

POLITECNICO DI MILANO

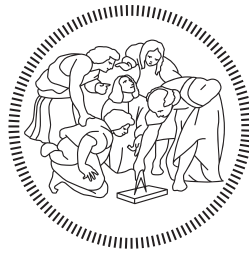
Facoltà di Ingegneria

Scuola di Ingegneria Industriale e dell'Informazione

Dipartimento di Elettronica, Informazione e Bioingegneria

Corso di Laurea Magistrale in

Ingegneria Biomedica



Prediction and Characterization of Acute Hypotensive Episodes in Intensive Care Unit

Relatore:

PROF. RICCARDO BARBIERI

Correlatore:

DR. LI-WEI H LEHMAN

Tesi di Laurea Magistrale di:

GIOVANNI ANGELOTTI

Matricola n. 841324

PIERANDREA MORANDINI

Matricola n. 841747

Anno Accademico 2016-2017

ACKNOWLEDGMENTS

The development of this thesis has started more than a year ago, it has been a long journey where we faced several obstacles, obstacles that we could never have overcome without the aid of the persons we had the great fortune to meet.

A Special Thanks goes to Professor Riccardo Barbieri for believing in us from the beginning and always since then, for giving us the chance to express our potential and introducing us to a world of challenges and opportunities.

Also a Special Thanks to Doctor Li-wei Lehman for guiding us throughout the process, for all the infinite support, patience and kindness who has always shown us.

We are grateful to Distinguished Professor Roger Mark for hosting us and letting us experience the beautiful environment that the LCP is: Thanks Chen, Mohammad, Leo, Benjamin, Alistair, Tom, Felipe, Rodrigo, Jerome, Miguel, Ken, Jesse, Kenneth and Cristina.

Thanks to all of You for welcoming us and always helping us.

We are grateful to the SPINlabS and Politecnico di Milano for giving us the infrastructures to pursue this work.

Mirco, Riccardo, Ester, Matteo, Carlo, Francesca, Luca, Donatella, Andrea, Matteo it has been a pleasure to share this experience with you.

CONTENTS

Abstract	xi
Estratto	xiii
1 INTRODUCTION	1
2 BACKGROUND	3
2.1 Acute Hypotensive Episodes	3
2.1.1 ANS	5
2.1.2 Medications	7
2.1.3 Septic Shock	7
2.2 State of The Art	8
2.3 MIMIC	17
2.3.1 MIMIC Clinical Database	18
2.3.2 MIMIC Waveform Database	19
2.3.3 MIMIC Matched Subset	20
3 METHODS	21
3.1 Experimental Settings	21
3.1.1 Definition of Hypotension	21
3.1.2 Data Window and Lead Time	22
3.2 Preprocessing	23
3.2.1 ECG	23
3.2.2 ABP	24
3.3 Signal Quality Index	25
3.3.1 Beat to beat distance analysis	26
3.3.2 Analysis of characteristics of signal	26
3.4 Exclusion Criteria	27
3.5 Annotations	28
3.5.1 RRI	28
3.5.2 Systole, Diastole and Percussion Wave Onset	29
3.5.3 Pulse Transit Times	30
3.5.4 Pulse Pressure	31
3.5.5 Relative Cardiac Output	32
3.5.6 Mean Arterial Pressure	32
3.6 Modeling	34
3.6.1 Point Process	34
3.6.2 Characterizing baroreflex sensitivity	36
3.7 Feature Engineering	41
3.7.1 Time Features	41
3.7.2 Spectral Features	43
3.7.3 Non Linear Features	45
3.7.4 Score Features	49
3.8 Feature Selection	51
3.8.1 PCA	52
3.8.2 Forward Selection	52

3.8.3	LASSO	53
3.9	Classification	54
3.9.1	Logistic Regression	55
3.9.2	Discriminant Analysis	55
3.9.3	K-nearest neighbor	57
3.9.4	Support Vector Machines	58
3.9.5	Classification Trees	59
3.9.6	Applied Models	62
3.10	Medications and Interventions	62
4	RESULTS	65
4.1	AHE Prediction	65
4.2	Autonomic Assessment	70
4.2.1	Limitations and Problematics in ANS assessment in ICU	70
4.2.2	Classification	70
4.3	Conclusions	75
4.3.1	Physiology of hypotension	75
4.3.2	Influence of interventions.	78
4.3.3	Machine Learning as a Decision Making tool in ICU	81
	Appendices	83
A	RELEVANT FEATURES DESCRIPTION	85
B	ACRONYMS	89
C	MEASURES OF PERFORMANCE	91
	BIBLIOGRAPHY	93

LIST OF FIGURES

Figure 0.1	Physionet & LCP. xiii	
Figure 0.2	Esempio di ipotensione da un record MIMIC. La linea verde rappresenta la pressione arteriosa media che diventa rossa durante l'attacco ipotensivo. L'intervallo di tempo tra le linee della griglia è di 30 minuti. La linea rosa intermittente contrassegna la soglia di 60 mmHg. xiv	
Figure 0.3	Performance table xvii	
Figure 0.4	Baroreflex processing steps. xviii	
Figure 2.1	Three elements windkessel model. The capacitive element C represents the arteries compliance, the parallel resistive element is the hydraulic resistance to blood flow of the arteries while the the resistance in series is the aortic characteristic impedance. 3	
Figure 2.2	RR-SBP-RESP interaction model. The model explains the different interactions between respiration, systolic pressure and heart rate. One can see that heart rate is influenced by respiration (RSA) through Htr and by systolic pressure (Baroreflex) through Hts plus a colored residual. On the other hand, systolic pressure is dependent form heart rate, the so called mechanical effect, respiration and vessels configuration (Baroreflex on vessels in figure). RR intervals and systolic pressure are heavily interconnected. 4	
Figure 2.3	Aortic arch on the left and carotid sinus to the right are the sites of baroreceptors. 6	
Figure 2.4	RSA model. 6	
Figure 2.5	Waveform variation prior to hypotension. 8	
Figure 2.6	Autocorrelation coefficients difference due to impending hypotension. 9	
Figure 2.7	GRNN multi-model workflow proposed by Henriques and Rocha, 2009. 10	
Figure 2.8	Workflow emphasizing the role of the HHT role. 13	
Figure 2.9	Simple example of a tree structured program for genetic programming. 14	
Figure 2.10	Classification logic. 15	
Figure 2.11	Predicted ABP versus actual ABP. 16	
Figure 2.12	Physionet & LCP. 17	
Figure 2.13	High-Level Structure of the MIMIC Database. 18	
Figure 2.14	Data Skeleton for the clinical Database 19	

Figure 2.15	Waveforms samples as seen from Lightwave.	20
Figure 3.1	Example of Hypoyension from a MIMIC record. The green line represent the MAP of the waveform, it turns red during the AHE. The time interval between grid lines is 30 minutes. The intermittent pink line labels the 60 mmHg threshold.	21
Figure 3.2	Data Window and Lead time.	22
Figure 3.3	Magnitude and phase response for the ECG filter and filtering effect on the QRS.	24
Figure 3.4	Exclusion criteria.	27
Figure 3.5	The waveforms studied by pan-tompkins.	29
Figure 3.6	PUD workflow.	30
Figure 3.7	Bisferiens pulse, a challenging waveform to automatically annotate.	30
Figure 3.8	Visual description of Systole, Diastole, Percussion Wave Onset (fiducial point in figure) and Pulse Transit Times.	31
Figure 3.9	Sample time series form Record s01418-2947-11-14-12-10.	33
Figure 3.10	Description of a point process in time domain: the process is a succession of events in continuous time.	35
Figure 3.11	Example of baroreflex from control subject s03279-2984-08-26-14-42.	37
Figure 3.12	Baroreflex processing steps.	39
Figure 3.13	Feature space. For TBP features see Section 3.7.4.	41
Figure 3.14	Line fitted using information from the data window to estimate pressure 30 minutes into the future. In this example it was used a data window of 80 minutes. The polynomial MAP features are highlighted in red.	42
Figure 3.15	Spectral decomposition of the three components in two scenarios: Rest and Tilt.	45
Figure 3.16	Pointcare plot.	46
Figure 3.17	Pointcare plot.	47
Figure 3.18	Correlation Dimension	48
Figure 3.19	Visualization of α_1 and α_2 give by the detrended fluctuation analysis.	49
Figure 3.20	TBP	50
Figure 3.21	Constrain region and square of error term for l_1 norm	54
Figure 3.22	Constrain region and square of error term for l_2 norm	54
Figure 3.23	Example of logistic curve	55

Figure 3.24	The result of classification is depending the number neighbors took in consideration. 57
Figure 3.25	Soft margin example 59
Figure 3.26	Example of a distribution of two classes that can be separate only using a non linear kernel. 59
Figure 4.1	ROC curve of classifiers took as examples for all types of models. 68
Figure 4.2	Median behaviour of spectral components computed using a Point Process framework. Figure 4.2d was calculated using the Welch's Periodogram. 71
Figure 4.3	Graphs showing the values of AUC in the optimal point, Sensitivity, Specificity and Accuracy for both classification with and without baroreflex features. 73
Figure 4.4	Skewness of the baroreflex peaks. Controls are orange, Hypotensives are blu. 74
Figure 4.5	Mean and standard Error for selected physiological variables. 75
Figure 4.6	Mean and standard Error for selected physiological variables. 76
Figure 4.7	Hierarchical clustering of systolic pressure and pulse pressure of the AHE-patients. In Figures 4.7a and 4.7c brighter colors mean that the value is above the population average in that instant, darker colors below average. Figures 4.7b and 4.7d are instead in absolute values. 77
Figure 4.8	Effect of different medications/interventions on mean arterial pressure. C are control, H are hypotensive patients. the number 1 indicates the group has received the medication, number 2 they did not. The strongest effect on MAP is given by pacemakers. 79

LIST OF TABLES

Table 0.1	Performance table xviii
-----------	---

Table 0.2	Tabella con indici di prestazione analizzate a seconda delle medicazioni. Ogni intervento ha due righe: la prima mostra gli indici di prestazione sui pazienti senza medicazione specifica (riga 'No'), e la seconda gli stessi indice dei pazienti che l'hanno ricevuta (riga 'Yes'). Questi valori sono il risultato di una cross-validazione 5-fold di un analisi con discriminante lineare utilizzando tutti i 442 pazienti del cohort. xix	
Table 2.1	Performance of Chen's Indices. 12	
Table 2.2	Comparison between single-compilation and multi-compilation performance one hour before hypotension onset. 16	
Table 3.1	Butterworth Bandpass cutoff frequencies. 23	
Table 3.2	Cohort Description, C stands for Control population while H for Hypotensive population. 28	
Table 3.3	Scoring of different algorithms. 30	
Table 3.4	Cohort Description. 38	
Table 3.5	Statistical description of different signals. DW means the interval for calculation was equal to that of the data window while lstm to the last available minutes before the end of the data window. V means that specific feature was computed X otherwise. 42	
Table 3.6	List of KNN variants. 57	
Table 4.1	Datasets Composition 65	
Table 4.2	Features analysis with LT=10min, DW = 10min in the training set of our Main Dataset (CR). 66	
Table 4.3	Features analysis with LT=30min, DW = 10min in the training set of our Main Dataset (CR). 66	
Table 4.4	The significance threshold is p-val<.05. 72	
Table 4.5	Performance table 73	
Table 4.6	Zero-mean normality distribution test for the population of peaks skewness for control and hypotensive records. The test rejects the null hypothesis at the 5% significance level. 74	
Table 4.7	Table with performance indices analyzed by interventions. Every intervention has two rows: the first shows the indeces about patients without the specific intervention (row 'No'), and the second indeces about patients with it (row 'Yes'). 78	

ABSTRACT

Acute Hypotensive Episode in the Intensive Care Unit represents a severe condition that, if not treated promptly, leads to irreversible organ damage (Bassale, 2001) and increased mortality (Physionet, 2009). Therefore, it would be of relevant importance if mathematical algorithms were able to provide clinicians with statistical predictions on the onset of such condition.

Other works addressed this issue with different methods, sometimes proposing new definition of hypotension. Of those, relevant importance are the studies of Chen et al., 2009 and Ghassemi, 2011. In 2009 Physionet even proposed a challenge on the subject.

Our work starts from developing methods to win the challenge to then apply those to a wider cohort of patients. In both cases we used data that can be found in MIMIC: a public database offered by Physionet.org, in which are stored waveforms and clinical data of more than 4000 patients (MIMIC II) in the Intensive Care Unit.

Unlike other works, what we present are algorithms that use information at a beat to beat level. It was necessary, therefore, to develop a peak detection logic able to find the actual data to work on. After building a large feature space, a process of feature selection made possible to select those with the higher predictive power to use in the classification step.

Following training, our classifiers were able to stand the comparison with the winner of the challenge. The extension to a wider cohort of subjects also brought good results: the best performances were achieved with Discriminant Analysis and Classification Trees.

The features used in this first analysis are relative to blood pressure amplitude in the moments preceding hypotension. To actually express the power of the beat to beat level information, it was necessary to reduce the dimension of the cohort to those waveform showing perfect continuity in the heart rate variability signal. On this selected cohort was possible to extract information about the baroreflex. We demonstrate here that these new informations are able to boost the classification results incrementing specificity from 0.739 to 0.796s and sensitivity from 0.745 to 0.834.

SOMMARIO

Episodi di ipotensione acuta nell'unità di terapia intensiva rappresentano una condizione grave che, se non trattata tempestivamente, porta a danni irreversibili agli organi e maggiore probabilità di mortalità. Pertanto, è di importanza significativa avere algoritmi matematici in grado di fornire al medico previsioni sull'esordio di tale condizione. Diversi studi hanno affrontato il problema con varie metodologie, a volte proponendo nuove definizioni di ipotensione. Di rilevante importanza sono gli studi di Chen et al., 2009 e Ghassemi, 2011. Nel 2009, Physionet ha persino indetto una Challenge sul tema.

Il nostro lavoro parte dallo sviluppo di metodi per vincere la Challenge per poi applicarli ad un più ampio dataset di pazienti. In entrambi i casi abbiamo utilizzato i dati del MIMIC: un database pubblico fornito da Physionet.org, in cui sono presenti forme d'onda e dati clinici di oltre 4000 pazienti (MIMIC II) in terapia intensiva.

A differenza di altri, presentiamo algoritmi che sfruttano informazione a livello battito-battito. È stato quindi necessario sviluppare una logica di rilevamento battiti in grado di trovare i dati reali su cui lavorare. Dopo aver creato un ampio dataset, un processo di selezione delle features ha permesso di selezionare quelle con maggiore potenza predittiva da utilizzare nella fase di classificazione.

Dopo il training, i nostri classificatori hanno sostenuto il confronto con il vincitore della Challenge. Anche l'estensione a una più ampio dataset di soggetti ha portato buoni risultati: le migliori prestazioni sono state ottenute con Analisi del Discriminante e Alberi di Classificazione. Le features utilizzate in questa prima analisi sono per lo più relative all'ampiezza della pressione sanguigna nei momenti precedenti all'ipotensione. Per esprimere efficacemente la potenza delle informazioni a livello battito battito, era necessario ridurre la dimensione della coorte a quelle forme d'onda che mostravano una perfetta continuità nel segnale di variazione della frequenza cardiaca. In questa coorte selezionata, è stato possibile estrarre informazioni sul baroriflesso. Dimostriamo qui che queste nuove informazioni sono in grado di aumentare i risultati della classificazione aumentando la specificità da 0,739 a 0,796 punti e la sensibilità da 0,745 a 0,834.

ESTRATTO

Nell'unità di terapia intensiva di un ospedale, gli operatori sanitari monitorano attentamente i pazienti con gravi lesioni, traumi post-operatori o condizioni di salute instabili utilizzando sistemi di monitoraggio del letto che registrano continuamente forme d'onda come elettrocardiogramma (ECG) e pressione sanguigna arteriosa (ABP). Gli episodi ipotensivi acuti (AHE) sono tra gli eventi più critici che si verificano nelle unità di terapia intensiva e richiedono un intervento rapido e efficace per evitare danni irreversibili agli organi e/o la morte del paziente. La predizione anticipata di AHE (entro poche ore) potrebbe essere di fondamentale importanza per fornire interventi terapeutici tempestivi e appropriati.

Grave ipotensione può essere il risultato di molteplici cause, che vanno da ipovolemia, iperkaliemia, ipotiroidismo e disfunzioni del sistema nervoso autonomo (ANS), sindrome da arresto cardiaco, insufficienza cardiaca congestizia, sepsi e altri. Quindi, l'origine dell'ipotensione può essere ricondotta a un ampio spettro di cause che evidenzia le complesse dinamiche dietro tali eventi.

Numerosi studi hanno già affrontato questo tema. Molti di questi fanno parte di studi svolti per affrontare la challenge indetta da Physionet.org. PhysioNet è uno dei repository più grandi, più completi e più ampiamente usati al mondo di segnali fisiologici registrati liberamente disponibili e dati clinici ad alta risoluzione provenienti dai reparti di terapia intensiva, con relativo software open source per la ricerca.

PhysioNet, il primo repository di questo tipo, è stato fondato nel 1999 per fornire risorse di dati e software in biomedicina alla comunità di ricerca.¹

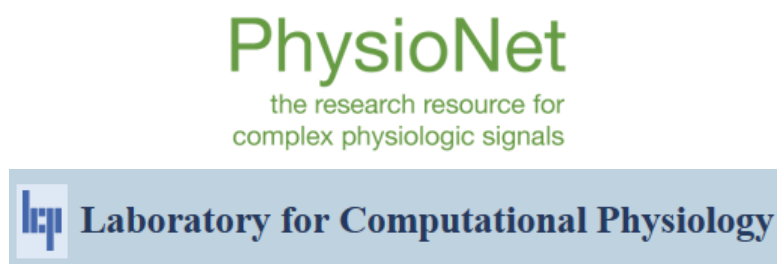


Figure 0.1: Physionet & LCP.

Annualmente Physionet propone delle challenge in cui i partecipanti devono presentare metodi per risolvere il problema oggetto dell'analisi. La challenge del 2009 aveva come focus la predizione di attacchi ipotensivi acuti che vengono definiti come una porzione di pressione arteriosa

¹ <http://lcp.mit.edu/physionet.shtml>

in cui la pressione media sia sotto i 60 mmHg per almeno 30 minuti. Per svolgere la predizione erano disponibili tracciati di ECG e ABP.

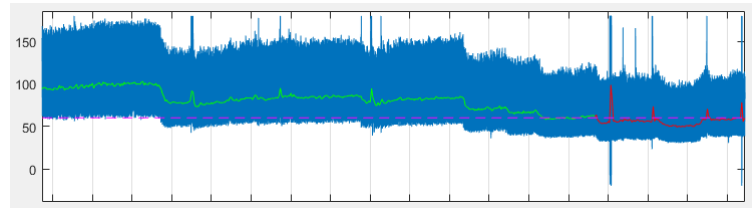


Figure 0.2: Esempio di ipotensione da un record MIMIC. La linea verde rappresenta la pressione arteriosa media che diventa rossa durante l'attacco ipotensivo. L'intervallo di tempo tra le linee della griglia è di 30 minuti. La linea rosa intermittente contrassegna la soglia di 60 mmHg.

Lo studio presentato in questa tesi parte dall'analisi della challenge per sviluppare metodi che verranno poi applicati su una coorte di pazienti piú grande in modo da verificare l'efficacia degli algoritmi sviluppati.

Un secondo studio andrà a focalizzarsi su un insieme piú ridotto di pazienti su cui verrà svolta un'analisi piú particolare: la predizione dell'ipotensione svolta con l'aiuto aggiuntivo dell'informazione portata dal baroriflesso.

Per il rilevamento dei complessi QRS (insieme di onde prodotte dalla depolarizzazione dei ventricoli), è stato utilizzato l'algoritmo di Pan-Tompkins, mentre per il rilevamento dei punti di sistole, diastole e punto fiduciario ², è stato utilizzato un algoritmo chiamato Pulse Waveform Delineator (PUD) (Li, Dong, and Vai, 2010). Analogamente a Pan-Tompkins, PUD funziona contemporaneamente su due canali, l'ABP stesso e la sua derivata. Cerca coppie di punti di flesso e zero crossing per identificare i picchi sistolici e le valli diastoliche usando soglie dinamiche, quindi cerca i nodi dicrotici.

Ottenuti i tracciati di sistogramma, diastogramma, fiduciogramma ³ e tacogramma è stato possibile costruire un ampio dataset per descrivere la fisiologia dei pazienti ipotensivi e di controllo. L'opera di predizione infatti si basa sull'identificare lo stato fisiologico di queste due classi di pazienti prima dell'onset della patologia sotto analisi, per poi effettuare una classificazione tra pazienti sani e quelli che andranno incontro all'evento ipotensivo.

Sull'ampio dataset costruito è stata fatta un'analisi per trovare le features che racchiudono il maggiore potere predittivo, in modo da ridurre la dimensionalità del problema e rendere piú interpretabili i modelli che sarebbero risultati. Per far questo è stato utilizzato l'algoritmo LASSO. Il dataset ottimale risulta infine composta da:

² Punto di massima derivata nel segnale ABP

³ Successione dei punti fiduciari

- ag70;
- m_diast;
- as10;
- minof;
- CO ⁴;

Una volta ottenuti i dataset ottimali é stato possibile cominciare con la classificazione. Sono stati addestrati 17 diversi classificatori per valutare il potere predittivo di diversi metodi e trovare quale dei diversi modelli offrissi il migliore fit ai dati in training e test set. I 17 classificatori utilizzati sono:

- Disc_Lin: Analisi del discriminante lineare;
- Discr_Quad: Analisi del discriminante quadratica;
- Discr_Sub: Analisi del discriminante migliorata con il Random Subspace method;
- KNN_Coarse: K-nearest neighbor con 100 neighbor e distanza euclidea;
- KNN_Medium: K-nearest neighbor con 10 neighbor e distanza euclidea;;
- KNN_Cosine: K-nearest neighbor con distanza cosinoidale;
- KNN_Mikowski: K-nearest neighbor con distanza di Mikowski;
- KNN_Weight: K-nearest neighbor con distanza euclidea usata come inverso del quadrato;
- KNN_Sub: K-nearest neighbor migliorata con il Random Subspace method;
- LogRegr: Regressione Logistica;
- SVM_Lin: Support Vector Machine con kernel lineare
- SVM_Quad: Support Vector Machine con kernelquadratico;
- SVM_Cub: Support Vector Machine con kernel cubico;
- Tree_AdaBoost: alberi di classificazione con Adaptative Boosting;
- Tree_RF: alberi di classificazione con Random Forest;

⁴ Si rimanda all'appendice A per i dettagli

- Tree_RUSBoost: alberi di classificazione con Random Under Sampling;
- Tree_TotalBoost: alberi di classificazione con Totally Corrective Boosting.

I risultati ottenuti sul test set della challenge Physionet sono molto buoni: con Support Vector Machine quadratiche e cubiche ed i Random Forrest Trees é stato possibile ottenere una classificazione con il 90% di accuracy (36 corrette classificazioni su 40). Anche con gli altri metodi é stato possibile ottenere buoni risultati.

L'applicazione di questi metodi ad un cohort di pazienti piú grande ha portato anch'essa a buoni risultati. Una coorte di 442 pazienti é stata divisa in train test e test set con una suddivisione rispettivamente di 75% e 25%. Con le performance sul training e test set, oltre agli alberi di classificazione, spiccano i classificatori che sfruttano l'analisi del discriminante. Questi ultimi, mostrano un bilancio molto buono tra sensibilità e specificità ed ottimi performance di classificazione. Come ultimo, é stato provato un utilizzo dei classificatori addestrati sul training set della coorte di 442 pazienti sul test set della Challenge Physionet. In questo caso le prestazioni sono state piú alte del caso precedente: si passa dal 90% al 92.5% di accuracy (37 corrette classificazioni). Questo risultato viene raggiunto con i Tree_RF. Le performance di tutti i classificatori sono riassunte nelle tabelle [0.3a](#) [0.3b](#) [0.3c](#) [0.3d](#).

Per il secondo studio sul baroriflesso é stato necessario scegliere con cura i pazienti con cui lavorare: nella coorte di 442 pazienti, numerosi artefatti sporcavano i segnali e non era possibile ottenere tracciati puliti di tacogramma e sistogramma. Dopo ispezione visiva, sono stati scelti 83 pazienti su cui effettuare questa seconda analisi.

Per estrarre il baroriflesso é stato necessario costruire un modello bivariato che fosse alimentato da tacogramma e sistogramma. Il modello creato é un modello stocastico che utilizza la metodologia del Point Process. Una volta costruito il modello bivariato é stato possibile calcolare il baroriflesso come modulo della funzione di trasferimento tra i due sistemi. Il baroriflesso trovato in questo modo é una funzione del tempo ma anche delle frequenze. Per ottenere un segnale é stato fatto un integrale nel dominio delle Low Frequency (0.04 – 0.15 Hz). Dal segnale cosí ottenuto sono state estratte feature atte a descrivere il diverso comportamento del segnale tra le due classi di pazienti. Sono stati utilizzati classici indici statistici ma anche parametri piú complessi come misure di entropia e caoticità del segnale. Altri indici particolari estratti sono stati gli indici statistici di linea di base e attività dei picchi in relazione alla linea di base.

L'ipotesi che sta alla base di questa scelta é che il baroriflesso puó essere considerato come composto da due meccanismi separati:

$$\text{baroreflex} = \text{baseline} + \text{bursts} \quad (0.1)$$

Classifier	AUC	AUC_OP	SENS	SPEC	ACC	Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.865	0.775	0.788	0.762	0.77	Discr_Lin	0.886	0.788	0.735	0.84	0.807
Discr_Quad	0.855	0.775	0.779	0.771	0.773	Discr_Quad	0.884	0.773	0.706	0.84	0.798
Discr_Sub	0.868	0.769	0.788	0.749	0.761	Discr_Sub	0.884	0.788	0.735	0.84	0.807
KNN_Coarse	0.853	0.773	0.779	0.767	0.77	KNN_Coarse	0.88	0.773	0.706	0.84	0.798
KNN_Cosine	0.84	0.757	0.721	0.793	0.77	KNN_Cosine	0.86	0.793	0.706	0.88	0.826
KNN_Medium	0.844	0.756	0.692	0.819	0.779	KNN_Medium	0.862	0.772	0.676	0.867	0.807
KNN_Mikowski	0.85	0.766	0.779	0.753	0.761	KNN_Mikowski	0.865	0.758	0.824	0.693	0.734
KNN_Sub	0.786	0.714	0.596	0.833	0.758	KNN_Sub	0.79	0.726	0.559	0.893	0.789
KNN_Weight	0.844	0.762	0.731	0.793	0.773	KNN_Weight	0.875	0.824	0.794	0.853	0.835
LogRegr	0.86	0.765	0.705	0.825	0.787	LogRegr	0.888	0.786	0.706	0.867	0.817
SVM_Cub	0.805	0.756	0.74	0.771	0.761	SVM_Cub	0.763	0.758	0.676	0.84	0.789
SVM_Lin	0.862	0.768	0.769	0.767	0.767	SVM_Lin	0.88	0.773	0.706	0.84	0.798
SVM_Quad	0.834	0.757	0.769	0.744	0.752	SVM_Quad	0.831	0.753	0.706	0.8	0.771
Tree_AdaBoost	0.835	0.756	0.644	0.868	0.798	Tree_AdaBoost	0.876	0.755	0.618	0.893	0.807
Tree_RF	0.855	0.767	0.702	0.833	0.792	Tree_RF	0.87	0.8	0.706	0.893	0.835
Tree_RUSBoost	0.831	0.755	0.663	0.846	0.789	Tree_RUSBoost	0.854	0.814	0.735	0.893	0.844
Tree_TotalBoost	0.852	0.797	0.721	0.872	0.825	Tree_TotalBoost	0.817	0.735	0.618	0.853	0.78

(a) Performance from the 10 fold cross validation of the models on the training set of our main dataset (CR). LT=10min DW=10min. (b) Performance of all classifiers trained on the main dataset training set (CR) and tested on the main dataset test set (CS). LT=10min DW=10min.

Classifier	AUC	AUC_OP	SENS	SPEC	ACC	Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.942	0.852	0.857	0.846	0.85	Discr_Lin	0.937	0.871	0.857	0.885	0.875
Discr_Quad	0.942	0.852	0.857	0.846	0.85	Discr_Quad	0.931	0.871	0.857	0.885	0.875
Discr_Sub	0.929	0.852	0.857	0.846	0.85	Discr_Sub	0.931	0.852	0.857	0.846	0.85
KNN_Coarse	0.948	0.852	0.857	0.846	0.85	KNN_Coarse	0.5	0.5	1	0	0.35
KNN_Cosine	0.911	0.854	0.786	0.923	0.875	KNN_Cosine	0.902	0.871	0.857	0.885	0.875
KNN_Medium	0.875	0.799	0.714	0.885	0.825	KNN_Medium	0.909	0.852	0.857	0.846	0.85
KNN_Mikowski	0.891	0.832	0.857	0.808	0.825	KNN_Mikowski	0.87	0.852	0.857	0.846	0.85
KNN_Sub	0.868	0.728	0.571	0.885	0.775	KNN_Sub	0.817	0.709	0.571	0.846	0.75
KNN_Weight	0.886	0.816	0.786	0.846	0.825	KNN_Weight	0.905	0.799	0.714	0.885	0.825
LogRegr	0.929	0.871	0.857	0.885	0.875	LogRegr	0.937	0.871	0.857	0.885	0.875
SVM_Cub	0.843	0.78	0.714	0.846	0.8	SVM_Cub	0.945	0.874	0.786	0.962	0.9
SVM_Lin	0.926	0.852	0.857	0.846	0.85	SVM_Lin	0.937	0.852	0.857	0.846	0.85
SVM_Quad	0.882	0.832	0.857	0.808	0.825	SVM_Quad	0.945	0.907	0.929	0.885	0.9
Tree_AdaBoost	0.9	0.854	0.786	0.923	0.875	Tree_AdaBoost	0.533	0.5	1	0	0.35
Tree_RF	0.915	0.926	0.929	0.923	0.925	Tree_RF	0.933	0.907	0.929	0.885	0.9
Tree_RUSBoost	0.897	0.835	0.786	0.885	0.85	Tree_RUSBoost	0.533	0.5	1	0	0.35
Tree_TotalBoost	0.951	0.874	0.786	0.962	0.9	Tree_TotalBoost	0.533	0.5	1	0	0.35

(c) Performance of classifiers trained on the training set of our dataset (CR) and tested on the test set of the Physionet challenge. (d) Performance of models trained on the Physionet Challenge training set (TR) and tested on the Physionet Challenge test set (TS).

Figure 0.3: Performance table

Dove la linea di base potrebbe rappresentare l'equilibrio dinamico del sistema nervoso autonomo mentre l'attività dei picchi è la risposta a una condizione specifica. In tale ipotesi, quindi, il valore assoluto di un picco non è paragonabile a quelli vicini a causa di cambiamenti della linea base, rendendo difficile fare ipotesi sul comportamento del baroriflesso e quindi su ipotetici cambiamenti emodinamici. Quello che, invece, è stato valutato come affidabile era l'ampiezza del burst rispetto alla linea di base, quindi quanto il picco di attività è in relazione all'attuale stato dinamico del baroriflesso. Purtroppo gli indici calcolati in questo modo non sono stati in grado di comprendere la differenza tra pazienti ipotensivi e no. È stato pensato, quindi, di fare il calcolo del baroriflesso

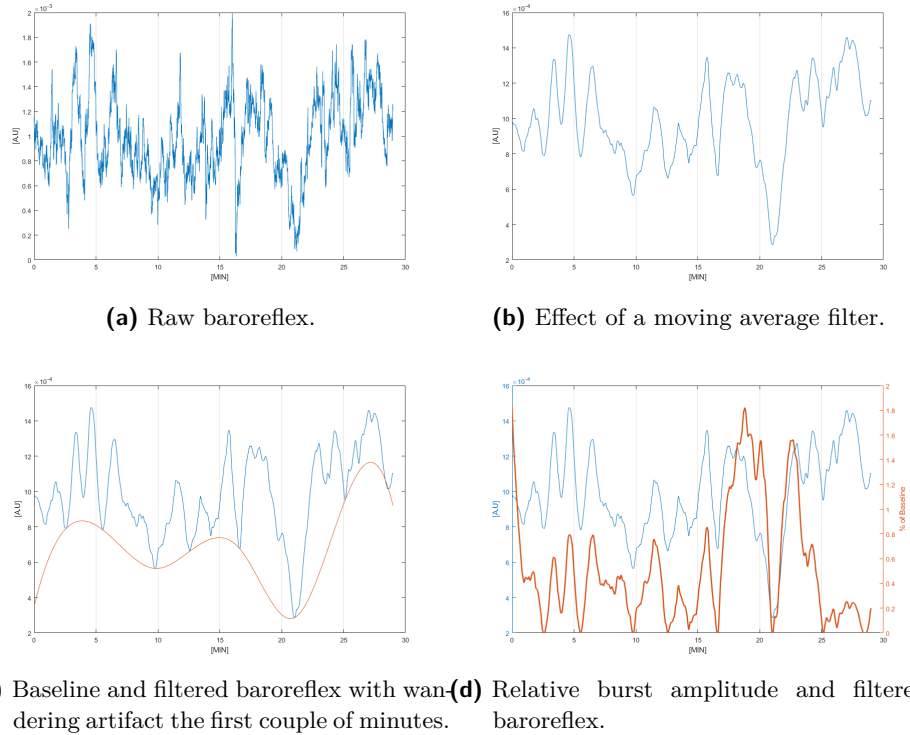


Figure 0.4: Baroreflex processing steps.

	AUC	Spec	Sens	Acc
$M5 + skewpks + kurtLW$	0.815 ± 0.013	0.834 ± 0.03	0.796 ± 0.03	0.811 ± 0.015
$M5$	0.742 ± 0.016	0.745 ± 0.09	0.739 ± 0.09	0.742 ± 0.026

Table 0.1: Performance table

nel dominio delle High Frequency (0.15 – 0.4 Hz). Gli indici statistici di linea di base e picchi relativi a essa del baroriflesso così calcolato, si sono rivelati molto utili nel miglioramento delle performance di classificazione. In particolare le features scelte per aumentare le prestazioni sono state la asimmetria (skewness) dei picchi e la kurtosis della linea di base. Usando un dataset costituito dalle cinque features già identificate nel precedente studio e le due nuove relative al baroriflesso ed un’altro con solo le prime cinque, sono stati ripetuti i processi di training e cross-validazione per cento volte. È stato possibile, in questo modo, quantificare l’effettivo aumento di prestazioni che il baroriflesso porta. Il risultato è visibile nella tabella 0.1.

È da tenere in considerazione come i segnali dei pazienti utilizzati in queste analisi non siano andati in contro a processi di esclusione secondo stato fisiologico o medicazioni. Questo è stato fatto di proposito per mantenere un approccio che richiami un ipotetico utilizzo nella vita reale di questi algoritmi. In un secondo momento è stata fatta un’analisi

delle prestazioni a seconda delle medicazioni ⁵ ricevute dai pazienti per rendersi conto di come il risultato di classificazione dipenda da queste.

Intervention		AUC	Sens	Spec	Acc
Vasopressors	Yes	0.786 ± 0.01	0.788 ± 0.01	0.785 ± 0.01	0.786 ± 0.01
	No	0.765 ± 0.01	0.848 ± 0.01	0.681 ± 0.01	0.75 ± 0.01
Sedative	Yes	0.786 ± 0.01	0.795 ± 0.01	0.776 ± 0.01	0.782 ± 0.01
	No	0.783 ± 0.01	0.821 ± 0.01	0.744 ± 0.01	0.772 ± 0.01
Ventilation	Yes	0.806 ± 0.01	0.803 ± 0.01	0.809 ± 0.01	0.807 ± 0.01
	No	0.766 ± 0.01	0.800 ± 0.01	0.732 ± 0.01	0.756 ± 0.01
Pacemaker	No	0.806 ± 0.01	0.794 ± 0.01	0.819 ± 0.01	0.811 ± 0.01
	Yes	0.735 ± 0.01	0.820 ± 0.01	0.651 ± 0.01	0.705 ± 0.01

Table 0.2: Tabella con indici di prestazione analizzate a seconda delle medicazioni. Ogni intervento ha due righe: la prima mostra gli indici di prestazione sui pazienti senza medicazione specifica (riga 'No'), e la seconda gli stessi indici dei pazienti che l'hanno ricevuta (riga 'Yes'). Questi valori sono il risultato di una cross-validazione 5-fold di un'analisi con discriminante lineare utilizzando tutti i 442 pazienti del cohort.

In Tabella 0.2 é possibile vedere quanto le prestazioni di classificazione varino a seconda se i pazienti hanno ricevuto o meno medicazioni. É possibile vedere come la differenza di prestazioni sia piú marcata tra pazienti che fanno uso di pacemaker e non. Anche la ventilazione meccanica mostra effetti simili ma meno marcati. Ciò potrebbe essere dovuto al fatto che i pazienti con ventilazione assistita, in genere, sono pazienti in condizioni critiche e/o pesantemente sedati. Per appurare l'effettiva causa saranno necessarie piú analisi. Vasopressori e sedativi, invece, sembrano non alterare troppo la classificazione creando solamente uno squilibrio tra sensibilità e specificità. Questo forse riflette una gravità piú lieve dello stato di salute di questi pazienti.

⁵ Somministrazioni di vasopressori, sedativi o uso di ventilazione meccanica o pacemaker

INTRODUCTION

In a hospital's Intensive Care Unit (ICU), healthcare professionals closely monitor patients with serious injuries, post-surgery trauma or unstable health conditions using bedside monitoring systems that continuously record electrocardiogram (ECG) and arterial blood pressure (ABP) waveforms. Acute hypotensive episodes (AHE) are among the most critical events that occur in ICUs and require effective, prompt intervention to avoid irreversible organ damage and death. Early prediction of AHE (within hours) could be of pivotal importance to establish timely and appropriate therapeutic interventions.

Severe hypotension can be the result of multiple causes, ranging from hypovolemia, hyperkalemia, hypothyroidism and ANS dysfunctions cardiac arrest syndrome, congestive heart failure, sepsis and others¹.

Hence, the origin of hypotension might be brought back to a broad spectrum of causes not considering comorbid ones, which highlight the complex dynamics behind such events.

Several studies tried to characterize hypotension and different models were proposed.

This thesis will examine the predictive power of the information gathered at the beat to beat level. In a first analysis we face the 2009 Physionet Challenge to develop methods to apply, in a second moment, to a much wider cohort of patients. This analysis will use features extracted from standard vital signs.

In a second analysis we will focus on a cohort formed by patients selected by the quality of their heart rate variability and systograms. On this restricted cohort we will extract features about the baroreflex to prove the predictive power of such kind of information.

In the second chapter are reviewed the nature of an AHE and the importance of AHE prediction in the ICU, and the work described in the literature related to the AHE prediction. The third chapter will describe the methods used to clean the raw data, annotate signals, extract features and classify patients. In the last chapter will be presented the performance of the methods and some consideration about the autonomic assessment in ICU and the influence of interventions on the classification results.

1 Yildiran et al., 2010

BACKGROUND

2.1 ACUTE HYPOTENSIVE EPISODES

Blood pressure is the pressure of flowing blood against the walls of vessels and is considered a vital sign. It is variable in time, showing long term patterns during the day cycle, which is considered an adaptive behavior to external and internal changes. A healthy person is thought to have a blood pressure lying in the 120/80mmHg¹, the first number indicating systolic pressure while the second diastolic. There is, though, great variability, where a certain pressure may result abnormal for a person while normal for another. Blood pressure is controlled by a range of different mechanisms: arterial stiffness, cardiac output, total peripheral resistance, moreover the short term response is also controlled by other physiological mechanisms, like baroreceptors. The ANS² has a strong influence in the vascular regulation, it has the ability to control vessels dilation and heart rate. There are several models describing the pressure behavior and the general control mechanism behind it.

Probably the oldest model describing the vascular system is the windkessel lumped model for the arterial properties. A typical windkessel model is the three elements windkessel model (wk3), taking as input ventricular pressure and aortic inflow from the heart. The wk3 model

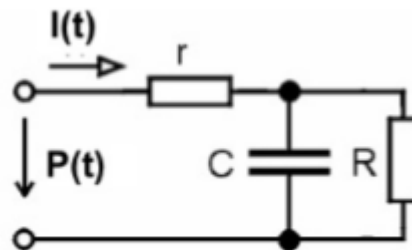


Figure 2.1: Three elements windkessel model. The capacitive element C represents the arteries compliance, the parallel resistive element is the hydraulic resistance to blood flow of the arteries while the the resistance in series is the aortic characteristic impedance.

is inspired from an RC electric analogue and of simple resolution. It models peripheral resistance and compliance and aortic compliance so to obtain the total impedance of the circulatory system. The windkessel model is a representation of the vascular mechanical system but

¹ De Backer and Vrints, 2017

² Autonomic Nervous System

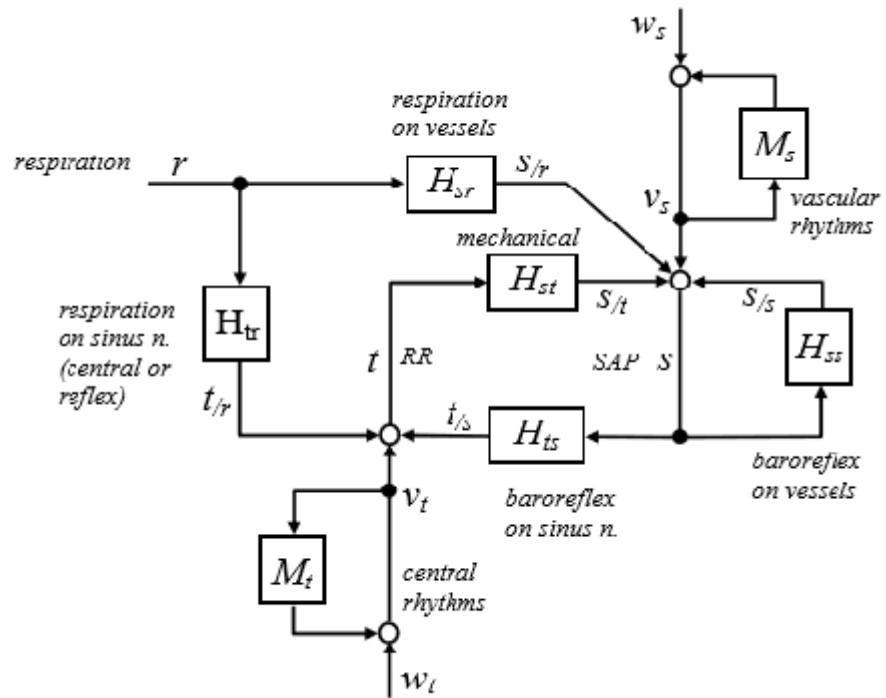


Figure 2.2: RR-SBP-RESP interaction model. The model explains the different interactions between respiration, systolic pressure and heart rate. One can see that heart rate is influenced by respiration (RSA) through H_{tr} and by systolic pressure (Baroreflex) through H_{ts} plus a colored residual. On the other hand, systolic pressure is dependent from heart rate, the so called mechanical effect, respiration and vessels configuration (Baroreflex on vessels in figure). RR intervals and systolic pressure are heavily interconnected.

it doesn't take into account ANS influence. Another approach, with broader reach, is the model proposed by Baselli. The Baselli model is a stochastic model which focuses into the interaction between heart rate, systolic pressure and respiration. It is a model fed by noise, and is able to examine all various interactions between the different variables. For instance the model provide the closed loop Systolic – Heart rate relations, so baroreflex sensitivity and mechanical effects, therefore including ANS influence. Blood pressure is essential in providing the correct amount of oxygen to the organs, therefore abnormal variations, being upward or downward variations, must be cautiously kept under control.

Blood pressure disorders belongs to two main groups: hypotension and hypertension. Hypertension increases heart workload and mechanical stress on vessels. Long standing hypertension might lead to heart and kidney failure. Hypertension is correlated with shortened life expectancy.

Hypotension is characterized by abnormally low pressure, after a certain point the blood flow is so weak that oxygen perfusion in the brain and other organs reach critical levels and might yield to permanent damage. It is therefore a critical condition to be carefully monitored.

Hypotension is not always a critical trait, for instance orthostatic hypotension due to postural variations is a short term condition which usually resolves by itself while the body adjusts for the gravitational pull. This thesis focuses into predicting and characterizing Acute Hypotensive Episodes in ICU, life threatening conditions.

Several studies are based on hypotension, trying to understand the physiology of this condition as well as identifying possible causes and solutions. As it turns out, hypotension is a complex condition, most of all because it is usually a symptoms of one or many other comorbid pathologies. Severe Hypotension might occur from blood loss, ANS disorders, heart conditions, medications, hypothyroidism, sinus dysfunctions, dehydration, cardiac output anomalies, bradycardia, septic shock and others (Kharod et al., 2014). Hypotension is significantly correlated with increased mortality, so that mortality of patients suffering from hypotension is more than twice that of normopressure subjects³.

2.1.1 ANS

The autonomic nervous system is composed by the sympathetic nervous system and parasympathetic nervous system. Sympathetic actions are associated with the flight and fight mechanism, increased pupil dilation, sweat, increased heart rate and blood pressure. The sympathetic response is abrupt and less delicate than its counterpart, the parasympathetic effect is that of heart rate relaxation and lowered blood pressure. Together, parasympathetic and sympathetic nervous system compose the sympathovagal balance. This is because when there is sympathetic activation there is parasympathetic deactivation and vice versa, they are complementary mechanism. ANS has influence on the sinus node, so it can influence heart rate, but it also has influence on baroreflex sensitivity. Baroreceptors are sensors mostly located in the carotid sinus and aortic heart; their purpose is to monitor pressure. The baroreflex is the response to pressure variations, for instance when there is a postural change, let's say from a supine position to standing, there is a sudden drop in blood pressure due to the gravitational pull.

Receptors detect this variation, resulting in vasoconstriction and increased heart rate for pressure normalization. Another example of baroreflex is the mechanical effect of respiration. During inspiration, pressure grows in the thoracic cavity; adding to that of blood pressure, resulting in a decrease in heart rate, the opposite occurs during expiration: pressure drops yielding to faster heart rate. This reflex is gen-

³ Physionet, 2009

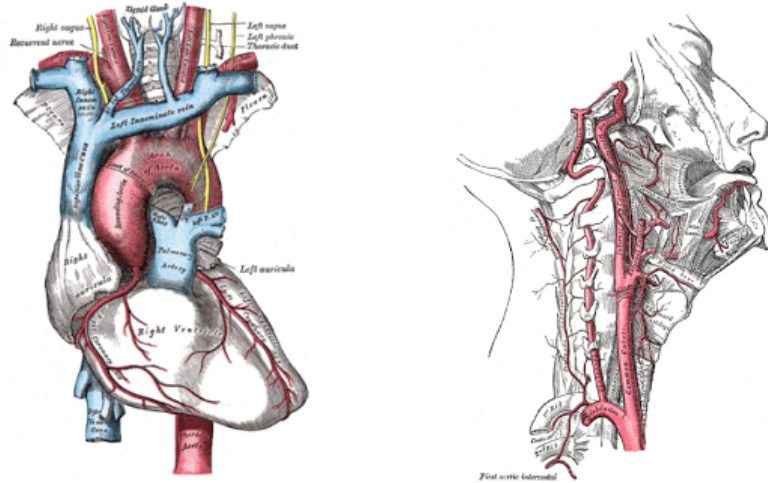


Figure 2.3: Aortic arch on the left and carotid sinus to the right are the sites of baroreceptors.

erally called Respiratory Sinus arrhythmia⁴. The blood pressure- heart

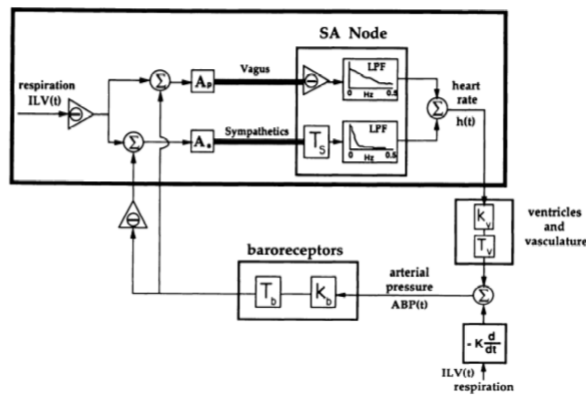


Figure 2.4: RSA model.

rate relation can be studied through a simple closed loop. Moreover by studying the interaction and spectra of given time series (tachogram and spectrogram) it is possible to identify some indicators. As indicated from the TaskForce, three spectral bands are associated with ANS activity, LF, HF and VLF. VLF are indicators of slow variations, maybe due to hormonal changes, while LF and HF might be due to parasympathetic balance. It has been shown for instance that in diabetic neuropathic patients, HF and LF presence is almost null, in accordance to the pathology. It comes not as a surprise, that a disorder in the ANS is reflected as disorders of vascular control, hence is a variable to keep under control.

⁴ RSA

2.1.2 *Medications*

Hypotension can result from response to medications. Vasopressors can cause hypotension due to adverse reactions. Medications for hypertension, like beta blockers, which impairs the normal control systems, might yield to hypotension. Without prior knowledge, it can be hard to distinguish between an ANS disorder and the effect of medications, both resulting in alterations of the underlying system.

2.1.3 *Septic Shock*

Septic Shock is a serious condition caused by infections. Sepsis naturally yields to severe low blood pressure caused by uneven distribution of blood in the system.

Other causes of hypotension are:

- Hypovolemia
- Pregnancy
- Anaphylaxis
- Nervous system damage
- Pulmonary embolism
- Anemia
- Lack of nutrients

2.2 STATE OF THE ART

Given the severity of the condition, in literature is possible to find several studies about the subject, each with different implications, methods and settings but with the shared goal to promptly alarm the clinician. Through the years many studies were proposed starting from simpler statistical descriptions to advanced deep learning methods.

Early studies were constrained by the lack of data, like the work of Bassale, 2001, which proposed an insightful waveform based analysis. The work of Jules Bassale inspects the behavior of the arterial blood pressure waveforms prior to the hypotension as mean to characterize hypotension itself in septic patients. The subjects pool is rather small, with only 5 patients, from which 28 hypotensive episodes were chosen. From this records the author had waveforms available for the analysis. Hypotension was defined as a drop in blood pressure of 20mmHg in less than 2 minutes. Each record consisted of 10 minutes of blood pressure waveform, 5 minutes before and 5 after the hypotension.

The author investigates how the shape of ABP changes as it approaches the event, and to do so he makes use of statistical indices and correlation functions. As a first step, waveforms are subdivided into 80 seconds windows from which than a distribution analysis is performed. The author highlight how the distribution of the window containing the hypotension is different from the Others, with the claim being supported by a significant difference in the skewness of the sample. Moreover, as shown in the paper, such variation can be visually assessed (See figure 2.5) as the waveform becomes less “smooth”. To

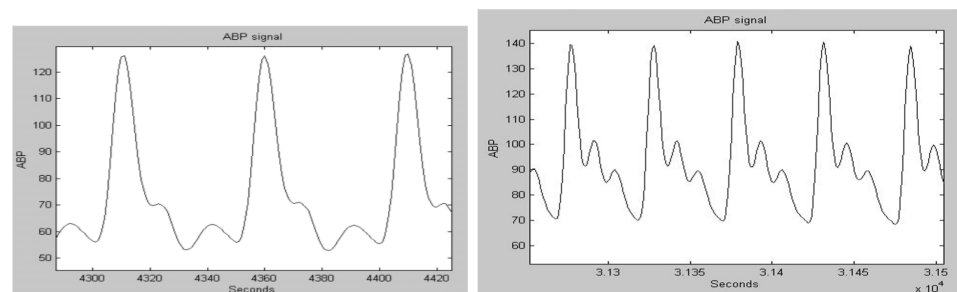


Figure 2.5: Waveform variation prior to hypotension.

even further characterize such changes the autocorrelation function is been used. The authors argue that a variation in the ABP shape will also induce a change in the linear coefficients describing the autoregression of the waveform, see figure 2.6. As final proof, the sminorv test shows effectively how shapes form the hypotensive window are actually part of a different population (p-value < .0001).

This paper was widely cited in different publications as it gives an insight of the hypotension phenomena.

Always in the contest of Sepsis, Crespo et al., 2002 evaluates variation of mean and variance as precursors of hypotension in 7 subjects of

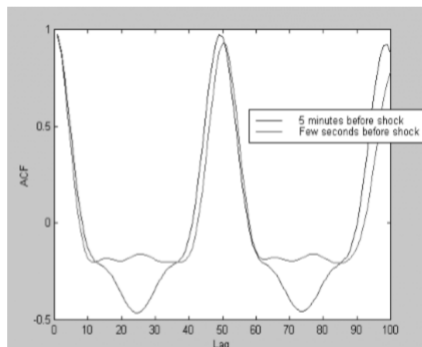


Figure 2.6: Autocorrelation coefficients difference due to impending hypotension.

young age (< 20 yrs). Of interest is the definition of hypotension: for each candidate episode ABP is divided in three intervals, a stable region of 300 seconds, a 10-120 seconds region of decreasing ABP and critical region, of 20 seconds, in which hypotension persists. An episode to be considered so, needs to have a blood pressure of maximum 80mmHg in the critical region, which also has to be at least 20% less than the pressure in the stable region and the drop must be more than 10 mmHg. Authors claims that this criteria were chosen to identify only critical hypotensions.

As for Bassale, 2001, many hypotensive episodes are recorded for each subject (with at least a 20 minutes separation).

Prior to the hypotension onset, ABP is sampled in five 20 seconds intervals distant 40 seconds from each other. Within each it is measured Pulse pressure, Peak to peak distance and the percussion wave slope (See Section 3.5). The idea is to prove that some of the described variables have a statistical significance ($p\text{-value} < .005$) difference between the fifth sampled interval and the remaining four. It was shown that ABP and percussion wave slope variances were significantly different.

By using bigger cohorts it possible to apply more complex models to the point where machine learning methods are an option. For instance the Laboratory of Computational Physiology -LCP- hosts annually a "challenge" where teams solves clinical problem using various classifiers, one of which focused on hypotension. Moreover the LCP, where part of this thesis was developed, is the creator of Physionet, a collection of large clinical databases and softwares. Apart from the database used for this study, which will be subject of a next chapter, Physionet is also the host of the yearly Physionet Challenge: An event that brings together data scientists, engineers and clinicians to discover new forms of cooperation. In 2009 Physionet organized the 10th Challenge in computing in Cardiology: Predicting Acute Hypotensive Episodes in ICU. The challenge was designed in two events, the first one about the identification of subjects under vasopressors medications, the second about

hypotensive episodes forecasts, the latter being the focus of our efforts. The dataset for the CinC⁵ consisted of 60 patients in the training set and 40 for the test, therefore can be considered a dataset larger, but still of medium size, compared to the ones previously described, and also doesn't comprehend multi-compilation observations in contrast to the previous analysis. As a result of this challenge, participants released their solutions to physionet, which can be easily accessed⁶. In the following sections are described some of the best approaches to the challenge.

One of this is Henriques and Rocha, 2009 solution. They attempt to solve the challenge with Generalized Regression Neural Network in an extremely interesting work, which also held the highest score ever for this event. A such can be probably considered the state of the art in Hypotension prediction.

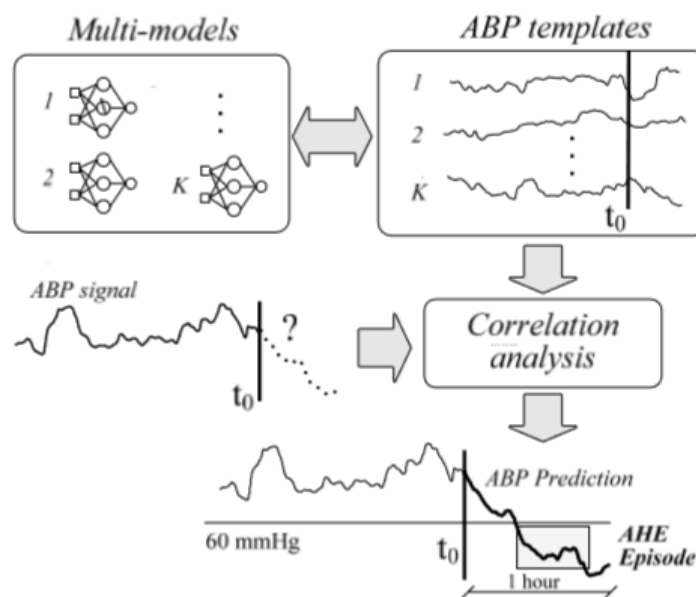


Figure 2.7: GRNN multi-model workflow proposed by Henriques and Rocha, 2009.

The proposed methodology applies generalized regression multi-models as an alternative to the simple autoregressive methods, which, quoting the authors, allows to “not recursively use model outputs as inputs for step ahead predictions. As result, prediction errors are not propagated over the forecast horizon and long-term predictions can be accurately

⁵ Computers in Cardiology.

⁶ <https://physionet.org/challenge/2009/papers/>

estimated". Multi-Models are discrete time non-linear auto regressive representations.

$$\text{step0} : y(k) = f_1(y(k-1), y(k-2), \dots, y(k-n)) \quad (2.1)$$

$$\text{step1} : y(k+1) = f_1(y(k), y(k-1), \dots, y(k+1-n)) \quad (2.2)$$

$$\text{step2} : y(k+2) = f_1(y(k+1), y(k), \dots, y(k+2-n)) \quad (2.3)$$

$$(2.4)$$

The expression in 2.4 can be rewritten:

$$\begin{aligned} y(k+2) &= f_1(f_1(y(k), \dots, y(k+1-n)), y(k), \dots, y(k+2-n)) \\ &= f_2(y(k), y(k-1), \dots, y(k+1-n)) \end{aligned}$$

Finally:

$$y(k+P) = f_p(y(k), y(k-1), \dots, y(k+1-n))$$

As demonstrated above, it can be proven that the P prediction step is not function of past outputs (no error propagation, Henriques and Rocha, 2009) but only the of the values till the current step. Instead, different autoregressive models are used for the prediction, to be precise one need P prediction models for a prediction of P steps ahead. Each of these models is approximated by a GRNN, thanks to the generalization capabilities of NN, moreover the great advantage of this methods, contrary to RNN, is that the back-propagation algorithm can be directly applied. This introduce a simplification, but an RNN might achieve even better results once learning difficulties are overcome.

To define the templates in the training phase, the designer needs to select a threshold, the size of the template and its order. The author propose that size and order are to be chosen so to minimize the prediction error. Hence, each record in the training phase has its own template. The model proposed is mono-variate, with only ABP, excluding clinical records and other waveforms. ABP is preprocessed and normalized. Then, the signal undergoes a correlation analysis with a series of templates elaborated in the training phase. The results will be many correlation coefficients relatives to the different templates. By using a fixed score, the most similar template, and the associated multi-models are chosen and used for the prediction. The power of this methodology is that it actually predicts the pressure values in the future, and not hypotension. Hypotension is defined afterwards.

While the work of Henriques and Rocha, 2009 focuses on a template base approach, Chen et al., 2009 proposed a simple index based classification for the challenge. Specifically all indices were drawn from blood pressure, yet again leaving out additional waveforms such as heart rate. Moreover apart from one index, all other are not optimized indices, meaning they are not optimally parametrized during the training phase. For the classification it self, each index is take singularly and used as the solely feature in the classification. The authors justifies the use of

simple indices taken from ABP only saying that most of the predictive information lies on pressure itself, and as such, they favor a pragmatic approach. Following the list with description of said indices:

- Index I: 5 minutes average of the mean arterial pressure before T0 (the forecast window)
- Index II: 5 minutes average of the ABP before the forecast window.
- Index III: Exponentially weighted ABP in the 10 hours prior to the forecast window. Of all features this is the only which is optimized in the training set, meaning that the time constant for the weighting was selected through ROC analysis to better classify training observations.
- Index IV: prediction of the ABP 30 minutes within the forecast window by means of linear interpolation.
- Index V: 5 minutes average of the diastolic pressure before the forecast window.
- Index VI: index function of index II and V, and acts as a voting system whereas when both this two indecies identify an hypotension then index VI is raised.

Accuracies and AUC in test and train for each index are in table 2.1.

Index	Training AUC	Test Accuracy
I	.76	.825
II	.75	.90
III	.82	.80
IV	.76	.90
V	.79	.925
VI	.75	.90

Table 2.1: Performance of Chen’s Indices.

Results are quite high in the test, although with big differences (from 32 to 37), moreover AUC in the training doesn’t seem to indicate a similar performance in the test (indeed the highest AUC in the training set yields to the lowest score in the test, maybe said feature doesn’t grasp the difference between the two classes or probably it is due to overfitting). Anyhow, Chen et al., 2009 proved that even a very simple method can yield to very good performance, and highlights the importance of ABP as a predictor of hypotension.

In the last section were described the best scores from the challenge they also shows two different approaches, a complex model and a simple index based classification. The two studies share the same dataset, the

challenge dataset, with a training set having the size of 60 patients. Overall the population size is not the largest, and it might not fully represent hypotension. Differently Jiang, Hu, and Wu, 2017 use a large dataset, with more than a thousand entries, for their study.

This paper proposes a classification algorithm based on the Hilbert-Huang Transform for feature extraction and Genetic Programming for the classification model. This method is applied on two datasets, the challenge dataset and a large dataset, always from MIMIC, with 2866 records. The fact that the challenge dataset is used comes in handy since it allows us to compare our model through the challenge, even if unfortunately the challenge test sets used from Jiang, Hu, and Wu, 2017 was incomplete (they could retrieve only 37 out of the 40 patients for the challenge test set).

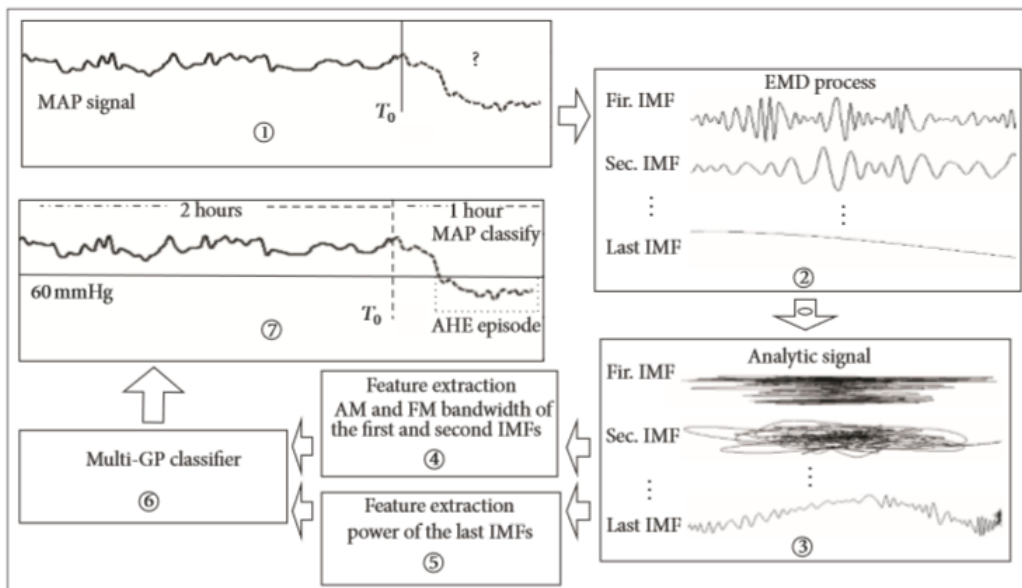


Figure 2.8: Workflow emphasizing the role of the HHT role.

The Hilbert-Huang transform is technique for time series decomposition into Intrinsic Mode Functions (IMF) in fashion similar to the Fourier transform and wavelet decomposition, but better suited for non-stationary and non-linear signals. The HHT is based on the empirical mode decomposition or EMD, able to break down the signal into a finite number of orthogonal basis, called IMF. The IMF roughly describe the harmonic component of the original signal, where the first component contains the highest frequency and so on. Each IMF than can be processed via the Hilbert Spectrum Analysis (or HSA) from which one can extract the Amplitude Modulation Bandwidth (AMB) and the Frequency Modulation Bandwidth (FMB) for each IMF. Along with the power of the last IMF, the AMB and FMB of each IMF acts as features to be fed to the genetic classifier for the classification of Hypotensive patients, which brings to the second Part of this paper: Genetic Programming.

Genetic Programming is born from evolutionary computation, and is method capable of “evolving” computer programs to achieve a designed purpose. Genetic Programming in itself is not particularly complex, indeed it interacts with tree structured computer programs by setting the tree structure in the optimal way (figure 2.9). Each tree is assigned

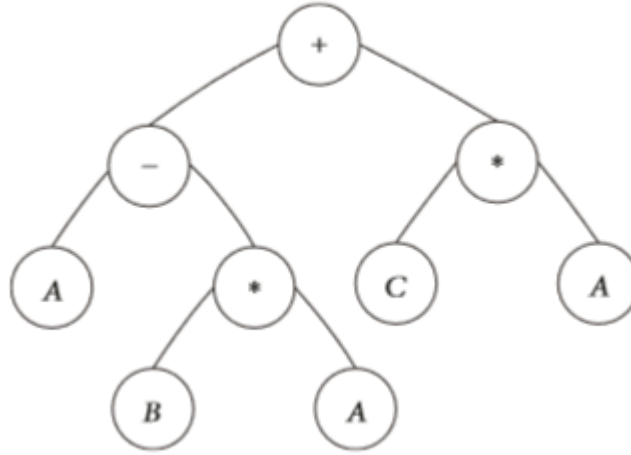


Figure 2.9: Simple example of a tree structured program for genetic programming.

with a set of actions, for instance the authors used:

$$+ \quad - \quad x \quad \div \quad \text{sqrt} \quad \text{exp} \quad \ln \quad ^2 \quad ^3 \quad ^{1/3} \quad \sin \quad \cos \quad \tan^{-1}$$

The algorithm will have to combine in certain way the features to achieve the goal following a cost function (fitness). The authors suggest that genetic programming “may be more powerful than neural networks and other machine learning techniques”. For the purpose of the study, as said earlier, features are the AMB and FMB of the different IMFs from the HHT. The training set is divided in many subsets (figure 2.10), on each subsets a genetic model is trained, and through a voting system, an observation is classified (if more than half of the models chose the observation as hypotensive, than it is classified as such). The training performance is applied with cross validation like methods, where the voting system is applied to one of these subsets of the training set itself. The results is a 34/37 accuracy in the challenge test set (91.89%) and of 82.41% in the large MIMIC dataset. What we think is remarkable about this study is the impressive sensitivity of around 85%, which is considered by us very high. Jiang, Hu, and Wu, 2017 proposed an effective method, on both a large sets and the challenge, although it is not clear to us their definition of hypotension, which we assume being the same of the challenge, moreover being that their challenge test set is not complete, we cannot compare our results with the one from the study. Also, although the high sensitivity is very important, it carries a lower specificity (although still very high) , hence an intrinsic unbalance in the classification performance in the two classes. For an

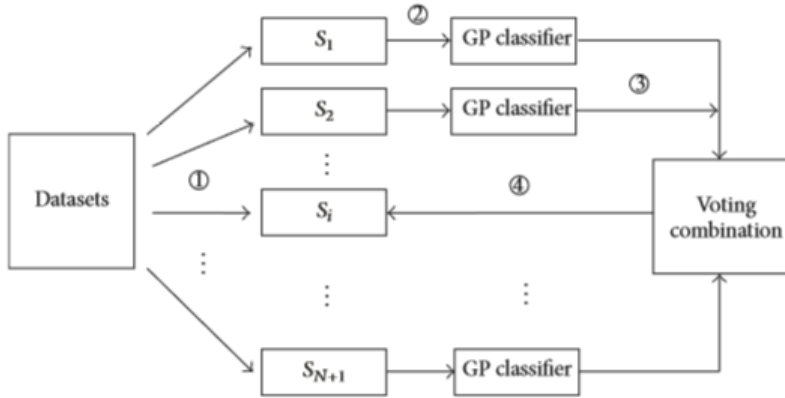


Figure 2.10: Classification logic.

overall assessment it would have been easier to have AUCs, but are not present in the paper. Overall, the performance achieved in the study is impressive.

In line with Jiang, Hu, and Wu, 2017 Work, Lee and Mark, 2010 cohort is made of 1311 records, therefore a quite large dataset. Data source was Always MIMIC. As it will be described in this paragraph, from the original pool of records are going to be selected up to 141000 subset with a mechanism called multi-compilation explained later. The design is to use a feed-forward classification network with three layers with the hidden one composed of 20 neurons. The choice of the model was justified by the fact that a three layers FNN is able to map any kind of input-output relation (Lee and Mark, 2010). Features were extracted from Systole, Diastole, Heart Rate, Pulse Pressure and relative cardiac output. Said features represents the different statistical moments (mean, standard deviation, skewness and kurtosis of each time series), wavelet coefficients, clinical information, such as age and medications and correlation at zero lag to estimate the coupling between time series. The large amount of features, though, prove to be challenging to be handled by a neural network because of the curse of dimensionality and collinearity (Haykin et al., 2009). Hence, PCA was applied for dimensionality reduction using 90% of explained variance.

Classification is performed in two datasets built differently based on the compilation method. With single compilation the author means that from each record it is taken at most one sample, hypotensive or normotensive, while with multiple compilation from each record are taken as many samples as possible using a sliding window. Quality of the samples was assessed using a simple interval of physiological limits. (ie. Pressure must be between 10mmHg and 200mmHg). It is clear that, by design, the multi-compilation dataset is much larger than the single compilation one, but also requires more attention in the train/test split. In the multi-compilation dataset samples from the same record could only belong to either the train or test set and not both,

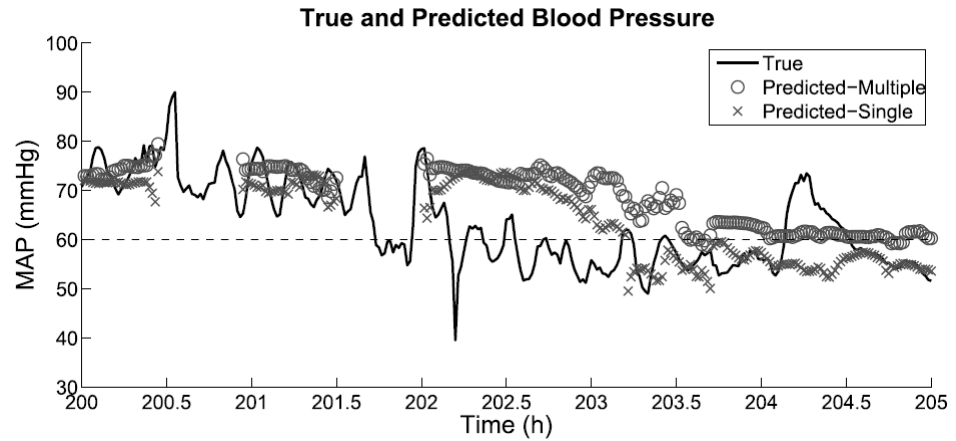


Figure 2.11: Predicted ABP versus actual ABP.

reason being that two samples taken from the same record and with relative small time apart would follow the same behavior, therefore classification performance would seem to overperform, while that is not true.

PCA⁷ was applied in the training set, and classification was performed in the PCA dimension with test observation being projected using the loading obtained from the training. The performance difference between single compilation and multi-compilation is high. On the multi-compilation set AUCs are 10% higher on average than in the simple compilation dataset. In the table 2.2 are shown the AUCs and sensitivities for the different methods.

	Single-compilation	Multi-compilation
AUC	.809 ± .042	.914 ± .018
Sensitivity	.745 ± .053	.812 ± .030
Specificity	.732 ± .036	.864 ± .014
Accuracy	.737 ± .026	.863 ± .014

Table 2.2: Comparison between single-compilation and multi-compilation performance one hour before hypotension onset.

But the study goes further, Lee also attempts to predict the value of ABP itself in the future, using the same NN architecture but with an hyperbolic activation function instead of a sigmoid one for dynamic range, see figure 2.11. The MSE achieved is of around 10% on average. Authors do not provide a comparison with the challenge.

By performance, this results are second only to Henriques and Rocha, 2009, although, as noted by the authors, the biggest limitation of the study is the quality assessment of the records, hence they did not excluded all corrupted data.

⁷ See Section 3.8.1

2.3 MIMIC

PhysioNet is one of the world's largest, most comprehensive, and most widely used repositories of freely available recorded physiologic signals and high resolution clinical ICU data, with related open-source software for research.

PhysioNet was established in 1999 to provide data and software resources in biomedicine to the research community.⁸

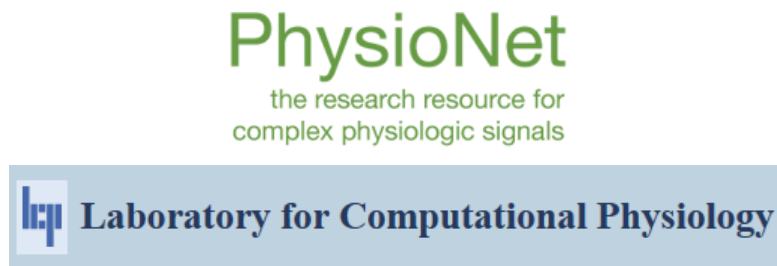


Figure 2.12: Physionet & LCP.

Physionet is the union of the following branches⁹:

- **PhysioBank** The Data archive, with clinical information and waveforms recorded from ICU patients.
- **PhysioToolkit** A collection of softwares, like the WFDB toolbox, for data driven analysis.
- **PhysioNetWorks** Incubation Laboratory.

The archive was created by the Laboratory For Computational Physiology, or LCP, where part of the development of this study took place. LCP is lead by Distinguished Professor Roger Mark and conducts research on improving health care through new data driven technologies and approaches.

Section 2.3 describes the MIMIC Database form which Records were pulled for this thesis.

Among the databases collected by PhysioNet, the biggest and most comprehensive ones are those of the MIMIC Family.

As of writing, MIMIC has reached its third upgrade, although for the thesis only a part of this last update was used while waveforms were pulled from the second version.

MIMIC is an acronym standing for: **Multi-parameter Intelligent Monitoring in Intensive Care.**

⁸ <http://lcp.mit.edu/physionet.shtml>

⁹ <http://lcp.mit.edu/physionet.shtml>

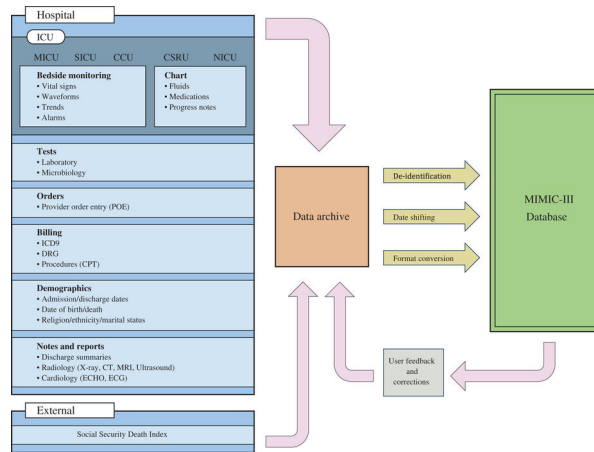


Figure 2.13: High-Level Structure of the MIMIC Database.

It contains Data gathered from the Intensive Care Unit with medications, vital signs, diagnosis, notes and patients' outcomes with approximately sixty thousand admissions¹⁰.

Before data was incorporated into the MIMIC-III database, it was first de-identified in accordance with HIPAA¹¹ standards using structured data cleansing and date shifting. Dates were shifted into the future by a random offset for each individual patient in a consistent manner to preserve intervals, resulting in stays which occur sometime between the years 2100 and 2200. Time of day, day of the week, and approximate seasonality were conserved during date shifting. The database has two main subsets, the Clinical Database and the Waveform Database.

2.3.1 MIMIC Clinical Database

The MIMIC Clinical Database is a relational database containing tables of data relating to patients who stayed within the intensive care units at Beth Israel Deaconess Medical Center.¹²¹³ The Clinical Database is made of 40 different tables, each bringing different information about a give subject, .

Within the Clinical Database were found valuable information for this study, such ICU admissions time and confounding factors, like vasopressors, sedation or mechanical ventilation.

Management and acquisition of Clinical Data from the Database is performed through PostgreSQL using SQL.

¹⁰ <https://physionet.org/physiobank/database/mimic3cdb/>

¹¹ Health Insurance Portability and Accountability Act

¹² <https://mimic.physionet.org/gettingstarted/overview/>

¹³ <http://www.bidmc.org/>

Table	Children	Parents	Columns	Rows	Comments
admissions	18	1	15	58,976	Hospital admissions associated with an ICU stay.
callout		2	24	34,499	Record of when patients were ready for discharge (called out), and the actual time of their discharge (or more generally, their outcome).
caregivers	7		4	7,567	List of caregivers associated with an ICU stay.
chartevents		5	15	330,712,463	Events occurring on a patient chart.
chartevents_1			15	38,033,561	Partition of charthevents. Should not be directly queried.
chartevents_10			15	9,584,888	Partition of charthevents. Should not be directly queried.
chartevents_11			15	470,141	Partition of charthevents. Should not be directly queried.
chartevents_12			15	265,413	Partition of charthevents. Should not be directly queried.
chartevents_13			15	39,066,570	Partition of charthevents. Should not be directly queried.
chartevents_14			15	100,075,138	Partition of charthevents. Should not be directly queried.
chartevents_2			15	13,116,197	Partition of charthevents. Should not be directly queried.
chartevents_3			15	38,657,533	Partition of charthevents. Should not be directly queried.
chartevents_4			15	9,374,587	Partition of charthevents. Should not be directly queried.
chartevents_5			15	18,201,026	Partition of charthevents. Should not be directly queried.
chartevents_6			15	28,014,688	Partition of charthevents. Should not be directly queried.
chartevents_7			15	255,967	Partition of charthevents. Should not be directly queried.
chartevents_8			15	34,322,082	Partition of charthevents. Should not be directly queried.
chartevents_9			15	1,274,692	Partition of charthevents. Should not be directly queried.
cpdevents		2	12	573,146	Events recorded in Current Procedural Terminology.
d_cpt			9	134	High-level dictionary of the Current Procedural Terminology.
d_icd_diagnoses	1		4	14,710	Dictionary of the International Classification of Diseases, 9th Revision (Diagnoses).
d_icd_procedures	1		4	3,898	Dictionary of the International Classification of Diseases, 9th Revision (Procedures).
d_items	8		10	12,487	Dictionary of non-laboratory-related charted items.
d_labitems	1		6	753	Dictionary of laboratory-related items.
datetimeevents		5	14	4,485,937	Events relating to a datetime.
diagnoses_icd		3	5	651,047	Diagnoses relating to a hospital admission coded using the ICD9 system.
diagcodes		2	8	125,557	Hospital stays classified using the Diagnosis-Related Group system.
housdays	8	2	12	61,532	List of ICU admissions.
inputevents_cv		4	22	17,527,935	Events relating to fluid input for patients whose data was originally stored in the CareVue database.
inputevents_mv		5	31	3,618,991	Events relating to fluid input for patients whose data was originally stored in the MetaVision database.
labevents		3	9	27,854,055	Events relating to laboratory tests.
microbiologyevents		5	16	631,726	Events relating to microbiology tests.
noteevents		3	11	2,083,180	Notes associated with hospital stays.
outpatievents		5	13	4,349,218	Outputs recorded during the ICU stay.
patients	19		8	46,520	Patients associated with an admission to the ICU.
prescriptions		3	19	4,156,450	Medicines prescribed.
procedureevents_mv		5	25	258,066	Procedure start and stop times recorded for MetaVision patients.
procedures_icd		3	5	240,095	Procedures relating to a hospital admission coded using the ICD9 system.
services		2	5	73,343	Hospital services that patients were under during their hospital stay.
transfers		3	13	261,897	Location of patients during their hospital stay.
40 Tables			634	728,656,685	

Figure 2.14: Data Skeleton for the clinical Database

2.3.2 MIMIC Waveform Database

The MIMIC Waveform Database contains recordings of multiple physiologic signals and time series of vital signs collected from bedside patient monitors in ICU. Most of the Records have (usually) multichannel ECG, ABP and less frequently respiration while in a best case scenario for each patient there can be up to eight physiological recordings at the same time.

Waveform quality is an issue as well as other recurrent artifacts in the Data. As reported from PhysioNet itself, major issues that could arise from MIMIC waveform analysis are:

- Gaps and patient identification: some records can have data from more than one patient. This is due to the bedside monitor not being reset after a patient discharge.
- Inter-waveform alignment problems: MIMIC was not designed for multi-waveform analysis¹⁴. Because of this there might be delays, mis-alignments and time varying delays between the different channels. Such artifact is quiet problematic for the evaluation of Pulse Transit Times¹⁵, and as explained in a next section, was used an exclusion criteria for records in the cohort.
- Varying Sampling Frequency in ECG: ECGs were originally sampled at 125Hz but later scaled using a technique called **peak-picking**. What happens is that the effective resolution of the ECG degrades from 12 bits to less than ten.

14 <https://physionet.org/physiobank/database/mimic3wdb/>

15 See appendix Pulse Transit Times or PTT definition.

Artifacts are dealt with in Section 3.2.



Figure 2.15: Waveforms samples as seen from Lightwave.

2.3.3 MIMIC Matched Subset

The MIMIC Waveform Database Matched Subset¹⁶ contains all Waveforms that have been associated with the Clinical Database records. This allow to cross to Clinical Database with Waveform Database. All used records belong to this database.

¹⁶ Saeed et al., 2011 Johnson et al., 2016b

METHODS

3.1 EXPERIMENTAL SETTINGS

3.1.1 Definition of Hypotension

An *Acute Hypotensive Episode* is defined as any period of 30 minutes in which at least 90% of the Mean Arterial Pressure (MAP¹) is below or equal to 60 mmHg².

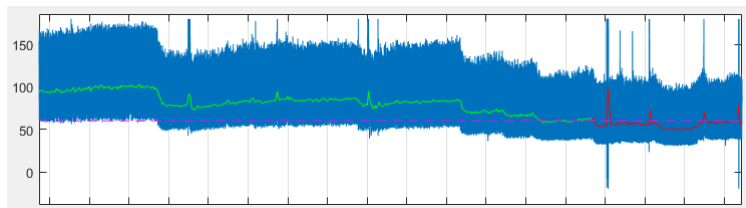


Figure 3.1: Example of Hypotension from a MIMIC record. The green line represent the MAP of the waveform, it turns red during the AHE. The time interval between grid lines is 30 minutes. The intermittent pink line labels the 60 mmHg threshold.

The above definition was suggested by the Physionet Challenge Team and it was chosen for this study for two reasons:

1. A considerable amount of literature is based on this definition, allowing for results comparison.
2. Methods can be tested directly on the Physionet Challenge dataset.

Moreover the simplicity of this AHE rationale lowers the likelihood of false AHE detections or mis-detection, specially in the contest of MIMIC waveforms where the noise influence during features extraction is already relevant.

On the other hand one could argue that such definition does not consider the normal pressure variability between patients: some people might naturally have very low blood pressure without any physiological damage or, on the contrary, relatively high pressure associated with the hypotensive condition: clearly the choice of a fixed threshold has its downsides.

However, on average pressure below 60 mmHg is considered hypotensive regardless of population diversity, often leading to irreversible tissue damage and death due to low organ perfusion (Bassale, 2001).

¹ See Section 3.5.6 for MAP definition.

² <https://physionet.org/challenge/2009/>

Another risen criticism toward this definition, but in the contest of AHE prediction, regarded the ease to identify AHE subject solely on their MAP, using the assumption that hypotensive patients already have very low blood pressure even much before the AHE onset. For instance a subject with a MAP in which 89% of its values are below 60 mmHg is likely to be hypotensive sometime in the future based on the current definition.

This last point is partially true: some subjects fit the "always-low-blood-pressure" profile, but it accounts only for a subset of AHE typologies that were observed in the cohort.

As shown in Section 2.2, different AHE definitions were already proposed: some tried to define a relative threshold (Crespo et al., 2002) others a more relaxed definition of the current one (Ghassemi, 2011). Given the complexity of the cardiovascular system, it will be challenging to find a universally accepted definition of hypotension that completely grasps its nature.

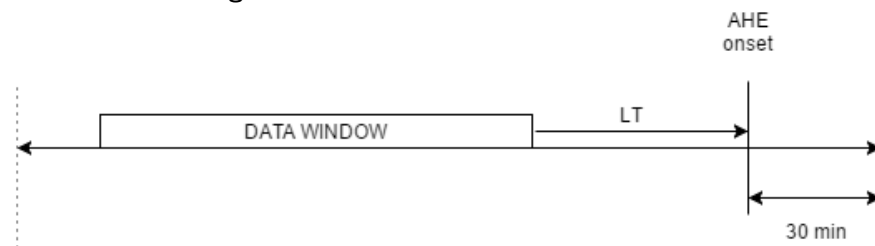
3.1.2 Data Window and Lead Time

The overall settings for the experiment rely on two parameters only: **Data Window and Lead Time**.

The Data Window (DW) is the time interval of a given size of the record from which the features for the classification are extracted. The time interval between the upper limit of the Data Window and the onset of the hypotension is the Lead Time (LT) and represents how far in time we want to predict the onset of the AHE. All the information contained in the LT is inaccessible.

As LT increases the harder the prediction³.

Figure 3.2: Data Window and Lead time.



For a better understanding of the phenomena in the general framework, various lead times and data windows were used. Precisely up to 30 datasets were built with different combinations of five different data windows and six different lead times.

The choice for an even higher number of combinations, specially for longer intervals, was constrained by the records length of only five hours: with a 120 min lead time the biggest allowed data windows

³ Chen et al., 2009

is of 180 minutes.

It was decided to investigate lead times and data windows with different time scales.

The influence of data windows and lead times will be discussed in Chapter 4.

3.2 PREPROCESSING

Waveforms were recorded with a sampling frequency of 125 Hz with 8, 10 or 12 bit resolution and stored with the WFDB (Silva and Moody, 2014) supported format.

All channels, but ECG in particular, present various degrees of noise, from 60 Hz noise, movements artifacts, trigger signals and so on. Moreover noise not always appears concurrently in all of the channels.

Waveforms were filtered, fed to a SQI⁴ logic and then annotated.

3.2.1 ECG

ECG was treated so to be fed to a Pan-Tompkins⁵ based algorithm, see Section 3.5 for details.

ECG is first filtered using a bandpass butterworth filter designed to get rid of the baseline wander and high frequency noise.

The lower and higher cutoff frequency were chosen to be 8% and 24% of the Nyquist frequency (62.5 Hz) at respectively 5 Hz and 15 Hz. The filter order was chosen to be 3 for performance reasons (while maintaining good results).

	f(Hz)	% of Fmax
FLow	5	8
FHigh	15	24

Table 3.1: Butterworth Bandpass cutoff frequencies.

To minimize shape deformations, *zero-phase distortion*, it was used the zero-phase digital filter technique⁶. The signal was filtered in both forward and reverse directions. This process generates the following effects:

- Zero-phase distortion
- The Transfer function of the final function equals the squared magnitude of the original
- The filter order is effectively doubled.

⁴ Signal Quality Index, See Section 3.3

⁵ Pan and Tompkins, 1985

⁶ Oppenheim, 1999

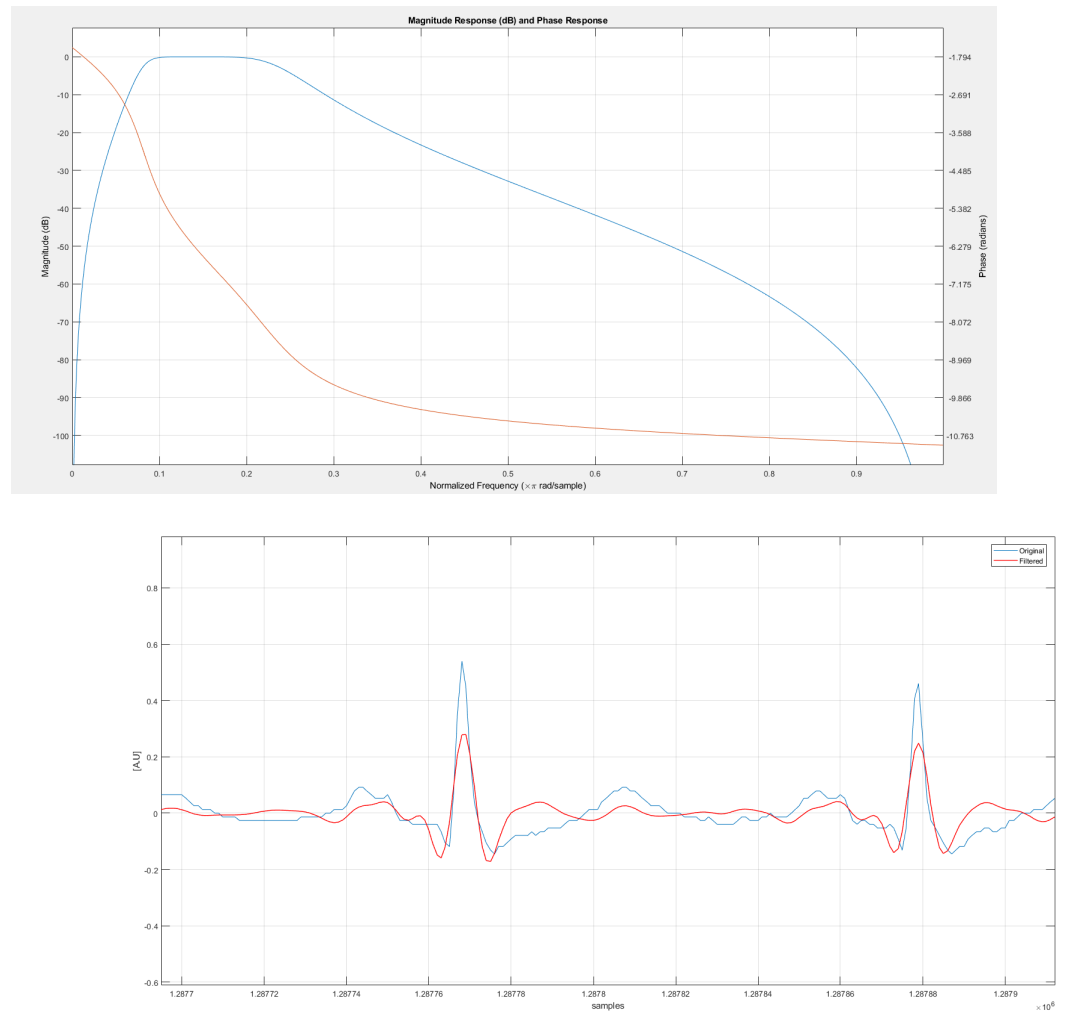


Figure 3.3: Magnitude and phase response for the ECG filter and filtering effect on the QRS.

Raw ECGs also present missing values. This can be handled in two ways depending on the size:

- Interval of at most a couple of samples: missing values are assigned to zero (common case).
- Longer intervals: missing values cannot be replaced, and the waveform must be treated accordingly.

3.2.2 ABP

Blood Pressure underwent a third order lowpass Bessel filter before the delineation process⁷. The cutoff frequency was chosen to be 25Hz. The advantage of the lowpass bessel filter is that delay up to the cutoff frequency is approximately constant limiting distortion effects. Compared

⁷ See Section 3.5, Automatic Delineator for Blood Pressure.

to the back-to-back filtering method this saves some computation time. Also, for blood pressure, missing values were problematic, even more than in the ECG: actual values of the ABP signal are important while in the QRS we were only interested in its position.

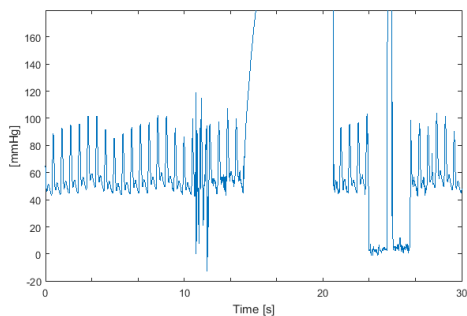
A simple but efficient fix for small intervals is to assign to missing values the average from the boundaries of the region by means of interpolation. For large intervals on the other hand, there is little that can be done since all values are lost. If the interval is large, the record has to be discarded.

3.3 SIGNAL QUALITY INDEX

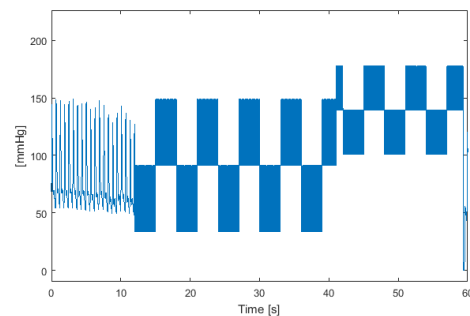
Since the signals we use come from the ICU, it is not unusual that waveforms show artifacts of different natures, with non Gaussian and non stationary properties. Both ECG and ABP are affected by such kind of artifacts.

The artifact that we want to identify are:

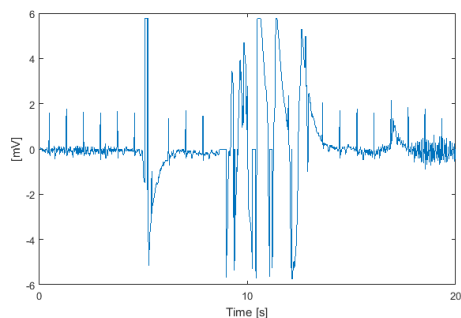
- * Saturation;
- * Movement artifact;
- * Trigger signal.



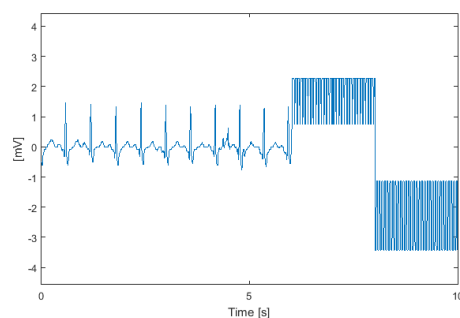
(a) Exemple of saturation in ABP signal



(b) Trigger signal in ABP



(c) Movement artifact in ECG



(d) Trigger signal in ECG

Since we use features extracted from the signal to forecast the onset of a pathology, it is crucial to be sure that the signal under analysis

is clear of artifacts that could alter the measure of the characteristics that describe the signal. To facilitate subsequent processing of peak detection algorithms, it was thought to use an algorithm that was able to figure out whether the annotations produced were reliable or not. This is also used to make the synchronization operation between ECG and ABP annotations easier. It was therefore necessary to develop an algorithm to ensure the goodness of the waveforms of both ECG and ABP signals.

The algorithm works with the same pattern on both signals:

- Beat to beat distance analysis;
- Analysis of characteristic of signal

3.3.1 *Beat to beat distance analysis*

Processing component common to both SQI algorithms, ECG and ABP, that is performed in the same way for both signals. The program analyzes the heart-rate variability and mark as suspicious the beats with temporal distances of less than 200 ms, and the signal portion where the distance between two beats is greater than three seconds (Pan and Tompkins, 1985).

3.3.2 *Analysis of characteristics of signal*

The program at this point makes a distinction on the processing to be made depending on whether the signal to be analyzed is either ECG or arterial pressure.

In case of the ECG, in addition to knowing the position of the R peak, it is not possible to extract other features that can be considered reliable and useful to determine the goodness of the trace under examination. It was decided therefore to consider the base line of the signal. The program performs a 5 Hz lowpass filtering on the signal. When the base line thus obtained passes an empirically determined threshold (1.5 mV) into the module, the non-reliability flag is raised for that signal portion. In case of arterial pressure, given the morphology of such signal, it is possible to individuate some characteristics that can be used to ascertain the presence of artifacts. These characteristic are:

- Diastolic pressure (P_{diast})
- Systolic pressure (P_{syst})
- Pulse pressure (PP)

The listed physical quantities are compared with physiological values for each beat in the track. A beat is labeled as non-physiological when:

- $P_{syst} > 170$ mmHg

- $P_{syst} < 50$ mmHg
- $P_{diast} > 110$ mmHg
- $P_{diast} < 15$ mmHg
- $PP < \overline{PP} - 4 \times \sigma_{PP}$

In which \overline{PP} and σ_{PP} are mean and standard deviation of Pulse Pressure.

3.4 EXCLUSION CRITERIA

Records were downloaded from the MIMIC Waveform Database Matched Subset (see Section 2.3 for details) using the WFDB Toolbox provided by PhysioNet⁸ and automatically annotated (Section 3.2).

Only records with five hours of contiguous ABP and ECG were chosen. Additionally, for hypotensive subjects only the first episode was considered. Finally all records were visually inspected and only those with no to little artifacts were selected to be included in the final cohort, see Figure 3.4.

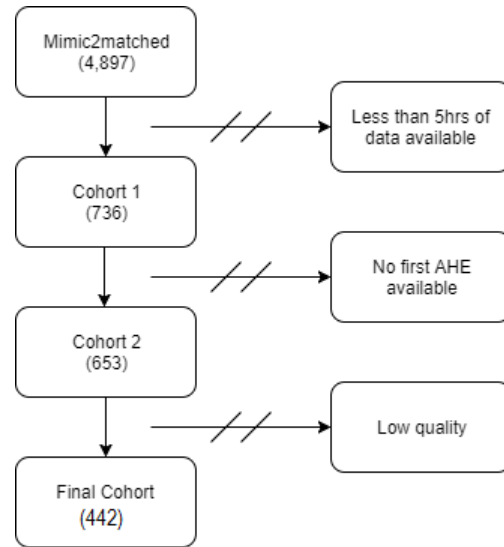


Figure 3.4: Exclusion criteria.

In the end, the cohort reached a size of 442 subjects. See Table 3.2 for an overall statistical description. The population characteristics are in line from what reported in (Li, Mark, and Clifford, 2009), interesting is to observe how mortality is more than twice for patients who suffered AHE at least once compared to the control population.

Injection of Vasopressors is statistically correlated ($p < .5$) to hypotension; this was expected as clinicians try to rise blood pressure in those with already low blood pressure.

⁸ <https://physionet.org/physiotools/matlab/wfdb-app-matlab/>

Age is significantly correlated with hypotension, although it doesn't seem to have predictive power when included for classification; moreover the scope of this study is toward a methodological approach to general hypotension prediction.

Pacemakers and sedation were equally present in both population, while ventilation is almost significant ($p < .09$).

	C	H	p-value	unit
%	68.33 (303)	31.45 (139)		%(num)
Age	62.01 ± 14.28	68.78 ± 13.70	> .0001	years
Gender	1.23	1.60	.2162	M/F
Vasopressors	14.57	22.46	.0413	%
Pacemakers	28.81	29.71	.8473	%
Sedation	19.54	23.91	.2957	%
Ventilation	50.33	59.42	.0767	%
Mortality	9.93	25.36	> .0001	%

Table 3.2: Cohort Description, C stands for Control population while H for Hypotensive population.

The final cohort is unbalanced class-wise with only 31.54% of records being hypotensive. This skew will require the need of some tweaks further in the classification chapter.

A second, smaller cohort, described in Table 3.4, was used to further investigate baroreflex sensitivity with a pool of waveforms of the highest quality.

3.5 ANNOTATIONS

3.5.1 *RRI*

For ECG peak detection were investigated two routines: Pan-Tompkins (Pan and Tompkins, 1985) and gQRS(*gQRS*). The former is a well know and simple peak detector while the second is an evolution of the first one. For the first part of the project it was used gQRS but then it was discovered that Pan-Tompkins for given tasks behaved better and it was more suited for the job, moreover it was much faster.

Pan-Tompkins is a three-layered algorithm:

- Linear Processing: Bandpass filtering, derivation and moving window integrator.
- Non-Linear Transforms: Amplitude squaring.
- Decision Rule Algorithm: Adaptive thresholds and T-wave discrimination.

The squaring of the derivative makes all the points positive and it emphasizes higher ECG frequencies. The window integration fuses the QRS in one single wave, easing the detection process; although it is important not to set a too large moving window since it would also fuse the T-wave. Therefore two signals are simultaneously scanned, the original ECG and the processed ECG, using two sets of thresholds, see Figure 3.5. Each set is composed by two limits, an higher one and a lower one. The latter is a search back threshold triggered when no R-peak was found.

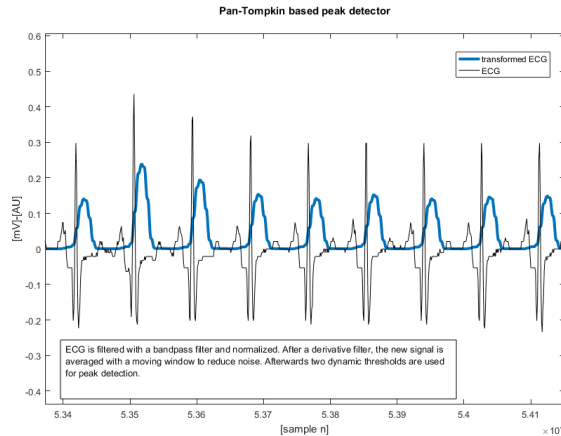


Figure 3.5: The waveforms studied by pan-tompkins.

3.5.2 *Systole, Diastole and Percussion Wave Onset*

One of the biggest challenges in the annotation process was to find a solid ABP waveform detector able to withstand MIMIC Records; indeed it took a great amount of time to find/build/train/test different algorithms; but we decided it was better to spend more time looking for a good solution rather than to relying on an approximate one and then use interpolation.

For the first attempt we relied on the first ABP derivative for markers; although this works in normal controlled conditions it is not robust enough for pathological patients or very noisy records. Afterwards, we used wavelet decomposition for diastolic valleys identification with much more promising results, only drawback: it was terribly slow, because of the Mallat algorithm requiring the application of filter banks. Then we designed a new algorithm (gpABP) based on a non-linear adaptive threshold function of RR-intervals and derivative peaks; the performance was similar to that of the wavelet analysis, but an order of magnitude faster, but still not sufficient.

Finally we were introduced to the Pulse Waveform Delineator (PUD) (Li, Dong, and Vai, 2010). Similarly to Pan-Tompkins, PUD works concurrently on two channels, the ABP itself and its derivative. It looks

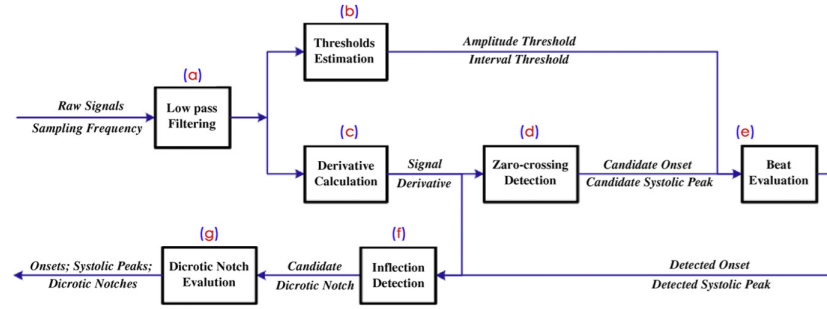


Figure 3.6: PUD workflow.

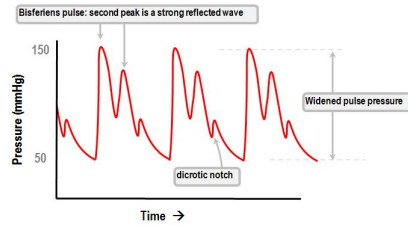


Figure 3.7: Bisferiens pulse, a challenging waveform to automatically annotate.

for pairs of inflection points and zero-crossing to identify systolic peaks and diastolic valleys using dynamic thresholds, then it searches back for dicrotic notches.

Of all the option we have investigated, PUD was the only capable to reliably annotate and recognize Bisferiens pulses⁹, plus it was reasonably fast, see Table 3.3 for comparisons.

Type	Controlled Conditions	Irregular Rhythms	Bisferiens Pulses	Robustness to noise	Speed
First Derivative Peak Detection	Good	Fail	Fail	Fail	Sufficient
Wavelet Decomposition	Excellent	Excellent	Sufficient	Good	Fail
gpABP	Excellent	Excellent	Fail	Sufficient	Excellent
Zong et al., 2003	Excellent	Excellent	Sufficient	Sufficient	Sufficient
PUD	Excellent	Excellent	Good	Good	Good

Table 3.3: Scoring of different algorithms.

3.5.3 Pulse Transit Times

Pulse Transit Times were defined as the time interval between the R-Peak and peak of the maximum derivative in the following pressure waveform. Although this definition was simple, the actuation of it, on the other hand, required great effort. Because the PTTs are the bridge between ABP and ECG, We had to be sure to make the correct assignment of R-peaks and ABP waveforms; step necessary for connectivity features like the Baroreflex.

⁹ <https://medical-dictionary.thefreedictionary.com/bisferiens+pulse>

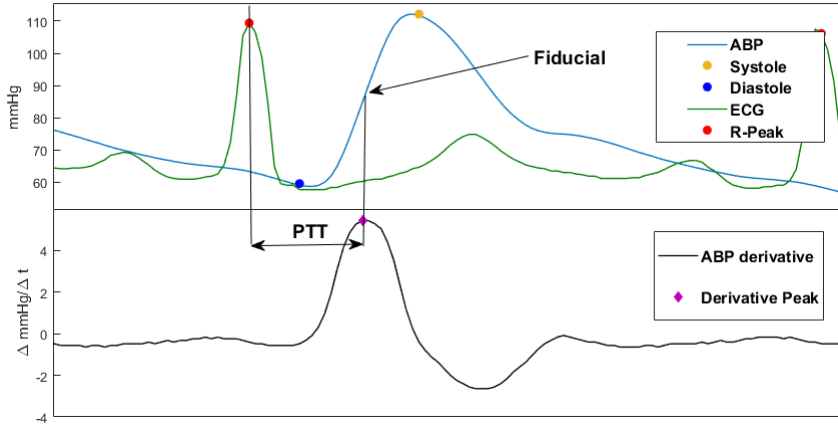


Figure 3.8: Visual description of Systole, Diastole, Percussion Wave Onset (fiducial point in figure) and Pulse Transit Times.

To assign to each R-peak the relative ABP waveform, we built a point process-like framework assuming Gaussianity¹⁰ to model the RR-distribution. We used Maximum Likelihood estimates for the parameters used than to compute the probabilities for each event.

At this point we had automated and streamlined all the annotation process form which we had R-peaks, Systoles, Diastoles and PTTs for each single beat in the five hour long record window.

Having those basic time series, we reached a level where we were confident we could extract high quality features. See Figure 3.8.

The biggest limitation regarding PTT is the quality of the features: the low sampling frequency¹¹ causes a relevant quantization error in the PTT estimation, artifact that can be seen in Figure 3.9e. The estimated SNR when considering quantization noise only is 25.9 dB, which is an optimistic estimate that ignores peak detectors and ABP-ECG synchronization errors. Assuming an average PTT of 200ms and considering a resolution of 8ms then the quantization error represent 4% of the value, which is acceptable for mean values estimates but completely destroys the dynamic of the signal¹².

3.5.4 Pulse Pressure

Pulse Pressure is difference between the Systolic Pressure and the Diastolic pressure for a given beat:

$$PP_i = SBP_i - DBP_i \quad (3.1)$$

Pulse Pressure is itself a useful indicator: high blood pressure is related to stiffness of the aorta (*What is pulse pressure? How important*

¹⁰ We knew that the best option was HDIG, but at price of higher computation times and moot performance increase.

¹¹ 125Hz.

¹² Only slow oscillations might be taken into account.

is pulse pressure to your overall health?); the greater the pulse pressure the more stiff are vessels. Low pulse pressure is a sign of decreased cardiac function. PP is also related to heart conditions and cardiovascular problems and can be used as a predictor of death (Yildiran et al., 2010).

3.5.5 Relative Cardiac Output

To compute Cardiac Output¹³ We relied on the Liljestrand nonlinear compliance both for its good estimation (Sun et al., 2005) and for the easier compliance with our data format¹⁴:

$$CO_i = K \frac{SBP_i - DBP_i}{SBP_i + DBP_i} \frac{1}{RR_i} = K \frac{PP_i}{SBP_i + DBP_i} f_i \quad (3.2)$$

3.5.6 Mean Arterial Pressure

Mean Arterial Pressure, MAP, was defined as the arterial blood pressure one minute average without overlaps between windows following Physionet guidelines¹⁵.

$$MAP_k = \frac{1}{\Delta k} \int_k^{k+\Delta k} abp(t) \delta t$$

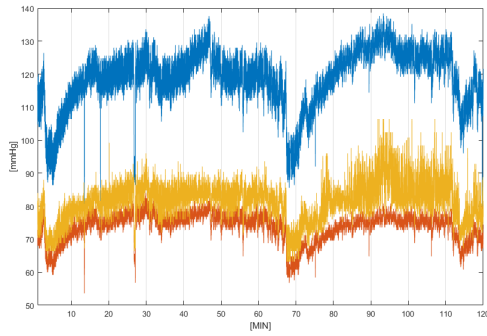
with $\Delta k = 60sec$.

AHE were defined on the MAP making it one of the most predictive time series under analysis. MAP dynamics are often hard to analyze both at the inter and intra patients level.

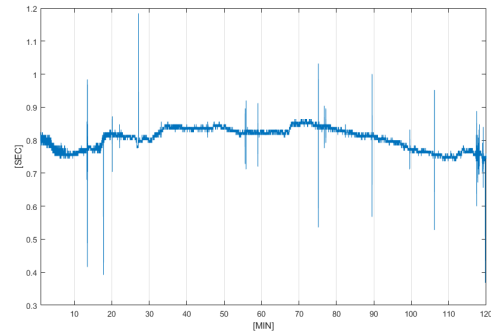
¹³ CO

¹⁴ The formula allowed us to use of all the beat-to-beat information resulting from our processing.

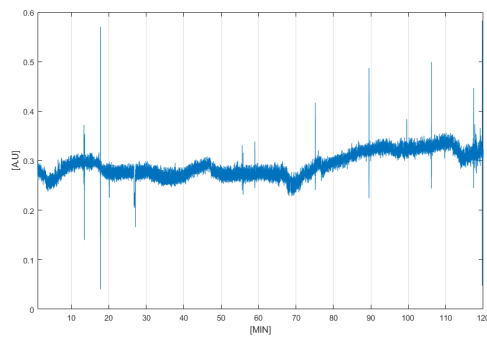
¹⁵ <https://physionet.org/challenge/2009/>



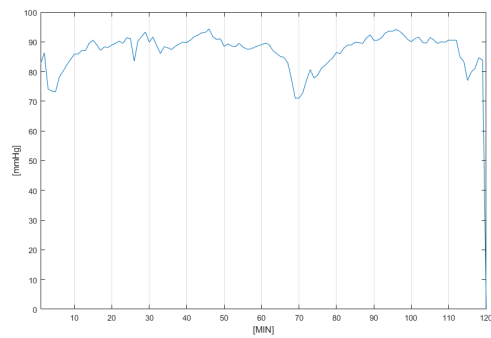
(a) Systole (Blu) , Diastole (Red) and Percussion Wave Slope Pressure (Yellow).



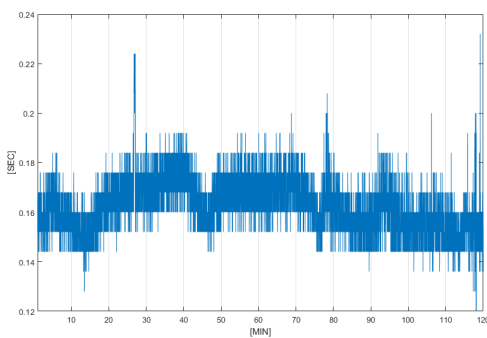
(b) RRI.



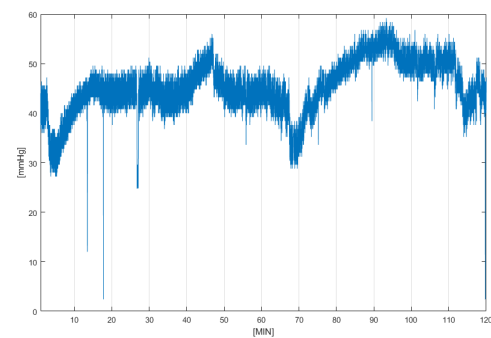
(c) Cardiac Output.



(d) MAP.



(e) Pulse Transit Times.



(f) Pulse Pressure.

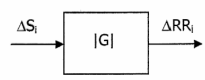
Figure 3.9: Sample time series form Record s01418-2947-11-14-12-10.

3.6 MODELING

3.6.1 Point Process

The baroreflex is an homeostatic mechanism that helps to regulate blood pressure: When it increases, the baroreceptors (sensors mainly present in the aortic arch) trigger an autonomic reflex that reduce the heart rate so that the blood pressure could return to a proper value. From a mathematical point of view the baroreflex can be seen as a coefficient that quantify the influence of the blood pressure on the heart rate.

There are various way to estimate this gain. One simple method to estimate this gain is measuring the spectral power of the heart rate variability and systogram in the Low Frequency band (LF = 0.04 - 0.15 Hz) and doing the ratio between the two:

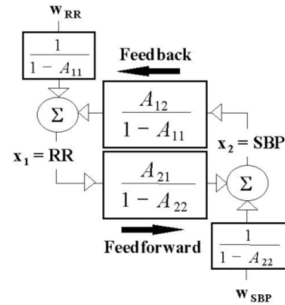
$$\alpha = \sqrt{\frac{P_{LF}(RR)}{P_{LF}(S)}} [ms/mmHg] \quad (3.3)$$


Where:

- $P_{LF}(RR)$ is the spectral power of the low frequencies of the tachogram;
- $P_{LF}(S)$ is the spectral power of the low frequencies of the systogram;

This is an evaluation of the baroreflex in a open loop system: the amplitude of systoles have an effect on the heart rate. In this situation there is no feedback from the variation the heart period ΔRR to a variation of blood pressure ΔP .

A more sophisticated model allows the closed-loop computation of the baroreflex gain α taking into account the feedbacks. In this case the baroreflex can be computed as the absolute value of the transfer function that describe the transfer of information from systole to heart period (H_{ts}).



$$\alpha = \left| \frac{A_{12}}{1 - A_{11}} \right| \quad (3.4)$$

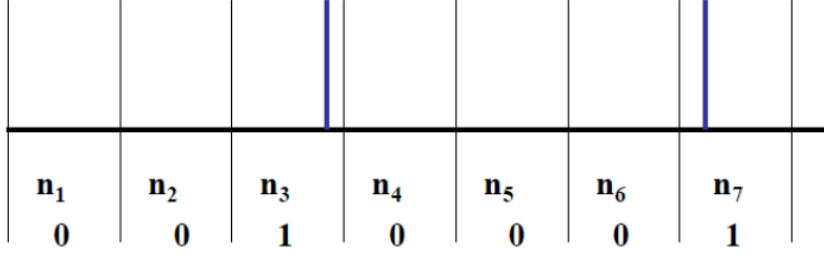


Figure 3.10: Description of a point process in time domain: the process is a succession of events in continuous time.

This requires to build a mathematical model that describe the dynamics between the two systems. Of course it is possible to do this in various way. The model chosen in this study is a Point Process model¹⁶: a point process is a discrete event happening in continuous time. It can be seen that this definition suit perfectly the succession of heart beats.

Assuming history dependence, the Point Process model that express the waiting time until the next heart beat is:

$$p(t) = \left[\frac{\theta}{2\pi t^3} \right]^{\frac{1}{2}} e^{-\frac{\theta(t-u_j-u_t)^2}{2\mu_t^2(t-u_j)}} \quad (3.5)$$

Where:

- u_j denotes the previous heart beat occurred before time t
- μ_t denotes the instantaneous heart beat distance value.

Since the effects of the sympathovagal balance occur on a millisecond timescale, but its effects last for several seconds, the interval must be modeled as dependent on the recent history of the Sino-Atrial node inputs

$$\mu_t \equiv \mu_{RR}(t) = a_0 + \sum_{i=1}^p a_i RR_{t-1} \quad (3.6)$$

We define the time interval between heart beats as the elapsed time RR between R peaks of the ECG.

Since the objective is to track the physiological changes along time, the evaluation of the parameters in 3.5 and 3.6 are adaptive. The measure of $\mu_{RR}(t)$ is time varying and is determined by the time varying AR coefficients $a_i(t)_{i=0}^p$.

The instantaneous variance of the inverse Gaussian model can be derived as

$$\sigma_{RR}^2(t) = \frac{\mu_{RR}^3(t)}{\theta(t)} \quad (3.7)$$

¹⁶ Barbieri et al., 2005

Heart Rate (HR) is defined as the reciprocal of the heart period RR . HR and Heart Rate Variability (HRV) indexes can be computed as

$$\mu_{HR} = \tilde{\mu}^{-1} + \tilde{\theta}^{-1} \quad (3.8)$$

$$\sigma_{HR} = \left[\frac{2\tilde{\mu} + \tilde{\theta}}{\tilde{\mu}\tilde{\theta}^2} \right] \quad (3.9)$$

Where:

$$- \tilde{\mu} = \frac{\mu_{RR}}{60} \frac{[s]}{[s/min]}$$

$$- \tilde{\theta} = \frac{\theta}{60} \frac{[]}{[s/min]}$$

Since to study the baroreflex is used the information brought by the systole S in ABP, the model has to take in consideration the influence of this covariate. To do this q coefficients are added to 3.6:

$$\mu_t \equiv \mu_{RR}(t) = a_0 + \sum_{i=1}^p a_i RR_{t-1} + \sum_{j=1}^q b_j S_{t-j} \quad (3.10)$$

Where S_{t-j} denote the previous j th measure of systole before time t . Now the mean RR is evaluated by a bivariate AR model.

It has to be highlighted that this model does not take in account the different occurring time of the two inputs: as it is known, the systole in the ABP happen after the R wave in the ECG. Anyway this does not affect the measures.

With a linear system assumption, baroreflex can be estimated as the absolute value of the transfer function of the built bivariate model. Given the parametric model in 3.10 it is possible to evaluate the frequency response of the baroreflex as

$$H_{12}(w) = \frac{\sum_{j=1}^q b_j(k) z^{-j}}{1 - \sum_{i=1}^p a_i(k) z^{-i}} \Bigg|_{z=e^{j2\pi f_s}} \quad (3.11)$$

Where f_s is the beat rate of the RR.

It is possible also to estimate the power spectrum or the gain in the frequency domain

$$P_{RR}(\omega, t) = \sigma_{RR}(t) |H_{11}(\omega, t)| \quad (3.12)$$

$$\text{Baroreflex}_{gain}(\omega, t) = |H_{12}(\omega, t)| \quad (3.13)$$

3.6.2 Characterizing baroreflex sensitivity

Baroreflex sensitivity is a measure of connectivity between heart rate and blood pressure, particularly the influence that systolic blood pressure has on heart rate. Although a very complex mechanism involving

many physiological variables, like respiration, with only the contribution of pressure is possible to make some general assessments.

Baroreflex has a very complex dynamic with no trivial behaviours not even in simple situations. For instance, when blood pressure drops, one could expect a rise in the magnitude of baroreflex sensitivity to increase the heart rate to then reach a plateau: heart rate increase to the maximum to compensate for the drop in pressure and maintain organ blood perfusion. The reality is that it is not this simple, with the signal showing bursting activity not easy to decipher. The analysis becomes even more complex in ICU, where one has to consider not only all the complications related to the patient pathology but also has to adjust for medications.

This study was performed in these very conditions.

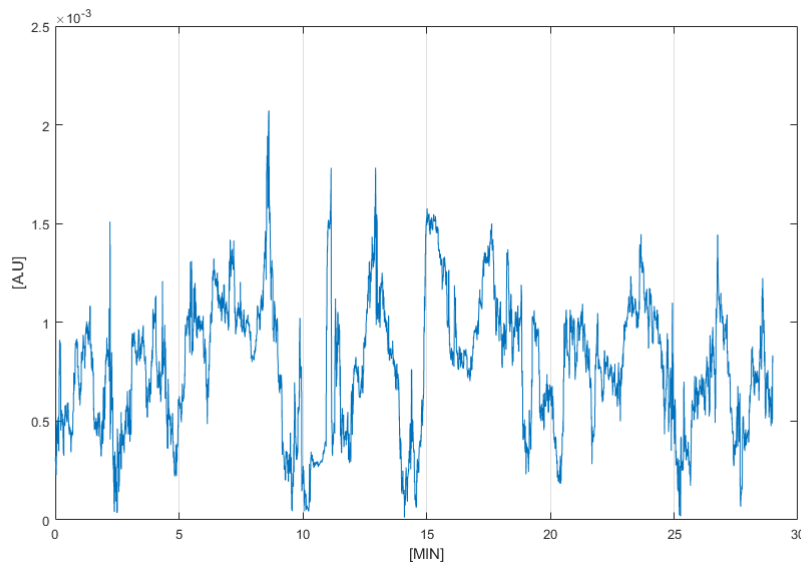


Figure 3.11: Example of baroreflex from control subject s03279-2984-08-26-14-42.

3.6.2.1 Cohort

For this second study, patients were pulled again from the MIMIC matched subset waveform database. Signals were processed and annotated in the same manner as 3.2 and 3.5, it is necessary for the application of the point process, it adds another layer of control to grant the best quality of the signals and minimizes the need to use interpolation techniques.

Only records with at least five hours of continuous ABP and ECG were chosen, if a patient was hypotensive, then its onset is always 30 minutes before the end of the observation. For control patient it was chosen the first available 5 hour window.

	C	H	p-val
% (num)	61.45 % (51)	38.55 % (32)	
Age	54.5 ± 13.80 yrs	63.28 ± 16.3 yrs	.0093
Vasopressors	17.65 %	25.00 %	.4258
Sedatives	33.33 %	21.88 %	.2679
Ventilation	52.94 %	46.88 %	.5966
Pacemakers	19.61 %	18.75 %	.9291
Gender(M)	52.94 %	68.75 %	.1581
Mortality	9.80 %	28.13 %	.0316

Table 3.4: Cohort Description.

Annotations results were visually inspected and those records with artifacts were discarded. Differently from the dataset described in Table 3.2, exclusion criteria were made much more strict this time as we kept only those records that showed extremely clean waveforms. In the end we obtained a cohort of 83 subject, see Table 3.4 for cohort description.

The cohort description is in line with that described from the 2009 Physionet Challenge, particularly mortality in hypotensive patients being more than twice that of control.

3.6.2.2 Background & Working Hypothesis

The task of identifying predictive features within the baroreflex has been challenging. All statistical descriptors failed to differentiate between the two classes even at the onset of hypotension. This is perhaps due to the high non linear non stationary dynamic of the signal. To our surprise, however, not even non linear and chaotic measures were able to grasp a significant difference, the reason for this behaviour is still to be understood.

In light of these findings it was then decided to simplify the problem by identifying specific well distinguishable markers: baseline and relative burst amplitude (RBA).

The hypothesis behind this choice was that the baroreflex is composed by two separated mechanism:

$$\text{baroreflex} = \text{baseline} + \text{bursts} \quad (3.14)$$

Where the baseline might represents the dynamic equilibrium of the ANS while the bursting activity the response to a specific condition change. In such hypothesis then, the absolute value of a peak is not comparable to those nearby due to baseline changes, making it difficult to make assumption on baroreflex behaviour and therefore to hypotensive hemodynamic changes. What, instead, was valued as reliable was the amplitude of the burst relative to the baseline, therefore how

strong the burst really is in the current dynamic state of the baroreflex. To realize what just stated it was needed to define the baseline:

- The baseline must be always lower and at most equal to the raw baroreflex.
- The baseline must be always greater and at most equal to zero.
- The baseline is computed by means of polynomial fitting by using the lowest possible polynomial order that respects the two previous conditions.
- Fitted points are local minima of the baroreflex.

In this study, barosensitivity was estimated using a point process framework. Thanks to this model it was possible to achieve a high signal resolution (200Hz) at the price of some high frequency noise. To solve for this, raw signals were treated with a simple moving average filter before being fed to the baseline computation algorithm.

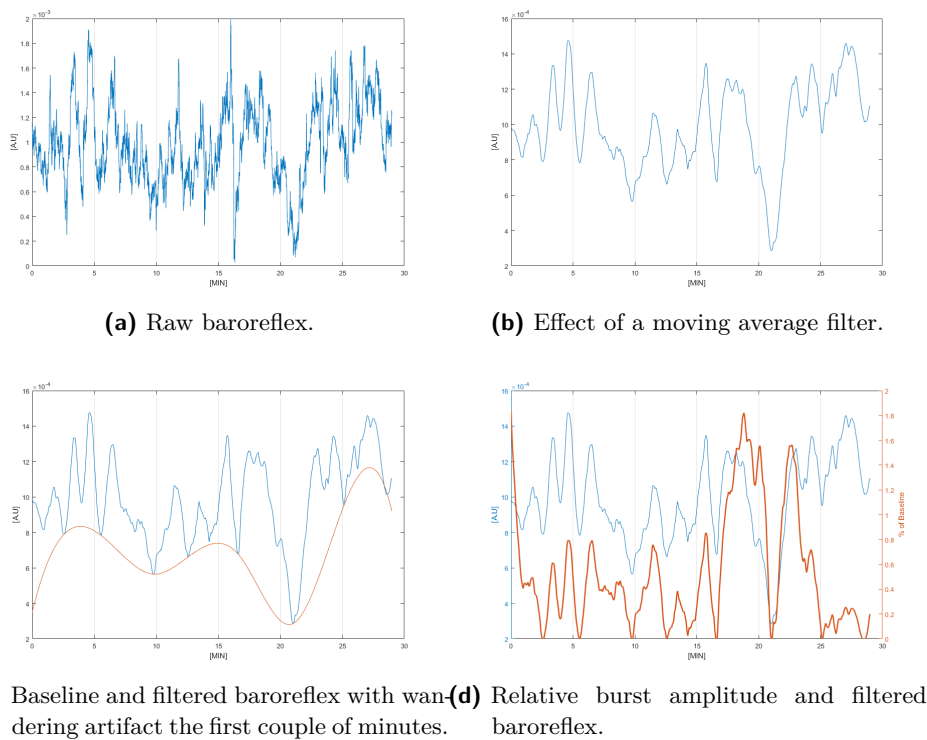


Figure 3.12: Baroreflex processing steps.

The algorithm implementation was straightforward since the baseline is defined with strong conditions. Baseline where found using a local minima search algorithm.

A drawback in the application of this routine is the baseline values at the extremes of the data window: Often the fitting algorithm generates wandering lines, see Figure 3.12c for an example. In the end, though,

this was just minor issues since for the analysis it was only used the central section of the baseline.

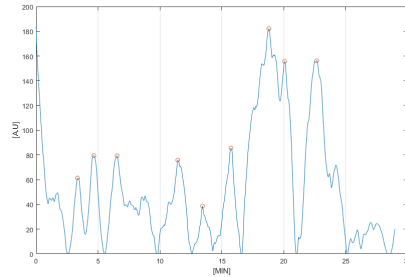
Relative Burst Amplitude -RBA- evaluation was trivial at this point. Given that there was interests only towards the peaks amplitude compared to baseline, it was just a matter of baroreflex normalization:

$$RBA = \frac{(baroreflex - baseline)}{baseline} \quad (3.15)$$

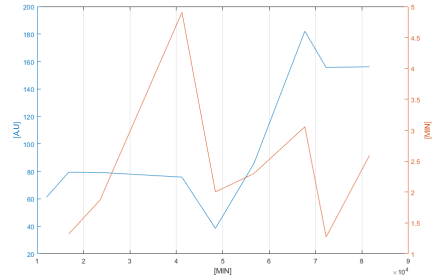
See Figure 3.12 for examples.

3.6.2.3 Feature Crafting

Features were crafted from the baseline and relative bursts only. Again some of them were simple statistical descriptors while others required some more crafting. It is the peaks amplitude compared to baseline and frequency we are interested in, but not of all detectable peaks: some of them could be noise or just minor variation and they can confound the analysis. Therefore in the peak detection process we restrained the research to only those that have a prominence equal or greater to 30% the baseline. With prominence it is meant the amplitude of a peak compared the closest local minima. See Figure 3.13a and Figure 3.13b.



(a) Detected peaks.



(b) Peaks amplitude (Blu) and frequency (Orange).

Within this framework, it was computed for peaks amplitudes and peaks frequency the first four statistical moments. From the baseline it was investigated the zero crossing and mean crossing in addition to the first four moments. With a simple ranksum analysis we observed statistical significance for the skewness and standard deviation of the peaks and the kurtosis of the normalized bursts activity, to our surprise¹⁷, in the HF band. This is, of course, a naive result, being not adjusted to confounding factors, but still the best result observed so far. See Chapter 4 for results.

We were not able to find any relevant feature in the LF band.

¹⁷ Despite the influence of respiration in this band.

3.7 FEATURE ENGINEERING

Differently from other studies(see Section 2.2) where features were crafted from either a single time series, like MAP, and different domains or from many series and few domains, like statistical descriptors and wavelet coefficients, this work extends the reach of the analysis to a multivariate and multi-domain model: The patient physiology was described with ninety-six features covering many different domains and time series.

The feature space contains non-linear and chaotic indices, spectral features, statistical descriptors, connectivity measures and different scores and flags, see Figure 3.13.

The choice to build a large set of regressors insures to reach a greater "field of view" increasing the likelihood to grasp the dynamics of the condition. The drawback of using a large number of features is the limited size of the cohort, therefore of the training set, and the presence of co-linearities and co-dependencies between predictors: feature selection/transform is unavoidable.

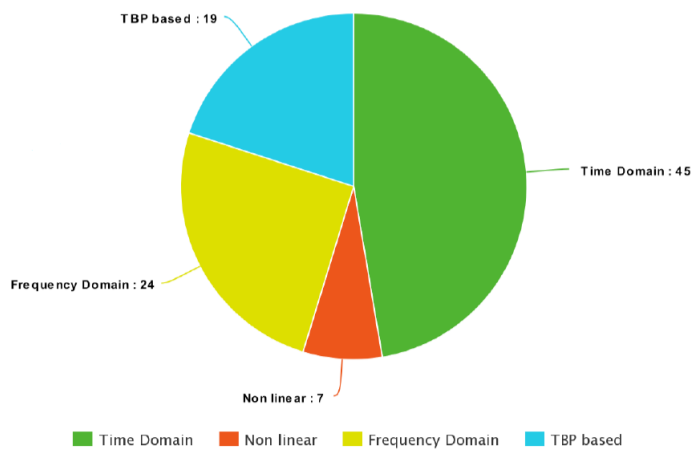


Figure 3.13: Feature space. For TBP features see Section 3.7.4.

3.7.1 Time Features

Time domain features are all the statistical descriptors, polynomial fittings and clinical measures. Amongst time features there are some of the most predictive for the task.

Statistical descriptors were computed for all time series up to the second order (mean and standard deviation) and within different time intervals. Particularly, mean measures closer to the hypotension computed in small intervals (in the couple of minutes order) turned out to be specially predictive in the classification, as was also noted by Chen et al., 2009. For a brief overview see Table 3.5.

Serie	Interval			
	DW		lstm	
	Mean	STD	Mean	STD
RRI	V	V	X	X
SBP	V	V	V	X
DBP	V	V	V	X
PP	V	V	V	X
MAP	V	V	V	X
CO	V	V	X	X
PTT	V	V	X	X

Table 3.5: Statistical description of different signals. DW means the interval for calculation was equal to that of the data window while lstm to the last available minutes before the end of the data window. V means that specific feature was computed X otherwise.

First order polynomial coefficients were computed to fit a line in the MAP. This operation yielded to three features: the slope, the intercept and a future estimate. The slope and the intercept are simply the coefficients from the fitting algorithm while the future estimate is a value representing the MAP 30 minutes in the future if the interpolating line were to remain the same. See Figure 3.14. This kind of estimation was inspired from the work of Chen et al., 2009 to solve the Physionet Challenge, although it worked in their attempt it didn't seem to be as powerful in our dataset. The intercept is generally the most predictive feature of the three, as all related to the MAP.

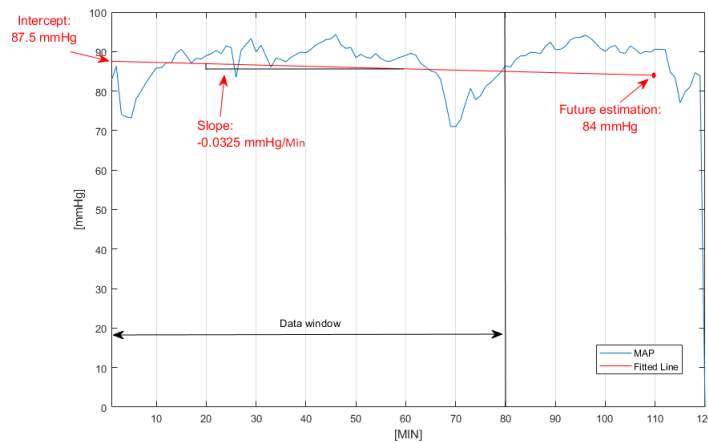


Figure 3.14: Line fitted using information from the data window to estimate pressure 30 minutes into the future. In this example it was used a data window of 80 minutes. The polynomial MAP features are highlighted in red.

Clinical indexes, only related to the HRV, are NN20, NN50 and RMSSD. The Root Mean Square of Successive Differences, or RMSSD, is a measure of short term variability of the RR intervals given by Eq.3.16.

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N-1} (RR_{j+1} - RR_j)^2} \quad (3.16)$$

NN20 and NN50 are similar in nature, they both measure the relative distance between successive RR intervals, specifically they count the number of successive interval which have a difference greater to a given threshold. The threshold is 20ms and 50ms respectively for NN20 and NN50. However it might be of greater interest to investigate NN20/NN50 to the total of beats, hence the introduction of pNN20 and pNN50 which are only a normalization of the previous features, see Eq.3.17 for a generalization.

$$\begin{aligned} sc_j &= \begin{cases} 1 & \text{if } (RR_{j+1} - RR_j) > x \\ 0 & \text{otherwise} \end{cases} \\ NNx &= \sum_{j=1}^{N-1} sc_j \\ pNNx &= \frac{NNx}{N-1} \times 100 \end{aligned} \quad (3.17)$$

Those described are the most relevant time features, other were applied to study the behaviour of spectral bands in time but with moot results and therefore were discarded for reasons explained in Subsection 3.7.2.

3.7.2 Spectral Features

From ECGs and ABPs we extracted heart rate variability, systograms and diastograms. It is well known the information that can be extracted from those signals (Camm et al., 1996): ANS activation. Therefore spectral features are those describing the autonomic activation.

3.7.2.1 Spectral bands

Three main spectral components are considered for ANS evaluation: VLF, LF and HF.

- VLF: This is the less defined component as its physiological explanation. Typically it is considered as that spectral band with frequencies lower than .04Hz. An additional spectral component can

be identified, the Ultra Low Frequencies -ULF-, with frequencies lower than .0033Hz. A broad system of variables may attributed to the power in the VLF band but it is typically associated to thermoregulatory processes, hemodynamic delays, mechanical effects of the breathing pattern, spinal reflexes and vascular autorythmicity. Some say that specially those rhythms in the ULF range might be correlated to hemodynamic abnormalities(Camm et al., 1996).

- LF: Low frequencies tend to reflect the baroreflex resonance frequency and belongs to the .04 to .15 Hz range. RSA may contaminate the rhythm. This factor become problematic for LF estimate validity since can render more complex the interpretation. LF activity might be attributed to symphatetic activation more so after normalization although this statement is discussed. Overall LF might contain three different rhythms: RSA, baroreflex oscillations and 30-s rhythms(Camm et al., 1996).
- HF: associated in the .15 to .4 Hz range. HF might be heavily contaminated from the RSA and it might be associated to parasymphatetic activity(Camm et al., 1996).
- LF/HF: It is usually associated to the autonomic balance. In normal subjects it takes values around two units, where an increase represent symphatetic activation and a decrease parasymphatetic activation. The feature is therefore a relative measure of the parasymphatetic balance.
- Open Loop Baroreflex: This is the simplest baroreflex estimation. We have also tested and measures a point process related baroreflex estimate, for this refer to the point process section. When computed in open loop, baroreflex gain is given by:

$$\alpha = \sqrt{\frac{LF_{RR}}{LF_{SBP}}} \quad (3.18)$$

Where LF_{RR} and LF_{SBP} are respectively the total power in the LF band of the HRV and sytogram.

3.7.2.2 PSD Estimation

MIMIC records are very noisy making direct FFT analysis rather complex and hardly interpretable. We then relied on parametric methods such as autoregressive models given the added robustness to noise. The Yule-Walker equations were used for AR coefficients estimates from which the PSD was extracted, see Eq.3.19.

$$PSD(f) = \frac{\sigma_Z^2}{|1 - \sum_{k=1}^p \phi_k e^{-i2\pi f k}|^2} \quad (3.19)$$

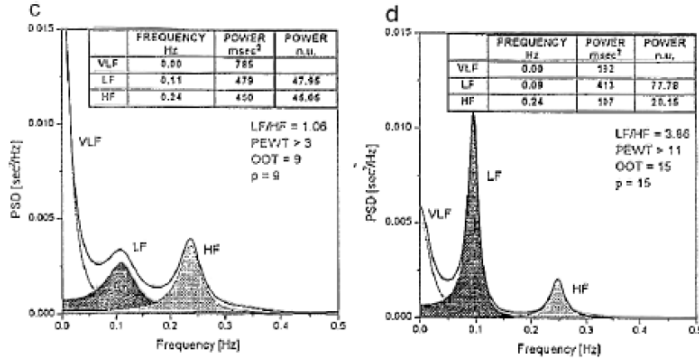


Figure 3.15: Spectral decomposition of the three components in two scenarios: Rest and Tilt.

$$P_{LS}(f) = \frac{1}{2\sigma^2} \left\{ \frac{\left[\sum_{k=1}^N (x_k - \bar{x}) \cos(2\pi f(t_k - \tau)) \right]^2}{\sum_{k=1}^N \cos^2(2\pi f(t_k - \tau))} + \frac{\left[\sum_{k=1}^N (x_k - \bar{x}) \sin(2\pi f(t_k - \tau)) \right]^2}{\sum_{k=1}^N \sin^2(2\pi f(t_k - \tau))} \right\}$$

Additionally we investigated the Lomb-Scargle Periodogram to account for the RR uneven sampling.

For non-parametric methods, differing from simple FFT, we investigated the performance of the Welch's Spectral estimate. The advantage of the technique is noise reduction in the estimate. The signal is split into many overlapping windows that are then windowed. Afterwards FFT is applied to each window and averaged.

Following the guidelines from Camm et al., 1996 we relied on intervals of different length for the different spectral bands. VLF were computed on forty minutes windows, LF on a two minutes window and HF on a one minute window.

3.7.3 Non Linear Features

Non-linear mechanisms are involved in the heart variability system. Some measures of chaos can be correlated to CHF and mortality, for instance to generally have a higher level of chaos in the HRV might be considered healthy. Several non-linear features were investigated, contributing to the bulk of processing power required for the computation of out features.

3.7.3.1 SD1 and SD2

Heart rate variability can be projected on the plane to show the relation between successive RR intervals. The Pointcare plot is the tool for

the task.

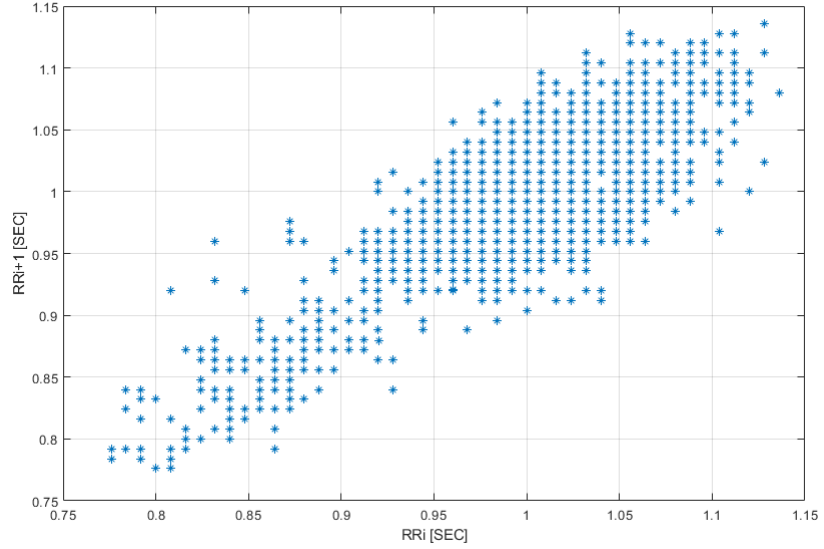


Figure 3.16: Pointcare plot.

The Pointcare plots RR_{i+1} as function of RR_i , the points distribution on the plane hides two interesting features called SD1 and SD2 which describe the shape the RR intervals take. In healthy subjects there should be a good dispersion on the plane as a result of good variability and therefore heart rate adaptability to everyday tasks. Pathological subjects show a broader spectrum of shapes which hopefully can be grasped by the crafted features.

SD1 and SD2 are a measure of dispersion of the plane: the standard deviations of the first and second principal components respectively. Therefore, once drawn, the Poincare plot undergoes a PCA transformation (see Section 3.8.1), after which is trivial to compute the two features.

3.7.3.2 Approximate Entropy

Approximate entropy or ApEn is an irregularity measure, the higher the more irregular the signal (Tarvainen et al., 2014).

Given the RRI signal, then lets define:

$$u_j = (RR_j, RR_{j+1}, \dots, RR_{j+m-1}) \quad \text{with } j = 1, 2, \dots, N - m + 1 \quad (3.20)$$

With N being the number of samples in the RR series and m the embedding dimension. m was chosen to be equal to 2 following the guidelines of (Tarvainen et al., 2014).

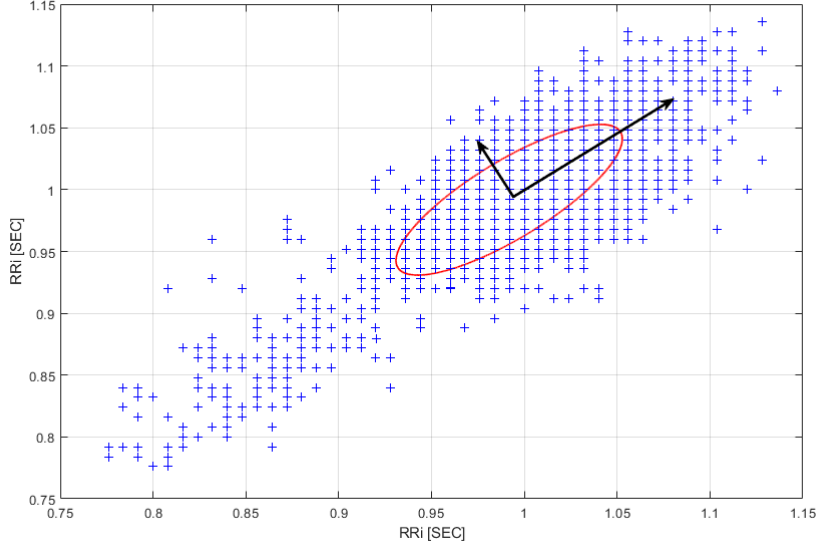


Figure 3.17: Pointcare plot.

Following for each pair of the $N-m+1$ u_j vectors we compute the pairwise difference and we search for the largest one. see Eq.3.24.

$$d(u_j, u_k) = \max\{|RR_{j+n} - RR_{k+n}|n = 0, 1, \dots, m-1\} \quad (3.21)$$

$$C_j^m = \frac{\text{Number of elements}\{u_k | d(u_j, u_k) \leq r\}}{N - m + 1} \quad \forall k \quad (3.22)$$

$$\varphi^m(r) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} \ln C_j^m(r) \quad (3.23)$$

$$ApEn(m, r, N) = \varphi^m(r) - \varphi^{m+1}(r) \quad (3.24)$$

Ideally N should be large so that the variable can reach the asymptotic value, moreover r should be equal to $0.2 \times SDNN$ where $SDNN$ is the standard deviation of the heart rate variability.

3.7.3.3 Sample Entropy

Another measure of irregularity similar to ApEn but with a slightly different formulation (Tarvainen et al., 2014).

$$C_j^m = \frac{\text{Number of elements}\{u_k | d(u_j, u_k) \leq r\}}{N - m} \quad \forall k \neq j \quad (3.25)$$

$$C^m = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} C_j^m(r) \quad (3.26)$$

$$SampEn(m, r, N) = \ln\left(\frac{C^m(r)}{C^{m+1}(r)}\right) \quad (3.27)$$

m and r are equal to those of the ApEn.

3.7.3.4 Correlation Dimension

The correlation dimension, D_2 , measures the complexity of a waveform (Tarvainen et al., 2014). As for ApEn and SampEn, the correlation dimension requires a similar steps for its computation:

$$d(u_j, d_k) = \sqrt[2]{\sum_{l=1}^m (u_j(l) - u_k(l))^2} \quad (3.28)$$

$$D_2(m) = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C^m(r)}{\log(r)} \quad (3.29)$$

Given that in reality is difficult to solve the above equations we had to rely on some approximations: the value of the correlation dimension is calculated as the slope of the linear part of the $\log C^m(r), \log(r)$ plot. See Figure 3.18.

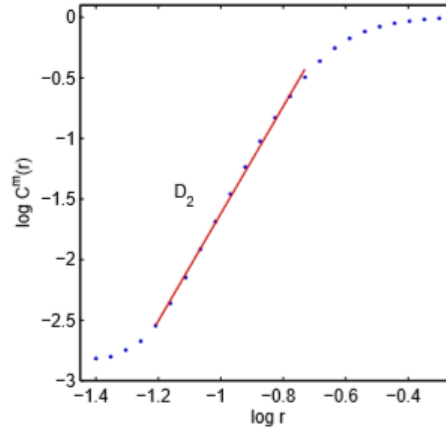


Figure 3.18: Correlation Dimension

3.7.3.5 Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis, DFA, measures the correlation of the signal with itself on different time scales (Tarvainen et al., 2014). The necessary steps to measure the DFA are the following:

$$y(k) = \sum_{j=1}^k (RR_j - \overline{RR}) \quad \text{for } k = 1, 2, \dots, N \quad (3.30)$$

$y(k)$ is split up into n non overlapping windows. Within each window it is fitted a regressing line. Given the n_{th} window then the respective regressing line is indicated by $y_n(k)$.

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - y_n(k))^2} \tag{3.31}$$

$$\tag{3.32}$$

F is evaluated for several values of n. Afterwards, two slopes in the log-log plot of F(n) and n are fitted generating the two features of interests: α_1 and α_2 . The former keeps track of short-term fluctuations while the latter long-term fluctuations, see Figure 3.19.

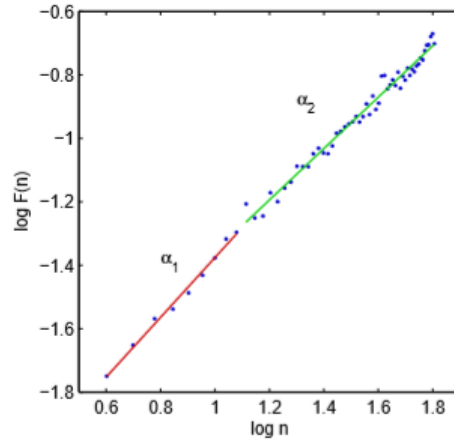


Figure 3.19: Visualization of α_1 and α_2 give by the detrended fluctuation analysis.

3.7.4 Score Features

As noted by Chen¹⁸, most of the predictive information resides in the ABP itself, so we defined a transformation which allowed us to have a simplified representation of the time series to better understand its long term dynamics.

Given the ABP time series as:

$$ABP_k = abp(k\Delta t_1) \quad \text{with } k = 1, 2, 3, \dots, n \tag{3.33}$$

Where $\Delta t_1 = 1/fs$, namely the sampling frequency. Let's then define a new time series TBP:

$$TBP_{j,\Delta t_2} = \frac{m}{n} \sum_{k=j-\frac{n}{2m}}^{j+\frac{n}{2m}} ABP_k \quad \text{with } j = 1, 2, 3, \dots, m \tag{3.34}$$

18 Chen et al., 2009

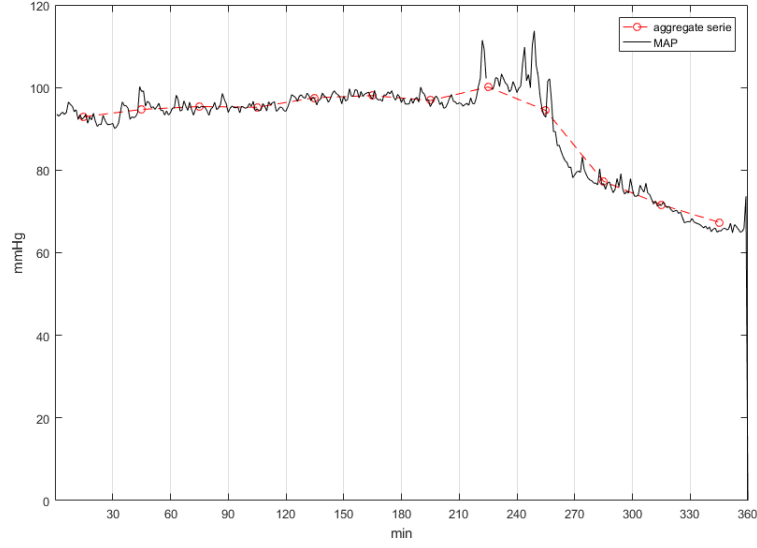


Figure 3.20: TBP

With the condition $m\Delta t_2 = n\Delta t_1$. Basically the signal is subdivided in m non-overlapping windows each of width Δt_2 and the average of every window is computed to form a TBP point. We tried different sizes of Δt_2 for different data windows. From this new series We compute a set of indices, like:

- `ag_area`: Weighted sum of the entire series using exponential weighting.

$$ag_area = \sum_{i=1}^m TBP_{i,\Delta t_2} w_i \quad (3.35)$$

- `d_ag`: Average TBP derivative.

$$d_ag = \frac{1}{m-1} \sum_{i=2}^m (TBP_{i,\Delta t_2} - TBP_{i-1,\Delta t_2}) \quad (3.36)$$

- `l_ag`: Exponentially weighted average TBP derivative.

$$l_ag = \frac{1}{\sum_{j=1}^{m-1} w_j} \sum_{i=2}^m (TBP_{i,\Delta t_2} - TBP_{i-1,\Delta t_2}) w_{i-1} \quad (3.37)$$

- `mean_ag`: Average TBP.
- `ag_70`: Number of elements in TBP lower than 70 mmHg.
- `b_80`: Number of elements in TBP lower than 80 mmHg.
- `agsudden`: Flag set to 1 when a `d_ag` is lower than one and the last available value in TBP is lower than 80.

- ag_80 : Number of elements in TBP with value between 70 and 80 mmHg.
- $up90$: Score defined as the product of the number of elements in TBP with value between 90 and 100 mmHg times the number of elements with value greater than 100 mmHg.

$$s90_{i,\Delta t_2} = \begin{cases} 1 & \text{if } 90 \text{ mmHg} < TBP_{i,\Delta t_2} < 100 \text{ mmHg} \\ 0 & \text{otherwise} \end{cases} \quad (3.38)$$

$$s100_{i,\Delta t_2} = \begin{cases} 1 & \text{if } TBP_{i,\Delta t_2} > 100 \text{ mmHg} \\ 0 & \text{otherwise} \end{cases} \quad (3.39)$$

$$up90_{\Delta t_2} = \left(\sum_{i=1}^m s90_i \right) \left(\sum_{j=1}^m s100_j \right) \quad (3.40)$$

3.8 FEATURE SELECTION

After feature engineering, the number of predictors in the model was ninety-six. To feed a classifier the integrity of this dataset can only yield to poor results:

- The number of feature is almost as the number of the observations (≈ 100 features for ≈ 450 observations), creating convergence problems.
- Existence of Co-linear and co-dependent features. In this case some classifier fail even to train, like quadratic discriminants. This phenomena is often present in our features: systolic, MAP and diastolic pressure, for example, are strictly related and so all the their features.
- Confounding features: features that don't carry any relevant information ends up only to confuse the model.
- Overfitting.
- Curse of dimensionality.

A dimensionality reduction method is therefore needed. In this study we investigated three methods to, in the end, chose the best one for our model.

3.8.1 PCA

The Principal Component Analysis is a transformation to convert observations in a set of linearly uncorrelated variables¹⁹. The resulting components are ordered by variance magnitude: the first component explains the biggest variance, the second the second biggest and so on. Apart from transforming the space in a linearly uncorrelated one, PCA can also be used as a mean of feature selection. The last components explain only a small percentage of the total variance, which could mean that they are not very informative, hence they can be discarded. Therefore PCA has a two fold advantage: feature selection and feature transformation.

Before applying the transformation the dataset needs to be normalized:

$$X = \frac{\tilde{X} - \bar{X}}{std(X)} \quad (3.41)$$

Where \bar{X} is the column wise mean of the dataset \tilde{X} . X is $m \times n$ matrix, with m observations and n features. Mathematically the PCA can be expressed as follows:

$$T = XW \quad (3.42)$$

Where W is a $n \times n$ matrix called the loading matrix and T are the principal components of the same dimension of X . The first component is given by:

$$w_1 = arg \max \left\{ \frac{w^T X^T X w}{w^T w} \right\} \quad (3.43)$$

Principal components are orthogonal to each other, therefore the remaining can be computed:

$$X_k = X - \sum_{s=1}^{k-1} X w_s w_s^t \quad (3.44)$$

$$w_k = arg \max \left\{ \frac{w^T X_k^T X_k w}{w^T w} \right\} \quad (3.45)$$

PCA is better suited for LDA, QDA and NN analysis, but not always. Moreover usually it leads to poor results when used with Trees and KNN, maybe because of the loss of some particular structure in the data. Overall the PCA was the worst performer amongst the feature selection techniques we have tested.

3.8.2 Forward Selection

The Forward Selection (FS) is greedy²⁰ subset selection algorithm. The FS is an iteration of different steps till convergence:

¹⁹ Jolliffe, 1986

²⁰ The algorithm at each stage settles for the locally optimal choice rather than the global.

1. Initialization with the original feature subset and an empty subset.
2. Each feature from the original subset is added in the subset and its performance is assessed by mean of 10fold CV on a given criterion.
3. The feature that has the biggest performance, if any, is added to the subset and the removed from the original, the algorithm goes to state 2. If no feature is beneficial the algorithm stops.

Forward selection worked well on Trees and KNN and on average yielded to better results across the board compared to PCA. The biggest drawback of Forward Selection is its greedy nature: the predictive power of features combinations is underestimated.

3.8.3 LASSO

LASSO (Least Absolute Shrinkage and Selection Operator) is an algorithm used both for regularization and feature selection purposes, (Tibshirani, 1996). The output of the LASSO algorithm are the coefficients β of the regression. To obtain this it could be sufficient to use the Ordinary Least Squares (OLS) estimate approach. Here the coefficients are obtained by minimizing the residual square error. LASSO addresses a drawback of OLS method related to model interpretation: with a large set of predictors it would be useful to identify a subset of predictors that contain most of the information needed to do a good prediction/classification.

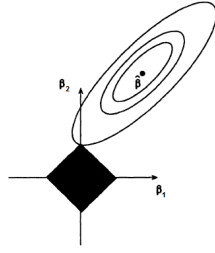
LASSO is able to set the coefficients of the weak predictors to zero reducing, in this way, the dimensionality of the problem.

LASSO constrain the sum of the regressors coefficients β to be less than a tuning parameter t fixed by the user (l_1 norm). This force some coefficient to be set to zero. The idea is similar to Ridge regression in which what is constrained to be less than a fixed value is the sum of squares of the coefficients (l_2 norm). In this case, though, only the amplitude of the coefficients is affected. Given a set of features $X = x_1, \dots, x_n$, that are supposed normalized, the problem addressed by LASSO is:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 \text{ subject to } \|\beta\|_1 \leq t \quad (3.46)$$

Where:

- t is a tuning parameter that describe the amount of shrinkage applied.
- $\|\beta\|_p = \left(\sum_{i=1}^N |\beta_i|^p \right)^{\frac{1}{p}}$ is the standard l_p norm.



$$\min_{\beta_0, \beta} \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 + t \sum_j^2 |\beta_j| \quad (3.47)$$

Figure 3.21: Constrain region and square of error term for l_1 norm

Using l_2 instead of l_1 norm lead to different shape of the constrain region. Using a bi-dimensional case, with l_1 norm, using Lagrange multiplier, the results is Eq. 3.47.

The diamond shape in the middle of figure 3.21 show the constrain region for β_1, β_2 . The elliptical contour plot represents sum of square error term. The optimal point is a point which is common point between ellipse and circle as well as gives a minimum value for the above function. There is a high probability that optimum point falls in the corner point of diamond region. An optimal point falls in the corner point, it means that one of the feature's estimate ($\beta_j = 0$) is zero. In Ridge regression is used the l_2 norm so the resulting constrain region will be elliptical (3.22). This means that there are no corner for the contours to hit, so solutions equal to zero will rarely happen.

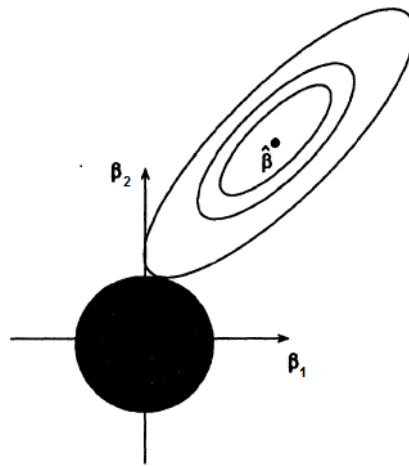


Figure 3.22: Constrain region and square of error term for l_2 norm

3.9 CLASSIFICATION

For the classification part Matlab functions have been used. A large number of classifiers have been trained to assess the predictive power

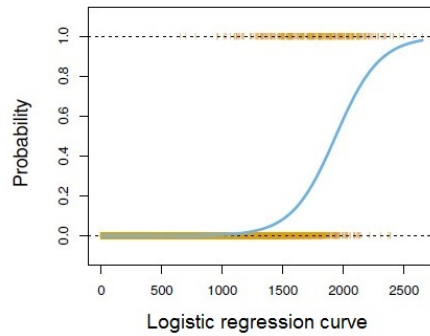


Figure 3.23: Example of logistic curve

of various methods that are re-proposed with different hyper parameter settings.

3.9.1 Logistic Regression

Logistic regression was one of the first approaches we tried. In statistics, logistic regression, or logit regression, or logit model is a regression model where the dependent variable Y is categorical. Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables X by estimating probabilities using a logistic function, which is the cumulative logistic distribution. Let's write $p(X) = Pr(Y = 1|X)$. Logistic regression uses the form:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (3.48)$$

It is easy to see that no matter what value β_0, β_1 or X take, $p(X)$ will have values between 0 and 1.

We use maximum likelihood to estimate the parameters.

$$L(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i:y_i=0} (1 - p(x_i)) \quad (3.49)$$

This likelihood gives the probability of the observed zeros and ones in the data (the joint probability). We pick 0 and 1 to maximize the likelihood of the observed data.

3.9.2 Discriminant Analysis

In discriminant analysis, the idea is to model the distribution of the feature space X in each of the classes separately and use the Bayes theorem to flip things around to get the probability of Y given X :

$Pr(Y|X)$. The Bayes theorem is a basis for discriminant analysis. The Bayes theorem for classification:

$$\begin{aligned} Pr(Y = k|X = x) &= \frac{Pr(X = x|Y = k) * Pr(Y = k)}{Pr(X = x)} \\ &= \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)} \end{aligned} \quad (3.50)$$

Where π_k is the prior probability for class k, and the marginal is the summing over all the classes $\sum_{l=1}^K \pi_l f_l(x)$. This formula is quite general where we can plug in any probability densities: $f_k(x) = P(Y = k|X = x)$ is the density for X in class k.

The Gaussian density has the form

$$f_k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu_k}{\sigma_k} \right)^2} \quad (3.51)$$

If we assume $\sigma_k = \sigma$, taking logs and discarding terms that do not depend on k, we see that this is equivalent to assigning x to the class with the largest discriminant score:

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) \quad (3.52)$$

Once we have estimates $\hat{\mu}_k(x)$ we can turn these into estimates for class probabilities:

$$\hat{Pr}(Y = k|X = x) = \frac{e^{\hat{\mu}_k(x)}}{\sum_{l=1}^K e^{\hat{\mu}_l(x)}} \quad (3.53)$$

This approach is quite general, and other distributions/density can be used including non-parametric approaches. By altering the forms for $f_k(x)$, we get different classifiers (ie classification rule). When $f_k(x)$ are Gaussian densities, with the same covariance matrix in each class, this leads to linear discriminant analysis.

When the variances of all X are different in each class, the cancellation doesn't occur because when the variances are different in each class, the quadratic terms don't cancel. In this case the discriminant function becomes 3.54:

$$\delta_k(x) = -\frac{1}{2} (x - \mu_k)^T \sum_k^{-1} (x - \mu_k) + \log(\pi_k) \quad (3.54)$$

And therefore, the discriminant functions are going to be quadratic functions of X. Quadratic discriminant analysis uses a different covariance matrix for each class. Quadratic discriminant analysis lead to good results if the number of variables is small.

In Our study we used both linear and quadratic approach.

Name	Num Neighbors	Distance	Distance Weight
KNN_Coarse	100	Euclidean	Equal
KNN_Medium	10	Euclidean	Equal
KNN_Cosine	10	Cosine	Equal
KNN_Mikowski	10	Mikowski	Equal
KNN_Weight	10	Euclidean	Square Inverse

Table 3.6: List of KNN variants.

3.9.3 *K*-nearest neighbor

In pattern recognition, the *k*-Nearest Neighbors algorithm (or *k*-NN for short) is a non-parametric method used for classification and regression. *K*-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms.

A peculiarity of the *k*-NN algorithm is that it is sensitive to the local structure of the data. Given a matrix of n observation and m variables, to decide what class assign to a new observation, the algorithm base its decision on the predominant class above the k neighbors in the m space created. The selected k neighbors taken in consideration to de-

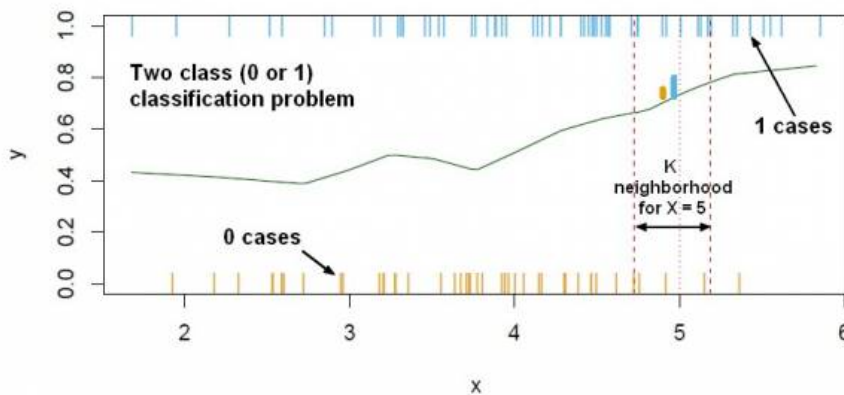


Figure 3.24: The result of classification is depending the number neighbors took in consideration.

cide the belonging class of the new observation, are those that show the smaller distance from the new point under analysis. There are various way to calculate this distance and various way to use it as a weight.

We used 6 type of *k*-NN, to test different numbers of neighbor and distances. In table 3.6 are shown used settings:

- Cosine: one minus the cosine of the included angle between observations (treated as vectors);
- Mikowski: $\left(\sum_{i=1}^N |x_i - y_i|^p\right)^{\frac{1}{p}}$;
- Square inverse: weight is $1/\text{distance}^2$

3.9.4 Support Vector Machines

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

To do this, an hyperplane able to divide the two classes is constructed. The hyperplane equation is

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0 \quad (3.55)$$

If $f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$, then $f(X) > 0$ for points on one side of the hyperplane and $f(X) < 0$ for point on the other.

The vector $w = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]$ is called normal vector. It point in a direction orthogonal to the surface of the hyperplane.

To find the values of w we have to solve the optimization problem known as Soft-margin Linear Support Vector Machine. In this type of problem points are allowed to violate the constrain, but pay a price ("loss") proportional to their distance from the margin boundary.

$$[\hat{w}, \hat{\beta}_0] = \arg \min_{[w, \beta_0]} \frac{1}{2} \|w\|^2 + C \sum_{t=1}^N l_h(r^t(\langle w, x^t \rangle - \beta_0)) \quad (3.56)$$

Where:

- l_h is the hinge loss;
- r^t define the class;
- C is a constant.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. In this way we are able to separate classes non linearly separable.

In our approach we used 3 types of Support Vector Machine with different polynomial kernels:

- Linear;
- Quadratic;
- Cubic.

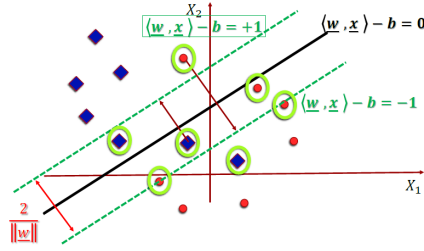


Figure 3.25: Soft margin example

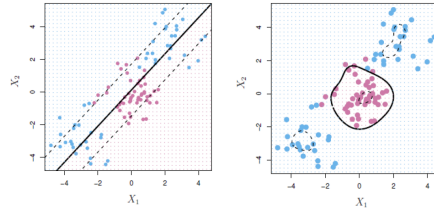


Figure 3.26: Example of a distribution of two classes that can be separated only using a non linear kernel.

3.9.5 Classification Trees

Tree based methods are simple and useful for interpretation. They can be applied to both regression and classification. A classification tree is composed by terminal nodes, leaves terminal part of the tree, and internal nodes, points along the tree where the predictor space is split. To build a tree means to divide the predictor space for X_1, X_2, \dots, X_p into J distinct and non-overlapping regions R_1, R_2, \dots, R_j . The goal is to find boxes that minimize the Residual Sum of Squares given by

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \tag{3.57}$$

Where \hat{y}_{R_j} is the mean response for the training observations within the j^{th} box.

The process however is likely to overfit the data. One possible alternative to this process is to grow the tree only so long as the decrease in the RSS due to the split exceeds some threshold. A better strategy is to grow a very large tree T_0 and then prune it back in order to obtain a subtree. Cost complexity pruning, known as weakest link pruning is used.

$$\sum_{m=1}^{|T|} \sum_{i: x_i \in R_m} (y_i - \hat{y}_{R_m}) + \alpha |T| \tag{3.58}$$

Where:

- α is a tuning parameter;
- $|T|$ indicates the number of terminal nodes of the tree T;

- R_m is the rectangle corresponding to the m^{th} terminal node;
- \hat{y}_{R_m} is the mean of the training observations in R_m .

The tuning parameter controls a trade-off between the subtree's complexity and its fit to the training data. We select an optimal $\hat{\alpha}$ using cross-validation.

In the classification settings, RSS cannot be used as a criterion for making the binary split. A natural alternative is the classification error rate: the fraction of the training observation in that region that do not belong to the most common class:

$$E = 1 - \max_k(\hat{p}_{mk}) \quad (3.59)$$

Where \hat{p}_{mk} represent the portion of training observation in the m^{th} region that are from the k^{th} class. Other indexes to measure the pureness of a leaf are:

- Gini index
- Cross-entropy

Classification trees are usually not competitive with the best supervised learning approaches in term of prediction accuracy, however, boosting methods are able to dramatically improve the performance. In this study ensemble boosting method have been used to improve the outcomes of this classification method. The common paradigm of the ensemble boosting method is to train an ensemble of weak learners to then build a final model that, hopefully, will lead to better classification results. Following are the boosting methods used to improve tree classification in this study.

3.9.5.1 Adaptive boosting

Known as AdaBoost²¹, this algorithm take as input an instance space $X = x_1, \dots, x_n$ and the associate class vector $Y = y_1, \dots, y_n$. Adaboost train a given weak learner algorithm repeatedly in a series of rounds $t = 1, \dots, T$. The weights on the training example on the round t is denoted $D_t(i)$. Initially all weights are equal to $1/n$. The weak learner at iteration t is used to find a weak hypothesis h_t appropriate for the distribution D_t . After this the weights of incorrectly classified examples of $h_t(x)$ are increased so that the weak learner is forced to focus on the hard examples in the training set.

²¹ Freund, Schapire, and Abe, 1999

3.9.5.2 *Totally corrective boosting*

Upgrade of Adaptive Boosting, here is improved the weight adjustment logic. In AdaBoost at every step t , the new weights distribution D_t is chosen to minimize test errors, (Warmuth, Liao, and Rätsch, 2006). In this way the next added weak learner is assumed to bring new information from the preceding one. However, there remains the possibility that this new information could be similar to the one brought from the $(t - 2)^{th}$ iteration. Totally corrective algorithms, optimize the updating values of weights, such that new added learners always bring new information. This can be accomplished by backfitting, linear programming or some other method.

3.9.5.3 *Random forest*

As in the Bagging method, the initial dataset $X = x_1, \dots, x_n$ is bootstrapped to obtain B different "new" dataset. A different classifier is then trained with every single bootstrapped dataset. After training, predictions for unseen samples x' can be made by averaging the predictions $f_b(x')$ from all the individual regression trees on x'

$$\hat{f}(x') = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (3.60)$$

or by taking the majority vote in case of classification trees.

Random forests provide an improvement over bagged trees by way of a small tweak that decorrelates trees. As in bagging, we build a number of decision trees on bootstrapped training samples. When building these decision trees, each time a split in a tree is considered, a random selection of m predictors is chosen as split candidates from the full set of p predictors. This reduces the variance when we average the trees, (Breiman, 2001). This means that while the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees is not, as long as the trees are not correlated.

3.9.5.4 *Random under-sampling boosting*

Classification methods designed for health application: in this fields it is usual to deal with imbalanced dataset, in which the pathologic patients are in minority in respect to the control class. For this reason classification methods are usually biased toward the control group.

In this method the majority class in randomly undersampled to create an artificial balanced dataset upon which machine learning algorithm will be applied (Dai and Hua, 2016). This process is repeated for several iteration generating a model. The final model is an aggregation of models over all iterations.

3.9.6 *Applied Models*

Following the list of models tested in this study:

- Disc_Lin: Linear Discriminant Analysis;
- Discr_Quad: Quadratic Discriminant Analysis;
- Discr_Sub: Discriminant Analysis boosted with Random subspace method;
- KNN_Coarse: K-nearest neighbor with 100 neighbors and Euclidean distance;
- KNN_Medium: K-nearest neighbor with 10 neighbors and Euclidean distance;
- KNN_Cosine: K-nearest neighbor with cosine distance;
- KNN_Mikowski: K-nearest neighbor with cosine Mikowski distance;
- KNN_Weight: K-nearest neighbor with Euclidean distance used an square inverse weight;
- KNN_Sub: K-nearest neighbor boosted with Random Subspace method;
- LogRegr: Logistic Regression;
- SVM_Lin: Support Vector Machine with linear kernel;
- SVM_Quad: Support Vector Machine with quadratic kernel;
- SVM_Cub: Support Vector Machine with cubic kernel;
- Tree_AdaBoost: tree boosted with Adaptive Boosting method;
- Tree_RF: tree boosted with Random Forest method;
- Tree_RUSBoost: tree boosted with Random Under Sampling method;
- Tree_TotalBoost: tree boosted with Totally Corrective Boosting method.

3.10 MEDICATIONS AND INTERVENTIONS

Table 3.2 and Table 3.4 provide a brief clinical description for the two cohorts used in this study.

In Section 4.3.2 is discussed the importance and influence of these factors in the model. This section has the purpose to describe and list what kind of medications the patients were given and the difficulties

encountered to identify them.

Starting with drugs, Vasopressors are powerful vasoconstriction medicines. There are two classes of vasopressors: Alpha adrenergic and Beta adrenergic. The former effects receptors on vascular walls inducing contraction, the latter increases heart contractility²². These drugs are usually given as last stance against critical pressure conditions while injection of fluids is preferred for milder situations, also because of side effects. Vasopressors considered in this study are:

- norepinephrine
- epinephrine
- phenylephrine
- vasopressin
- dopamine
- dobutamine

Sedatives depresses patient's awareness in ICU, relieving the sensation of painful procedures. Moreover sedatives ease the applications of certain procedures, such as mechanical ventilation²³. Hypotension is a possible side effect of these drugs. Sedatives considered for this study are:

- midazolam
- lorazepam
- ativan
- fentanyl
- dilaudid
- propofol
- dexmedetomidine
- precedex

Mechanical-ventilation is that procedure which, partially or totally, substitute the physiological respiration process in the patient. Due to increased intra-thoracic pressure, venous return is diminished which as a consequence might reduce cardiac output. As a result hypotension is again a possible side effect for these kind of interventions²⁴.

Finally subjects with pacemakers were included in the study to extend the casuistry even though it's the class of subject where is hardest to

²² *Use of vasopressors and inotropes*

²³ *Sedation in ICU*

²⁴ *Mechanical Ventilation: Physiologic effects of mechanical ventilation.*

predict hypotension.

Information from the interventions and medications were extracted from the MIMIC Clinical Database²⁵, after which their times of application were crossed with waveform dates. Time resolution for medication is of 15 minutes, meaning that the drug might have been injected at any point within the window of application.

²⁵ Section 2.3

RESULTS

4.1 AHE PREDICTION

To split train and test set we used stratified sampling to maintain the Hypotensive/Control proportion fixed in both sets. Samples in each set were chosen randomly in a 75%-25% Train-Test split. Afterwards the 2009 Physionet Challenge test set was used as benchmark for validation of the algorithms. See Table 4.1. For brevity we named the different train and test set with labels.

Table 4.1: Datasets Composition

	Main Dataset	
	train - CR	test - CS
C	228 - 68.47%	75 - 68.81%
H	105 - 31.53%	34 - 31.19%
	Challenge Dataset	
	train - TR	test - TS
C	25 - 54.35%	26 - 65%
H	21 - 45.65%	14 - 35%

Regardless of the Data Window, the best performers are always those with the smallest LT. This is expected since the shortest the time to the event the easiest the prediction. Generally data window size doesn't seem to influence the quality of the prediction, although it can be seen a recurrent pattern in which shorter data windows behave better with smaller LT and bigger data windows with longer prediction times. We set up a voting system to select the DW given the LT. For instance, given six classifiers and computed their training AUCs¹ for each LT, we ranked their performance based on the DW. With LT=10 min the majority of classifiers achieved higher AUCs with a DW of 10 min, hence for the LT of 10min the ideal DW was chosen to be 10 min.

With LT equal to 10min, all predictors chosen by LASSO are features somewhat related to the magnitude of the blood pressure. Table 4.2 and Table 4.3 shows the p-values and odd ratios of these features respectively for LT equals to 10 minutes and 30 minutes. The farthest in the time the prediction, the less relevant ABP related features are. This might be due to the fact that the closer we move to the AHE the more likely is to have a lower blood pressure or a decreasing blood

¹ Using 10-fold crossvalidation

pressure in hypotensive subjects.
For features description refer to the appendix.

Feature	AUC	p-value	OddRatio	LowerOdds	UpperOdds
ag70	0.81	3.6e-19	1.3492	1.2635	1.4406
m_diast	0.8511	2.14e-15	0.8532	0.8204	0.8873
as10	0.8376	2.59e-15	0.8876	0.8607	0.9136
minof	0.6924	5.25e-8	126.7743	22.1656	725.0740
CO	0.6569	1.59e-5	161.2032	16.0193	1.6222e+03

Table 4.2: Features analysis with LT=10min, DW = 10min in the training set of our Main Dataset (CR).

Feature	AUC	p-value	OddRatio	LowerOdds	UpperOdds
ag70	0.7663	2.8e-14	1.2633	1.1893	1.3418
as10	0.7991	1.74e-12	0.9107	0.8873	0.9347
as9	0.8040	5.59e-12	0.9098	0.8856	0.9346
m_diast	0.7923	7.42e-12	0.9044	0.8787	0.9308
minof	0.6358	4.55e-6	2245.8	82.9007	6.0839e+04
CO	0.6510	2.52e-5	125.7553	13.2628	1.1924e+03

Table 4.3: Features analysis with LT=30min, DW = 10min in the training set of our Main Dataset (CR).

Since the dataset is not balanced, during the training phase it was necessary to use weights to equilibrate sensitivity and specificity. As seen in figure 4.1a, without some form of misclassification cost the model tends to sacrifice sensitivity for higher specificity. On the other hand if we weight more hypotensive observations by means of misclassification cost, we reach an equivalence between sensitivity and specificity (see figure 4.1b). The tables 4.1a 4.1b 4.1c 4.1d show in details the performance of every classifier trained using weights.

Discriminant Analysis, overall, behaved the best with the best AUCs and equilibrium in the classification of both classes (Table 4.1a). When considering accuracy, trees have the upper hand. Of course, this cannot be considered a good measure for evaluating the model given the ratio of the two classes. Table 4.1b shows results on the test set; Discriminant Analysis, again, leads to better results in term of AUC. In this case, the sensitivity/specificity balance is degraded compared to the training, but still satisfying. As it concerns KNN and SVM, we can say that they achieve pretty good classifications with reasonable results in term of AUC and accuracy in both validation and test set.

Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.865	0.775	0.788	0.762	0.77
Discr_Quad	0.855	0.775	0.779	0.771	0.773
Discr_Sub	0.868	0.769	0.788	0.749	0.761
KNN_Coarse	0.853	0.773	0.779	0.767	0.77
KNN_Cosine	0.84	0.757	0.721	0.793	0.77
KNN_Medium	0.844	0.756	0.692	0.819	0.779
KNN_Mikowski	0.85	0.766	0.779	0.753	0.761
KNN_Sub	0.786	0.714	0.596	0.833	0.758
KNN_Weight	0.844	0.762	0.731	0.793	0.773
LogRegr	0.86	0.765	0.705	0.825	0.787
SVM_Cub	0.805	0.756	0.74	0.771	0.761
SVM_Lin	0.862	0.768	0.769	0.767	0.767
SVM_Quad	0.834	0.757	0.769	0.744	0.752
Tree_AdaBoost	0.835	0.756	0.644	0.868	0.798
Tree_RF	0.855	0.767	0.702	0.833	0.792
Tree_RUSBoost	0.831	0.755	0.663	0.846	0.789
Tree_TotalBoost	0.852	0.797	0.721	0.872	0.825

Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.886	0.788	0.735	0.84	0.807
Discr_Quad	0.884	0.773	0.706	0.84	0.798
Discr_Sub	0.884	0.788	0.735	0.84	0.807
KNN_Coarse	0.88	0.773	0.706	0.84	0.798
KNN_Cosine	0.86	0.793	0.706	0.88	0.826
KNN_Medium	0.862	0.772	0.676	0.867	0.807
KNN_Mikowski	0.865	0.758	0.824	0.693	0.734
KNN_Sub	0.79	0.726	0.559	0.893	0.789
KNN_Weight	0.875	0.824	0.794	0.853	0.835
LogRegr	0.888	0.786	0.706	0.867	0.817
SVM_Cub	0.763	0.758	0.676	0.84	0.789
SVM_Lin	0.88	0.773	0.706	0.84	0.798
SVM_Quad	0.831	0.753	0.706	0.8	0.771
Tree_AdaBoost	0.876	0.755	0.618	0.893	0.807
Tree_RF	0.87	0.8	0.706	0.893	0.835
Tree_RUSBoost	0.854	0.814	0.735	0.893	0.844
Tree_TotalBoost	0.817	0.735	0.618	0.853	0.78

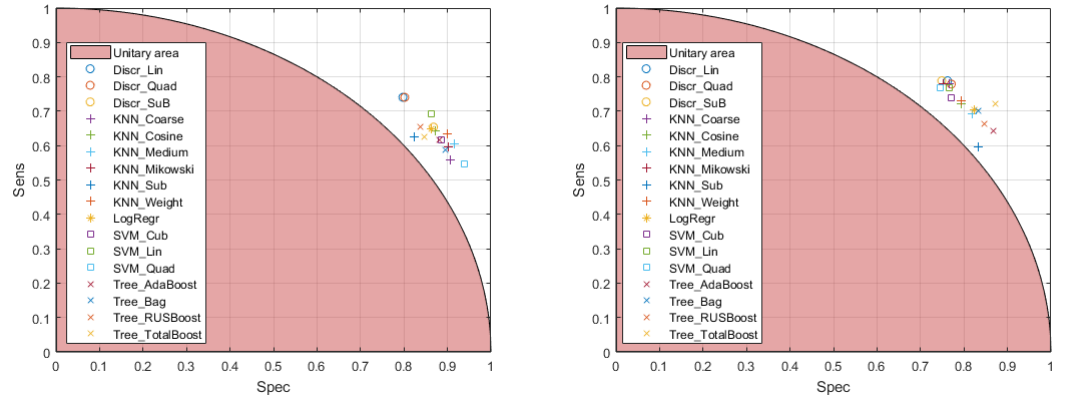
- (a) Performance from the 10 fold cross validation of the models on the training set of our main dataset (CR). LT=10min DW=10min.
- (b) Performance of all classifiers trained on the main dataset training set (CR) and tested on the main dataset test set (CS). LT=10min DW=10min.

Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.942	0.852	0.857	0.846	0.85
Discr_Quad	0.942	0.852	0.857	0.846	0.85
Discr_Sub	0.929	0.852	0.857	0.846	0.85
KNN_Coarse	0.948	0.852	0.857	0.846	0.85
KNN_Cosine	0.911	0.854	0.786	0.923	0.875
KNN_Medium	0.875	0.799	0.714	0.885	0.825
KNN_Mikowski	0.891	0.832	0.857	0.808	0.825
KNN_Sub	0.868	0.728	0.571	0.885	0.775
KNN_Weight	0.886	0.816	0.786	0.846	0.825
LogRegr	0.929	0.871	0.857	0.885	0.875
SVM_Cub	0.843	0.78	0.714	0.846	0.8
SVM_Lin	0.926	0.852	0.857	0.846	0.85
SVM_Quad	0.882	0.832	0.857	0.808	0.825
Tree_AdaBoost	0.9	0.854	0.786	0.923	0.875
Tree_RF	0.915	0.926	0.929	0.923	0.925
Tree_RUSBoost	0.897	0.835	0.786	0.885	0.85
Tree_TotalBoost	0.951	0.874	0.786	0.962	0.9

Classifier	AUC	AUC_OP	SENS	SPEC	ACC
Discr_Lin	0.937	0.871	0.857	0.885	0.875
Discr_Quad	0.931	0.871	0.857	0.885	0.875
Discr_Sub	0.931	0.852	0.857	0.846	0.85
KNN_Coarse	0.5	0.5	1	0	0.35
KNN_Cosine	0.902	0.871	0.857	0.885	0.875
KNN_Medium	0.909	0.852	0.857	0.846	0.85
KNN_Mikowski	0.87	0.852	0.857	0.846	0.85
KNN_Sub	0.817	0.709	0.571	0.846	0.75
KNN_Weight	0.905	0.799	0.714	0.885	0.825
LogRegr	0.937	0.871	0.857	0.885	0.875
SVM_Cub	0.945	0.874	0.786	0.962	0.9
SVM_Lin	0.937	0.852	0.857	0.846	0.85
SVM_Quad	0.945	0.907	0.929	0.885	0.9
Tree_AdaBoost	0.533	0.5	1	0	0.35
Tree_RF	0.933	0.907	0.929	0.885	0.9
Tree_RUSBoost	0.533	0.5	1	0	0.35
Tree_TotalBoost	0.533	0.5	1	0	0.35

- (c) Performance of classifiers trained on the training set of our dataset (CR) and tested on the test set of the Physionet challenge.
- (d) Performance of models trained on the Physionet Challenge training set (TR) tested on the Physionet Challenge test set (TS).

Models trained on the training set from the challenge for further details.



(a) Sensitivity Vs. Specificity for different classifiers on the training set of our main dataset. The various models didn't use weight during the training phase. (b) Sensitivity Vs. Specificity for different classifiers on the training set of our main dataset. LT= 10min DW=10min. Classifiers with sensitivity/specificity inside the unit circle were discarded.

Some of the trained models seemed to be badly parametrized: four models classify every observation as hypotensive. This might be caused by the excessive weight given to observations of this class in the training phase. The KNN_Sub has troubles to recognize hypotensive patients. This is may due to the fact that the subspace ensemble method rely on the instability of the classifiers to improve their performance, as KNN is fairly stable with respect to resampling, these methods fail in their attempt to improve the performance of KNN classifier.

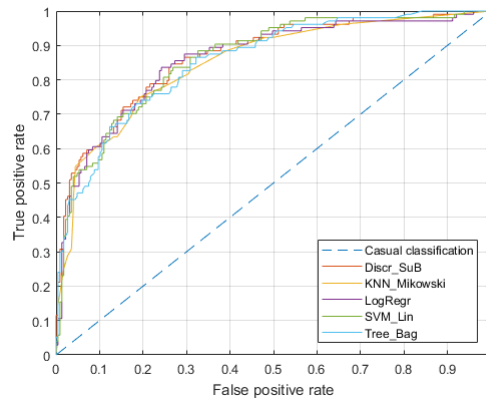


Figure 4.1: ROC curve of classifiers took as examples for all types of models.

In addition, classifiers trained on our dataset were tested on the Test Set of the challenge (TS) as well.

Of those, accuracy drops below .85 (34/40) only once (KNN_Sub) and

peaks at 0.925 (37/40) with the Tree_RF (see Table 4.1c), again. This is a satisfactory result, being that is tied to the currently highest challenge score².

For clarity we will call the pool of LASSO selected features (those from Table 4.2) as M5 as it will be a topic of the following Section³.

² <https://physionet.org/challenge/2009/final-scores>

³ Section 4.2.

4.2 AUTONOMIC ASSESSMENT

4.2.1 *Limitations and Problematics in ANS assessment in ICU*

Intensive Care Unit patients suffers from severe pathologies and are, more often then not, under the effect of several treatments. These two factors made the classical spectral bands analysis (see Section 3.7.2) particularly difficult. Moreover subject with pacemaker were characterized by controlled heart rate variabilities hiding their natural sinus rhythm. An overall analysis shows that LF and HF bands, either in absolute or normalized units, do not carry relevant information for the differentiation of the classes. Also, we couldn't find any statistically significant difference between the two groups at any point before or during the hypotension. For the aforementioned reasons also baroreflex sensitivity didn't reflect any pattern, at least with a statistical analysis⁴.

Figure 4.2a,4.2b and 4.2c picture the median behaviour of the LF,HF and LF/HF computed using a PP routine for the two classes thirty minutes before the hypotension onset. As it can be seen, there is no difference in the value and in the pattern. Interestingly, although not significant, the LF/HF ratio from an open loop baroreflex model is consistently lower for hypotensive subjects (Figure 4.2d), suggesting impaired sympathatetic activation for this group. However, given the susceptibility of these model to noise, we are skeptical on the reliability of these results, fact highlighted by the absence of classification performance increase when included in the final prediction model.

4.2.2 *Classification*

This phase was split in three distinct analysis:

1. Univariate Logistic Regression to evaluate the predictive power of single features using only baroreflex regressors⁵.
2. Multivariate Logistic Regression for overall performance evaluation using only baroreflex regressors.
3. Predictive model implementation using a large feature space with features from baroreflex, arterial blood pressure and heart rate variability.(metti citazione a noi stessi)

Predictions were performed for all three cases ten minutes before the hypotension onset (lead time) using a ten minutes worth of data (data window) for feature extraction.

⁴ Although we were still able to find promising results using a closed Point Process model, see next section.

⁵ skewpks and LWkurtosis, see Section 3.6.2.3.

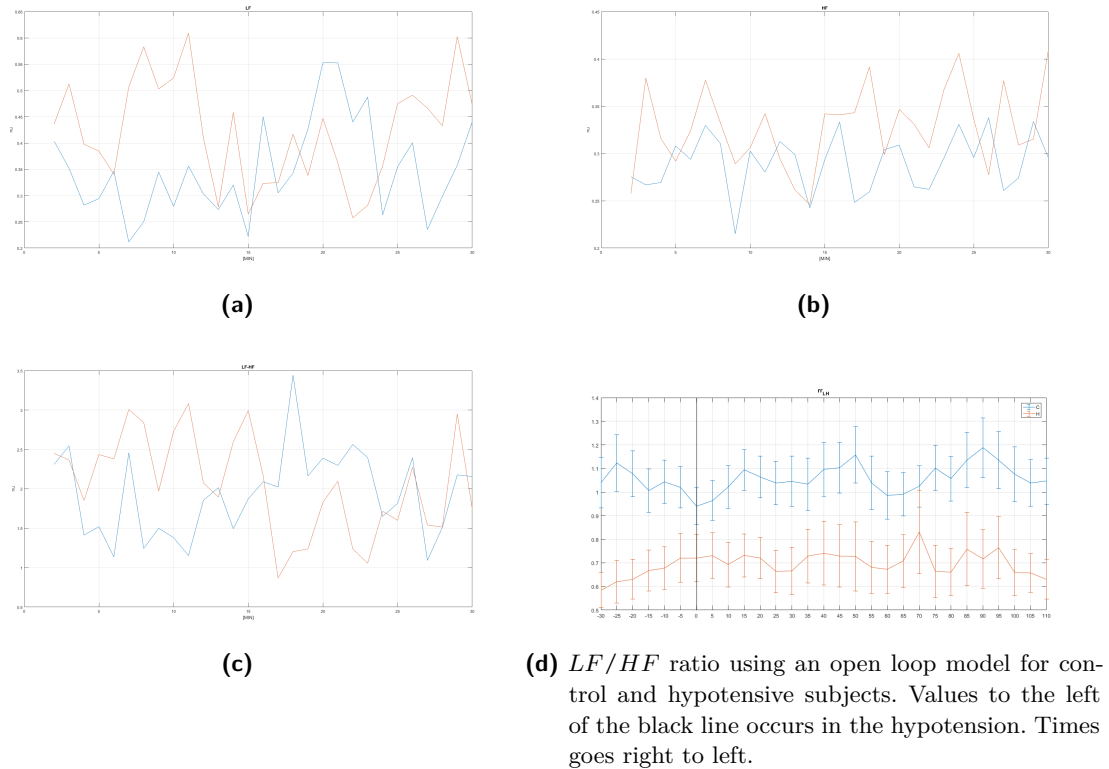


Figure 4.2: Median behaviour of spectral components computed using a Point Process framework. Figure 4.2d was calculated using the Welch's Periodogram.

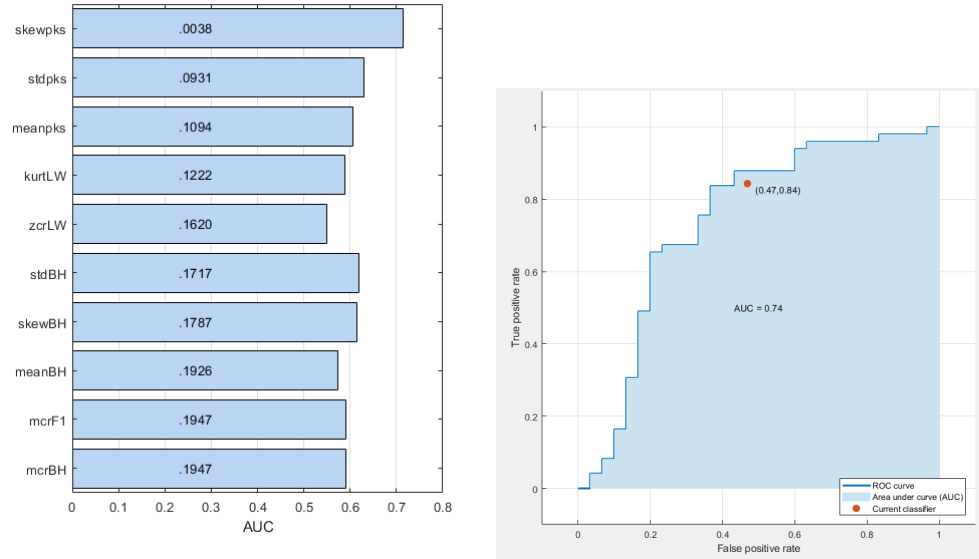
From univariate logistic regression we discovered that the skewness of detected peaks, see 3.6.2.3 for explanation, when used alone is able to discriminate to a certain degree the two classes, reaching an AUC of .71. This is by itself a very good results because it remarks on the ranksum outcome but also achieves a good classification performance: all other tested features failed to obtain this level. Moreover the peaks skewness also shows a p-value of .0038, but still without confoundings considerations.

All other features did not surpass the null hypothesis and also achieved much lower AUCs. See Figure 4.3a.

In the multivariate case, first we tried to fed the entire dataset to the classifier. Results were poor, with AUC below .6 using ten fold cross-validation. This outcome was somewhat expected because the number of feature is almost a half the number of observations, moreover predictors might be co-linear.

To overcome the issue and to identify relevant features we used a feature selection algorithm. With LASSO we identified two features: Baseline kurtosis and, again, the skewness of the peaks.

With ten-fold cross-validation the AUC was .74, see Figure 4.3b, which is more than peak skewness alone, by a 3% margin, but still not accept-



(a) The first ten features by pvalue tested with univariate logistic regression. Horizontal bars are the AUCs for respective features, while the value lying in each bar is the associated p-value. skewpks is the peaks skewness. (b) Multivariate Logistic Regression after LASSO on the baroreflex only feature space.

Feature	C	H	p-val	OddsRatio
skew_pks	-0.0247 ± 0.5566	0.3896 ± 0.5668	.0064	4.0375

Table 4.4: The significance threshold is p-val < .05.

able: Sensitivity was below 60% and Accuracy of 73.5%. Probably by assigning weights, one could increase the sensitivity/specificity balance at cost of lower accuracy, but, overall, with only marginal gains in performance.

After the identification of the features with the highest predictive power, it was time to make sure that this new information helps the features selected in the previous analysis to discriminate between the two classes. Two classification were performed using two different dataset: the first composed with the same five features used previously (ag70, m_diast, as10, minof and CO), and the second one equal to the first with skewness and kurtosis. As classifier was used the Linear Discriminant Analysis, since it has already demonstrated to be one of the most performing algorithms.

The training and cross-validation of the classifier was performed 100 times to have a better estimation of the values of AUC of the optimal point, Sensitivity, Specificity and Accuracy.

As shown form Figure 4.3, whit the addition of the two baroreflex related features the classification improves significantly in every aspect:

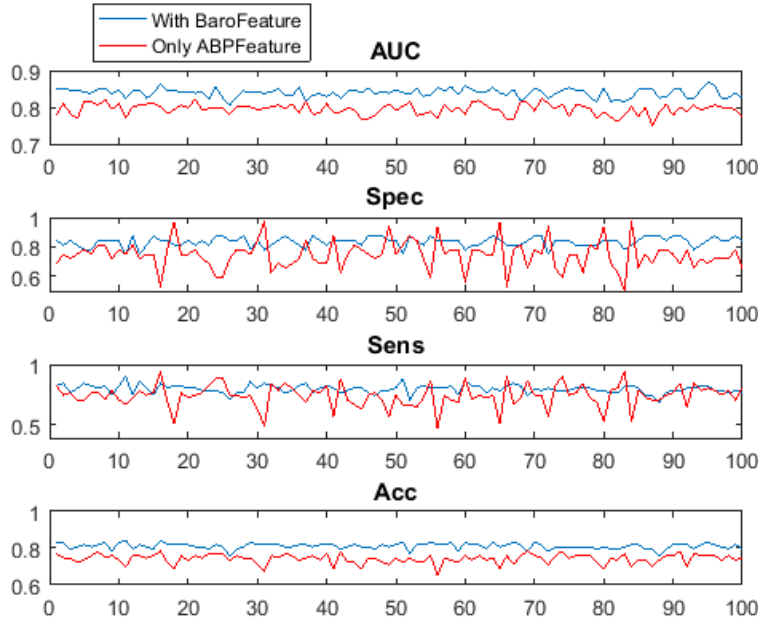


Figure 4.3: Graphs showing the values of AUC in the optimal point, Sensitivity, Specificity and Accuracy for both classification with and without baroreflex features.

	AUC	Spec	Sens	Acc
$M5 + skewpks + kurtLW$	0.815 ± 0.013	0.834 ± 0.03	0.796 ± 0.03	0.811 ± 0.015
$M5$	0.742 ± 0.016	0.745 ± 0.09	0.739 ± 0.09	0.742 ± 0.026

Table 4.5: Performance table

every single graph shows a non negligible improvement of the specific quantity. It is important to notice that by using skewness and kurtosis the graphs relative to Sensibility and Specificity, not only show a much higher mean value, but also a more steady value along the different iterations. In table 4.5 is possible to quantify the improvement: the AUC of the optimal point has an improvement of more than 6 percentage points. Sensitivity improves by 5 percentage points and Specificity jumps from 0.745 to 0.834, an improvement of nine percentage points.

With a deeper analysis we investigated the influence of interventions to understand if we were missing something trivial. We used the logistic regression to assess the influence of medications and treatments: of the new features, only skewpks is really significant once adjusted for confoundings, with a p-value of .0064.

The distribution of skewpks between hypotensive and control is shown in Figure 4.4. The zero-mean normality test tells us that control patients follow a Gaussian distribution with a p-value of .9 while sick subjects not (Table 4.6). Even though the number of samples is not

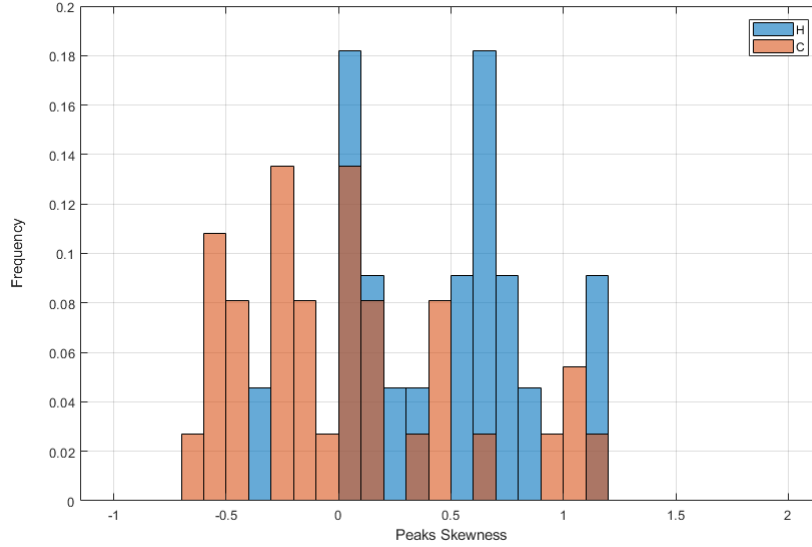


Figure 4.4: Skewness of the baroreflex peaks. Controls are orange, Hypotensives are blue.

	pval	H
C	.9656	0
H	.0028	1

Table 4.6: Zero-mean normality distribution test for the population of peaks skewness for control and hypotensive records. The test rejects the null hypothesis at the 5% significance level.

the largest, it is easy to see how hypotensive subjects (blue in Figure 4.4) are shifted towards positive skewness values compared to controls. In terms of values, controls have a slightly negative mean skewness (-0.0247, Table 4.4) when hypotensives average at 0.3896. Standard deviations are similar for both groups at around .5. Hypotensive subjects seem to follow a bimodal distribution, one peak assuming zero value the other just more than .5. Perhaps a subset of the AHE-group suffers from a different type of hypotension we are yet to identify.

In the end we are satisfied with the results, having improved the overall performance of the model.

4.3 CONCLUSIONS

4.3.1 *Physiology of hypotension*

By inspecting the mean behavior of single annotations, one could make some general considerations. First of all, some were of little surprise: Figure 4.5b displays the relative cardiac output⁶ for control (blu line) and hypotensive (orange line) subjects starting almost two hours before hypotension onset to 30 minutes after it. Mean Cardiac output is one of the most predictive features and we were expecting to see a lower value compared to the controls. As it turns out, AHE-patients on average have a relative cardiac output .45 while control of .4; moreover neither class shows relevant changes in variability hinting to very different dynamics or common patterns. The drop is concurrent to the hypotension onset, leaving no warnings beforehand, but always maintaining a value greater than the healthy counterpart.

Another interesting behavior is the increase of pulse transit times after hypotension. Any feature related to PTTs is not significant for the differentiation of the two groups, still they show a pattern. This is in line from what reported by Kounalakis and Geladas, 2009, hence a negative correlation with systolic pressure and cardiac output and positive correlation with arterial stiffness.

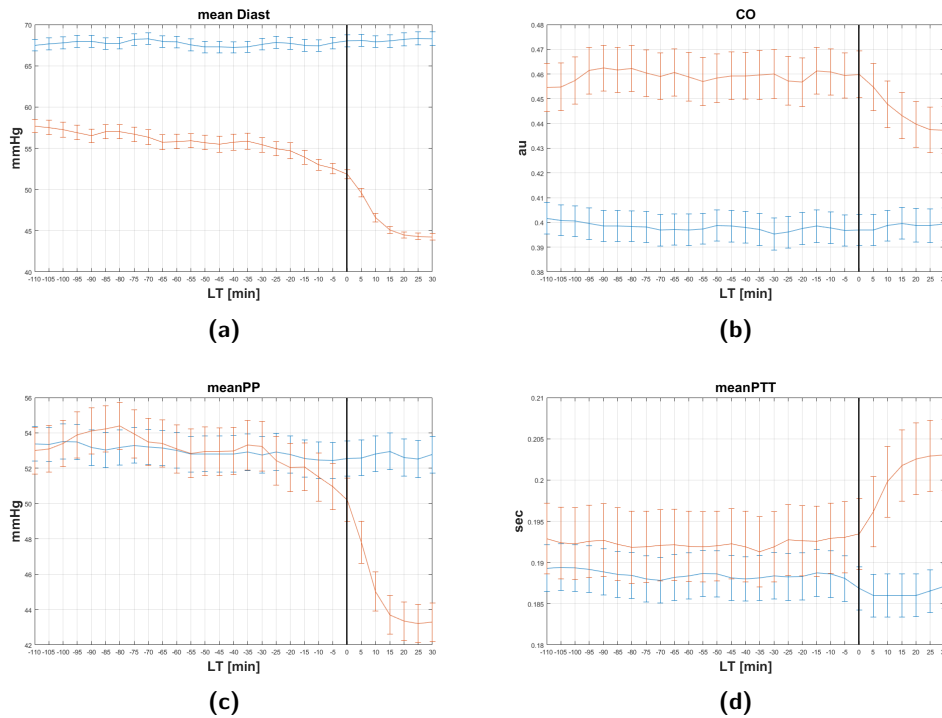


Figure 4.5: Mean and standard Error for selected physiological variables.

⁶ See Eq.3.2

In contrast with relative cardiac output and PTT, pulse pressure shows signs of instability before the insurgence of the condition, anticipating it by a couple of minutes. Pulse pressure narrows sharply with hypotension, perhaps due to a drop in stroke volume or blood loss. Also, this suggests that the rate at which systolic pressure drops is higher than diastolic.

In Figure 4.6b is shown the mean behavior of the RRI: heart rate becomes more bradycardic with hypotension, but again, with no definite pattern.

In the end, the most relevant features for the goal of prediction are those extracted from blood pressure: MAP, systolic and diastolic pressure. All these features were shown to be the most predictive (Tables 4.2, 4.3), particularly the ones related to diastolic pressure and MAP. Score features, features extracted from the TBP transform⁷, had the highest predictive power: predictors like ag70 (See Figure 4.6d) create a good contrast at any point time.

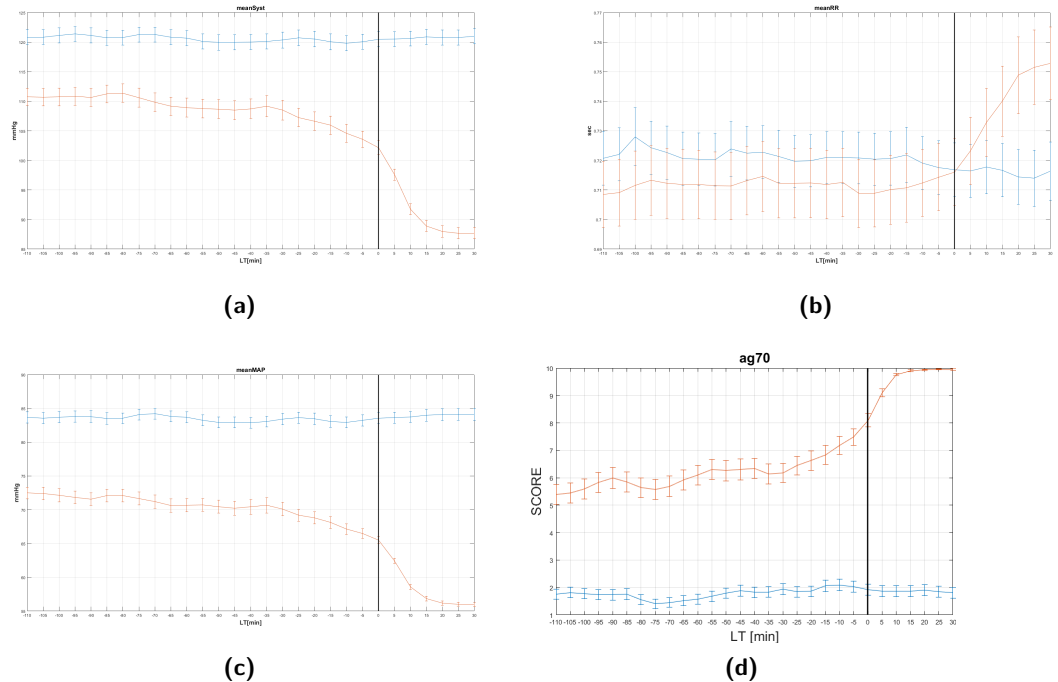


Figure 4.6: Mean and standard Error for selected physiological variables.

To look for patterns we used hierarchical clustering with euclidean distance. Figures 4.7a and 4.7c are the result of the clustering on the signals using euclidean distance as measure for similarity. The analysis was performed on hypotensive subjects only, since controls did not show any relevant trend.

The clusters of the systole (closely related to that of the diastole) shows

⁷ Section 3.7.4.

something that we have already noticed by looking through the data: there are several manifestations of hypotension. We identified three main kinds of hypotension:

- **Slow AHE:** The pressure drops slowly, taking hours to reach critical levels. This is the easiest kind to predict, showing a well defined trend. The reason behind the event might maybe be related to sepsis mechanisms. In Figure 4.7a those records are the one portrayed in the upper interval of the figure.
- **Sudden AHE:** Pressure is relatively stable to drop suddenly before hypotension. These are the observations in the mid-lower interval of Figure 4.7a.
- **Inverse AHE:** Pressure is rising before rapidly falling to critical levels. They can be seen near the lower bezel of the systolic pressure cluster figure.

Talking with clinicians we have been suggested that case two and three might be an allergic reaction to a medication. Case two and three are also the hardest to predict, since they give little or none notice. The frequency of occurrence of the two is lower than the first type.

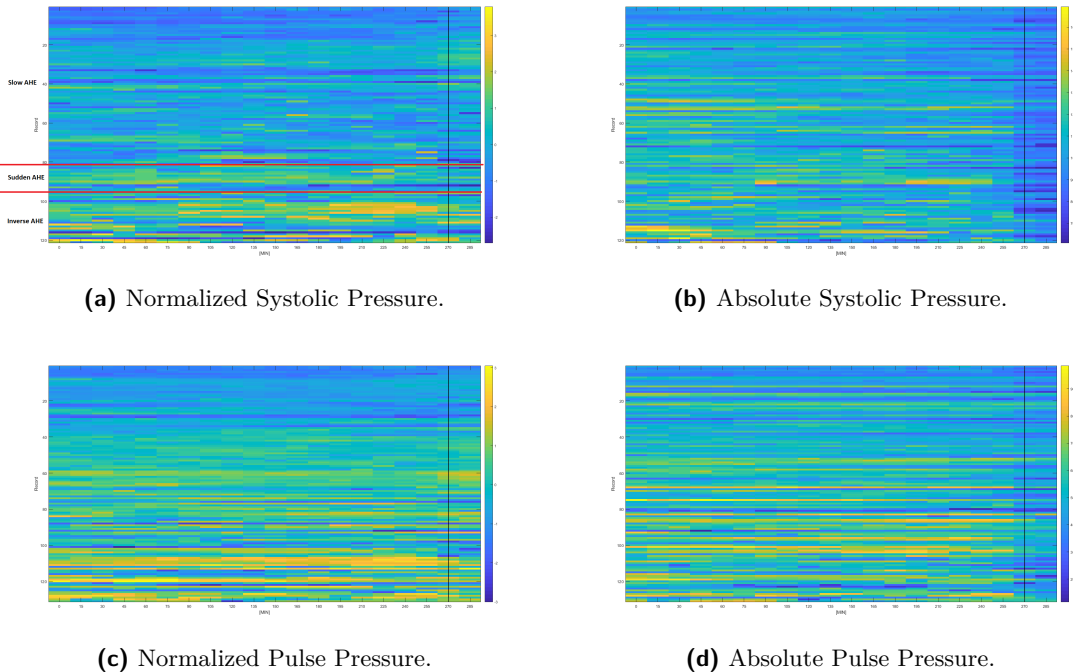


Figure 4.7: Hierarchical clustering of systolic pressure and pulse pressure of the AHE-patients. In Figures 4.7a and 4.7c brighter colors mean that the value is above the population average in that instant, darker colors below average. Figures 4.7b and 4.7d are instead in absolute values.

4.3.2 Influence of interventions.

The patients that have been used to train and test the selected models didn't encounter any kind of selection process regarding health state or medication. In this study the approach is to simulate a real case scenario in which any patient could be included in the analysis. The obtained performance are still good and this highlight the solidity of the trained models.

An analysis was conducted to investigate the classification of patients with and without specific kind of interventions. In table 4.7 is possible to see the different classifications performance obtained with the Linear Discriminant Analysis, repeating one hundred times the process of training and cross-validation on the cohort of 442 subjects. This algorithm was choose because it showed more consistency and better performance during the various analysis.

Intervention		AUC	Sens	Spec	Acc
Vasopressors	No	0.786 ± 0.01	0.788 ± 0.01	0.785 ± 0.01	0.786 ± 0.01
	Yes	0.765 ± 0.01	0.848 ± 0.01	0.681 ± 0.01	0.75 ± 0.01
Sedative	No	0.786 ± 0.01	0.795 ± 0.01	0.776 ± 0.01	0.782 ± 0.01
	Yes	0.783 ± 0.01	0.821 ± 0.01	0.744 ± 0.01	0.772 ± 0.01
Ventilation	No	0.806 ± 0.01	0.803 ± 0.01	0.809 ± 0.01	0.807 ± 0.01
	Yes	0.766 ± 0.01	0.800 ± 0.01	0.732 ± 0.01	0.756 ± 0.01
Pacemaker	No	0.806 ± 0.01	0.794 ± 0.01	0.819 ± 0.01	0.811 ± 0.01
	Yes	0.735 ± 0.01	0.820 ± 0.01	0.651 ± 0.01	0.705 ± 0.01

Table 4.7: Table with performance indices analyzed by interventions. Every intervention has two rows: the first shows the indeces about patients without the specific intervention (row 'No'), and the second indeces about patients with it (row 'Yes').

It can be seen that in all the four categories there is a different behavior between who received the medication and who did not. The intervention that less affected the classification outcomes are the sedatives. The drop of performance is in fact less marked than the others and the main affect is an imbalance between Sensitivity and Specificity. Vasopressors show a stronger but similar behavior.

The interventions that mostly affect the classifications are mechanical ventilation and pacemakers. In mechanical ventilation Specificity is the quantity that shows the strongest loss of more than 7 percentage points: patients without ventilation shows a specificity of 0.81 while for patients with ventilation, it drops at 0.73. The Sensitivity seems to be invariant but the general effect is a weakening of AUC and overall Accuracy. This could be caused, not directly by the presence of assisted ventilation but by the concauses of this intervention: this kind of patients are often in critical condition and/or heavily sedated.

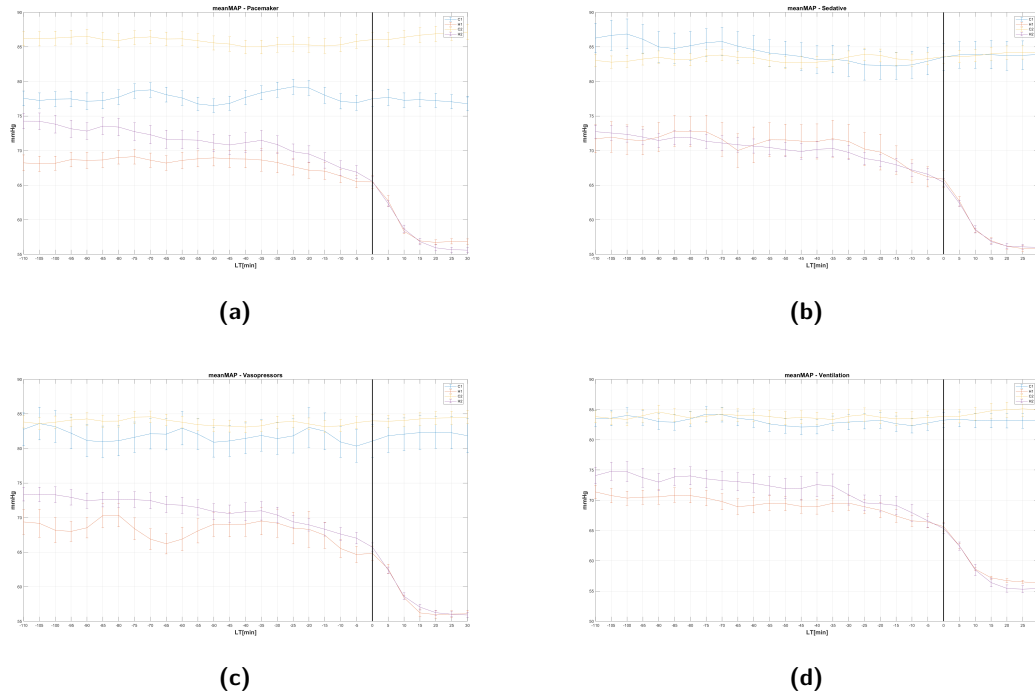


Figure 4.8: Effect of different medications/interventions on mean arterial pressure. C are control, H are hypotensive patients. the number 1 indicates the group has received the medication, number 2 they did not. The strongest effect on MAP is given by pacemakers.

Pacemakers show the same effects: the Specificity of patients that use this intervention is 0.65. Patients without need of pacemaker are much easier to classify, in fact they show a Specificity of 0.82. This is caused by the different physiology caused by this intervention: when the heart rhythm is controlled by a pacemaker, the control feedback from ABP to heart beat does no longer exist.

Like in Mechanical Ventilation, Sensitivity does not vary. Due to the decreased Specificity, the classification results of patients with Pacemaker are lower than in patients without them.

The same analysis can be performed even on the selected cohort for the analysis of baroreflex. It is possible to see in table 4.9a that, also in this case, interventions have a strong impact on classification performance. However their influence seems to be quite different: the presence of Vasopressor, Sedative and Ventilation still alter the balance between Sensitivity and Specificity but to a lesser extent. Even more, the effect of these medications seems to increase the AUC and Accuracy by a non negligible quantity. This could be caused by the fact that the analysis of the effect of every intervention is made comparing the performance of the patients using as control group those patients that didn't receive the intervention under analysis. It is possible however that this "control group" could have been undergone one or more of the other interven-

tions.

However, it is possible to see that Pacemakers have the same effect as before: they alter the physiology of the interested patients and consequently they lower the classification performance.

This highlight the fact that an analysis made taking into account the various types of interventions would lead to better results.

Intervention		AUC	Sens	Spec	Acc
Vasopressors	No	0.793 ± 0.02	0.789 ± 0.02	0.796 ± 0.02	0.794 ± 0.02
	Yes	0.874 ± 0.02	0.855 ± 0.02	0.892 ± 0.02	0.875 ± 0.02
Sedative	No	0.803 ± 0.02	0.765 ± 0.02	0.842 ± 0.02	0.81 ± 0.02
	Yes	0.851 ± 0.04	0.949 ± 0.04	0.754 ± 0.04	0.811 ± 0.04
Ventilation	No	0.753 ± 0.02	0.727 ± 0.02	0.779 ± 0.02	0.757 ± 0.02
	Yes	0.872 ± 0.02	0.902 ± 0.02	0.841 ± 0.02	0.861 ± 0.02
Pacemaker	No	0.824 ± 0.01	0.815 ± 0.01	0.833 ± 0.01	0.826 ± 0.01
	Yes	0.751 ± 0.03	0.762 ± 0.03	0.741 ± 0.03	0.748 ± 0.03

(a) Performance indices on the selected cohort for the analysis on baroroflexes.

Intervention		AUC	Sens	Spec	Acc
Vasopressors	No	0.749 ± 0.02	0.739 ± 0.02	0.76 ± 0.02	0.752 ± 0.02
	Yes	0.757 ± 0.06	0.656 ± 0.06	0.858 ± 0.06	0.763 ± 0.06
Sedative	No	0.756 ± 0.02	0.721 ± 0.02	0.79 ± 0.02	0.762 ± 0.02
	Yes	0.728 ± 0.05	0.706 ± 0.05	0.751 ± 0.05	0.737 ± 0.05
Ventilation	No	0.731 ± 0.03	0.719 ± 0.03	0.743 ± 0.03	0.733 ± 0.03
	Yes	0.76 ± 0.03	0.716 ± 0.03	0.805 ± 0.03	0.775 ± 0.03
Pacemaker	No	0.79 ± 0.02	0.746 ± 0.02	0.834 ± 0.02	0.799 ± 0.02
	Yes	0.57 ± 0.06	0.57 ± 0.06	0.569 ± 0.06	0.569 ± 0.06

(b) Table that show performance indeces on classification performed with a Linear Discriminant Analysis on the selected cohort for baroreflex analysis. The used features are: ag70, m_diast, as10, minof and CO.

In table 4.9b are shown the results of the classification done without using the features related to baroreflex. These data show the same behavior: interventions alter the classification results, and in particular pacemaker negatively skew the outcomes.

4.3.3 *Machine Learning as a Decision Making tool in ICU*

We couldn't find a single feature embodying all types of hypotensions. The condition is the result of many alterations affecting an already complex system. In addition each subject has its own clinical history with distinctive characteristics making extremely hard to generalize patterns. Moreover the poor quality of raw data lower even more the accuracy of the outcomes.

The task, therefore, requires the development of predictive models for long sequences of non-stationary multivariate time series in sub-optimal conditions.

However, thanks to the advancement in technology and its expansion in critical care, researcher have at their disposal an always increasing amount of health data. The rise of relational databases, such as MIMIC, provide them with resources to finally apply those techniques, once constrained by the lack data, falling under term of machine learning. These models can analyze bigger and more complex data, grasp hidden factors and provide useful decision making tools. A possible solution might be the coupling of reinforcement learning models, Deep Q-Learning before all, with advanced signal processing techniques like the Point Process. The latter enables the extraction of strong measurements of the autonomic activation at the beat-to-beat level, where others, instead, fail. The importance of having a strong base for feature extraction is shown in this work: the inclusion of features originating from the Point Process increased substantially the quality of the model, a model already using most of the predictors seen in literature, in a real-life scenario typical of the ICU environment. On the other hand, Temporal Difference Learning could be a key into merging information from the many with patterns from the single to enter the era of **precision medicine**⁸: to tailor the medical treatment to the individual characteristics of each patient.

8 Johnson et al., 2016a

Appendices



RELEVANT FEATURES DESCRIPTION

- AVNN: Average RR interval computed in seconds;
- SDNN: Standard Deviation of the RR intervals in seconds;
- NN20: NN20 count;
- pNN20: NN20 expressed as percentage of the total;
- NN50: NN50 count;
- pNN50: NN50 expressed as percentage of the total
- RMSSD: Square root of the mean squared differences between successive RR in ms;
- meanABP: mean arterial blood pressure, as mean of the whole signal;
- stdABP: standard deviation of arterial blood pressure;
- slopeABP: slope of the line used to interpolate the ABP signal;
- intrcptABP: intercept of the line used to interpolate the ABP signal;
- m_syst: mean systole amplitude;
- st_syst: standard deviation of systole;
- m_diast: mean diastole amplitude;
- st_diast; standard deviation of diastole;
- meanPP: pulse pressure mean amplitude;
- stdPP: pulse pressure standard deviation
- meanPS: mean time between fiducial point and the following systole;
- stdPS: standard deviation of time between fiducial point and the following systole;
- meanSD: mean time between systole and diastole;
- stdSD: standard deviation of time between systole and diastole;
- minof: percentage ABP signal of points below 60 mmHg;

- minof70: percentage ABP signal of points below 70 mmHg;
- minof80: percentage ABP signal of points below 80 mmHg;
- minof90: percentage ABP signal of points below 90 mmHg;
- est_P: see chapter [3.7.1](#)
- MAP: mean arterial pressure;
- CO: Cardiac Output. See [3.5.5](#)
- MAPslope: slope of the line that interpolate MAP signal;
- MAPintcpt: intercept of the line that interpolate MAP signal;
- stslope: slope of the linear regression of the last 10% of available MAP;
- stint: intercept of the linear regression of the last 10% of available MAP;
- slow: Flag set to one when stslope and MAPslope are lower than zero.
- sudden1: Flag set to one when stslope and MAPslope are greater than zero.
- sudden2: Flag set to one when stslope is greater than zero and MAPslope is lower than zero.
- sudden3: Flag set to one when stslope is lower than zero and MAPslope is greater than zero.
- meanPTT: mean Pulse Transit Time;
- stdPTT: standard deviation of PTT;
- meanD: mean diastole;
- stdD: standard deviation of diastole;
- meanPO: mean Pulse Onset (fiducial point);
- stdPO: standard deviation of Pulse Onsets;
- rr_apen: Sample Entropy of *RR* distances. See [3.7.3](#);
- rr_D2: Correlation Dimension of *RR* distances. See [3.7.3](#);
- ss_apen: Sample Entropy of systogram. See [3.7.3](#);
- ss_D2: Correlation Dimension of systogram;
- dd_apen: Sample Entropy of diastogram;

- dd_D2: Correlation Dimension of diastogram;
- rr_TOTPWR: tachogram total spectral power. See [3.7.2](#)
- rr_ULF: tachogram spectral power of Ultra Low Frequency;
- rr_VLF: tachogram spectral power of Very Low Frequency;
- rr_LF: tachogram spectral power of Low Frequency;
- rr_HF: tachogram spectral power of High Frequency;
- rr_L_H: ratio of tachogram low frequency and high frequency power;
- BRFX: estimation of baroreflex;
- po_TOTPWR: Total power of the fiducial pressure.
- po_ULF: ULF of the fiducial pressure*.
- po_VLF: VLF of the fiducial pressure*.
- po_LF: LF of the fiducial pressure*.
- po_HF: HF of the fiducial pressure*.
- po_L_H: LF/HF of the fiducial pressure*.
- ss_TOTPWR: systogram total spectral power;
- ss_ULF: systogram spectral power of Ultra Low Frequency;
- ss_VLF: systogram spectral power of Very Low Frequency;
- ss_LF: systogram spectral power of Low Frequency;
- ss_HF: systogram spectral power of High Frequency;
- ss_L_H: ratio of systogram low frequency and high frequency power;
- dd_TOTPWR: Total power of the diastogram.
- dd_ULF: ULF power of the diastogram.
- dd_VLF: VLF power fo the diastogram.
- dd_LF: LF power of the diastogram.
- dd_HF: HF power of the distogram.
- dd_L_H: LF/HF ratio of the diastogram.
- PPslope: See section [3.7.1](#).
- PPintrcp: See section [3.7.1](#).

- as1: First TBP element.
- as2: Second TBP element.
- as3: Third TBP element.
- as4: Fourth TBP element.
- as5: Fifth TBP element.
- as6: Sixth TBP element.
- as7: Seventh TBP element.
- as8: Eighth TBP element.
- as9: Ninth TBP element.
- as10: Tenth TBP element.
- ag_area: See Equation 3.35.
- d_ag: See Equation 3.36.
- l_ag: See Equation 3.37.
- meanag: Mean TBP.
- ag70: See section 3.7.4.
- b80: Elements in TBP lower than 80mmHg.
- agsudden: Flag set to one when ag10 is lower than 80mmHg and d_ag is lower than one.
- ag80: Number of elements in TBP with values between 70mmHg and 80mmHg.
- scr90: Flag set to one when up90 is greater than 3.
- up90: See Equation 3.40.
- lstABP: Mean ABP of the last two minutes and half available in the Data Window.

*fiducial pressure is that time series composed by the percussion wave slope pressure for each beat, see Section 3.5.

ACRONYMS

- ANS: Autonomic Nervous System
- AUC: Area Under the Curve
- RBA: Relative Bursts Amplitude
- PP: Pulse Pressure
- RRI: RR Intervals
- ABP: Arterial Blood Pressure
- SBP: Systolic Blood Pressure
- DBP: Diastolic Blood Pressure
- PTT: Pulse Transit Times
- MAP: Mean Arterial Pressure
- ROC: Receiver Operating Characteristic
- SNR: Signal to Noise Ratio
- MIMIC: Medical Information Mart for Intensive Care
- CO: (relative) Cardiac Output
- AHE: Acute Hypotensive Episode
- LASSO: Least Absolute Shrinkage and Selection Operator
- PCA: Principal Component Analysis
- DFA: Detrended Fluctuation Analysis
- RMSSD: Root Mean Square of Successive Differences
- SQI: Signal Quality Index
- GRNN: General Regression Neural Network
- LCP: Laboratory for Computational Physiology
- PSD: Power Spectral Density
- ICU: Intensive Care Unit
- RSA: Respiratory Sinus Arrhythmia

- NN: Neural Network
- FNN: Feed forward Neural Network
- RNN: Recurrent Neural Network
- GRNN: Generalized Regression Neural Network
- HHT: Hilbert-Huang Transform
- IMF: Intrinsic Mode Functions
- EMD: Empirical Mode Decomposition
- HSA: Hilbert Spectrum Analysis
- AMB: Amplitude Modulation Bandwidth
- FMB: Frequency Modulation Bandwidth
- DW: Data Window
- LT: Lead Time

MEASURES OF PERFORMANCE

$$\text{Specificity or TNR} = \frac{TN}{TN + FP} \quad (\text{C.1})$$

$$\text{Sensitivity or TPR} = \frac{TP}{TP + FN} \quad (\text{C.2})$$

$$\text{False Positive Rate FPR} = 1 - TNR \quad (\text{C.3})$$

$$\text{AUC} = \int_{\infty}^{-\infty} TPR(T)(-FPR'(T))\delta T \quad (\text{C.4})$$

$$\text{Accuracy} = \frac{(TP + TN)}{TP + TN + FP + FN} \quad (\text{C.5})$$

$$(\text{C.6})$$

Where T is the Threshold.

BIBLIOGRAPHY

- Barbieri, Riccardo, Eric C Matten, AbdulRasheed A Alabi, and Emery N Brown (2005). “A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability.” In: *American Journal of Physiology-Heart and Circulatory Physiology* 288.1, H424–H435 (cit. on p. 35).
- Bassale, Jules (2001). “Hypotension prediction arterial blood pressure variability.” In: *Tecnical report* (cit. on pp. xi, 8, 9, 21).
- Breiman, Leo (2001). “Random forests.” In: *Machine learning* 45.1, pp. 5–32 (cit. on p. 61).
- Camm, A John, Marek Malik, JT Bigger, G Breithardt, Sergio Cerutti, RJ Cohen, P Coumel, EL Fallen, HL Kennedy, RE Kleiger, et al. (1996). “Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology.” In: *Circulation* 93.5, pp. 1043–1065 (cit. on pp. 43–45).
- Chen, X, D Xu, G Zhang, and R Mukkamala (2009). “Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform.” In: *Computers in Cardiology, 2009. IEEE*, pp. 545–548 (cit. on pp. xi, xii, 11, 12, 22, 41, 42, 49).
- Crespo, Cristina, James McNames, Mateo Aboy, Jules Bassale, Miles Ellenby, Susanna Lai, and Brahm Goldstein (2002). “Precursors in the arterial blood pressure signal to episodes of acute hypotension in sepsis.” In: *Proceedings of the 16th International EURASIP Conference BIOSIGNAL*. Vol. 16, pp. 206–8 (cit. on pp. 8, 22).
- Dai, Dong and Shaowen Hua (2016). “Random Under-Sampling Ensemble Methods for Highly Imbalanced Rare Disease Classification.” In: *Proceedings of the International Conference on Data Mining (DMIN)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p. 54 (cit. on p. 61).
- De Backer, De Backer and Vrints (2017). *Blood pressure*. Ed. by European Society of Cardiology. URL: <https://www.escardio.org/Education/ESC-Prevention-of-CVD-Programme/Treatment-goals/Risk-factor-control/blood-pressure> (cit. on p. 3).
- Downham, Jonathan. *Mechanical Ventilation: Physiologic effects of mechanical ventilation*. URL: <http://www.jonathandownham.com/mechanical-ventilation-physiologic-effects-mechanical-ventilation/> (cit. on p. 63).

- Freund, Yoav, Robert Schapire, and Naoki Abe (1999). “A short introduction to boosting.” In: *Journal-Japanese Society For Artificial Intelligence* 14.771-780, p. 1612 (cit. on p. 60).
- Ghassemi, Marzyeh (2011). “Methods and models for acute hypotensive episode prediction.” In: *PhD diss., MSc Thesis* (cit. on pp. xi, xii, 22).
- Haykin, Simon S, Simon S Haykin, Simon S Haykin, and Simon S Haykin (2009). *Neural networks and learning machines*. Vol. 3. Pearson Upper Saddle River, NJ, USA: (cit. on p. 15).
- Henriques, JH and TR Rocha (2009). “Prediction of acute hypotensive episodes using neural network multi-models.” In: *Computers in Cardiology, 2009*. IEEE, pp. 549–552 (cit. on pp. 10, 11, 16).
- Jiang, Dazhi, Bo Hu, and Zhijian Wu (2017). “Prediction of acute hypotensive episodes using EMD, statistical method and multi GP.” In: *Soft Computing* 21.17, pp. 5123–5132 (cit. on pp. 13–15).
- Johnson, Alistair EW, Mohammad M Ghassemi, Shamim Nemati, Katherine E Niehaus, David A Clifton, and Gari D Clifford (2016a). “Machine learning and decision support in critical care.” In: *Proceedings of the IEEE* 104.2, pp. 444–466 (cit. on p. 81).
- Johnson, Alistair EW, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark (2016b). “MIMIC-III, a freely accessible critical care database.” In: *Scientific data* 3 (cit. on p. 20).
- Jolliffe, Ian T (1986). “Principal Component Analysis and Factor Analysis.” In: *Principal component analysis*. Springer, pp. 115–128 (cit. on p. 52).
- Kharod, Shivam, Candice Norman, Matthew Ryan, and Robyn M Hoelle (2014). “Severe Unexplained Relative Hypotension and Bradycardia in the Emergency Department.” In: *Case reports in emergency medicine* 2014 (cit. on p. 5).
- Kounalakis, Stylianos N and Nickos D Geladas (2009). “The role of pulse transit time as an index of arterial stiffness during exercise.” In: *Cardiovascular Engineering* 9.3, pp. 92–97 (cit. on p. 75).
- Lee, Joon and Roger G Mark (2010). “An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care.” In: *Biomedical engineering online* 9.1, p. 62 (cit. on p. 15).
- Li, Bing Nan, Ming Chui Dong, and Mang I Vai (2010). “On an automatic delineator for arterial blood pressure waveforms.” In: *Biomedical Signal Processing and Control* 5.1, pp. 76–81 (cit. on pp. xiv, 29).
- Li, Qiao, Roger G Mark, and Gari D Clifford (2009). “Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator.” In: *Biomedical engineering online* 8.1, p. 13 (cit. on p. 27).

- Manaker, Scott. *Use of vasopressors and inotropes*. URL: <https://www.uptodate.com/contents/use-of-vasopressors-and-inotropes> (cit. on p. 63).
- Moody, George B. *gQRS*. URL: <https://physionet.org/physiotools/wag/gqrs-1.htm> (cit. on p. 28).
- Nickson, Chris. *Sedation in ICU*. URL: <https://lifeinthefastlane.com/ccc/sedation-in-icu/> (cit. on p. 63).
- Oppenheim, Alan V (1999). *Discrete-time signal processing*. Pearson Education India (cit. on p. 23).
- Pan, Jiapu and Willis J Tompkins (1985). “A real-time QRS detection algorithm.” In: *IEEE transactions on biomedical engineering* 3, pp. 230–236 (cit. on pp. 23, 26, 28).
- Physionet (2009). *A challenge from PhysioNet and Computers in Cardiology 2009*. URL: <https://physionet.org/challenge/2009/> (cit. on pp. xi, 5).
- Saeed, Mohammed, Mauricio Villarroel, Andrew T Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H Kyaw, Benjamin Moody, and Roger G Mark (2011). “Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database.” In: *Critical care medicine* 39.5, p. 952 (cit. on p. 20).
- Sheldon G. Sheps, M.D. *What is pulse pressure? How important is pulse pressure to your overall health?* URL: <https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/expert-answers/pulse-pressure/faq-20058189> (cit. on p. 31).
- Silva, Ikaro and George B Moody (2014). “An open-source toolbox for analysing and processing physionet databases in matlab and octave.” In: *Journal of open research software* 2.1 (cit. on p. 23).
- Sun, JX, AT Reisner, M Saeed, and RG Mark (2005). “Estimating cardiac output from arterial blood pressure waveforms: a critical evaluation using the MIMIC II database.” In: *Computers in Cardiology, 2005*. IEEE, pp. 295–298 (cit. on p. 32).
- Tarvainen, Mika P, Juha-Pekka Niskanen, Jukka A Lipponen, Perttu O Ranta-Aho, and Pasi A Karjalainen (2014). “Kubios HRV—heart rate variability analysis software.” In: *Computer methods and programs in biomedicine* 113.1, pp. 210–220 (cit. on pp. 46–48).
- Tibshirani, Robert (1996). “Regression shrinkage and selection via the lasso.” In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288 (cit. on p. 53).
- Warmuth, Manfred K, Jun Liao, and Gunnar Rätsch (2006). “Totally corrective boosting algorithms that maximize the margin.” In: *Proceedings of the 23rd international conference on Machine learning*. ACM, pp. 1001–1008 (cit. on p. 61).
- Yildiran, Tansel, Mevlut Koc, Abdi Bozkurt, Durmus Yildiray Sahin, Ilker Unal, and Esmeray Acarturk (2010). “Low pulse pressure as a

- predictor of death in patients with mild to advanced heart failure.”
In: *Texas Heart Institute Journal* 37.3, p. 284 (cit. on pp. 1, 32).
- Zong, W, T Heldt, GB Moody, and RG Mark (2003). “An open-source algorithm to detect onset of arterial blood pressure pulses.” In: *Computers in Cardiology, 2003*. IEEE, pp. 259–262 (cit. on p. 30).