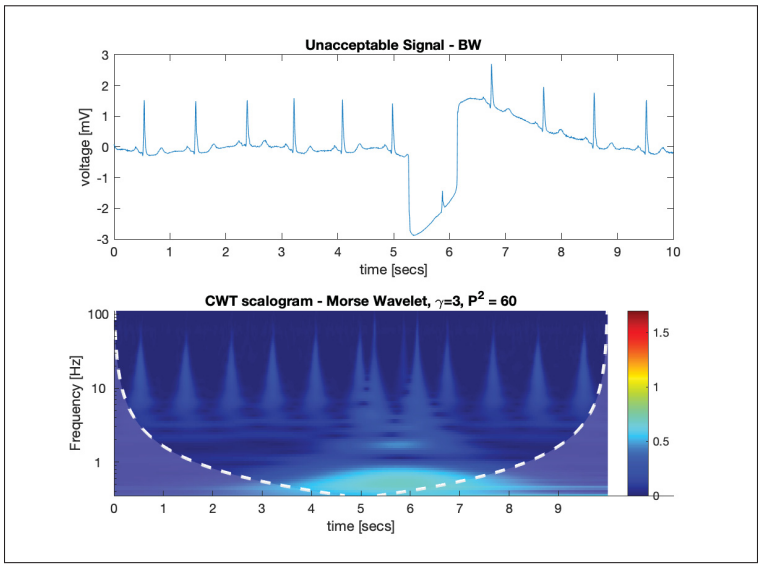


Figure 6.1 Example of an unacceptable ECG signal with BW interference, represented by a CWT scalogram, using the Morse *mother wavelet* with $\gamma = 3$ and $P^2 = 60$

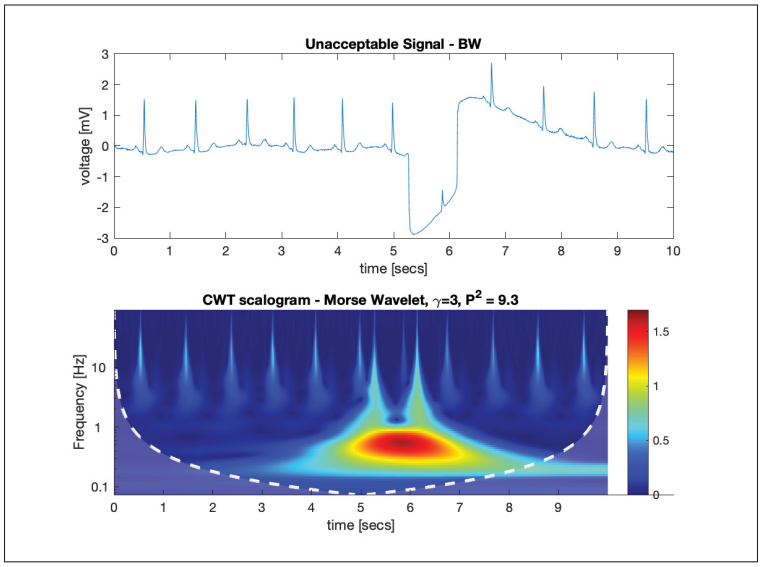


Figure 6.2 Example of an unacceptable ECG signal with BW interference, represented by a CWT scalogram, using the Morse *mother wavelet* with $\gamma = 3$ and $P^2 = 9.3$

According to (Torrence & Compo, 1998b, p. 65), the *wavelet power spectrum* is defined as the squared norm of the detail coefficients. Examples of *wavelet power spectrum* images are

illustrated in figures 6.3 and 6.4. The coefficients are codified to an interval of 8 bits (integers from 1 to 256) for illustration purposes.

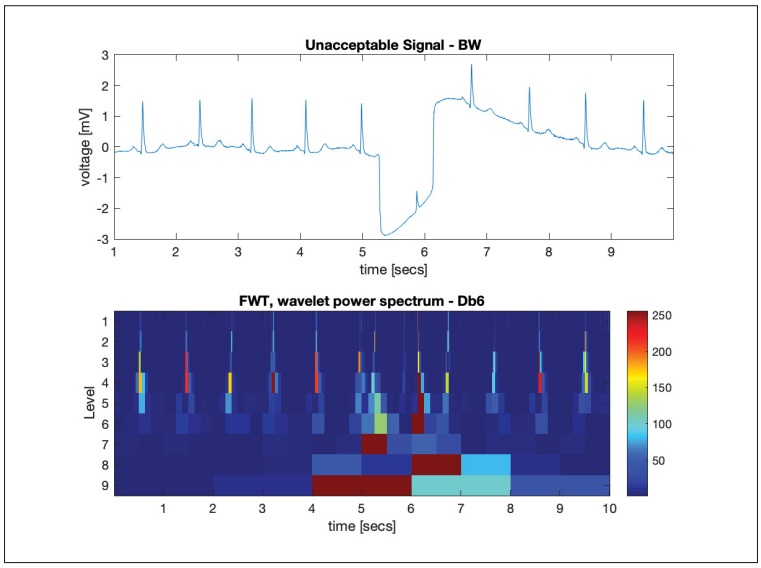


Figure 6.3 Example of an unacceptable ECG signal with BW interference, represented by a DWT *wavelet power spectrum*, using the Db6 *mother wavelet*

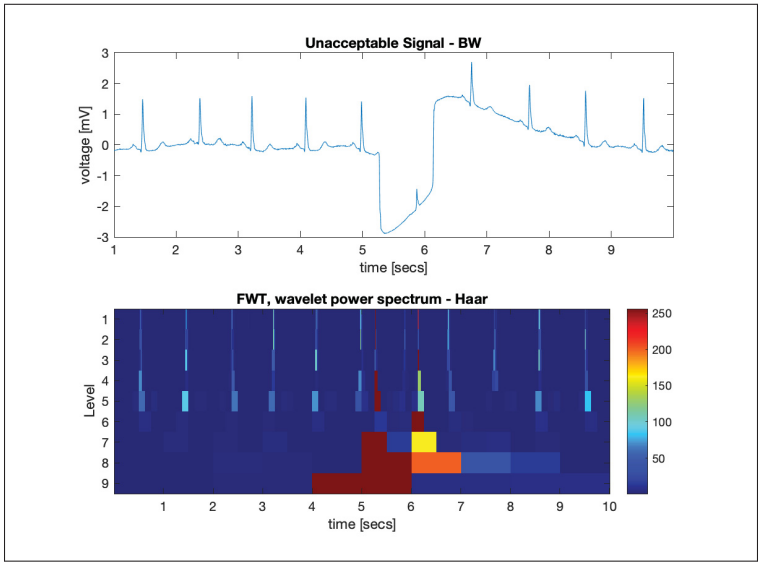


Figure 6.4 Example of an unacceptable ECG signal with BW interference, represented by a DWT *wavelet power spectrum*, using the Haar *mother wavelet*

Notice that such representation is no longer redundant, compared to the CWT, but still is less compact, compared to the *time-frequency analysis* suggested in Chapter 3. An example of a Haar wavelet *time-frequency analysis* representation is presented in Figure 6.5. Just by inspecting Figure 6.5, it seems that such representation does not look very much different from Figure 3.19.

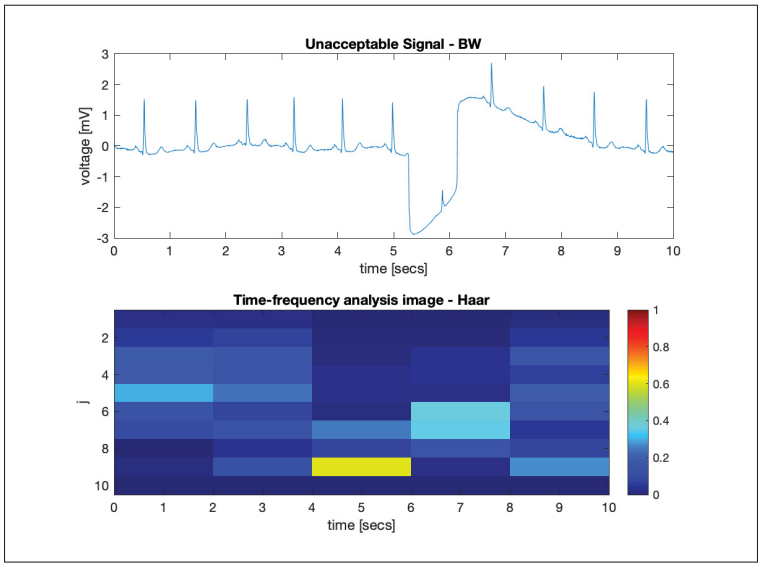


Figure 6.5 Example of an unacceptable ECG signal with high frequency interference, represented by the *time-frequency analysis* using the Haar *mother wavelet*, and demeaning each analyzed signal window individually

Moreover, notice the Haar *wavelet power spectrum* representation contains less information compared to the Db6 *wavelet power spectrum* representation. However, the computation of the Haar wavelet coefficients requires no multiplications. This is a great advantage. Therefore, it is suggested to explore the use of Haar wavelets to effectively represent ECG signals, especially in beat classification problems.

Furthermore, it is suggested to investigate the use Haar wavelets and the *time-frequency analysis* representation, or DWT *wavelet power spectrum* representation, in the same problem of ECG SQA classification, where a CNN, ART2 network, or even a recurrent network (due to time correlation) can classify these images. The reason Haar wavelet representation results were

not presented in this work, is that previous trials, using the Haar wavelets and the *frequency analysis* representation, were not as successful as results from the Db6 *frequency analysis* representation trials, for the problem of single lead ECG SQA. However, this might be different with an effective use of time information, maintaining a good computational cost, since the Haar wavelet coefficients require zero multiplications to be computed.

Finally, the SQA system proposed by this thesis can be used as a control mechanism for other adaptive systems, such as adaptive filters or heart rate estimators, in order to minimize or avoid the introduction of artifacts. Depending on the application, it is necessary to evaluate when filtering is necessary or not. Compromising filtering with SQA can be a quite powerful solution. Therefore, it is recommended to investigate the design of hybrid SQA filtering systems. For instance, this idea is discussed in the work of Oster & Clifford (2015), where SQIs are used to control the uncertainty parameter of a Kalman filter based heart rate estimation algorithm.

APPENDIX I

MOVING AVERAGE CUTOFF FREQUENCY ESTIMATION

This current chapter simply aims to prove equation (4.3). The impulse response of the moving average filter is given as

$$h(n) = \frac{1}{N}\sum_{k=0}^{N-1}\delta(n-k). \tag{A I-1}$$

Where, $\delta(n)$ is a *unit sample* function. The discrete-time Fourier transform (DTFT) of $h(n)$ is:

$$\begin{aligned} H(\omega) &= \sum_{n=-\infty}^{\infty} h(n)e^{-j\omega n} \\ &= \frac{1}{N}\sum_{n=0}^{N-1} e^{-j\omega n} \\ &= \frac{1}{N}\frac{1-e^{-j\omega N}}{1-e^{-j\omega}} \\ &= \frac{1}{N}\frac{e^{-j\omega N/2}}{e^{-j\omega/2}}\frac{e^{j\omega N/2}-e^{-j\omega N/2}}{e^{j\omega/2}-e^{-j\omega/2}} \\ &= \frac{1}{N}\frac{e^{-j\omega N/2}}{e^{-j\omega/2}}\frac{\sin\left(\frac{\omega N}{2}\right)}{\sin\left(\frac{\omega}{2}\right)}. \end{aligned} \tag{A I-2}$$

The squared magnitude of $H(\omega)$ is:

$$|H(\omega)|^2 = \frac{\sin^2\left(\frac{\omega N}{2}\right)}{N^2\sin^2\left(\frac{\omega}{2}\right)}. \tag{A I-3}$$

The cutoff frequency ω_c, given in rad, is defined as:

$$\frac{\sin^2\left(\frac{\omega_c N}{2}\right)}{N^2\sin^2\left(\frac{\omega_c}{2}\right)} = \frac{1}{2}. \tag{A I-4}$$

There is no analytical solution for equation (A I-4). However, it is still possible to approximate ω_c. The relationship between and the frequency Ω, given in rad s^{-1}, and its normalized version (ω), given in rad, is $\Omega T_s = \omega$. Where, T_s is the sampling period. It is known that $T_s = 1/f_s$ and $\Omega = 2\pi f$. Where, f_s is the sampling frequency in Hz, and f is the frequency in Hz. Hence,

$$\omega_c = \frac{2\pi f_c}{f_s}. \tag{A I-5}$$

Where f_c is the cutoff frequency in Hz. Therefore, equation (A I-4) is redefined as:

$$\frac{\sin^2(\alpha N)}{N^2\sin^2(\alpha)} = \frac{1}{2}. \tag{A I-6}$$

Where $\alpha = \frac{\pi f_c}{f_s}$. Furthermore, the order of magnitude of N should be around 10^3. Therefore, $N >> 1$. Moreover, $sin^2(\alpha)$ should only require 1 term from the Taylor series expansion, since $\alpha << 1$, while $sin^2(\alpha N)$ should require more terms, because $\alpha << \alpha N$. Hence, $sin^2(\alpha) \approx \alpha^2$. While,

$$\begin{aligned}
\sin^2(\alpha N) &= \frac{1}{2}(1-\cos(2\alpha N)) \\
&\approx \frac{1}{2}\left(1-\left(1-\frac{(2\alpha N)^2}{2!}+\frac{(2\alpha N)^4}{4!}-\frac{(2\alpha N)^6}{6!}\right)\right) \\
&\approx \frac{1}{2}\left(\frac{(2\alpha N)^2}{2!}-\frac{(2\alpha N)^4}{4!}+\frac{(2\alpha N)^6}{6!}\right) \\
&\approx \alpha^2 N^2 - \frac{\alpha^4 N^4}{3} + \frac{2\alpha^6 N^6}{45}.
\end{aligned} \tag{A I-7}$$

Thus, equation (A I-6) is rewritten as:

$$\frac{2}{45}\beta^2 - \frac{1}{3}\beta + \frac{1}{2} = 0. \tag{A I-8}$$

Where $\beta = \alpha^2 N^2$. It is clear that one of the solutions of equation (A I-8) should provide an approximation for f_c. Therefore, the approximation of f_c should written as:

$$f_c \approx \frac{\alpha_s f_s}{\pi N}. \tag{A I-9}$$

Where α_s is the solution that better approximates f_c. The solutions of equation (A I-8) are $\beta_1 = 2.07295$ and $\beta_2 = 5.42705$. Since, α and N are both positive, $\alpha_1 N = 1.43977$ and $\alpha_2 N = 2.32960$. Among the values of α, the use of α_2 better approximates the true value of f_c, compared to the use of α_1. However, α_2 provides an approximation that is biased. Hence, the true value of f_c was inspected graphically by using the MATLAB function "freqz" and verifying the value of ω_c for $20\log_{10}(|H(\omega_c)|) = 10\log_{10}(\frac{1}{2}) \approx -3.01$ dB. Given that, the term α_2 is adjusted by a multiplication factor. Finally, the approximation of the cutoff frequency f_c, for a large N, is finally given as:

$$f_c \approx \frac{0.3362 f_s}{N}. \tag{A I-10}$$

APPENDIX II

ARTIFICIAL NEURAL NETWORKS

This chapter aims to explain the artificial neural network (ANN) architectures that were used in the application presented by this thesis (Chapter 5). Such architectures consisted on the multilayer perceptron (MLP), the support vector machine (SVM) and the adaptive resonance theory 2 (ART2).

1. Definitions

1. **Pattern:** The feature vector itself. In this section it is represented by $\mathbf{x}_m$ (bold lower case letter), where m stands for the pattern number.

2. **Pattern set or dataset:** defined as the set of feature vectors or patterns. Usually represented by a matrix, containing M patterns column vectors, as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m, \cdots, \mathbf{x}_M]$ (bold upper case letter).

3. **Prototype vector:** defined the class vector where each pattern vector should be mapped to. Such is also known as *target* vector, represented by $\mathbf{d}_m$.

4. **Prototype set:** analogous to the pattern set, a prototype set is a matrix containing all prototype vectors. Hence, $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_m, \cdots, \mathbf{d}_M]$.

5. **Epoch:** An epoch indicates the number of passes of the entire training dataset the machine learning algorithm has completed.

2. Support vector machines

The SVM classifier, similarly to the Perceptron algorithm, is motivated by the optimization of a single neuron network hyperplane separator. In the problem of one-class classification, also known as *anomaly detection*, the ANN tries to identify objects of a specific class among all objects. In this problem, the hyperplane is defined as

$$\mathcal{H} : \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_m) + b = 0. \tag{A II-1}$$

Where $\boldsymbol{\phi}(\cdot)$ maps its argument vector to a more complex non-linear feature space. For mathematical convenience purposes, a variable y_n is introduced, such that $y_{n+} = 1$ and $y_{n-} = -1$. Where, $n+$ stands for positive samples (class 1), and $n-$ stands for negative samples (class 2). It is important to mention that n iterated through the *support vectors*. The *support vectors* are denoted by $\boldsymbol{\phi}(\mathbf{x}_n)$, and are constrained to:

$$\mathcal{M} : y_n[\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n) + b] = 1. \tag{A II-2}$$

The *support vectors* are located in what is so called *margin*. Given the condition (A II-2), the distance between *margins* is maximized according to:

$$\underset{\mathbf{w}}{\operatorname{argmax}} \left[d_{\mathcal{M}} \left(\boldsymbol{\phi} \left(\mathbf{x}_{n+} \right), \boldsymbol{\phi} \left(\mathbf{x}_{n-} \right) \right) \right] = \underset{\mathbf{w}}{\operatorname{argmax}} \left[\frac{\left(\boldsymbol{\phi}(\mathbf{x}_{n+}) - \boldsymbol{\phi}(\mathbf{x}_{n-}) \right) \mathbf{w}}{||\mathbf{w}||_2} \right] = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{2}{||\mathbf{w}||_2}. \tag{A II-3}$$

Where, $||\mathbf{w}||_2$ is the euclidean norm of $\mathbf{w}$, and $d_{\mathcal{M}}(\cdot)$ is a function that takes *support vectors* as argument and computes the distance between *margins*. It turns out that, for mathematical convenience, finding the maximum distance between *margins* can be rewritten as:

$$\underset{\mathbf{w}}{\operatorname{argmax}} \frac{2}{||\mathbf{w}||_2} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{1}{||\mathbf{w}||_2} = \underset{\mathbf{w}}{\operatorname{argmin}} ||\mathbf{w}||_2 = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2} ||\mathbf{w}||_2^2. \tag{A II-4}$$

Hence, optimal solution $\mathbf{w}^{opt}$ is found by using Lagrange multipliers, for minimizing $\frac{1}{2}||\mathbf{w}||_2^2$, subjected to the constrain $y_n[\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n) + b] - 1 = 0$. The derivative of the *Lagrangian* in respect to $\mathbf{w}$ is equalized to zero, and the derivative of the *Lagrangian* in respect to b is also equalized to zero, in order to find the optimal solution $\mathbf{w}^{opt}$.

It turns out that the optimization depends on the *Lagrange multipliers* α_n, the variables y_n, and the dot product of pairs of *support vectors*. Thus, there is no need for mapping the inputs into a more complex non-linear feature space, because such dot products are replaced by non-linear functions called *kernels*, represented by $k(\mathbf{x}_n, \mathbf{x}_m)$.

Unlike the Perceptron, SVMs are capable of achieving non-linear separation through the use of *kernels*. By reformulating the solution, found with the use of *Lagrange multipliers*, to the so called *dual form*, predictions are computed as

$$y_m = \sum_n \alpha_n y_n k(\mathbf{x}_n, \mathbf{x}_m) + b. \tag{A II-5}$$

Where y_m is the class prediction, given the pattern $\mathbf{x}_m$. If $y_m \geq 0$, the input $\mathbf{x}_m$ is considered to be a positive sample. Otherwise it is considered as a negative sample. In this thesis, a Radial-basis function *kernel* is used. Such *kernel* is given as

$$k(\mathbf{x}_n, \mathbf{x}_m) = \exp\left(-||\mathbf{x}_n - \mathbf{x}_m||_2^2\right). \tag{A II-6}$$

3. Multilayer perceptron ANNs

The forward-propagation, computed in an ascending order, from $l = 1$ till $l = L$, is be represented by a matrix form, according to:

$$\boldsymbol{\upsilon}_l(m) = \begin{cases} \mathbf{W}_l \begin{bmatrix} 1 \\ \mathbf{x}_m \end{bmatrix} & \text{for } l = 1; \\ \\ \mathbf{W}_l \begin{bmatrix} 1 \\ \boldsymbol{o}_{l-1}(m) \end{bmatrix} & \text{for } 1 < l \leq L. \end{cases} \tag{A II-7}$$

Where

$$\boldsymbol{o}_l(m) = \boldsymbol{\varphi}_l(\boldsymbol{v}_l(m)). \tag{A II-8}$$

Furthermore, $\mathbf{W}_l$ is the weight of layer l. Each weight matrix has its size equal to $u \times (v + 1)$, where u is the number of neurons in layer l, and v is the number of neurons in layer $l - 1$. Meaning that each matrix includes the bias term. In the case of $l = 1$, v is the number of inputs of the network. Since in the application presented by this thesis $L = 2$, the output vector is $\boldsymbol{o}_2(m)$. Moreover, $\boldsymbol{\varphi}_l(\cdot)$ is known as the *activation function* of the neurons from the l_{th} layer. Such *activation function*, according to Zurada (1992, p. 38), can be described as:

$$\boldsymbol{\varphi}_l(\cdot) = \begin{bmatrix} g_l(\cdot) & 0 & \cdots & 0 \\ 0 & g_l(\cdot) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_l(\cdot) \end{bmatrix}. \tag{A II-9}$$

Where $g_l(\cdot)$ is some generic non-linear function, belonging to the l_{th} layer, that receives a scalar as input. In this thesis, a standard three layer feedforward MLP is used. Where, the input layer (containing N_{in} inputs) is fully connected to the hidden layer (containing N_h nodes), and the hidden layer is fully connected to the output layer (containing N_o output nodes). Hence, $L = 2$.

3.1 Classic back-propagation algorithm

The back-propagation algorithm is a learning method that uses the principle of *gradient descent* to find the optimal solution of the weight coefficients. After computing the forward-propagation, the error is back-propagated by calculating the local gradient vectors using equation (A II-10), at descending order, from $l = L$ till $l = 1$. This process is repeated through the entire M size dataset ($\mathbf{X}$), for E_{poch} epochs, or until the algorithm reaches a stopping criteria. Hence,

$$\boldsymbol{\delta}_l(m) = \begin{cases} \begin{bmatrix} 0 \\ \mathbf{e}(m) \circ \boldsymbol{\varphi}_l'(\boldsymbol{\upsilon}_L(m)) \end{bmatrix} & \text{for } l = L; \\ \boldsymbol{\varphi}_l'(\boldsymbol{\upsilon}_l(m)) \circ \mathbf{W}_{l+1}^T(m)\boldsymbol{\beta}_{l+1}(m) & \text{for } 1 \leq l < L. \end{cases} \tag{A II-10}$$

Where

$$\boldsymbol{\beta}_l(m) = \left[\delta_{l,21}(m), \delta_{l,21}(m), \cdots, \delta_{l,K1}(m)\right]^T, \tag{A II-11}$$

and,

$$\mathbf{e}(m) = \mathbf{d}_m - \boldsymbol{o}_L(m). \tag{A II-12}$$

The element $\delta_{l,ij}(m)$ is the i_{th} row and j_{th} of the column of the vector $\boldsymbol{\delta}_l(m)$, and η is known as the *learning rate* constant. Furthermore, "$\circ$" represent the Hadamard (element wise) product and K stands for the number of rows at $\boldsymbol{\delta}_l(m)$. The weight matrix updates follow the following rules:

$$\mathbf{W}_l(m+1) = \mathbf{W}_l(m) + \Delta\mathbf{W}_l(m) = \begin{cases} \mathbf{W}_l(m) + \eta\boldsymbol{\beta}_l(m) \begin{bmatrix} 1 \\ \boldsymbol{o}_{l-1}(m) \end{bmatrix}^T & \text{for } 1 < l \leq L; \\ \mathbf{W}_l(m) + \eta\boldsymbol{\beta}_l(m) \begin{bmatrix} 1 \\ \mathbf{x}_m \end{bmatrix}^T & \text{for } l = 1. \end{cases} \tag{A II-13}$$

3.2 Back-propagation algorithm with momentum

In order to speed-up and stabilize the learning process, a momentum term is added to the weight update expression. Thus,

$$\mathbf{W}_l(m+1) = \begin{cases} \mathbf{W}_l(m) + \alpha\Delta\mathbf{W}_l(m-1) + \eta\boldsymbol{\beta}_l(m) \begin{bmatrix} 1 \\ \boldsymbol{o}_{l-1}(m) \end{bmatrix}^T & \text{for } 0 < l \leq L; \\ \mathbf{W}_l(m) + \alpha\Delta\mathbf{W}_l(m-1) + \eta\boldsymbol{\beta}_l(m) \begin{bmatrix} 1 \\ \mathbf{x}_m \end{bmatrix}^T & \text{for } l = 0. \end{cases} \quad \text{(A II-14)}$$

Where α is the momentum constant, and $\Delta\mathbf{W}_l(m-1)$ is the previous adjustment of the matrix $\mathbf{W}_l(m)$, for the l_{th} layer.

3.3 Back-propagation algorithm with adaptive learning rate

In this algorithm, the values of η is dynamically controlled, depending on the direction of the gradient. The control algorithm is very simple. At each epoch, if the performance decreases toward the goal (acceptable MSE value), then the learning rate is increased by the factor lr_{inc}. If the performance increases by more than the factor max_{inc}, the learning rate is adjusted by the factor lr_{dec} and the change that increased the performance is not made.

3.4 Softmax layer

In the application presented in this thesis, the output layer of the MLP consists on a *softmax* transfer function, aiming to provide a probabilistic interpretation of the results. Hence, the classifier outputs a probability value $P(A|m)$, where the event A states the input $\mathbf{x}_m$ belongs to

the class of acceptable signals, and a probability value $P(B|m)$, where the event B states the input $\mathbf{x}_m$ belongs to the class of unacceptable signals.

For the presented MLP ANN architecture, the target vector $\mathbf{d}_m$ can be either $[1,0]^T$ or $[0,1]^T$. Hence, $P(A|m) = (1 - P(B|m))$, and it is defined as:

$$P(A|m) = \frac{\exp\left(o_{L,1}(m)\right)}{\exp\left(o_{L,1}(m)\right) + \exp\left(o_{L,2}(m)\right)}. \tag{A II-15}$$

Where $o_{L,1}(m)$ is the first row element of the vector $\boldsymbol{o}_L(m)$, and $o_{L,2}(m)$ is the second row element of the vector $\boldsymbol{o}_L(m)$.

4. Adaptive resonance theory 2

The adaptive resonance theory (ART) network was initially inspired by the developmental physiology of the human visual system, through the development of the *Grossberg network* (Demuth, Beale, De Jess & Hagan, 2014, p. 18_3). Moreover, the ART network is known to serve the purpose of cluster discovery (Zurada, 1992, p. 432). Unlike the previous presented architectures, the ART network uses unsupervised learning for discovering clusters, while both SVM and MLP make the use of supervised learning only.

The first ART architecture, known as ART1 or *Capenter / Grossberg classifier* (Lippmann, 1987, p. 11), was designed with the purpose of clustering and classifying binary images (see Zurada (1992, p. 432-444) and Lippmann (1987, p. 12)). The concept is very simple: images, that are originally formulated in a matrix form, are vectorized and fed to the ART network as an input and, subsequently, assigned to some specific class or cluster (output).

Unlike both the MLP and SVM, ART networks are adaptable and stable, thanks to their *short-term memory* (STM) and *long-term network memory* (LTM) encoded vectors. The problem of the system being able to be receptive to significant new patterns and yet remain stable in response to

old patterns is referred by Grossberg as the *stability / plasticity dilemma* (Demuth *et al.*, 2014, p. 19_2).

Many of the competitive learning networks, developed prior to the ART network, tended to erode the previous learned patterns, when presented to new patterns. Hence, networks developed previous to the ART, tended to lose *learning stability* due to its capability of learning new classes, know as adaptability (or *plasticity*) (Demuth *et al.*, 2014, p. 19_2). This "forgeting" issue is known as *learning instability* (Demuth *et al.*, 2014, p. 19_2).

Both SVM and MLP architectures, when trained with a dataset $\mathbf{X}_1$, and then retrained with a dataset $\mathbf{X}_2$, they tend to forget the encoded memory constructed by $\mathbf{X}_1$. Hence they are subjected to the *stability / plasticity dilemma*. On the other hand, the ART networks were designed to overcome this issue. They are capable of being trained to respond to a new dataset (*plasticity*), say $\mathbf{X}_2$, that does not necessarily contain all the patterns, or classes, encountered in a previously trained dataset, say $\mathbf{X}_1$, while maintaining a stable response learned from $\mathbf{X}_1$ (*learning stability*).

The second generation of ART networks, known as ART2, was designed to overcome the binary input limitation. Hence, such architecture is capable of clustering and classifying grayscale images, with colors encoded as float values in the range of 0 to 1 for example.

It is important to mention that, in the application presented by this thesis, the output of the ART2 is altered by appending a *Grossberg outstar layer* to the output, following the same principle of the counterpropagation network (Zurada, 1992, p. 410-414). Such layer serves the purpose of converting unsupervised learning into supervised learning. In other words, it is responsible of tagging each cluster into some specific class.

The basic ART model is unsupervised in nature and consist on the following modules:

1. F1 layer, comparison or feature representation field. Where inputs are processed.

2. F2 layer, recognition or category representation field. It consists on the clustering units.

3. The reset module, that acts as a control mechanism by deciding if a comparison is matched with a cluster, if the searching procedure needs to be continued, or if a new cluster creation is necessary.

4.1 The training algorithm of the ART2 network

The implementation of the algorithm in MATLAB is described by Sivanandam, Sumathi & Deepa (2006, p. 299-309). Here, the implementation is slightly altered, in order to provide an automatic number of clusters defined by the vigilance parameter, and track training instability with a counting variable. The following table describes the variables used in the algorithm.

Table-A II-1 Summary variables used in the ART2 algorithm

Parameters	**Description**	**Typical value**
N	Number of input units (length of vectorized image).	User defined.
K	Number of cluster units.	Automatically defined.
a	Fixed weight in F1 layer.	10
b	Fixed weight in F1 layer.	10
c	Fixed weight used when testing for a reset.	0.1
d	Activation of winning F2 unit.	0.9
e	Parameter to prevent division by zero.	Very small value.
θ	Noise suppression parameter.	$\frac{1}{\sqrt{N}}$
α	Learning rate.	Small value, less then 1.
ρ	Vigilance parameter.	Small value, less then 1.
$\mathbf{t}_j$	j_{th} cluster memory or column of the top-down LTM traces (**T**).	$\left[0\right]$ (initialized as a vector of zeros)
$\mathbf{b}_j$	j_{th} cluster memory or column of the bottom-up LTM traces (**B**).	$\left[\frac{1}{(1-d)\sqrt{N}}\right]$ (initialized as a vector of constant values)
β	Instability tracker.	Counting variable.

Table-A II-1 Summary variables used in the ART2 algorithm *(cont'd)*

Parameters	**Description**	**Typical value**
γ	Reset module.	Boolean variable.
N_{it}	Number iterations during memorization.	1 (for slow learning). 10 (for faster learning).

The steps that describe the ART2 training algorithm are enumerated bellow.

Step 1: Initialize the parameters. In particular, the parameters c and d are constrained to

$$\frac{cd}{1-d} \leq 1.$$

In this thesis, parameters are initialized as:

$$a = 10 \quad b = 10 \quad c = 0.1 \quad d = 0.9 \quad e = 0.001 \qquad \alpha = 0.6 \quad \rho = 0.94$$
$$N = 50 \quad K = 1 \quad \theta = \frac{1}{\sqrt{N}} \quad \mathbf{T} = \left[0\right] \quad \mathbf{B} = \left[\frac{1}{(1-d)\sqrt{N}}\right] \quad N_{it} = 1.$$

It is important to mention that both **T** and **B** have both the size of $N \times K$.

Step 2: Begin the epoch by permuting the columns of the pattern set **X**. Assign $i = 1$.

Step 3: Assign **x** to be equal to the i_{th} column of **X**. Moreover, assign $\beta = 0$.

Step 4: Update the F1 layer by using the equations listed below, in the same order as presented. Equations listed side-by-side can be computed in parallel.

$$\mathbf{s} = \frac{\mathbf{x}}{e + ||\mathbf{x}||_2}$$

$$\mathbf{v} = \mathbf{\Gamma}(\mathbf{s})$$

$$\mathbf{u} = \frac{\mathbf{v}}{e + ||\mathbf{v}||_2}$$

$$\mathbf{w} = \mathbf{x} + a\mathbf{u} \qquad\qquad \mathbf{p} = \mathbf{u}$$

$$\mathbf{s} = \frac{\mathbf{w}}{e + ||\mathbf{w}||_2} \qquad\qquad \mathbf{q} = \frac{\mathbf{p}}{e + ||\mathbf{p}||_2}$$

$$\mathbf{v} = \mathbf{\Gamma}(\mathbf{s}) + b\mathbf{\Gamma}(\mathbf{q}).$$

Where

$$\mathbf{\Gamma}(\cdot) = \begin{bmatrix} g(\cdot) & 0 & \cdots & 0 \\ 0 & g(\cdot) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g(\cdot) \end{bmatrix}.$$

Typically,

$$g(x) = \begin{cases} x, & \text{for } x \geq \theta; \\ 0, & \text{for } x < \theta. \end{cases}$$

Step 5: Compute the F2 layer by executing $\mathbf{y} = \mathbf{B}^T\mathbf{p}$ and set $\gamma = 1$.

Step 6: Find the row where the maximum element of $\mathbf{y}$ is located. If there is a tie, choose the lowest value. Hence, $j = \max_{\text{row}}(\mathbf{y})$.

Step 7: Update the F1 layer elements by computing the following:

$$\mathbf{u} = \frac{\mathbf{v}}{e + ||\mathbf{v}||_2}$$
$$\mathbf{p} = \mathbf{u} + d\mathbf{t}_j$$
$$\mathbf{r} = \frac{\mathbf{u} + c\mathbf{p}}{e + ||\mathbf{u}||_2 + c||\mathbf{p}||_2}.$$

Step 8: If $||\mathbf{r}||_2 \geq (\rho - e)$, compute the following:

$$\mathbf{w} = \mathbf{x} + a\mathbf{u}$$
$$\mathbf{s} = \frac{\mathbf{w}}{e + ||\mathbf{w}||_2} \qquad \mathbf{q} = \frac{\mathbf{p}}{e + ||\mathbf{p}||_2}$$
$$\mathbf{v} = \mathbf{\Gamma}(\mathbf{s}) + b\mathbf{\Gamma}(\mathbf{q}) \qquad \gamma = 0.$$

Step 9: Else, if $||\mathbf{r}||_2 < (\rho - e)$, -1 is assigned to the j_{th} row of $\mathbf{y}$. Hence, $y_j = -1$. After assigning such value, if all elements of $\mathbf{y}$ are equal to -1, then a new class needs to be created by computing the following:

$$\beta^{new} = \beta + 1 \qquad K^{new} = K + 1$$
$$\mathbf{t}_K = \left[0\right] \qquad \mathbf{b}_K = \left[\frac{1}{(1-d)\sqrt{N}}\right] \qquad j = K$$
$$\mathbf{T}^{new} = \left[\mathbf{T} \quad \mathbf{t}_K\right] \qquad \mathbf{B}^{new} = \left[\mathbf{B} \quad \mathbf{b}_K\right].$$

It is important to emphasize that the length of both vectors $\mathbf{t}_K$ and $\mathbf{b}_K$ is N.

Step 10: If $\beta > 1$, an instability occurred. Thus, the training failed, the parameters need to be readjusted and the training should be restarted from the very beginning. That being said, usually the value of either θ or ρ, or both, can be attenuated to guarantee a stable training.

Step 11: If $\gamma = 1$, the algorithm goes back to **Step 6**. Else, the algorithm continues to **Step 12**, while assigning $n_o = 1$.

Step 12: The weights are updated by computing the following:

$$\mathbf{t}_j^{new} = \alpha d\mathbf{u} + (1 + \alpha d(d-1))\,\mathbf{t}_j \qquad \mathbf{b}_j^{new} = \alpha d\mathbf{u} + (1 + \alpha d(d-1))\,\mathbf{b}_j.$$

Step 13: If $n_o < N_{it}$, assign $n_o^{new} = n_o + 1$, update the F1 layer, then return to **Step 12**. The F1 layer is updated by computing the following:

$$\mathbf{u} = \frac{\mathbf{v}}{e + ||\mathbf{v}||_2}$$

$$\mathbf{w} = \mathbf{x} + a\mathbf{u}$$

$$\mathbf{p} = \mathbf{u} + d\mathbf{t}_j \qquad\qquad \mathbf{s} = \frac{\mathbf{w}}{e + ||\mathbf{w}||_2}$$

$$\mathbf{q} = \frac{\mathbf{p}}{e + ||\mathbf{p}||_2}$$

$$\mathbf{v} = \mathbf{\Gamma}(\mathbf{s}).$$

Step 14: Assign $i^{new} = i + 1$. If i is less or equal to the number of columns from $\mathbf{X}$, return to **Step 3**. Else, start a new epoch by returning to **Step 2**. The training algorithm stops whenever the maximum number of epochs is reached.

4.2 The recall procedure of the ART2 network

The recall procedure is very similar to the training algorithm. Where the only difference is that the "vigilance test" is no longer present, neither the weights are updated (Zurada, 1992, p. 437). It consists on the following steps:

Step 1: An input vector $\mathbf{x}$ is assumed to be given as an input. Update the F1 layer by using the equations listed bellow, in the same order as presented. Hence,

$$\mathbf{s} = \frac{\mathbf{x}}{e + ||\mathbf{x}||_2}$$
$$\mathbf{v} = \mathbf{\Gamma}(\mathbf{s})$$
$$\mathbf{u} = \frac{\mathbf{v}}{e + ||\mathbf{v}||_2}$$
$$\mathbf{p} = \mathbf{u}.$$

Step 2: Compute the F2 layer calculating $\mathbf{y} = \mathbf{B}^T\mathbf{p}$.

Step 3: Find the row where the maximum element of $\mathbf{y}$ is located. If there is a tie, chose the lowest value. Hence, the output is $j = \max_{\text{row}}(\mathbf{y})$.

4.3 Unsupervised to supervised learning conversion

The conversion of unsupervised learning into supervised learning is performed by appending a *Grossberg outstar layer* to the output of the ART2 network. Here, the *Grossberg outstar* weight matrix $\mathbf{W}_g$ has a size of $K \times 2$. Hence, the target vector $\mathbf{d}$ can be either $[1, 0]^T$ or $[0, 1]^T$. $\mathbf{W}_g$ is updated by using the *outstar learning rule* (Zurada, 1992, p. 413). The matrix $\mathbf{W}_g$ has the following format:

$$\mathbf{W}_g = \left[\mathbf{w}_{g,1}, \quad \mathbf{w}_{g,2}, \quad \cdots, \quad \mathbf{w}_{g,K}\right]. \tag{A II-16}$$

Given a trained $\mathbf{W}_g$, the output class of the ART2 network, given an input $\mathbf{x}$, must be encoded in a vector form. The row (j), where the maximum value of $\mathbf{y}$ is located, is encoded to a vector z, using a method called *one hot encoding*. The classification output of the *Grossberg outstar layer* appended to the ART2 network, given the vector z, is described as

$$\text{class}_{out} = \max_{\text{row}}(\mathbf{W}_g^T \mathbf{z}). \tag{A II-17}$$

The learning procedure is executed according to the following steps:

Step 1: Initialize $\mathbf{W}_g$ as a $K \times 2$ matrix of random numbers. Assign a very small value to Δ. Here $\Delta = 0.0001$.

Step 2: Begin the epoch by permuting the columns of the pattern set $\mathbf{X}$ and the prototype set $\mathbf{D}$ at the same order. Assign $i = 1$.

Step 3: Assign $\mathbf{x}$ to be equal to the i_{th} column of $\mathbf{X}$ and $\mathbf{d}$ to be equal to the i_{th} column of $\mathbf{D}$.

Step 4: Input $\mathbf{x}$ to the ART2 network and output the number j (the row where the maximum value of $\mathbf{y}$ is located).

Step 5: Normalize all columns of $\mathbf{W}_g$.

Step 6: Update $\mathbf{W}_g$ by using the *outstar learning rule*, according to:

$$\mathbf{w}_{g,j}^{new} = \mathbf{w}_{g,j} + \Delta(\mathbf{d} - \mathbf{w}_{g,j}).$$

Step 7: Assign $i^{new} = i + 1$. If i is less or equal to the number of columns from $\mathbf{X}$, return to **Step 3**. Else, start a new epoch by returning to **Step 2**. The training algorithm stops whenever the maximum number of epochs is reached.

5. Conclusion

This chapter presented several ANN architectures that are used in the application presented by Chapter 5. The main objective of this chapter was to clarify the mathematics and procedures used in the implementation of the machine learning based classification SQA systems.

APPENDIX III

PATIENTS CONDITIONS FROM THE SIG.TEST DATA-SET

Table-A III-1 summarizes a description of patients conditions from the SIG.test data-set. Each condition was analyzed by a specialized cardiology during a clinical study performed by SIG.NUM.

Table-A III-1 Description of patients conditions from the SIG.test data-set

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
3	ABNORMAL ECG LONG QT INTERVAL SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
4	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
5	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
6	ABNORMAL ECG INFERIOR INFARCT AGE UNDETERMINED NONSPECIFIC ST and T WAVE ABNORMALITY WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	ATRIAL FIBRILLATION
7	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
8	ABNORMAL ECG LEFT ANTERIOR FASCICULAR BLOCK SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
9	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
10	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
11	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
12	ABNORMAL ECG POSSIBLE LEFT VENTRICULAR HYPERTROPHY TALL T WAVES, POSSIBLE HYPER-KALEMIA SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
13	ABNORMAL ECG INFERIOR IN-FARCT, POSSIBLY ACUTEST DEPRES-SION ST ELEVATION CONSISTENT WITH EPICARDIAL INJURY, PERI-CARDITIS, OR EARLY REPOLARIZA-TION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
14	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
15	ABNORMAL ECG SINUS RHYTHM WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM
16	ABNORMAL ECG POSSIBLE AN-TEROLATERAL INFARCT ST and T WAVE ABNORMALITY CONSISTENT WITH ANTEROLATERAL ISCHEMIA SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
17	ABNORMAL ECG INCOMPLETE RIGHT BUNDLE BRANCH BLOCK LEFT ANTERIOR FASCICULAR BLOCK SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
18	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
19	LEFT AXIS DEVIATION ABNORMAL ECG NONSPECIFIC INTRAVENTRICULAR CONDUCTION DELAY POSSIBLE RIGHT VENTRICULAR HYPERTROPHY NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
20	LEFT AXIS DEVIATION ABNORMAL ECG NONSPECIFIC INTRAVENTRICULAR CONDUCTION DELAY LATERAL INFARCT, AGE UNDETERMINED LONG QT INTERVAL NONSPECIFIC ST and T WAVE ABNORMALITY WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	ATRIAL FIBRILLATION
21	LEFT AXIS DEVIATION ABNORMAL ECG INFERIOR INFARCT, AGE UNDETERMINED SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
22	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
23	ABNORMAL ECG RIGHT BUNDLE BRANCH BLOCK LONG QT INTERVAL SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
24	ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
25	LEFT AXIS DEVIATION ABNORMAL ECG POSSIBLE LEFT VENTRICULAR HYPERTRO PHYANTERIOR INFARCT, AGE UNDETERMINED NONSPECIFIC ST and T WAVE ABNORMALITY	ABNORMAL ECG	ATRIAL FLUTTER WITH VARYING CONDUCTION
26	ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK ATRIAL RHYTHM	ABNORMAL ECG	ATRIAL RHYTHM
27	LEFT AXIS DEVIATION ABNORMAL ECG INFERIOR INFARCT, POSSIBLY ACUTE LONG QT INTERVAL ST DEPRESSION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
28	NORMAL ECG INFERIOR INFARCT POSSIBLE SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
29	ABNORMAL ECG ANTERIOR INFARCT, POSSIBLY ACUTEST ELEVATION CONSISTENT WITH EPICARDIAL INJURY, PERICARDITIS, OR EARLY REPOLARIZATION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
30	ABNORMAL ECG LOW QRS VOLTAGE IN LIMB LEADS INFERIOR INFARCT	ABNORMAL ECG	SUPRA-VENTRICULAR TACHYCARDIA
31	ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK LONG QT INTERVAL NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM WITH FIRST DEGREE AV BLOCK WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM
32	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
33	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
34	ABNORMAL ECG RIGHT BUNDLE BRANCH BLOCK LONG QT INTERVAL SINUS RHYTHM WITH FIRST DEGREE AV BLOCK	ABNORMAL ECG	SINUS RHYTHM
35	ABNORMAL ECG NONSPECIFIC INTRAVENTRICULAR CONDUCTION DELAY LONG QT INTERVAL	ABNORMAL ECG	SINUS BRADYCARDIA (LESS THAN 40)
36	ABNORMAL ECG ANTEROLATERAL INFARCT EVOLVING SINUS RHYTHM POSSIBLE LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS RHYTHM
37	ABNORMAL ECG POSSIBLE ANTERIOR INFARCT ST and T WAVE ABNORMALITY CONSISTENT WITH ANTEROLATERAL ISCHEMIA ST and T WAVE ABNORMALITY CONSISTENT WITH INFERIOR ISCHEMIA SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
38	ABNORMAL ECG ST DEPRESSION SINUS TACHYCARDIA	ABNORMAL ECG	SINUS TACHYCARDIA
39	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
40	ABNORMAL ECG ST DEPRESSION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
41	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM ELECTRONIC VENTRICULAR PACEMAKER	ABNORMAL ECG	SINUS RHYTHM
42	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM ELECTRONIC VENTRICULAR PACEMAKER	ABNORMAL ECG	SINUS RHYTHM
43	ABNORMAL ECG NON SPECIFIC INTRAVENTRICULAR CONDUCTION DELAY	ABNORMAL ECG	ATRIAL FIBRILLATION
44	NORMAL ECG SINUS RHYTHMWITH SUPRAVENTRICULAR PREMATURE COMPLEXES	NORMAL ECG	SINUS RHYTHM
45	ABNORMAL ECG POSSIBLE LEFT VENTRICULAR HYPERTROPHY ST DEPRESSION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
46	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
47	LEFT AXIS DEVIATION ABNORMAL ECG SINUS RHYTHM LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
48	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
49	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
50	ABNORMAL ECG LEFT VENTRICULAR HYPERTROPHY WITH REPOLARIZATION ABNORMALITY NONSPECIFIC ST ELEVATION SINUS RHYTHM WITH FIRST DEGREE AV BLOCK	ABNORMAL ECG	SINUS RHYTHM
51	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY	ABNORMAL ECG	ATRIAL FIBRILLATION (WITH IDIOVENTRICULAR RHYTHM)
52	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
53	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
54	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
55	ABNORMAL ECG	ABNORMAL ECG	ATRIAL FIBRILLATION
56	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
57	ABNORMAL ECG	ABNORMAL ECG	ATRIAL FIBRILLATION
58	ABNORMAL ECG INFERIOR INFARCT, AGE UNDETERMINED SINUS RHYTHM WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM
59	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
60	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
61	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
62	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
63	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
64	ABNORMAL ECG SINUS RHYTHM LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
65	ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
66	LEFT AXIS DEVIATION ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK LONG QT INTERVAL SINUS RHYTHM WITH FIRST DEGREE AV BLOCK	ABNORMAL ECG	SINUS RHYTHM
67	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
68	LEFT AXIS DEVIATION ABNORMAL ECG LEFT BUNDLE BRANCH BLOCK	ABNORMAL ECG	ATRIAL FIBRILLATION
69	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
70	ABNORMAL ECG INFERIOR INFARCT WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	ATRIAL FIBRILLATION
71	ABNORMAL ECG LEFT ANTERIOR FASCICULAR BLOCK POSSIBLE LEFT VENTRICULAR HYPERTROPHY ANTERIOR INFARCT, AGE UNDETERMINED	ABNORMAL ECG	ATRIAL FIBRILLATION
72	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
73	ABNORMAL ECG	ABNORMAL ECG	ATRIAL FIBRILLATION
74	RIGHT AXIS DEVIATION ABNORMAL ECG SINUS RHYTHM WITH FIRST DEGREE AV BLOCK LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS RHYTHM
75	LEFT AXIS DEVIATION ABNORMAL ECG SINUS TACHYCARDIA	ABNORMAL ECG	SINUS TACHYCARDIA
76	ABNORMAL ECG LEFT ANTERIOR FASCICULAR BLOCK ANTEROLATERAL INFARCT, POSSIBLY ACUTE SINUS TACHYCARDIA LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS TACHYCARDIA (POSSIBLY ACUTE)
77	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
78	ABNORMAL ECG ANTERIOR INFARCT	ABNORMAL ECG	ATRIAL FLUTTER
79	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
80	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
81	ABNORMAL ECG RIGHT BUNDLE BRANCH BLOCK SINUS RHYTHM WITH ATRIAL PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM
82	ABNORMAL ECG SINUS RHYTHM WITH FIRST DEGREE AV BLOCK WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM
83	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
84	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
85	RIGHT AXIS DEVIATION ABNORMAL ECG RIGHT VENTRICULAR HYPERTROPHY WITH REPOLARIZATION ABNORMALITY SINUS TACHYCARDIA POSSIBLE LEFT ATRIAL ENLARGEMENT	ABNORMAL ECG	SINUS TACHYCARDIA
86	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
87	ABNORMAL ECG SINUS RHYTHM WITH VENTRICULAR PREMATURE COMPLEXES	ABNORMAL ECG	SINUS RHYTHM

Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
88	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
89	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
90	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM
91	ABNORMAL ECG LONG QT INTERVAL SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
92	ABNORMAL ECG LEFT ANTERIOR FASCICULAR BLOCK RIGHT BUNDLE BRANCH BLOCK	ABNORMAL ECG	TACHYCARDIA
93	ABNORMAL ECG SINUS TACHYCARDIA	ABNORMAL ECG	SINUS TACHYCARDIA
94	ABNORMAL ECG SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
95	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
96	ABNORMAL ECG NONSPECIFIC INTRAVENTRICULAR CONDUCTION DELAY SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
97	NORMAL ECG SINUS RHYTHM	NORMAL ECG	SINUS RHYTHM

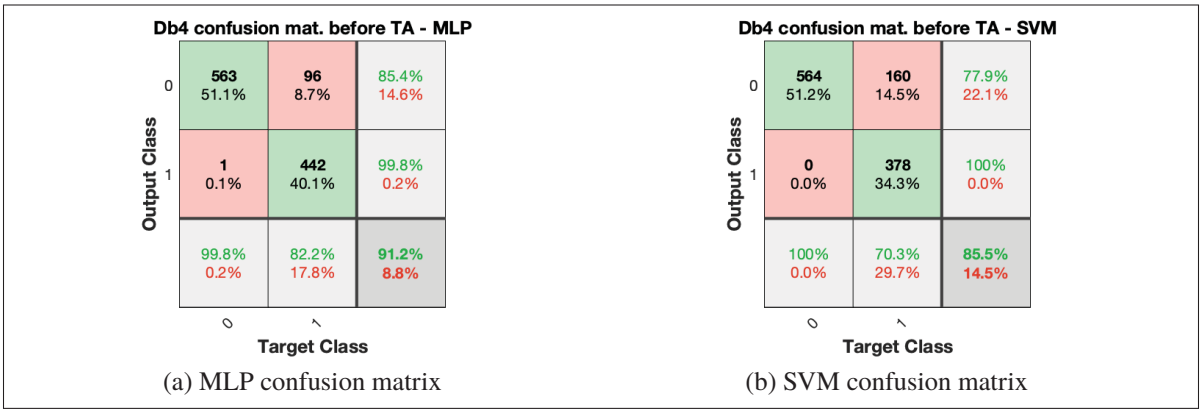
Table-A III-1 Description of patients conditions from the SIG.test data-set *(cont'd)*

Patient number	**ECG specification**	**Normal / abnormal**	**Rhythm**
98	ABNORMAL ECG ST DEPRESSION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
99	ABNORMAL ECG NON SPECIFIC ST ELEVATION SINUS TACHYCARDIA	ABNORMAL ECG	SINUS TACHY-CARDIA
100	ABNORMAL ECG RIGHT BUNDLE BRANCH BLOCK SINUS RHYTHM WITH 2ND DEGREE AV BLOCK, MO-BITZ TYPE II	ABNORMAL ECG	SINUS RHYTHM
102	ABNORMAL ECG ANTEROLATERAL INFARCT, POSSIBLY ACUTEST EL-EVATION CONSISTENT WITH EPI-CARDIAL INJURY, PERICARDITIS, OR EARLY REPOLARIZATION SINUS RHYTHM	ABNORMAL ECG	SINUS RHYTHM
103	ABNORMAL ECG NONSPECIFIC ST and T WAVE ABNORMALITY	ABNORMAL ECG	ATRIAL FIBRIL-LATION (WITH RAPID VENTRIC-ULAR RE-SPONSE)

APPENDIX IV

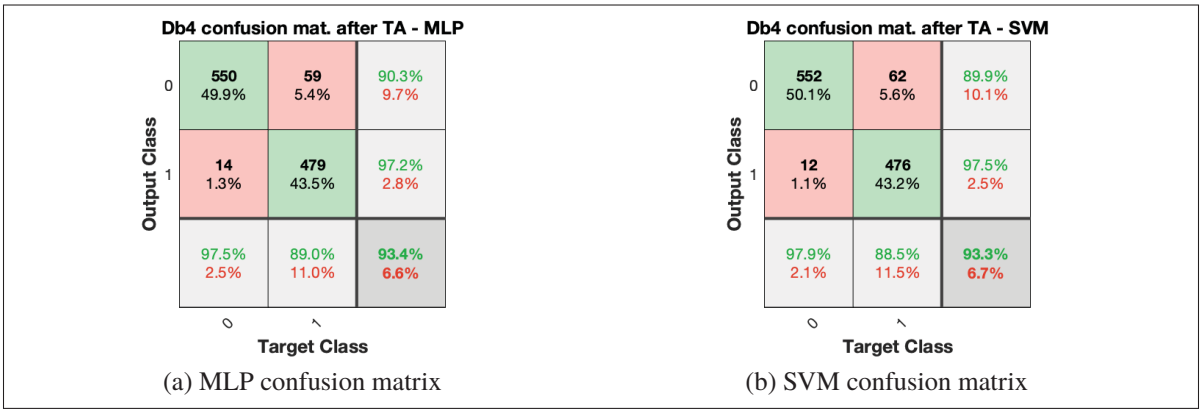
ADDITIONAL RESULTS

1. Best trained nets (ANNs) using the Db4 *frequency analysis* feature representation



(a) MLP confusion matrix (b) SVM confusion matrix

Figure-A IV-1 Confusion matrix from the best trained MLP or SVM, before FP and TA, using a Db4 *frequency analysis* feature representation, tested on the *SIG.test data-set*



(a) MLP confusion matrix (b) SVM confusion matrix

Figure-A IV-2 Confusion matrix from the best trained MLP or SVM, before FP, after TA, using a Db4 *frequency analysis* feature representation, tested on the *SIG.test data-set*

2. Impact of threshold change in probability values

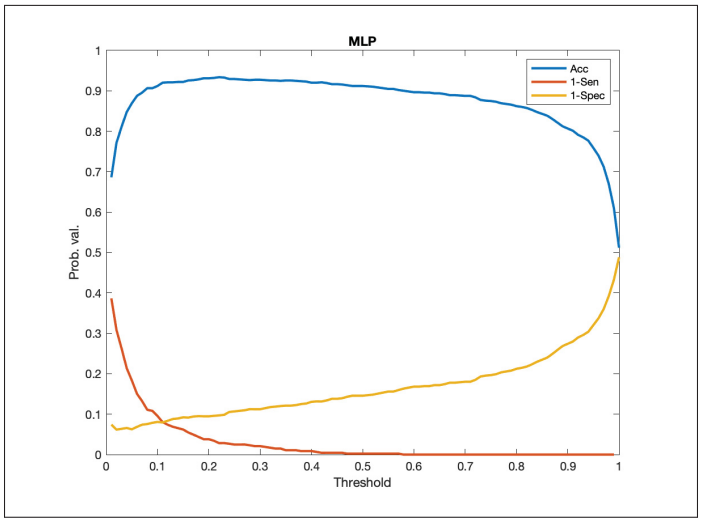


Figure-A IV-3 Probability values vs Db4 *frequency analysis* representation MLP classifier threshold

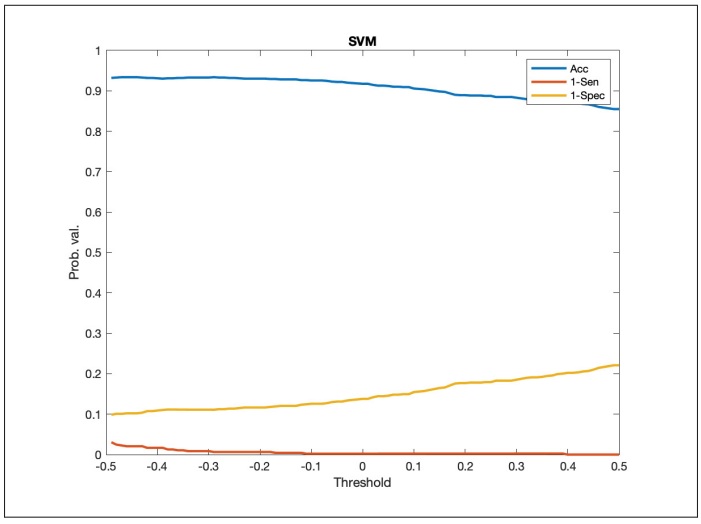


Figure-A IV-4 Probability values vs Db4 *frequency analysis* representation SVM classifier threshold

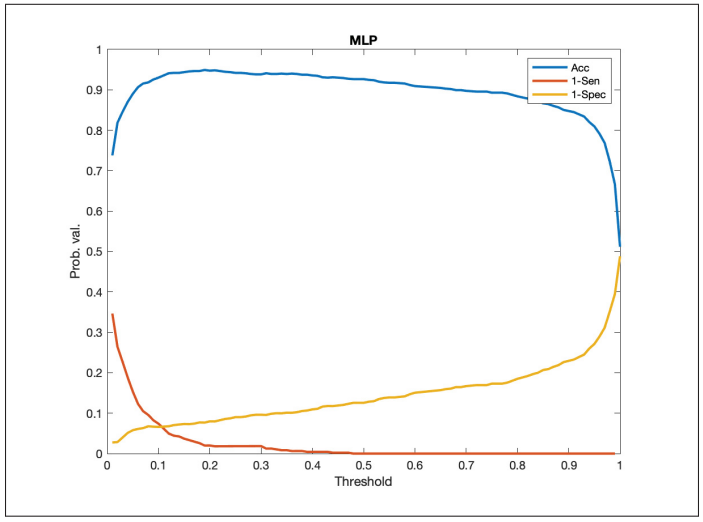


Figure-A IV-5 Probability values vs Db6 *frequency analysis* representation MLP classifier threshold

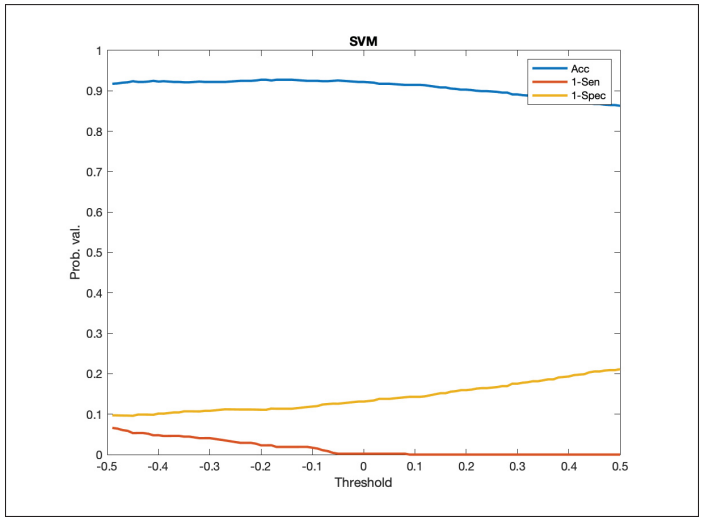


Figure-A IV-6 Probability values vs Db6 *frequency analysis* representation SVM classifier threshold

LIST OF REFERENCES

Addison, P. S. (2005). Wavelet transforms and the ECG: a review. *Physiological Measurement*, 26(5), R155-R199. doi: 10.1088/0967-3334/26/5/R01.

Ariño, M. A., Morettin, P. A. & Vidakovic, B. (2004). Wavelet scalograms and their applications in economic time series. *Brazilian Journal of Probability and Statistics*, 18(1), 37–51. Consulted at http://www.jstor.org/stable/43601038.

Baba, A. & Burke, M. J. (2008). Measurement of the electrical properties of ungelled ECG electrodes. *International Journal of Biology and Biomedical Engineering*, 2(1), 89–97. Consulted at http://www.naun.org/journals/bio/bio-17.pdf.

Behar, J., Oster, J., Li, Q. & Clifford, G. D. (2012). A single channel ECG quality metric. *2012 Computing in Cardiology*, 39, 381–384.

Behar, J., Oster, J., Li, Q., Clifford, G. D., Qiao Li, Clifford, G. D., Li, Q., Clifford, G. D., Qiao Li & Clifford, G. D. (2013). ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Transactions on Biomedical Engineering*, 60(6), 1660–1666. doi: 10.1109/TBME.2013.2240452.

Blahut, R. E. (2010). *Fast Algorithms for Signal Processing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511760921.

Blanco-Velasco, M., Weng, B. & Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*, 38(1), 1–13. doi: 10.1016/j.compbiomed.2007.06.003.

Böck, C. (2015). *ECG signal analysis based on the wavelet transform*. (Ph.D. thesis). Consulted at https://www.jku.at/fileadmin/gruppen/183/Docs/Finished_Theses/Master_Thesis_Boeck.pdf.

Bohm, A., Antink, C. H., Leonhardt, S. & Teichmann, D. (2015, 9). Determining the connection between capacitively coupled electrocardiography data and the ground truth. *2015 Computing in Cardiology Conference (CinC)*, 42, 677–680. doi: 10.1109/CIC.2015.7411001.

Brunton, S. L. & Kutz, J. N. (2019). *Data-Driven Science and Engineering*. Cambridge University Press. doi: 10.1017/9781108380690.

Can Ye, Kumar, B. V. K. V. & Coimbra, M. T. (2012). Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals. *IEEE Transactions on Biomedical Engineering*, 59(10), 2930–2941. doi: 10.1109/TBME.2012.2213253.

Castro, I. D., Morariu, R., Torfs, T., Van Hoof, C. & Puers, R. (2016, 5). Robust wireless capacitive ECG system with adaptive signal quality and motion artifact reduction. *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 53(9), 1–6. doi: 10.1109/MeMeA.2016.7533802.

Castro, I. D., Varon, C., Torfs, T., Van Huffel, S., Puers, R. & Van Hoof, C. (2018). Evaluation of a Multichannel Non-Contact ECG System and Signal Quality Algorithms for Sleep Apnea Detection and Monitoring. *Sensors (Basel, Switzerland)*, 18(2), 577. doi: 10.3390/s18020577.

Chang, K.-M. & Liu, S.-H. (2011). Gaussian Noise Filtering from ECG by Wiener Filter and Ensemble Empirical Mode Decomposition. *Journal of Signal Processing Systems*, 64(2), 249–264. doi: 10.1007/s11265-009-0447-z.

Chi, Y. M., Jung, T.-P. & Cauwenberghs, G. (2010). Dry-Contact and Noncontact Biopotential Electrodes: Methodological Review. *IEEE Reviews in Biomedical Engineering*, 3, 106–119. doi: 10.1109/RBME.2010.2084078.

Chun-lin, L. (2010). A Tutorial of the Wavelet Transform.

CINTRA, R. J. d. S. (2005). *Aproximação espectral e construção de wavelets com aplicações em eletrogastrografia*. (Ph.D. thesis, Universidade Federal de Pernambuco). Consulted at https://repositorio.ufpe.br/handle/123456789/5625.

Clifford, G. D., Behar, J., Li, Q. & Rezek, I. (2012). Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological Measurement*, 33(9), 1419–1433. doi: 10.1088/0967-3334/33/9/1419.

Clifford, G. D., Azuaje, F. & McSharry, P. (2006). Advanced methods and tools for ECG data analysis. Boston: Artech House. Consulted at http://www.books24x7.com/marc.asp?bookid=27221.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2, 303–314.

Demuth, H. B., Beale, M. H., De Jess, O. & Hagan, M. T. (2014). *Neural Network Design* (ed. 2nd). Stillwater, OK, USA: Martin Hagan.

Freitas, D. R. R., Inocêncio, A. V. M., Lins, L. T., Alves, G. J. & Benedetti, M. A. (2019). A Parallel Implementation of the Discrete Wavelet Transform Applied to Real-Time EEG Signal Filtering (pp. 17–23). doi: 10.1007/978-981-13-2517-5{_}3.

Friesen, G., Jannett, T., Jadallah, M., Yates, S., Quint, S. & Nagle, H. (1990). A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1), 85–98. doi: 10.1109/10.43620.

Geddes, L. A. & Valentinuzzi, M. E. (1973). Temporal changes in electrode impedance while recording the electrocardiogram with "Dry" electrodes. *Annals of Biomedical Engineering*, 1(3), 356–367. doi: 10.1007/BF02407675.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-k. & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23). doi: 10.1161/01.CIR.101.23.e215.

Gonzalez, R. C. & Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc.

Gonzalez, R. C. & Woods, R. E. (2008). Summary for Policymakers. In Intergovernmental Panel on Climate Change (Ed.), *Climate Change 2013 - The Physical Science Basis* (vol. 3rd Editio, pp. 1–30). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107415324.004.

Häggström, M. (2012). Placement of the precordial electrodes. Consulted at https://en.wikipedia.org/wiki/Electrocardiography#/media/File:Precordial_leads_in_ECG.png.

Hamilton, P. S. (2002). Open Source ECG Analysis Software Documentation. Consulted at http://eplimited.com/osea13.pdf.

Haykin, S. S. (2009). *Neural networks and learning machines* (ed. Third). Pearson Education.

Haykin, S. S. (2013). *Adaptive Filter Theory* (ed. 5th). Pearson. Consulted at https://www.pearson.com/uk/educators/higher-education-educators/program/Haykin-Adaptive-Filter-Theory-International-Edition-5th-Edition/PGM1033861.html?tab=overview.

Hayn, D., Jammerbund, B. & Schreier, G. (2012). QRS detection based ECG quality assessment. *Physiological Measurement*, 33(9), 1449–1461. doi: 10.1088/0967-3334/33/9/1449.

He, Z. (2016). *Wavelet Analysis and Transient Signal Processing Applications for Power Systems*. Singapore: John Wiley & Sons Singapore Pte. Ltd. doi: 10.1002/9781118977019.

Heuer, S., Chiriac, S., Kirst, M., Gharbi, A. & Stork, W. (2011). Signal Quality Assessment for Capacitive ECG Monitoring Systems using Body-sensor-impedance. *BIOSIGNALS 2011 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pp. 454–458.

Hou, Z., Xiang, J., Dong, Y., Xue, X., Xiong, H. & Yang, B. (2018). Capturing Electrocardiogram Signals from Chairs by Multiple Capacitively Coupled Unipolar Electrodes. *Sensors*, 18(9), 2835. doi: 10.3390/s18092835.

Jarchi, D., Rodgers, S. J., Tarassenko, L. & Clifton, D. A. (2018). Accelerometry-Based Estimation of Respiratory Rate for Post-Intensive Care Patient Monitoring. *IEEE Sensors Journal*, 18(12), 4981–4989. doi: 10.1109/JSEN.2018.2828599.

Jerônimo da Silva Júnior, G. (2008). *Banco de filtros e wavelets sobre corpos finitos*. (Ph.D. thesis, Universidade Federal de Pernambuco). Consulted at https://repositorio.ufpe.br/handle/123456789/5247.

Kido, K., Tamura, T., Ono, N., Altaf-Ul-Amin, M., Sekine, M., Kanaya, S. & Huang, M. (2019). A Novel CNN-Based Framework for Classification of Signal Quality and Sleep Position from a Capacitive ECG Measurement. *Sensors*, 19(7), 1731. doi: 10.3390/s19071731.

Koenig, A., Rehg, T. & Rasshofer, R. (2015). Statistical sensor fusion of ECG data using automotive-grade sensors. *Advances in Radio Science*, 13, 197–202. doi: 10.5194/ars-13-197-2015.

Kuzilek, J., Kremen, V., Soucek, F. & Lhotska, L. (2014). Independent component analysis and decision trees for ECG holter recording de-noising. *PLoS ONE*, 9(6), 1–9. doi: 10.1371/journal.pone.0098450.

Leicht, L., Skobel, E., Knackstedt, C., Mathissen, M., Sitter, A., Wartzek, T., Mohler, W., Reith, S., Leonhardt, S. & Teichmann, D. (2019). Capacitive ECG Monitoring in Cardiac Patients During Simulated Driving. *IEEE Transactions on Biomedical Engineering*, 66(3), 749–758. doi: 10.1109/TBME.2018.2855661.

Lessard-Tremblay, M., Weeks, J., Morelli, L., Cowan, G., Gagnon, G. & Zednik, R. J. (2020). Contactless Capacitive Electrocardiography Using Hybrid Flexible Printed Electrodes. *Sensors*, 20(18), 5156. doi: 10.3390/s20185156.

Li, Q., Mark, R. G. & Clifford, G. D. (2008). Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, 29(1), 15–32. doi: 10.1088/0967-3334/29/1/002.

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2), 4–22. doi: 10.1109/MASSP.1987.1165576.

Lundberg, K. (2002). Noise sources in bulk CMOS. *Unpublished paper*, 1–12. doi: 10.1.1.151.7933.

Mcsharry, P. & Cifford, G. (2004). Open-source software for generating electrocardiogram signals. *Proceedings of the 3rd IASTED International Conference on Biomedical Engineering 2005*.

McSharry, P., Clifford, G., Tarassenko, L. & Smith, L. (2003). A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3), 289–294. doi: 10.1109/TBME.2003.808805.

Moody, G. & Mark, R. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45–50. doi: 10.1109/51.932724.

Moody, G. B., Muldrow, W. K. & Mark, R. G. (1984). *Noise Stress Test for Arrhythmia Detectors*. Computers in Cardiology.

Npatchett. (2015). Derivation of the limb leads. Consulted at https://en.wikipedia.org/wiki/Electrocardiography#/media/File:Limb_leads_of_EKG.png.

Oh, S. (2004). A NEW QUALITY MEASURE IN ELECTROCARDIOGRAM SIGNAL.

Olkkonen, J. (2011). *Discrete Wavelet Transforms - Theory and Applications*. InTech. doi: 10.5772/649.

Orphanidou, C. (2018). *Signal Quality Assessment in Physiological Monitoring*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-68415-4.

Oster, J. & Clifford, G. D. (2015). Signal quality indices for state space electrophysiological signal processing and vice versa. In Chen, Z. (Ed.), *Advanced State Space Methods for Neural and Clinical Data* (pp. 345–366). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139941433.016.

Ottenbacher, J., Kirst, M., Jatoba, L., Huflejt, M., Grossmann, U. & Stork, W. (2008, 8). Reliable motion artifact detection for ECG monitoring systems with dry electrodes. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1695–1698. doi: 10.1109/IEMBS.2008.4649502.

Ottenbacher, J. & Heuer, S. (2009). Motion Artefacts in Capacitively Coupled ECG Electrodes. In *IFMBE Proceedings* (vol. 25, pp. 1059–1062). doi: 10.1007/978-3-642-03882-2{_}282.

Outram, N. J. (1998). *Intelligent Pattern Analysis of the Foetal Electrocardiogram*. (Ph.D. thesis). Consulted at d:%5CLiteratur%5CArbeiten%5CDissertation%5Cn-%5CnIntelligent%5CnPattern%5CnAnalysis%5Cnof%5Cnthe%5CnFoetal%5CnElectrocardiogram.pdf.

Owen, M. (2020). Calculate exp() and log() Without Multiplications. Consulted at https://www.quinapalus.com/efunc.html.

Perez Alday, E. A., Gu, A., Shah, A., Robichaux, C., Wong, A.-K. I., Liu, C., Liu, F., Rad, A. B., Elola, A., Seyedi, S., Li, Q., Sharma, A., Clifford, G. D. & Reyna, M. A. (2020). Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *medRxiv*, 2020.08.11.20172601. doi: 10.1101/2020.08.11.20172601.

Rai, H. M., Trivedi, A. & Shukla, S. (2013). ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier. *Measurement*, 46(9), 3238–3246. doi: 10.1016/j.measurement.2013.05.021.

Ren, A., Du, Z., Li, J., Hu, F., Yang, X. & Abbas, H. (2017). Adaptive interference cancellation of ECG signals. *Sensors (Switzerland)*, 17(5). doi: 10.3390/s17050942.

Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M. & Başar, E. (2001). Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105(1), 65–75. doi: 10.1016/S0165-0270(00)00356-3.

Saha, S. & Ghorai, S. (2015, 2). Effect of feature fusion for discrimination of cardiac pathology. *Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT)*, pp. 1–6. doi: 10.1109/C3IT.2015.7060202.

Santos, E. J. P. (2011). *Eletrônica Analógica Integrada e Aplicações*. Editora Livraria da Física.

Saritha, C., Sukanya, V. & Murthy, Y. (2008). ECG Signal Analysis Using Wavelet Transforms. *Bulgarian Journal of Physics*, 35.

Schneider, J., Koellner, C. & Heuer, S. (2012, 6). An approach to automotive ECG measurement validation using a car-integrated test framework. *2012 IEEE Intelligent Vehicles Symposium*, pp. 950–955. doi: 10.1109/IVS.2012.6232289.

Schumm, J., Arnrich, B. & Troster, G. (2012). ECG Monitoring in an Airplane Seat: Appraising the Signal Quality. *IEEE Pervasive Computing*, 11(4), 28–34. doi: 10.1109/MPRV.2011.40.

Seung Min Lee, Ko Keun Kim & Park, K. S. (2008, 8). Wavelet approach to artifact noise removal from Capacitive coupled Electrocardiograph. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2944–2947. doi: 10.1109/IEMBS.2008.4649820.

Sivanandam, S. N., Sumathi, S. & Deepa, S. N. (2006). *Introduction to neural networks using MATLAB 6.0*. New Delhi: McGraw Hill Education (India) Private Limited.

Sohmyung Ha, Chul Kim, Chi, Y. M., Akinin, A., Maier, C., Ueno, A. & Cauwenberghs, G. (2014). Integrated Circuits and Electrode Interfaces for Noninvasive Physiological Monitoring. *IEEE Transactions on Biomedical Engineering*, 61(5), 1522–1537. doi: 10.1109/TBME.2014.2308552.

Such, O. & Muehlsteff, J. (2006, 9). The Challenge of Motion Artifact Suppression in Wearable Monitoring Solutions. *2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, pp. 49–52. doi: 10.1109/ISSMDBS.2006.360094.

Texas Instruments. (2020). INA12x Precision, Low-Power Instrumentation Amplifiers. Texas Instruments. Consulted at https://www.ti.com/lit/ds/symlink/ina128.pdf?ts=1596039676224&ref_url=https%253A%252F%252Fwww.google.com%252F.

Torrence, C. & Compo, G. P. (1998a). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1), 61–78. doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

Torrence, C. & Compo, G. P. (1998b). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79, 61–78.

Wang, T., Lu, C., Sun, Y., Yang, M., Liu, C. & Ou, C. (2021). Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network. *Entropy*, 23(1), 119. doi: 10.3390/e23010119.

Wang, Y. (2018). *An automated ECG signal quality assessment method with supervised learning algorithm*. (Ph.D. thesis, Delft University of Technology).

Wartzek, T., Lammersen, T., Eilebrecht, B., Walter, M. & Leonhardt, S. (2011). Triboelectricity in Capacitive Biopotential Measurements. *IEEE Transactions on Biomedical Engineering*, 58(5), 1268–1277. doi: 10.1109/TBME.2010.2100393.

Wedekind, D., Kleyko, D., Osipov, E., Malberg, H., Zaunseder, S. & Wiklund, U. (2018). Robust methods for automated selection of cardiac signals after blind source separation. *IEEE Transactions on Biomedical Engineering*, 65(10), 2248–2258. doi: 10.1109/TBME.2017.2788701.

Weeks, J., Elsaadany, M., Lessard-Tremblay, M., Targino, L., Liamini, M. & Gagnon, G. (2020, 7). A Novel Sensor-Array System for Contactless Electrocardiogram Acquisition. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4122–4125. doi: 10.1109/EMBC44109.2020.9175666.

Wiklund, U., Karlsson, M., Östlund, N., Berglin, L., Lindecrantz, K., Karlsson, S. & Sandsjö, L. (2007). Adaptive spatio-temporal filtering of disturbed ECGs: A multi-channel approach to heartbeat detection in smart clothing. *Medical and Biological Engineering and Computing*, 45(6), 515–523. doi: 10.1007/s11517-007-0183-0.

Wu, S., Shen, Y., Zhou, Z., Lin, L., Zeng, Y. & Gao, X. (2013). Research of fetal ECG extraction using wavelet analysis and adaptive filtering. *Computers in Biology and Medicine*, 43(10), 1622–1627. doi: 10.1016/j.compbiomed.2013.07.028.

Xu, L., Rabotti, C., Zhang, Y., Harpe, P. J., Mischi, M., Meftah, M. & Ouzounov, S. (2018). Adaptive motion-artifact reduction in capacitive ECG measurements by using the power-line interference. *MeMeA 2018 - 2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings*. doi: 10.1109/MeMeA.2018.8438751.

Xu, L., Rabotti, C., Zhang, Y., Ouzounov, S., Harpe, P. J. A. & Mischi, M. (2019). Motion-Artifact Reduction in Capacitive Heart-Rate Measurements by Adaptive Filtering. *IEEE Transactions on Instrumentation and Measurement*, 68(10), 4085–4093. doi: 10.1109/TIM.2018.2884041.

Zhao, Z. & Zhang, Y. (2018). SQI Quality Evaluation Mechanism of Single-Lead ECG Signal Based on Simple Heuristic Fusion and Fuzzy Comprehensive Evaluation. *Frontiers in Physiology*, 9(JUN), 1–13. doi: 10.3389/fphys.2018.00727.

Zong, W. (2010). wqrs: Single-lead QRS detector based on length transform. PhysioNet. Consulted at https://archive.physionet.org/physiotools/wfdb/app/wqrs.c.

Zurada, J. (1992). *Introduction to Artificial Neural Systems*. USA: West Publishing Co.