

Advanced tools for ambulatory ECG and respiratory analysis

Jonathan Moeyersons

Supervisors:

Prof. dr. ir. S. Van Huffel

Prof. dr. R. Willems

Prof. dr. ir. C. Varon

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

March 2021

Advanced tools for ambulatory ECG and respiratory analysis

Jonathan MOEYERSONS

Examination committee:

Prof. dr. ir. P. Wollants, chair

Prof. dr. ir. S. Van Huffel, supervisor

Prof. dr. R. Willems, supervisor

Prof. dr. ir. C. Varon, supervisor

Prof. dr. ir. J. Suykens

Prof. dr. P. Hespel

Prof. dr. ir. D. Schreurs

Prof. dr. ir. JM. Aerts

(KU Leuven)

Prof. dr. R. Vullings

(Technische Universiteit Eindhoven)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

March 2021

© 2021 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Jonathan Moeyersons, Kasteelpark Arenberg 10 - box 2446, B-3001 Leuven
(Belgium), B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden
door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande
schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm,
electronic or any other means without written permission from the publisher.

Preface

A typical PhD is four years, sometimes a bit less, sometimes a bit more. At the end of that PhD you need to bundle all your work in a booklet and defend yourself for a jury. Afterwards, you take the credit as if you were the only one that did all the work. However, this could not be further from the truth. My PhD is the result of teamwork, and I would have never gotten here if that was not the case.

I started this PhD with two complementary supervisors, Sabine and Rik. Sabine was the technical brain and Rik the clinical brain. I would like to thank you for giving me the opportunity and confidence to develop into the researcher that I am today. I am very grateful that I could work in an environment that is built on honest and open communication. I will never forget the dinner conversations during the conferences we did together or the many pizza evenings.

Somewhere along the way, I had the privilege to add another supervisor. She started as my 'nothing', but we both knew from the start that she was a lot more than that. Carolina, I cannot describe how grateful I am for your guidance and support. You have made this work so much better than it could have been and that in the most pleasant atmosphere possible. Thank you for inviting Lene and myself to your wedding in Colombia. We cherish the moments that we could spend with you, Steven and your family! I hope that you can start a new chapter of happiness and fulfilment with your new house and that you can continue to find joy in the little things. *Un abrazo.*

I would like to thank the rest of my examination committee. Prof. Patrick Wollants for accepting the invitation to chair this defence. Prof. Dominique Schreurs for showing me a different perspective. Prof. Peter Hespel for the honest and to the point feedback. Prof. Johan Suykens for challenging me on a mathematical level. Prof. Jean-Marie Aerts for supervising my master thesis and now judging this PhD. Prof. Rik Vullings for the thorough assessment and for judging me from the Netherlands.

During my PhD, I had the chance to go on a research stay to the BSICoS research group in Zaragoza. Besides learning a lot, I also had the opportunity to meet a lot of very nice people. Javi, gracias por invitarme a su casa y presentarme a su familia. I feel blessed to have been a small part of your family while I was there. Pablo A., throughout the years, I have seen your confidence grow and I am sure that you will do many great things in the future. Thank you for all the nice moments in Spain, Belgium and France. Spyros, you introduced me to the wonder that juepincho is. I am eternally grateful for that. David, thank you for sharing me your data and knowledge. Jesus, you have shown me how to be a hard working family man. Thank you. Raquel, Juan-Pablo and Pablo L., thank you for inviting me to work in your group and making me feel welcome. It was a pleasure to discuss new insights with all of you.

Omwille van mijn ervaring en achtergrond in sport wetenschappen, werd ik mee betrokken in de uitlopers van het Sensor-based Platform for the Accurate and Remote monitoring of Kine(ma)tics Linked to E-health, ofte SPARKLE project. Hierbij wil ik graag mijn dankbaarheid uitdrukken aan Thijs en Evelyne. Thijs, bedankt voor je onaflatende energie en enthousiasme. Evelyne, bedankt om al mijn tools mee te testen. Het was heel fijn om met jou samen te werken.

Due to the collaboration between our group and the cardiology department, I could closely work together with some world class cardiologists. Bert, you have inspired me to make impossible things possible. Without you, R-DECO would have still been a concept in the back of my head, instead of a tool that is published and used by many. Sebastian, thank you for all the nice conversations during meetings and dinners. Matthew, thank you for pushing me to make better tools, to think critically about algorithms and teaching me a small part of your medical knowledge.

Throughout the years, our research group has changed a lot. Many people left BIOMED and were replaced by others. However, each and every one of them left behind some memories. I would like to thank all of you for making BIOMED such a warm working environment.

Thomas, thank you for being such a calm, steady presence in the office. Your joke will always echo through the hallways of ESAT. Neetha, thank you for your kindness and generosity. Alex, my brother from another mother, and probably father. Thank you for all the cake, but mostly for welcoming me into the group from day one. You were always like an older, balder version of myself, but an example nonetheless.

Amir, you guided me through the CNN world. I learned a lot from you. Pooya, I feel sad that we never had the opportunity to play badminton together. Let's hope that we can still make up for that. Rob, bedankt om zo'n visueel genie

te zijn. Je slaagde er altijd in om je eigen standaard telkens te verhogen en trok de rest van ons met je mee omhoog. Je legendarische feestjes gaan nooit vergeten worden. Jasper, bedankt voor de leuke ping-pong sessies, voor de gesprekken tijdens en na het werk en voor alle wijze raad die je deelde. Het was een privilege om mijn hele doctoraat samen met jou te mogen doen. Ik hoop dat we hierna nog veel tijd samen kunnen doorbrengen. Bori, thank you for all your kindness.

Laure, hopelijk kunnen we in de toekomst nog veel dagen zoals die ene dag op Couleur Café beleven. Griet. Oh Griet, waar moet ik beginnen. Jij bent de reden waarom ik aan dit doctoraat ben begonnen. Je hebt me begeleid, gesteund en bij momenten vooruit gesleurd. Zonder jou zou ik nu niet staan waar ik ben. Ik kan je niet genoeg bedanken voor alle warmte en vriendschap. Dorien, Doske, Dostank, bedankt om er te zijn, altijd. Om te babbelen, zeveren en helpen waar nodig en dat altijd met de glimlach. Ik heb je nog altijd een cantus te goed. Margot, bedankt om een voorbeeld te zijn van hoe het allemaal zou moeten. Ik heb altijd heel hard naar jou opgekeken. Nick, zonder jou waren auto encoders, nog altijd vreemde wezens. Bedankt voor het delen van je kennis, een klein stukje dan toch, en al je geduld.

Nico en Martijn, jullie hebben de wereld der tensoren een stuk minder angstaanjagend gemaakt. Hopelijk kunnen we nog veel age of empires spelen. Christos, you showed me how being passionate and being polite can coexist. Cem, although I was terrible at the whole Spyfall game, I really enjoyed playing it with you.

Jonathan Dan, your changing appearances always lit up the room. Abhi, I thought that I knew a lot about football, but your knowledge is unmatched. Kaat, ik herinner me nog altijd een examen bij de bio-ingenieurs waar ik twee keer op gebuisd was en jij doodleuk zei dat het eigenlijk gewoon een beetje puzzelen was. Ik denk dat we allebei sinds dat moment enorm zijn veranderd en gegroeid. Ik ben heel dankbaar dat ik heel deze weg met jou heb kunnen afleggen.

Barath, thank you for being my badminton partner. It was a pleasure playing together. Ofelie, je was een plezier om mee samen te werken en een uitdaging om tegen te badmintonnen. Ik hoop dat we mekaar nogeens tegenkomen in de toekomst. Mario, at one of the earliest meetings in our PhD at the sleeplab, we sat together in the car and we started talking about our doubts and future. And, truth be told, we never stopped doing that afterwards. You were my loud, not so solid rock at the office and I still miss you. Elisabeth and Laura, heel veel respect om mee te kunnen met Mario zijn enthousiasme en energie. Het is niet velen gegeven. Ik vond het heel fijn om jullie te zien groeien van student naar onderzoeker.

Simon G, bedankt om mijn gebrekkige wiskundige kennis bij te willen schaven. Je heb me heel veel bijgebracht. Als we nog eens pingpongen, laat me dan ook nog eens winnen aub. Servaas, ik hoop dat we snel nog eens samen kunnen gaan klimmen. Adrian, big guy, you left a hole when you left. Mostly because you are tall, but also because you are such a kind soul.

Amalia, your enthusiasm, smile and positivity made it pure joy to work with you. When I started my PhD, I corrected your thesis and now, almost five years later, you have corrected mine. What a turn of events. I am grateful that I can call you my friend, gracias por todo. John, thank you for being a friend, a teacher and a companion during my PhD. I wish that there was more time that we could spend together, tinkering about codes, talking at the coffee machine or complaining about students. Dries, I still remember the day that you started, Genk's finest in tha house. You were full of energy and new insights. I really enjoyed playing pingpong and organizing the Kinderuniversiteit with you. Andrea, thank you for all your patience when I was running late with something again. I am sure that you will do great. For sure, you have the best supervisor possible.

Lieven, hoewel je al even weg bent, is het gat dat je hebt achtergelaten nog steeds niet opgevuld. Er lopen namelijk geen twee personen zoals jij rond op deze wereld. Bedankt voor je geduld als ik weer eens ongeduldig was en je steun wanneer ik het even niet zag zitten. Btw. ik vind het nog altijd jammer dat je Aurélie niet Sinathan hebt genoemd. Bij je volgende kind krijg je een herkansing. Simon, ik vind het moeilijk om woorden te vinden wat jij voor mij hebt betekend de afgelopen jaren. We hebben samen gelachen, gesport, gediscussieerd en nog zoveel meer. Bedankt voor een geweldige vrijgezellen. Ik ga het nooit vergeten. Tim, je bent de laatste van onze bureau die overblijft, maar ik weet dat je de eer hoog gaat houden. Bedankt voor al je grappen en prachtige Vlaamse woorden.

To all the 'new' people in the group, Nithin, Kenneth, Joran, Miguel, Luis, Konstantinos and all the others that have recently arrived. I wish you all the best of luck. You have landed in the best possible group and I hope that you can have at least the same amount of fun that I had.

John, Elsy, Ida, Wim and Aldona, thank you for always answering my questions or correcting me whenever I made mistakes, which was a lot. BIOMED would not be the same without you.

Kirsten, Marc, Dine en Steven, bedankt om mij bij te staan met raad en daad en lekker eten. Ik kon altijd bij jullie terecht voor eerlijke feedback en wijze raad.

Ben, Marika, Sander en Eni, bedankt om mijn gezaag te aanhoren als het weer eens niet ging zoals ik het wou. Ik ben heel dankbaar voor al jullie steun.

Thieu, merci om te luisteren naar mij, ookal was het vaak Chinees. Uw aanwezigheid was vaak al voldoende om mijn gedachten te verzetten.

Mama en papa, bedankt om van mij de persoon te maken die ik nu ben. Ik had hier nooit gestaan zonder jullie onaflatende steun en geloof in mij.

Lene, mijn bolleke, bedankt voor de steun, de stampen onder mijn gat, de motivatie en de liefde. Zonder u was dit doctoraat waarschijnlijk nog altijd niet af. Onze Beninging is met zijn, of haar, schattige poep in de boter gevallen met u als mor.

Abstract

The electrocardiogram or ECG is a relatively easy-to-record signal that contains an enormous amount of potentially useful information. It is currently mostly being used for screening purposes. For example, pre-participation cardiovascular screening of young athletes has been endorsed by both scientific organisations and sporting governing bodies.

A typical cardiac examination is taken in a hospital environment and lasts 10 seconds. This is often sufficient to detect major pathologies, yet this small sample size of the heart's functioning can be deceptive when used to evaluate one's general condition. A solution for this problem is to monitor the patient outside of the hospital, during a longer period of time. Due to the extension of the analysis period, the detection rate of cardiac events can be highly increased, compared to the cardiac exam in the hospital. However, it also increases the likelihood of the signals being exposed to noise, which could decrease the diagnostic capabilities of the signals.

Therefore, in the first part of this work, we present novel quality indication algorithms for cardiac and respiratory signals which could aid in cardiac-, respiratory- and cardiorespiratory analysis. These algorithms have shown good results on newly labelled datasets that were recorded with different recording devices. Additionally, we have shown that a transfer learning approach could be used to optimize an artefact detection algorithm that was trained with contact ECG towards non-contact ECG.

When signal quality is ensured, most often the next step in cardiac analysis is the detection of heartbeats. This sounds like a straightforward task, and in many cases it is, but in some situations it could be a challenge. Hence, for such situations it is recommended to visually inspect and review each signal before further analysis.

Many of the existing ECG analysis toolboxes assume that all heartbeats are correctly annotated and as such, do not provide any correction tools.

Furthermore, the ones that do provide these, are not user friendly. Hence, in the second part of this work, we present a toolbox that can be used to detect heartbeats, visualize these together with the raw signal, and correct possible wrong annotations. Furthermore, we extended this toolbox for beat-to-beat variability of repolarization (BVR) analysis. We used this extension to investigate the temporal evolution in BVR before spontaneous non-sustained ventricular tachycardia (nsVT) in patients with ischaemic heart disease (IHD). Our preliminary results suggest that temporal changes in pre-arrhythmic BVR could be used to predict imminent nsVT events in IHD patients.

In the last part of this work, we used the quality indication algorithm and R-peak detection and correction tool to investigate the strength of the cardiorespiratory coupling during exercise. The presented pipeline could be used similarly for other applications. Lastly, we have shown that the combination of ECG criteria with demographic and body composition features can be used to accurately estimate left ventricular mass in endurance athletes.

Beknopte samenvatting

Het electrocardiogram of ECG is een relatief eenvoudig op te nemen signaal dat een enorme hoeveelheid potentieel bruikbare informatie bevat. Het wordt momenteel vooral gebruikt voor screening doeleinden. De cardiovasculaire screening vóór deelname van jonge atleten aan sport is bijvoorbeeld goedgekeurd door zowel wetenschappelijke organisaties als sportbesturen.

Een typisch hartonderzoek wordt afgenomen in een ziekenhuisomgeving en duurt doorgaans 10 seconden. Dit is vaak voldoende om belangrijke pathologieën op te sporen, maar dit korte onderzoek kan misleidend zijn wanneer het wordt gebruikt om iemands algemene toestand te evalueren. Een oplossing voor dit probleem is om de patiënt buiten het ziekenhuis, gedurende langere tijd te monitoren. Door de verlenging van de onderzoeksperiode kan de detectiegraad van abnormale gebeurtenissen sterk worden verhoogd in vergelijking met een typisch hartonderzoek in het ziekenhuis. Het verhoogt echter ook de kans dat de signalen worden blootgesteld aan ruis, wat de diagnostische mogelijkheden van de signalen zou kunnen verminderen.

In het eerste deel van dit werk presenteren we daarom nieuwe algoritmes voor kwaliteitsindicatie van cardiale en respiratoire signalen die kunnen helpen bij cardiale, respiratoire en cardiorespiratoire analyses. Deze algoritmes behaalden goede resultaten op nieuw gelabelde datasets die zijn opgenomen met verschillende opnameapparaten. Bovendien hebben we aangetoond dat transfer learning kan worden gebruikt om een artefact detectie algoritme dat getraind is op contact ECG te optimaliseren naar non-contact ECG.

Wanneer de kwaliteit van de signalen is verzekerd, is de meest voorkomende volgende stap in de ECG analyse de detectie van hartslagen. Dit klinkt als een eenvoudige taak, en in veel gevallen is het dat ook, maar in sommige situaties kan het een uitdaging zijn. Daarom wordt het voor dergelijke situaties aanbevolen om elk signaal visueel te inspecteren en te beoordelen alvorens verder te analyseren.

Veel van de bestaande toolboxes voor ECG-analyse gaan ervan uit dat alle hartslagen correct zijn gedetecteerd en bieden als dusdanig geen correctietools aan. Bovendien zijn de toolboxes die dit wel aanbieden niet gebruiksvriendelijk. Daarom presenteren we in het tweede deel van dit werk een toolbox die kan worden gebruikt om hartslagen te detecteren, deze samen met het onbewerkte signaal te visualiseren en om mogelijke foute annotaties te corrigeren. Verder hebben we deze toolbox uitgebreid voor beat-to-beat variabiliteit van repolarisatie (BVR) analyse. We hebben deze extensie gebruikt om de temporele evolutie in BVR te onderzoeken vóór spontane niet-aanhoudende ventriculaire tachycardie (nsVT) bij patiënten met ischemische hartziekte (IHD). Onze voorlopige resultaten suggereren dat tijdelijke veranderingen in pre-aritmische BVR kunnen worden gebruikt om op handen zijnde nsVT-gebeurtenissen bij IHD-patiënten te voorspellen.

In het laatste deel van dit werk hebben we het kwaliteitsindicatie-algoritme en de R-piek detectie- en correctietool gebruikt om de sterkte van de cardiorespiratoire koppeling tijdens inspanning te onderzoeken. De gepresenteerde pijplijn kan op dezelfde manier worden gebruikt voor andere toepassingen. Ten slotte hebben we aangetoond dat de combinatie van ECG-criteria met demografische en lichaamssamenstellingskenmerken kan worden gebruikt om de linker ventrikel massa bij duursporters nauwkeurig te schatten.

Nomenclature

Abbreviations

Acc	Accuracy
ACF	Autocorrelation Function
ACM	Accelerometer
AE	Autoencoder
AF	Atrial Fibrillation
ANN	Artificial Neural Network
ANS	Autonomic Nervous System
AUC	Area Under the ROC-Curve
bAcc	Balanced Accuracy
BioZ	Bio-impedance
BMC	Bone Mineral Content
BR	Breathing Rate
BSA	Body Surface Area
BVR	Beat-to-beat Variability of Repolarization
BW	Baseline Wander
CC	Capacitatively Coupled
CCA	Canonical Correlation Analysis
ccECG	Capacitively Coupled ECG
CinC	Computing in Cardiology
CMR	Cardio Magnetic Resonance
CNN	Convolutional Neural Networks
CNS	Central Nervous System
CO	Cardiac Output
Conv	Convolutional Layer
COPD	Chronic Obstructive Pulmonary Disease
CP	Cornell Product
CRC	Cardio Respiratory Coupling

CRD	Chronic Respiratory Diseases
CRF	Cardio Respiratory Fitness
CV	Cornell Voltage
CVD	Cardiovascular disease
DT	Decision Tree
DXA	Dual-energy X-ray Absorptiometry
ECG	Electrocardiogram
EDF	European Data Format
EDR	ECG Derived Respiration
EM	Electrode Motion
EMD	Empirical Mode Decomposition
EMG	Electromyogram
FIR	Finite Impulse Response
FM	Fat Mass
FMin	First (local) Minimum
FN	False Negatives
FP	False Positives
FPR	False Positive Rate
GDF	General Data Format
GUI	Graphical User Interface
HCM	Hypertrophic Cardiomyopathy
HDF5	Hierarchical Data Format 5
HH	Hand Held
HR	Heart Rate
HRV	Heart rate variability
ICA	Independent Component Analysis
ICD	Implantable Cardioverter-Defibrillator
ICU	Intensive Care Unit
IHD	Ischaemic Heart Disease
LA	Left Arm Electrode
LASSO	Least Absolute Shrinkage and Selection Operator
LL	Left Leg Electrode
LM	Lean Mass
LV	Left Ventricle
L VH	Left Ventricle Hypertrophy

LVM	Left Ventricular Mass
MA	Muscle Artefact
MAmp	Maximum Amplitude at 35 ms
MIP	Maximal Static Inspiratory Pressure
mRMR	Minimum Redundancy Maximum Relevance
NN	Normal-to-Normal
NPV	Negative Predictive Value
NSTDB	Noise Stress Test Database
nsVT	Non-Sustained Ventricular Tachycardia
NYHA	New York Heart Association
PCA	Principal Component Analysis
PSD	Power Spectral Density
PSG	Polysomnographic
PPV	Positive Predictive Value
PVC	Premature Ventricular Contraction
RA	Right Arm Electrode
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RL	Right Leg Electrode
RMSE	Root-Mean-Squared-Error
ROC	Receiver Operating Characteristic
RUS	Random Under Sampling
RUSBoost	Random Under Sampling Boosting
SA	Sinoatrial
SD	Standard Deviation
Se	Sensitivity
Sim	Similarity
SLP	Sokolow-Lyon Product
SLV	Sokolow-Lyon Voltage
SM	Spectral method
SNR	Signal-to-Noise Ratio
Sp	Specificity
SV	Stroke Volume
SVM	Support Vector Machines
STV	Short-Term Variability
TN	True Negatives

TP	True Positives
TPR	True Positive Rate
TWA	T-Wave Alternans
VT	Ventricular Tachycardia
Vx	Chest electrodes V1 to V6
wAUC	Weighted Area Under the ROC-curve
wSe	Weighted Sensitivity
wSp	Weighted Specificity

Contents

Abstract	vii
Beknopte samenvatting	ix
List of Abbreviations	xv
List of Symbols	xvii
Contents	xvii
List of Figures	xix
List of Tables	xxi
1 Introduction	1
1.1 Research motivation	1
1.2 Background	2
1.2.1 Physiology of the heart	3
1.2.2 The electrocardiogram	5
1.2.3 Ambulatory monitoring	10
1.2.4 Artefacts	10
1.2.5 Cardiorespiratory coupling	12
1.3 State of the art	13
1.4 Research objectives	17
1.5 Chapter overview and main contributions	17
1.5.1 Part I. Background	18
1.5.2 Part II. Signal quality	19
1.5.3 Part III. Tools for ECG analysis	20
1.5.4 Part IV. Sport applications	20
1.6 Collaborations	21

2	Machine learning	26
2.1	Introduction	27
2.2	Supervised classification	28
2.2.1	Support vector machines	28
2.2.2	RUSBoost	30
2.2.3	Convolutional neural networks	33
2.2.4	Performance evaluation	37
2.3	Regression techniques	43
2.3.1	Support vector machines	43
2.3.2	Performance evaluation	44
2.4	Conclusion	45
3	Sports and the heart	47
3.1	Introduction	47
3.2	Acute cardiorespiratory response to exercise	48
3.2.1	Research question	50
3.3	Cardiovascular adaptations to exercise	50
3.3.1	Research question	51
4	ECG quality	54
4.1	Introduction	55
4.2	Data	56
4.2.1	Polysomnographic dataset (PSG)	56
4.2.2	Hand held dataset (HH)	56
4.2.3	Stress dataset (STR)	57
4.2.4	(Re)labelling of the data	57
4.3	Methods	58
4.3.1	Pre-processing	58
4.3.2	Feature extraction	58
4.3.3	Classification	61
4.3.4	Quality indication	63
4.4	Results	65
4.4.1	Inter-rater agreement	65
4.4.2	Model performance	65
4.4.3	Quality assessment index	66
4.5	Discussion	68
4.6	Conclusion	71
5	Transfer Learning for Modality-Specific Artefact Detection	73
5.1	Introduction	74
5.2	Data	75
5.2.1	Capacitively coupled dataset (CC)	76
5.3	Methods	77

5.3.1	Pre-processing	77
5.3.2	Features	77
5.3.3	Classification	78
5.3.4	Performance evaluation	82
5.4	Results	82
5.5	Discussion	86
5.5.1	Features & base classifier	86
5.5.2	Transfer learning	88
5.5.3	Limitations	90
5.6	Conclusions	91
6	Respiration quality	93
6.1	Introduction	94
6.2	Data	96
6.2.1	Subjects	96
6.2.2	Data acquisition	96
6.2.3	Respiratory protocol	98
6.3	Methods	99
6.3.1	Pre-processing	99
6.3.2	Labelling	100
6.3.3	Classification	101
6.3.4	Performance evaluation	108
6.4	Results	108
6.5	Discussion	112
6.5.1	Features	112
6.5.2	Classification	113
6.5.3	Limitations	114
6.6	Conclusions	114
7	R-DECO	116
7.1	Introduction	117
7.2	Computational methods	118
7.2.1	R-peak detection	118
7.3	Software description	124
7.3.1	Input data formats	124
7.3.2	User interface	125
7.3.3	Save and export results	131
7.3.4	Data browser	131
7.3.5	Preferences	132
7.4	Sample run	133
7.5	Potential of future growth	134
7.6	Conclusion	135

8	Beat-to-beat variability of repolarization	137
8.1	Introduction	138
8.2	BVR-tool	139
8.2.1	T-wave end detection	139
8.2.2	QRS-onset detection	142
8.2.3	Software description	145
8.2.4	Export results	148
8.2.5	Potential of future growth	149
8.3	Predicting non-sustained ventricular tachycardia in ischemic heart disease patients	149
8.3.1	Data	149
8.3.2	Methods	149
8.3.3	Results	151
8.3.4	Discussion	151
8.4	Conclusion	153
9	Cardiorespiratory strength during exercise	156
9.1	Introduction	157
9.2	Data	158
9.3	Methods	159
9.3.1	Lead selection	159
9.3.2	HRV derivation	160
9.3.3	ECG derived respiration (EDR)	160
9.3.4	Cardiorespiratory coupling (CRC)	161
9.4	Results and discussion	162
9.4.1	EDR	162
9.4.2	CRC	163
9.5	Conclusion	166
10	Left ventricular mass estimation	169
10.1	Introduction	170
10.2	Data	170
10.3	Methods	171
10.3.1	Feature collection	171
10.3.2	Correlation and regression analysis	173
10.3.3	Feature selection	173
10.3.4	Development of LVM estimation model	173
10.4	Results	174
10.5	Discussion	174
10.6	Conclusion	176
11	Conclusions and future directions	179
11.1	Conclusions	180

11.1.1	Accurate artefact detection and quality indication of wearable biomedical signals	181
11.1.2	Develop user friendly tools that aid in analysing ECG signals	182
11.1.3	Explore the added value of ECG and cardiorespiratory analysis in sport applications	184
11.2	Future directions	184
11.2.1	Accurate artefact detection and quality indication of wearable biomedical signals	185
11.2.2	Develop user friendly tools that aid in analysing ECG signals	186
11.2.3	Explore the added value of ECG and cardiorespiratory analysis in sport applications	186
11.2.4	General	187
	Bibliography	189
	Curriculum	211

List of Figures

1.1	Schematic representation of the conduction within the human heart	4
1.2	Standard 12-lead electrode placement	7
1.3	Derivation of the (augmented) limb leads	8
1.4	Characteristic waves and segments of a normal ECG	9
1.5	Examples of noise signals	12
1.6	ECG analysis pipeline	13
1.7	Manuscript overview	18
2.1	Mapping function	29
2.2	RUSBoost	33
2.3	General example of an ROC-curve	40
4.1	Comparison between a clean and a contaminated ECG segment	59
4.2	An example of an ECG signal that contains a large artefact . .	60
4.3	Comparison between a clean and noisy ECG signal, together with the ACF's of their respective sliding windows	61
4.4	Impact of Electrode Motion on ECG signal quality	64
4.5	The mean squared classification error of the 10-fold cross-validation of the PSG dataset versus the number of weak learners	66
4.6	Feature space of the three training datasets	67
4.7	Boxplot of the score of the clean class plotted against the amount of agreeing annotators	68
4.8	The quality of the EM and MA	69
5.1	Example of the mattress and chair implementation of the multi-channel ccECG acquisition system	76
5.2	Comparison between a clean and noisy segment for both contact and non-contact signals	78
5.3	Performance evaluation base classifiers	80

5.4	Feature space of the three datasets	83
5.5	Comparison of the performance on the CC dataset without and with transfer learning	84
5.6	The effect on performance of different subset sizes and the two sampling techniques	85
5.7	The difference in entropy for the clean and noisy samples of the CC subsets	86
5.8	The Se of the CC base classifier when applied on the HH dataset	87
6.1	Wearable BioZ device	97
6.2	Representation of the tetra-polar electrode configurations	97
6.3	The inspiratory threshold loading protocol	99
6.4	An example of the result of the pre-processing steps. The blue line indicates a raw signal of one minute and the red line indicates the filtered version. The baseline wander and high frequency noise is mostly removed.	100
6.5	Graphical user interface for visualizing, commenting and labelling each signal	101
6.6	A clean and contaminated BioZ signal during normal breathing and the resulting ACF	104
6.7	The PSD of the signal shown in Figure 5.2	105
6.8	Workflow per fold of the feature based approach	106
6.9	Feature space of the most frequently selected combination of three features	111
6.10	Comparison of the ROC curves of the ten folds from the SVM (a) and CNN (b) approach. The average ROC curves of the SVM (blue line) and CNN (red line) approach is depicted in (c) . Their respective AUC's are 92.77 ± 2.95 % and 92.51 ± 1.74 %	112
7.1	Flattening procedure	119
7.2	Procedure to select R-peaks	120
7.3	Sensitivity of the performance to the choice of window width	121
7.4	The graphical user interface of R-DECO.	126
7.5	Example of a high pass filter	127
7.6	The R-peak parameter selection window	128
7.7	The template selection window	130
7.8	The data browsing options of R-DECO	132
7.9	Example of an nsVT segment without correction	134
7.10	Example of an nsVT segment with correction	134
8.1	Graphical representation of the semi-automatic T-wave end detection method	141

8.2	Graphical representation of the semi-automatic QRS-onset detection method	144
8.3	The GUI of the BVR tool	146
8.4	Poincaré plot from a representative patient with IHD	151
8.5	Results of the BVR analysis	152
9.1	Comparison between the peak frequency of the EDR and the measured breathing rate	163
9.2	Comparison between (a) \mathcal{P}_x and VO_2 and (b) \mathcal{P}_x and RER	164
9.3	Comparison between (a) \mathcal{P}_x and VO_2 and (b) \mathcal{P}_x and RER of the subjects with a correlation coefficient higher than 0.99	165
9.4	Correlation coefficients with increasing frequency correlation	166
10.1	Boxplots LV mass	176
11.1	Cardiorespiratory analysis pipeline	180

List of Tables

2.1	Example of a confusion matrix for a binary problem	38
4.1	Overview of the HH and STR dataset (re)labelling	65
4.2	Classification performance on independent test sets	67
5.1	Overview of all datasets.	77
5.2	Results obtained for the base classifiers	83
6.1	Demographics of the COPD patients	96
6.2	Overview of the different class labels	102
6.3	Overview of the network architecture	107
6.4	Overview of the dataset	109
6.5	Overview of the performances of the SVM models on the test folds	110
7.1	Performance of the R-peak detection algorithm on the Physionet MIT/BIH dataset.	122
8.1	Clinical data of patients with IHD	150
8.2	nsVT characteristics	151
9.1	Overview of the population demographics	158
9.2	SubC protocol	159
9.3	Results cardio-respiratory analysis	164
10.1	Population demographics	171
10.2	Overview features	172
10.3	Correlation and regression analysis for each feature	175

Chapter 1

Introduction

1.1 Research motivation

A typical cardiac examination is taken in a hospital environment and lasts 10 seconds. This is often sufficient to detect major pathologies, yet this small sample size of the heart's functioning can be deceptive when used to evaluate one's general condition.

A solution for this problem is to monitor the patient outside of the hospital, during a longer period of time. Due to the extension of the analysis period, the detection rate of cardiac events can be highly increased, compared to the cardiac exam in the hospital. However, it also increases the likelihood of the signals being exposed to noise. During a cardiac examination in the hospital, the patient is asked to lay still in a supine position and as a result, the recorded signals are generally of very high quality. This is no longer the case for ambulatory recordings, where the diagnostic capabilities of the signals can be reduced by the presence of artefacts.

Most data driven support tools assume clean data. Since this is much less straightforward for signals recorded with an ambulatory device, we need algorithms that can quantify the contamination level of the resulting signals. These could be used to define a different processing methodology, remove segments that are too contaminated to use, or serve as a reliability metric for the support tool. During this PhD we aimed to develop quality indication algorithms for cardiac and respiratory signals which could aid in cardiac-, respiratory- and cardiorespiratory analysis.

When signal quality is ensured, the next step in most cardiac analysis tasks is the detection of heartbeats. This sounds like a straightforward task, and in many cases it is, but in some situations it could be a challenge. Hence, for such situations it is recommended to visually inspect and review each signal before further analysis. Many of the existing ECG analysis toolboxes assume that all heartbeats are correctly annotated and as such, do not provide visualization and correction tools. Furthermore, the ones that do provide these, are not user friendly. Hence, we aimed to develop a toolbox that can be used to detect heartbeats, visualize these together with the raw signal, and provide user friendly ways to correct possible miss annotations.

Once all heartbeats are correctly located, more elaborate analysis can be performed, for example cardiorespiratory coupling. The cardiorespiratory coupling is perhaps best typified by the occurrence of the respiratory sinus arrhythmia (RSA). This is characterized by a heart rate (HR) increase during inspiration and a HR decrease during expiration. Despite the limited understanding of the mechanisms and function of RSA, it has been suggested as a potential biomarker for people's health status. However, the role of RSA during exercise is not yet fully understood. In this work, we investigated the relationship between the strength of the RSA and different physiological parameters that are monitored during exercise.

In summary, we aimed to

- develop quality indication algorithms for cardiac and respiratory signals,
- develop a user friendly toolbox for ECG analysis,
- explore the added value of ECG and cardiorespiratory analysis in sport applications

1.2 Background

In order to take a step forward in the development of algorithms and tools for analysing the functioning of the heart, one has to understand how it operates and interacts with other systems. Therefore, we take a step back, and start with the basics.

1.2.1 Physiology of the heart

What is the main function?

The heart is a muscular pump that provides the necessary force to circulate blood to all the tissues in the body. Here it delivers oxygen and nutrients and removes carbon dioxide and other waste products in the process. Its function is vital, because even a small interruption of the blood flow could lead to tissue degeneration and, eventually, death.

What are the structural parts?

The heart consists of four chambers: two atria in the upper part and two ventricles in the lower part. The two atria are thin-walled chambers that receive blood from the veins and the two ventricles are thick-walled chambers that forcefully pump blood out of the heart. In order to keep the blood flowing in the correct direction, the heart has two types of valves. The atrioventricular valves, which, as the name suggests, are located between the atria and the ventricles, and the semilunar valves. These are located at the bases of the large arteries that leave the ventricles.

When the ventricles contract, the atrioventricular valves close to prevent blood from flowing back into the atria. When the ventricles relax, the semilunar valves close to prevent blood from flowing back into the ventricles.

How does the blood flow?

The heart pumps blood through two pathways: the pulmonary circuit and the systemic circuit. Note that blood runs simultaneously through these two pathways. The right atrium receives oxygen-poor blood from the superior and inferior vena cava and pumps it to the right ventricle.

The right ventricle pumps the blood through the pulmonary arteries to the lungs to receive oxygen and release carbon dioxide. This sequence is called the pulmonary circuit. The oxygen-rich blood flows from the lungs through the pulmonary veins to the left atrium and then to the left ventricle. From there, it is pumped to the systemic circuit where it is delivered to the tissue. Finally, the resulting oxygen-poor blood returns through the veins to the right atrium of the heart and the process can start all over.

During one cardiac cycle, the four chambers contract and relax in an alternating fashion. The contraction phase is called systole and the relaxation, or filling,

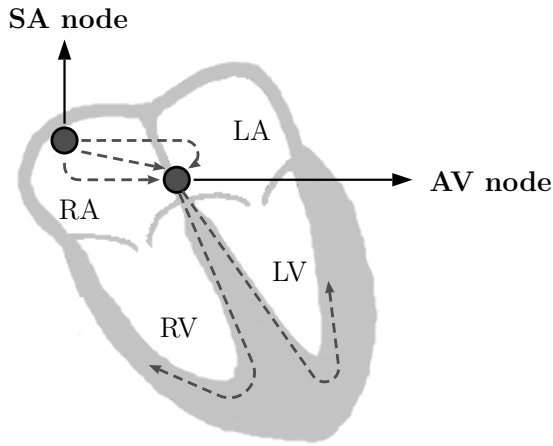


Figure 1.1: Schematic representation of the conduction within the human heart. The sinoatrial (SA) node initiates the action potential. After reaching the atrioventricular (AV) node, there is a delay of approximately 100 ms that allows the atria to complete empty before the impulse is transmitted to the AV bundle. Following this delay, the impulse travels through the AV bundle and bundle branches to the Purkinje fibers. Then, it spreads to the contractile fibers of the ventricle and initiates ventricular contraction. Figure taken from [190]

phase is called diastole. This alternation is coordinated by the electrical conduction system of the heart.

How does the heart's electrical system work?

The cardiac cycle starts with the electrical stimulation of the sinoatrial (SA) node. This node is also called the cardiac pacemaker, since it initiates the normal electric pattern, known as sinus rhythm. The train of action potentials that is generated in the SA node propagates through the atrioventricular (AV) node, the AV bundle (bundle of His) and the AV bundle branches, until it reaches the Purkinje fibers (Figure 1.1). The rapid transmission of the action potential ensures that the entire heart contracts in one coordinated motion, thereby creating a heartbeat.

What controls the heart rate?

The heart rate varies between 40 and 210 bpm, and sometimes even higher, according to the intensity of the activity performed [12]. For example, during exercise, the active muscles require more oxygen to function. Hence, the HR needs to increase significantly to deliver more oxygen to these muscles. Conversely, during periods of sleep, the need for oxygen is low and thus the heart rate decreases. These processes are mainly controlled by the autonomous and central nervous systems (ANS, CNS).

The ANS is the part of the CNS that regulates involuntary physiologic processes including HR, blood pressure (BP), respiration, digestion, and sexual arousal. Due to its peripheral nerves, it has the remarkable quality of quickly modifying the functioning of several organs within the body.

The HR is modulated by the interplay between the two branches of the ANS: the sympathetic and the parasympathetic branch. These branches differ significantly both from an anatomical and functional standpoint.

The sympathetic branch is responsible for the fight-or-flight response. This is a state of elevated activity and attention as a reaction to a stressful situation. Sympathetic responses include an increase in HR, BP and cardiac output (CO), but also an increase in breathing rate (BR) and bronchiolar dilation [90]. The parasympathetic branch is responsible for the rest-and-digest actions. It is largely concerned with the conservation and restoration of energy. Parasympathetic actions include a reduction in HR and BP, among others.

1.2.2 The electrocardiogram

How can we measure the electrical impulses?

The electrical impulse generated in the SA can travel through the heart due to the electrical properties of the cardiac cells. This propagation can be measured as potential differences by electrodes that are placed on the skin. The resulting signal is known as the electrocardiogram or ECG.

The ECG signal can be measured in-subject, on-subject and off-subject, by placing electrodes directly on the heart, on the body surface and in close proximity to the subject, respectively. The most accurate measurements would be in-subject, since the signals are directly recorded from the source and are not altered by any tissue or fabric. However, since the subject needs to undergo surgery to do so, this approach is not very practical.

The most common measurement set-up is the on-subject approach whereby multiple electrodes are placed on the chest and/or limbs. The electrical potential difference between two, or more, electrodes placed on specific points on the body is called an ECG lead or channel.

The standard measurement setup in clinical practice is a 12-lead ECG configuration. As the name suggests, it implies 12 leads, which are derived from 10 electrodes. These electrodes are located on standardized places: four electrodes on the limbs and six on the chest (Figure 1.2). The name of each electrode is derived from its location on the body surface:

- LA = Left Arm Electrode
- RA = Right Arm Electrode
- LL = Left Leg Electrode
- RL = Right Leg Electrode
- V_x = Chest Electrodes V1 to V6

Note that the limb leads are typically not placed on the limbs, but on a location on the chest near the limbs. This avoids the inclusion of muscle artefacts, which could decrease the signal quality [114].

Each ECG lead shows the electrical activity from one spatial angle. Together, the leads completely characterize the electrical activity of the heart and give a comprehensive three-dimensional view. The chest leads record the different angles in the horizontal plane, while the limb leads and augmented leads provide information about the vertical plane.

The simplest leads are composed with only two electrodes, where one electrode is defined as the exploring (positive) and the other as the reference (negative) electrode. However, in most leads, the reference is a combination of multiple electrodes.

The standard 12-lead ECG consists of three limb leads, three augmented limb leads and six precordial leads. Leads I, II and III are called the limb leads. The electrodes that form these signals are located on the limbs, one on each arm and one on the left leg, and form Einthoven's triangle. Leads aVR, aVL, and aVF form the augmented limb leads. The 'a' stands for augmented, 'V' for voltage and 'R', 'L' and 'F' refer to the right arm, left arm and left foot. These leads are derived from the same three limb electrodes as the first three leads, but instead of using one electrode as reference, they use the average of the non-exploring electrodes. This is known as the Goldberger's central terminal

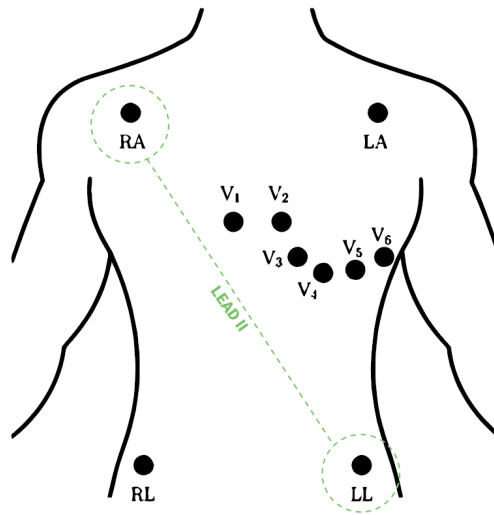


Figure 1.2: Standard placement of electrodes for a 12-lead ECG recording. Six electrodes on the chest, and four on the limbs. The term 'lead' refers to the electrical potential difference between two, or more, electrodes placed on specific points on the body. The most commonly used lead is lead II: a bipolar lead with electrodes on the right arm and left leg. Figure adapted from [190].

[71]. The derivation of these six leads and Einthoven's triangle are depicted in Figure 1.3. Lastly, the six precordial electrodes act as the exploring electrodes for the six corresponding precordial leads: (V1, V2, V3, V4, V5 and V6). The reference is a virtual electrode called Wilson's Central Terminal. This is defined by the average of the signals from electrodes LA, RA and LL and corresponds to the electrical center of the heart.

Characteristic ECG waveforms

The ECG signal provides a graphical representation of the cardiac cycle. An example of two cardiac cycles is depicted in Figure 1.4. As stated before, a cardiac cycle is initiated by an electrical impulse, generated by the SA node. This impulse propagates through the atria, which causes them to depolarize and contract. On the surface ECG this can be observed as the P-wave. The following iso-electrical PQ-segment corresponds to the slowing down of the signal in the AV node. This avoids the atria and ventricles to contract simultaneously, which would affect the blood flow between the chambers. The impulse then spreads

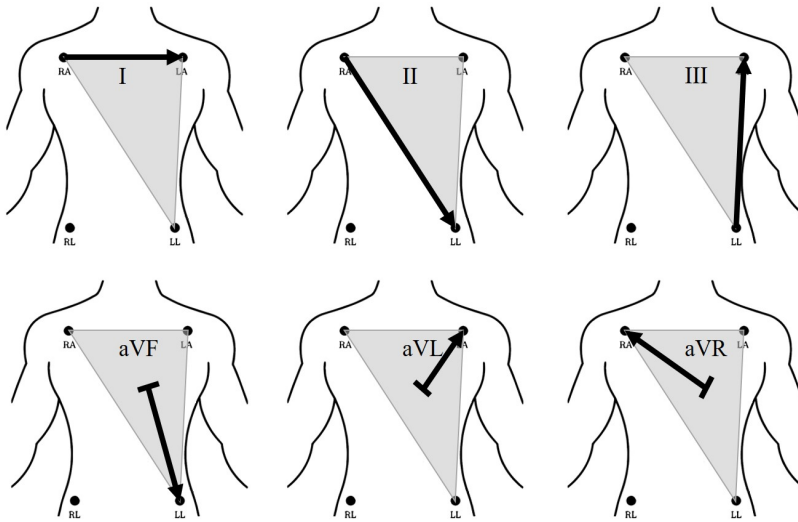


Figure 1.3: Derivation of the (augmented) limb leads and the Einthoven's triangle. The center of the triangle is referred to as Wilson's central terminal.

out in the ventricles through the bundles of His, the bundle branches and the Purkinje fibers. This initiates the depolarisation of the ventricles and thus the ventricular contraction. It can be observed as the largest deflection of the ECG, namely the QRS-complex. At the same time, the repolarization of the atria takes place. Since it coincides with the ventricular depolarization, it is masked by the QRS-complex. Lastly, the ventricles repolarize. This is represented in the ECG as the T-wave and as such, the end of the T-wave indicates the end of the cardiac cycle.

Heartbeat detection

A crucial step in the analysis of the ECG is the detection of heartbeats. These can be detected in the ECG signal by locating the QRS-complexes. These complexes are the most prominent waveforms in the ECG and contain an enormous amount of information about the state of the heart. This is why the detection of the QRS-complexes constitutes the basis for almost all automated ECG analysis algorithms [89]. Once they have been identified, more elaborate analyses, such as heart rate variability (HRV), can be conducted.

Throughout the past four decades, researchers have developed a variety of

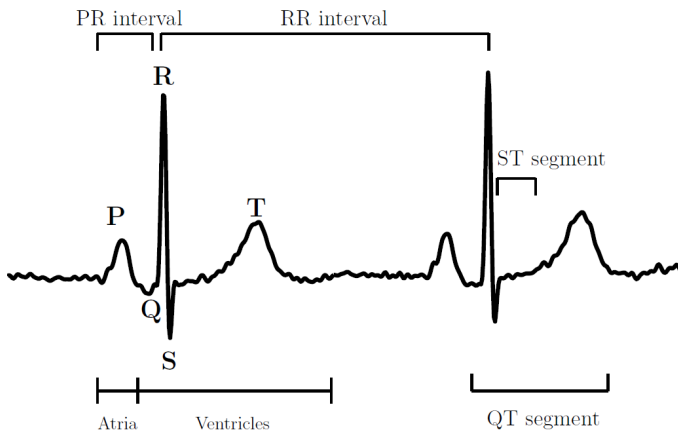


Figure 1.4: Characteristic waves and segments of a normal ECG. Two cardiac cycles are depicted. Figure taken from [190]

automatic methods to detect QRS-complexes within the ECG signal. These methods are based on, among others, derivatives, digital filters, wavelet-transforms and classifiers[146, 56, 65, 170, 39]. However, none of the proposed methodologies have proven to be completely flawless. In a review paper by Elgendi et al., they have compared the results of 22 beat detection algorithms on the MIT-BIH arrhythmia database [61]. When they compared the results of the automated algorithms with the expert annotations, they showed that many algorithms obtained excellent accuracy. However, none of the algorithms reached perfection. This means that, no matter how good the QRS detection algorithm is, it is highly likely that not all annotations are correct. Since advanced decision support tools rely on accurately detected QRS-complexes, it is recommended to visually inspect and review each signal before further analysis [149].

Many of the existing ECG toolboxes have focussed on the derivation of HRV-analysis parameters from RR-intervals. This makes sense, since most of the available hardware include some kind of QRS-complex detection algorithm. However, this does not necessarily mean that the output of these devices are the raw RR-intervals. Many of these devices have a built-in post-processing algorithm, which compensates for false detections by averaging over a certain range of RR-intervals [139, 149, 195]. However, for some analyses, such as ECG derived respiration (EDR) or beat-to-beat variability of repolarization (BVR), it is of the utmost importance that the actual R-peak of the QRS-complex is

detected. Therefore, it is necessary to visualize the actual R-peak positions in the ECG signal and allow the possibility to make manual adaptations.

During this work, we proposed an advanced tool consisting of a novel algorithm and an easy-to-use graphical user interface (GUI) for the detection and correction of R-peaks. This is discussed in the third part of the manuscript.

1.2.3 Ambulatory monitoring

Innovations in sensor technology have made it possible to record electrical signals from the heart outside of the hospital [15]. This allows patients to be monitored for several days without interfering in their daily life activities. Moreover, the extended analysis period allows to detect events that only occur occasionally and thus cannot be captured during a cardiac examination in the hospital. Nowadays, several wearable ECG recording systems exist that allow patients to record their ECG signals with, for example, a patch or a smartphone [15].

Taking the recording procedure out of the hospital also introduces some challenges. As mentioned in Section 1.1, the main challenges of ambulatory monitoring are:

- dealing with large datasets,
- dealing with contaminated signals.

Due to the presence of artefacts, the output of different data driven support tools can be corrupted. For example, when classifying segments of the signal or deriving respiration from the ECG, an artefact can be interpreted as an anomaly. Therefore, in order to obtain accurate conclusions from any ECG analysis, it is very important to filter and eliminate any type of erroneous segments.

1.2.4 Artefacts

Artefacts are electrical signals that are measured by the ECG-electrodes, but do not originate from the heart. They may be either of physiological or non-physiological origin. Muscle contractions and respiration noise are examples of physiological artefacts and electrode movement, power-line or electromagnetic interference are examples of non-physiological artefacts. All these artefacts affect the signal in a different way. For instance, muscle activity and power line interference respectively cause abrupt and continuous alterations of the

signal. Since these artefacts can have large amplitudes, they can alter the normal morphology of the ECG signal. This in turn, could lead to misdiagnosis or inappropriate treatment decisions, which we would like to avoid at all cost.

Typical examples of noise and artefacts contaminating the ECG are:

- Power line interference: noise generated by the power line electromagnetic field. In terms of frequency, it is a stationary artefact, which exhibits its peak at 50 Hz in Europe and 60 Hz in the USA. Its amplitude can go up to 50% of the peak-to-peak amplitude of the ECG signal.
- Electrode contact noise: electrode movement causes deformation of the skin around the electrode site, which could result in loss of contact between the electrodes and the skin of the patient. These artefacts can occur intermittently and they are measured as drastic changes in the ECG signal with amplitudes larger than the peak-to-peak ECG amplitude.
- Motion artefacts: these are caused by movement of the patient or motion of the electrode with respect to the skin that results in impedance changes. These changes can be observed as baseline variations in the ECG.
- Electromyogram (EMG) artefacts: these artefacts are also called muscle noise [116], since they are caused by muscle contractions. Typically, this is zero-mean Gaussian noise.
- Baseline drift/wander: this is a low frequency artefact that is mostly due to electrode-skin impedance changes. They occur as a result of breathing, perspiration, and others. The frequency content of baseline wander is typically around 0.5 Hz.

Figure 1.5 shows four of these typical ECG artefacts. In the Figure, they are depicted separately, but in reality they might coexist.

Artefacts with no spectral overlap with the physiological components of the ECG signal can be easily removed by applying the appropriate filter. For instance, power line interference (Figure 4.1c) can be removed with a notch filter around the peak frequency and baseline wander (Figure 4.1d) can be removed with a high pass filter. Artefacts with spectral overlap present a more difficult challenge. An example of this type of artefacts are motion artefacts. Due to their spectral overlap and non-stationarity, they cannot be effectively removed by traditional filtering alone.

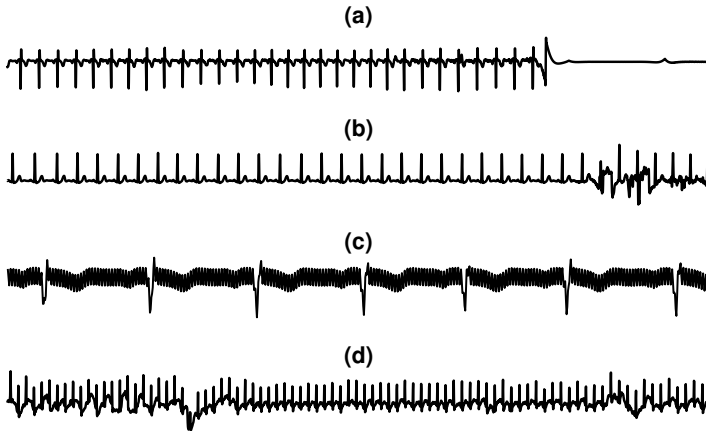


Figure 1.5: Examples of ECG segments contaminated by artefacts. (a) Electrode contact noise, (b) EMG artefact, (c) Power line interference, (d) Baseline wander

1.2.5 Cardiorespiratory coupling

The two main goals of the respiratory system are to deliver oxygen (O_2) to the cells, at a rate adequate to satisfy their metabolic needs, and remove the produced carbon dioxide (CO_2) out of the body. This is achieved through the action of respiration. The latter consists of four vital parts:

1. Pulmonary ventilation
Breathing air into and out of the lungs.
2. Pulmonary gas exchange
The exchange of O_2 and CO_2 between the lungs and the blood.
3. Respiratory gas transport
The transportation of O_2 via the blood between the lungs and the cells.
4. Peripheral gas exchange
The exchange of O_2 and CO_2 between the blood and the cells.

The need for O_2 depends on what we are doing. For example, during exercise, the need for O_2 inhalation and CO_2 exhalation increases, which results in an

increased BR. These changes are mainly modulated by the ANS, however, it is possible to consciously override its effect.

The neuronal control mechanisms of the HR and BR are closely coupled. This cardiorespiratory coupling (CRC) is perhaps best typified by the occurrence of the respiratory sinus arrhythmia (RSA) [66]. The RSA is characterized by a HR increase during inspiration and a HR decrease during expiration. The mechanism of RSA and the factors affecting it have been studied extensively [3, 176, 90, 16], yet the physiological function of RSA remains under debate.

Hayano et al. hypothesized that the physiological function of RSA is to improve the energetic efficiency of gas exchange in the lungs by matching perfusion and ventilation [75]. However, recent evidence does not support this hypothesis, instead it suggests that RSA assists the heart in reducing its workload while maintaining healthy oxygen levels [16, 135].

In Chapter 9 of this manuscript we explored whether the strength of the CRC could be linked to a quantitative measure, such as VO_2 , during exercise.

1.3 State of the art

We can divide the ECG signal analysis pipeline into five major blocks: signal acquisition, pre-processing, heartbeat detection, feature extraction and applications (Figure 1.6).

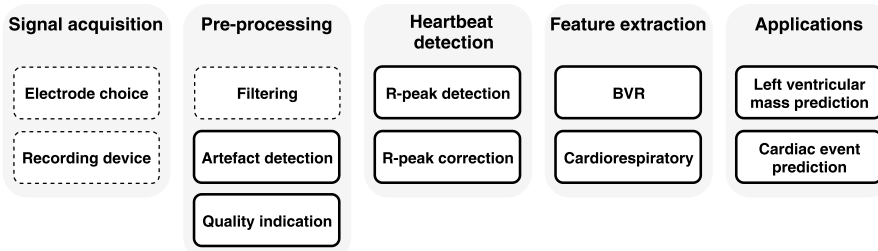


Figure 1.6: ECG analysis pipeline. The full and striped lines, respectively, indicate that the topic was or was not investigated during this work.

During this PhD, the main focus was on the four last blocks, which include the algorithmic part of the analysis pipeline.

After the signals are acquired, they need to be prepared for heartbeat detection and/or feature extraction. Since most data driven support tools assume clean data it is of utmost importance that the to-be-analysed signals are as clean as possible. Ambulatory signals present an additional challenge, since they inherently contain more noise, compared to signals measured in the hospital. Therefore, the first objective of this work was to **develop robust and accurate artefact detection and quality indication algorithms**. We focussed on ECG and respiratory signals.

The studies that address this objective can be broadly divided into four groups. The first group is focused on instrumentation. They aim to provide alternative instruments for signal recording, such as dry electrodes and smaller sensors, as well as different electrode placements that are less susceptible to artefacts [144].

The second group uses blind source separation techniques, which separate the sources that contain physiological information from the ones containing artefacts. An example can be found in the research of He et al. where Independent Component Analysis (ICA) was used to remove artefacts [76]. Many variants on this approach exist, such as Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA) and Empirical Mode Decomposition (EMD) [38]. These methods achieve good performances in removing baseline drifts, estimating in-band noise and recovering the morphology of the ECG [119], but they have clear disadvantages. For example, they require multichannel ECG recordings. These are most often available in a clinical setting, however in an ambulatory setting, the ECG is typically measured using a single lead setup. In [118], a solution was proposed for this problem by combining empirical-mode decomposition and ICA. However, the selection of the number of components that characterize the noise remains arbitrary. Additionally, these methods are very sensitive to small changes in either the signal or the noise [160].

The third group uses adaptive filters to remove the artefacts. For example, Tong et al. demonstrated that motion artefacts can be reduced by an adaptive filter with the output of an accelerometer (ACM) sensor on the left arm as the reference input [181]. Additionally, many other types of adaptive filters have been applied to this problem including least mean squares [81] and recursive least squares among others [109, 153]. The main disadvantage of adaptive filters is that they do not only alter the artefacts, but they also adapt the signal of interest. In particular, since some artefacts resemble the signal of interest, an adaptive filter that successfully models the artefact will also affect the physiological component of the ECG in clean segments.

If none of the above techniques managed to effectively clean the signal, it is safe to state that the signal is contaminated beyond repair. The only option left, is to detect the contaminated segments and remove these for further analysis.

This is the focus of the fourth and last group.

A variety of signal quality indices and algorithms were proposed in line with this approach as a result of the Physionet/Computing in Cardiology (CinC) challenge of 2011 [202], [204], [132], [43]. The challenge aimed at encouraging the development of software for mobile phones by recording an ECG and providing useful feedback about its quality [171]. The best performing algorithm was developed by Xia et al. [202]. In line with other proposed methods, their algorithm consists of multiple stages, namely: flat line detection, missing channel identification and auto- and cross-correlation thresholding. Amplitude features such as, minimum and maximum amplitude or range and differences with previous samples were also frequently used. However, due to different saturation levels between recording devices, these features restrict the usability of the algorithms.

Most of these approaches consist of a simple binary, clean or contaminated, classification. One example of where this might be used is at the front end of the acquisition process. The user could be provided with rapid, binary feedback and, if required, make adjustments in the recording set-up or, worst case scenario, re-start the recording. However, different study objectives require different quality levels. For instance, HRV studies do not require the same, high, quality level as beat classification studies. This prompted a number of authors to propose a multi-level quantification of the signal quality [154], [183], [102]. For example, Li et al. proposed a five level signal quality classification algorithm which divided the signals into five bins: clean, minor noise, moderate noise, severe noise and extreme noise [102].

In this PhD, we proposed a novel method for the detection of artefacts and quantification of the signal quality. In order to avoid recording device specific features, we used the autocorrelation function (ACF) to characterize the ECG signal. This has shown to facilitate the separation of clean and contaminated segments [194]. From the ACF, three descriptive features are extracted and fed to a RUSBoost classifier. The main novelty is the new approach to ECG signal quality assessment. We suggested to exploit the posterior class probability of the RUSBoost classifier and to use the probability for the clean class as a novel quality assessment index. This allows users to identify periods of data with a pre-defined level of quality, depending on the task at hand.

Research into artefact detection for respiratory signals is rather limited. Previous studies have focused mainly on removing the cardiac component of BioZ signals [167, 166, 120] or removing motion artefacts using adaptive filtering approaches [5, 157]. Detecting and removing contaminated segments has only recently picked up interest. For example, in [121, 36] the authors handcrafted characteristic features and used heuristics to separate clean from contaminated

segments. We believe that machine learning algorithms could improve upon these methodologies. Therefore, in this PhD, we investigated the use of machine learning algorithms to separate clean from noisy BioZ signals.

When signal quality is ensured, the next step in most cardiac analysis tasks is the detection of heartbeats [89]. This sounds like a straightforward task, but after four decades of automated heartbeat detection research this still remains a challenge.

In a review paper, Elgendi et al. compared the results of 22 beat detection algorithms on the MIT-BIH arrhythmia database [61]. When evaluating the results of the automated algorithms with expert annotations, they have shown that many algorithms obtained excellent accuracy. However, none of the algorithms reached perfection. This means that, no matter how good the QRS detection algorithm is, it is highly likely that not all annotations are correct. Therefore, it is recommended to visually inspect and review each signal before further analysis [149].

Many of the existing ECG toolboxes have focused on the derivation of HRV-analysis parameters from RR-intervals, the time between subsequent R-peaks. This makes sense, since most of the available hardware include some kind of QRS-complex detection algorithm. However, this does not necessarily mean that the output of these devices are the raw RR-intervals. Many of these devices have a built-in post-processing algorithm, which compensates for false detections by averaging over a certain range of RR-intervals [139, 149, 195]. However, for some analyses, such as EDR or BVR, it is of utmost importance that the actual R-peak of the QRS-complex is detected. Hence, as a second objective, we wanted to **develop user friendly tools that aid in analysis ECG signals**, both for R-peak detection and correction and more extended analysis.

Once all heartbeats are correctly located, more elaborate analysis can be performed, for example cardiorespiratory coupling. This coupling is perhaps best typified by the occurrence of the RSA, which is characterized by a HR increase during inspiration and a HR decrease during expiration [66]. Despite the limited understanding of the mechanisms and function of RSA, it has been suggested as a potential biomarker for people's health status [192]. In this PhD, we **explored the added value of both ECG and cardiorespiratory analysis in sport applications**.

Currently, the role of RSA during exercise is not yet fully understood. In this PhD, we explored the relationship between the strength of the RSA and different physiological parameters that are monitored during a submaximal exercise test.

Both ECG and respiration can be recorded at the same time using state-of-the-art ambulatory monitoring devices. However, when recorded during daily

life activities, their quality could be degraded by artefacts. The latter can be resolved by integrating the signal quality algorithms developed for the first objective. Moreover, the accuracy of the R-peak detection algorithm could be investigated with toolbox developed for the second objective.

1.4 Research objectives

1. Accurate artefact detection and quality indication for ECG and respiratory signals.

The aim was to develop algorithms that are able to accurately distinguish clean from noisy wearable signals. Additionally, we also investigated a way to quantify the level of contamination. The developed algorithms were compared with the state-of-the-art using freely accessible and newly labelled datasets.

2. Develop user friendly tools that aid in analysing ECG signals.

During this work, we developed many analysis tools for ECG and other biomedical signals. The tool that is furthest in development is R-DECO: a tool for detecting and correcting R-peaks. It is freely accessible and is currently being used both within and outside of our group.

3. Explore the added value of ECG and cardiorespiratory analysis in sport applications.

In this part, we used the algorithms and tools from the previous parts to explore the strength of the cardiorespiratory coupling during a submaximal exercise test. Additionally, we investigated the possibility to predict the left ventricular mass (LVM) of young endurance athletes with ECG derived features and the added value of dual-energy X-ray absorptiometry (DXA) and demographic features.

1.5 Chapter overview and main contributions

We divided the manuscript into four main parts. Part I consists of two background Chapters, which provide the reader some insight into the effects that sports have on the heart and an introduction to the machine learning algorithms that are used throughout this manuscript. Part II focuses on deriving signal quality metrics for both ECG and respiratory signals. It also includes a sidestep to transfer learning. Part III describes the most important ECG analysis tools that were developed during this work. Part IV investigate the added value of

ECG and cardiorespiratory analysis in sport applications. The structure of this manuscript is illustrated in Figure 1.7.

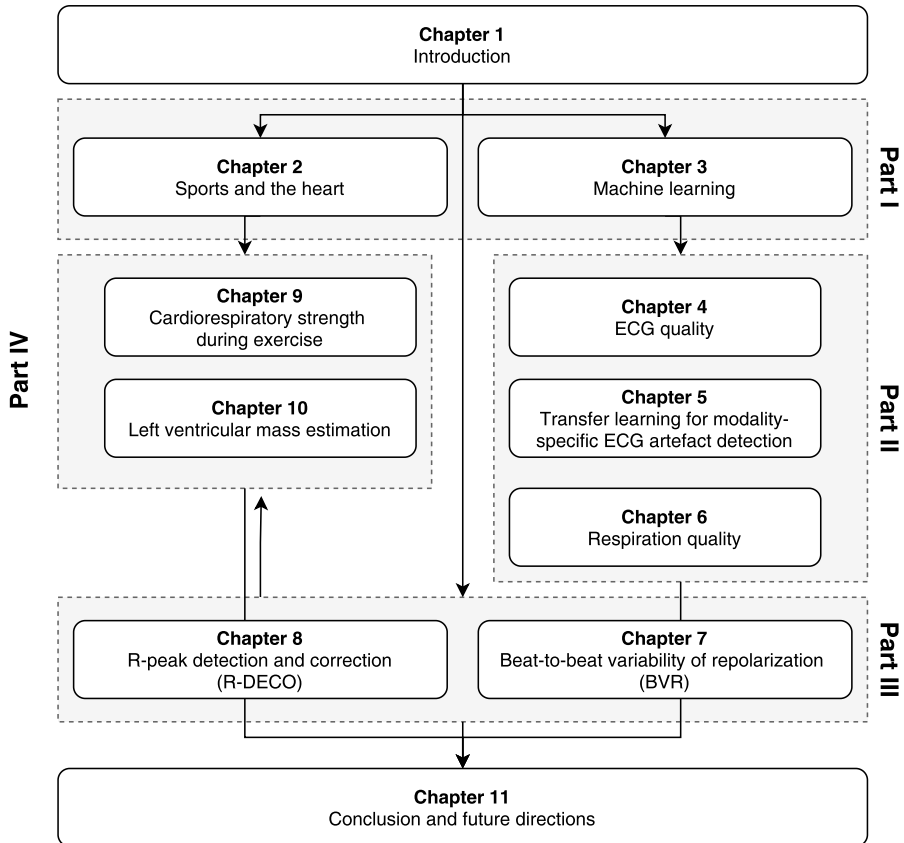


Figure 1.7: Graphical overview of the structure of this manuscript.

1.5.1 Part I. Background

The first part of this manuscript consists of two Chapters. These Chapters provide some clinical and mathematical background to the topics that are discussed throughout this PhD.

Chapter 2 describes the machine learning techniques that are used throughout this PhD. The majority of the studies use classification methods to group signals

into two classes, but we also used one regression method. The methods described in this Chapter are support vector machines (SVM), random under sampling boosting (RUSBoost) and convolutional neural networks (CNN). Additionally, we described the performance measures that we used to compare the different methods.

Chapter 3 summarizes the acute and chronic cardiovascular responses to exercise. It highlights the benefits, but also the risks associated with long-term endurance training. Additionally, we formulated three research questions, which we tried to answer in part III.

1.5.2 Part II. Signal quality

The second part of this manuscript contains three Chapters which are all related to biomedical signal quality estimation. The first two Chapters focus on ECG signals, recorded with different modalities, and the last Chapter focusses on respiratory signals.

Chapter 4 discusses a novel approach to detect artefacts in ECG signals. We used the autocorrelation function to highlight the differences between clean and noisy segments and used RUSBoost as a classifier. The model is tested on three different datasets and compared to both a heuristic and a more elaborate machine learning algorithm. Additionally, we proposed to use the probability of the clean class, as given by the model, as a novel metric for signal quality. This allows users to choose a desired quality level, depending on the problem at hand. This work is published in [128].

Chapter 5. The rapid development of new ECG recording hardware, such as capacitatively coupled ECG (ccECG), allows patients to be monitored outside the hospital. However, it also introduces new artefacts, which are unknown to previously trained artefact detection models. A solution would be to adapt an existing model, so that it is able to detect these new artefacts. This approach is called transfer learning. In this Chapter we adapted our artefact detection algorithm with a transfer learning approach designed for SVM models. The results of this work are published in [127].

Chapter 6. Numerous approaches have been presented to detect and remove noisy ECG segments, among which the approach presented in the previous Chapter, but respiratory quality indication is a relatively unexplored area of research. In this Chapter we present two novel algorithms for the detection of artefacts in respiratory signals.

1.5.3 Part III. Tools for ECG analysis

During my PhD, I came into contact with a lot of medical doctors and other researchers working in the field of ECG signal processing. Despite having a lot of data available and a lot of ideas to test, the necessary tools were not always available or did not yet exist. Therefore, I created a couple of general and tailored tools for both doctors and other researchers to aid in testing their ideas.

Chapter 7 describes a novel tool to detect and correct R-peaks. Many existing R-peak detection algorithms obtain excellent accuracy, but despite decades of research, none of the algorithms is flawless. The presented tool includes an R-peak detection algorithm, the performance of which is comparable with the state-of-the-art, and also tools to manually and semi-automatically adapt wrongly labelled beats. This work is published in [124].

Chapter 8 originated from the hypothesis that spontaneous non-sustained ventricular tachycardia (nsVT) in patients with ischaemic heart disease (IHD) could be predicted by BVR changes. However, this idea could not be investigated, since a tool was lacking that could derive BVR values in a (semi-)automated way. Therefore, we developed an extension to the tool presented in the previous Chapter which is able to compute the BVR in a semi-automated way. It allows to (semi-)automatically detect and correct the QRS-onset and the end of the T-wave, which are needed to derive the BVR. We used this tool to investigate the original hypothesis. The results of this work are published in [125].

1.5.4 Part IV. Sport applications

Chapter 9. In this Chapter, we implemented the algorithms and tools from the previous parts into our analysis pipeline with the aim to explore how the strength of the cardiorespiratory coupling changes during a maximal exercise test. Additionally, we conducted a preliminary study to explore whether this coupling is related to cardiorespiratory fitness parameters.

Chapter 10 discusses the possibility to estimate the left ventricular mass (LVM) of young endurance athletes with regularly measured features derived from ECG, DXA and demographics. The obtained estimations were compared with the gold standard LVM measurements, namely cardio magnetic resonance (CMR) imaging.

1.6 Collaborations

This PhD research was conducted in the biomedical data processing research group (BIOMED), which is part of the STADIUS center for dynamical systems, signal processing and data analytics in the Department of Electrical Engineering of KU Leuven, under the supervision of professor Sabine Van Huffel and co-supervision of professor Rik Willems and professor Carolina Varon. I collaborated with several people both within and outside the BIOMED group, leading to different conference and journal publications.

Within BIOMED, headed by Sabine Van Huffel, the first major collaboration was with Carolina Varon, John Morales and Amalia Villa. Together we have annotated over 10000 ECG signals on quality. These labels were used as ground truth in conference and journal publications [130, 129, 128]. The collaboration with John also resulted in a journal paper on model-based evaluation of methods for respiratory sinus arrhythmia estimation, which was published in IEEE Transactions on Biomedical Engineering [136].

Additionally, I collaborated with Mario Lavanga during his search for an automated perinatal stress calculator. The results of this study were presented during the IEEE Engineering in Medicine and Biology Society (EMBC) [97]. Hereafter, I assisted Mario and Elisabeth Heremans in their study on the development of premature infants' ANS. This study was recently published in Frontiers in Physiology [96].

Lastly, I worked together with Nick Seeuws for the development of a new ECG quality indicator using an auto-encoder approach. Moreover, he introduced me into the world of neural networks and aided me to complete the Chapter on respiratory quality.

During the first three years of my PhD, we closely collaborated with the research department of Imec, headed by Chris Van Hoof, more in particular with Walter De Raedt, Elena Smets and Ivan Castro. This was part of the joint B-project 02-DS-SMART diagnostics from long-term cardiorespiratory monitoring using wearable sensors (2017-2019). At the time, they were finishing a major stress study, called SWEET, which included the recording of biomedical parameters, such as ECG. Our group was given the opportunity to analyse the data. Part of the results of that analysis is included in Chapter 4. Additionally, Ivan worked on a very similar topic as myself, namely signal quality indication, but he used ccECG signals [31]. This presented the ideal opportunity to test the transfer learning approach developed in our group by Thomas De Cooman [51]. The results of this analysis are published in Sensors [127].

Different clinical partners were involved in this thesis. Each of them helped to

broaden my clinical knowledge and interpret the obtained results from a clinical point of view. This was only possible due to a strong collaboration between our group and the department of Cardiology of the University Hospitals Leuven. Prof. dr. Rik Willems, co-supervisor of this PhD, and dr. Matthew Amoni were involved in the clinical interpretation and analysis of most of the results of this thesis.

The collaboration with Matthew was mostly focussed on BVR analysis. Matthew presented the initial results of this work with a poster at the European Heart Rhythm Association conference in 2018 [2]. And I presented an extension of this work at the Computing in Cardiology conference of 2018 [125]. The results of this collaboration are described in Chapter 8.

The ongoing collaboration with prof. dr. Rik Willems and my profound interest in sport analysis resulted also in an invitation to join the ProAtHeart study. This is a longitudinal cohort study in elite endurance athletes aiming to phenotype the structural and functional cardiovascular adaptations associated with high volume exercise. In this study, young, competitive endurance athletes are examined regularly during a long-term follow-up. It is a collaboration between the university hospitals of Leuven (Belgium) and the Baker Heart and Diabetes Institute (Australia). As part of this collaboration, we investigated the possibility to predict left ventricular mass from a multitude of parameters. The results of this work were presented at the Computing in Cardiology conference of 2020 [126] and are described in Chapter 10.

Besides the department of Cardiology, our group also has a strong relationship with the division of Rheumatology in the University Hospitals. This was mainly due to the extensive collaboration during the setup of the Sensor-Based Platform for the Accurate and Remote monitoring of Kine(ma)tics Linked to E-health or SPARKLE project. As an extension of the work done in this project, I successfully supervised two design projects and two master theses. Additionally, I extended the annotation and segmentation toolbox (SegmentLab) of Lieven Billiet.

Throughout my PhD, I contributed to the EU Interreg VA project Nano4Sports (C14/17/07). This project aimed to stimulate technological innovation in the sports world. I mainly contributed to the work package that focused on data analysis. The ECG analysis tools in Chapter 7 fit within this work package.

In my third year, I had the opportunity to go on a research stay to the Biomedical Signal Interpretation and Computational Simulation or BSICoS group in Zaragoza. My work there was guided by prof. dr. Pablo Laguna, prof. dr. Juan-Pablo Martínez and prof. dr. Raquel Bailón. It was mainly focussed on investigating the added value of T-wave alternans during exercise

testing and the changes in cardiorespiratory coupling during exercise. The results of the latter analysis can be found in Chapter 9. Additionally, our common interest in cardio-respiratory coupling led to the construction of a cardio-respiratory toolbox by John Morales, Pablo Armanac and myself. Pablo used this framework to analyse weaning readiness of intensive care unit (ICU) patients in [6].

The guidance of master thesis students became an important part of my PhD after the first year of my PhD. I was the main supervisor of five master students. I supervised my first master thesis student, Felipe Novillo, together with Simon Van Eyndhoven. The main aim of this thesis was to extract respiration from magnetic resonance imaging at rest and during exercise. The results of this thesis were published in *Physics in Medicine & Biology* [140]. The last master thesis student that I guided was Ben Jacobs. This was in co-supervision with Amalia Villa. In this thesis, we investigated whether Laplacian eigenmaps could be used to monitor Tetralogy of Fallot patients. The results were presented at the Computing in Cardiology conference of 2020 [79]. Additionally, I also supervised the master theses of Marijke Coenen, Daphne Daems, Ellen Copmans and Jolien De Donder.

The last year of my PhD, we have been working closely together with the research group of Imec-NL that is focussed on wearable monitoring and algorithms in the clinical domain. This was part of the AAA79-project on cardiorespiratory monitoring (2020), supervised by Willemijn Groenendaal. The main result of this collaboration from my part was the development of an artefact detection tool for respiration (Chapter 6).

I would like to acknowledge and thank all the people that I worked and collaborated with. Their insight, knowledge and support heavily influenced the quality of the work presented in the following Chapters.

Part I

Background

Chapter 2

Machine learning

This Chapter introduces the machine learning techniques that are used in this thesis. In the first Section, we introduced three types of supervised classifiers: support vector machines, RUSBoost and convolutional neural networks. In the second Section we introduced support vector machines for regression. Each Section includes a subsection that describes how the results of these models can be evaluated.

2.1 Introduction

Machine learning is the field of study in which algorithms learn to detect specific patterns or predict certain events through a systematic discovery of statistically significant patterns in the available data. It was first defined in 1959 by Arthur Samuel who described it as "*a field of study that gives computers the ability to learn without being explicitly programmed*" [161]. Today, we can find examples of machine learning all around us. For example, Spotify proposes songs based on what you have listened to, Facebook uses your search history to show tailored advertisements and your cellphone estimates your sleep stages using the embedded accelerometer as input.

We can distinguish two main types of learning, namely supervised and unsupervised learning [74]. In supervised learning, the algorithm is exposed to a training dataset for which the ground truth is known. It can then use this prior knowledge to optimize the model based on the performance of the training set. Unsupervised learning does not require a ground truth. The algorithm learns the underlying structure of the data itself. The advantage of this type of methods is that there is no need to annotate the training data, which makes it easier to collect and analyse large training data sets. However, the disadvantage is that when a ground truth is available, they tend to be outperformed by their supervised counterparts. In this manuscript, we used only supervised techniques.

Supervised learning techniques aim to learn a function that optimally maps an input to an output. Consider a training set $\{x_i, y_i\}_{i=1}^N$, with $x_i \in \mathbb{R}^D$, as the input training data, y_i as the corresponding output target values, N as the length of the feature vector and D as the dimension of the feature space. The goal is then to learn a mapping function $f(\cdot)$, such that $y = f(x)$.

These techniques can be further divided into regression and classification problems. In regression problems, the predicted outcome is continuous, whereas in classification problems, the predicted output is a discrete value, which represents a class or label. The most used supervised learning technique in this work is classification.

This Chapter describes the machine learning techniques that are used throughout this work. They include and are limited to support vector machines (SVM), random under sampling boosting (RUSBoost) and convolutional neural networks (CNN). Section 2.2 describes the classification techniques and Section 2.3 describes the regression techniques.

2.2 Supervised classification

Here, the target value y corresponds to discrete class labels and the function $f(\cdot)$ maps the input data onto a space where the decision boundary could be easily defined. In order for this technique to work, it requires a database of signals with known class labels. A multitude of techniques exist to determine the decision boundary. In the following Sections we described the techniques that we used in this work.

2.2.1 Support vector machines

Support vector machines (SVM) are a set of supervised learning methods used mainly for classification and regression [49]. It has been shown that they obtained mostly good performances on classification and regression tasks in many types of applications [117]. They are also known to have good generalization properties.

The main goal of SVMs is to find a hyperplane that best separates the different classes in the dataset. This is obtained by maximizing the margin: the distance between the hyperplane and the points closest to the decision boundary. These points are also called the support vectors.

SVMs were originally designed for linearly separable data [186]. Hereafter, the technique was extended to non-separable cases [44]. In this formulation, misclassifications are allowed by introducing a soft margin constant, a slack variable, that penalizes points that lie on the wrong side of the margin. Finally, the technique was extended to non-linear classification [188].

In the non-linear formulation, the input data is mapped to a higher dimensional feature space where the samples can be linearly separated by the hyperplane $w^T \phi(x) + b$, with $\phi(\cdot)$ as the mapping function, w as the unknown weight vector and b as an unknown bias term (Figure 2.1).

The optimization problem of the non-linear, non-separable SVM classification problem is formulated as

$$\begin{aligned} \min_{w,b,\xi} \mathcal{J}_{\mathcal{P}}(w, \xi) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad &\begin{cases} y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, \forall i \in [1, N], \end{aligned} \tag{2.1}$$

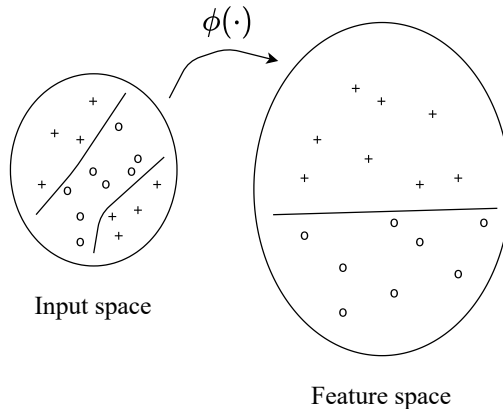


Figure 2.1: The input space is mapped to a high dimensional feature space where the classes can be linearly separated. This corresponds to a non-linear separation in the input space.

with w the weight vector, C a regularization constant, and ξ_i the slack variables introduced to relax the constraints.

The dual formulation of this optimization becomes the following quadratic programming problem:

$$\begin{aligned} \max_{\alpha} \mathcal{J}_{\mathcal{D}}(\alpha) &= -\frac{1}{2} \sum_{i,j=1}^N y_i y_j K(x_i, x_j) \alpha_i \alpha_j + \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad \sum_{i=1}^N \alpha_i y_i &= 0, \quad 0 \leq \alpha_i \leq C, \quad , \forall i \in [1, N], \end{aligned} \quad (2.2)$$

When solving equation 2.2, many of the Lagrange multipliers α_i are equal to zero. This way, sparsity is added to the problem. The samples with $\alpha_i \neq 0$ are the support vectors and only the distances between those samples and the hyperplane are maximized. The final formulation of the SVM is then:

$$y(x) = \text{sign} \left[\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right]. \quad (2.3)$$

As can be seen in equation 2.2 and 2.3, neither the dual problem, nor the classification function contain the mapping function $\phi(\cdot)$. Only the inner

products of the feature maps are required. If we have a positive definite kernel function:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j), \quad (2.4)$$

then we do not need to compute the mapping function. This is known as the kernel trick. The main advantage of the kernel trick is that it avoids the cost-intensive calculation of the high-dimensional feature space.

Some of the most used kernel functions are:

- The linear kernel: $K(x, x_i) = x_i^T x$.
- The polynomial kernel: $K(x, x_i) = (\mathcal{T} + x_i^T x)^d$, with \mathcal{T} a positive constant and d the degree of the polynomial [184].
- The Radial Basis Function (RBF) kernel: $K(x, x_i) = \exp(-\frac{\|x-x_i\|_2^2}{\sigma^2})$, with σ as a hyperparameter that defines the non-linearity of the kernel. The higher σ , the more non-linear the decision boundary.

2.2.2 RUSBoost

Most learning algorithms assume that training sets used for learning are balanced. However, this is not always the case in real world data. The samples in one class could largely exceed the samples of the other class. This is known as the class imbalance problem and is often reported as an obstacle to the training of a well performing Machine Learning model.

An example of this problem can be found in the outcome of the non-invasive prenatal test. This a non-invasive screening test for the detection of several syndromes, performed 12 weeks after gestation. The majority of the babies does not have any syndrome, which results in an imbalanced dataset. If you would train a classification model on such a dataset, than the over-representation of the majority class could lead to a tendency of the model to classify most samples as belonging to the majority class [165]. Such a model would be ineffective at identifying examples of the minority class, which is frequently the class of interest, especially regarding syndrome detection.

Weiss listed many methods for alleviating the problem of class imbalance, such as, among others, data sampling and boosting [200]. The main objective of data sampling is to balance the class distribution. This is done either by adding samples to the minority class (oversampling) or removing samples from

the majority class (undersampling). The simplest technique for resampling a dataset, is random sampling. In the case of oversampling this means that random samples of the minority class are duplicated until a desired class ratio is obtained. Similarly, random undersampling (RUS) removes random samples from the majority class.

Both data sampling approaches have their benefits and drawbacks. The main drawback associated with undersampling is the loss of information that comes with removing samples from the majority class. This could result in a decrease of classifier performance. It has the benefit, however, of accelerating the training process, since the size of the training data set is reduced. Opposite, we have oversampling. While no information is lost during this procedure, it could increase the likelihood of overfitting, since it duplicates samples from the minority class. Additionally, the inclusion of more training samples could also increase the model training time.

Boosting is another technique that can be used to improve the classification performance on imbalanced datasets. This technique is not specifically designed to overcome the class imbalance problem, but can be used to improve the performance of any weak classifier [64].

Boosting iteratively builds an ensemble of models. During each iteration, the sample weights are altered based on the performance of the classifier of that iteration. Samples that were misclassified have their weights increased, while those that were correctly classified have their weights decreased. This forces the classifier to focus more on the misclassified samples in the next iteration.

Given the class imbalance problem, this technique is particularly effective since the minority class is most likely to be classified wrongly [200]. This results in increasingly higher sample weights during subsequent iterations and, as a consequence, an improved classification performance.

Seiffert et al. proposed to combine RUS and boosting into a hybrid, ensemble classification algorithm named RUSBoost [165]. The main drawback of RUS, the loss of information, is mostly overcome by combining it with boosting. While certain information may be absent during a given iteration of boosting, it will likely be included when training models during other iterations.

Let x_i be a point in the feature space X and y_i be a class label in a set of class labels Y . Each of the N examples in the training data set S can be represented by the tuple (x_i, y_i) . Let t be an iteration between one and the maximum number of iterations T . This corresponds to the maximum number of classifiers in the ensemble. Let h_t be the hypothesis of a classification algorithm, *WeakLearn*, on iteration t , and $h_t(x_i)$ be the output of hypothesis h_t , for x_i . Let $D_t(i)$ be the weight of the i^{th} example on iteration t and P the desired

percentage of minority samples. The algorithm is then constructed as follows:

1. Initialize $D_1(i) = \frac{1}{N}$, for all i .
2. Do for $t = 1, 2, \dots, T$
 - (a) Create a subset S'_t with weights D'_t using RUS. For example, if $P = 50\%$, then the majority class examples are randomly removed until the numbers of majority and minority class samples are equal.
 - (b) Pass S'_t and D'_t to *WeakLearn* and train the classifier.
 - (c) Calculate the pseudo-loss ϵ_t as

$$\epsilon_t = \sum_{(i,y):y_i \neq y} D_t(i)(1 - h_t(x_i, y_i) + h_t(x_i, y)). \quad (2.5)$$

- (d) Calculate the weight update parameter α_t as

$$\alpha_t = \frac{\epsilon_t}{1 - \epsilon_t}. \quad (2.6)$$

- (e) Update the weights for the next iteration so that

$$D_{t+1}(i) = D_t(i)\alpha_t^{1+h_t(x_i, y_i)-h_t(x_i, y:y \neq y_i)} \quad (2.7)$$

and normalize them such that

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{\sum_i D_{t+1}(i)}. \quad (2.8)$$

3. After T iterations, the final hypothesis $H(x)$ is returned as a weighted vote of the T weak hypotheses:

$$H(x) = \operatorname{argmax}_{y \in Y} \sum_{t=1}^T h_t(x, y) \log \frac{1}{\alpha_t}. \quad (2.9)$$

The methodology is graphically displayed in Figure 2.2.

Previous research has shown great reference results in case of class imbalance [137]. A detailed overview of the algorithm can be found in [165].

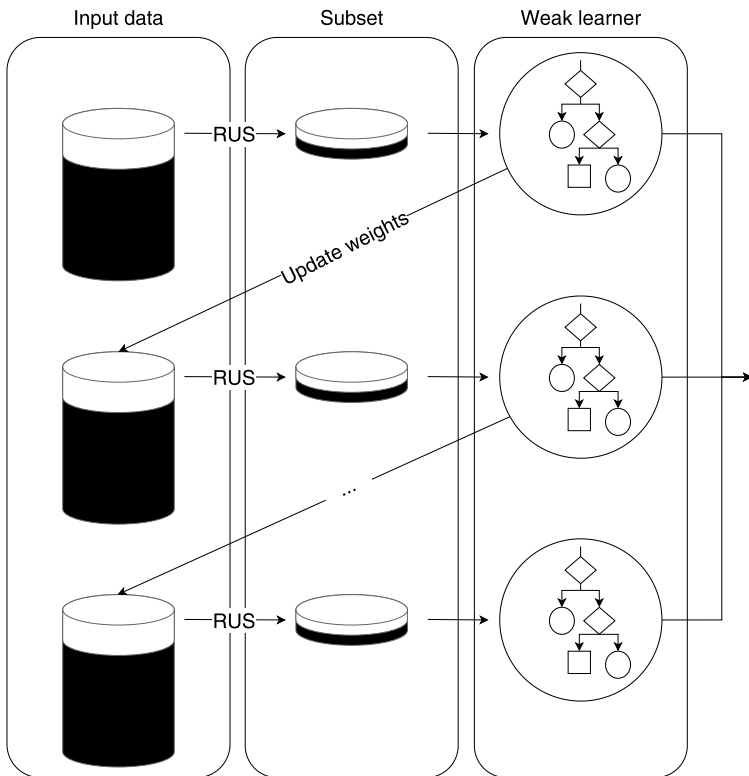


Figure 2.2: After selecting a subset using RUS, a weak learner is trained. The output of this weak learner is then used to update the weights of the input data. This procedure is repeated T times. The final hypothesis $H(x)$ is returned as a weighted vote of the T weak hypotheses.

2.2.3 Convolutional neural networks

Most classification models, such as the ones described in the previous Sections, are built using a combination of hand-crafted features and automatic learning techniques. The input data is transformed so that it can be represented by low dimensional vectors that can be easily matched or compared. This feature crafting procedure usually requires prior knowledge and is task specific. Therefore, one of the main problems with this approach is that the classification performance is largely determined by the ability of the researcher to come up with an appropriate set of features. This turns out to be a daunting task which unfortunately must be redone for each new problem [99]. This problem can be

solved by using neural networks.

Convolutional neural networks (CNNs), also known as ConvNet, are classified as a special type of feed-forward artificial neural networks (ANNs). The architecture of these networks are inspired by the organization of the mammalian visual cortex. Each neuron is stimulated only by a restricted region of the visual field, this field is known as the receptive field [168]. A collection of such fields overlap to cover the entire visual area. The main advantage of a CNN, compared to more traditional algorithms, is that there is no need to select any features manually. It is capable of building features from raw time series and images and automatically optimizes the features and classifier at the same time.

CNNs usually have a deeper structure compared to ANNs, because they need fewer parameters for the same number of hidden layers. This parameter reduction makes the computation more efficient and helps to prevent overfitting. It is achieved by three key features: local receptive fields, parameter sharing and pooling. Using local receptive fields means that each neuron is only connected to a limited region of the input volume, in contrast with fully connected neurons. As a result, the network is able to learn local spatial/temporal correlations from the input.

Parameter sharing means that all neurons located in one layer have the same parameters. This allows the network to detect a shifted target pattern, which is known as the shift invariant property. The last characteristic that makes the CNN different from other ANNs is pooling. It progressively reduces the size of the data representation and improves translational invariance.

Classical CNNs consist of multiple stacked layers of 3 different types: convolutional layer (Conv), a pooling layer and a nonlinear layer. Their functions are shortly summarized in the following sections.

Conv layer

Central to the CNN is the Conv layer. This layer performs a convolution operation. In the context of a CNN, a convolution is a linear operation that involves the multiplication of a set of weights with the input. Given that the technique was designed for two-dimensional inputs, the multiplication is performed between an array of input data and a two-dimensional array of weights, called a filter or a kernel.

The filter size is smaller than the input data. This is intentional, since it allows the same filter to be multiplied by the input array multiple times at different points. Specifically, the filter is applied systematically to each overlapping part

or filter-sized patch of the input data. Mathematically this can be formulated as follows:

$$O(i, j, k) = \sum_{p=1}^P \sum_{n=1}^N \sum_{m=1}^M f_k(m, n, p) I(i - m, j - n, p) \quad (2.10)$$

where $I(i, j, p)$ is the input of the Conv layer, (i, j, p) represent the dimensionality of the input data, and $f_k(m, n, p)$ are the coefficients of the k^{th} filter which consists of $M \times N \times P$ coefficients, where M and N represent the size of the filters, and P represents the number of filters in the previous layer. $O(i, j, k)$ is the output of Conv layer.

The filter coefficients, f_k , are the only unknown parameters of the CNN which should be found in the training process by a backpropagation method. However, the size of filters (M, N), known as receptive field, as well as the number of filters of each Conv layer should be predefined in the design process.

In this work, we only used one-dimensional input signals. The resulting formula is then the following:

$$O(i, k) = \sum_{p=1}^P \sum_{m=1}^M f_k(m, p) I(i - m, p) \quad (2.11)$$

where $I(i, p)$ is the input of the Conv layer, (i, p) represent the dimensionality of the input data, with $p = 1$ for the first layer. $f_k(m, p)$ are the coefficients of the k^{th} filter which consists of $M \times P$ coefficients.

Comparing the finite impulse response (FIR) filters of the Conv layer with common neurons in ANN shows that each filter is like a layer of simple linear neurons with two important characteristics: 1) The weights of all neurons located in the layer are shared between the neurons. 2) Neurons only connect to a limited number of inputs with overlap. Applying these two characteristics on a layer of simple neurons converts the layer to the mentioned convolutional FIR filter.

The output $O(i, k)$ from the convolution operation is called the feature map. Once a feature map is created, we can pass each value in the feature map through a nonlinear layer.

Pooling layer

A limitation of the feature map output of the Conv layers is that they record the precise position of features in the input. This means that small translations of the input will result in a different feature map.

A common approach to addressing this problem is called down sampling. This is where a lower resolution version of an input signal is created that still contains the large or important structural elements, without the fine detail that may not be as useful to the task.

The main aim of pooling layers is to create a lower resolution version of the feature map that still contains the important structural elements. In practice, taking the maximum and averaging are the two most common operations.

The output volume size of this layer in each mode equals

$$\lfloor \frac{S_{input} - S_{filter}}{stride} \rfloor + 1 \quad (2.12)$$

where S represents the size and $\lfloor \cdot \rfloor$ is the floor function. The stride of pooling should be predefined as a hyper-parameter. Note that pooling layers are specified, rather than learned. Hence, they have no trainable parameters.

Non-linear layer

Typical Conv layers are linear systems, as their outputs are affine transformations of their inputs. Due to their linear nature, they lack the ability to express non-linearities that may be present in the input data. To overcome this, non-linear layers with variable activation functions have been introduced inside CNNs.

A non-linear layer consists of an activation function that takes the feature map and generates an activation map. The activation function is an element-wise operation and thus the dimensions of the input and the output are identical.

The most commonly used activation function in CNNs is the rectified linear unit (ReLU) which is defined as follows:

$$O(x_i) = \max(0, x_i) \quad (2.13)$$

where x_i is the input value and $O(\cdot)$ is the output. In other words, this function is a half-wave rectifier which replaces the output values of the Conv layer with zero. It does not affect the size of data.

Two other commonly used activation functions are the (logistic) Sigmoid and the softmax function. The Softmax function can be written as

$$O(x_i) = \frac{1 + e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (2.14)$$

with N as the number of inputs. Hence, the Softmax output is the exponent of the individual input, divided by the sum of the exponents of all inputs. This ensures that:

- Negative inputs are converted into positive outputs.
- Each output is in the interval $[0, 1]$.

The Sigmoid function can be written as

$$O(x_i) = \frac{1}{1 + e^{-x_i}}. \quad (2.15)$$

Both functions are mostly used in the very last layers of the CNN, which are performing the classification task. This is mainly because they map the input in the interval $[0, 1]$. Hence, they can convert a real valued input into an output that can be interpreted as a probability.

2.2.4 Performance evaluation

After the implementation of a classification model, the next step is to evaluate its performance. In a supervised setting, this is done by comparing the predicted labels, as given by the model, with the true labels. Generally, this is done on a separate test set, which should not overlap with the training dataset in order to prevent bias. For example, in the case of patient data it is best to work with an inter patient approach where the patients either belong to the training or the test set.

The choice of metric depends on the given dataset and problem. The next subsections discuss the used metrics throughout this manuscript. They can be distinguished in metrics derived from the confusion matrix and the Receiver Operating Characteristic (ROC) curve.

Confusion matrix

The confusion matrix is one of the most intuitive and easiest metrics used for visualizing the correctness and accuracy of the model. It is used for classification problems where the output can be of two or more classes. Table 2.1 illustrates an example of a confusion matrix for a binary classification problem with positive and negative class values. The rows indicate the predicted labels and the columns the true labels. The number of rows and columns is equal and is defined by the number of classes, in this case two.

Table 2.1: Example of a confusion matrix for a binary problem. Correctly labelled samples are indicated in green and misclassified samples are indicated in red.

		Actual class		
		Positive	Negative	
Predicted class	Positive	TP	FP	→ PPV
	Negative	FN	TN	→ NPV
		↓ Se	↓ Sp	

The following terms are associated with the confusion matrix:

- True Positives (TP): the actual class of the data point is 1 and the predicted class is also 1. This is called "True" because the model has predicted correctly.
- True Negatives (TN): the actual class of the data point is 0 and the predicted class is also 0.
- False Positives (FP): the actual class of the data point is 0 and the predicted class is 1. This is called "False" because the model has predicted incorrectly.
- False Negatives (FN): the actual class of the data point is 1 and the predicted class is 0.

The ideal scenario would be a model that has 0 False Positives and 0 False Negatives. However, in real life this will never be obtained, since no model will be 100% accurate all the time.

The confusion matrix in itself is not a performance measure as such, but a lot of the performance metrics are based on the confusion Matrix and the numbers inside it. We can compute the following performance metrics from the confusion matrix:

- Sensitivity (Se), Recall or True Positive Rate (TPR): $\frac{TP}{TP+FN}$
- Specificity (Sp): $\frac{TN}{TN+FP}$
- False Positive Rate (FPR): $\frac{FP}{TN+FP}$
- Positive Predictive Value (PPV) or precision: $\frac{TP}{TP+FP}$
- Negative Predictive Value (NPV): $\frac{TN}{TN+FN}$
- Accuracy (Acc): $\frac{TP+TN}{TP+TN+FP+FN}$

Acc represents the fraction of samples that are correctly classified. It is the most commonly used metric for classification tasks. It is a good metric for balanced, or nearly balanced, problems. However, when the prior class probabilities are very different, the use of this measure might lead to misleading conclusions, since the Acc is strongly biased towards the majority class. For example, imagine a two-class problem where 99% of the entries belong to the positive class. If you would classify all entries to the positive class, you would get an Acc of 99%, but none of your negative entries would be correctly classified.

Some solutions to alleviate this problem were already described in Section 2.2.2, but another solution would be to use the balanced accuracy (bAcc) [26]. This can be defined as the average Acc obtained on either class. It can be written as

$$bAcc = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right). \quad (2.16)$$

If the classifier performs equally well on either class, the bAcc is equal to the conventional Acc. However, if the conventional Acc is only high due to an imbalanced test set, then the bAcc will drop to the chance level.

Receiver Operating Characteristic curve

A Receiver Operating Characteristic (ROC) curve characterizes the performance of a binary classification model across all possible trade-offs between the classifiers TPR and FPR.

It is a straightforward way to visualize and compare the performance of different models and select the optimal one. Furthermore, it can also be used to define a good decision threshold, since changing the threshold results in a different prediction and corresponding performance. The ideal threshold can be

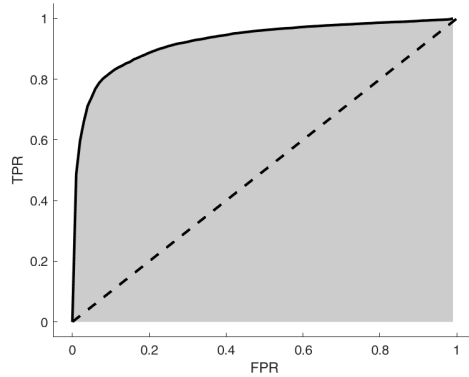


Figure 2.3: General example of an ROC-curve. The full black line indicates the ROC-curve. The striped line indicates the 50 % line. The x-axis indicates the False Positive Rate (FPR) and the y-axis indicates the True Positive Rate (TPR).

determined by selecting the point in the upper left corner of the curve (Figure 2.3).

The Area Under the ROC-Curve (AUC) is a measure of the capability of a model to distinguish between classes. A perfect classification method has an AUC value of 1, while a method that makes random guesses has an AUC of 0.5. AUC values of 0.5 are theoretically possible but are in practice indicative of wrong labelling or an incorrect training strategy.

The main advantage of this type of performance evaluation is that it does not have any bias towards models that perform well on the minority class at the expense of the majority class. This is a very useful property when dealing with imbalanced datasets.

Multiple ground truths

The aforementioned metrics are applicable when there is only one ground truth defined. However, during this PhD we (re)labelled some datasets with four or more annotators, which inevitably leads to uncertainty of the ground truth. A clear and objective set of rules could alleviate this problem, but does not lead to a guaranteed consensus. One way of quantifying the agreement is to calculate the overall percentage of agreement. However, this metric does not take the agreement into account that would be expected due to chance. Hence, in order

to have a correct estimation of the level of agreement, only the agreement beyond that expected by chance should be taken into account.

The Fleiss' kappa is a relatively simple, yet powerful, metric that considers the possibility that the agreement has occurred by chance [131]. It can be computed as

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}, \quad (2.17)$$

where \bar{P} represents the actual observed percentage of agreement and \bar{P}_e the expected agreement due to chance. \bar{P} can be computed as

$$\bar{P} = \frac{1}{Nn(n-1)} \left(\sum_{i=1}^N \sum_{j=1}^k n_{ij}^2 - Nn \right), \quad (2.18)$$

where N is the total number of segments, n is the number of annotations per segment and k is the number of categories, in this case two. n_{ij} represents the number of annotators who assigned the i^{th} subject to the j^{th} category. \bar{P}_e can be computed as

$$\bar{P}_e = \sum_{j=1}^k P_j^2, \quad (2.19)$$

where P_j is the proportion of all observations (n/N) that was assigned to the j^{th} category.

Kappa values can vary between -1 and 1, but are higher than 0 if the observed agreement is due to more than chance alone. Landis and Koch proposed a rule of thumb for the interpretation of the kappa values [95] and it is used in this manuscript to evaluate the (re)labelling results. The rule is as follows: ≤ 0 =poor, 0.01-0.20=slight, 0.21-0.40=fair, 0.41-0.60=moderate, 0.61-0.80=substantial and 0.81-1.00=almost perfect or perfect agreement.

After establishing an acceptable level of agreement, we can use the obtained labels to perform classification. In cases like this, here are different ways to measure the performance which we explain below:

- **Majority voting**

In this approach, the most frequent label is defined as the aggregated label for each segment separately. For example, a sample for which 60 % of

the annotators voted positive is labelled as positive. The main issue with this method is that the minority votes are not taken into account. Hence, the previously discussed sample has the same effect on the performance evaluation as a sample for which 100% of the annotators voted positive. This is contra intuitive, since we can assume that samples with a high percentage of agreement are more reliable than the controversial ones.

- **Weighted metrics**

Ansari et al. proposed a new set of metrics that take the agreement of the annotators into account [4]. The key aspect of this approach is to take the minority votes into account when evaluating the performance. This is done by quantifying the overall level of agreement of the annotators using Fuzzy logic.

Assuming the samples are labelled as positive (+1) and negative (-1), then two corresponding fuzzy sets can be derived and subsequently each sample has two membership values: μ_C and μ_N , which show the membership of the sample to each set. These values are defined based on the percentage of their corresponding labels. Thus, $\mu_C + \mu_N = 1$. For instance, μ_C and μ_N of a sample which is labelled as [+1, +1, -1, +1, -1] by 5 annotators are respectively 0.6 and 0.4.

The authors extended the previously mentioned classical metrics as follows:

- Weighted sensitivity (wSe)

$$wSe = \frac{\sum_{p=1}^P (\mu_C(p) O^{Sample}(p))}{\mu_C(p)} \times 100 \quad (2.20)$$

with P the total number of samples and $O^{Sample}(p)$ the predicted label of sample p .

- Weighted specificity (wSp)

$$wSp = \frac{\sum_{p=1}^P (\mu_N(p) (1 - O^{Sample}(p)))}{\mu_N(p)} \times 100 \quad (2.21)$$

Note that if only one annotator labelled the data or there is 100 % agreement among the annotators, the weighted and classical metrics are equal. Thus, these weighted metrics can be seen as a generalization of the classical metrics for multi-annotator purposes.

2.3 Regression techniques

The goal of these techniques is to predict a continuous variable. One of the best known formulations is least-squares linear regression. The aim is to find a linear function $f(\cdot)$ so that $f(x) = w^T x + b$, with w as weight vector and b offset. A solution is then found by minimizing the quadratic error with regard to the label vector y , i.e., $\min_w \sum_{i=1}^N \|w^T x_i + b - y_i\|_2^2$.

The optimization problem is often required to satisfy certain additional conditions, which can be obtained by introducing penalty terms. This is termed regularized least squares. One such condition is sparsity, a measure of model simplicity. The goal is to obtain a well-performing model by focusing on as few variables as possible. This can be obtained by using support vector machines.

2.3.1 Support vector machines

In addition to classification, the support vector methodology can also be used for linear and non-linear function estimation problems [187]. We restricted the description of the methodology to the linear implementation, since we only used a linear kernel.

The following primal problem can be formulated:

$$\begin{aligned} \min_{w, b, \xi, \xi^*} \mathcal{J}_{\mathcal{P}}(w, \xi, \xi^*) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} \quad \begin{cases} y_i - w^T x_i - b \leq \epsilon + \xi_i & , \forall i \in [1, N] \\ w^T x_i + b - y_i \leq \epsilon + \xi_i^* & , \forall i \in [1, N] \\ \xi_i, \xi_i^* \geq 0 & , \forall i \in [1, N], \end{cases} \end{aligned} \quad (2.22)$$

with ϵ as the Vapnik ϵ -insensitive loss function and $C > 0$ as a regularization constant that determines the level of tolerance with respect to the desired ϵ accuracy.

The dual formulation of this optimization becomes the following quadratic programming problem:

$$\begin{aligned} \max_{\alpha, \alpha^*} \mathcal{J}_{\mathcal{D}}(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i^T x_j \\ & -\epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (2.23)$$

$$s.t. \quad \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C],$$

The properties of this solution are comparable with the results on classification. The solution is global and unique. Moreover, sparsity is obtained, since many of the elements in the solution vector will be equal to zero.

2.3.2 Performance evaluation

Similarly to classification, after the implementation of a regression model, the next step is to evaluate its performance. This is done by comparing the predicted labels, as given by the model, with the true labels. The choice of metric depends on the given dataset and problem.

In this work, we used only two statistics to measure the goodness-of-fit, namely

- The coefficient of determination R^2 , which is computed as

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2.24)$$

with \hat{y} as the predicted label and \bar{y} as the average of the true labels. In case of a perfect prediction, $R^2 = 1$.

- The root mean square error (RMSE), which is computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}. \quad (2.25)$$

The lower the RMSE, the better the prediction.

2.4 Conclusion

In this Chapter, we provided an overview of the different machine learning techniques that were used in this research. We introduced three supervised classification methods, one regression method, and discussed several performance evaluation metrics. In the next Chapters, the various techniques will be used to classify and predict biomedical signals.

Chapter 3

Sports and the heart

3.1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of morbidity and mortality worldwide. In the United States, for instance, CVDs account for 600.000 deaths, or 25% of the total death count, each year [17], and after a continuous decline over the last 5 decades, its incidence is increasing again [159]. This puts a heavy economic burden on the health care system at both a global and a national scale [29].

Multiple risk factors are attributed to causing CVD. The following are some of the most significant risk factors: age, sex, family history, smoking, sedentary lifestyle, obesity, arterial hypertension and diabetes mellitus [58]. These risk factors can be divided in two categories: non-modifiable and modifiable risk factors. Non-modifiable risk factors consist of those conditions that a person cannot alter, for example age. Modifiable risk factors are conditions that can be altered by making certain lifestyle changes.

A sedentary lifestyle is one of these modifiable risk factors. It can be characterized by consistently low levels of physical activity. Besides being a risk factor for CVD, it is also one of the key determinants of the growing rates of obesity, another risk factor for CVD [113]. Conversely, regular physical activity is associated with remarkable widespread health benefits and a significantly lower CVD risk.

Several long-term studies have shown that increased physical activity, at any intensity, is associated with a reduction in all-cause mortality [60, 20, 83].

Additionally, it may modestly increase life expectancy, an effect which is strongly linked to a decline in the risk of developing cardiovascular and respiratory diseases [155]. Consistent with this notion, death rates among men and women have been found to be inversely related to cardiorespiratory fitness levels, even in the presence of other modifiable risk factors such as smoking, hypertension, and hyperlipidemia [175, 19].

Despite the fact that moderate levels of exercise are consistently associated with a reduction in cardiovascular disease risk, there is evidence that suggests that continuous high levels of exercise could have detrimental effects on cardiovascular health [93]. For example, a study that compared healthy controls with endurance athletes has estimated that the AF incidence among the athletes was 5.3%, compared to the 0.9% among the control subjects [84]. These potential associations between training induced cardiac remodelling and arrhythmias is still a subject of debate.

The ECG is currently mostly being used for screening purposes. For example, pre-participation cardiovascular screening of young athletes has been endorsed by both scientific organisations and sporting governing bodies [55]. During this PhD we investigated the changes in ECG parameters due to cardiovascular training adaptations.

In the next Sections, we discuss the acute cardiovascular response and the long term adaptations induced by endurance exercise. Each section is accompanied by a research question, which we tried to answer in part II of this manuscript.

3.2 Acute cardiorespiratory response to exercise

When a person exercises at progressively higher work loads, there is a linear increase in oxygen uptake to match the demand of the active skeletal muscles. This increase continues until the maximal oxygen consumption (VO_2max) is reached. Hence, the VO_2max is an indicator of the system's maximal oxygen uptake, as well as a biomarker of cardiorespiratory fitness (CRF). The determinants of oxygen consumption are summarized in the Fick equation as follows [62]:

$$VO_2 = CO(CaO_2 - CVO_2) \quad (3.1)$$

with VO_2 as the oxygen consumption, CO as the cardiac output, and $CaO_2 - CVO_2$ as the O_2 difference between arterial and venous blood ($a-vO_2$ difference). The Fick Equation states that VO_2max equals the amount of blood pumped

per heart beat, at maximum heart rate, times the amount of oxygen the working muscles are able to extract from the blood passing through them. Thus, there are two major factors that influence VO_{2max} : cardiac output and the capacity of the active muscles to extract oxygen from arterial blood. The VO_{2max} is also influenced by other factors, such as arterial O_2 , but these are out of the scope of this PhD.

During exercise, muscle and coronary blood flow increase, whilst cerebral blood flow is maintained constant and splanchnic flow diminishes. Resting blood flow to the muscles is usually 2–4 ml·100 g/muscle/min, but might increase to nearly 100 ml·100 g/muscle/min during maximal exercise [28]. This increase in blood flow to the muscles requires an increase in the CO . Since CO is the product of the stroke volume (SV) and the heart rate (HR), it is increased by both a rise in the HR and the SV . This attributes to a more complete emptying of the heart by a forcible systolic contraction.

At the onset of exercise, HR and SV increase so that CO closely matches the metabolic demand of the working skeletal muscles. Hereafter, the CO increases linearly with exercise intensity, from a resting value of approximately 5 L/min to 15-40 L/min, depending on the gender and the level of conditioning [59].

In untrained individuals, the SV increases with increasing work load up to approximately 40%-50% of the VO_{2max} and then plateaus [9]. The rest of the increase in CO with increasing exercise intensity is primarily mediated by an increasing HR . In endurance athletes, the CO is mediated by the SV , rather than an increase in HR . Gledhill et al. have found that the SV progressively increases throughout all increasing work loads [70]. This pronounced effect of exercise training on maximal CO can only be obtained by a structurally and functionally improved heart in the trained state.

Since the maximal HR decreases with exercise training [205], the increase in cardiac output associated with exercise training has to be the result of a larger SV . During exercise, the SV rises as a result of increases in LV end-diastolic volume and, to a lesser extent, sympathetically mediated reduction in end-systolic volume. LV end-diastolic volume is determined by diastolic filling, which is determined by a complex interplay of factors. An increase in SV with exercise training could, therefore, potentially be a result of changes in one or more of these variables.

The $a - vO_2$ difference is an indicator of the effectiveness of active muscles in extracting oxygen from arterial blood. At rest, approximately 20%-30% of the blood's original oxygen content is extracted at the tissue level. In a healthy subject, CaO_2 remains constant at rest and during exercise, while the CVO_2 is reduced during exercise. Thus, the $a - vO_2$ difference increases progressively

with exercise intensity. During maximal exercise, up to 50-60 % of the available O_2 in the blood is extracted by the skeletal muscles [111]. Therefore, it appears that the delivery of O_2 through increased blood flow is the most influential factor on the VO_2max [28].

3.2.1 Research question

The VO_2max could be almost twice as high in trained athletes compared to sedentary individuals because of their enhanced SV , improved myocardial function, and higher capacity for oxidative metabolism in active muscles.

However, the role of the cardiorespiratory coupling within this framework is not well understood. It is well-known that cardiovascular and pulmonary functions are coupled at rest, but few attempts have been made to study the strength of this coupling during exercise. In this context, we conducted a preliminary analysis to investigate whether the strength of the cardiorespiratory coupling could be related to a quantitative measure, such as VO_2 or the respiratory exchange ratio (RER).

The results of this study are discussed in Chapter 9.

3.3 Cardiovascular adaptations to exercise

Endurance training elicits many adaptations in skeletal muscles and in the metabolic and cardiorespiratory systems. Together, these adaptations improve the ability to deliver and utilize oxygen for energy production, the capacity for prolonged exercise, sports performance, and the performance of activities of daily living without undue fatigue.

One of the hallmarks of endurance training is a decrease of the resting HR (sinus bradycardia) [24]. Since, systemic oxygen demand and CO at rest are largely unaffected by exercise training, it is highly likely that the bradycardia is a secondary effect of the training induced increase in SV . Additionally, an increase in parasympathetic activity and, to a lesser extent, a reduction in sympathetic activity also decrease the HR at rest [158].

The HR during exercise at any absolute level of submaximal oxygen uptake also decreases, but the extent of this decrease is largely defined by genetics [151]. Similar to resting conditions, CO at a given absolute submaximal workload is not significantly different after exercise training, since the oxygen demand is unaltered. The unchanged CO is the result of a larger SV and lower HR .

Endurance training also leads to an increase in LV end-diastolic diameter, LV wall thickness and LV mass (LVM), such that the heart of some highly trained endurance athletes is more than double the size of non-athletes [93]. The result of these cardiac adaptations is called "athlete's heart".

3.3.1 Research question

Increased cardiac mass and changes in cardiac geometry are well-established risk factors for cardiovascular events in non-athletic populations. Currently, there is an ongoing debate whether people with increased cardiac size and hypertrophy due to intense exercise training are also at elevated risk of cardiac events. Regardless, due to the similarities in cardiac adaptations, it is very difficult to distinguish normal training adaptations from pathological changes [53]. For instance, the mainstay of the diagnosis of concentric hypertrophic cardiomyopathy (HCM) is an increased LV wall thickness.

The stakes of resolving such diagnostic ambiguity are high since, on the one hand, false reassurance may lead to an increased risk of sudden cardiac death. On the other hand, a diagnosis of HCM may imply a life-long ban from competitive sports, regular medical follow-up, screening of family members and psychological stress [53]. Hence, accurate phenotyping, including quantification of cardiac dimensions, is crucial to help distinguish normal training adaptations from pathological changes.

The gold standard for the assessment of cardiac volumes and mass is CMR. Structural measures by CMR are both accurate and reproducible [92]. However, cost and operational considerations tend to limit its utility in large-scale population studies and clinical trials [152].

We wanted to investigate whether we could use regularly measured features, such as ECG derived features, demographic and body composition features to predict LVM. The results of this study are depicted in Chapter 9.

Part II

Signal quality

Chapter 4

ECG quality

In this Chapter, we describe the development of a novel algorithm for the quantification of ECG signals. We can divide this process into two steps: the creation of an algorithm for the detection of contaminated segments and the extension of this algorithm towards a novel signal quality index.

We evaluated the performance of each of these algorithms extensively on different datasets and the results of these experiments are published in **MOEYERSONS, J.**, SMETS, E., MORALES, J., VILLA, A., DE RAEDT, W., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., VAN HUFFEL, S. AND VARON, C. Artefact detection and quality assessment of ambulatory ECG signals. *Computer Methods and Programs in Biomedicine* 182, 105050, (2019), <https://doi.org/10.1016/j.cmpb.2019.105050>.

4.1 Introduction

Ambulatory devices allow patients to be monitored during daily life, instead of in a protected hospital environment. This results in a drastic increase of the detection window and, thereby, the likelihood of identifying dysfunctions. However, taking the recording procedure out of the hospital increases the exposure to noise. Noise can originate from a variety of sources, such as the power line, muscle activity or electrode movement. All of these sources affect the recording in a different way. These alterations are categorized as artefacts. Since artefacts could profoundly reduce the diagnostic quality of the recording, it is important to deal with them accordingly.

Silva et al. stated that: "Rigorous quality control is essential for accurate diagnosis, since alterations during the actual recording might result in inappropriate treatment decisions" [171]. To mitigate this problem, automated algorithms are needed to detect artefacts and to quantify the quality of the recorded signal [78]. Two closely related approaches can be distinguished to assess the acceptability of the recorded ECG: detection and quantification. The first approach consists of a simple binary, clean or contaminated, classification. One example of where this might be used is at the front end of the acquisition process. The user could be provided with rapid, binary feedback and, if required, make adjustments in the recording set-up or, worst case scenario, re-start the recording.

The rationale for the second approach, quality quantification, is that different study objectives require different quality levels. For instance, HRV studies do not require the same, high, quality level as beat classification studies. This prompted a number of authors to propose a multi-level quantification of the signal quality [154], [183], [102]. For example, Li et al. proposed a five level signal quality classification algorithm which divided the signals into five bins: clean, minor noise, moderate noise, severe noise and extreme noise [102].

In this study, we propose a novel method for the detection of artefacts and quantification of the signal quality. The autocorrelation function (ACF) is used to characterize the ECG signal, since it facilitates the separation of clean and contaminated segments [194]. From the ACF, descriptive features are extracted and fed to a RUSBoost classifier. We propose to use the posterior probability of the clean class as an indication of the quality of the signal.

The first objective of this study is to evaluate the robustness of the proposed artefact detection method by training it on one dataset and testing it on two independent datasets. The second objective is to evaluate the novel signal quality index by comparing it to the agreement of the annotators and the level of noise.

4.2 Data

We used ECG recordings from three different sources, acquired with three different recording devices: (1) a polysomnographic dataset, recorded in the University hospitals Leuven (UZ Leuven), Belgium, (2) a hand held dataset, available in Physionet and recorded for the Computing in Cardiology (CinC) Challenge of 2017 and (3) a stress dataset, recorded with a wearable developed by imec. In the following paragraphs, we outlined the datasets.

4.2.1 Polysomnographic dataset (PSG)

This dataset was recorded in the sleep laboratory of the University hospitals Leuven, UZ Leuven, Belgium. The study protocol was approved by the Ethical Committee of the University Hospitals of Leuven (S53746). The dataset consists of 16 single-lead (lead II) ECG recordings from 16 patients, accounting for a total of 152 hours and 12 minutes of data. A sampling frequency of 200 Hz was used.

In sleep apnea research, it is common practice to first divide the recordings into one minute segments and subsequently analyse each recording on a minute-by-minute basis [194]. Implementation of this procedure resulted in a total of 9132 one-minute segments.

A medical doctor evaluated each minute for the presence of artefacts and assigned a binary label (+1 = clean, -1 = noisy). This procedure resulted in a total of 9132 one-minute signals of which 295, or 3.2%, are noisy. These labels have been previously used for artefact detection in [194].

4.2.2 Hand held dataset (HH)

The PhysioNet/Computing in Cardiology Challenge of 2017 was intended for the differentiation of atrial fibrillation (AF) from noise, normal and other rhythms in short term ECG recordings [42]. All signals were recorded by volunteers with an AliveCor's single-channel ECG device. They were instructed to hold one electrode in each hand, thereby creating a lead I (LA-RA) equivalent ECG. The signals lasted between 9 and 60 seconds, were sampled at 300 Hz, and stored as 16-bit files with a bandwidth of 0.5–40 Hz.

In total, the dataset consists of 12186 ECG recordings, of which 8528 recordings were provided for training. Only the normal rhythm and noisy class were

selected for this paper. They accounted for a total of 5334 recordings. We refer the reader to [42] for a more detailed description of the dataset.

4.2.3 Stress dataset (STR)

This dataset was recorded to detect and quantify stress in daily life using different physiological metrics, including ECG. The single-lead (lead II) ECG signals were recorded with the Health Patch (imec, Leuven, Belgium), which has a sampling frequency of 256 Hz. The device was previously used in another stress study to measure ECG and acceleration [77].

The subjects were instructed to wear this patch for five consecutive days and to remove it only while working out. They were also provided with a waterproof cover to protect the patch while showering. One day of one subject was selected to test the proposed algorithm.

We divided the ECG signal into segments of 30 seconds, which resulted in a total of 2879 segments.

4.2.4 (Re)labelling of the data

During the CinC challenge, participants noted that some recordings of the non-noisy classes were actually very noisy [42]. Therefore, the organizers decided to re-check all recordings and provide new labels if necessary. However, despite these adjustments, the organizers of the CinC challenge stated that: "*The kappa value, κ , between many data remained low even after relabelling, indicating that the training data could be improved.*" [42]. Additionally, no labelling of the STR dataset was present prior to this study. Therefore, we opted to have our own (re)labelling procedure for the CinC and Stress datasets.

Each segment was (re)labelled by four independent annotators according to the following rule:

"If the annotator is able to confidently distinguish all R-peaks in the segment, then the segment is labelled as clean. Otherwise, it is labelled as contaminated."

The annotators were all engineers working in the signal processing domain, but with varying experience in ECG analysis. The most experienced annotator has seven years of experience and the least experienced has approximately one year of experience. In order to avoid any bias, the annotators had no knowledge of the original labels.

To quantify the agreement between the different annotators, we computed two different metrics: (1) the percentage of agreement and (2) the Fleiss' kappa. The segments were classified in three classes: 'Perfect agreement', when all four annotators agreed, 'Moderate agreement', when three annotators agreed, or 'Disagreement', when no majority voting was possible.

4.3 Methods

4.3.1 Pre-processing

We filtered all signals by means of a zero phase, 2nd order high- and 4th order low-pass Butterworth filter with cut-off frequencies at 1 Hz and 40 Hz, respectively. These filters ensure the removal of baseline wander and high frequency noise, without altering the information contained in the characteristic waveforms of the ECG.

4.3.2 Feature extraction

In theory, it would be possible to derive features from the whole segment (e.g. one minute) and use these to characterize the segment. In practice however, this might cause smaller, more subtle artefacts to go undetected. For a similar problem, Varon et al. showed that a window reduction causes an improvement of the algorithm's resolution. This is due to the fact that less information from clean parts of the signal could be included in the contaminated segments. In that study, they obtained the best results with a window width of five seconds [194]. Moreover, visual analysis of the noisy segments showed that often only small parts of the segments are contaminated. Therefore, we decided to use a window width of five seconds and a sliding window approach with 80% overlap.

As a first step, we moved the sliding window over the ECG signals and, as a second step, we computed the ACF from each sliding window. This was obtained by computing the inverse Fourier transform of its corresponding power spectral density (PSD) [25]. We used a maximum time lag of 250 ms to exclude consecutive heartbeats and to ensure the inclusion of all characteristic waveforms.

The most characteristic waveform of an ECG signal is the QRS-complex. It consists of three consecutive deflections, namely: a downward deflection, the Q-wave; an upward deflection, the R-peak; and another downward deflection, the S-wave. Since we use a maximum time lag of 250 ms, we can assume that the

shape of the ACF is primarily defined by a shift of this complex. An example of the ACF of two contact and non-contact signals with a different quality level can be seen in Figure 4.1. From the ACF we can derive three intuitive features:

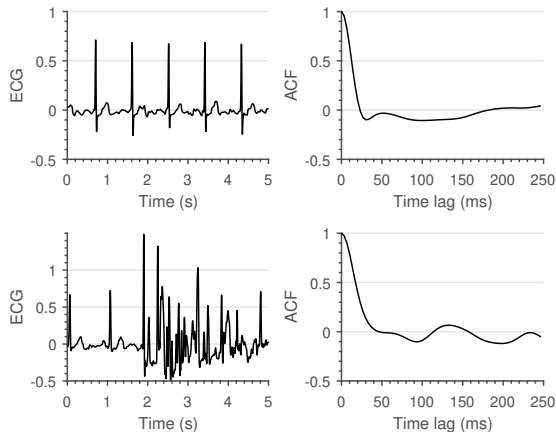


Figure 4.1: Comparison between a clean (top) and a contaminated (bottom) segment. A clear difference between the shape of both ACF's can be observed.

1. First (local) minimum (FMin):

Since the QRS-complex consists of three opposite deflections, we can assume that by shifting the QRS-complex, we reach a time lag where the opposite deflections overlap. Due to the high amplitude of the QRS-complex, this shift should coincide with the first local minimum in the ACF.

A large shift, could be an indication of a flat line, while a small shift could be an indication of a sharp, high amplitude, artefact. In Figure 4.2, a large, wide artefact is depicted. This results in a large FMin, which indicates that the lag is a lot larger than is normally expected.

To derive a feature for the entire segment, we first selected the location of every first local minimum for all sliding windows. Afterwards, the overall minimum of the whole segment was computed. This results in a single value per segment.

2. Maximum amplitude at 35ms (MAmp):

For the previous feature, we looked at the shift that results in a local minimum and we assumed that this should align with a time lag where the opposite deflections overlap. However, this is not always the case. Hence, for the second feature, we estimated that the first local minimum

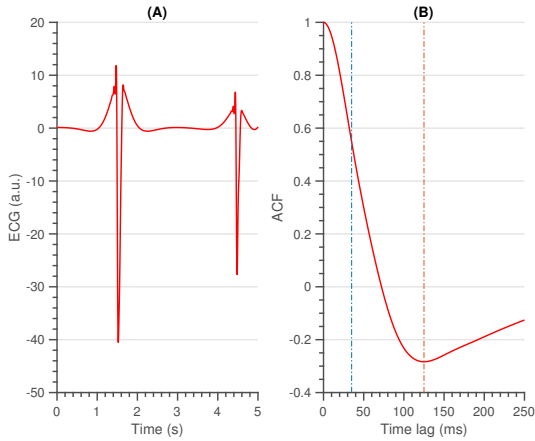


Figure 4.2: An example of an ECG signal that contains a large artefact. The dotted red line in plot (B) indicates the first local minimum and the dotted blue line indicates the maximum amplitude at 35 ms.

of an average clean segment is around a time lag of 35 ms and take the amplitude at this time lag.

High amplitudes, as well as large time lags of the previous feature, could indicate a constant value. It could also indicate a technical artefact with a high amplitude and large width (Figure 4.2).

To have a general feature for the entire segment, we select the amplitude at 35 ms in the ACF of all sliding windows and we take the maximum over all obtained values.

3. **Similarity (Sim):** An ECG signal is a very repetitive, quasi-stationary, signal. While the heart rate may differ along the signal, the morphology of the heartbeat remains comparable. Therefore, it is safe to assume that a measure of similarity would be a good indication of the quality of an ECG signal.

From the ACF's of all sliding windows, we selected the interval between time lags 30 and 115 ms. The boundaries of these intervals were empirically defined. The amplitudes before 30 ms could be of interest, but are already taken into account in the first two features. Additionally, we observed no added value in terms of classification performance by extending the interval beyond 115 ms. Hereafter, we computed the pairwise euclidean distance between all the ACF's in this interval and used the maximum of

these distances as a measure of similarity. We can observe the difference between a clean and a noisy non-contact ECG signal in Figure 4.3.

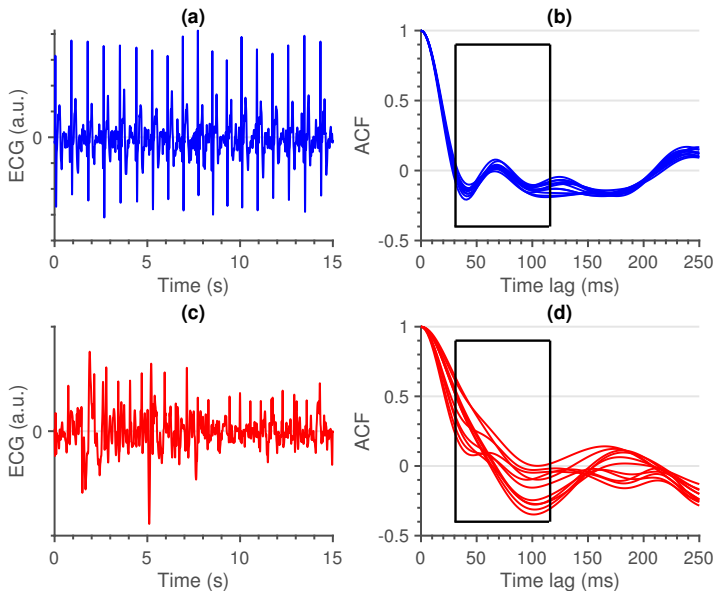


Figure 4.3: Comparison between a clean (blue) and noisy (red) ECG signal (a-c), together with the ACF's of their respective sliding windows (b-d). The difference between the ACF's within the search window, as depicted by the black box, is clearly higher for the noisy ECG signal.

A large value of this feature indicates the presence of a divergent ACF. It could also be an indication of variable ECG morphologies within that segment. In essence, the larger the value of this feature, the less similar the ACF's in that interval are and the more likely that the quasi-stationarity, that is expected in an ECG signal, is destroyed by the presence of an artefact.

4.3.3 Classification

Instead of using a random training set selection, we used fixed size sampling, which is based on Rényi's quadratic entropy (H_r) [49], to construct the training and test set. Rényi's quadratic entropy is defined as

$$H_r(x) = -\log \int p(x)^2 dx \quad (4.1)$$

with p as the probability density function. This can be estimated as

$$\int \hat{p}(x)^2 dx = \frac{1}{N^2} \mathbf{1}_N^T \Omega \mathbf{1}_N, \quad (4.2)$$

with N as the number of samples, Ω as the kernel matrix and vector $\mathbf{1}_N$ as a $N \times 1$ dimensional vector of ones [69].

This sampling strategy creates a subset with an entropy that approximates the entropy of the entire dataset set. It consists of the following steps:

First, a subset with a pre-defined size is randomly selected. Second, a random sample of the subset is swapped with a new, randomly selected, sample of the training set. If Rényi's quadratic entropy of the subset increases, then the new sample replaces the old sample in the subset. Otherwise, the exchange is rejected. These steps are repeated until there is no more increase in entropy or after a predefined number of iterations, which is set to 1000. This results in an entropy value of the training set that approximates the entropy of the entire dataset.

The first objective of this study is to test the robustness of the proposed algorithm. Hence, one dataset was used to train the classifier and the other two were used for independent testing. The PSG dataset was used for the initial classifier training, since a gold standard was provided by a medical doctor.

The percentage of contaminated segments in the PSG dataset is substantially lower compared to the clean percentage, 3.2% vs. 96.8%. Therefore, as indicated in Chapter 2, the implementation of a classification algorithm that is able to effectively identify the rarely occurring contaminated segments is required. In this study we used RUSBoost. This is a combination of random under sampling and boosting.

Decision trees were selected as weak learners, since they are well-suited for ensemble classifiers, such as the one we use [67]. Each decision tree was trained with the CART algorithm and deep trees were used with a minimal leaf size of five. The learning rate of the ensemble was set at 0.1, which requires more learning iterations, but often achieves better accuracy.

In addition to the selection of the right weak learner, one of the most important hyper-parameters is the number of weak learners of the ensemble. We used a standard 10-fold cross-validation approach to select the number of weak learners.

The maximum number of weak learners was fixed at 500 and the mean squared classification error of the folds was used as decisive metric.

To evaluate the robustness of the proposed approach, we compared the classification performance of two models: the first model is trained on the PSG dataset and the second model is trained on the entire dataset. Additionally, we compared the performance of the proposed algorithm, with two other algorithms: a heuristic [143] and a machine learning algorithm [41].

4.3.4 Quality indication

When removing contaminated data segments, an important balance should be kept. On the one hand, it is desired to remove as many contaminated segments as possible. On the other hand, it is desired to retain as much valuable information as possible. Nevertheless, keeping this balance is a very challenging task when using a simple binary classification (i.e. clean vs. noisy segments).

RUSBoost is a probabilistic classifier, which means that it defines the class of each input sample based on the probability of it belonging to each class. The sample is assigned to the most probable class. This probability is defined by the relative weight across the set of decision trees.

For each sample, each class and each decision tree, this relative weight is obtained by dividing the number of training samples of each class by the total number of training samples in the selected leaf. These weights are averaged over all decision trees in the ensemble to obtain the overall class weight. When we normalize these weights, we obtain a probability for each class ranging between 0 and 1.

Mounce et al. used the class probabilities to produce a more fine-grained and focussed assessment of the risk of iron failure in drinking water [137]. In this study, we propose a similar approach, but applied to the level of quality of ECG signals. The probability of the clean class was transformed to a quality score ranging from 0% to 100%, or in other words, from too contaminated to process to perfectly clean.

The obtained score is taken as the quality level of the signal. Since this score reflects the certainty of the algorithm, it is expected that it closely relates to the certainty/agreement of the annotators.

To evaluate the proposed algorithm, we created a dataset with known Signal-to-noise ratio (SNR) levels. We corrupted the clean signals of the PSG dataset with recorded artefacts from the MIT-BIH Noise Stress Test Database (NSTDB) [134]. This database contains samples of three types of noise: electrode motion

(EM), baseline wander (BW) and muscle artefact (MA). Only EM and MA were considered, since baseline wander is usually not a cause for erroneous R-peak detection. We generated contaminated signals with five contamination levels, similarly to [102].

To enlarge the amount of segments, we segmented the clean signals and artefact samples in 10 seconds epochs. For each clean ECG epoch, an artefact epoch was randomly selected and a calibrated amount was added to the clean epoch. The SNR of the resulting signals was defined as described in [41]. An example of the resulting signals is shown in Figure 4.4

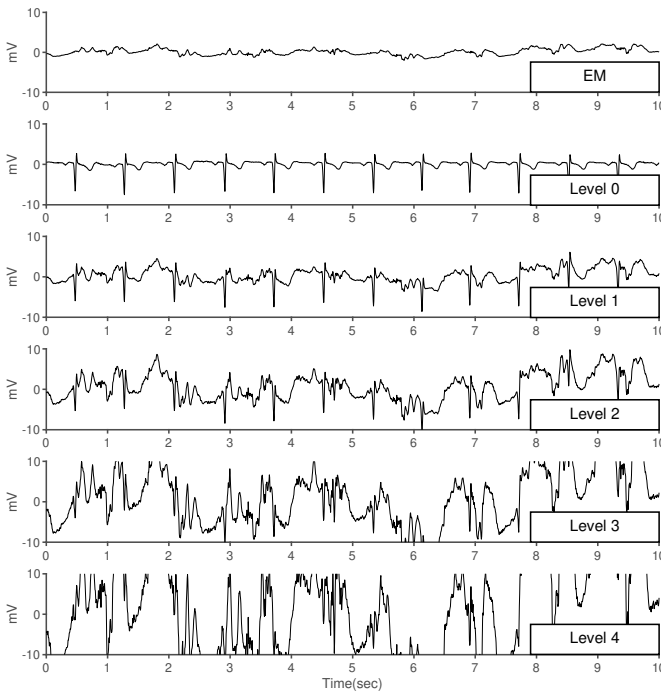


Figure 4.4: Impact of Electrode Motion on ECG signal quality.

EM: Electrode Motion noise, Level 0: Clean ECG signal, Level 1: Minor contamination, Level 2: Moderate contamination, Level 3: Severe contamination, Level 4: Extreme contamination. The quality of the ECG signal decreases with the increase of electrode motion (EM) noise increases.

4.4 Results

In this section we present the results of the experiments. It is divided in three subsections, the first subsection describes the inter-rater agreement metrics for the (re)labelling of the HH and STR dataset. The next subsection describes the performance of the model and the last subsection evaluates the proposed quality assessment index. All analysis were performed using `MATLABR2017a` (The Mathworks, Natick, MA, USA).

4.4.1 Inter-rater agreement

Table 4.1 displays the results of the (re)labelling of the HH and STR dataset, including the percentage of 'Perfect agreement', 'Moderate agreement', 'Disagreement', and the Fleiss' kappa. Despite the great difference in experience, the annotators were in agreement on 72.3% of the HH and 91.9% of the STR dataset. Additionally, the kappa values ranged from substantial (>0.61) for the HH dataset, to almost perfect (>0.81) for the STR dataset. This indicates that the high level of agreement was not due to chance.

Table 4.1: Overview of the HH and STR dataset (re)labelling

	HH			STR		
	Clean	Cont.	All	Clean	Cont.	All
Perfect	2404	1453	3857 (72%)	1817	829	2646 (92%)
Moderate	518	523	1041 (20%)	118	66	184 (6%)
Disagree			436 (8%)			49 (2%)
Fleiss' kappa			0.686			0.901

4.4.2 Model performance

To evaluate the robustness of the proposed approach, we compared the classification performance of two models. The first model is trained on the PSG dataset and the second model is trained on a combined subset of the three datasets. For the Stress and CinC datasets, only the segments in the 'Perfect agreement' subset were considered.

The result of the 10-fold cross-validation for the determination of the amount of weak learners for the PSG dataset is shown in Figure 4.5.

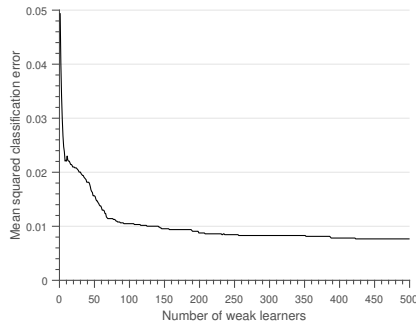


Figure 4.5: The mean squared classification error of the 10-fold cross-validation of the PSG dataset versus the number of weak learners. We can observe a stagnation of the decrease in classification error around 140 weak learners.

The model that was trained on the PSG dataset achieved a mean squared classification error of under 2% when 40 or more weak learners were used. Additionally, the classification error stagnated at 140 weak learners. Therefore, we retained 140 weak learners and removed the excessive weak learners. This is likely to have minimal effect on the predictive performance of the model, since the selected weak learners give near-optimal accuracy.

We used the same methodology for the second model, which resulted in a model consisting of 90 weak learners, corresponding to an average classification error of 4.3%.

From Table 4.2 we can observe that similar performances were obtained with the first model compared to the second model. All performance metrics were above 0.89. Regarding NPV, better results were obtained for the HH and STR sets for the second model. This might be due to a different spread of the features in the three datasets. In Figure 4.6, we can observe that the features of the CinC dataset have a larger spread, compared to the other two datasets. Overall, both models outperformed the method of [143], which had a particularly low NPV on the PSG dataset, and achieved similar performance scores compared to the method of [41].

4.4.3 Quality assessment index

Since the performance of the second model was slightly better, compared to the first model, we used that model for the evaluation of the quality assessment index.

Table 4.2: Classification performance on independent test sets. Model 1: Trained on the PSG dataset, Model 2: Trained on the three datasets, Orphanidou et al. [143], Clifford et al. [41]

	Test PSG				Test HH				Test STR			
	Model 1	Model 2	Orphanidou	Clifford	Model 1	Model 2	Orphanidou	Clifford	Model 1	Model 2	Orphanidou	Clifford
Se	1.000	1.000	0.892	0.999	0.932	0.977	0.945	0.981	0.993	1.000	0.987	0.993
Sp	0.966	0.910	0.876	0.828	0.966	0.947	0.88	0.892	1.000	0.996	1.000	0.996
NPV	1.000	1.000	0.211	0.965	0.896	0.960	0.906	0.998	0.984	1.000	0.972	0.997
AUC	0.999	1.000	/	0.827	0.988	0.993	/	0.875	1.000	1.000	/	0.989

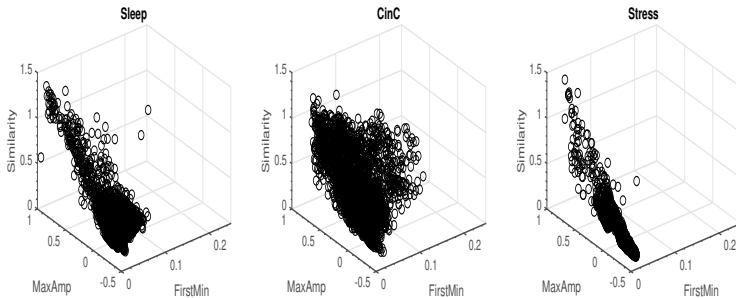


Figure 4.6: Feature space of the three training datasets. A larger spread can be observed in the CinC dataset compared to the other datasets.

When adding the 'Moderate agreement' subsets, we observed a decrease in the weighted performance metrics with $wSe=0.794$, $wSp=0.787$ and $wAUC=0.928$ for the HH dataset and $wSe=0.966$, $wSp=0.848$ and $wAUC=0.970$ for the STR dataset.

These results indicate that the segments in the 'Moderate agreement' subset are more difficult to classify. We hypothesized that this difficulty is not only reflected in the performance of the classifier, but also in the probability of the predicted class.

The boxplots in Figure 4.7 show an increase of the score of the clean class with respect to the amount of agreeing annotators. A Pearson correlation coefficient of 0.8131 was obtained. This confirms the assumption that the score of the clean class is closely related to the certainty/agreement of the annotators.

The last experiment consisted of testing the proposed quality indication on a dataset with known SNR levels. We added two types of realistic ECG noise, EM and MA, at different SNR levels from the NSTDB to create a simulated dataset. The quality values of the EM and MA noise decreased significantly with the increasing noise levels (Figure 4.8). Additionally, the inter-noise quality values

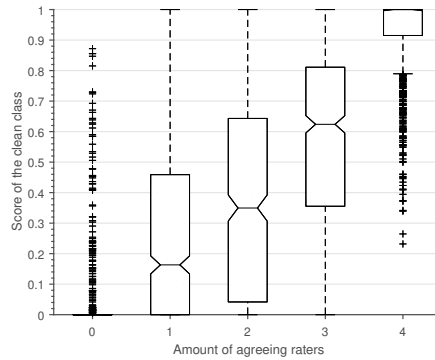


Figure 4.7: Boxplot of the score of the clean class plotted against the amount of agreeing annotators. The box plots represent the median value together with the interquartile range. Outliers are shown as plus signs.

also differed significantly ($p < 0.01$).

4.5 Discussion

In this study we presented an automated method for artefact detection and continuous signal quality assessment, based on the ACF and a RUSBoost classifier. The main novelty is the new approach to signal quality assessment, using the posterior probability of the clean class of the RUSBoost classifier as an estimation of the quality of the ECG segments.

The first objective of this study was to evaluate the robustness of the proposed algorithm by training it on one dataset and testing it on two independent datasets. Due to the low level of agreement between the annotators of the CinC dataset and the lack of labelling in the Stress dataset we performed our own (re)labelling procedure. Relabelling (parts) of a freely available dataset was already done by other authors [41, 47, 177].

The newly labelled datasets were used to test the robustness of the proposed model. Firstly, the model was trained on the Sleep dataset and subsequently tested on the 'Perfect agreement' subset of the other two datasets. Very high performance metrics on the test sets were obtained, indicating that the model can be applied on data obtained in different recording conditions without substantial performance loss. Furthermore, this highlights the robustness of the

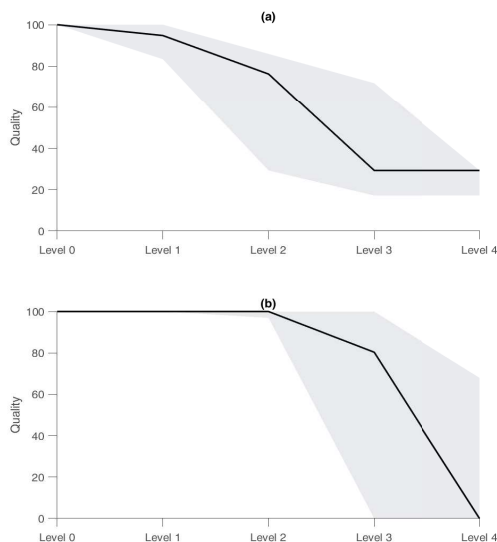


Figure 4.8: The quality of the EM (a) and MA (b) noise both significantly decrease, to a different extent, with the increasing noise level. The boundaries of the gray area indicate the 25th and 75th percentiles, and the solid line the median.

model, since the data was recorded with different sampling frequencies, acquired with different systems and segmented with different segment lengths.

Secondly, the model was trained on a training set derived from all datasets and tested on a test subset of the 'Perfect agreement' subset. In this set-up all recording conditions were taken into account when training the model. Although only slightly, the overall performance of the model increased. This is particularly observable for the NPV in the CinC dataset. This might be due to a different spread of the feature space of the three datasets, as shown in Figure 4.6.

The difference between the datasets could be explained by a couple of factors:

1. **The amount of subjects** The HH dataset consists of ECG signals of many different subjects, while the PSG and STR dataset consist of 16 and 1 subject respectively.
2. **Clean vs. Contaminated** The HH dataset contains more contaminated segments, compared to the other two datasets, as shown in Table 4.1.

This might be due to the way of recording. The HH dataset recordings were made by a hand held device, while the other two were made with electrodes attached to the body.

The performance of the proposed model was compared with a heuristic [143] and a machine learning algorithm [41]. We have shown that the proposed model outperforms the model proposed in [143]. This algorithm is intrinsically linked to the performance of a peak detector, which hinders the performance of the method. Since our method is solely based on features derived from the ACF of small ECG segments, it is not subjected to this type of problems.

A similar performance is obtained compared to the machine learning algorithm proposed in [41]. This method also uses peak detection, but, instead of using the obtained RR-intervals, it compares the output of two peak detection algorithms. This is a more robust methodology, compared to the method proposed in [143].

The advantage of the method proposed in this study is that only three features are needed to obtain an accurate performance. However, a drawback of this study is that the model was trained and tested on a dataset which mostly consisted of normal sinus rhythm. Previous studies have shown that signal quality classification algorithms experience a reduction in performance on an arrhythmia database when it was not explicitly retrained using signals containing arrhythmia episodes [102]. We expect the model to be only marginally influenced by abnormal rhythms, but more affected by abnormalities which change the shape of the heartbeat significantly, e.g. pre-mature ventricular contractions. This hypothesis needs to be further investigated in future work.

The obtained Pearson correlation coefficient of 0.8131 confirms the hypothesis that the certainty of the algorithm, quantified by the score of the clean class, is closely related to the certainty/agreement of the annotators. This motivated the proposal to use the score of the clean class as a new quality assessment index. One of the big advantages of this method is that no additional parameters have to be tuned, after the model is trained. This is in contrast with [101], where an extra coefficient needs to be chosen when spectral or statistical noise is present.

As a last experiment, we tested the proposed quality indication on a dataset with known SNR levels. As a first observation, we noticed an overall quality level of 100% (25th percentile=100% and 75th percentile=100%) for the clean signals. This indicates that the selected clean signals are indeed clean, according to the proposed algorithm. Secondly, we observed significant quality decreases per noise level for both types of added noise. However, when the clean signals are contaminated with EM noise, the quality values do not converge to 0% for the most heavily contaminated segments. This might be due to the similar

morphology of the EM noise and the QRS complex. A comparable issue was reported by Li et al. when testing their signal quality indices [101].

Despite these promising findings, we also found significant inter-noise differences. This might indicate that the algorithm reacts differently to different types of noise. An explanation might be that, although the SNR values are chosen according to the clinical usage of the signals, the shape of the different noise types remain different. On more elaborate algorithms, such as the one of [102], where more features are used the effect is limited, but in the proposed method, only three features derived from the ACF are used. All three features are (QRS-)shape dependent and are thus heavily influenced by the shape of the noise. It should be noted that this might not be the optimal set of features. On the contrary, it is very probable that the quality assessment can be improved by including statistical, frequency or other metrics.

One of the limitations of this study is that the window size for the initial feature extraction was not tuned to produce optimal results. Additional investigation into the optimal window size might result in better performance of the model. Since, for research purposes, it is desired to remove as few data points as possible, the model would benefit from a smaller window size.

4.6 Conclusion

We proposed an automated, accurate and robust artefact detection algorithm, based on the ACF and the RUSBoost algorithm. We have shown that, when training the classification model on one specific dataset, it is robust enough to accurately predict the labels in other datasets. Although, as expected, the prediction is more accurate when we model the other conditions as well.

The main novelty of this study is the new approach to ECG signal quality assessment. We suggest to exploit the posterior class probability of the RUSBoost classifier and to use the probability for the clean class as a novel quality assessment index. This allows users to identify periods of data with a pre-defined level of quality, depending on the task at hand.

Chapter 5

Transfer Learning for Modality-Specific Artefact Detection

In this Chapter, we optimized the artefact detection model proposed in the previous Chapter towards another signal modality, using transfer learning. To the best of our knowledge, this is the first study that showed that an artefact detection model for one signal modality, namely contact ECG, can be optimized with transfer learning towards another signal modality, namely capacitively coupled ECG (ccECG).

We evaluated the performance of each of this optimized algorithm extensively on different datasets and the results of these experiments are published in **MOEY-ERSONS, J.**, MORALES, J., VILLA, A., CASTRO, I., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., VAN HUFFEL, S. AND VARON, C. Supervised SVM Transfer Learning for Modality-Specific Artefact Detection in ECG. *Sensors* 21, 2, (2021), 662. <https://doi.org/10.3390/s21020662>.

5.1 Introduction

During a cardiac examination in the hospital, the patient is asked to lay still in a supine position and thus, as a result, the recorded signals are generally of very high quality. This is no longer the case for ambulatory recordings. The diagnostic capabilities of the signals can be reduced by the presence of artefacts in the signal. These contaminants can be caused by electrode motion, contact noise and electromyography among others [10].

One of the most commonly used wearable devices, is a Holter. It continuously records the electrical activity of the heart for a period of, typically, 24 or 48 hours. Although typical electrodes are used, the device can be worn around the hip. This allows the patient to participate in daily life activities, in contrast to the "snap-shot" ECG in the hospital. Although the extension of the analysis period highly increases the detection rate of specific cardiac events, it is still not enough to detect a large amount of cardiac conditions [46, 105]. It is stated in [87] that recordings acquired during up to 3 months are beneficial for this task.

ccECG sensing is an unobtrusive monitoring technique which can be used to extend the analysis period further. It enables ECG signals to be recorded through insulating materials, such as clothing, without direct skin contact. This allows the capacitive electrodes to be integrated into daily life objects, such as mattresses or (car) seats [100, 197]. On the upside, this provides more comfort for the patient. However, on the downside, this introduces new, modality specific, artefacts originating from electromagnetic changes in the surroundings, movement of the patient and poor electrode coupling. These artefacts can be different from those in contact ECG [31]. Therefore, it might be that an artefact detection model, trained on contact ECG, cannot accurately distinguish clean from noisy ccECG signals without any adaptation.

One way to solve this, is to train and test a classifier on labelled ccECG data, similar to the previous Chapter. Multiple artefact detection algorithms have been developed in recent years. Examples of the used methodologies are: QRS amplitude ratios in combination with PCA [198], wavevelet decompositions [169] and frequency features in combination with ICA [199]. Moreover, in [163], the authors derived basic features from pressure and ccECG signals to accurately separate clean from noisy segments.

Castro et al. obtained an accurate separation by using only signal quality features for contact ECG, such as kurtosis and spectral density ratio's [31]. A downside of this approach is the requirement of a lot of ground truth data. This is usually obtained by manually labelling the signals, which is a tedious and time consuming task. A better and faster solution would be to train a classifier on contact ECG data and adapt it for ccECG data. Since many labelled contact

ECG datasets exist, this presents a more feasible solution.

We propose to use transfer learning to extend a model from contact ECG to other sensor modalities, like ccECG, without the need of relabelling an entire new dataset. This technique exploits the knowledge obtained from a previous task to optimize the performance on a similar target task, for which fewer labelled data points are available [147]. It thereby circumvents the need to retrain a new classifier from scratch and requires only a limited amount of labelled data. Applied to our case, it means that an artefact detection model for one signal modality, namely contact ECG, can be adapted with transfer learning towards another signal modality, ccECG in this case. Of course, this would only be needed if an artefact detection model, trained on contact ECG, performs poorly on a ccECG dataset.

The first goal of this study was to investigate the necessity of transfer learning. We used features that were previously used for artefact detection in contact ECG [128], to construct an artefact detection model. Then, we investigated its ability to make an accurate distinction between clean and noisy ccECG samples. The second goal was to optimize the performance of this artefact detection model towards ccECG. To this end, we made use of the transfer learning formulation described in [51], to create a modality-specific classifier by using only a limited amount of modality-specific data. Lastly, we investigated whether an active sampling strategy based on entropy could decrease the amount of samples that are needed to optimize the model.

To the best of our knowledge, this is the first study that showed that an artefact detection model for one signal modality, namely contact ECG, can be optimized with transfer learning towards another signal modality, namely ccECG in this case.

5.2 Data

We used ECG recordings from three different sources, acquired with three different recording devices: (1) a polysomnographic dataset, recorded in the University hospitals Leuven (UZ Leuven), Belgium, (2) a hand held dataset, available in Physionet and recorded for the Computing in Cardiology (CinC) Challenge of 2017 and (3) a capacitively coupled dataset, which originated from multiple sources. The first two datasets were already outlined in Chapter 4. They consist of contact ECG, albeit recorded with a different device. The third dataset consists of ccECG. This dataset is not described yet. Therefore, it is outlined below. A complete overview of the datasets is given in Table 5.1.

5.2.1 Capacitively coupled dataset (CC)

This dataset consists of 10000 ccECG signals originating from two recording devices. The first recording device is described in [34, 35] and was used to record 15 subjects [33, 32]. An example of the mattress and chair implementation can be observed in Figure 5.1. The second one was used to create the publicly available UnoVis dataset [197] it contains recordings from 26 subjects. The dataset comprises ccECG signals from different scenarios, such as signals recorded while sitting in a static car seat, while driving a car, while sleeping and while working on an office chair. Each signal was 15 seconds long with a sampling frequency of 512 Hz.

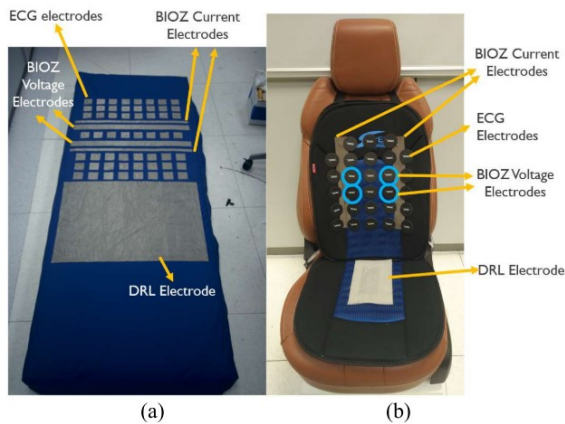


Figure 5.1: Example of the mattress (a) and chair (b) implementation of the multi-channel ccECG acquisition system. Figure taken from [34] with permission.

All signals were labeled by five annotators with experience in ECG signal analysis. Three classes were used:

1. Useless or no ccECG
2. ccECG with artefacts that may affect the detection of two to five heartbeats
3. ccECG useful for HRV analysis and possibly morphology analysis

In total, 90 signals with strong annotation disagreement, labelled as both 1 and 3 by different annotators, were discarded. This resulted in a dataset of 9910 signals on which at least three annotators agreed. A Fleiss' Kappa of 0.80 was obtained.

The three quality levels were assigned to two binary classification problems: one with a 'low threshold' (level 1 vs level 2-3) and another with a 'high threshold' (level 1-2 vs level 3). The high threshold labelling matches the best with the labelling procedure in the previous Chapter. Therefore, we conducted all the analysis with these labels. The selected labelling procedure resulted in 80.5% noisy signals and 19.5% clean signals.

Table 5.1: Overview of all datasets.

Dataset	Data source	Sampling frequency (Hz)	Scenario	# segments
PSG	University hospitals Leuven	200	During sleep	9132
HH	CinC Challenge of 2017 [42]	300	Unknown	3857
CC	Recordings of system presented in [31]	512	Static car seat Bed Office chair Driving a car	2500 2500 1520 480
	UnoVis database [197]	512	Bed Driving a car Armchair	1000 1000 1000

5.3 Methods

5.3.1 Pre-processing

We filtered all signals with a zero phase 2nd order high- and 4th order low-pass Butterworth filter with cut-off frequencies at 1 and 40 Hz, respectively. This ensures the removal of baseline wander and high frequency noise, and retains the relevant physiological information.

5.3.2 Features

We used the same feature crafting strategy as described in Chapter 4. As a first step, we moved a sliding window over the ECG signals with a width of five seconds and 80 % overlap. As a second step, we computed the ACF from each sliding window. We used a maximum time lag of 250 ms to exclude consecutive heartbeats and to ensure the inclusion of all characteristic waveforms. An example of the ACF of two contact and non-contact signals with a different quality level can be seen in Figure 5.2. From these ACF, we derived three features: FMin, MAmp and Sim.

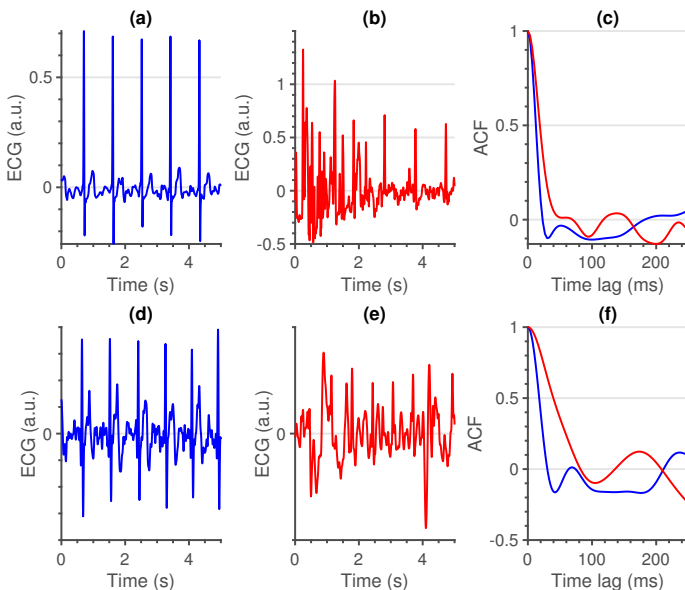


Figure 5.2: Comparison between a clean (blue) and noisy (red) segment for both contact (a-b) and non-contact (d-e) signals. A clear difference in ACF shape can be observed between the two signals.

5.3.3 Classification

In this section, we describe the base classifier and the theory behind the optimization of that classifier using transfer learning. Additionally, we explain the active sampling strategy based on entropy that we want to investigate.

Base classifier

Once each segment was characterized by the set of 3 features an SVM classification model was used to distinguish between clean and noisy segments. This model was used as the base classifier for the transfer learning.

Consider a training set $\{x_i, y_i\}_{i=1}^N$, with $x_i \in \mathbb{R}^3$, as the input data, and $y_i \in \{-1, +1\}$ as the corresponding labels. In this case, class +1 ($y_i=1$) represents clean samples and class -1 ($y_i=-1$) noisy samples. The SVM maps the input data to a high dimensional feature space using a mapping function $\phi(x)$. This transformation allows to separate the datapoints in a higher dimensional

space using the hyperplane $w^T \phi(x) + b$, with w the unknown weight vector and b an unknown bias term.

The SVM can be formulated in the primal space as:

$$\begin{aligned} \min_{w,b,\xi} \mathcal{J}(w, \xi) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad &\begin{cases} y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, \forall i \in [1, N], \end{aligned} \tag{5.1}$$

with w as the weight vector of the base classifier and ξ_i as the classification error of the model on data point x_i . The regularization constant C provides a balance between regularization and minimization of the number of missclassifications: when C is too high, this will lead to overfitting, and when C is too low, this causes underfitting. In this paper, the hyperparameters are tuned using Bayesian optimization in combination with five fold cross-validation [172].

Since the datasets are rather small, we opted to use five fold cross-validation to evaluate the classifiers. The average result was used as a reference for the transfer learning classifier.

Additionally, since we are also interested in the inter dataset performance, we evaluated the classification model of each fold of one dataset, on all folds of the other datasets (Figure 5.3). This results in 25 performance evaluations (5×5 folds). It provides an indication of the similarity of the datasets and the overall performance of the classification model.

Transfer learning

We optimized each base classifier for the other datasets by applying the transfer learning approach described in [51]. The key concept of this approach is the modification of the objective function of the SVM. This is constructed so that it minimizes both the classification error on the new training data and the discrepancy between the adapted and base classifier.

Let $\{\tilde{x}_k, \tilde{y}_k\}_{k=1}^M$ be the M training points of the new dataset, with $\tilde{x}_k \in \mathbb{R}^3$ as the input data, and $\tilde{y}_k \in \{-1, +1\}$, the corresponding labels. De Cooman et al. proposed the following optimization problem for an adapted SVM classifier [51]:

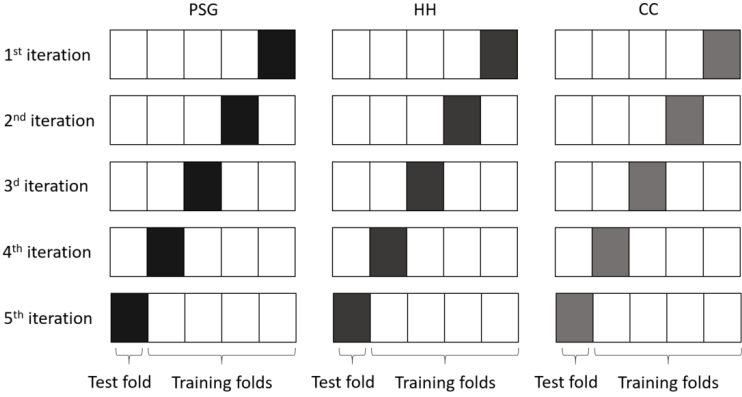


Figure 5.3: We tested the base classifiers using 5 fold cross-validation. Additionally, each base classifiers was also tested on all folds of the other datasets. This results in a total of 25 performance evaluations.

$$\begin{aligned}
 \min_{\tilde{w}, \tilde{b}, \tilde{\xi}} \mathcal{J}(w, \xi) &= \frac{1}{2} \|\tilde{w} - w\|^2 + D \sum_{k=1}^M \tilde{c}_k \tilde{\xi}_k \\
 s.t. \quad &\begin{cases} \tilde{y}_k(\tilde{w}^T \phi(\tilde{x}_k) + \tilde{b}) \geq 1 - \tilde{\xi}_k \\ \tilde{\xi}_k \geq 0 \end{cases}, \forall k \in [1, M],
 \end{aligned}
 \tag{5.2}$$

with \tilde{w} the weight vector of the adapted classifier, D the regularization constant, similarly to C in equation 5.1, \tilde{c}_k an additional weight constant and $\tilde{\xi}_k$ the error of the model on data point \tilde{x}_k . The factor \tilde{c}_k was used in [51] to counter the class imbalance in the dataset. In this study we always selected an equal amount of samples from both classes, so that the subsets for transfer learning were always balanced. This results in $\tilde{c}_k = 1, \forall k \in [1, M]$.

Hyperparameter D allows to balance between minimizing the errors for the new training data points and minimizing the discrepancy with the base classifier. De Cooman et al. empirically adjusted D based on the amount of new data points, for example, if more than 10 new datapoints were available, D was set to 100 [52]. Therefore, since we always have more than 10 new datapoints available, we also used a value of 100. This value might still be fine tuned, but an in-depth analysis of the optimal D is outside the scope of this paper.

We can define the dual representation of 5.2 as

$$\begin{aligned}
\min_{\tilde{\alpha}} &= \sum_{i=1}^N \sum_{k=1}^M y_i \tilde{y}_k K(x_i, \tilde{x}_k) \alpha_i \tilde{\alpha}_k - \sum_{k=1}^M \tilde{\alpha}_k + \frac{1}{2} \sum_{k=1}^M \sum_{l=1}^M \tilde{y}_k \tilde{y}_l K(\tilde{x}_k, \tilde{x}_l) \tilde{\alpha}_k \tilde{\alpha}_l \\
s.t. & \begin{cases} \sum_{k=1}^M \tilde{\alpha}_k \tilde{y}_k = 0 \\ 0 \leq \tilde{\alpha}_k \leq D, \end{cases} \quad , \forall k \in [1, M],
\end{aligned} \tag{5.3}$$

with $\tilde{\alpha}_k$ as a Langrangian multiplier. We can note that

$$\tilde{w} = x + \sum_{k=1}^M \tilde{\alpha}_k \tilde{y}_k \phi(\tilde{x}_k) \tag{5.4}$$

which indicates that \tilde{w} , the weight vector of the adapted classifier, is a combination of information of the original and new recording device.

The advantage of transfer learning is that it allows to adapt an existing classifier with a limited amount of data [147]. Therefore, instead of using the whole training set $\{\tilde{x}_k, \tilde{y}_k\}_{k=1}^M$, we can use a subset $\{\tilde{x}_l, \tilde{y}_l\}_{l=1}^P$, with $P \ll M$. Ideally, this subset represents the main characteristics of the whole training set. In [51], a random subset selection was used. Hence to objectively evaluate the results on the given problem, we applied the same sampling procedure here.

We used 500 randomly selected samples of the training folds of the CC dataset to update the base classifiers of the other two datasets. These were selected 10 times for generalization of the results.

Subset selection

The amount of training samples available for transfer learning can significantly affect the performance of the algorithm. The smaller this training set, the less likely that the underlying distribution of the whole dataset is fairly represented. With this in mind, we investigated the added value of the fixed size sampling technique, which is based on Rényi's quadratic entropy [49]. This was previously described in Section 4.3.3.

We compared the random sampling and fixed size sampling for a varying number of subset sizes. More specifically, we used a subset of 20, 50, 100, 200 and 500 randomly selected samples from the new training set to adapt the classifier. The classes were equally divided over each set. We hypothesize that the fixed

size sampling approach results in a steeper learning curve, compared to the random sampling approach. In order to make a general conclusion, we repeated this procedure 10 times for each subset size. This results in a total of 250 performance evaluations for each subset size. Namely 25 folds, similar to the performance evaluation of the base classifiers, times 10 repetitions per subset size.

5.3.4 Performance evaluation

We assessed the performance of the algorithms using the Acc, percentage of correctly detected samples, Se, percentage of correctly detected clean samples and Sp, percentage of correctly detected noisy samples. Additionally, since the class distribution of the datasets is unbalanced, we used the balanced accuracy (bAcc) to compare the performances [27]. This is computed by taking the average of the Se and Sp. We used paired t-tests to compare the two sampling methodologies. A $p < 0.05$ was considered significant.

5.4 Results

The feature spaces of the three datasets are depicted in Figure 5.4. It can be observed that the feature space of the PSG is more compact compared to the other two. Additionally, the three plots also show a shift in the data distribution. The PSG contains mostly clean samples (96.8%), while the HH contains only 62.3% clean samples, and the CC contains mostly noisy samples (80.5%). The latter is due to the inclusion of 'floating measurements', in the dataset. These are measurements in which the device was turned on, but without a user present [31]. These two observations pose an extra difficulty for the base classifier, since it assumes the same class distribution in the test set, with respect to the training set.

Table 5.2 gives an overview of the results of the base classifiers. We obtained very good results for both contact and capacitive ECG databases (bAcc: 87.5-90.3), when the training and testing subsets are part of the same dataset. However, these similar bAcc's mask the underlying distribution of the Se and Sp values. The balance between Se and Sp changes drastically over the three datasets. We can observe high Se values for the PSG dataset and high Sp values for the CC dataset. Similarly, the CC model obtains a higher Sp compared to Se when tested on the other datasets.

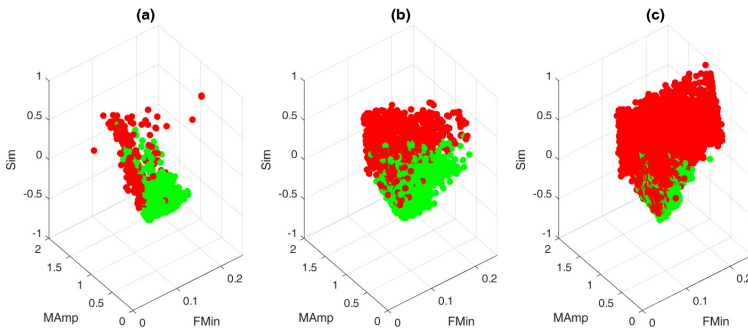


Figure 5.4: Feature space of the three datasets. Green dots indicate clean samples, red dots indicate noisy samples: (a) PSG dataset. (b) HH dataset. (c) CC dataset.

Table 5.2: Results obtained for the base classifiers. The columns indicate the training datasets and the rows the test datasets. The cells in gray indicate the intra dataset results when using five fold cross-validation. The other cells indicate the inter dataset results, averaged over 25 folds. In other words, when the classifier is trained on five training sets and, each time, applied on the five testing sets of the other datasets. The results are shown as mean \pm standard deviation.

	PSG			HH			CC		
	Se	Sp	bAcc	Se	Sp	bAcc	Se	Sp	bAcc
PSG	99.5 \pm 0.1	81.0 \pm 11.9	90.3 \pm 6.0	99.5 \pm 0.12	73.9 \pm 5.3	86.7 \pm 2.7	91.9 \pm 3.0	97.3 \pm 2.1	94.6 \pm 1.0
HH	91.7 \pm 3.4	83.4 \pm 5.8	87.6 \pm 2.6	95.8 \pm 1.3	84.4 \pm 2.4	90.1 \pm 1.1	61.4 \pm 3.7	98.3 \pm 0.5	79.9 \pm 1.9
CC	95.6 \pm 2.5	81.5 \pm 6.9	88.5 \pm 2.7	98.9 \pm 0.5	80.6 \pm 2.2	89.8 \pm 1.1	80.0 \pm 2.7	94.9 \pm 1.2	87.5 \pm 1.8

Figure 5.5 shows the average performance of the base classifiers on the CC dataset compared to the transfer learning approach. For the latter, 500 randomly selected samples were used to update the base classifiers and the whole procedure was repeated 10 times. These results show that the transfer learning classifiers outperform the base classifiers for almost all performance metrics. Only the Se of the base classifier, trained on the HH dataset, decreased. The strongest increase can be observed for the Sp.

As a next step, we investigated the influence of the number of included training samples. Figure 5.4 shows the effect of the number of samples included for training (x-axis) on the different performance metrics (y-axis). We only show the performances up to a training set of 200 samples, since no improvement was

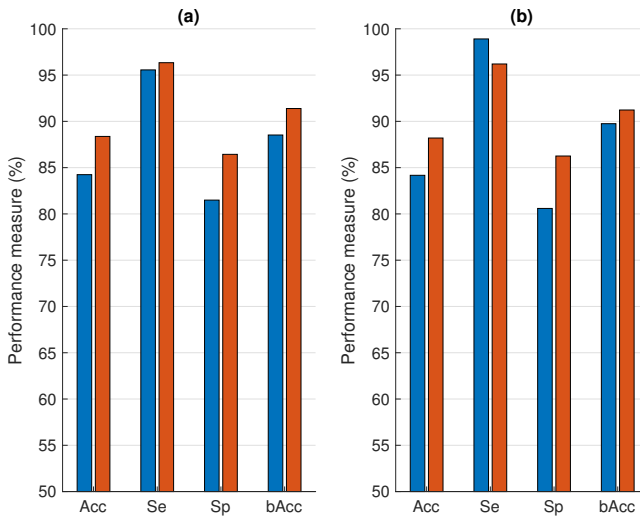


Figure 5.5: Comparison of the performance on the CC dataset without (blue bars) and with (red bars) transfer learning. **(a)** Trained on the PSG dataset. **(b)** Trained on the HH dataset.

obtained hereafter. Despite a clear difference in entropy (Figure 5.7), the only significant improvements of the fixed-size sampling on the PSG base classifiers could be observed in the Se when 100 and 200 samples were included. Moreover, significantly lower Acc and Sp values were obtained with a subset of only 20 samples. No significant differences were obtained for the HH base classifiers.

The Se of the PSG base classifier decreased drastically when 20 training samples were included. From this value, the sensitivity of the PSG base classifier increased significantly until the training subset contained 200 samples. No significant changes occurred hereafter. The Sp increased significantly up to 50 samples, and seized to increase hereafter. The resulting bAcc followed the same V-shaped pattern as the Se. After an initial drop at 20 training samples, it increased significantly for every addition of extra samples. Starting from a subset of 100 samples, the median bAcc of the transfer learning classifier was higher than the base classifier. Additionally, the variability of all performance metrics decreased for every addition of extra samples to the subset. This can be observed by the narrowing of the interquartile ranges.

We observed a similar pattern for the base classifier trained on the HH dataset. For example, the Se also showed a V-shaped pattern, albeit not so outspoken

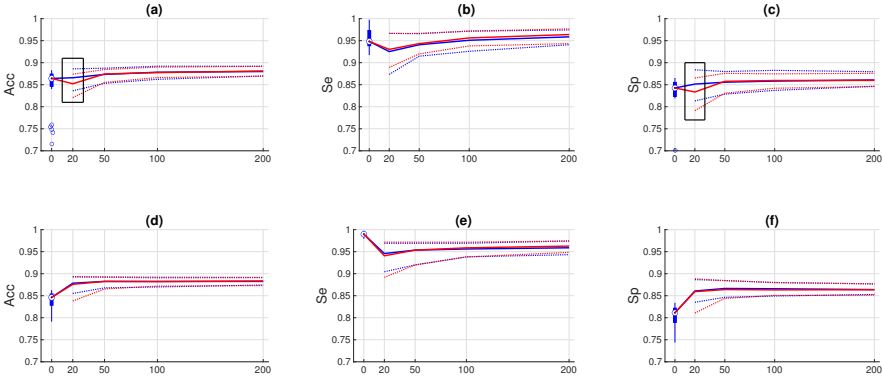


Figure 5.6: The effect on performance of different subset sizes and the two sampling techniques. The first row corresponds to the base classifier trained on the PSG dataset and the second row when trained on the HH dataset. The performance at 0 indicates the performance of the base classifier. Since this is evaluated only 25 times instead of the 250 times for the transfer learning performances, we used a blue boxplot. The blue circles below the boxplot indicate outlier values. The blue line that starts in the median value at 0, corresponds to the random sampling procedure and the red line to the fixed size procedure. The dotted lines indicate the interquartile ranges. These provide insight in the variability of the classifier performance. The black boxes indicate a significantly lower performance of the active sampling strategy, compared to random sampling.

as the other base classifier. When more samples were added to the subset, the Se increased significantly up to a subset of 100 samples. However, it never reaches the same Se as the base classifier. A significant increase in Sp could only be observed between a subset of 50 and 500 samples. The balanced accuracy increased with every addition of more samples, except between 100 and 200 samples. An increase could be observed between these two values, but it was not significant. In contrast to the other base classifier, here the median bAcc of the transfer learning classifier was already higher after the inclusion of 20 samples in the training subset.

As an extra experiment, we wanted to see how much the transfer learning approach could improve the lowest performance metric from Table 5.2. The lowest value was obtained for the Se of the CC base classifier, when tested on

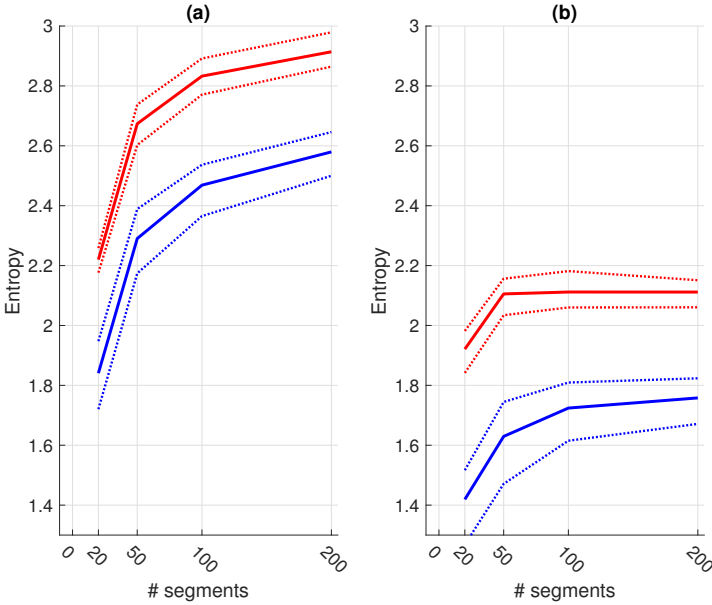


Figure 5.7: The difference in entropy for the clean (a) and noisy (b) samples of the CC subsets. The blue line indicates the random sampling and the red line indicates the fixed-size sampling approach. The full line indicates the median values and the dotted lines indicate the interquartile ranges. We can observe that the entropy is consistently higher for the fixed-size sampling.

the HH dataset. Hence, we applied the transfer learning approach on the CC classifier and incrementally added more training samples from the HH dataset. A strong increase could already be observed after including only 20 samples. This increase continued with the inclusion of more samples, together with a decrease in Se variability.

5.5 Discussion

5.5.1 Features & base classifier

The primary goal of this study was to optimize the performance of a simple artefact detection model, that is trained on contact ECG, towards ccECG. To

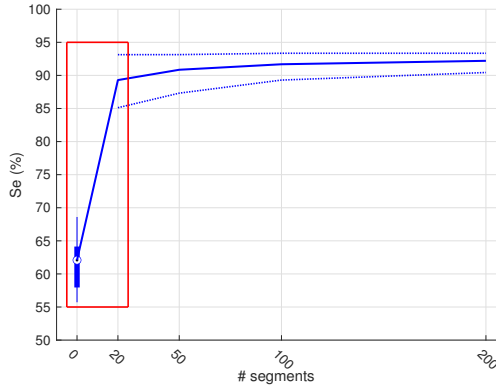


Figure 5.8: The Se of the CC base classifier when applied on the HH dataset. The full line corresponds to the median values and the dotted lines indicate the interquartile ranges. We used a different graphical representation for the results of the base classifiers at zero, since these results originate from only 25 folds, compared to the 250 folds of the transfer learning approach.

this end, we used three features that have previously shown to produce good results for artefact detection in contact ECG [128].

In order for a transfer learning approach to be successful, we need to be sure that the proposed features are able to make a distinction between clean and noisy samples. Visually, this seems to be the case for the PSG dataset, since little to no overlap between the two classes can be observed in the feature space (Figure 5.4). However, this overlap increases for the HH dataset and even more for the CC dataset. Quantitatively, the little overlap of the PSG dataset results in a high performance of the base classifier on the test set of the PSG dataset (Table 5.2). We obtained a bAcc of 90.27 ± 6.0 . Remarkably, despite the visible overlap in the feature space, a similar bAcc was obtained for the two other datasets. This indicates that the proposed features are able to make an accurate distinction between clean and noisy samples, regardless of the recording modality. These results confirm the findings in [128], where accurate results were obtained when a classifier was trained on a subset of the PSG dataset and tested on an independent subset of the HH dataset. However, the results in this study allow for a more trustworthy conclusion, due to the robustness of the analysis.

The bAcc of the CC base classifier, when tested on the CC testing folds is on average 87.5 ± 1.8 (Se= 80.0 ± 2.7 , Sp= 94.9 ± 1.2). Despite this high accuracy, it

is lower compared to the results obtained on the same dataset in [31]. There, a bAcc of 94.02 (Se=95.19, Sp=92.85) with a linear SVM classifier was obtained. In that paper, only three features, which were selected with a threshold-based one-level decision tree (DT) approach, were used. However, one important difference exists between these studies. Castro et al. used only one random training and testing split. Hence, it is hard to draw general conclusions from these results. In this study we performed a 5-fold cross-validation, which produces results that should generalize better.

In the introduction, we hypothesised that an artefact detection model, trained on contact ECG, and without any adaptation, is not able to accurately distinguish clean from noisy ccECG signals. We assumed that this would be due to the artefact characteristics of the different recording modalities [31]. The resulting Acc values seem to confirm this hypothesis, since they are remarkably lower, compared to the values of the original classifier. This is due to a substantial decrease in Sp, since the Se values increased. These findings indicate that the clean ccECG segments are accurately detected, but that the noisy segments are not. Hence, one could assume that this confirms the initial hypothesis and is indeed caused by a difference in artefact characteristics. However, that assumption would ignore another important factor, namely the difference in class distribution between the different datasets. For example, the distribution of the PSG dataset is drastically different from the CC dataset. The PSG dataset contains predominantly clean segments (96.8%), while the opposite is true for the CC dataset (19.5%). This information does not exclude the difference in artefact characteristics as an influencing performance factor, but highlights the fact that the class distribution is also a major performance influencing factor.

5.5.2 Transfer learning

The transfer learning approach used in this paper has previously shown its added value for the two previously stated performance influencing factors. For instance, in [51], it was used to adapt a patient-independent seizure detection algorithm towards a patient-dependent solution. De Cooman et al. have shown that the inclusion of a limited number of patient-specific seizures can result in a strong increase of performance for most patients [51]. In other words, a patient-independent model, which performance is otherwise too low for practical use, was successfully optimized with the proposed transfer learning approach, using patient-specific seizure characteristics.

The approach that they used was an adaptation of the transfer learning approach developed in [203]. Originally, it was intended to adapt existing classifiers to a new set of data with a different distribution, much like the problem in this study.

They have shown that their method outperforms several standard methodologies, when it is applied on an artificial and a benchmark classification task. Hence, regardless of the reason why the base classifiers fail to accurately detect the noisy segments of the CC dataset, the transfer learning approach should be able to improve the classification performance.

From Figure 5.5, we can observe that this is indeed the case. The performance of the PSG and HH base classifiers on the CC dataset significantly increased when transfer learning was applied. This indicates that the transfer learning approach successfully improved the performance of a classifier trained on contact ECG. The only performance metric that failed to increase was the Se of the HH dataset. This could be explained by the very high Se value for the base classifier. Visually, this means that the decision boundary of the HH base classifier is directed more towards the centroid of the noisy class in the CC dataset (Figure 5.4). This results in a very high Se and a low Sp. However, when transfer learning is used, this decision boundary presumably shifts towards the middle between the two classes. This shift results in a higher Sp and a lower Se.

In this work, we used a simple approach with only three features to show the relevance of the transfer learning. Future work should consider the addition of other features in order to improve the performance even further.

Subset selection

We have shown that transfer learning allows to train a device-specific classifier, while using only a limited amount of device-specific data. More specifically, we used 500 labelled samples to obtain the results depicted in Figure 5.5. Since labelling is a very tedious and time-consuming task, it is important to know the minimum amount of samples that are necessary to have an acceptable accuracy. To this end, we performed an experiment in which we varied the amount of randomly selected training samples from 20 to 200. Additionally, we performed the same experiment, using an active sampling approach, based on entropy. We hypothesized that the active sampling approach would result in a steeper learning curve, compared to the random sampling approach.

Remarkably, the only significant improvements of the PSG base classifiers by the active sampling could be observed in the Se when 100 and 200 samples were included. No significant differences could be observed for the other sampling sizes and performance metrics. On the contrary, significantly lower Acc and Sp values were obtained with a subset of only 20 samples. Hence, despite a higher entropy for all training subset sizes, the proposed active sampling strategy did not have a steeper learning curve and did not reach the performance plateau faster.

This could be because the samples at the edges of the feature space are more likely to be selected since they increase the Rényi quadratic entropy of the subset. However, these samples might poorly represent the underlying distribution of the whole dataset. Tong et al. have shown that selecting samples close to the decision boundary can improve classification performance [182]. Hence, in future research, we should investigate such approaches.

From the current analysis, we can state that an increase in entropy does not result in an increase of informative samples. Moreover, we can conclude that active sampling based on entropy does not provide an added value for transfer learning, compared to random sampling.

In Figure 5.4 we can observe similar patterns for the Se and Sp for the two base classifiers. Namely, a V-shaped pattern for the Se and a mirrored version for the Sp . This pattern can be explained by an initial overfitting, when a small training subset is used, and an incremental increase in performance when more training samples are included. However, this increase seems to stagnate when more training samples are included. In other words, the performance reaches a plateau. For both base classifiers, this appears to be when the training subset contains 200 samples (100 samples of each class).

5.5.3 Limitations

A limitation of this study is the composition of the CC dataset. We used data from two ccECG recording devices, which recorded signals in various scenarios, including floating sensors. We made the assumption that the difference in artefact characteristics can be related to the distinction between contact and non-contact ECG. However, it might be that the recording devices and scenario's in which they were used also influence these characteristics. In a future study this should be investigated into more detail.

Another limitation is the lack of demographic information from the recorded subjects. Albeit we believe that for artefact detection this is not of major influence, it could provide more insight into the differences between the datasets.

Due to the small amount of either clean or noisy samples, depending on the dataset, no deep learning approaches were tested. Future research could investigate the added value of such an approach on a larger dataset.

5.6 Conclusions

We showed that the features that were crafted for contact ECG can be used for artefact detection in ccECG signals. This finding allowed us to successfully construct a classifier for contact ECG and optimize it towards ccECG using transfer learning. Moreover, we showed that the performance of the base classifiers could be significantly improved using only a limited amount of samples.

No added value was obtained by using an active sampling approach based on entropy maximisation compared to random sampling. Hence, we do not recommend to use this approach for transfer learning in the future.

In conclusion, transfer learning is able to optimize a modality independent artefact detection model, when only a limited amount of modality-specific samples is available. This results in less labelling and more accurate results.

Chapter 6

Respiration quality

In this chapter we investigated the possibility to detect noisy respiratory segments using two machine learning approaches. The first approach is the more classical feature-based approach and the second approach is a representation learning approach based on neural networks. We have shown that both approaches are suitable for artefact detection. To the best of our knowledge, it is the first time that this is investigated for impedance pneumography signals. The results of the study are reported in **MOEYERSONS, J.**, MORALES, J., SEEUWS, N., VAN HOOFF, C., HERMELING, E., GROENENDAAL, W., WILLEMS, R., VAN HUFFEL, S., AND VARON, C., Artefact detection in impedance pneumography signals: a machine learning approach, Internal Report 21-18, ESAT-STADIUS, KU Leuven (Leuven, Belgium), 2021 (submitted for publication).

6.1 Introduction

Chronic respiratory diseases (CRD) are among the leading causes of morbidity and mortality worldwide. Among all types of CRDs, chronic obstructive pulmonary disease (COPD) ranks the highest worldwide [107]. It is a type of obstructive lung disease characterized by long-term breathing problems and poor airflow. The main symptoms include intermittent coughing and shortness of breath. These symptoms may be mild at first, but can become more constant up to a point where it becomes difficult to breathe.

The standard respiratory function test for case detection of COPD is spirometry [82]. This is a maximum breathing test, which is used to objectively determine the ventilatory capacity of the lungs. It is highly reproducible, practical and safe, but requires large, stationary equipment and trained personnel. The latter makes it impossible to be conducted outside of the clinical environment. Additionally, since the subjects are asked to breathe through a mouthpiece or wear a face mask, the normal breathing pattern of the subjects could be altered [8]. More comfortable methods, such as bio-impedance (BioZ), inductance plethysmography or electromyography could solve this issue. These techniques are minimally- or non-invasive, but currently lack validation in clinical applications.

Thoracic BioZ measurement has been suggested as a convenient and comfortable ambulatory technique for healthcare monitoring of respiratory diseases in normal and restrictive breathing [23]. It is a non-invasive technique that measures changes in the electrical impedance of a person's thorax by injecting a high-frequency low-amplitude sinusoidal current herein. The passage of current through the tissue results in a voltage difference which can be measured. These impedance variations are mainly a result of gas volume changes in relation to fluid volume displacement while breathing. The conductance path is similarly affected by inspiration and expiration movements [73, 22]. Compared to spirometry, this technique does not interfere with the breathing pattern of the patient and allows a longer analysis period. Several studies reported a strong linear relation with respiratory volume when measured in the thorax [167]. Since it can be recorded with a wearable device, it could be used for preventive screening and patient follow-up [30].

Despite the obvious advantages, BioZ also presents some disadvantages. For instance, it remains unclear how all thoracic components, such as fluids and organs, contribute to the measurement [22]. Moreover, due to the ambulatory nature of the technique, the recordings are prone to different kinds of noise sources. These result in the presence of artefacts which could alter the quality of the signal. Some of these artefacts, such as baseline wander, are easy to remove,

but motion artefacts, for example, remain a challenge [178]. Since signal quality has a huge impact on the performance of data-driven decision support tools, it is important that segments of poor quality are identified and removed [91].

Previous studies have focused mainly on removing the cardiac component of BioZ signals [167, 166, 120] or removing motion artefacts using adaptive filtering approaches [5, 157]. However, only very little research has been conducted to detect and remove segments that are contaminated beyond repair. In [121, 36] the authors handcrafted characteristic features and used heuristics to separate clean from contaminated segments. We believe that machine learning algorithms could improve upon these methodologies.

In this study, we investigated the use of machine learning algorithms to separate clean from noisy BioZ signals. We compared two approaches. The first approach is a feature-based classification model. Here, characteristic features that represent quality aspects of the signal are hand-crafted. These are then fed into a classification model, which separates them into classes of different quality. The advantage of this approach is that the obtained results are interpretable. However, the downside of this approach is that the performance of the classifier is limited by the effectiveness of the features to capture relevant changes due to noise and artefacts.

The second approach uses a convolutional neural network (CNN). This technique learns characteristic features directly from the raw data and uses these to classify the signals. The difficulty is to select an adequate network architecture so that the model is able to learn accurate representations of the data.

To the best of our knowledge, this is the first study that compares a feature-based classifier and a neural network approach for the classification of thoracic BioZ signals.

The outline of this study is as follows. In Section 6.2, we described the subjects and the data recording protocol. In Section 6.3, we described the proposed methodologies. The results are shown in Section 6.4 and discussed in Section 6.5. The conclusions are drawn in Section 6.6.

6.2 Data

6.2.1 Subjects

Forty seven COPD patients were recruited at Ziekenhuis Oost-Limburg (Genk, Belgium), during their consultation or rehabilitation session (Table 6.1). All

patients provided written consent before they participated in the study. The study followed the World Medical Association's Declaration of Helsinki on Ethical Principles for Medical Research Involving Humans Subjects and was approved by the local institutional medical ethics committee from Ziekenhuis Oost-Limburg. The reference number is 18/0047U.

Table 6.1: Demographics of the COPD patients. FEV1: forced expiratory volume in one second. FVC: forced vital capacity. The data are presented as mean \pm standard deviation.

Age (years)	64.6 \pm 6.5
BMI (kg/m^2)	26.2 \pm 4.9
Male	36
Female	11
FEV1 (% predicted)	55.6 \pm 18.8
FVC (% predicted)	91.8 \pm 27.4
FEV1/FVC (%)	48.7 \pm 13.8

6.2.2 Data acquisition

Each patient was equipped with a wearable, as well as with a standard wired acquisition system to record their respiration. The latter was used as gold standard and allowed the wearable signals to be labelled for quality. We summarized the details of both devices below.



Figure 6.1: The wearable BioZ device (ROBIN, Stichting imec the Netherlands, Eindhoven, the Netherlands) was used to measure respiration. Figure taken from [185] with permission.

The wearable device (ROBIN, Stichting imec the Netherlands, Eindhoven, the Netherlands) uses bioimpedance (BioZ) to measure respiration (Figure 6.1). It has a sampling frequency of 16 Hz. Eight stress test AG/AgCl electrodes (Kendall H92SG, Covidien Inc., MA, USA) were placed on different locations on the thorax (Figure 6.2) and were used to create a tetra-polar electrode configuration. The electrodes were placed symmetrical at both sides of the thorax. In this tetra-polar configuration, two leads are used for injecting the excitation current (I) and the other two measure the generated voltage (V). More details about this configuration can be found in [23, 22, 21]. The injection current amplitude was $110 \mu A$ at a frequency of 80 kHz. This setup ensures a good signal-to-noise ratio and a linear correlation of the BioZ signal with the volume changes [21].

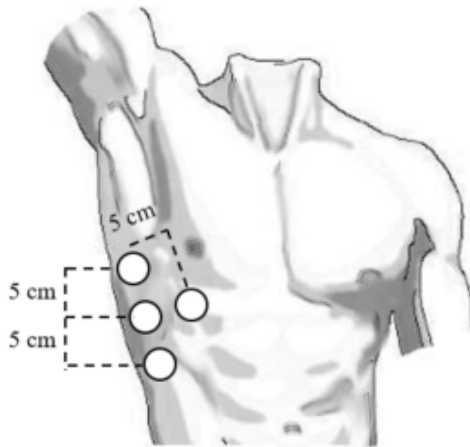


Figure 6.2: Representation of the tetra-polar electrode configurations. Only the right side is shown because the configurations were symmetric from the midsternal line.

We used the Biopac wired acquisition system as gold standard. This system measures respiratory airflow with an airflow transducer (pneumotach transducer TSD107B, Biopac Systems, Inc., Goleta, CA, USA). The signals were digitized with a sampling frequency of 10 kHz. The subjects were instructed to breathe through a disposable mouth piece, with a bacterial filter (AFT36, Biopac Systems, Inc.), which was attached to the pneumotach device, and they were wearing a nose clip.

6.2.3 Respiratory protocol

Each subject performed an inspiratory threshold loading protocol, during which their respiration was recorded. This kind of protocol initiates changes in breathing mechanics, similarly to an airway obstruction. For example, it is known to affect the breathing pattern and diaphragm fatigue [57]. The same protocol was previously used in [23] and [108] to validate the BioZ device in a clinical setting.

The protocol consisted of two minutes of quiet tidal breathing, followed by imposing five inspiratory threshold loads to the subjects while breathing. The threshold values were defined as increasing percentage values from the maximal static inspiratory pressure (MIP) [1]. This reflects the maximal pressure that a person can generate during inspiration and it indicates respiratory muscle strength. Hence, before the start of the protocol, the subjects performed a maximal volitional test to derive the MIP. An inspiratory muscle trainer device (POWERbreathe KH2, POWERbreathe International Ltd, Southam, UK) was used for the MIP derivation and to impose the threshold loads.

The five load values corresponded to a progressively increasing percentage of each subject's MIP. The following threshold values were utilized: 12 %, 24 %, 36 %, 48 % and 60 % of the subject's MIP. The subjects were instructed to breathe 30 times per load and each load was followed by a two-minute resting period to return to baseline. The load protocol is shown in Figure 6.3. If a subject failed to complete one of the thresholds, then that recording was excluded from further analysis. The subjects remained seated in an upright position during the entire protocol and wore a nose clip to block nasal breathing.

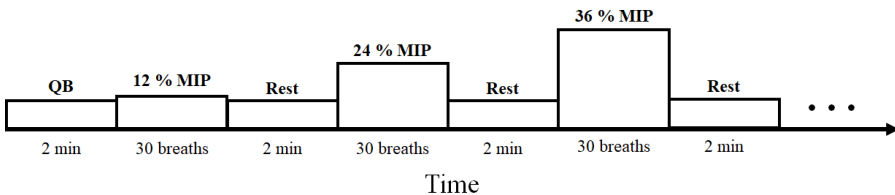


Figure 6.3: The inspiratory threshold loading protocol. After two minutes of load-less quiet breathing (QB), the load values, expressed as percentages of the maximal static inspiratory pressure (MIP), are progressively increased. Each loading task is followed by two minutes of rest.

6.3 Methods

6.3.1 Pre-processing

Prior to signal labelling, we implemented two pre-processing steps: filtering and segmentation. Firstly, we filtered all signals, both wearable and gold standard, with a high-pass zero-phase Butterworth filter with a cut-off frequency of 0.05 Hz (3 breaths/minute). This ensures the removal of baseline oscillations, but prevents the loss of physiological information from the respiration signal. Hereafter, we low-pass filtered both signals with a zero-phase Butterworth filter with a cut-off frequency of 0.70 Hz (42 breaths/minute). This removes high frequency content that is not related to breathing.

Secondly, since not every recording was equally long, 2 minutes for the resting phase vs. 30 breaths for the loading phase, we implemented an additional segmentation procedure. Before the actual segmentation, we removed the first three seconds and the last second of each recording, because visual inspection showed distortions due to the on/off setting of the device within these time spans. In the second step, we divided each recording into non-overlapping one-minute segments. This allowed to extract features from equally long segments.

Note that each recording is measured with four electrode configurations. Due to the fact that electrode movement could be configuration specific, we regarded these configurations as different recordings, albeit with the same gold standard.

In total, we extracted 1896 one-minute segments, sampled at 16 Hz, whereof the most obvious noise sources (i.e., baseline wander and powerline interference) were removed. An example of the output of the pre-processing steps is shown in Figure 6.4 with both a raw and a filtered respiratory signal.

6.3.2 Labelling

Each recording was labelled by four independent annotators with experience in biomedical signal analysis. To visualize, comment and label each signal as easily as possible, we developed a GUI in MATLAB (Figure 6.5).

The GUI contains some useful tools to counter some known issues. For example, on several occasions the BioZ signal was delayed or advanced, compared to the airflow signal. This was especially the case during high loads. In order to improve the alignment of the two signals, the tool allows to manually shift the signal with a maximum of 2 seconds in both directions. Additionally, an accurate comparison might be hampered by the difference in polarity of the

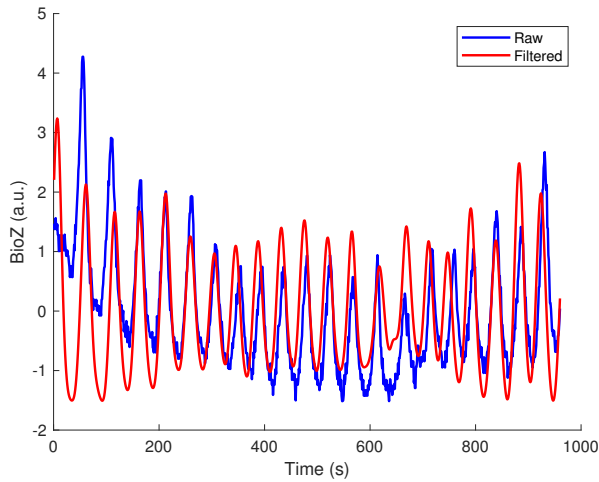


Figure 6.4: An example of the result of the pre-processing steps. The blue line indicates a raw signal of one minute and the red line indicates the filtered version. The baseline wander and high frequency noise is mostly removed.

BioZ signal and the airflow signal. This can also be adapted in the GUI by inverting the polarity of the BioZ signal.

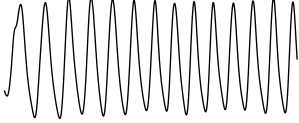
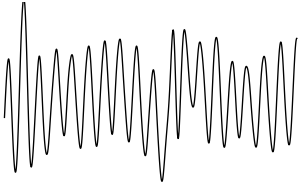

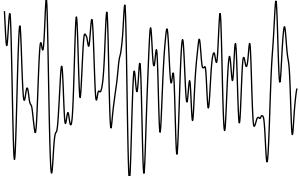
These tools allowed a proper alignment of the two signals, which aided in correctly annotating the BioZ signals. However, note that the adaptations of the annotators might not always be the same. These were not stored and therefore also not compared. Thus, it might be that the annotators made their annotation based on a slightly different visual representation of the two signals. Nevertheless, due to the quasi-periodic nature of the respiration signal, these slightly different annotation set-ups do not affect the final labelling results.

We categorized the signals in five classes. If the gold standard is of bad quality, due to the subject not breathing through the reference device, motion artefacts or signal saturation, then the signal is categorized in class (5), "Bad reference quality". We described the other classes in Table 6.2.

6.3.3 Classification

We compared two classification approaches, namely a feature-based SVM model and a CNN. The approaches are described in the following paragraphs.

Table 6.2: Overview of the different class labels. Note that these rules were created for one-minute segments.

Classes	Examples
<p>(1) Excellent signal quality</p> <ul style="list-style-type: none"> • all breaths can be identified • the same number of prominent peaks as the gold standard • the amplitude differences between peaks are relatively the same <ul style="list-style-type: none"> – a difference of 20% is allowed 	
<p>(2) Good signal quality</p> <ul style="list-style-type: none"> • all breaths can be identified • the same number of prominent peaks as the gold standard 	
<p>(3) Average signal quality</p> <ul style="list-style-type: none"> • at most 10 seconds of the signal is allowed to be corrupted <ul style="list-style-type: none"> – missing beats – additional bumps in the signal 25% higher than the expected amplitude 	
<p>(4) Bad signal quality</p> <ul style="list-style-type: none"> • more than 10 seconds of the signal is corrupted 	

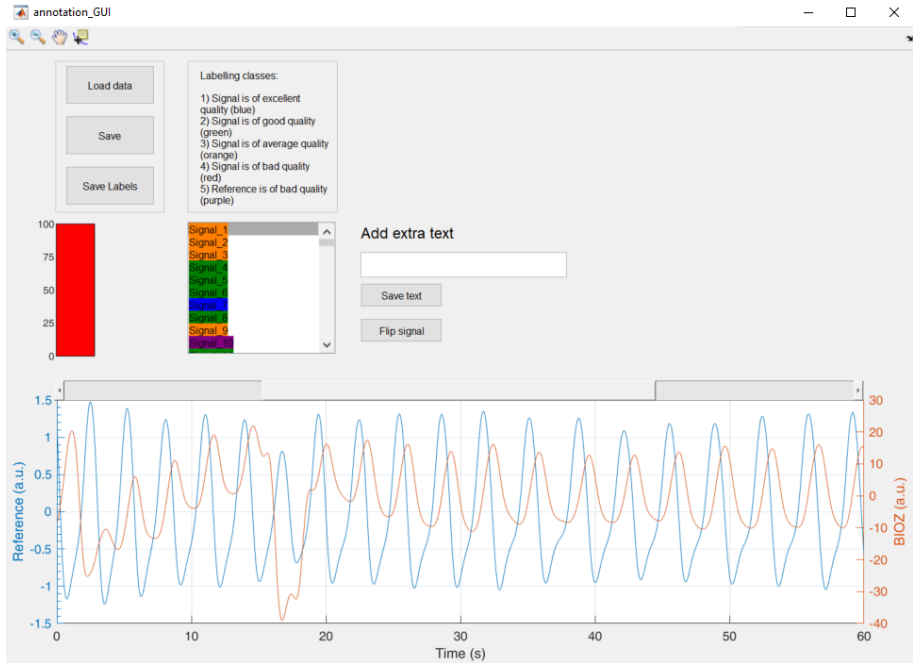


Figure 6.5: Graphical user interface for visualizing, commenting and labelling each signal. A.u. stands for arbitrary units.

SVM

Firstly, we computed features from the entire 60 seconds segment. These provide a general view of the signal, but could have difficulties capturing local variations. Hence, secondly, we divided the signal into non-overlapping 15 seconds segments and computed the same features. The mean and standard deviation of 4 consecutive features were used to characterize each 60-seconds segment. We used these as separate features.

Two data representation methods were used to highlight the differences between the classes: (1) the auto correlation function (ACF) and (2) the power spectral density (PSD) estimation.

1. ACF

The ACF can be used to inspect the (quasi-)periodic components within a time series. For a series of breaths during normal breathing, the amplitudes of the first and second peak in the ACF represent phase shifts equal to one and two breaths, respectively.

Figure 6.6 shows an unbiased ACF of a clean and contaminated BioZ signal during normal breathing. A semi-sinusoidal signal, such as the respiratory signal, should reach ACF values close to one when the lag is equal to a multiplication of the breathing period. Under this assumption, the first and second peak of the ACF should represent a lag of, respectively, one and two breaths. The amplitude at the first peak (Ap1) can be regarded as an expression of the regularity of the respiratory signal [122]. A low Ap1 generally indicates a low regularity and thus, a higher likelihood that an artefact is present. In this case, the amplitude at the second peak (Ap2) will also be low. Thus, closeness of both Ap1 and Ap2 to one reflects breathing regularity. Additionally, the ratio of Ap1 and Ap2 is a measure of symmetry, such that an Ap1/Ap2 close to one reflects a repetitive signal.

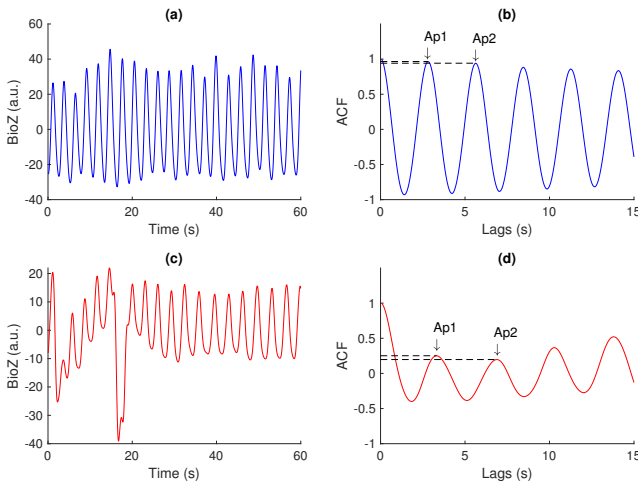


Figure 6.6: **(a, c)** A clean and noisy BioZ signal during normal breathing. **(b, d)** The resulting ACFs. It can be observed that the amplitude of the first and second peak for the clean signal is higher, compared to the noisy signal.

2. PSD

The PSD is the measure of signal's power content versus frequency. It is often used in biomedical signal analysis to calculate the proportion of the spectral distribution within certain physiological frequency bands.

In the case of respiratory signals, most spectral energy is concentrated around the breathing frequency. The more regular the breathing pattern, the more energy is contained around this peak. Hence, we could use the bandwidth of the signal as a feature for signal quality. This is defined by the frequencies where the gain drops below 70.79 % (-3 dB) relative to the main peak (Figure 6.7).

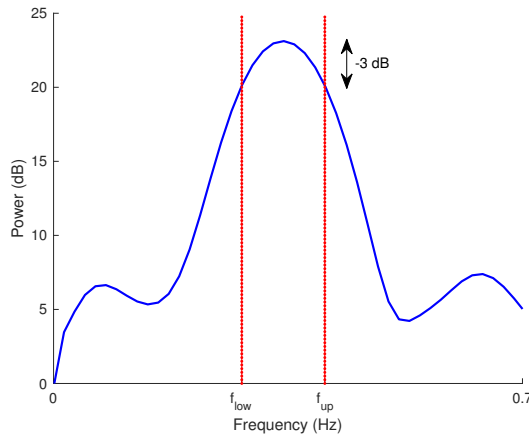


Figure 6.7: The PSD of the signal shown in Figure 5.2. The blue line represents the PSD and the dotted red lines represent the -3 dB bandwidth around the largest peak. f_{low} indicates the lower bound and f_{up} indicates the upper bound.

In this study, we calculated the PSD of the BioZ signal using the Welch periodogram and a Hamming window of 30 seconds with 50 % overlap. In the case of the smaller, 15 seconds, segments, we used the entire segment without overlap.

We defined four features hereof: the frequency of the lower (f_{low}) and upper (f_{high}) bounds, the bandwidth and the normalized power contained in the bandwidth. The normalization was done with respect to the power between 0.05 Hz and 0.70 Hz (3 to 42 breaths/minute). Ultimately, this normalization results in a value between zero and one. The closer this value is to one, the more regular the respiration signal is. Therefore, a low value is an indication of an irregular breathing pattern (i.e. broadband respiratory signal) or the presence of artefacts.

We used the resultant 21 features, 7 from the entire one minute segments and 14 from the smaller segments (i.e. 7 means and 7 standard deviations), to train an SVM model to classify the data as clean (1) or noisy (-1). The features derived from the smaller segments are further referred to as the *grained* features. We used an RBF kernel and the hyperparameters were automatically tuned using Bayesian optimization and 5-fold cross-validation on the training set.

Initially, we trained the SVM model only with the features derived from the entire one minute segment. This can be considered the simplest approach, since no extra segmentation of the recording is needed. Then, we retrained the model with the *grained* features derived. The hypothesis is that these features are better at capturing variations within the recording due to non-stationarities and thus should result in a better classification performance.

We trained each model 10 times with a cross-subject approach with 70 % of the subjects in the training set and the remaining 30 % in the test set. For generalization purposes, this division was done 10 times at random.

In order to identify the most relevant feature set from all features, we used the minimum redundancy, maximum relevance (mRMR) feature selection algorithm [150]. This algorithm minimizes the redundancy of a feature set, while maximizing the relevance to the response variable, in this case the corresponding class. As a result, it ranks the features based on mutual importance.

For each training fold, we ranked all features. We then used the highest ranked features to train an SVM model and compared their respective performances. A graphical overview of the different steps per fold is given in Figure 6.8.

CNN

CNNs are a class of deep neural networks, frequently used in image analysis. They use convolution operations instead of general matrix multiplications as in multi-layer perceptrons. The main advantage of this type of network is the ability to learn the features automatically from the data. This decreases the need for prior knowledge and human effort in feature design. CNNs have been successfully used to detect artefacts in ECG, EEG and other (biomedical) signals [206, 138, 88].

The proposed architecture of the 1-dimensional CNN is shown in Table 6.3. We trained the CNN on input signals of 1 minute, which contain 960 samples. Each input signal is first normalized by subtracting its mean and dividing by its standard deviation. Hereafter, it is passed through two blocks of convolutional layers. Each layer of the first block consists of 10 filters and operates on 32 samples along the temporal axis. This equals a window of 2

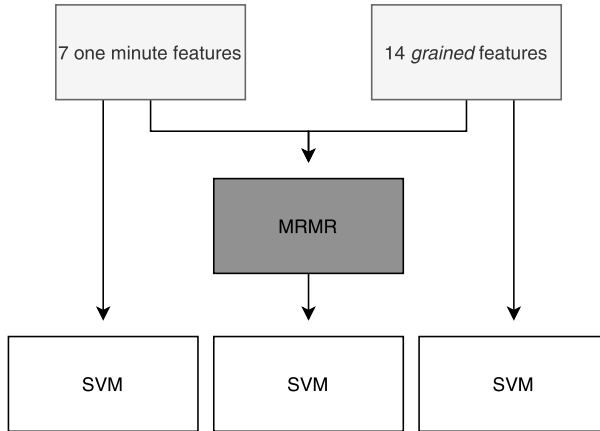


Figure 6.8: Workflow per fold of the feature based approach. First we derive the one minute and *grained* features. These are used separately to train an SVM model. Then, we use the mRMR algorithm to rank all features and retrain the SVM model with the highest ranked features.

seconds which, if expressed in breathing frequency, is 30 breaths per minute. Since an average adult at rest has a respiratory rate between 8 to 14 breaths per minute approximately, this window ensures the inclusion of at least one breathing cycle per input sequence for the first convolutional layer. The result of the convolution is passed through a rectified linear unit (ReLU) activation function. The hyperparameters of the layers of the second convolutional block are equal to the first block, but here, only 5 filters were used instead. This was done to create a feature map with five features, similar to the SVM approach.

Table 6.3: Overview of the network architecture.

	Layer info	Output size	
	Input:	(960,	1)
1	Conv (32) x 10, stride=2	(480,	10)
2	ReLU	(480,	10)
3	Conv (32) x 10, stride=2	(240,	10)
4	ReLU	(240,	10)
5	Conv (32) x 10, stride=2	(120,	5)
6	ReLU	(120,	5)
7	Conv (32) x 10, stride=2	(60,	5)
8	ReLU	(60,	5)
9	Global avg pooling	(1,	5)
10	SoftMax	(1,	2)
Total number of parameters:		5962	

A CNN usually consists of alternating convolutional and pooling layers. However, Springenberg et al. have shown that pooling layers can simply be replaced by convolutional layers with increased stride, without a loss in performance [174]. Hence, in order to keep the network as simple as possible, we used strided convolutions to downsample the signal. The stack of convolutional layers results in a temporal axis that is downsampled to 60 time samples and 5 features per time sample (60×5). Hereafter, we added a global average pooling layer. This layer generates a single feature vector by taking the average of all feature maps over the temporal axis. The main advantage of the global average pooling is the robustness to temporal translations of the input [103]. Another advantage is that there are no parameters to optimize for this computation, thus overfitting is avoided at this layer. The resulting vector is fed directly into a fully-connected 2-node output layer. This final layer uses a softmax function to convert the output vector to a vector of categorical probabilities.

We used the Adam version of stochastic gradient descent for 50 epochs and a batch size of 100. The categorical cross-entropy is used as loss function. During training, the loss of a separate validation set is tracked and the weights with the best validation loss are retained. These weights are then used to evaluate the model on the test set.

We trained each model 10 times with a cross-subject approach with 70 % of the subjects in the training set and the remaining 30 % in the test set. For generalization purposes, this division was done 10 times at random. To get a correct comparison, we used the same splits for both approaches (SVM and

CNN).

6.3.4 Performance evaluation

We used four common performance metrics to evaluate and compare the performance of the proposed approaches: the Se, which measures the proportion of clean signals that are correctly identified as clean, the Sp, which measures the proportion of noisy signals that are correctly identified as noisy, the Acc, the proportion of signals that are correctly classified and the AUC.

We measured the statistical differences ($p < 0.05$) with a Student's t-test or Wilcoxon rank sum test, depending on the normality of the data. The latter was measured with a Lilliefors test.

6.4 Results

The goal of this study was to examine whether machine learning algorithms could be used to separate clean from noisy thoracic BioZ signals. Hence, the first four classes were binarized per rater into a clean (1) and a noisy (-1) class, where class 1 and 2 were considered clean and class 3 and 4 were considered noisy. Hereafter, majority voting among raters was performed to create a single label per signal.

The annotators fully agreed on 79.01 % of the signals, the majority agreed on an additional 18.29 % and no agreement was obtained for 6.70 % (Table 6.4). This resulted in a Fleiss Kappa of 0.75, which excludes the hypothesis of random annotation and confirms a strong agreement between the raters.

Due to the bad quality of the reference signals in class 5, we decided not to use them for further analysis, regardless of the actual BioZ quality. From the other classes, we used only the segments where at least three out of four raters agreed (93.30 %). This resulted in a total of 1471 one minute segments.

Table 6.4: Overview of the dataset. The rows represent the level of agreement. Agreement: all annotators agreed; Majority: three out of four annotators agreed; Disagreement: the annotators disagreed. The columns indicate the class to which the signals belong.

	Clean (1)	Noisy (-1)	Bad reference quality (5)	All
Agreement	1007	234	257	1498
Majority	111	119	53	283
Disagreement				115
Total	1118	353	310	1896

Initially, we trained the SVM model using only the features derived from the entire one minute segments. This can be considered the simplest approach, since no extra segmentation of the recording is needed. This resulted in an Acc of 83.72 ± 2.99 % and an AUC of 84.26 ± 5.12 % (Table 6.5). When we compared this approach with the results when only the features derived from the smaller segments were considered, further referred to as the *Grained* approach, we could observe some significant differences. The Acc, Sp and AUC all significantly improved ($p < 0.05$), while no significant changes could be observed for the Se. Since artefacts usually do not span the entire segment, this supports the initial assumption that features derived from the entire segment cannot accurately capture local variations or non-stationarities due to artefacts.

The best combination of features was investigated using the mRMR algorithm. For each training fold, we ranked all features. We then used the selected feature sets to train an SVM model and compared their respective performances. Given the simple nature of the respiratory signal, we only included up to five features. This was done in order to avoid overfitting. We depicted the average performance of each model in Table 6.5. Note that the feature combination could change each training run.

The AUC tends to increase when more features are included. However, the only significant increases were obtained when we compared the inclusion of one and two features with five features. Moreover, no significant changes in AUC could be noted after the inclusion of three features. The Se values did not increase significantly for any inclusion of more features. This means that the gain in performance was obtained by a better detection of noisy segments. Despite the apparent increase in the average Sp, the Sp of the combination of five features (69.95 ± 7.31 %) was only significantly higher compared to the Sp when trained with only one feature (57.90 ± 11.64 %). No significant difference could be obtained with the inclusion of two or more features. Due to the overall better performance of the model with five features, we opted to use it to compare with

the CNN. The resulting average ROC curve of these two classification models can be observed in Figure 6.10.

Table 6.5: Overview of the performances of the SVM models on the test folds. The results are shown as mean±standard deviation. One, Two,... and Five indicate the performance of the SVM model with the inclusion of the highest, second highest,... and five highest ranked features. The overall highest AUC is obtained when the model is trained with the five highest ranked features.

	Per minute	Grained	One	Two	Three	Four	Five
Acc	83.72 ± 2.99	88.76 ± 2.25	83.91 ± 3.84	85.08 ± 3.87	86.29 ± 2.50	86.00 ± 2.23	87.77 ± 2.64
Se	90.76 ± 5.88	93.52 ± 3.54	93.41 ± 5.89	92.90 ± 4.17	93.53 ± 4.30	92.87 ± 3.72	94.09 ± 3.82
Sp	63.73 ± 10.35	75.30 ± 4.45	57.90 ± 11.64	62.99 ± 8.33	65.41 ± 10.18	67.06 ± 9.41	69.95 ± 7.31
AUC	84.26 ± 5.12	92.63 ± 1.59	87.98 ± 5.63	87.33 ± 5.16	90.64 ± 3.57	90.53 ± 3.53	92.77 ± 2.95

The most relevant feature, as defined by the mRMR algorithm, is the amplitude of the first peak of the ACF of the entire segment. This was selected 9/10 times. The feature that is most complementary to the latter is the lower bound of the bandwidth of the entire signal. This was selected 6/9 times. Hereafter, the standard deviation of the amplitude of the first peak of the ACF of the smaller segments was selected 4/6 times. The last two features in the most used feature combination were the average Ap1/Ap2 and the standard deviation of the bandwidth of the smaller segments.

The feature space of the three most complimentary features is depicted in Figure 6.9. It can be observed that the Ap1 of clean signals tends to be higher compared to their noisy counterparts, indicating a more regular signal. Moreover, more variation is present in the noisy signals, since the standard deviation of the Ap1 for the smaller segments is higher for noisy signals. The lower bound of the -3 dB bandwidth is often low for noisy signals. This could indicate the presence of low frequency components with large amplitudes.

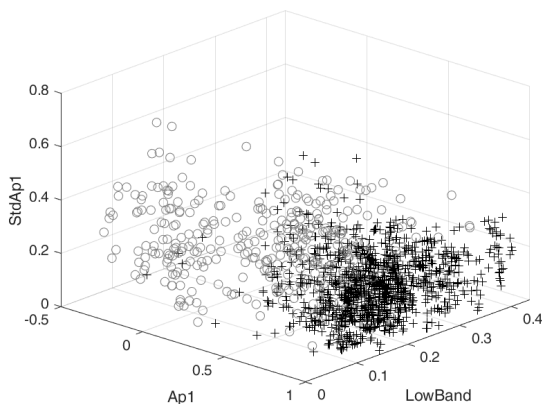


Figure 6.9: Feature space of the most frequently selected combination of three features: The amplitude of the first peak of the ACF of the entire segment (Ap1), the lower boundary of the bandwidth of the entire segment (LowBand) and the standard deviation of the Ap1 for the smaller segments (StdAp1). The gray circles indicate the noisy signals and the black plus signs indicate the clean signals.

The selected approach obtained an Acc of 88.32 ± 2.61 % on the test sets. Compared to the Acc of the classification model with the per segment features, this is not significantly different. Moreover, although the Se (94.69 ± 3.59 %) appears higher and the Sp (70.43 ± 6.48 %) appears lower, we could not observe any significant changes. The resulting receiver operating characteristic (ROC) curve of the latter classification model can be observed in Figure 6.10. The AUC is 93.87 ± 2.12 %, which is not also statistically different from the AUC when the per segment features are used.

The CNN model obtained an Acc of 87.20 ± 2.78 , a Se of 93.25 ± 3.24 and a Sp of 70.40 ± 8.16 . The AUC was 92.51 ± 1.74 . No significant performance differences existed between the two approaches for any of the performance metrics. Figure 6.10 shows the average ROC curves of the feature-based method with five features and the CNN model. It can be observed that they are very similar.

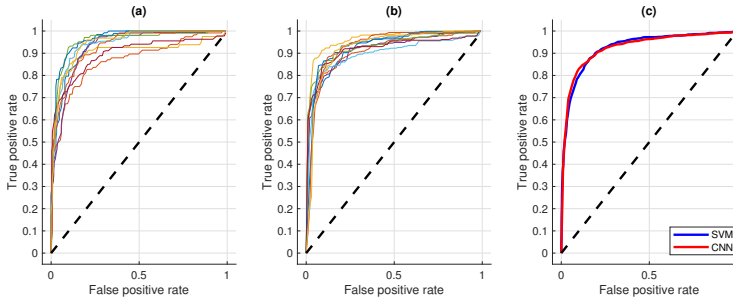


Figure 6.10: Comparison of the ROC curves of the ten folds from the SVM (a) and CNN (b) approach. The average ROC curves of the SVM (blue line) and CNN (red line) approach is depicted in (c). Their respective AUC's are 92.77 ± 2.95 % and 92.51 ± 1.74 %.

6.5 Discussion

The goal of this study was to examine whether machine learning algorithms could be used to separate clean from noisy thoracic BioZ signals. This is of particular interest for patients suffering from CRDs, since they require long time follow-up during daily living conditions. In this study, we compared a feature-based approach and a neural network approach. We have shown that both approaches are able to accurately separate clean and noisy signals.

6.5.1 Features

The first challenge of the feature-based approach was to craft meaningful features. We used two transformation methods, the ACF and PSD, to highlight artefacts in the raw signal. The most relevant feature, as defined by the mRMR algorithm, is the Ap1, the amplitude of the first peak of the ACF of the entire segment. This feature was previously proposed by Moe-Nilssen et al. for the estimation of gait regularity [122]. They indicated two reasons for a low Ap1: low regularity between steps and a systematic asymmetry between left and right steps. For obvious reasons, the latter does not apply on the current problem. In this study, a difference in Ap1 could be observed between signals of different quality. The Ap1 of noisy signals was, in general, lower, compared to that of clean signals. This indicates that the presence of artefacts results in a less regular signal, which can be measured with the Ap1.

Due to the lack of irregular breathing patterns, it is still an open question whether this feature is able to make a distinction between noisy and irregular breathing signals. Future research could tackle this by including more irregular, but clean, respiratory signals, for instance during apnea episodes.

The feature selection experiment has shown that only one feature is sufficient to obtain a high sensitivity and a specificity of above 50 %. Moreover, the results show that the detection of noisy segments can be improved by adding more features to the model, while at the same time the sensitivity remains unaltered. This indicates that expanding the dimension of the feature space is particularly beneficial for the detection of noisy segments.

6.5.2 Classification

We obtained the highest Sp values when all grained features were included to train the feature-based model. However, a large discrepancy remained between the Sp and the Se. This is probably due to the data distribution. The dataset that we used contains 76.00 % clean signals, which prompts the classification model to assign more weight to clean segments. Therefore, a higher number of false positives is obtained, which results in lower Sp values. This problem could be alleviated by adjusting the cost function of the classifier. However, since the class distribution presents an extra challenge for both approaches, we did not adjust the cost function of any of the models.

To the best of our knowledge, this is the first study in which a CNN is used to detect contaminated BioZ respiration signals. On average, the proposed CNN-based method performs similar to the feature-based method. However, in this study, less than 1000 samples were available for training, which is considered a very small dataset for a CNN. We expect that the performance of the CNN could be improved by extending the training database.

Finding discriminative features is a challenge and has a big influence on the performance. For the given problem, some researchers have proposed novel features to enhance the classification performance [121, 36]. The main advantage of the CNN approach is that there is no need to select any features manually. Moreover, since the CNN is an integrated approach, both the features as well as the classifier, are optimized simultaneously.

In this study, we focused on 1-dimensional signals. However, CNN's were originally developed for higher dimensional inputs. Therefore, it could be hypothesised that the performance of the CNN could be improved by providing a higher dimensional input. Recently, Zhang et al. obtained some promising results when using the time-frequency spectrum as input for a CNN for ECG

artefact detection [206]. A similar pipeline could be implemented for the given problem. Since we wanted to investigate the prediction capabilities of the CNN with the raw data as input, we did not investigate this approach here. However, we do intend to investigate this in future research.

6.5.3 Limitations

The current research population is limited to COPD patients. In future research we should expand the population towards patients suffering from other CVDs or, more broadly, all patients that require respiratory follow-up. Due to the quasi-stationary pattern of the respiration, the proposed model should be able to accurately detect artefacts during daily life. However, the effect of influencing factors, such as speaking, eating and drinking, still needs to be investigated.

A related limitation is that the dataset in this study is recorded during rest. Hence, we could state that a whole spectrum of possible artefacts due to motion was not included. In future research, we need to validate the proposed methodologies on a dataset that is recorded during daily life activities.

6.6 Conclusions

In this study, we investigated the use of machine learning algorithms to separate clean from noisy respiratory BioZ signals. We compared two approaches: a feature-based classification model and a CNN.

We have shown that both models accurately distinguish clean from noisy samples and observed no significant differences in performance. This shows that a data-driven approach could be used for the task of artefact detection in respiratory BioZ signals.

Part III

Tools for ECG analysis

Chapter 7

R-DECO

In this Chapter, we present R-DECO, a **MATLAB** based GUI for the detection and correction of R-peaks in ECG signals. The goal of R-DECO is to provide a complete workflow from the raw ECG signal to the tachogram. It includes an accurate R-peak detection algorithm, the performance of which is comparable to literature, and allows the user to graphically correct wrong or missing R-peaks. Additionally, R-DECO supports a variety of ECG input file formats, which allows the processing of recordings from multiple recording devices. This makes it a tool that can be used both by engineers, and clinicians.

The tool was published in **MOEYERSONS, J.**, AMONI, M., VAN HUFFEL, S., WILLEMS, R., AND VARON, C. R-DECO: an open-source Matlab based graphical user interface for the detection and correction of R-peaks. *PeerJ Computer Science* 5, e226, (2019), <https://doi.org/10.7717/peerj-cs.226>. and later also on Physionet as **MOEYERSONS, J.**, AMONI, M., VAN HUFFEL, S., WILLEMS, R., AND VARON, C. R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks (version 1.0.0). *PhysioNet*, (2020), <https://doi.org/10.13026/x6j7-sp58>.

The software is freely available at:

- http://homes.esat.kuleuven.be/jmoeyers/R_DECO.zip or
- <https://physionet.org/content/r-deco/1.0.0/>.

7.1 Introduction

A crucial step in the study of the ECG is the location of the QRS-complexes. These complexes are the most prominent waveforms in the ECG, and contain an enormous amount of information about the state of the heart. This is why the detection of the QRS-complexes constitutes the basis for almost all automated ECG analysis algorithms [89]. Once these have been identified, more elaborated analyses can be performed, such as HRV.

Four decades of automated QRS detection research has resulted in a variety of methods using different approaches. These methods can be stratified based on derivatives, digital filters, wavelet-transforms, classifiers, etc [146, 56, 65, 170, 39]. Despite the wide methodological variety, most of these QRS detectors have the same algorithmic structure. This can be divided in two steps: pre-processing and decision making [89].

In the pre-processing step the QRS-complex is highlighted and the other signal components are suppressed to facilitate the detection. The resulting signal is then used to detect the occurrence of QRS-complexes in the decision making step. This is done by using either fixed or adaptive thresholds. Despite high detection rates, some QRS-complexes remain undetected. Reasons for this might be small amplitudes, wide complexes or contamination by noise [7]. Therefore, in many algorithms an extra post-processing step is added for the exact determination of the temporal location of the detected QRS-complex.

One of the most established QRS detection algorithms is the Pan-Tompkins algorithm [146]. Although it was developed in the eighties, it achieves comparable performance to many more elaborate algorithms [61]. In this study, an envelope-based procedure that enhances the QRS-complexes and flattens the rest of the ECG is used in combination with an adapted version of the threshold-based approach of the Pan-Tompkins algorithm. This method, which was proposed by our group in [191], combines the simplicity of an envelope-based procedure, while maintaining the accuracy of many more elaborate methods.

In a review paper, Elgendi et al. have compared the results of 22 beat detection algorithms on the MIT-BIH arrhythmia database [61]. When comparing the results of the automated algorithms with expert annotations, they have shown that many algorithms obtained excellent accuracy. However, none of the algorithms reached perfection. This means that, no matter how good the QRS detection algorithm is, it is highly likely that not all annotations are correct. Therefore, it is recommended to visually inspect and review each signal before further analysis [149].

Many of the existing ECG toolboxes have focused on the derivation of HRV-analysis parameters from RR-intervals, the time between subsequent R-peaks. This makes sense, since most of the available hardware include some kind of QRS-complex detection algorithm. However, this does not necessarily mean that the output of these devices are the raw RR-intervals. Many of these devices have a built-in post-processing algorithm, which compensates for false detections by averaging over a certain range of RR-intervals [139, 149, 195]. However, for some analyses, such as EDR or BVR, it is of utmost importance that the actual R-peak of the QRS-complex is detected. Therefore, it is necessary to visualize the actual R-peak positions in the ECG signal and allow the possibility to make manual adaptations.

In this Chapter, we present R-DECO, a MATLAB based GUI for the detection and correction of R-peaks. This user interface includes the developed R-peak detection algorithm and provides the user with the possibility to correct possible false detections in a very straightforward way.

7.2 Computational methods

7.2.1 R-peak detection

We developed an R-peak detection algorithm that is based on an enveloping procedure. It achieved a 99.60% sensitivity and 99.69% positive predictive value on the MIT/BIH Arrhythmia Database [133]. The algorithm can be divided in three steps: pre-processing, decision and post-processing.

Pre-processing

In the presented approach, the filtering, derivative, squaring and integration steps of the Pan-Tompkins algorithm are replaced by one flattening step, which enhances the QRS-complexes and flattens the rest of the ECG [191]. A visual explanation of the method is shown in Figure 7.1.

First, the upper (U) and lower (L) envelopes are computed from the ECG signal by the secant method. This method selects the segment with the steepest positive and negative slope in a user-defined window with length t . Once U and L are obtained, they are used to derive a flattened version of the ECG signal (F): $F = U - L$. Since L is subtracted from U , the baseline is eliminated and only a positive signal, F , remains.

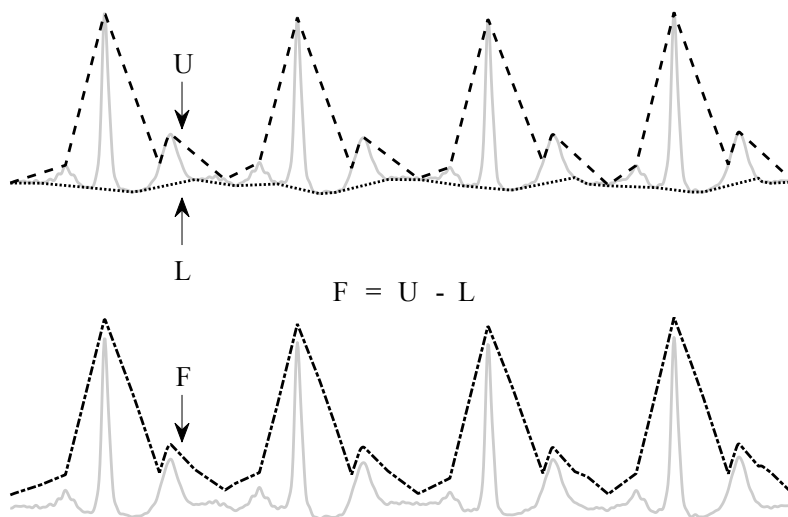


Figure 7.1: Flattening procedure. The flattened ECG (F) is constructed by subtracting the lower envelope (L) from the upper envelope (U). This enhances the QRS-complex and flattens the rest of the ECG signal.

Decision

The locations of the QRS-complexes are found by detecting the peaks in the flattened ECG. These peaks are detected in three stages. First, all samples with an amplitude lower than the amplitude of the sample 80 ms further are selected. The 80 ms step size was experimentally defined. This results in the selection of the upward slopes. As a second step, only the upward slopes that are longer than the step size are selected in order to exclude small peaks. Finally, the maximum is selected in a window, with a length equal to the step size, that starts from the last selected sample of the upward slope. A graphical representation of this process is shown in Figure 7.2.

On this selection of peaks, the adaptive thresholding procedure of the Pan-Tompkins algorithm is applied to define the peaks that correspond to the QRS-complexes.

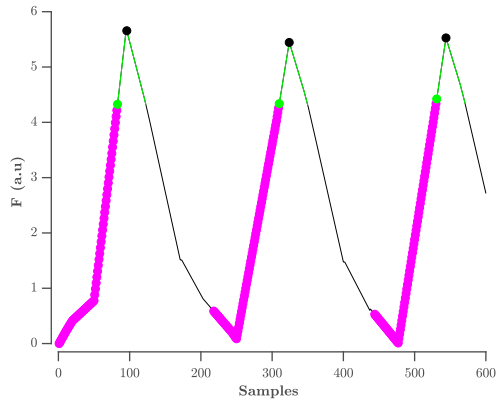


Figure 7.2: Procedure to select R-peaks. The resulting flat ECG denoted F is indicated by the black line. The samples with an amplitude lower than the sample 80 ms further are indicated by the magenta circles, with the last sample indicated by the green circle. The search window is indicated by the green line. The selected R-peaks are indicated by the black circles. A.u. stands for arbitrary units.

Post-processing

The Pan-Tompkins algorithm generally produces satisfactory results in the detection of the QRS-complexes. However, some of the automatically generated RR-intervals might be physiologically unreasonable and need to be removed for further analysis. A slightly modified version of the search-back procedure as proposed in [50] was used for this purpose.

Once the positions of the QRS-complexes are identified, the original ECG is used to find the exact location of the R-peaks. The search for an R-peak is performed up to 50 ms from the peak in the flattened signal. This extra search is necessary, because the presence of large S-waves might shift the peak in the flattened signal towards the valley of the S-wave.

Evaluation on the Physionet MIT/BIH arrhythmia database

We used the MIT/BIH arrhythmia database to evaluate the proposed algorithm [133]. This dataset consists of 48 half-hour ECG signals, which were recorded in the Boston's Beth Israel Hospital between 1975 and 1979. All recordings were annotated by two independent cardiologists who also made a distinction between

normal and abnormal beats. In total, 110 122 heartbeats were annotated, of which 89 133 were labelled as normal. Each recording contains two channel ECG signals with a sampling frequency of 360 Hz. In most records, one channel is lead II and the other channel is V1. However, we only used the first channel for the evaluation.

As mentioned previously, the pre-processing consists of a flattening step of the ECG with a user-defined window width. To evaluate the sensitivity of the performance to the choice of the width we have tested multiple window widths. As can be observed from Figure 7.3, comparable results were obtained for window widths between 250 and 350 ms.

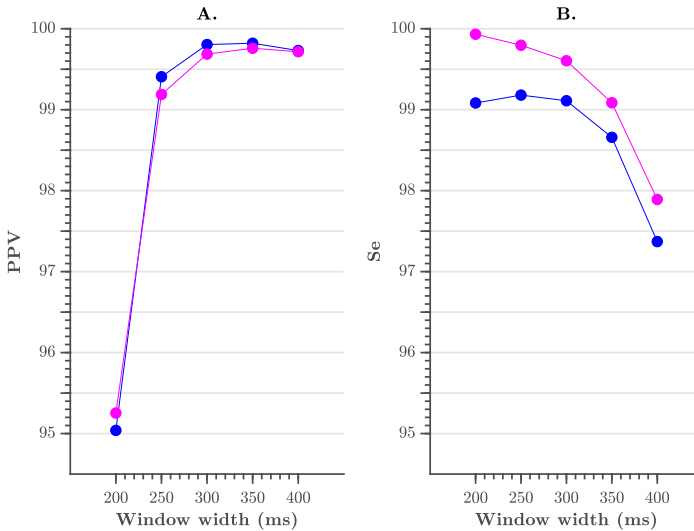


Figure 7.3: **A.** PPV of the performance to the choice of window width. **B.** PPV of the performance to the choice of window width. Blue: with post-processing, Magenta: without post-processing. A window width between 250 and 350 ms results in the best performance.

In Table 7.1 we listed the R-peak detection results of the proposed algorithm. With an envelope width of 300 ms and without post-processing, we obtained an overall Se of 99.60% and PPV of 99.69%. With the post-processing, we obtain an overall Se of 99.09% and PPV of 99.80%. These results are comparable with those in literature, especially with the Pan-Tompkins algorithm, which reaches a Se of 99.76 % and a PPV of 99.56% [61].

Table 7.1: Performance of the R-peak detection algorithm on the Physionet MIT/BIH dataset.

Record	Total (beats)	TP (beats)	FP (beats)	FN (beats)	Se (%)	PPV (%)
100	2273	2273	0	0	100	100
101	1865	1864	4	1	99.95	99.79
102	2187	2187	0	0	100	100
103	2084	2084	0	0	100	100
104	2229	2227	3	2	99.91	99.87
105	2572	2542	41	30	98.83	98.41
106	2027	2023	2	4	99.80	99.90
107	2137	2130	0	7	99.67	100
108	1763	1739	80	24	98.64	95.60
109	2532	2532	0	0	100	100
111	2124	2123	2	1	99.95	99.91
112	2539	2539	0	0	100	100
113	1795	1795	0	0	100	100
114	1879	1877	6	2	99.89	99.68
115	1953	1953	0	0	100	100
116	2412	2388	2	24	99	99.92
117	1535	1535	0	0	100	100
118	2278	2278	1	0	100	99.96
119	1987	1987	2	0	100	99.90
121	1863	1861	12	2	99.89	99.36
122	2476	2476	0	0	100	100
123	1518	1518	0	0	100	100
124	1619	1619	0	0	100	100
200	2601	2595	8	6	99.77	99.69
201	1963	1958	0	5	99.75	100
202	2136	2120	14	16	99.25	99.34
203	2980	2749	20	231	92.25	99.28
205	2656	2641	2	15	99.44	99.92
207	1860	1855	8	5	99.73	99.58
208	2955	2941	2	14	99.53	99.93
209	3005	3005	0	0	100	100
210	2650	2582	3	68	97.43	99.88
212	2748	2748	1	0	100	99.96
213	3251	3250	0	1	99.97	100
214	2262	2259	3	3	99.87	99.87
215	3363	3354	0	9	99.73	100

Continued on next page

Table 7.1 – continued from previous page

Record	Total (beats)	TP (beats)	FP (beats)	FN (beats)	Se (%)	PPV (%)
217	2208	2202	0	6	99.73	100
219	2154	2154	1	0	100	99.95
220	2048	2047	0	1	99.95	100
221	2427	2425	2	2	99.92	99.92
222	2483	2475	21	8	99.68	99.16
223	2605	2605	0	0	100	100
228	2053	2044	59	9	99.56	97.19
230	2256	2256	0	0	100	100
231	1571	1571	0	0	100	100
232	1780	1780	17	0	100	99.05
233	3079	3070	0	9	99.71	100
234	2753	2753	1	0	100	99.96
Total	109 494	108 989	317	505	99.60	99.69

While these results are very promising, we can also observe that for some recordings only moderate detection results are obtained. This decrease in performance is generally due to poor signal quality, unusual morphology or stretches of extremely irregular rhythms. For instance, recording 116 and 208 contain stretches where the signal is lost in the first channel. However, the recordings with the highest amount of false detections are 108, 203 and 210. Correctly detecting the R-peaks in recording 108 has been proven difficult for many algorithms [146]. It contains a lot of baseline wander and additionally, very tall and sharp P-waves. These characteristics make it difficult to distinguish P-waves from R-peaks and thus result in a high false positive count. However, the highest amount of false positives is observed in recording 203 (92.25% sensitivity). This might be explained by the extremely high percentage of premature ventricular contractions (PVC) present in the recording, almost 15%. Since the envelope width was fixed during the detection process, one may assume that the performance could be improved if manual adjustments were permitted. This holds as well for other records with PVC's, such as record 210.

The noise tolerance of the algorithm was evaluated with the NSTD Database that was also used in Chapter 3 [134]. We observed that both median PPV and Se remained around 100% above an SNR of 6 dB. From this threshold the performance of the algorithm decreased significantly.

From the analysis of the results, we could deduce two main factors that influence the results of the algorithm: (1) the envelope width and (2) the RR-post

processing. The number of samples in the envelope is important, since it can be regarded as a filter of the RR-intervals. Smaller envelope widths might result in the enhancement of more peaks than only the R-peaks. This might be beneficial in the case of small R-peaks, but might also enhance artefact peaks. Larger envelope widths might cause adjacent R-peaks to be merged in the flattened signal. In practice, this might result in the failure of detecting premature heartbeats, which appear shortly after the previous heartbeat. In summary, a larger envelope width results in less false positives and more false negatives and the opposite is true for a small envelope width. A similar effect can be observed when the RR-intervals are post-processed. This increases the certainty of detection of the algorithm and thus results in less false positives. The downside is that, in the presence of abnormal rhythms, it also results in more false negatives.

7.3 Software description

The algorithms have been implemented with `MatlabR2018a` (The Mathworks, Natick, MA, USA). We used GUIDE, MATLAB's GUI development environment, to design the GUI of R-DECO. The current subsection describes the possible input data formats and the user interface.

7.3.1 Input data formats

The standard input of the toolbox is raw or filtered ECG data. This can be both single- or multichannel ECG. Since a plethora of open formats exist for storing the ECG, it would be impossible to write supporting software for all formats [139]. Therefore, we focussed on the data formats that are most commonly used by our clinical partners in the cardiology department of the UZ Leuven, Belgium. The following file formats are supported:

- ISHNE-Holter files (*.ecg)
- MATLAB files (*.mat)
- European Data Format (*.edf)
- Text files (*.txt)
- Excel files (*.xls or *.csv)

An International Society for Holter and Noninvasive Electrocardiology (ISHNE)-Holter file is organized in a header record, followed by a data block that contains all digital ECG samples. This file format was developed to facilitate data exchange and research in the field of Holter [11]. The software automatically extracts all ECG channels and also the sampling frequency.

A `MATLAB` formatted file can contain one variable, up to an entire workspace. Therefore, if the file contains more than one variable, the user is prompted to select the variable containing the ECG signal. In the specific case that the selected file is a structure, the software allows to search within the structure until the ECG signal is selected.

A standard European Data Format, EDF, file consists of a header record and the data records [86]. It was originally intended for the digital storage and exchange of EEG and polysomnogram recordings, but currently it can store a variety of annotations and signals, such as EMG, ECG and many more [85]. Since not all EDF files have the same standard labels, the user is prompted to identify the ECG channel(s). Additionally, the software attempts to identify the sampling frequency of the selected signal by scanning the file.

As an extra feature, the software allows the user to load a session. When the current session is interrupted, the session can be saved as a `MATLAB` file. It includes all the analysis parameters, the ECG signal and the RR-intervals, if computed. When a previous session is loaded, the software restores the entire user interface to the moment on which the session was saved. This allows the user to pause and continue, whenever wanted.

7.3.2 User interface

The strength of this toolbox is that everything is operated through a single GUI. As shown in Figure 7.4, it can be divided in five segments: Data, Filter, Analysis Period, R-peak Detection and R-peak Correction. All segments are described below.

Data

In the data panel, the user has the option to load data in two different ways: from a file or from the `MATLAB` workspace. Both can be accessed via the respective pushbuttons.

When a file is selected, the software visualizes a small segment of the ECG signal. If the signal is inverted, the user can indicate this and the file will be

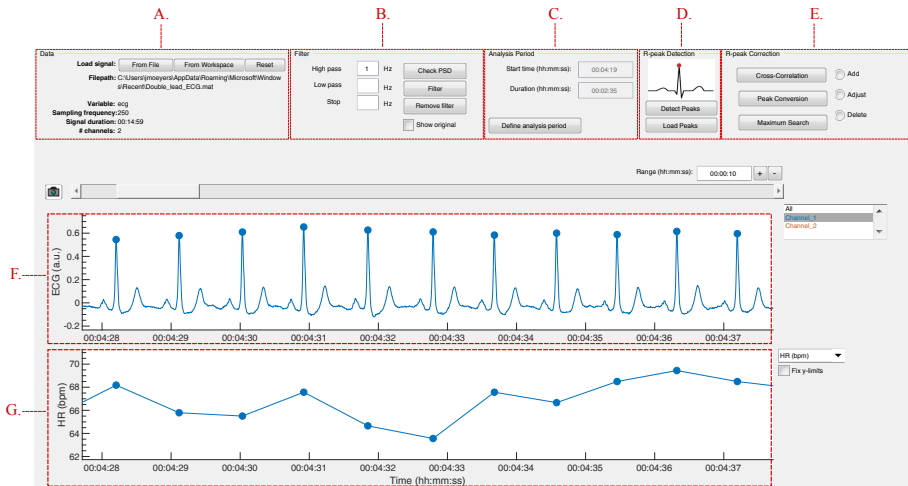


Figure 7.4: The graphical user interface of R-DECO. The user interface can be divided in five segments: A. Data, B. Filter, C. Analysis Period, D. R-peak Detection and E. R-peak Correction. The ECG signal and the resulting tachogram are shown in respectively F. and G. The detected R-peaks are depicted as small blue circles.

inverted before further analysis. The software also scans the selected file for the sampling frequency. If this is not found, the software prompts the user to manually indicate the sampling frequency.

Finally, the data panel also contains a reset button. This button allows the user to reset the entire GUI. It empties all plots, restores all default variables and deletes all results. If the user has not yet saved the current analysis, the user is prompted to confirm the reset action to prevent unwanted loss of information.

Filter

Since ECG signals can be contaminated with noise, filtering is often essential for further analysis. R-DECO provides three basic filters: high pass, low pass and a notch filter.

The high pass filter consists of a zero phase, second order Butterworth filter. The low pass filter consists of a zero phase, fourth order Butterworth filter. Finally, also a zero phase notch filter is included. The latter could be used to

remove the power-line interference. Important to note is that filtering actions are always executed on the original signal to ensure repeatability.

In order to aid the user in the selection of appropriate frequency threshold(s), R-DECO is able to compute and display the power spectrum. An estimate of the power spectrum is computed using the Welch method [201]. As a default, we used a window of 500 samples with 60% overlap.

To visualize the effect of the filtering in the frequency domain, R-DECO displays both the filtered and the original power spectrum. Furthermore, the effect of the filtering in the time domain can also be investigated by checking the “Show Original” checkbox. This will overlay the original signal on top of the filtered signal (Figure 7.5).

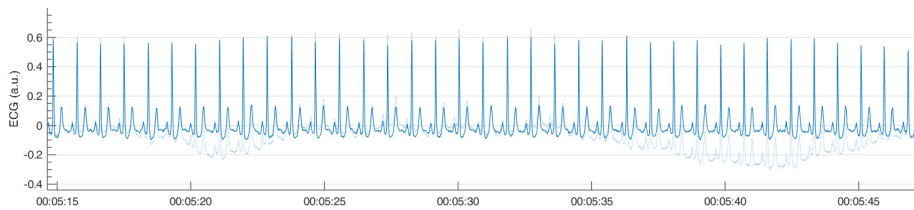


Figure 7.5: Example of a high pass filter. By clicking the "Show Original" checkbox, the original signal (light blue) is overlaid on the filtered signal (blue).

Analysis period

In the Analysis Period panel the user has the possibility to define an analysis window. After pushing the “Define analysis period” button, the user has to select a window by clicking, dragging and releasing the mouse. The window is shown as a transparent patch over the data and can be enlarged, shrunk or moved with the mouse. After the initial window is drawn, the user can further modify the analysis window by changing the start time and/or the duration of the window.

When a window is selected, the user can accept the analysis window by pressing the "Apply changes" button. This prompts the x-limits of the graph to match the analysis window and disables the window modifications. From here on, all further analysis will be performed solely on the selected window.

All the above is very useful when the to-be-selected time period is known in advance, however this is not always the case. Sometimes the selection of the

analysis window is depending on certain patterns in the tachogram, hence the tachogram has to be constructed first. Therefore, if R-peaks have been detected already, the user is prompted to indicate whether he/she would like to keep the detected R-peaks.

R-peak Detection

The execution of the R-peak detection algorithm, as described in the first section, is initiated when the “Detect Peaks” button is pushed. However, before the actual algorithm is executed, the user is able to adjust the default parameters of the algorithm.

In Figure 7.6, an epoch of 10 seconds of the ECG signal and its respective enveloped signal is shown as an example of the flattening procedure. The user can adjust the envelope size to the desired value and enact the changes by pressing the "Apply" button. Additionally, an estimation of the average heart rate can be defined to compute the boundaries of normal-to-normal RR-intervals. Lastly, the user can select the automatic post-processing of the RR-intervals.

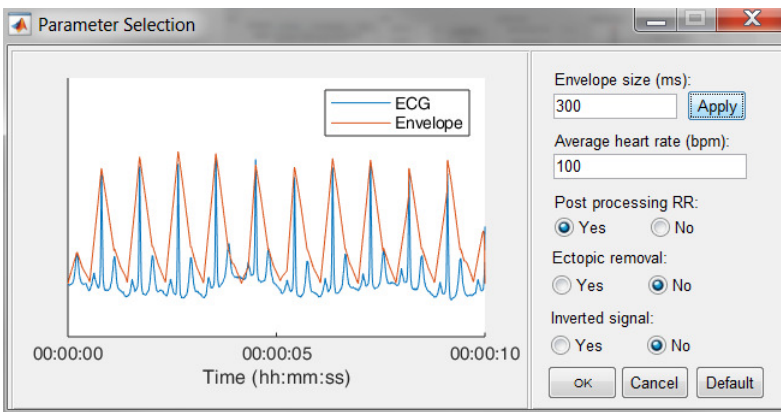


Figure 7.6: The R-peak parameter selection window. The user can adjust the envelope size, the average heart rate and indicate if RR post-processing is necessary. Pressing the Default button restores the default values.

Since some devices have built-in QRS detection algorithms and some researchers have their own preferred QRS detection algorithm, the software allows to load R-peak locations. These will be displayed the same way the R-peaks of the algorithm are displayed.

R-peak Correction

In case of heart rate variability studies, only the normal-to-normal RR-intervals need to be taken into account. This can be achieved by selecting the ectopic removal option. This option corrects ectopic beats, without altering the normal RR-intervals.

After finishing the detection process, either by detecting or loading the R-peaks, it is still possible that not all R-peaks are accurately detected. In R-DECO, the user can make manual and (semi-)automatic adjustments to the R-peak locations.

The manual methods are: add, adjust and delete. These methods can be activated by selecting the specific radiobuttons or via a context menu, which is linked to each R-peak. These manual methods allow the user to correct wrong or missing annotations either in all or in individual leads.

- Add: When this radiobutton is active, new R-peaks can be added by clicking in the ECG graph. The program selects the signal sample that is closest to the mouse position in a symmetric window of 300 ms around the mouse position. Upon mouse release, a new R-peak is added and the tachogram is adapted.
- Adjust: While hovering over the ECG graph, the R-peak closest to the mouse location is selected. After clicking on the desired R-peak, it can be moved by dragging the mouse. The movement of the selected R-peak is restricted by the previous and next R-peak. While adjusting the R-peak location, the tachogram is automatically updated. Upon mouse release, the new R-peak location is saved.
- Delete: While hovering over the ECG graph, the R-peak closest to the mouse location is selected. After clicking on the desired R-peak, more to-be-deleted R-peaks can be selected by dragging the mouse. Upon mouse release, the selected R-peaks are removed and the tachogram is adapted.

The three (semi-)automatic R-peak correction methods are: cross-correlation, peak conversion and maximum search.

- Cross correlation: For this method, a symmetrical window of 300 ms around each R-peak is selected. Then, all heartbeats are normalized by subtracting the mean and dividing it by the standard deviation. Then, a trimmed average QRS-complex is computed of all the positive and 'negative' R-peaks. In this work a 'negative' R-peak is understood as the

absence of an R-peak or the presence of a very prominent S-wave, also described as RS-complex.

The user is prompted to select either the positive or the 'negative' average heartbeat. If necessary, the user can also adjust the location of the R-peak on the selected template. This is all graphically displayed as shown in Figure 7.7. Finally, the cross-correlation of every heartbeat is computed with the trimmed average and the R-peak is re-located, based on the highest correlation value.

- **Peak conversion:** The absolute amplitude of every R-peak annotation is compared with the previous and next sample's absolute amplitude. If it is bigger than the previous and smaller than the next, the location of the R-peak will be shifted forward, until an extremum is obtained. If it is the other way around, the location of the R-peak will be shifted backwards.

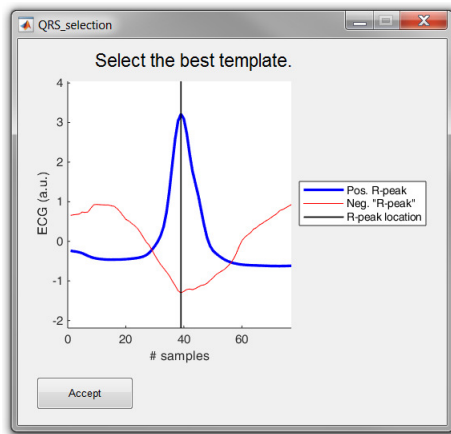


Figure 7.7: The template selection window. The user can select either the positive or 'negative' R-peak template and can shift the location of the R-peak if necessary.

This functionality avoids the 'jumping' of R-peaks from e.g. an actual R-peak to a pre-mature ventricular contraction, which might be the case when window search is applied. Furthermore, pressing the button more than once will not affect the relocation after one correction.

- **Maximum search:** Firstly, a symmetric window of 300 ms around each R-peak is selected. Secondly, the user is prompted to select either the maximum, minimum or absolute maximum. Based on this selection, the respective extremum within the window is selected as new R-peak location.

7.3.3 Save and export results

By default, the results are saved as a **MATLAB** file. This file includes the R-peak locations and the RR-intervals, which can further be used for HRV-, BVR- or any other analysis that requires R-peak detection. Additionally, the software can export the results in two different ways: 1) an Excel file and 2) a **MATLAB** file.

1. Excel file (*.xls): A new workbook is created with on the first sheet a general overview of the file: the number of channels, the sampling frequency, the duration of the signal and the duration of the analysis period. The number of additional sheets is defined by the number of channels, since for every channel, a separate sheet is created. This contains the R-peak locations, the RR-intervals and a number of basic metrics, such as the mean heart rate.
2. **MATLAB** file (*.mat): This file contains a single structure named *data*. In accordance to the structure of the Excel file, a structure is created per channel. This contains the signal in the analysis window, the R-peak locations and the RR-intervals. This option is especially useful for further analysis in **MATLAB**.

7.3.4 Data browser

In order to graphically correct the R-peaks it is important to have a clear view of the segment to be investigated. Therefore, after the R-peaks are detected, the software immediately displays the ECG signal with the R-peak annotations and the respective tachogram.

The window width of the x-axis can be adjusted in three different ways: 1) the range edit box, 2) the plus and minus buttons or 3) by using the zoom button. All three methods also adjust the width of the scroll bar.

The scroll bar can be used to slide through the signal. Since both axes are linked, both slide at the same time. The y-axis range in both axes is adjusted automatically according to the data within the selected range. However, if the “Fix y-limits” checkbox is selected, the range of the y-axis of the tachogram is fixed to the current limits. Since some users favour a tachogram that displays the RR-intervals, while others favour HR values, we made it possible to switch between the two.

An ECG recording with multiple channels might result in axes that become unclear. Therefore, R-DECO provides the user to switch view between different

channels. This way the user can select one, multiple or all channels. If the channel labels are not present in the signal file, R-DECO names and numbers the channels itself: Channel 1, Channel 2, etc. However, it also provides the possibility to change these names. When the user clicks twice on a channel name in the listbox, a dialog box pops up that allows the user to choose a new name. This will be adjusted in the listbox, and also in the output files.

An extra feature is the possibility to take 'pictures' of the axes. This saves both axes in a format of choice, without including any of the buttons or bars from the user interface. In order to design the axes to the user's taste, R-DECO allows to change the line colors and the grid lines.



Figure 7.8: The data browsing options of R-DECO.

7.3.5 Preferences

The analysis settings of R-DECO can be adjusted via the Preferences menu. Note that the changes only apply for the current session and are not saved for the next session. This will be adjusted in a future release.

The Preference menu can be divided in four segments: Power Spectrum, Filter, Detection and Correction

- Power Spectrum: All variables of the Welch method can be adjusted here.
- Filter: The type and order of the filter can be defined here.

- Detection: The default input parameters for the R-peak detection algorithm can be adjusted here. Whenever the default button is pressed in the parameter selection window, see Figure 7.6, all parameters are reset to the values defined in this segment.
- Correction: For all correction methods, the user can define the window size in which the new R-peak is supposed to be located.

7.4 Sample run

As a sample run, we used a 24 hour digital Holter signal that was recorded from a male subject with ischaemic heart disease. The idea was to investigate the temporal evolution in beat-to-beat variability of repolarization before spontaneous non-sustained ventricular tachycardia (nsVT). Before analysis could be performed, the nsVT episodes needed to be identified and the R-peaks needed to be detected.

Once the signal was loaded and the sampling frequency was defined, we first had a look at the power spectrum. According to the plot, no power line interference was present, since we could not observe a peak at 50 or 60 Hz. However, it was clear that most of the power was situated in the lower frequency bands. This could indicate the presence of baseline wander. Therefore, we high-pass filtered the signal with a cut-off frequency of 0.66 Hz.

Next, the nsVT episodes needed to be identified. However, the time stamps of the episodes were not known. Therefore, it was necessary to detect the R-peaks first. This way the nsVT episodes could be identified from the tachogram. An example of an nsVT episode, taken with the picture button, can be observed in Figure 7.9.

Based on the example signal, we selected an envelope size of 300 ms, which provided the best results for this signal. This envelope size ensures the enhancement of the QRS-complexes, without skipping any beats. Additionally, we indicated that no post-processing of the RR-intervals is wanted, since we wanted to be able to detect nsVT segments as well.

The nsVT episodes were identified based on the resulting RR-intervals. From the start of one of these episodes, we selected 30 consecutive heartbeats [179]. Only normal-to-normal intervals should be taken into account for BVR-analysis. Hence, (ventricular) ectopic and post-extrasystolic beats were removed for further analysis, as can be observed in Figure 7.10.

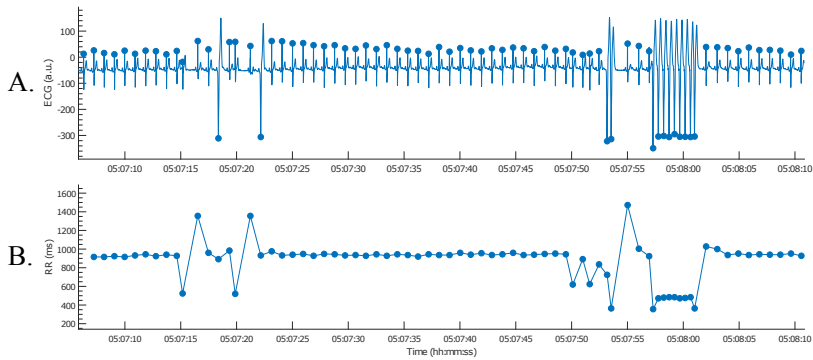


Figure 7.9: Example of an nsVT segment without correction. The resulting tachogram is shown in B.

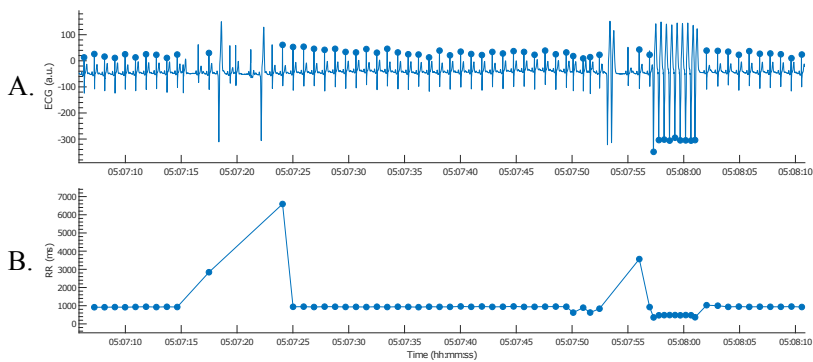


Figure 7.10: Example of an nsVT segment with correction. The resulting tachogram is shown in B.

After the RR-intervals in the wanted analysis window were selected, the analysis results were saved. This was done by selecting "Save Results" on the menu bar and entering a file name. The results are then saved as a MATLAB file and can be loaded for the following BVR-analysis. Note that you can also export the results as an Excel file.

7.5 Potential of future growth

R-DECO is the first step towards a complete ECG processing tool. At the moment, it focusses on accurate R-peak detection and intuitive correction

options. Therefore, it is a complementary tool for existing HRV-analysis toolboxes, which tend to focus on the computation of HRV metrics from RR-intervals. Although some toolboxes already provide the possibility to detect R-peaks, their possibility to correct R-peak annotations is rather limited [149, 156, 195]. Additionally, to the best of our knowledge, none of the existing toolboxes provide filtering, R-peak detection and correction all together.

The main advantage of R-DECO is the easy-to-use, intuitive GUI. All actions are performed in one window, which simplifies the use and reduces the learning time.

Several extra features are being developed and will be released in future versions. We intend to add support for other input file formats, such as Hierarchical Data Format 5 (HDF5) files, General Data Format (GDF) files, etc. However, most improvements will be in the amount of analysis options. Some of the first extra analysis options will be automatic signal quality detection, EDR and HRV-analysis.

7.6 Conclusion

R-DECO is a MATLAB based GUI for detecting and correcting R-peaks in ECG signals. The goal of R-DECO is to provide a complete workflow from the raw signal to the tachogram. It includes an accurate R-peak detection algorithm, the performance of which is comparable to literature, and allows the user to graphically correct wrong or missing detections. Additionally, R-DECO supports a variety of ECG input file formats, which allows the processing of recordings directly from the recording device. This makes it a tool that can be used both by engineers, and clinicians.

We included some basic pre-processing options, such as three filters and the possibility to select an analysis window. The analysis results can be exported to the MATLAB workspace or Excel for later analysis.

Chapter 8

Beat-to-beat variability of repolarization

Beat-to-beat variability of repolarization (BVR) is an ECG-based biomarker that captures the repolarization instability. It has shown to be superior to standard metrics in the screening of patients with heart failure. A limitation of the analysis of BVR is the requirement of separate software and dedicated personnel. In this chapter we present an extension to R-DECO for BVR analysis.

We used this tool to investigate the temporal evolution in BVR before spontaneous non-sustained ventricular tachycardia (nsVT). The results of this analysis are published as **MOEYERSONS, J.**, AMONI, M., VANDENBERK, B., VARON, C., SIPIDO, K., HUFFEL, S., AND WILLEMS, R. Temporal beat-to-beat variability of repolarization changes predict non-sustained ventricular tachycardia in ischemic heart disease patients. *Computing in Cardiology 45*, Maastricht, The Netherlands, (Sep 23-26, 2018).

8.1 Introduction

Ischaemic heart disease (IHD) is the leading cause of mortality worldwide, accounting for 11.2% of all deaths globally in 2011 [141]. A significant portion of these deaths are caused by ventricular tachycardia's (VT) [106].

Currently, risk stratification focusses on identifying patients eligible for an implantable cardioverter-defibrillator (ICD) [54]. However, this stratification is almost solely based on left ventricular function assessment and lacks sensitivity and specificity. Non-invasive, ECG-based, biomarkers may improve the management and outcome of IHD patients by better risk stratification and allowing preventive care of imminent VT.

The prolongation of myocardial repolarization, quantified as the QT-interval, has been used broadly as a non-invasive biomarker to predict mortality. However, there are several clinical and experimental studies that question the predictive value of a prolonged QT-interval. Already in 1997, Gilmour et al. concluded that the repolarization prolongation itself was not sufficient for the initiation of ventricular arrhythmias [68]. Therefore, in an attempt to address this issue, numerous biomarkers have been proposed to improve the predictive value with respect to the prolongation of repolarization.

BVR is a relatively new ECG-based biomarker that captures the repolarization instability [179]. It has shown to improve the diagnostic screening of patients with documented drug-induced proarrhythmia, heart failure and congenital long-QT syndrome [189]. Furthermore, it seems to be superior to the QT-interval in terms of identifying patients at risk for ventricular arrhythmias [189].

One of the current limitations of the analysis of BVR is the requirement of separate software and dedicated personnel [189]. We tried to overcome this limitation by developing an easy-to-use graphical user interface for semi-automated analysis of the ECG signals. The main objectives can be summarized as follows: select accurate delineation algorithms, implement these algorithms in a user interface and create a readable output format. The outline of this tool is depicted in Section 8.2.

In Section 8.3, we investigated the temporal evolution in BVR before spontaneous non-sustained ventricular tachycardia (nsVT) in patients with IHD. The newly developed tool was used to derive the BVR values.

8.2 BVR-tool

8.2.1 T-wave end detection

The great morphological variation in ECG signals makes it hard to design an automated and widely applicable algorithm for T-wave end detection [110]. Whereas the QRS-complex is easily detected, because of its sharpness, it can be quite challenging to determine the end of the T-wave, since it gradually merges with the baseline [45]. Furthermore, the presence of U-waves might cause additional difficulties. Large U-waves, fused with the T-wave, should be included in the measurement, in contrast to small and/or separate U-waves which should not be included. During exercise, the problems with T-wave end detection are even more distinct, since at fast heart rates, the T-wave might fuse with the following P-wave [37]. These facts make it difficult to automatically detect the end of the T-wave.

Despite all these challenges, several algorithms have been developed using different methodologies. Since these different algorithms can differ in T-wave end detection by 10 to 20 ms it is important that a correct detection method is selected for our tool [148]. Therefore, in [123], we compared four (semi-)automated methods.

The first method is a semi-automatic method, based on a template matching algorithm in which the user manually selects, hence the semi, the beginning and the end of the QT interval on a template beat [18]. The second method determines the end of the T-wave using the tangent of the steepest point of the descending limb of the T-wave. The third and fourth method perform a maximum area search of, respectively, a trapezium [196] and the area under the curve, referring to the T-wave [207]. The annotations of each method were compared to the manually annotated T-wave ends of the PhysioNet QT database [94].

This database was designed to serve as a reference for the validation and comparison of automatic algorithms for QT measurement. Despite recent critique, it is still considered the benchmark dataset for ECG delineation algorithms [72]. It contains 105 records of 15 min two-lead ECG signals, sampled at 250 Hz, and a total of 3944 annotated beats. The records consist of a wide variety of T wave morphologies, so that they would be useful to evaluate the performance of automatic methods. Two expert cardiologists manually annotated the signals. The first cardiologist annotated 3542 heartbeats and the second only 402. At least 30 beats per record were annotated. For each annotated beat, the following fiducial points were annotated: the beginning, peak, and end of the P-wave; the beginning and end of the QRS-complex (the

QRS-peak was given by a QRS-detector); the peak and the end of the T-wave.

In order to evaluate the accuracy and repeatability of the proposed algorithms, the average and standard deviation of the detection errors, that is the time difference in ms between the manually and automatically detected T-wave ends, were computed for the four methods in the two ECG leads. Since the dataset consists of two leads per recording, Martinez et al. proposed to apply the algorithm on both leads and select the a posteriori best result per heartbeat [112]. However, in clinical practice it is not possible to constantly switch leads. Hence, we adapted the approach of Jané et al. who selected the best lead, based on the previous approach, and computed the performance per lead [80].

We have found that the fourth method, the one with the area under the curve, performs best in terms of average accuracy (3.9 ms), but that the first method has the lowest standard deviation (17.6 ms). Hence, we included both methods in the BVR tool. The outline of these methods is described in the following sections.

Semi-automatic method [18]

The first step of this algorithm is to create a representative heartbeat. In order to do so, we select 0.35 s before and 0.75 s after each R-peak and store each heartbeat in a matrix. From these beats, we exclude the samples that have the upper and lower 10 % percent amplitudes of each time point. However, this is only meaningful if all heartbeats are perfectly aligned. Hence, the R-peaks need to be accurately annotated beforehand.

Hereafter, the user manually annotates the end of the T-wave on the template. Given that the time location of the R-peak, $R(k)$, is known, we can construct a segment $\phi(n)$ between the R-peak and the T-wave end annotation, where n stands for the sample number. We can write this as

$$\phi(n) = x(n) \quad \text{for } n = n_0 : n_1 \quad (8.1)$$

where $x(\cdot)$ is the ECG signal, n_0 is the beginning of the interval, in this case the S-wave, and n_1 is the manually selected end point of the ST-interval. The duration of the interval is N samples.

For the purpose of matching all other beats to the template, we use only the region of the template from $n = R(k) + n_{\nabla}$ to $n = n_1$, with n_{∇} equal to 50 ms.

Per beat i , we can define an error function $\epsilon_i(a)$ as

$$\epsilon_i(a) = \sum_{j=n_{\nabla}}^{n_1-R(k)} [\phi(R(k) + j) - x(R(i) + aj)]^2 \quad (8.2)$$

where a is the time-stretching factor and $R(i)$ is the R-peak under investigation. The result is the sum of squared differences between the template T-wave and the stretched or compressed T-wave for beat i . A progressive search in the interval $[0.8 \ 1.2]$ is conducted in order to find the value of a that minimizes $\epsilon_i(a)$. The best value of a is denoted \hat{a}_i and the ST-interval of the i^{th} beat is defined as

$$ST(i) = \hat{a}_i N. \quad (8.3)$$

The i^{th} T-wave end is then defined by the sum of $ST(i)$ and $R(i) + 50$ ms.

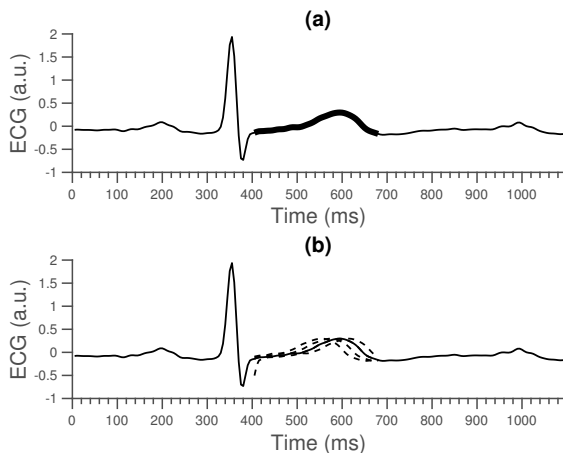


Figure 8.1: The operator selects the T-wave end on the template (a). For each of the other beats in the epoch, multiple time-compressed or time-stretched versions of the ST-interval are generated (b). These will be used to compare with the template ST-interval to derive the optimal time-stretching factor \hat{a} .

The working principle of this method is demonstrated graphically in Figure 8.1. The top panel presents the template beat. Here, the manually selected end of the T-wave is indicated. The region highlighted in bold is the segment used to compute the error function. In the next panel, several time-compressed/stretched versions of the T-wave of the investigated beat

are superimposed on the template. The area between the template T-wave and the uncompressed T-wave of the beat is then calculated and the optimal value of a is determined.

In summary, the algorithm finds the ST-interval for each beat such that the T-wave shape best matches the template T-wave under the time-stretched model.

Area under the curve method [207]

Consider k_1 and k_2 as the beginning and end of a T-wave and $L = k_1 - k_2$ as the length of the T-wave. L is generally an unknown value. The proposed algorithm mainly consists of the computation of an indicator $A(k)$, which is computed through an integration operation in a sliding window. The window width w should be defined such that $0 < w < L$. This can be roughly chosen based on the known RR-interval. At each time instant k , the indicator is defined as:

$$A(k) = \sum_{j=k-w+1}^k (s(j) - \bar{s}(k)) \quad (8.4)$$

where $s(j)$ indicates the signal value at time point j and $\bar{s}(k)$ is the mean value of the signal in a small window around k . The latter is chosen over $s(k)$ to reduce the effect of acquisition noise. $A(k)$ reaches its maximal value when $k = k_2$.

8.2.2 QRS-onset detection

The QRS-complex in an ECG signal represents the depolarisation of the ventricles that precedes the mechanical contraction. QRS onset and offset detection is difficult, because of the physiological variability of the QRS complex.

We implemented two methods: a heuristic method and a similar method to the previously described semi-automatic method. We described both below.

Heuristic method

For each k^{th} heartbeat, a search window w is defined based on the distance between the current and previous R-peak, $R(k)$ and $R(k-1)$. This window is defined so that $w = \frac{RR}{10}$, with $RR = R(k) - R(k-1)$. The right bound of

this window is equal to $R(k)$ and the left bound is defined as $R(k) - w$. As a first step, the right bound is redefined. In order to do so, the maximum and minimum within the search window are computed and the location of the steepest down slope is detected. If the difference between the amplitude at $R(k)$ and at the steepest slope is larger than half the maximum minus the minimum, then this is taken as the new right bound. Otherwise, the right bound is shifted to the left, until that point is reached. Within these new bounds, the QRS-onset is detected.

Within this window, we detect all points that are smaller than the two surrounding points. If there is only one such point, then this point is selected as the QRS-onset. If there are no such points, then we compute the amplitude difference of every point in the search window with the amplitude at $R(k)$. Each difference is then weighted, based on its distance from $R(k)$. The QRS-onset is then defined by the point with the maximal weighted amplitude difference. In the case of multiple points, the point with the maximal weighted amplitude difference is selected as the QRS-onset.

Note that this algorithm selects the deepest point of the Q-wave, rather than the actual QRS-onset.

Semi-automatic method

The first step of this algorithm is to create a representative heartbeat. This is done similarly as in Section 8.2.1.

Hereafter, the user manually annotates the onset of the QRS-complex on the template. Given that the time location of the R-peak, $R(k)$, is known, we can construct a segment $\phi(n)$ between the QRS-onset and the R-peak, where n stands for the sample number. We can write this as

$$\phi(n) = x(n) \quad \text{for } n = n_0 : n_1 \quad (8.5)$$

where $x(\cdot)$ is the ECG signal, n_0 is the beginning of the interval, in this case the manually selected QRS-onset, and n_1 is known R-peak location. The duration of the interval is N samples.

In contrast to the method described in Section 8.2.1, we now use the entire QR-interval to match with. Hence, per beat i , we can define an error function $\epsilon_i(a)$ as

$$\epsilon_i(a) = \sum_{j=0}^{R(k)-n_0} [\phi(R(k) - j) - x(R(i) - aj)]^2 \quad (8.6)$$

where a is the time-stretching factor and $R(i)$ is the R-peak under investigation. The result is the sum of squared differences between the template QR-segment and the stretched or compressed QR-segment for beat i . A progressive search in the interval $[0.8 \ 1.2]$ is conducted in order to find the value of a that minimizes $\epsilon_i(a)$. The best value of a is denoted \hat{a}_i and the QR interval of the i^{th} beat is defined as

$$QR(i) = \hat{a}(i)N. \quad (8.7)$$

The i^{th} QRS-onset is then defined by the subtraction of $QR(i)$ from $R(i)$.

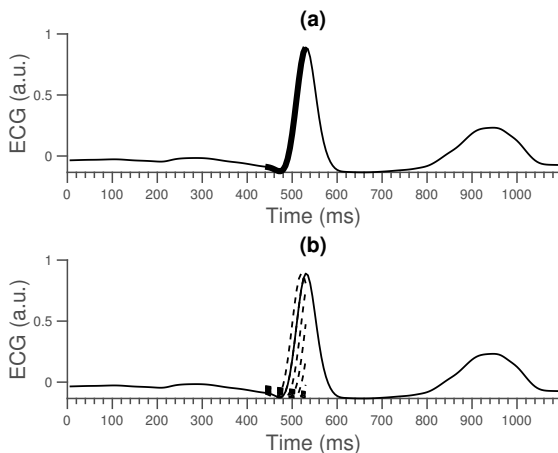


Figure 8.2: The operator selects the QRS-onset on the template (a). For each of the other beats in the epoch, multiple time-compressed or time-stretched versions of the QR interval are generated (b). These will be used to compare with the template QR interval to derive the optimal time-stretching factor \hat{a} .

The working principle of this method is demonstrated graphically in Figure 8.2. The top panel presents the template beat. Here, the manually selected QRS-onset is indicated. The region highlighted in bold is the segment used to compute the error function. In the next panel, several time-compressed/stretched versions of the QR-interval of the investigated beat are superimposed on the template.

The area between the template QR-interval and the uncompressed QR-interval of the beat is then calculated and the optimal value of a is determined.

8.2.3 Software description

The algorithms have been implemented with *MATLABR2018a*. We used GUIDE, *MATLAB*'s GUI development environment, to design the GUI. This Section describes the possible input data formats and the user interface.

Input data formats

Currently, the BVR GUI is an extension to the R-DECO toolbox. This means that it can only be called from R-DECO. There are two options: you could call it after the R-peaks are annotated or you could load the output file of R-DECO as an input file. In future work, we could extend this to a stand alone tool that allows multiple file formats as input.

User interface

The tool can be divided in four segments: Data browser, Selection and detection, Adjustment and BVR quantification (Figure 8.3). All segments are described below.

Data browser

In order to graphically correct the annotations, it is important to have a clear view of the to be investigated segment. Therefore, the software displays the ECG signal with all annotations, the respective tachogram and the QT-plot. We chose to use two y-axis, one for the tachogram and one for the QT-plot, to be able to better detect outliers.

The window width of the x-axis can be adjusted in three different ways: 1) the range edit box, 2) the plus and minus buttons and 3) the zoom button. All three methods also adjust the width of the scroll bar.

The scroll bar can be used to slide through the signal. Since both axes are linked, they slide at the same time. The y-axis range in both axes is adjusted automatically according to the data within the selected range.

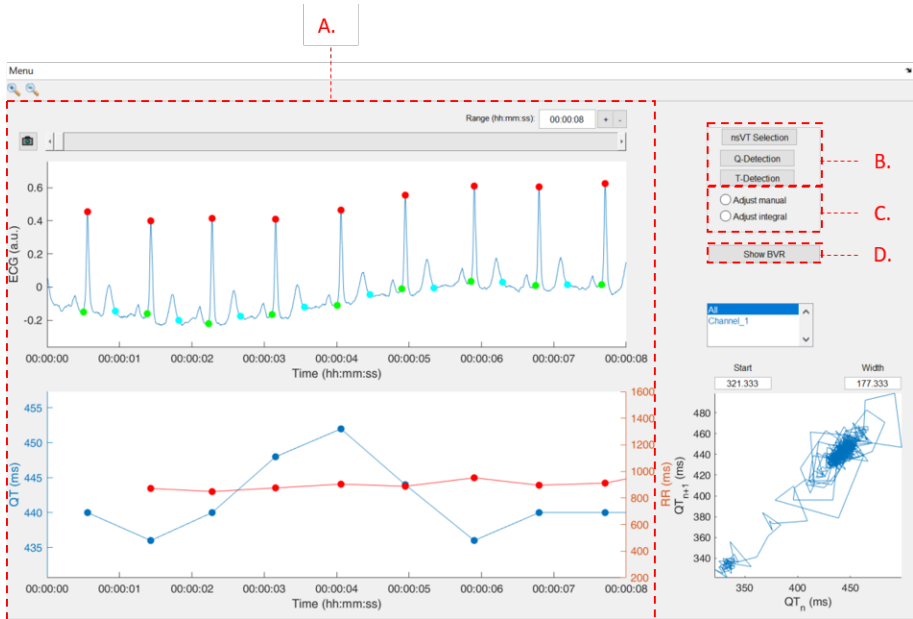


Figure 8.3: The GUI of the BVR tool. The user interface can be divided in four segments: A. Data browser, B. Selection and detection, C. Adjustment and D. BVR quantification.

An ECG recording with multiple channels might result in axes that become unclear. Therefore, the user can switch between the different channels by selecting the desired channel(s) in the list box.

Similarly to R-DECO, we also included the option to take 'pictures' of the axes. This saves both axes in a format of choice, without including any of the buttons or bars from the user interface.

Selection and detection panel

The selection and detection panel includes one selection and two detection actions. All actions are described below.

- **nsVT selection:** This button allows the user to define an nsVT segment of interest. This can be done both before and after QT-delineation. After pushing the 'nsVT Selection' button, the user is prompted to select a

window by clicking, dragging and releasing the mouse. The window is shown as a transparent patch over the data and can be enlarged, shrunk or moved with the mouse. As a result, the BVR will only be computed within this window.

- **Q-detection:** Pressing this button prompts a list box in which the user has to indicate which of the previously described methods he/she would like to use to detect the QRS-onsets. After selecting the desired method, the QRS-onsets are detected and visualized on the top plot. If the T-wave ends are already detected, the QT- and Poincaré plot are constructed and visualized as well.
- **T-detection:** Pressing this button prompts a list box in which the user has to indicate which of the previously described methods he/she would like to use to detect the T-wave ends. After selecting the desired method, the T-wave ends are detected and visualized on the top plot. If the QRS-onsets are already detected, the QT- and Poincaré plot are constructed and visualized as well.

Adjustment panel

This panel includes two methods to adjust misplaced annotations: adjust manual and adjust integral.

- **Adjust manual:** Selecting this button allows the user to select and move any of the annotation by clicking and dragging the mouse. While hovering over the ECG plot, the annotation closest to the mouse location is highlighted. After clicking on the desired annotation, it can be moved by dragging the mouse. Note that the movement of each annotation is limited by the annotation on the left- and right-hand side. While adjusting the location, the tachogram or QT plot is automatically updated, depending on the annotation that is moved. Upon mouse release, the new location is stored.
- **Adjust integral:** When selecting this button and clicking on a T-wave, the T-wave location is going to be recomputed using the 'area under the curve' method. This is especially useful if the other method was used for the detection of the T-wave end, but you are not satisfied with the result of a particular annotation. After releasing the mouse, the QT plot is automatically updated.

BVR quantification

BVR can be quantified in a multitude of ways. Here, it is quantified by the short-term variability of the QT interval (STV), as first described by Thomsen et al. [179]. This can be assessed based on the dimensions of a Poincaré plot. This is a scatter graph that visualizes the correlation between two consecutive data points in a time-series. In the context of BVR quantification, each repolarization duration (D_{n+1}) is plotted against the repolarization duration of each previous heartbeat (D_n) for a predetermined number of consecutive beats (N) (Figure 1A). Each deviation from the diagonal, which is referred to as the line of identity, reports the occurrence of a difference in repolarization duration between 2 subsequent beats. The STV is defined as the mean orthogonal distance from the line of identity to all points on the Poincaré plot. This corresponds to the following formula:

$$STV = \frac{\sum_{n=1}^N |D_{n+1} - D_n|}{N \times \sqrt{2}}. \quad (8.8)$$

In 2004, Thomsen et al. used 30 consecutive beats under stable idioventricular focus to derive the STV. Since this approach has been adopted by many other authors, we also implemented the formula with $N = 30$. The resulting BVR can be visualized by pushing the 'Show BVR' button. This opens a new figure which displays all BVR values starting from the 30th R-peak.

8.2.4 Export results

The software can export the results in two different ways: 1) an Excel file and 2) a MATLAB file.

1. Excel file (*.xls): Similarly to R-DECO, a new workbook is created with on the first sheet a general overview of the file: the number of channels, the sampling frequency, the duration of the signal and the duration of the analysis period. The number of additional sheets is defined by the number of channels. Each sheet contains the R-peak locations, the RR-intervals, the QRS-onset locations, the T-wave end locations, the QT-intervals, the BVR values and a number of basic metrics, such as the mean heart rate.
2. MATLAB file (*.mat): This file contains a single structure named *data*. In accordance to the structure of the Excel file, a structure is created per channel, which contains the same information as the Excel file.

8.2.5 Potential of future growth

Currently, the BVR-tool is an add-on to the R-DECO GUI. This is mostly reflected in the fact that the tool can only use the output format of R-DECO as input format. Therefore, the most obvious improvement would be to broaden the input format possibilities. This would allow researchers who have their ECG signals and R-peak locations in a different format or structure to also use the tool. Eventually, this could expand the tool from a simple add-on to a stand alone tool.

Although the implemented annotation algorithms have shown the capability to accurately detect the QRS-onset and T-wave end, it might be that more accurate algorithms exist. For instance, Martinez et al. used a wavelet based approach to delineate the heartbeats [112]. They have shown higher accuracy on the same database. The advantage of the current framework is that it is very easy to interchange or add other algorithms. The reason why we implemented the current algorithms is that those were readily available and had an acceptable performance. Moreover, any inaccuracies that might be present could be resolved with the provided correction tools.

8.3 Predicting non-sustained ventricular tachycardia in ischemic heart disease patients

8.3.1 Data

The study population included 20 patients with IHD. The study protocol was approved by the Ethical Committee of the University Hospitals of Leuven (S56074). From each patient, a 24 hour digital 2-lead Holter recordings was collected prior to ICD implantation. The mean age was 65.6 ± 8.9 years and 15 % were female. The severity of each patient's disease is classified according to the New York Heart Association (NYHA) Functional Classification. The details of the clinical data are shown in Table 8.1.

8.3.2 Methods

ECG analysis

We performed the analysis mostly on lead A, which corresponds to lead II of the standard 12-lead clinical ECG. Only if the QRS-complexes were too small

Table 8.1: Clinical data of patients with IHD. Values are shown as mean \pm standard deviation.

Group size (#)	20
Age (years)	65.6 \pm 8.9
Women/men	3/17
LVEF (%)	32.9 \pm 12.6
NYHA class	
I	6
II	8
III	6

or the signal quality was not sufficient in the region of interest, we analysed lead B, which corresponds to lead VI.

Since a 24 hour Holter recording inevitably contains a fair amount of noise, pre-processing was necessary to improve the signal quality. We removed baseline wander and power line interference with, respectively, a high-pass Butterworth filter with a cut-off frequency of 0.5 Hz and a notch filter at 50 Hz.

We detected the R-peaks using the R-DECO interface and made sure that all heartbeats were correctly detected. After this detection, a cardiologist reviewed the annotations for any leftover (ventricular) ectopic beats. If any were present, then they are removed. This is done because only Normal-to-Normal (NN) intervals should be taken into account for BVR analysis.

Hereafter, we derived the QT-intervals with the newly developed BVR tool. All annotations were examined and, if need be, adjusted by a cardiologist. The drag-and-drop feature of the tool was used for this purpose.

nsVT episodes were semi-automatically identified based on the tachogram and the raw signal. Hereafter, the BVR was assessed at time points 1, 5 and 30 minutes prior to an nsVT episode, and at rest: a fixed moment during sleep (03.00am).

Statistical analysis

All variables are reported as mean \pm SD. We used repeated measures ANOVA to assess the significance levels between BVR values at different points in time prior to nsVT. P-values lower than 0.05 were accepted as statistically significant. All programming and calculations were done using *MATLAB R2018a*.

8.3.3 Results

An overview of the nsVT characteristics is given in Table 8.2. An average nsVT lasted 49 heartbeats, spread over 20 s. This equals a heart rate of 146 bpm.

Table 8.2: nsVT characteristics. Values are shown as mean \pm standard deviation.

HR (bpm)	146 \pm 9
Cycle length (ms)	424 \pm 24
# beats	49 \pm 7
Duration (s)	20 \pm 3

The QT Poincaré plots from a representative patient at 03.00am and one minute prior to nsVT are displayed in Figure 8.4. We have found that resting BVR (7.26 ± 3.88 ms) was significantly increased at 5 minutes (14.40 ± 7.61 ms) and 1 minute (18.01 ± 6.48 ms), but not at 30 minutes (8.90 ± 4.93 ms) prior to nsVT (Figure

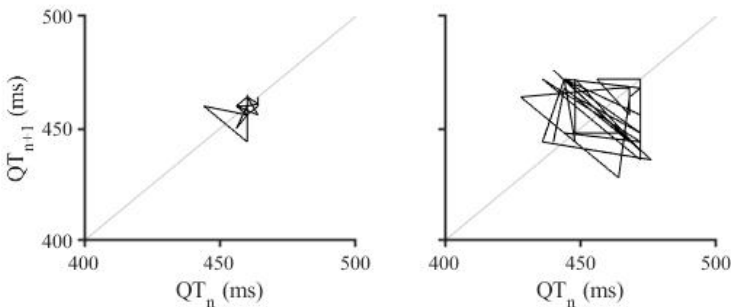


Figure 8.4: Poincaré plot from a representative patient with IHD. Increased dispersion of the repolarization time, and thus the BVR values, from 03.00am (a) to 1 minute prior to nsVT (b) is clearly visible.

8.3.4 Discussion

We have shown that nsVT events in IHD patients are preceded with a significant increase in their BVR. This indicates that the assessment of BVR could improve risk stratification of IHD patients and identifies a novel prediction method for impending VT.

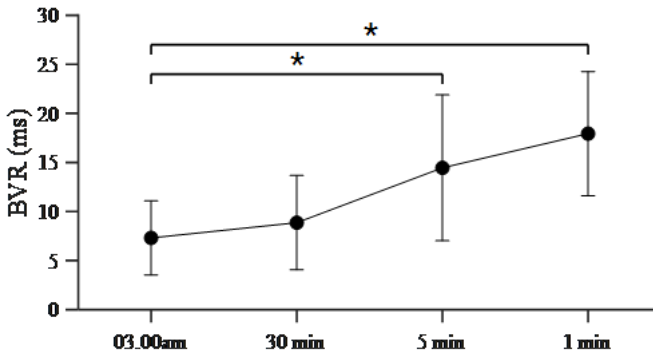


Figure 8.5: Results of the BVR analysis. A significant increase in BVR could be observed for 1 minute and 5 minutes prior to nsVT. * $p < 0.05$ vs. 03.00am.

BVR has proven to be a sensitive predictor of lethal ventricular arrhythmias in animal models [162], where it reflects the electrical substrate of repolarization heterogeneity and impaired reserve that precedes re-entrant arrhythmia [179]. The observed temporal changes in BVR likely reflect the dynamic substrate of repolarization heterogeneity preceding arrhythmic re-entry.

Temporal variations in BVR have been recently described in relation to diurnal rhythms in high risk patients, and suggested as an assessment marker [173]. We have used a more clinically relevant concept of impending nsVT to demonstrate the clinical relevance of temporal BVR in improving risk stratification of IHD patients.

Changes in cardiac cycle duration frequently precede ventricular arrhythmias and detailed analysis of these changes that have clinical value have been proposed for integration into monitoring devices, such as ICD software. We have shown that a significant increase in BVR already occurs at 5 minutes prior to the nsVT event. As such, monitoring BVR could allow for some much needed time to respond to impending arrhythmia. It might be enough for a doctor to take action in a hospital or to alarm relatives when at home.

In order to confirm these preliminary findings, the obtained BVR values should be compared to the BVR values throughout an entire day. This way, it could be investigated if the observed increases in BVR are solely present before a nsVT or occur also at different events.

We used a robust, semi-automated fiducial point detection technique with manual corrections. Future research should investigate the influence of fully automated fiducial point detection algorithms on the BVR analysis. If accuracy

and robustness can be guaranteed, this would allow BVR to be monitored in real-time and strengthen the case to be integrated in wearable devices.

The major limitation of this study is the small patient cohort. Therefore, while the results presented in this study are promising, the findings are still preliminary. In order to have more clinical value, they need to be verified in larger patient cohorts and other high-risk patients.

8.4 Conclusion

One of the current limitations of BVR analysis is the requirement of separate software and dedicated personnel. We tried to overcome this limitation by developing an easy to use graphical user interface for semi-automated analysis of the ECG signals. It includes a semi-automated and automated QRS-onset and T-wave end detection algorithm. In the case of wrong detections, the user can graphically correct or remove them. Currently it has not been published yet, but the idea would be to publish it as part of a large ECG toolbox.

We used this tool to investigate the temporal evolution in BVR before spontaneous nsVT. The presented results suggest that temporal changes in pre-arrhythmic BVR could be used to predict imminent nsVT events in IHD patients. These preliminary results set a strong precedent that both reinforces the value of BVR analysis in the risk stratification of IHD patients and identifies a novel method of impending VT prediction that could be used for real-time analysis and monitoring. The obtained results need to be verified in larger patient cohorts and other high-risk patients.

Part IV

Sport applications

Chapter 9

Cardiorespiratory strength during exercise

In this Chapter, we used the algorithms and tools described in the previous Chapters to explore the strength of the cardiorespiratory coupling during a maximal exercise test.

To extract both cardiac and respiratory information, we used only the ECG signal. Hence it was of utmost importance that the signal was as clean as possible. We used the methodology described in Chapter 4 to get an indication of the quality. Hereafter, we used the R-peak detection and correction tool from Chapter 7 to detect the R-peaks and correct wrong annotations.

Since we used a derived respiration signal, instead of the real respiration, the results of this Chapter should be taken as preliminary. However, it does serve as an excellent example of how the previously presented tools and algorithms could be used in an analysis pipeline.

9.1 Introduction

Cardio-pulmonary exercise testing provides valuable diagnostic and prognostic information for patients with cardiovascular and pulmonary disease [14]. It includes ventilatory gas exchange measurements which provide a wide array of unique and clinically useful information. For example, it measures breath-by-breath oxygen uptake (VO_2), carbon dioxide output (VCO_2), and ventilation. Moreover, it has demonstrated high reproducibility of VO_{2max} values [40].

A reduced exercise performance is mostly due to a combination of multiple factors such as musculoskeletal, ventilatory, cardiovascular and gas exchange or simply due to deconditioning [98]. However, it is unclear what the role of the cardiorespiratory coupling is in this regard.

It is well-known that cardiovascular and pulmonary functions are coupled, but few attempts have been made to study the strength of this coupling during exercise. In this study, we investigated whether the strength of the cardiorespiratory coupling is related to a quantitative measure, such as the VO_2 or RER , during exercise.

Normally, cardiorespiratory analysis requires two modalities: ECG and respiration. However, in the case of non-availability of continuous respiratory signals, which is the case here, we need to estimate the respiratory signal from another modality. Previous studies have shown that EDR can be used to derive the respiratory rate and the respiratory wave morphology [50, 193]. This technique is based on the observation that the positions of the ECG electrodes on the chest surface move relative to the heart as the lungs fill and empty [104]. Moreover, the change in volume also causes transthoracic impedance variations. These two effects result in a measurable change in the ECG morphology due to respiration. Hence, in this study, we used only the ECG signal to compute the cardiorespiratory coupling.

It has been shown that EDR becomes more accurate when two or more ECG signals are available, but a single lead is also sufficient [104]. Therefore, we adopted a single lead approach in order to mimic an ambulatory approach.

In Section 9.2, we described the dataset, in Section 9.3, we described the methodology and in Section 9.4, we described and discussed the results of this study. Conclusions are drawn in Section 9.5.

9.2 Data

The database consists of 25 healthy male volunteers. They were not taking any medication, had normal blood pressure levels and no abnormal ECG patterns. All subject participated in sports activities at least three days a week. The study protocol was approved by the ethics committee of Aragón and was in accordance with the Declaration of Helsinki. Written informed consent was obtained from each subject. Table 9.1 shows the study population demographics.

Table 9.1: Overview of the population demographics. Values are shown as mean \pm standard deviation.

Age (years)	33.4 ± 5.2
Height (cm)	178.0 ± 5.5
Weight (kg)	74.8 ± 7.0
Body mass index (kg/m^2)	23.6 ± 2.1
$\dot{V}O_{2max}$ (ml O_2 /min)	4695.9 ± 604.5
$\dot{V}O_{2max}/kg$ (ml O_2 /min/kg)	63.0 ± 9.8

All tests were conducted at the University of Zaragoza (Spain). The subjects completed three sessions in consecutive weeks: a maximal test on a treadmill (MaxT), a submaximal test on a cycle ergometer (SubC) and a submaximal test on a treadmill (SubT). We performed the cardiorespiratory analysis only on the SubC recordings. The reasoning behind this decision is that the signals that are recorded during the cycle ergometer test most likely contain less noise, compared to the treadmill test. This is due to the fact that the upper body remains more stable during the cycle ergometer test. The entire SubC protocol is depicted in Table 9.2.

All tests were divided in three phases: a resting phase, an exercise phase and a recovery phase. The resting phase was the same for all three tests. Namely, the subjects were monitored while they remained seated for five minutes. They were asked to refrain from talking and moving. This phase was used to measure baseline cardiorespiratory variables.

In the SubC test, the exercise phase started at 75 W and increased with 25 W/min. The pedalling frequency was fixed at 80 rpm. When the subject reached 90% of the maximum HR, they were asked to continue pedalling for two more minutes at the resultant power.

The recovery phase followed after the exercise phase. During this phase, the subjects were asked to remain running or cycling at the respective initial speed or power for about three to five minutes. However, despite the protocol, the

Table 9.2: SubC protocol. During the resting phase, when the subject reached 90% of the maximum HR, they were asked to continue pedalling for two more minutes at the resultant power.

Phase	Duration (min)	Power (W)
Rest	5	0
Exercise	1	75
Exercise	1	100
Exercise	1	125
Exercise	1	150
Exercise	1	175
⋮	⋮	⋮
Exercise	2	Power at 90% HR_{max}
Recovery	3 - 5	75

subjects behaved very inconsistently. Especially after the MaxT test, some subjects needed a break to breath, while others could continue, but at a slower pace.

The ECG signals were recorded with a high-resolution Holter device (Mortara 48-hour H12+, Mortara Instrument, Milwaukee, Wisconsin) at a sampling frequency of 1000 Hz. Nine leads were recorded: I, II, III, aVL, aVR, aVF, V4, V5 and V6.

An open-circuit sampling system (Oxycon Pro, Jaeger-Viasys Healthcare, Hoechberg, Germany) was used to record the following variables in a breath-by-breath manner: respiratory rate, oxygen consumption ($\dot{V}O_2$), carbon dioxide production ($\dot{V}CO_2$) and the respiratory exchange ratio (RER). The latter is computed as $\frac{\dot{V}CO_2}{\dot{V}O_2}$, in other words, the ratio of carbon dioxide production and oxygen consumption.

9.3 Methods

9.3.1 Lead selection

First, we filtered all signals by means of a zero phase, 2nd order high- and 4th order low-pass Butterworth filter with cut-off frequencies at 1 Hz and 40 Hz, respectively. These filters ensure the removal of baseline wander and high

frequency noise, without altering the information contained in the characteristic waveforms of the ECG.

Hereafter, since we only used a single lead to estimate the cardiorespiratory coupling, it was essential that this lead contains the least amount of noise possible. Hence, we investigated the quality of each lead using the methodology described in Chapter 4.

We divided each recording into non-overlapping segments of 10 seconds and computed the quality score per segment. Then, we computed an average quality score per lead, per recording. The lead with the highest overall quality is used for further analysis.

Over all subjects and segments, V5 obtained the highest quality score: 90%. Therefore, if some of the leads had an equal quality score, and those leads include V5, then V5 was selected for further analysis. Otherwise, a random selection was made.

9.3.2 HRV derivation

We detected the R-peak locations on the selected lead using the algorithm described in Chapter 7. Each location was then manually verified using the developed user interface. Hereafter, we located the ectopic heartbeats with a threshold based method developed by the BSICoS group [115] and used the integral pulse frequency modulation (IPFM) model to generate a HRV representation that is sampled at 4 Hz. The details of the IPFM model can be found in [13].

9.3.3 ECG derived respiration (EDR)

A recent study by Varon et al. compared ten single-lead EDR techniques under multiple conditions [193]. They have found that the simplest methods, namely methods that explore the amplitude changes between the R- and S-wave or the change in slope range of the QRS complex, can be used to accurately estimate the respiratory signal from the ECG [193]. Hence, in this study we used the amplitude differences between the R- and S-wave to derive the respiration. The R-wave amplitude is defined as the maximum amplitude within the QRS-complex and the amplitude of the S-wave is defined as the minimum amplitude in a 80 ms window after the R-wave [145]. We used cubic spline interpolation to resample the EDR to 4 Hz, similarly to the HRV signal.

After we generated an HRV representation and the EDR for the entire recording, we divided these signals into segments of one minute according to the power levels of the ergometer. Since these power levels were not recorded, we had to do this division manually, based on visual inspection. From each minute, we computed the cardiorespiratory coupling.

9.3.4 Cardiorespiratory coupling (CRC)

A common approach to evaluate RSA in a non-invasive way is through HRV. Under resting conditions, it was shown that power spectral parameters of the HRV are able to accurately quantify the RSA. Typically, the total power in the high frequency (HF, 0.15-0.4 Hz) band of the HRV's power spectrum is used. The downside of this quantification is that it assumes that the respiration frequency remains within this band. Unfortunately, this is not always the case. For example, during maximal exercise, the subjects breath at a rate that is higher than the upper limit.

A recent study by Morales et al. compared seven RSA estimation methods on simulated and real-life data [136]. They have shown that estimates based on frequency bands perform poorly when the respiration frequency is not within the boundaries of the respective bands or when the respiratory spectrum has a broad bandwidth. One of the best performing methods is the normalized power of the HRV that is linearly related to the respiration [192]. Hence, we used this method for this study. It decomposes the HRV signal into two components: one respiration component, which describes all dynamics (linearly) modulated by respiration, and a residual component, which describes all dynamics modulated by other mechanisms and possible non-linear respiratory effects. The decomposition was obtained using orthogonal subspace projections (OSP).

Consider two physiological signals $X = \{x_i\}_{i=1}^N$ and $Y = \{y_i\}_{i=1}^N$, where X corresponds to the respiratory signal, Y corresponds to the HRV signal, and N is the length of the signals.

In order to extract all dynamics of the heart rate that are linearly related to respiration, Y can be projected onto a subspace spanned by $V \in \mathbb{R}^{(N-L+1) \times L}$. This is a time-delayed embedding of the respiration using L delays [191].

We can then use V to project the HRV signal Y onto the respiratory subspace by means of a projection matrix P . This is defined as

$$P = V(V^T V)^{-1} V^T \quad (9.1)$$

Hereafter, the component of the HRV that is linearly correlated to the respiration (Y_X) can be derived as

$$Y_X = PY, \quad (9.2)$$

The RSA is then estimated as the relative power of the linear respiratory influences on the HRV, computed as,

$$\mathcal{P}_X = \frac{Y_X^T Y_X}{Y^T Y}. \quad (9.3)$$

9.4 Results and discussion

In this Section we present and discuss the results of this study. First, we discuss the EDR and then the CRC results.

9.4.1 EDR

Since continuous respiratory signals were not available, we needed to estimate the respiratory signal from the ECG. In order to do so, we used a method based on R- and S-wave amplitude differences to derive the respiration. However, without a reference signal, there is no straightforward way to verify whether the resulting EDR signal correctly represents the real respiration. In an attempt to get some notion of the success rate of the used methodology, we compared the average measured respiration rate (BR) and the peak frequency of the EDR on a minute-by-minute basis. In Figure 9.1, we plotted the resulting values, averaged over all subjects. In order to have an equal number of samples for all subjects, we included only the resting phase and 11 minutes of the exercise phase.

During the resting phase of the test, which takes place during the first five minutes, a large discrepancy can be observed. This is an odd finding, since the subjects were not pedalling yet, so the influence of muscle noise and cable motion should be limited to a minimum. After the resting phase, this discrepancy seems to reduce up to minute 12. At this point, the power has been raised to 250 W, which might prompt the subjects to gain more power by moving their upper body. This, in turn, results in more noisy signals and thus a less accurate estimation of the respiration. The average overall correlation coefficient, ρ , is 0.695 ± 0.279 , which indicates that there are large inter-subject differences.

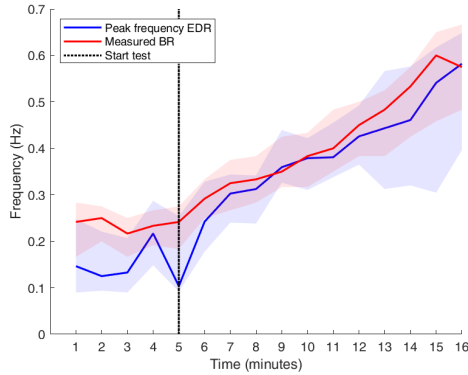


Figure 9.1: Comparison between the peak frequency of the EDR and the measured breathing rate (BR). The dotted line indicates the start of the exercise phase. The thick line indicates the median and the boundaries indicate the 25th and 75th percentiles.

9.4.2 CRC

Figure 9.2 present the average \mathcal{P}_x , VO_2 and RER values over all subjects during the resting phase and 11 minutes of the exercise phase. The VO_2 remains more or less constant during the resting phase with very little inter-subject variability. During the exercise phase, it increases up to a median value of 45.90 ml O_2 /min/kg.

The \mathcal{P}_x is characterized by a large inter-subject variability, visible by the large difference between the 25th and 75th percentiles. We cannot observe any clear trend, however, surprisingly, we obtained a negative significant correlation ($\rho = -0.56$). Lastly, the RER mirrors the pattern of the VO_2 with the exception that when the exercise phase starts, it decreases first, before increasing.

Given the conclusions on the comparison of the recorded BR and that estimated from the EDR, we assumed that the negative correlation between \mathcal{P}_x and VO_2 could be due to a poor estimation of the respiration signal during the exercise phase. Hence, as an extra experiment, we selected the subjects with a nearly perfectly estimated respiration rate ($\rho > 0.99$). This included only two subjects. The obtained values are shown in Figure 9.3.

No significant correlation was obtained between \mathcal{P}_x and VO_2 during the combined resting and exercise phase. However, we did obtain a significant correlation when only the exercise phase is included ($\rho = 0.78$). Concerning the

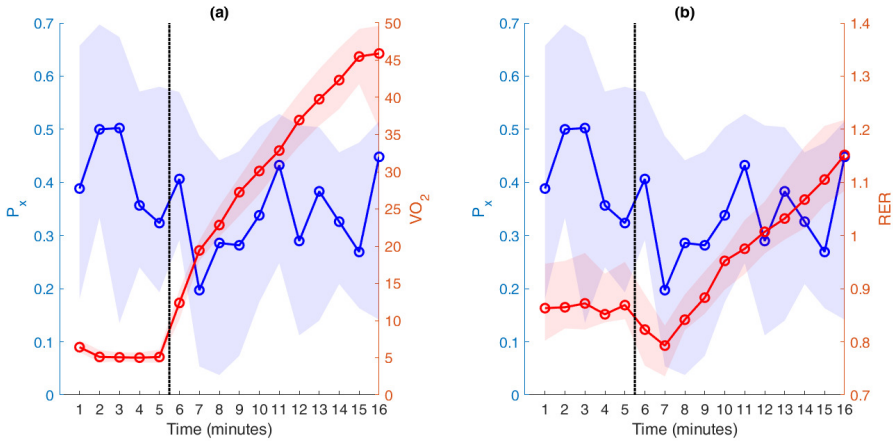


Figure 9.2: Comparison between (a) \mathcal{P}_x and VO_2 and (b) \mathcal{P}_x and RER . The dotted line indicates the start of the exercise phase. The thick line indicates the median and the boundaries indicate the 25th and 75th percentiles.

Table 9.3: Results of the correlation analysis between P_X , VO_2 and RER for all subjects and only the subjects for which the estimated and measured breathing rate correlated almost perfectly ($\rho > 0.99$). * indicates $p < 0.05$.

	P_X (All)		P_X ($\rho > 0.99$)	
	All	Exercise	All	Exercise
VO_2	-0.56*	0.12	0.05	0.78*
RER	-0.09	0.36	0.68*	0.81*

RER , we obtained a significant correlation both when the phases are combined and when only the exercise phase is included ($\rho = 0.68$ and $\rho = 0.81$). All correlation values are summarized in Table 9.3.

These findings prompted the question of how the correlation coefficients between \mathcal{P}_x and VO_2 , and \mathcal{P}_x and RER relate to the quality of the EDR as measured by the correlation coefficient between the peak frequency of the EDR and the measured respiration frequency. This was investigated in a last experiment by comparing the correlations during the exercise phase.

From Figure 9.4 we can observe that, the higher the desired correlation between the peak frequency of the EDR and the measured respiration frequency, the

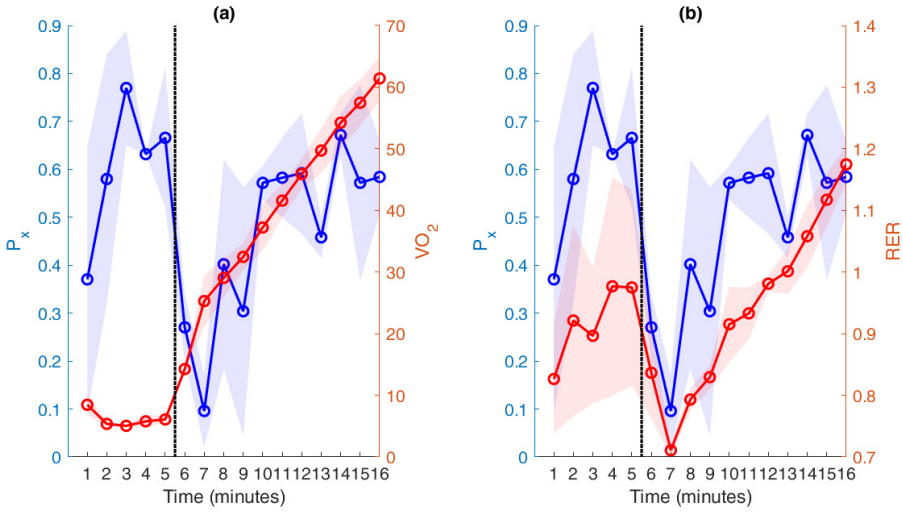


Figure 9.3: Comparison between (a) P_x and VO_2 and (b) P_x and RER of the subjects with a correlation coefficient higher than 0.99. The dotted line indicates the start of the exercise phase. The thick line indicates the median and the boundaries indicate the 25th and 75th percentiles.

lesser subjects are included. This is in accordance with the large standard deviation that we observed in Section 9.4.1. Only 19 recordings registered a significantly positive correlation ($\rho > 0.39$). For these recordings, we could not observe a significant correlation between P_x and VO_2 ($\rho = 0.22$), however, we could observe a significant correlation between P_x and RER ($\rho = 0.47$). Additionally, we could observe that for both metrics, the values increased when the correlation between the peak frequency of the EDR and the measured respiration frequency increased. The downside is that this is at the cost of included subjects.

These preliminary results indicate that P_x follows a similar pattern as the VO_2 and RER during the exercise phase of a submaximal exercise test. However, due to poor EDR quality, we could not confirm these findings for all subjects. In order to fully understand the relationship between the cardiorespiratory coupling and measurable physiological signals, we should repeat this analysis with a recorded respiration signal.

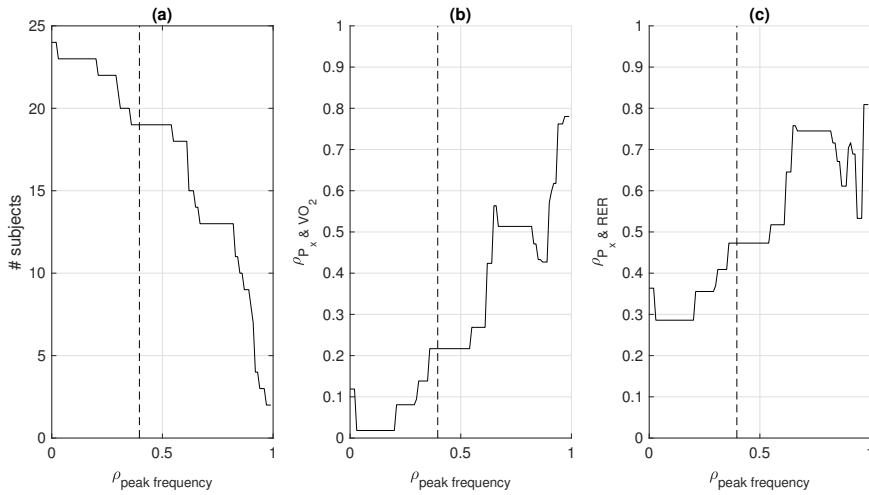


Figure 9.4: **(a)** The number of subjects, compared to the correlation coefficient between the peak frequency of the EDR and the measured respiration frequency. **(b)** The correlation coefficient between \mathcal{P}_x and VO_2 . **(c)** The correlation coefficient between \mathcal{P}_x and RER . The striped line indicates the threshold where correlation coefficient between the peak frequency of the EDR and the measured respiration frequency is significant.

9.5 Conclusion

In this Chapter we explored if the strength of the cardiorespiratory coupling is related to CRF parameters during exercise.

Due to a lack of modality specific recordings, we could not fully answer this question. Since we did not have a continuous respiratory signal, we had to derive the respiration from the ECG. In a noise free situation, this EDR is rather accurate, but this is no longer true during exercise. We have shown large inter-subject variability between the estimated and measured respiration rate, which indicates that the EDR was not able to accurately estimate the respiratory signal.

When we incrementally increased the desired correlation between the estimated and measured respiration rate, we could observe that the correlation between \mathcal{P}_x and VO_2 and RER increased during the exercise phase. When we only included the subjects with the best EDR, we found that \mathcal{P}_x follows a similar pattern as

the VO_2 and RER during the exercise phase of a submaximal exercise test.

These findings are the first step towards a non-invasive estimation of the physical condition of a person, based on cardio-respiratory interactions. However, as shown in this study, an accurate estimation of the respiration is essential. Hence, the performance of EDR algorithms during exercise needs to be improved before we could take a next step.

These preliminary results need to be confirmed on a dataset where a continuous respiration signal is recorded. This would allow for an accurate derivation of \mathcal{P}_x and a fair investigation of the physiological value of this parameter during exercise.

Chapter 10

Left ventricular mass estimation

This chapter describes a methodology to estimate left ventricular mass (LVM) using features derived from ECG, DXA and demographics. The gold standard for the assessment of cardiac mass is CMR. However, it is costly and requires specific expertise. ECG criteria could provide a low-cost solution, but have shown to be poorly correlated with LVM in athletes. We hypothesized that this poor correlation could be overcome by taking body measurements (length, weight) and composition (fat mass, lean mass and bone mass) into account.

The results of this study were published in **MOEYERSONS, J., DE BOSSCHER, R., DAUSIN, C., CLAESSEN, G., LA GERCHE, A., BOGAERT, J., WILLEMS, R., VAN HUFFEL, S., VARON, C.** Predicting Left Ventricular Mass Using ECG, Demographic and DXA Features. *Computing in Cardiology* 47, Rimini, Italy, (Sep 13-16, 2020)..

10.1 Introduction

Increased wall thickness, LV mass (LVM) and chamber dilation are the primary features of exercise induced cardiac remodeling [48]. Distinguishing these physiological adaptations from concentric hypertrophic cardiomyopathy (HCM) can be challenging [53].

The gold standard for the assessment of cardiac volumes and mass is CMR. Structural measures by CMR are both accurate and reproducible [92]. However, cost and operational considerations tend to limit its utility in large-scale population studies and clinical trials [152].

ECG criteria for left ventricle hypertrophy (LVH) provide a low-cost solution, but have low sensitivity in the general population. In an attempt to resolve this issue, the combination of multiple criteria was suggested. This approach increased the sensitivity for detection of LVH, but came at the expense of a lowered specificity [142]. A more sophisticated solution was the use of multivariate statistical models for the estimation of LVM. This way, the LVM could be estimated on a continuous scale, instead of a binary classification. Good results have been obtained in an older population [152], but this has not been done in a young athlete population.

The ECG is measured at the body surface. The conduction of the electrical signal of the heart to the skin is influenced by the composition of the different organ tissues separating the electrodes from the heart. Including body composition to an ECG-based estimation of LVM might therefore improve its accuracy. In this study, we first combined ECG features in a regression model to estimate LVM. Hereafter, we added demographic and body composition features. The latter are measured with DXA. We hypothesize that adding these features will improve the estimation.

The first objective of this study is to identify features from the different modalities that are highly correlated with LVM. The second objective is to assess whether adding demographic and/or DXA features could improve an ECG-based regression model for the estimation of LVM in athletes.

10.2 Data

This study was conducted using data of the ProAtHeart study. This is a longitudinal cohort study in elite young endurance athletes aiming to phenotype the structural and functional cardiovascular adaptations associated with high volume exercise. It was approved by the Ethical Committee of the University

Hospitals Leuven. 107 young competitive endurance athletes (19 ± 2 years; 35 female), were included in the KU Leuven (Belgium) and the Baker Heart and Diabetes Institute (Australia). The full demographics are shown in Table 10.1.

Table 10.1: Population demographics. Values are shown as median and interquartile range.

	All	Men	Women
Number	107	72	35
Age (years)	19.00 (2.00)	19.00 (2.00)	18.00 (2.75)
Height (m)	1.78 (0.13)	1.82 (0.10)	1.70 (0.09)
Weight (kg)	66.50 (12.20)	69.33 (10.43)	58.36 (10.01)
LVM (g)	142.80 (54.54)	159.50 (45.80)	106.05 (32.35)
V_{O_2max} (ml O_2/min/kg)	61.00 (15.95)	64.18 (11.75)	51.47 (8.79)

10.3 Methods

10.3.1 Feature collection

Every athlete underwent a 12-lead ECG, DXA scan and CMR. A subset of features, which have shown potential as single features or as combinations in earlier studies, was selected from each modality for further analysis.

The data from all 12-lead ECGs were automatically processed and measurements of each lead were averaged over the ECG to obtain representative metrics. We selected: SV1, SV3, RV5, RV6, RI, RaVL, SaVL and QRS duration; where S is the absolute value of the S-wave amplitude and R is the R-wave amplitude in the respectively indicated leads.

Additionally, we included 4 ECG criteria for LVH: Cornell voltage (CV) and product (CP), Sokolow-Lyon voltage (SLV) and product (SLP). CV is considered the most accurate criterion for LVH. It is computed as the amplitude of R in aVL plus the amplitude of S in V3 (RaVL+SV3). The SLV is computed by the amplitude of S in V1 and the amplitude of R in V5 or V6, whichever is the larger (SV1+RV5/6). Both products are obtained by multiplying the voltages with the QRS duration.

Five demographic variables were used: age, sex, weight (kg), height (m) and body surface area (BSA, m^2). BSA was computed with the Dubois and Dubois equation [53] and was included because of its frequent use as a scaling parameter for LVM.

From the DXA scan, we selected: bone mineral content of the trunk (BMCTrunk), fat mass of the trunk (FMTrunk), lean mass of the trunk (LMTrunk), total fat mass (FMTotal) and total lean mass (LMTotal). All features are depicted in Table 10.2.

For calculation of LVM, the volume of the myocardium measured by CMR was multiplied by the specific gravity of the myocardium (1.05 g/mL). The volume of the myocardium was obtained by subtracting the LV end-diastolic volume from the epicardial volume, both drawn from apex to basal short axis views.

Table 10.2: Overview features. Values are shown as median and interquartile range.

	All	Men	Women
SV1 (mm)	9.0 (6.6)	12.0 (6.8)	6.5 (3.0)
SV3 (mm)	13.0 (13.4)	17.9 (12.3)	7.0 (7.2)
RV5 (mm)	20.0 (9.3)	21.0 (9.5)	17.0 (8.3)
RV6 (mm)	16.0 (7.2)	16.5 (6.5)	13.6 (7.4)
RI (mm)	4.4 (3.0)	4.4 (3.0)	4.3 (2.1)
RaVL (mm)	1.1 (1.2)	1.1 (1.7)	1.3 (0.9)
SaVL (mm)	4.0 (4.4)	4.0 (5.5)	4.0 (3.3)
SIII (mm)	0.0 (1.5)	0.2 (2.0)	0.0 (0.9)
QRS (ms)	98.0 (12.0)	100.0 (9.0)	90.0 (9.5)
CV (mm)	15.0 (14.6)	19.5 (12.5)	8.0 (7.3)
CP (mm×ms)	1525.0 (1435.7)	2028.5 (1159.1)	722.2 (688.3)
SLV (mm)	30.5 (13.5)	33.0 (13.4)	24.8 (10.7)
SLP (mm×ms)	2976.0 (1406.3)	3396.0 (1304.9)	2277.0 (922.7)
Age (year)	19.0 (2.0)	19.0 (2.0)	18.0 (2.8)
Height (cm)	177.6 (13.1)	182.0 (9.8)	170 (9.1)
Weight (kg)	66.5 (12.2)	69.3 (10.4)	58.4 (10.0)
BSA (m²)	653.5 (81.9)	665.2 (58.5)	595.3 (59.2)
BMCTrunk (g)	732.0 (216.8)	785.0 (257.5)	652.0 (184.5)
FMTrunk (g)	3376.0 (1902.9)	3787.5 (1426.8)	4073.3 (3304.5)
LMTrunk (g)	26912.0 (6231.7)	28477.5 (4213.5)	22220.0 (4101.3)
FMTotal (g)	9021.0 (4824.4)	7977.4 (3292.5)	11450.0 (6047.7)
LMTotal (g)	54339.4 (12709.8)	57057.4 (8290.8)	44725.0 (8168.8)

10.3.2 Correlation and regression analysis

We quantified the linear correlation between each of the features and the LVM on the entire dataset using the Pearson correlation coefficient, ρ .

To identify the predictive power of each feature we performed a univariate linear regression analysis. The subjects were divided into a training (70%) and a test (30%) set. The training set was used to build the model, while the performance was assessed on the test set. For statistical robustness, the whole process was repeated 100 times.

10.3.3 Feature selection

Features with a non-significant ($\rho < 0.19$) correlation with the LVM were removed for further analysis. Hereafter, we created four feature subsets. The first subset consists solely of ECG features (ECG). The second subset is comprised of both ECG and demographic features (ECG+Demo). The third subset consists of both ECG and DXA features (ECG+DXA) and the last subset contains all available features (All).

The selection of the most relevant features from the different subsets was carried out using the least absolute shrinkage and selection operator (LASSO) algorithm [180]. This is a particular case of the penalized least squares regression with L1-penalty function. It has one major drawback however: if there is a group of highly correlated variables, LASSO tends to randomly select only one variable from that group [208]. Therefore, we performed the LASSO algorithm 100 times and sorted each feature based on majority voting.

10.3.4 Development of LVM estimation model

Since we have only 107 subjects, we opted to keep the model complexity low. Therefore, we constructed a linear SVM regression model and used only the five highest ranked features from every subset. The latter to avoid overfitting.

To be able to statistically compare the 4 models, we used the same approach as for the univariate regression analysis. We randomly split the subjects 100 times into a training (70%) and test (30%) set. These sets were then used to train and test the model. The hyperparameters were automatically tuned using 5-fold cross-validation and Bayesian optimization.

The coefficient of determination, R^2 , and the root-mean-squared-error (RMSE) were used as statistics to measure the goodness-of-fit of the models. The

median and interquartile range of the performance metrics are always reported. Depending on the normality, we tested the difference between models with a one-way ANOVA or with a Kruskal-Wallis test. Normality was tested with the Lilliefors test and a p-value <0.05 was considered significant. All analysis were performed using *MATLABR2017a*.

10.4 Results

The highest correlated variables ($\rho > 0.5$) are SLP, age, sex, height, weight and LMTrunk, LMTotal. In contrast, RaVL, SaVL, SIII, FMTrunk and FMTotal are not significantly correlated to the LVM. Hence, these features were removed before the feature selection (Table 10.3).

The univariate regression analysis showed that ECG derived features estimate LVM rather poorly, compared to features derived from other modalities. The best performing ECG features are SLV and SLP. The five highest ranked features by the LASSO algorithm are SLP, Age, Sex, Weight and LMTrunk. All of these features were selected at least 83 out of 100 times. The resulting RMSE of the regression models of each of these features separately ranged between 26.66-35.46 g.

The performance of all four models are shown in Figure 10.1. The best performing model was constructed with the 5 highest ranked features from all modalities ($R^2 = 0.67$ (0.14), RMSE = 23.08 (4.42) g). The model outperformed the ECG and ECG+DXA based models, but did not differ from the ECG+Demo model. The ECG based model performed significantly worse compared to all other models ($R^2 = 0.28$ (0.17), RMSE = 34.33 (5.63) g).

The ECG+Demo ($R^2 = 0.65$ (0.10), RMSE = 23.56 (3.65) g) and the ECG+DXA model ($R^2 = 0.61$ (0.14), RMSE = 25.21 (4.10) g) performed similarly.

10.5 Discussion

The first objective was to identify features which correlate well with LVM. LMTrunk and LMTotal correlate best ($\rho > 0.7$) and have the best goodness of fit ($R^2 > 0.5$, RMSE < 28 g). This was expected, since the cardiovascular system has to efficiently distribute oxygen to metabolic active tissue such as lean muscle mass, especially during exercise [53].

Table 10.3: Correlation and regression analysis for each feature. ρ = correlation coefficient; R^2 = coefficient of determination; RMSE = root mean squared error; # LASSO = amount of times each feature was selected by LASSO when taking all features into account.

	ρ	R^2 (IQR)	RMSE (g) (IQR)	# LASSO
SV1	0.26	0.09 (0.12)	37.30 (4.92)	1
SV3	0.32	0.12 (0.09)	35.96 (5.65)	1
RV5	0.43	0.20 (0.17)	34.96 (5.87)	73
RV6	0.24	0.08 (0.13)	37.55 (5.44)	1
RI	0.20	0.05 (0.08)	37.67 (5.25)	1
RaVL	0.04	0.01 (0.03)	38.33 (5.21)	
SaVL	-0.01	0.01 (0.04)	38.45 (4.11)	
SIII	0.09	0.02 (0.04)	38.02 (5.32)	
QRS	0.44	0.21 (0.19)	35.32 (4.74)	32
CV	0.32	0.11 (0.08)	35.94 (5.62)	0
CP	0.37	0.15 (0.10)	35.28 (5.65)	10
SLV	0.47	0.23 (0.14)	34.07 (4.37)	0
SLP	0.54	0.30 (0.19)	32.49 (4.15)	83
Age	0.43	0.19 (0.13)	35.46 (5.78)	95
Sex	-0.61	0.30 (0.15)	30.56 (4.55)	93
Height	0.58	0.33 (0.15)	31.65 (3.88)	2
Weight	0.68	0.46 (0.17)	28.74 (4.26)	86
BSA	0.69	0.48 (0.14)	28.54 (4.16)	7
BMCTrunk	0.60	0.32 (0.18)	31.70 (5.34)	73
FMTrunk	0.09	0.01 (0.04)	38.12 (5.10)	
LMTrunk	0.73	0.55 (0.16)	26.66 (4.66)	99
FMTTotal	-0.04	0.01 (0.03)	38.15 (5.33)	
LMTTotal	0.72	0.53 (0.17)	27.30 (4.84)	23

Additionally, this high correlation has also been observed in other studies [53, 63]. Since the metabolic demand of FM is rather low, this could also explain the poor correlation between FM and LVM.

ECG criteria correlate poorly with LVM ($R^2 < 0.3$). The SLV and SLP are the best predicting ECG features with a median R^2 of respectively 0.23 and 0.30.

The best performing model used all possible features. The five highest ranked features which were used for the construction of the model are LMTrunk, Age, Sex, Weight and SLP. The presence of features from all modalities indicates that all modalities contribute for an accurate estimation of LVM. The added value of Sex with LMTrunk has previously been observed where a higher LVM

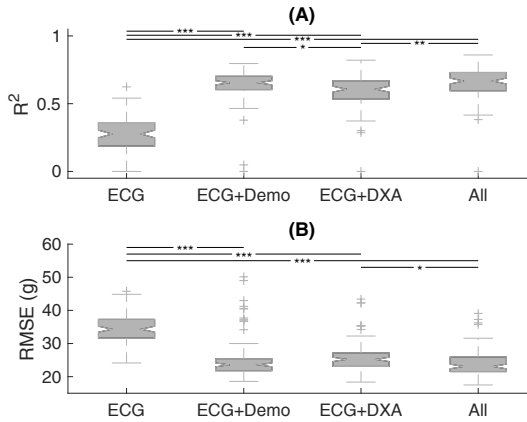


Figure 10.1: (A) R^2 and (B) RMSE of all 4 models. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

is measured in boys compared to girls [63]. However, sex-specific predictive equations were used for the quantification of lean body mass. We performed a more accurate assessment of LVM using a DXA-scan.

One limitation of this study is the limited number of data points. This prevented the creation of an independent test set. In this study we used all data points for feature selection and again for the learning algorithm design. Ideally, those steps should be performed separately to prevent overfitting. Future research is needed to validate this model in a larger cohort of athletes, as well as to see whether such a model could help differentiate mild concentric left ventricular hypertrophy from physiological adaptations in athletes.

Furthermore the added value of other modalities should be assessed. Maximal oxygen consumption (VO_{2max}) as measure by cardiac pulmonary exercise testing has shown strong correlation with LVM with an R^2 of 0.71 [92], hence could serve as a potential add-on in predictive models for LVM.

10.6 Conclusion

An ECG-based regression model poorly estimates LVM in endurance athletes. Adding demographic and body composition features significantly improves the model enabling accurate estimation of LVM. Further research is needed to firstly

validate the model and secondly assess its discriminative performance between healthy athletes and those with LVH.

Chapter 11

Conclusions and future directions

In the introduction, we defined three major research objectives for this manuscript. In this final Chapter we summarized the conclusions related to each of these goals and provided suggestions for future research directions.

11.1 Conclusions

Ambulatory monitoring allows to monitor a patient's vital signs, such as heart rate, outside of the hospital, during a longer period of time. This extension of the analysis period highly increases the detection rate of cardiac events, compared to a cardiac exam in the hospital. However, it often yields two problems:

- low quality of the signals,
- no medical expert to interpret the data.

Since most data driven support tools assume clean data, it is of utmost importance that an indication is given of the quality of the signals. In this work, we developed quality indication algorithms for cardiac and respiratory signals which could aid in cardiac-, respiratory- and cardiorespiratory analysis.

Hereafter, we developed a toolbox that can be used by researchers, engineers and medical doctors alike, to detect heartbeats, visualize these together with the raw signal, and provide user friendly ways to correct possible miss annotations. Moreover, it facilitates the correct annotation of medical data which is of essential importance for more advanced analysis techniques.

An example of where this is important is in the investigation of the cardiorespiratory coupling. In this work, we used the previously described quality indication algorithm and R-peak detection and correction tool to investigate the strength of the cardiorespiratory coupling during exercise. The presented pipeline could be used similarly for other applications (Figure 11.1).

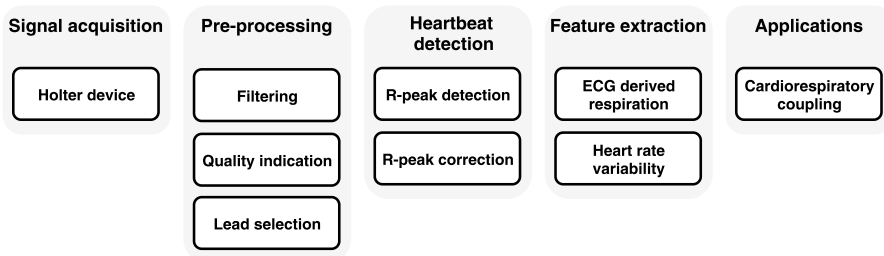


Figure 11.1: Cardiorespiratory analysis pipeline.

The manuscript was organized around three main objectives:

1. Accurate artefact detection and quality indication of wearable biomedical signals.
2. Develop user friendly tools that aid in analysing ECG signals.
3. Explore the added value of ECG and cardiorespiratory analysis in sport applications.

The methods presented in the previous Chapters all contributed to one or more of these research objectives. The next Sections summarize the major findings related to each objective and list the most important conclusions.

11.1.1 Accurate artefact detection and quality indication of wearable biomedical signals

The first objective of this PhD was the development of algorithms that are able to accurately distinguish clean from noisy wearable signals. In Chapters 4 and 5 we focussed on ECG signals and in Chapter 6, we focussed on respiratory signals.

Chapter 4 can be divided in two parts: artefact detection and quality indication. First, we proposed an algorithm for ECG artefact detection, based on features derived from the ACF and a RUSBoost classifier. In order to evaluate this algorithm, we annotated a large dataset, consisting of two smaller datasets, with four annotators. We trained and tested the algorithm on these two datasets and, additionally, also on a dataset that was annotated by a medical doctor. The obtained results show that the proposed artefact detection algorithm is able to accurately distinguish clean from noisy signals, regardless of the dataset it was trained on. Although, as expected, the prediction is more accurate when we include signals from the other datasets as well. The obtained performance is similar, compared to the machine learning algorithm proposed in [41].

One of the advantages of the proposed method is that only three features are needed to obtain an accurate performance, compared to the seven features in [41]. Moreover, it does not require the detection of R-peaks as in [143], which is computationally a lot more demanding than the computation of the ACF. An additional advantage of the use of the ACF is the robustness towards different recording devices. Algorithms, such as the one described in [202], heavily rely on amplitude based features which are affected by the properties of the recording device. The obtained results indicate that this is not the case for our algorithm.

The main novelty within Chapter 4 is the new approach to ECG signal quality assessment. We suggested to exploit the posterior class probability of the

RUSBoost classifier and to use the probability for the clean class as a novel quality assessment index. This allows users to identify periods of data with a pre-defined level of quality, depending on the task at hand. We have shown a strong correlation of the proposed metric with the agreement of the annotators on the newly annotated datasets. Additionally, we evaluated the proposed metric on a dataset with known SNR levels. We found that the proposed metric significantly decreased when the SNR decreased. This approach is currently used by other research groups within KU Leuven and other universities, such as TU Delft and Zaragoza.

In Chapter 5, we have shown that the features that we have crafted for contact ECG in Chapter 4 can be used for artefact detection in ccECG signals. This finding allowed us to successfully construct a classifier for contact ECG and optimize it towards ccECG using transfer learning. Moreover, we have shown that we only need a limited amount of ccECG samples to do so. Hence, by using this approach, we circumvent the need to retrain a new classifier from scratch and require only a limited amount of labelled data to do so.

In Chapter 6, we shifted the focus towards respiratory signals, measured with a bio-impedance device. Here also, we attempted to separate clean from noisy signals using a machine learning approach. We compared two different approaches: a feature based SVM and a CNN. The first approach can be regarded as a classical approach where hand-crafted features are fed to a classification model, which is then trained and evaluated on a separate test set. The main challenge for this approach is to craft meaningful features. The second approach circumvents this challenge by learning the features directly from the raw data. Besides an adequate network architecture, the main requirement is a lot of labelled data.

To the best of our knowledge, this is the first study that compares a feature based classifier and a supervised neural network for the classification of respiratory BioZ signals. We have shown that both models accurately distinguish clean from noisy samples and observed no significant differences in performance. This shows that a data-driven approach could be used for the task of artefact detection in respiratory BioZ signals. These algorithms will be shared with imec and, after further validation, be implemented in their analysis pipeline.

11.1.2 Develop user friendly tools that aid in analysing ECG signals

The second part of this work focussed on developing user friendly tools for ECG signal analysis. We can divide the ECG signal analysis pipeline in four major

blocks: signal acquisition, pre-processing, R-peak detection, feature extraction and applications. All blocks could benefit from user friendly tools, but during this PhD, we focussed on the third and fourth block.

R-DECO is a MATLAB based GUI for detecting and correcting R-peaks in ECG signals. The main improvement that we made, compared to existing toolboxes, is that we included user friendly tools to correct wrongly annotated R-peaks. Most of the existing ECG-analysis toolboxes do not provide these and if they do, they are not user friendly. This is because they assume that the R-peaks have been correctly annotated. With this GUI, we intended to overcome this limitation.

The main novelty is the implementation of an interactive drag-and-drop function. This allows the user to select an annotation and drag it to the desired location. Furthermore, we included the possibility to delete multiple annotations at once. This speeds up the process in case the signal is heavily distorted for a long period of time and the annotations are not reliable. The semi-automated correction methods allow for the annotations of the whole signal, or a smaller segment, to be adjusted similarly. For example, we have implemented a correction method based on cross-correlation in which the annotations are corrected based on their correlation with a template heartbeat.

Since some devices have built-in QRS-detection algorithms and some researchers have their own preferred QRS detection algorithm, the software allows to load R-peak locations. These can be corrected as if they were detected by the implemented algorithm.

One of the main added values is that the code of the GUI is open source. This allows users to implement their own favourite algorithms or use the correction functions for different purposes. Additionally, the code could be used as a framework so that similar tools could be built faster in the future.

Concerning BVR analysis, one of the current limitations is the requirement of separate software and dedicated personnel. We used the framework of R-DECO to overcome this limitation and developed a GUI for semi-automated BVR analysis. The main objectives can be summarized as follows: select accurate delineation algorithms, implement these algorithms in a user interface and create a readable output format.

11.1.3 Explore the added value of ECG and cardiorespiratory analysis in sport applications

In Chapter 9, we explored if the strength of the cardiorespiratory coupling is related to CRF parameters during exercise. However, due to a lack of modality specific recordings, we could not fully answer this question. Since we did not have a continuous respiratory signal, we had to derive the respiration from the ECG. In a noise free situation, this EDR is rather accurate, but it is not so much during exercise.

When we incrementally increased the desired correlation between the estimated and measured respiration rate, we could observe that the correlation between \mathcal{P}_x and VO_2 and RER increased during the exercise phase. When we only included the subjects with the best EDR, we found that \mathcal{P}_x follows a similar pattern as the VO_2 and RER during the exercise phase of a submaximal exercise test.

These findings are the first step towards a non-invasive estimation of the physical condition of a person, based on cardio-respiratory interactions. However, as shown in this study, an accurate estimation of the respiration is essential. Hence, the performance of EDR algorithms during exercise needs to be improved before we could take a next step.

In Chapter 10, we used ECG criteria for left ventricular hypertrophy and standard ECG features to construct a regression model for the estimation of LVM. Unfortunately, we came to the conclusion that an ECG-based regression model poorly estimates LVM in endurance athletes. However, we could significantly improve the model performance by adding demographic and body composition features.

11.2 Future directions

We can identify different extensions and future directions of the presented research. Similarly to the previous Section, we have organized this Section by objective and added a general view of how this PhD could contribute to future research.

11.2.1 Accurate artefact detection and quality indication of wearable biomedical signals

The main drawback of the first study is that the model was trained and tested on a dataset which mostly consisted of normal sinus rhythm. Previous studies have shown that signal quality classification algorithms experience a reduction in performance on an arrhythmia database when it was not explicitly retrained using signals containing arrhythmia episodes [102]. We expect the model to be only marginally influenced by abnormal rhythms, but more affected by abnormalities which change the shape of the heartbeat significantly, e.g. pre-mature ventricular contractions. This hypothesis needs to be further investigated.

The proposed methodology uses only three features. This is an advantage for ambulatory systems, since the computation time of the algorithm is very low. However, all three features somehow try to characterize changes in the QRS-complex, which possibly neglects other characteristic ECG waves. A straightforward extension of this work would be the investigation of the added value of other features. This would be at the cost of computational speed, but could alleviate the performance in more challenging environments. For example, heart failure patients that experience a lot of abnormal heartbeats.

Instead of adding new features, we could also circumvent feature crafting as a whole by mimicking the same approach as in Chapter 6. For example, we could use an auto-encoder to perform unsupervised artefact detection. Besides feature crafting, this also eliminates the need for labelled data. In [164], they obtained promising results with this approach. Additionally, they also proposed to modify such auto-encoders towards the task of unsupervised quality assessment in order to have a continuous score.

Given the results in Chapter 6, the main drawback is the size of the dataset and the fact that it consists only of COPD patients. We are currently working on extending the dataset to a new patient group, which would allow more training and evaluation options. For instance, we could investigate whether the disease has an influence on the proposed algorithms and whether the neural network based approaches benefit from the availability of more data. The algorithms that were developed in this Chapter, will be shared with imec and the goal is to implement it in their analysis pipeline.

11.2.2 Develop user friendly tools that aid in analysing ECG signals

R-DECO is the first step towards a complete ECG analysis toolbox. At this moment, it focusses on accurate R-peak detection and intuitive correction options. This makes it a complementary tool for existing HRV-analysis toolboxes, which tend to focus on the computation of HRV metrics from RR-intervals.

Several extra features could be added in future versions. We intend to add support for other input file formats, such as Hierarchical Data Format 5 (HF) files, General Data Format (GDF) files, etc. Additionally, we also intend to include the algorithms developed in Part II, so that the user can detect and remove signal segments that contain artefacts. Hereafter, we could include tools to derive HRV metrics and include ECG derived respiration algorithms.

The ultimate goal is to construct an ECG toolbox that can be used for pre-processing, quality indication, R-peak detection, EDR computation and feature extraction. The latter aimed at HRV and cardiorespiratory metrics.

The largest bottleneck from the current toolbox is the fact that it is written in MATLAB. Since this is a paid program, this excludes users that do not have access to it. Hence, in order to allow everyone to use the toolbox, it should be rewritten in Python.

11.2.3 Explore the added value of ECG and cardiorespiratory analysis in sport applications

The main extension to the cardiorespiratory analysis is to try to confirm the preliminary findings on a dataset where a continuous respiration signal is recorded. This would allow for an accurate derivation of \mathcal{P}_x and a fair investigation of the physiological value of this parameter during exercise.

The model that was developed in Chapter 10 was our first attempt to construct a regression model for LVM estimation using regularly measured features. In order for this model, or an extension of this model, to be used in clinical practice we first need to validate it on a larger cohort. Hereafter, we can investigate if it is accurate enough to discriminate between healthy athletes and those with LVH. Another extension to this work would be to really mine the ECG and use a neural network approach to extract features from it.

11.2.4 General

New wearable devices open up a world of possibilities to improve health and well-being in the general population. Moreover, with the exponential increase in wearable technology, there might be a 'tsunami' of medical data coming. Hence, a way of quantifying the signal quality level is necessary to fully exploit the potential of these wearable devices. In this PhD, I have made a first attempt to develop artefact detection and quality indication algorithms for ambulatory signals using a very simple and computationally inexpensive transformation technique.

I hope that my findings pave the way for other researchers to continue on this path, so that the impact of these new wearable technologies on healthcare applications can be maximised.

Bibliography

- [1] Ats/ers statement on respiratory muscle testing. *American Journal of Respiratory and Critical Care Medicine* 166, 4 (2002), 518–624.
- [2] AMONI, M., VANDENBERK, B., MOEYERSONS, J., VAN HUFFEL, S., SIPIDO, K., AND WILLEMS, R. Temporal beat-to-beat variability of repolarization (BVR) changes predict imminent non-sustained ventricular tachycardia in ischaemic heart disease patients. In *European Heart Rhythm Association* (2018), vol. 20.
- [3] ANREP, G., PASCUAL, W., AND RÖSSLER, R. Respiratory variations of the heart rate—ii—the central mechanism of the respiratory arrhythmia and the inter-relations between the central and the reflex mechanisms. *Proceedings of the Royal Society of London. Series B-Biological Sciences* 119, 813 (1936), 218–230.
- [4] ANSARI, A. H., J., C., CAICEDO, A., JANSEN, K., DEREYMAEKER, A., DE WISPELAERE, L., DIELMAN, C., VERVISCH, J., GOVAERT, P., DE VOS, M., NAULAERS, G., AND VAN HUFFEL, S. Weighted performance metrics for automatic neonatal seizure detection using multi-scored eeg data. *IEEE Journal of Biomedical and Health Informatics* 22, 4 (2018), 1114–1123.
- [5] ANSARI, S., WARD, K., AND NAJARIAN, K. Epsilon-tube filtering: reduction of high-amplitude motion artifacts from impedance plethysmography signal. *IEEE Journal of Biomedical and Health Informatics* 19, 2 (Mar 2015), 406–417.
- [6] ARMAÑAC, P., VARÓN, C., HERNANDO, D., DE HARO, C., MAGRANS, R., ORINI, M., MORALES, J., MOEYERSONS, J., SARLABOUS, L., LÓPEZ-AGUILAR, J., LAGUNA, P., GIL, E., BAILÓN, R., AND BLANCH, L. Cardiopulmonary coupling indices to improve weaning readiness criteria. *Scientific Reports PP*.

- [7] ARZENO, N. N., DENG, Z. D., AND POON, C. S. Analysis of first-derivative based qrs detection algorithms. *IEEE Transactions on Biomedical Engineering* 55, 2 (2008), 478–484.
- [8] ASKANAZI, J., SILVERBERG, P. A., FOSTER, R. J., HYMAN, A. I., MILIC-EMILI, J., AND KINNEY, J. M. Effects of respiratory apparatus on breathing pattern. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*. 48, 4 (1980), 577–580.
- [9] ASTRAND, P.-O., CUDDY, T. E., SALTIN, B., AND STENBERG, J. Cardiac output during submaximal and maximal work. *Journal of Applied Physiology* 19, 2 (1964), 268–274.
- [10] AZUAJE, F., CLIFFORD, G., AND MCSHARRY, P. *Advanced Methods and Tools for ECG Data Analysis*. Artech, 2006.
- [11] BADILINI, F. The ishne holter standard output file format. *Annals of Noninvasive Electrocardiology* 3, 3 (1998), 263–266.
- [12] BAILEY, J. J., BERSON, A. S., GARSON, A., HORAN, L. G., MACFARLANE, P. W., MORTARA, D. W., AND ZYWIETZ, C. Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association. *Circulation* 81, 2 (Feb 1990), 730–739.
- [13] BAILON, R., LAOUINI, G., GRAO, C., ORINI, M., LAGUNA, P., AND MESTE, O. The integral pulse frequency modulation model with time-varying threshold: application to heart rate variability analysis during exercise stress testing. *IEEE Transactions on Biomedical Engineering* 58, 3 (Mar 2011), 642–652.
- [14] BALADY, G. J., ARENA, R., SIETSEMA, K., MYERS, J., COKE, L., FLETCHER, G. F., FORMAN, D., FRANKLIN, B., GUAZZI, M., GULATI, M., KETEVIAN, S. J., LAVIE, C. J., MACKO, R., MANCINI, D., AND MILANI, R. V. Clinicians guide to cardiopulmonary exercise testing in adults. *Circulation* 122, 2 (2010), 191–225.
- [15] BANSAL, A., AND JOSHI, R. Portable out-of-hospital electrocardiography: A review of current technologies. *Journal of Arrhythmia* 34, 2 (04 2018), 129–138.
- [16] BEN-TAL, A., SHAMAILOV, S. S., AND PATON, J. F. Evaluating the physiological significance of respiratory sinus arrhythmia: looking beyond

- ventilation-perfusion efficiency. *Journal of Physiology* 590, 8 (Apr 2012), 1989–2008.
- [17] BENJAMIN, E. J., MUNTNER, P., ALONSO, A., BITTENCOURT, M. S., CALLAWAY, C. W., CARSON, A. P., AND ET AL., A. M. C. Heart disease and stroke statistics update: A report from the american heart association. *Circulation* 139, 10 (2019), e56–e528.
- [18] BERGER, R. D., KASPER, E. K., BAUGHMAN, K. L., MARBAN, E., CALKINS, H., AND TOMASELLI, G. F. Beat-to-beat QT interval variability: novel evidence for repolarization lability in ischemic and nonischemic dilated cardiomyopathy. *Circulation* 96, 5 (Sep 1997), 1557–1565.
- [19] BLAIR, S. N., KAMPERT, J. B., KOHL, H. W., BARLOW, C. E., MACERA, C. A., PAFFENBARGER, R. S., AND GIBBONS, L. W. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *Journal of the American Medical Association* 276, 3 (Jul 1996), 205–210.
- [20] BLAIR, S. N., AND MORRIS, J. N. Healthy hearts—and the universal benefits of being physically active: physical activity and health. *Annals of Epidemiology* 19, 4 (Apr 2009), 253–256.
- [21] BLANCO-ALMAZAN, D., GROENENDAAL, W., LOZANO-GARCIA, M., ESTRADA-PETROCELLI, L., LIJNEN, L., SMEETS, C., RUTTENS, D., CATTHOOR, F., AND JANE, R. Combining bioimpedance and myographic signals for the assessment of COPD during loaded breathing. *IEEE Transactions on Biomedical Engineering* (2020), 298 – 307.
- [22] BLANCO-ALMAZÁN, D., GROENENDAAL, W., CATTHOOR, F., AND JANÉ, R. Chest movement and respiratory volume both contribute to thoracic bioimpedance during loaded breathing. *Scientific Reports* 9 (12 2019), 20232.
- [23] BLANCO-ALMAZÁN, D., GROENENDAAL, W., CATTHOOR, F., AND JANÉ, R. Wearable bioimpedance measurement for respiratory monitoring during inspiratory loading. *IEEE Access* 7 (2019), 89487–89496.
- [24] BLOMQUIST, C. G., AND SALTIN, B. Cardiovascular adaptations to physical training. *Annual Review of Physiology* 45 (1983), 169–189.
- [25] BOX, G., AND JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

- [26] BRODERSEN, K. H., ONG, C. S., STEPHAN, K. E., AND BUHMANN, J. M. The balanced accuracy and its posterior distribution. In *International Conference on Pattern Recognition* (2010), vol. 20, pp. 3121–3124.
- [27] BRODERSEN, K. H., ONG, C. S., STEPHAN, K. E., AND BUHMANN, J. M. The balanced accuracy and its posterior distribution. In *International Conference on Pattern Recognition* (2010), vol. 20, pp. 3121–3124.
- [28] BURTON, D. A., STOKES, K., AND HALL, G. M. Physiological effects of exercise. *Continuing Education in Anaesthesia Critical Care and Pain* 4, 6 (12 2004), 185–188.
- [29] BUTTAR, H. S., LI, T., AND RAVI, N. Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation. *Experimental & Clinical Cardiology* 10, 4 (2005), 229–249.
- [30] CASTRO, I., PATEL, A., DEVIANE, M., HUYSMANS, D., BORZÉE, P., BUYSE, B., TESTELMANS, D., VAN HUFFEL, S., VARON, C., AND TORFS, T. Unobtrusive, through-clothing eeg and bioimpedance monitoring in sleep apnea patients. *Computing in Cardiology* 47 (2020).
- [31] CASTRO, I., VARON, C., MOEYERSONS, J., VILLA, A., MORALES, J., DEVAIENE, M., TORFS, T., VAN HUFFEL, S., PUERS, R., AND VAN HOOFF, C. Data quality assessment of capacitively-coupled eeg signals. In *Computing in Cardiology* (2019), IEEE, pp. 1–4.
- [32] CASTRO, I. D., MERCURI, M., PATEL, A., PUERS, R., VAN HOOFF, C., AND TORFS, T. Physiological driver monitoring using capacitively coupled and radar sensors. *Applied Sciences* 9, 19 (2019).
- [33] CASTRO, I. D., MERCURI, M., TORFS, T., LORATO, I., PUERS, R., AND VAN HOOFF, C. Sensor fusion of capacitively coupled eeg and continuous-wave doppler radar for improved unobtrusive heart rate measurements. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8, 2 (2018), 316–328.
- [34] CASTRO, I. D., PATEL, A., TORFS, T., PUERS, R., AND VAN HOOFF, C. Capacitive multi-electrode array with real-time electrode selection for unobtrusive eeg and bioz monitoring. In *IEEE Engineering in Medicine and Biology Society* (2019), vol. 41, pp. 5621–5624.
- [35] CASTRO, I. D., VARON, C., TORFS, T., VAN HUFFEL, S., PUERS, R., AND VAN HOOFF, C. Evaluation of a multichannel non-contact eeg system and signal quality algorithms for sleep apnea detection and monitoring. *Sensors* (2018).

- [36] CHARLTON, P. H., BONNICI, T., TARASSENKO, L., CLIFTON, D. A., BEALE, R., WATKINSON, P. J., AND ALASTRUEY, J. An impedance pneumography signal quality index: Design, assessment and application to respiratory rate monitoring. *Biomedical Signal Processing and Control* 65 (2021), 102339.
- [37] CHAUHAN, V. S., KRAHN, A. D., WALKER, B. D., KLEIN, G. J., SKANES, A. C., AND YEE, R. Sex differences in QTc interval and QT dispersion: dynamics during exercise and recovery in healthy subjects. *American Heart Journal* 144, 5 (Nov 2002), 858–864.
- [38] CHAWLA, M. Pca and ica processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison. *Applied Soft Computing* 11, 2 (2011), 2216 – 2226.
- [39] CHEN, S.-W., CHEN, H.-C., AND CHAN, H.-L. A real-time qrs detection method based on moving-averaging incorporating with wavelet denoising. *Computer Methods and Programs in Biomedicine* 82, 3 (2006), 187–195.
- [40] CHIDNOK, W., DIMENNA, F. J., BAILEY, S. J., BURNLEY, M., WILKERSON, D. P., VANHATALO, A., AND JONES, A. M. .VO2max is not altered by self-pacing during incremental exercise. *European Journal of Applied Physiology* 113, 2 (Feb 2013), 529–539.
- [41] CLIFFORD, G. D., BEHAR, J., LI, Q., AND REZEK, I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological Measurement* 33, 9 (2012), 1419–1433.
- [42] CLIFFORD, G. D., LIU, C., MOODY, B., LEHMAN, L. H., SILVA, I., LI, Q., JOHNSON, A. E., AND MARK, R. G. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In *Computing in Cardiology* (2017), pp. 1–4.
- [43] CLIFFORD, G. D., LOPEZ, D., LI, Q., AND REZEK, I. Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments. *Computing in Cardiology* (2011), 285–288.
- [44] CORTES, C., AND VAPNIK, V. *Support-Vector Networks*, vol. 20. Kluwer Academic Publishers, USA, 1995.
- [45] COUDERC, J.-P., AND ZAREBA, W. *Assessment of Ventricular Repolarization From Body-Surface ECGs in Humans*. 01 2005, pp. 107–129.

- [46] DAGRES, N., KOTTKAMP, H., PIORKOWSKI, C., WEIS, S., ARYA, A., SOMMER, P., BODE, K., GERDS-LI, J., KREMASTINOS, D., AND HINDRICKS, G. Influence of the duration of holter monitoring on the detection of arrhythmia recurrences after catheter ablation of atrial fibrillation: Implications for patient follow-up. *International journal of cardiology* 139, 3 (2010), 305–306.
- [47] DALUWATTE, C., JOHANNESSEN, L., GALEOTTI, L., VICENTE, J., STRAUSS, D., AND SCULLY, C. Assessing ecg signal quality indices to discriminate ecgs with artefacts from pathologically different arrhythmic ecgs. *Physiological Measurement* 37, 8 (2016), 1370–1382.
- [48] D’ANDREA, A., LIMONGELLI, G., CASO, P., SARUBBI, B., DELLA PIETRA, A., BRANCACCIO, P., CICE, G., SCHERILLO, M., LIMONGELLI, F., AND CALABRÓ, R. Association between left ventricular structure and cardiac performance during effort in two morphological forms of athlete’s heart. *International Journal of Cardiology* 86, 2-3 (Dec 2002), 177–184.
- [49] DE BRABANTER, K., DE BRABANTER, J., SUYKENS, J., AND DE MOOR, B. *Optimized Fixed-Size Kernel Models for Large Data Sets*, vol. 54. Elsevier Science Publishers B. V., NLD, jun 2010.
- [50] DE CHAZAL, P., HENEGHAN, C., SHERIDAN, E., REILLY, R., NOLAN, P., AND O’MALLEY, M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering* 50, 6 (June 2003), 686–696.
- [51] DE COOMAN, T., VANDECASTEELE, K., VARON, C., HUNYADI, B., CLEEREN, E., VAN PAESSCHEN, W., AND VAN HUFFEL, S. Personalizing heart rate-based seizure detection using supervised svm transfer learning. *Frontiers in Neurology* 11 (2020), 145.
- [52] DE COOMAN, T., VARON, C., VAN DE VEL, A., CEULEMANS, B., LAGAE, L., AND VAN HUFFEL, S. Semi-supervised one-class transfer learning for heart rate based epileptic seizure detection. In *Computing in Cardiology* (2017), pp. 1–4.
- [53] DEWEY, F. E., ROSENTHAL, D., MURPHY, D. J., FROELICHER, V. F., AND ASHLEY, E. A. Does size matter? Clinical applications of scaling cardiac size and function for body size. *Circulation* 117, 17 (Apr 2008), 2279–2287.
- [54] DEYELL, M. W., KRAHN, A. D., AND GOLDBERGER, J. J. Sudden cardiac death risk stratification. *Circulation research* 116, 14 (2015), 1907–1918.

- [55] DHUTIA, H., AND MACLACHLAN, H. Cardiac Screening of Young Athletes: a Practical Approach to Sudden Cardiac Death Prevention. *Current Treatment Options Cardiovascular Medicine* 20, 10 (Aug 2018), 85.
- [56] DOHARE, A. K., KUMAR, V., AND KUMAR, R. An efficient new method for the detection of qrs in electrocardiogram. *Computers and Electrical Engineering* 40, 5 (2014), 1717–1730.
- [57] EASTWOOD, P. R., HILLMAN, D., AND FINUCANE, K. E. Ventilatory responses to inspiratory threshold loading and role of muscle fatigue in task failure. *Journal of Applied Physiology* 76, 1 (1985), 185–195.
- [58] EBRAHIM, S., TAYLOR, F., WARD, K., BESWICK, A., BURKE, M., AND DAVEY SMITH, G. Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane Database Systematic Reviews* 19, 1 (Jan 2011).
- [59] EKBLUM, B., AND HERMANSEN, L. Cardiac output in athletes. *Journal of Applied Physiology* 25, 5 (1968), 619–625. PMID: 4879852.
- [60] EKELUND, U., TARP, J., STEENE-JOHANNESSEN, J., HANSEN, B. H., JEFFERIS, B., FAGERLAND, M. W., WHINCUP, P., DIAZ, K. M., HOOKER, S. P., CHERNOFSKY, A., LARSON, M. G., SPARTANO, N., VASAN, R. S., AND DOHRN, I. M. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *British Medical Journal* 366 (08 2019), l4570.
- [61] ELGENDI, M., ESKOFIER, B., DOKOS, S., AND ABBOTT, D. Revisiting qrs detection methodologies for portable, wearable, battery-operated, and wireless ecg systems. *PLOS ONE* 9, 1 (01 2014), 1–18.
- [62] FICK, A. Uber die blutquantums in den herzventrikeln. *Sitzungsberichte der Physikalisch-medicinischen Gesellschaft Wurzburg* 16 (1870).
- [63] FOSTER, B. J., KHOURY, P. R., KIMBALL, T. R., MACKIE, A. S., AND MITSNEFES, M. New Reference Centiles for Left Ventricular Mass Relative to Lean Body Mass in Children. *Journal of the American Society of Echocardiography* 29, 5 (05 2016), 441–447.
- [64] FREUND, Y., AND SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55, 1 (1997), 119–139.

- [65] FUJII, T., NAKANO, M., YAMASHITA, K., KONISHI, T., IZUMI, S., KAWAGUCHI, H., AND YOSHIMOTO, M. Noise-tolerant instantaneous heart rate and R-peak detection using short-term autocorrelation for wearable healthcare systems. vol. 35, pp. 7330–7333.
- [66] GARCIA, A. J., KOSCHNITZKY, J. E., DASHEVSKIY, T., AND RAMIREZ, J. M. Cardiorespiratory coupling in health and disease. *Autonomic Neuroscience* 175, 1-2 (Apr 2013), 26–37.
- [67] GASHLER, M., GIRAUD-CARRIER, C. G., AND MARTINEZ, T. R. Decision tree ensemble: Small heterogeneous is better than large homogeneous. *International Conference on Machine Learning and Applications* 7 (2008), 900–905.
- [68] GILMOUR, R. J., RICCIO, M., LOCATI, E., MAISON-BLANCHE, P., COUMEL, P., AND SCHWARTZ, P. Time- and rate-dependent alterations of the QT interval precede the onset of torsade de pointes in patients with acquired QT prolongation. *Journal of the American College of Cardiology* 30, 1 (1997), 209–217.
- [69] GIROLAMI, M. Orthogonal series density estimation and the kernel eigenvalue problem. *Neural Computation* 14, 3 (Mar 2002), 669–688.
- [70] GLEDHILL, N., COX, D., AND JAMNIK, R. Endurance athletes' stroke volume does not plateau: major advantage is diastolic function. *Med & Science in Sports & Exercise* 26, 9 (Sep 1994), 1116–1121.
- [71] GOLDBERGER, E. A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. *American Heart Journal* 23, 4 (1942), 483–492.
- [72] GONZÁLEZ, F., ALCARAZ, R., AND RIETA, J. J. The physionet qt database: Study on the reliability of p-wave manual annotations under noisy recordings. In *Computing in Cardiology* (2017), pp. 1–4.
- [73] GUPTA, A. K. Respiration rate measurement based on impedance pneumography. *Data Acquisition Products, Texas Instruments, Application Report, SBAA181* (2011).
- [74] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [75] HAYANO, J., YASUMA, F., OKADA, A., MUKAI, S., AND FUJINAMI, T. Respiratory sinus arrhythmia. A phenomenon improving pulmonary gas exchange and circulatory efficiency. *Circulation* 94, 4 (Aug 1996), 842–847.

- [76] HE, T., CLIFFORD, G., AND TARASSENKO, L. Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing and Applications* 15 (04 2006), 105–116.
- [77] HUYSMANS, D., SMETS, E., DE RAEDT, W., VAN HOOF, C., BOGAERTS, K., DIEST, I., AND HELIC, D. Unsupervised learning for mental stress detection - exploration of self-organizing maps. *Conference on Biomedical Engineering Systems and Technologies 4* (2018), 26–35.
- [78] IMTIAZ, S., MARDELL, J., AND S. SAREMI-YARAHMADI, E. R.-V. Ecg artefact identification and removal in mhealth systems for continuous patient monitoring. *Healthcare Technology Letters* 3, 3 (2016), 171–176.
- [79] JACOBS, B., MOEYERSONS, J., VILLA, A., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Can Laplacian Eigenmaps be used for differentiation between healthy subjects and patients with corrected Tetralogy of Fallot? *Computing in Cardiology* 47 (2020).
- [80] JANE, R., BLASI, A., GARCIA, J., AND LAGUNA, P. Evaluation of an automatic threshold based detector of waveform limits in holter ecg with the qt database. In *Computers in Cardiology* (1997), pp. 295–298.
- [81] JEONG, D., AND KIM, S. Development of a technique for cancelling motion artifact in ambulatory ecg monitoring system. In *International Conference on Convergence and Hybrid Information Technology* (2008), vol. 3, pp. 954–961.
- [82] JOHNS, D. P., WALTERS, J. A., AND WALTERS, E. H. Diagnosis and early detection of copd using spirometry. *Journal of thoracic disease* 6, 11 (2014), 1557–1569.
- [83] JOYNER, M. J., AND GREEN, D. J. Exercise protects the cardiovascular system: effects beyond traditional risk factors. *Journal of Physiology* 587, Pt 23 (Dec 2009), 5551–5558.
- [84] KARJALAINEN, J., KUJALA, U. M., KAPRIO, J., SARNA, S., AND VIITASALO, M. Lone atrial fibrillation in vigorously exercising middle aged men: case-control study. *British Medical Journal* 316, 7147 (Jun 1998), 1784–1785.
- [85] KEMP, B., AND OLIVAN, J. European data format ‘plus’ (edf+), an edf alike standard format for the exchange of physiological data. *Clinical Neurophysiology* 114, 9 (2003), 1755–1761.
- [86] KEMP, B., VÄRRI, A., ROSA, A. C., NIELSEN, K. D., AND GADE, J. A simple format for exchange of digitized polygraphic recordings.

- Electroencephalography and Clinical Neurophysiology* 82, 5 (1992), 391–393.
- [87] KINLAY, S., LEITCH, J. W., NEIL, A., CHAPMAN, B. L., HARDY, D. B., AND FLETCHER, P. J. Cardiac event recorders yield more diagnoses and are more cost-effective than 48-hour holter monitoring in patients with palpitations: A controlled clinical trial. *Annals of Internal Medicine* 124(1) (1996), 16–20.
- [88] KLEMPÍŘ, O., KRUPÍČKA, R., BAKŠTEIN, E., AND JECH, R. Identification of microrecording artifacts with wavelet analysis and convolutional neural network: An image recognition approach. *Measurement Science Review* 19, 5 (2019), 222 – 231.
- [89] KOHLER, B. U., HENNIG, C., AND ORGLMEISTER, R. The principles of software qrs detection. *IEEE Engineering in Medicine and Biology Magazine* 21, 1 (Jan 2002), 42–57.
- [90] KOOPMAN, F. A., STOOF, S. P., STRAUB, R. H., VAN MAANEN, M. A., VERVOORDELDONK, M. J., AND TAK, P. P. Restoring the balance of the autonomic nervous system as an innovative approach to the treatment of rheumatoid arthritis. *Molecular Medicine* 17, 9-10 (2011), 937–948.
- [91] KRISTIENSEN, S., HUGAAS, M. S., GOEBEL, V., PLAGEMANN, T., NIKOLAIDIS, K., AND LIESTØL, K. Data mining for patient friendly apnea detection. *IEEE Access* 6 (2018), 74598–74615.
- [92] LA GERCHE, A., BURNS, A. T., TAYLOR, A. J., MACISAAC, A. I., HEIDBÜCHEL, H., AND PRIOR, D. L. Maximal oxygen consumption is best predicted by measures of cardiac size rather than function in healthy adults. *European Journal of Applied Physiology* 112, 6 (Jun 2012), 2139–2147.
- [93] LA GERCHE, A., AND HEIDBÜCHEL, H. Can intensive exercise harm the heart? You can get too much of a good thing. *Circulation* 130, 12 (Sep 2014), 992–1002.
- [94] LAGUNA, P., MARK, R. G., GOLDBERG, A., AND MOODY, G. B. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology* (1997), pp. 673–676.
- [95] LANDIS, J. R., AND KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (1977), 159–174.

- [96] LAVANGA, M., HEREMANS, E., MOEYERSONS, J., BOLLEN, B., JANSEN, K., ORTIBUS, E., NAULAERS, G., VAN HUFFEL, S., AND CAICEDO, A. Maturation of the autonomic nervous system in premature infants: Estimating development based on heart-rate variability analysis. *Frontiers in Physiology* 11, 581250 (2021).
- [97] LAVANGA, M., WEL, O. D., CAICEDO, A., DEVIAENE, M., MOEYERSONS, J., VARON, C., BOLLEN, B., JANSEN, K., ORTIBUS, E., NAULAERS, G., AND VAN HUFFEL, S. The implementation of an apnea-based perinatal stress calculator. In *IEEE Engineering in Medicine and Biology Society* (2019), vol. 41, pp. 6000–6003.
- [98] LAVENEZIANA, P., DI PAOLO, M., AND PALANGE, P. The clinical value of cardiopulmonary exercise testing in the modern era. *European Respiratory Review* 30, 159 (Mar 2021).
- [99] LECUN, Y., BOTTOU, L., BENGIO, Y., AND HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
- [100] LEE, H. J., HWANG, S. H., YOON, H. N., AND LEE, W.K.AND PARK, K. S. Heart rate variability monitoring during sleep based on capacitively coupled textile electrodes on a bed. *Sensors* 15, 2 (2015).
- [101] LI, Q., MARK, R. G., AND CLIFFORD, G. D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter. *Physiological Measurement* 29, 1 (2008), 15–32.
- [102] LI, Q., RAJAGOPALAN, C., AND CLIFFORD, G. D. A machine learning approach to multi-level ecg signal quality classification. *Computer Methods and Programs in Biomedicine* 117, 3 (2014), 435–447.
- [103] LIN, M., CHEN, Q., AND YAN, S. Network in network, 2014.
- [104] LIPSITZ, L. A., HASHIMOTO, F., LUBOWSKY, L. P., MIETUS, J., MOODY, G. B., APPENZELLER, O., AND GOLDBERGER, A. L. Heart rate and respiratory rhythm dynamics on ascent to high altitude. *Heart* 74, 4 (1995), 390–396.
- [105] LIPSKI, J., LARRY, C., JAIME, E., MICHAEL, M., AND SIMON, D. Value of holter monitoring in assessing cardiac arrhythmias in symptomatic patients. *American Journal of Cardiology* 37, 1 (1976), 102 – 107.
- [106] LO, R., CHIA, K., AND HSIA, H. Ventricular tachycardia in ischemic heart disease. *Cardiac Electrophysiology Clinics* 9, 1 (2017), 25–46.

- [107] LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R., AHN, S., ALVARADO, M., ANDERSON, H., ANDERSON, L., ANDREWS, K., ATKINSON, C., BADDOUR, L., BARKER, S., BARTELS, D., BELL, M., AND MEMISH, Z. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the global burden of disease study 2010. *The Lancet* 380 (12 2012), 2095–2128.
- [108] LOZANO-GARCÍA, M., SARLABOUS, L., MOXHAM, J., RAFFERTY, G. F., TORRES, A., JANÉ, R., AND JOLLEY, C. J. Surface mechanomyography and electromyography provide non-invasive indices of inspiratory muscle force and activation in healthy subjects. *Scientific Reports* 8 (2018), 16921.
- [109] LUO, S., AND TOMPKINS, W. J. Experimental study: brachial motion artifact reduction in the ecg. In *Computers in Cardiology* (1995), pp. 33–36.
- [110] MANRIQUEZ, A. I., AND ZHANG, Q. An algorithm for QRS onset and offset detection in single lead electrocardiogram records. *IEEE Engineering in Medicine and Biology Society* (2007), 541–544.
- [111] MARTIN, D. S., COBB, A., MEALE, P., MITCHELL, K., EDSSELL, M., MYTHEN, M. G., GROCOTT, M. P. W., FOR THE XTREME ALPS RESEARCH GROUP, ADAMS, T., BISEKER, L., BOOTH, A., BURDALL, O., COBB, A., CUMPSTEY, A., DAUNCEY, S., EDSSELL, M., FARRANT, J., FEELISCH, M., FERNANDEZ, B., FIRTH, O., GILBERT, E., GRANT, D., GROCOTT, M., HENNIS, P., JACKSON, L., JENNER, W., VAN DER KAAIJ, J., KHOSRAVI, M., KORTEKAAS, E., LEVETT, D., MAHOMED, Z., MARTIN, D., MEALE, P., MILLEDGE, J., MITCHELL, K., MOLE, D., MOSES, O., MYTHEN, M., RIGAT, F., O'DOHERTY, A., SALAM, A., SANBORN, M., SHEPERDIGIAN, A., SHRUBB, F., SIMPSON, J., TALBOT, N., WANDRAG, L., WIJESINGHA, S., WILLIAMSON, W., WOOLLEY, T., AND YOW, HENG, F. T. X. A. R. G. Systemic oxygen extraction during exercise at high altitude. *BJA: British Journal of Anaesthesia* 114, 4 (12 2014), 677–682.
- [112] MARTINEZ, J. P., ALMEIDA, R., OLMOS, S., ROCHA, A. P., AND LAGUNA, P. A wavelet-based ecg delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering* 51, 4 (2004), 570–581.
- [113] MARTINEZ-GONZALEZ, M. A., MARTINEZ, J. A., HU, F. B., GIBNEY, M. J., AND KEARNEY, J. Physical inactivity, sedentary lifestyle and

- obesity in the European Union. *International Journal of Obesity Related Metabolic Disorders* 23, 11 (Nov 1999), 1192–1201.
- [114] MASON, R. E., AND LIKAR, I. A new system of multiple-lead exercise electrocardiography. *American Heart Journal* 71, 2 (1966), 196 – 205.
- [115] MATEO, J., AND LAGUNA, P. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on Biomedical Engineering* 50, 3 (Mar 2003), 334–343.
- [116] MATTEO, D., LONGO, A., AND RIZZI, M. Noisy ecg signal analysis for automatic peak detection. *Information* 10 (01 2019), 35.
- [117] MEYER, D., LEISCH, F., AND HORNİK, K. The support vector machine under test. *Neurocomputing* 55, 1 (2003), 169–186.
- [118] MIJOVIĆ, B., DE VOS, M., GLIGORIJEVIĆ, I., TAELEMAN, J., AND VAN HUFFEL, S. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE Transactions on Biomedical Engineering* 57, 9 (2010), 2188–2196.
- [119] MILANESI, M., MARTINI, N., VANELLO, N., POSITANO, V., SANTARELLI, M. F., PARADISO, R., DE ROSSI, D., AND LANDINI, L. Multichannel techniques for motion artifacts removal from electrocardiographic signals. In *IEEE Engineering in Medicine and Biology Society* (2006), pp. 3391–3394.
- [120] MLYNCZAK, M., AND CYBULSKI, G. Decomposition of the cardiac and respiratory components from impedance pneumography signals. In *Biosignals* (Feb 2017), vol. 4, pp. 26–33.
- [121] MLYNCZAK, M., AND CYBULSKI, G. Motion artifact detection in respiratory signals based on teager energy operator and accelerometer signals. In *IEEE Engineering in Medicine and Biology Society* (Singapore, 2018), H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen, Eds., Springer, pp. 45–48.
- [122] MOE-NILSSEN, R., AND HELBOSTAD, J. L. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics* 37, 1 (2004), 121–126.
- [123] MOEYERSONS, J. Effect of food and meals on ECG dynamics. Master’s thesis, KU Leuven, 2016.

- [124] MOEYERSONS, J., AMONI, M., VAN HUFFEL, S., WILLEMS, R., AND VARON, C. R-DECO: An open-source matlab based graphical user interface for the detection and correction of R-peaks. *PeerJ Computer Science* 5 (2019), e226.
- [125] MOEYERSONS, J., AMONI, M., VANDENBERK, B., VARON, C., SIPIDO, K., VAN HUFFEL, S., AND WILLEMS, R. Temporal beat-to-beat variability of repolarization changes predict non-sustained ventricular tachycardia in ischemic heart disease patients. *Computing in Cardiology* 45 (2018), 1–4.
- [126] MOEYERSONS, J., DE BOSSCHER, R., DAUSIN, C., CLAESSEN, G., LA GERCHE, A., BOGAERT, J., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Predicting left ventricular mass using ECG, demographic and DXA features. *Computing in Cardiology* 47 (2020), 1–4.
- [127] MOEYERSONS, J., MORALES, J., VILLA, A., CASTRO, I., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Supervised svm transfer learning for modality-specific artefact detection in ecg. *Sensors* 21 (2021), 662.
- [128] MOEYERSONS, J., SMETS, E., MORALES, J., VILLA, A., DE RAEDT, W., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., AND VAN HUFFEL, S. Artefact detection and quality assessment of ambulatory ecg signals. *Computer methods and programs in biomedicine* 182 (2019), 105050.
- [129] MOEYERSONS, J., TESTELMANS, D., BUYSE, B., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Evaluation of a continuous ecg quality indicator based on the autocorrelation function. In *Computing in Cardiology* (2018), vol. 45, IEEE, pp. 1–4.
- [130] MOEYERSONS, J., VARON, C., TESTELMANS, D., BUYSE, B., AND VAN HUFFEL, S. Ecg artefact detection using ensemble decision trees. In *Computing in Cardiology* (2017), IEEE, pp. 1–4.
- [131] MOMENI, A., PINCUS, M., AND LIBIEN, J. *Introduction to Statistical Methods in Pathology*. Springer International Publishing, 2018.
- [132] MOODY, B. E. Rule-based methods for ecg quality control. *Computing in Cardiology* (2011), 361–363.
- [133] MOODY, G. B., AND MARK, R. G. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* 20, 3 (May 2001), 45–50.

- [134] MOODY, G. B., MULDROW, W. K., AND MARK, R. G. A noise stress test for arrhythmia detectors. *Computers in Cardiology 1* (1984), 381–384.
- [135] MORALES, J., DEVAEENE, M., MILAGRO, J., TESTELMANS, D., BUYSE, B., BAILÓN, R., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Respiratory sinus arrhythmia in apnea patients with apnea associated comorbidities. In *Computing in Cardiology Conference* (2018), vol. 45, pp. 1–4.
- [136] MORALES, J., MOEYERSONS, J., ARMANAC, P., ORINI, M., FAES, L., OVEREEM, S., VAN GILST, M., VAN DIJK, J., VAN HUFFEL, S., BAILON, R., AND VARON, C. Model-Based Evaluation of Methods for Respiratory Sinus Arrhythmia Estimation. *IEEE Transactions on Biomedical Engineering PP* (Oct 2020).
- [137] MOUNCE, S. R., ELLIS, K., EDWARDS, J. M., SPEIGHT, V. L., JAKOMIS, N., AND BOXALL, J. B. Ensemble decision tree models using rusboost for estimating risk of iron failure in drinking water distribution systems. *Water Resources Management 31*, 5 (2017), 1575–1589.
- [138] NEJEDLY, P., CIMBALNIK, J., KLIMES, P., PLESINGER, F., HALAMEK, J., KREMEN, V., VISCOR, I., BRINKMANN, B. H., PAIL, M., BRAZDIL, M., WORRELL, G., AND JURAK, P. Intracerebral EEG Artifact Identification Using Convolutional Neural Networks. *Neuroinformatics 17*, 2 (04 2019), 225–234.
- [139] NISKANEN, J.-P., TARVAINEN, M. P., RANTA-AHO, P. O., AND KARJALAINEN, P. A. Software for advanced hrv analysis. *Computer Methods and Programs in Biomedicine 76*, 1 (2004), 73–81.
- [140] NOVILLO, F., VAN EYNDHOVEN, S., MOEYERSONS, J., BOGAERT, J., CLAESSEN, G., LA GERCHE, A., VAN HUFFEL, S., AND CLAUS, P. Unsupervised respiratory signal extraction from ungated cardiac magnetic resonance imaging at rest and during exercise. *Physics in Medicine and Biology 64*, 6 (03 2019), 065001.
- [141] NOWBAR, A. N., HOWARD, J. P., FINEGOLD, J. A., ASARIA, P., AND FRANCIS, D. P. 2014 global geographic analysis of mortality from ischaemic heart disease by country, age and income: Statistics from world health organisation and united nations. *International Journal of Cardiology 174*, 2 (2014), 293 – 298.
- [142] OKIN, P. M., HILLE, D. A., KJELDSSEN, S. E., AND DEVEREUX, R. B. Combining ECG Criteria for Left Ventricular Hypertrophy Improves Risk Prediction in Patients With Hypertension. *Journal of the American Heart Association 6*, 11 (Nov 2017).

- [143] ORPHANIDOU, C., BONNICI, T., CHARLTON, P., CLIFTON, D., VALLANCE, D., AND TARASSENKO, L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. *IEEE Journal of Biomedical and Health Informatics* 19, 3 (2015), 832–838.
- [144] OTTENBACHER, J., KIRST, M., JATOB, L., HUFLEJT, M., GROSSMANN, U., AND STORK, W. Reliable motion artifact detection for ECG monitoring systems with dry electrodes. *IEEE Engineering in Medicine and Biology Society* 30 (2008), 1695–1698.
- [145] PALLAS-ARENY, R., AND CANALS RIERA, F. Recovering the respiratory rhythm out of the ecg. *Medical & Biological Engineering & Computing* 23, Supplement, Part I (1985), 338–339.
- [146] PAN, J., AND TOMPKINS, W. J. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering* 32, 3 (1985), 230–236.
- [147] PAN, S. J., AND YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.
- [148] PANICKER, G. K., KARNAD, D. R., NATEKAR, M., KOTHARI, S., NARULA, D., AND LOKHANDWALA, Y. Intra- and interreader variability in QT interval measurement by tangent and threshold methods in a central electrocardiogram laboratory. *Journal of Electrocardiology* 42, 4 (2009), 348–352.
- [149] PICHOT, V., ROCHE, F., CELLE, S., BARTHÉLÉMY, J.-C., AND CHOUCOU, F. Hrvanalysis: A free software for analyzing cardiac autonomic activity. *Frontiers in Physiology* 7 (2016), 557.
- [150] RADOVIC, M., GHALWASH, M., FILIPOVIC, N., AND OBRADOVIC, Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics* 18, 9 (2017).
- [151] RANKINEN, T., SUNG, Y. J., SARZYNSKI, M. A., RICE, T. K., RAO, D. C., AND BOUCHARD, C. Heritability of submaximal exercise heart rate response to exercise training is accounted for by nine SNPs. *Journal of Applied Physiology* 112, 5 (Mar 2012), 892–897.
- [152] RAUTAHARJU, P. M., MANOLIO, T. A., SISCOVICK, D., ZHOU, S. H., GARDIN, J. M., KRONMAL, R., FURBERG, C. D., BORHANI, N. O., AND NEWMAN, A. Utility of new electrocardiographic models for left ventricular mass in older adults. *Hypertension* 28, 1 (1996), 8–15.

- [153] RAYA, M. A. D., AND SISON, L. G. Adaptive noise cancelling of motion artifact in stress ecg signals using accelerometer. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society* [Engineering in Medicine and Biology (2002), vol. 2, pp. 1756–1757 vol.2.
- [154] REDMOND, S. J., XIE, Y., CHANG, D., BASILAKIS, J., AND LOVELL, N. H. Electrocardiogram signal quality measures for unsupervised telehealth environments. *Physiological Measurement* 33, 9 (2012), 1517–1533.
- [155] REIMERS, C. D., KNAPP, G., AND REIMERS, A. K. Does physical activity increase life expectancy? A review of the literature. *Journal of Aging Research* 2012 (2012), 243958.
- [156] RODENHAUSER, A., GOOD, W. W., ZENGER, B., TATE, J., ARAS, K., BURTON, B., AND MACLEOD, R. S. Pfeifer: Preprocessing framework for electrograms intermittently fiducialized from experimental recordings. *Journal of Open Source Software* 3, 21 (2018), 472.
- [157] ROSELL, J., COHEN, K., AND WEBSTER, J. Reduction of motion artifacts using a two-frequency impedance plethysmograph and adaptive filtering. *IEEE Transactions on Biomedical Engineering* 42, 10 (October 1995), 1044–1048.
- [158] ROSENWINKEL, E. T., BLOOMFIELD, D. M., ALLISON ARWADY, M., AND GOLDSMITH, R. L. Exercise and autonomic function in health and cardiovascular disease. *Cardiology Clinics* 19, 3 (2001), 369–387.
- [159] ROTH, G. A., FOROUZANFAR, M. H., MORAN, A. E., BARBER, R., NGUYEN, G., FEIGIN, V. L., NAGHAVI, M., MENSAH, G. A., AND MURRAY, C. J. Demographic and epidemiologic drivers of global cardiovascular mortality. *New England Journal of Medicine* 372, 14 (2015), 1333–1341. PMID: 25830423.
- [160] SAMENI, R., SHAMSOLLAHI, M. B., JUTTEN, C., AND CLIFFORD, G. D. A nonlinear bayesian filtering framework for ecg denoising. *IEEE Transactions on Biomedical Engineering* 54, 12 (2007), 2172–2185.
- [161] SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3, 3 (1959), 210–229.
- [162] SARUSI, A., RÁROSI, F., SZÜCS, M., CSÍK, N., FARKAS, A. S., PAPP, J. G., VARRÓ, A., FORSTER, T., CURTIS, M. J., AND FARKAS, A. Absolute beat-to-beat variability and instability parameters of ECG intervals: biomarkers for predicting ischaemia-induced ventricular fibrillation. *British Journal of Pharmacology* 171, 7 (Apr 2014), 1772–1782.

- [163] SCHUMM, J., AXMANN, S., ARNRICH, B., AND TRÖSTER, G. Automatic signal appraisal for unobtrusive ECG measurements. *International Journal of Bioelectromagnetism* 12, 4 (2010), 158–164.
- [164] SEEUWS, N., DE VOS, M., AND BERTRAND, A. Electrocardiogram quality assessment using unsupervised deep learning. *IEEE Transactions on Biomedical Engineering* (Under review).
- [165] SEIFFERT, C., KHOSHGOFTAAR, T., VAN HULSE, J., AND NAPOLITANO, A. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 40, 1 (2010), 185 – 197.
- [166] SEPPÄ, V. P., HYTTINEN, J., AND VIIK, J. A method for suppressing cardiogenic oscillations in impedance pneumography. *Physiological Measurement* 32, 3 (Mar 2011), 337–345.
- [167] SEPPÄ, V. P., VIIK, J., AND HYTTINEN, J. Assessment of pulmonary flow using impedance pneumography. *IEEE Transactions on Biomedical Engineering* 57, 9 (Sep 2010), 2277–2285.
- [168] SERRE, T., KOUH, M., CADIEU, C., KNOBLICH, U., KREIMAN, G., AND POGGIO, T. A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *AI Memo 2005-036, CBCL Memo* (03 2007).
- [169] SEUNG MIN LEE, KO KEUN KIM, AND PARK, K. S. Wavelet approach to artifact noise removal from capacitive coupled electrocardiograph. In *IEEE Engineering in Medicine and Biology Society* (2008), vol. 30, pp. 2944–2947.
- [170] SHARMA, L. D., AND SUNKARIA, R. K. A robust qrs detection using novel pre-processing techniques and kurtosis based enhanced efficiency. *Measurement* 87 (2016), 194–204.
- [171] SILVA, I., MOODY, G. B., AND CELI, L. Improving the quality of ecgs collected using mobile phones: The physionet/computing in cardiology challenge 2011. *Computing in Cardiology* (2011), 273–276.
- [172] SOLLICH, P. Bayesian methods for support vector machines: Evidence and predictive class probabilities. *Machine Learning* 46 (12 2000).
- [173] SPRENKELER, D. J., TUINENBURG, A. E., RITSEMA VAN ECK, H. J., MALIK, M., ZABEL, M., AND VOS, M. A. Circadian pattern of short-term variability of the qt-interval in primary prevention icd patients - eu-cert-icd methodological pilot study. *PLOS ONE* 12, 8 (08 2017), 1–11.

- [174] SPRINGENBERG, J. T., DOSOVITSKIY, A., BROX, T., AND RIEDMILLER, M. A. Striving for simplicity: The all convolutional net. *CoRR abs/1412.6806* (2015).
- [175] SUI, X., LAMONTE, M. J., AND BLAIR, S. N. Cardiorespiratory fitness and risk of nonfatal cardiovascular disease in women and men with hypertension. *American Journal of Hypertension* 20, 6 (Jun 2007), 608–615.
- [176] TAHA, B. H., SIMON, P. M., DEMPSEY, J. A., SKATRUD, J. B., AND IBER, C. Respiratory sinus arrhythmia in humans: an obligatory role for vagal feedback from the lungs. *Journal of Applied Physiology* 78, 2 (Feb 1995), 638–645.
- [177] TEIJEIRO, T., GARCÍA, C., CASTRO, D., AND FÉLIX, P. Arrhythmia classification from the abductive interpretation of short single-lead ecg records. *Computing in Cardiology* 44 (2017).
- [178] THILL, M., DÄUBENER, S., KONEN, W., AND BÄCK, T. H. W. Anomaly detection in electrocardiogram readings with stacked LSTM networks. vol. 19, pp. 17–25.
- [179] THOMSEN, M. B., VERDUYN, S. C., STENGL, M., BEEKMAN, J. D., DE PATER, G., VAN OPSTAL, J., VOLDERS, P. G., AND VOS, M. A. Increased short-term variability of repolarization predicts d-sotalol-induced torsades de pointes in dogs. *Circulation* 110, 16 (2004), 2453–2459.
- [180] TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 1 (1996), 267–288.
- [181] TONG, D. A., BARTELS, K. A., AND HONEYAGER, K. S. Adaptive reduction of motion artifact in the electrocardiogram. In *Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society* [Engineering in Medicine and Biology (2002), vol. 24, pp. 1403–1404.
- [182] TONG, S., AND KOLLER, D. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research* (2001), 45–66.
- [183] VAGLIO, M., ISOLA, L., GATES, G., AND BADILINI, F. Use of ecg quality metrics in clinical trials. *Computing in Cardiology* (2010), 505–508.
- [184] VAN GESTEL, T., SUYKENS, J., BAESENS, B., VIAENE, S., VAN THIENEN, J., DEDENE, G., DE MOOR, B., AND, M., AND VANDEWALLE, J.

- Benchmarking least squares support vector machine classifiers. *Machine Learning* 54 (06 2002).
- [185] VAN STEENKISTE, T., GROENENDAAL, W., DREESEN, P., LEE, S., KLERKX, S., DE FRANCISCO, R., DESCHRIJVER, D., AND DHAENE, T. Portable Detection of Apnea and Hypopnea Events Using Bio-Impedance of the Chest and Deep Learning. *IEEE Journal of Biomedical and Health Informatics* 24, 9 (09 2020), 2589–2598.
- [186] VAPNIK, V. N. Pattern recognition using generalized portrait method. *Automation and Remote Control* 24 (1963), 774–780.
- [187] VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin, Heidelberg, 1995.
- [188] VAPNIK, V. N. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [189] VARKEVISSER, R., WIJERS, S. C., VAN DER HEYDEN, M. A., BEEKMAN, J. D., MEINE, M., AND VOS, M. A. Beat-to-beat variability of repolarization as a new biomarker for proarrhythmia in vivo. *Heart Rhythm* 9, 10 (Oct 2012), 1718–1726.
- [190] VARON, C. *Mining the ECG: Algorithms and Applications*. PhD thesis, KU Leuven, 2015.
- [191] VARON, C., CAICEDO, A., TESTELMANS, D., BUYSE, B., AND HUFFEL, S. V. A novel algorithm for the automatic detection of sleep apnea from single-lead ecg. *IEEE Transactions on Biomedical Engineering* 62, 9 (Sep. 2015), 2269–2278.
- [192] VARON, C., LÁZARO, J., BOLEA, J., HERNANDO, A., AGUILÓ, J., GIL, E., VAN HUFFEL, S., AND BAILÓN, R. Unconstrained estimation of hrv indices after removing respiratory influences from heart rate. *IEEE Journal of Biomedical and Health Informatics* 23, 6 (2019), 2386–2397.
- [193] VARON, C., MORALES, J., LAZARO, J., ORINI, M., DEVIAENE, M., KONTAXIS, S., TESTELMANS, D., BUYSE, B., BORZÉE, P., SÖRNMO, L., LAGUNA, P., GIL, E., AND BAILON, R. A Comparative Study of ECG-derived Respiration in Ambulatory Monitoring using the Single-lead ECG. *Scientific Reports* 10, 1 (03 2020), 5704.
- [194] VARON, C., TESTELMANS, D., BUYSE, B., SUYKENS, J., AND VAN HUFFEL, S. Robust artefact detection in long-term ecg recordings based on autocorrelation function similarity and percentile analysis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2012), 3151–3154.

- [195] VICENTE, J., JOHANNESSEN, L., GALEOTTI, L., AND STRAUSS, D. G. Ecglab: User friendly ecg/vcg analysis tool for research environments. *Computing in Cardiology* (Sep. 2013), 775–778.
- [196] VÁZQUEZ-SEISDEDOS, C. R., NETO, J. E., MARAÑÓN REYES, E. J., KLAUTAU, A., AND LIMÃO DE OLIVEIRA, R. C. New approach for t-wave end detection on electrocardiogram: performance in noisy conditions. *Biomedical Engineering Online* 10 (2011), 77.
- [197] WARTZEK, T., CZAPLIK, M., AND ANTINK, C. H. Unovis: The mediterranean public unobtrusive vital signs database. *Health Information Science and Systems* 3 (2015).
- [198] WARTZEK, T., EILEBRECHT, B., LEM, J., LINDNER, H., LEONHARDT, S., AND WALTER, M. Ecg on the road: Robust and unobtrusive estimation of heart rate. *IEEE Transactions on Biomedical Engineering* 58 (2011), 3112–3120.
- [199] WEDEKIND, D., MALBERG, H., AND ZAUNSEDER, S. Cascaded output selection for processing of capacitive electrocardiograms by means of independent component analysis. In *Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)* (2013), pp. 1–6.
- [200] WEISS, G. M., AND PROVOST, F. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research* 19, 1 (2003), 315–354.
- [201] WELCH, P. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 15, 2 (June 1967), 70–73.
- [202] XIA, H., GARCIA, G. A., MCBRIDE, J. C., SULLIVAN, A., BOCK, T. D., BAINS, J., WORTHAM, D. C., AND ZHAO, X. Computer algorithms for evaluating the quality of ecgs in real time. *Computing in Cardiology* (2011), 369–372.
- [203] YANG, J., YAN, R., AND HAUPTMANN, A. G. Adapting svm classifiers to data with shifted distributions. In *IEEE International Conference on Data Mining Workshops* (2007), vol. 7, pp. 69–76.
- [204] ZAUNSEDER, S., HUHLE, R., AND MALBERG, H. Cinc challenge - assessing the usability of ecg by ensemble decision trees. *Computing in Cardiology* (2011), 277–280.

- [205] ZAVORSKY, G. S. Evidence and possible mechanisms of altered maximum heart rate with endurance training and tapering. *Sports Medicine* 29, 1 (Jan 2000), 13–26.
- [206] ZHANG, Q., FU, L., AND GU, L. A cascaded convolutional neural network for assessing signal quality of dynamic ecg. *Computational and Mathematical Methods in Medicine 2019* (10 2019), 1–12.
- [207] ZHANG, Q., MANRIQUEZ, A. I., MÉDIGUE, C., PAPELIER, Y., AND SORINE, M. An algorithm for robust and efficient location of T-wave ends in electrocardiograms. *IEEE Transactions on Biomedical Engineering* 53, 12 (Dec 2006), 2544–2552.
- [208] ZOU, H., AND HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67, 2 (2005), 301–320.

Curriculum

Jonathan Moeyersons is born in Duffel on September 14, 1991. In 2013 he obtained the degree of Master of Science in Physical Education and Kinesiology, cum laude, from the KU Leuven. Hereafter, in 2016, he obtained the degree of Master of Bioscience Engineering: Human Health Engineering, cum laude, also from the KU Leuven.

In September 2016 he started a PhD at the STADIUS Center of Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven under the supervision of prof. dr. ir. Sabine Van Huffel, prof. dr. Rik Willems and prof. dr. ir. Carolina Varon. In 2018, he went on a brief research stay to the Biomedical Signal Interpretation & Computational Simulation research group at the University of Zaragoza.

List of publications

Articles in international journals

1. **MOEYERSONS, J.**, MORALES, J., VILLA, A., CASTRO, I., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., VAN HUFFEL, S. AND VARON, C. Supervised SVM transfer learning for modality-specific artefact detection in ECG. *Sensors* 21, 2, (2021), 662. <https://doi.org/10.3390/s21020662>.
2. LAVANGA, M., HEREMANS, E., DEVIAENE, M., **MOEYERSONS, J.**, BOLLEN, B., JANSEN, K., ORTIBUS, E., NAULAERS, G., VAN HUFFEL, S., AND CAICEDO, A. Maturation of the autonomic nervous system in premature infants: estimating development based on heart-rate variability analysis. *Frontiers in Physiology* 11, 581250, (2021), <https://doi.org/10.3389/fphys.2020.581250>.
3. MORALES, J., **MOEYERSONS, J.**, ARMANAC, P., ORINI, M., FAES, L., OVEREEM, S., VAN GILST, M., VAN DIJK, J., VAN HUFFEL, S., BAILON, R., AND VARON, C. Model-based evaluation of methods for respiratory sinus arrhythmia estimation. *IEEE Transactions on Biomedical Engineering PP*, (2020), <https://doi.org/10.1109/TBME.2020.3028204>.
4. **MOEYERSONS, J.**, AMONI, M., VAN HUFFEL, S., WILLEMS, R., AND VARON, C. R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks (version 1.0.0). *PhysioNet*, (2020), <https://doi.org/10.13026/x6j7-sp58>.
5. **MOEYERSONS, J.**, SMETS, E., MORALES, J., VILLA, A., DE RAEDT, W., TESTELMANS, D., BUYSE, B., VAN HOOFF, C., WILLEMS, R., VAN HUFFEL, S. AND VARON, C. Artefact detection and quality assessment of ambulatory ECG signals. *Computer Methods and Programs in Biomedicine* 182, 105050, (2019), <https://doi.org/10.1016/j.cmpb.2019.105050>.

6. **MOEYERSONS, J.**, AMONI, M., VAN HUFFEL, S., WILLEMS, R., AND VARON, C. R-DECO: an open-source Matlab based graphical user interface for the detection and correction of R-peaks. *PeerJ Computer Science* 5, e226, (2019), <https://doi.org/10.7717/peerj-cs.226>.
7. NOVILLO, F., VAN EYNDHOVEN, S., **MOEYERSONS, J.**, BOGAERT, J., CLAESSEN, G., LA GERCHE, A., VAN HUFFEL, S., AND CLAUS, P. Unsupervised respiratory signal extraction from ungated cardiac magnetic resonance imaging at rest and during exercise. *Physics in Medicine and Biology* 64, 6, (2019), <https://doi.org/10.7717/10.1088/1361-6560/ab02cd>.

Papers in proceedings of international conferences

1. JACOBS, B., **MOEYERSONS, J.**, VILLA, A., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Can Laplacian Eigenmaps be used for differentiation between healthy subjects and patients with corrected Tetralogy of Fallot? *Computing in Cardiology 47*, Rimini, Italy, (Sep 13-16, 2020).
2. **MOEYERSONS, J.**, DE BOSSCHER, R., DAUSIN, C., CLAESSEN, G., LA GERCHE, A., BOGAERT, J., WILLEMS, R., VAN HUFFEL, S., VARON, C. Predicting left ventricular mass using ECG, demographic and DXA features. *Computing in Cardiology 47*, Rimini, Italy, (Sep 13-16, 2020).
3. LAVANGA, M., WEL, O. D., CAICEDO, A., DEVIAENE, M., **MOEYERSONS, J.**, VARON, C., BOLLEN, B., JANSEN, K., ORTIBUS, E., NAULAERS, G., AND VAN HUFFEL, S. The implementation of an apnea based perinatal stress calculator. *IEEE Engineering in Medicine and Biology Society 41*, Berlin, Germany, (Jul 23-27, 2019).
4. CASTRO, I., VARON, C., **MOEYERSONS, J.**, GOMEZ, A. V., MORALES, J., DEVIAENE, M., TORFS, T., VAN HUFFEL, S., PUERS, R., AND VAN HOOFF, C. Data quality assessment of capacitively-coupled ECG signals. *Computing in Cardiology 46*, Singapore, Singapore, (Sep 8-11, 2019).
5. **MOEYERSONS, J.**, AMONI, M., VANDENBERK, B., VARON, C., SIPIDO, K., HUFFEL, S., AND WILLEMS, R. Temporal beat-to-beat variability of repolarization changes predict non-sustained ventricular tachycardia in ischemic heart disease patients. *Computing in Cardiology 45*, Maastricht, The Netherlands, (Sep 23-26, 2018).

6. **MOEYERSONS, J.**, TESTELMANS, D., BUYSE, B., WILLEMS, R., VAN HUFFEL, S., AND VARON, C. Evaluation of a continuous ECG quality indicator based on the autocorrelation function. *Computing in Cardiology 45*, Maastricht, The Netherlands, (Sep 23-26, 2018).
7. **MOEYERSONS, J.**, VARON, C., TESTELMANS, D., BUYSE, B., AND VAN HUFFEL, S. ECG artefact detection using ensemble decision trees. *Computing in Cardiology 44*, Rennes, France, (Sep 24-27 2017).
8. **MOEYERSONS J.**, GOOVAERTS G., HUIJGHEBAERT S., VANDENBERK B., WILLEMS R., VAN HUFFEL S. Automated T wave end detection methods: comparison of four different methods for T wave end detection, *International Joint Conference on Biomedical Engineering Systems and Technologies 10*, Porto, Portugal, (Feb 21-23, 2017).

Abstracts in proceedings of (inter)national conferences

1. COENEN, M., JACOBS, B., SWINNEN, T., VANAUTGAERDEN, E., VAN GILS, N., DE BOODT, P., JACOBS, S., **MOEYERSONS, J.**, VAN HUFFEL, S., STAES, F., SERMON, A. HIPS DON'T LIE: sensor-based hip fracture data to promote functioning and survival. *Presented at the Symposium "Technologie in onderzoek en ontwikkeling voor Veerkrachtig Ouder worden"*, Leuven, Belgium, (Feb 1, 2020).
2. **MOEYERSONS, J.**, MARTINEZ, J.P., YEBRA, A.M., LAGUNA, P., VAN HUFFEL, S., WILLEMS, R., BAILON, R., VARON, C. Cardiorespiratory coupling and gas exchange during exercise, bridging the gap. *Belgian Day on Biomedical Engineering 18*, Brussels, Belgium, (Nov 29, 2019).
3. AMONI, M., VANDENBERK, B., **MOEYERSONS, J.**, HUFFEL, S., SIPIDO, K., AND WILLEMS, R. Temporal beat-to-beat variability of repolarization (bvr) changes predict imminent non-sustained ventricular tachycardia in ischaemic heart disease patients. *European Heart Rhythm Association 20*, Barcelona, Spain, (Mar 18-20, 2018).
4. **MOEYERSONS, J.**, GOOVAERTS, G., HUIJGHEBAERT, S., VANDENBERK, B., WILLEMS, R., VAN HUFFEL, S., AERTS, J. Effect of food and meals on ECG dynamics. *Belgian Day on Biomedical Engineering 15*, Brussels, Belgium, (Mar 4, 2016).

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
BIOMED

Kasteelpark Arenberg 10 - box 2446, B-3001 Leuven (Belgium)
B-3001 Leuven

jonathan.moeyersons@esat.kuleuven.be

<http://www.stadius.esat.kuleuven.be>

