Multi-sensor data fusion : Application to the prediction of ARDS

Aline Taoum

Thèse
de doctorat
de l'UTT

Aline TAOUM

Multi-sensor Data Fusion: Application to the Prediction of ARDS

Champ disciplinaire :
Sciences pour l'Ingénieur

2019TROY0006 **Année 2019**

Thèse en cotutelle avec l'Université Libanaise – TRIPOLI – LIBAN

Université Libanaise

utt
UNIVERSITÉ DE TECHNOLOGIE
TROYES

THESE

pour l'obtention du grade de

DOCTEUR
de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

EN SCIENCES POUR L'INGENIEUR

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Aline TAOUM

le 24 janvier 2019

Multi-sensor Data Fusion: Application to the Prediction of ARDS

JURY

M. M. SAYED MOUCHAWEH	PROFESSEUR - HDR	Président
M. J. BOUDY	PROFESSEUR - HDR	Rapporteur
Mme C. MAILHES	PROFESSEURE DES UNIVERSITES- HDR	Rapporteure
M. P. HONEINE	PROFESSEUR DES UNIVERSITES - HDR	Examinateur
M. H. AMOUD	PROFESSEUR	Directeur de thèse
Mme F. CHEHADE	MAITRE DE CONFERENCES - HDR	Directrice de thèse
M. Z. FAWAL	PROFESSEUR	Directeur de thèse

Acknowledgements

It would not have been possible to write this doctoral dissertation without the help and support of the kind people around me, only some of whom it is possible to mention here.

First, I would like to express my sincere gratitude to my supervisor in France, Farah CHEHADE, for her continuous support of my Ph.D study and related research as well as her patience, motivation, and immense knowledge. Her guidance helped me throughout the entire duration of the research and writing of this thesis.

I am also indebted to my supervisor in Lebanon, Hassan AMOUD, who not only encouraged me to pursue a doctoral degree but also provided endless support and encouragement throughout the course of this thesis. I am also grateful to Ziad FAWAL for his insightful comments and encouragement. I could not have imagined better advisors and mentors for my Ph.D study.

Next, I would like to thank the members of my defense committee for their time and dedication. I thank the reviewers, Corinne MAILHES and Jérôme BOUDY, who patiently read the manuscript and provided detailed comments and insightful suggestions. I also thank the examiners, Paul HONEINE and Moamar SAYED MOUCHAWEH. Their helpful feedback greatly improved the contents of this manuscript.

In my daily work, I have been blessed with the friendly and cheerful group of fellow students and members of both The Azm Center and UTT. I would like to thank all the members of the M2S Laboratory and Azm Center for their support and friendship. My gratitude extends to Bernadette ANDRE, Véronique BANSE and Jana EL HAJJ for their availability, logistic support, organizational talents and kind spirits. I also thank the UTT and UL doctoral schools. In particular, the directors Khemais SAANOUNI, Fawaz EL OMAR and Mohamad KHALIL and the secretaries Pascale DENIS, Isabelle LECLERCQ , Thérèse KAZARIAN and Zeinab IBRAHIM.

I am thankful to all my friends for their continued help and support. I thank not only all the new ones for the life experiences that we have shared

here in France but also those back home who have stayed present and full of support despite the distance.

I am thankful to my beloved parents and brothers for their ever present love and all kinds of support. Without the tremendous sacrifices they have made for me, I would not have had the chance to pursue this doctoral degree. I would like to thank Bacel for his love, support, tolerance and for always being there ready to help. My last thanks goes to the spirit of my dear sister, Rita.

Dedicated to

the memory of my sister, Rita.
You are gone but your belief in me has made this journey possible.

Abstract

Nowadays, the elderly population represents 13% of the world's population. It is estimated that this rate will increase very rapidly leading to the phenomenon of the population ageing. This phenomenon increases the incidence of various diseases, such as cardiopulmonary pathologies. In this thesis, we are interested in the pathology known as Acute Respiratory Distress Syndrome (ARDS). ARDS is a fatal lung condition occurring in critically ill patients representing a high mortality rate. The aim of this thesis is to develop methods for the prediction of ARDS in real-time using physiological signals, such as heart rate, respiratory rate, peripheral arterial oxygen saturation and mean arterial blood pressure. During the course of this thesis, we developed three original methods for the prediction of ARDS. The first method is based on the novelty detection on each physiological signal, followed by different techniques of decision fusion. The second method consists of the extraction of parameters from the signals, selecting the most relevant ones and fusing them using the theory of belief functions. Finally, we proposed a model that combines both kernel methods and belief function theory for the prediction of ARDS. We also proposed the extraction of new parameters to complete the list used in the previous method.

Keywords:

- Multisensor data fusion
- Dempster-Shafer theory
- Ridge regression (Statistics)
- Respiratory distress syndrome, Adult

Résumé

La population âgée représente actuellement 13% de la population mondiale. Ce taux est estimé d'augmenter très rapidement entrainant le phénomène de vieillissement de la population. Ce phénomène augmente la fréquence d'occurrence de diverses maladies, telles que les pathologies cardio-pulmonaires. Dans cette thèse, nous nous intéressons à la pathologie connue sous le nom de syndrome de détresse respiratoire aigüe (SDRA). Il s'agit d'une forme très sévère de défaillance pulmonaire aiguë, consécutive à une altération de la perméabilité capillaire. Le SDRA représente un taux de mortalité élevé. Cette thèse a pour objectif de développer des méthodes de prévention du SDRA en temps réel en utilisant des signaux physiologiques, telles que le rythme cardiaque, le rythme respiratoire, le taux d'oxygène dans le sang et la moyenne de la pression artérielle. Au cours de cette thèse, nous avons développé trois méthodes originales pour la prévention du SDRA. La première méthode est basée sur la détection d'anormalité dans chaque signal physiologique, suivie d'une fusion de décisions, selon différentes approches. La deuxième méthode consiste à extraire des caractéristiques des signaux, à sélectionner les paramètres les plus pertinents et à les fusionner à l'aide de la théorie des fonctions de croyance. Finalement, nous avons proposé un modèle combinant à la fois les méthodes à noyaux et la théorie des fonctions de croyance pour la prévention du SDRA. Nous avons également proposé l'extraction de nouveaux paramètres.

Mots-clés :

- Fusion multicapteurs
- Théorie de Dempster-Shafer
- Ridge régression (statistique)
- Syndrome de détresse respiratoire aiguë de l'adulte

Contents

Acknowledgements **i**

Abstract **v**

Résumé **vii**

1 Acute Respiratory Distress Syndrome **1**
1.1 The elderly population 2
1.2 The respiratory system 3
1.2.1 External anatomy 4
1.2.2 Bronchi 4
1.2.3 Bronchioles 5
1.2.4 Alveoli 5
1.2.5 Respiratory conditions 6
1.3 Acute Respiratory Distress Syndrome 7
1.3.1 Pathophysiology 8
1.3.2 Epidemiology 10
1.3.3 Diagnosis and monitoring 11
1.3.4 Risk factors of ARDS 13
1.3.5 Different studies on ARDS 14
1.4 Problem statement 16
1.4.1 Objective 16
1.4.2 Medical devices and sensors 17
1.4.3 Vital Signs 18
1.4.3.1 Heart rate 18
1.4.3.2 Respiratory rate 19
1.4.3.3 Peripheral oxygen saturation 19
1.4.3.4 Mean arterial blood pressure 20
1.5 MIMIC II database 20
1.5.1 Description of the database 20
1.5.2 Subjects Selection 21
1.5.2.1 Invasive ventilated subjects 22

1.5.2.2 Non-Invasive ventilated subjects 23
1.6 Organization of the manuscript 25
1.7 Publications 26

2 Information Fusion Techniques **29**
2.1 Introduction 30
2.2 Information fusion 30
2.3 Decision fusion techniques 31
2.3.1 Decision fusion procedure 31
2.3.2 Majority Voting 33
2.3.3 Weighted Majority voting 34
2.4 Feature fusion techniques 36
2.4.1 Probabilistic fusion 38
2.4.2 Belief functions theory 39
2.4.2.1 Basic concept 40
2.4.2.2 Discounting 40
2.4.2.3 Combination 41
2.4.2.4 Decision 42
2.5 Machine learning and kernel methods 43
2.6 Conclusion 45

3 ARDS Prediction using Novelty Detection **47**
3.1 Introduction 48
3.2 Description of the approach 49
3.3 Novelty Detection Algorithm 51
3.4 Decision fusion 55
3.4.1 Linear fusion based on performance indexes 56
3.4.2 Linear fusion based on error rates 57
3.4.3 Kernel Ridge Regression-based fusion 58
3.5 Parameters estimation 60
3.5.1 Optimization of D and τ 60
3.5.2 Optimization of kernel parameters 61
3.6 Missing signals 61
3.7 Results 62
3.7.1 Materials 62
3.7.2 Invasive ventilated subjects 64
3.7.2.1 Results of Definition 1 64
3.7.2.2 Results of the D distance-based approach . . 65
3.7.2.3 Missing signals 67

3.7.2.4 Early predictions . . . 68
3.7.3 Non-invasive ventilated subjects . . . 69
3.7.3.1 Results of the proposed approach . . . 70
3.7.3.2 Missing signals . . . 70
3.7.3.3 Early predictions . . . 70
3.7.4 Comparison to state-of-the-art methods . . . 71
3.8 Discussions . . . 73
3.9 Conclusion . . . 75

4 ARDS Prediction using the Belief Functions Theory 77
4.1 Introduction . . . 78
4.2 Evidence-based approach for ARDS prediction . . . 79
4.2.1 Description of the approach . . . 79
4.2.2 Parameters Extraction . . . 81
4.2.3 ARDS prediction with the belief functions theory . . . 84
4.2.3.1 Basic concept . . . 85
4.2.3.2 Discounting . . . 87
4.2.3.3 Combination . . . 88
4.2.3.4 Decision making . . . 89
4.3 Evidence-based approach for four-group prediction . . . 89
4.3.1 Identification of groups . . . 90
4.3.2 The belief functions framework . . . 91
4.4 Selection of parameters . . . 93
4.5 Results . . . 94
4.5.1 Statistical analysis . . . 95
4.5.2 Parameters Selection . . . 97
4.5.3 Selection of the alert generation threshold . . . 99
4.5.4 Selection of window size . . . 99
4.5.5 Performance Evaluation . . . 100
4.5.6 Early predictions . . . 102
4.5.7 Comparison to the state-of-the-art methods . . . 102
4.5.8 Multi-class classification . . . 103
4.6 Discussions . . . 105
4.7 Conclusion . . . 107

5 ARDS Prediction using Belief Functions and Kernel Methods 109
5.1 Introduction . . . 110
5.2 Literature review on masses definition . . . 110
5.3 Evidence-based approach using kernel methods . . . 112

5.3.1 Description of the approach . . . 112
5.3.2 The recurrence quantification analysis . . . 113
5.3.3 Belief functions approach using kernels . . . 115
5.3.3.1 Mass assignment by kernel methods . . . 115
5.3.3.2 Kernel-based regression methods . . . 117
5.3.3.3 Discounting . . . 121
5.3.3.4 Combination . . . 122
5.3.3.5 Decision making . . . 122
5.3.4 Selection of parameters . . . 122
5.4 Results . . . 123
5.4.1 Statistical analysis . . . 124
5.4.2 Optimization of the KRR parameters . . . 124
5.4.3 Selection of parameters . . . 125
5.4.4 Selection of the alert generation threshold . . . 126
5.4.5 Performance evaluation . . . 127
5.4.6 Early predictions . . . 128
5.4.7 Comparison with previous models . . . 129
5.5 Discussion . . . 130
5.6 Conclusion . . . 131

6 Conclusion and Perspectives **133**
6.1 Summary of Contributions . . . 134
6.2 Perspectives . . . 136

Appendix A Phase Space and Recurrence Quantification Analysis **139**
A.1 Reconstructed phase space . . . 140
A.1.1 Estimation of the time delay . . . 141
A.1.2 Choosing the embedding dimension . . . 141
A.2 Recurrence quantification analysis . . . 142
A.2.1 Recurrence plot . . . 142
A.2.2 Quantitative analysis of the RP . . . 143

Appendix B Résumé en Francais **147**
B.1 Le syndrome de détresse respiratoire aigue . . . 149
B.1.1 Introduction . . . 149
B.1.2 Le système respiratoire . . . 149
B.1.3 Le syndrome de détresse respiratoire aigüe . . . 150
B.1.4 Objectif . . . 152
B.1.5 La base de données MIMIC II . . . 152

B.1.6 Contributions 153
B.2 La prévention du SDRA par la détection d'anomalies 154
B.2.1 Introduction 154
B.2.2 Approche proposée 155
B.2.2.1 Détection d'anomalies 156
B.2.2.2 Fusion des décisions individuelles 157
B.2.2.3 Optimisation des paramètres 160
B.2.3 Résultats 160
B.2.3.1 Sujets ventilés invasifs 161
B.2.3.2 Sujets ventilés non-invasifs 161
B.2.3.3 Comparaison à d'autres méthodes de classification 162
B.2.4 Conclusion 162
B.3 La prévention du SDRA par les fonctions de croyance 162
B.3.1 Introduction 162
B.3.2 Prévention binaire 163
B.3.2.1 Extraction de paramètres 163
B.3.2.2 Fonctions de croyance 164
B.3.3 Classification multi-classe 166
B.3.4 Identification des groupes 166
B.3.5 Modèle des fonctions de croyance 167
B.3.6 Résultats 168
B.3.6.1 Exemple sur un ensemble d'apprentissage . . 168
B.3.6.2 Evaluation des performances 169
B.3.6.3 Comparaison à d'autres méthodes 169
B.3.6.4 Classification multi-classes 170
B.3.7 Conclusion 171
B.4 La prévention du SDRA par les croyances et les noyaux 171
B.4.1 Introduction 171
B.4.2 Approche proposée 172
B.4.2.1 Analyse de quantification de récurrence . . . 172
B.4.2.2 La théorie des fonctions de croyance 173
B.4.2.3 Sélection des paramètres 177
B.4.3 Résultats 177
B.4.3.1 Exemple sur un ensemble d'apprentissage . . 178
B.4.3.2 Evaluation des performances 179
B.4.4 Conclusion 179
B.5 Conclusion et perspectives 179

B.5.1 Contributions principales 180
B.5.2 Perspectives 181

Bibliography **183**

List of Figures

1.1 The growth of the worldwide elderly population in millions from 1990 to 2017 and projections for the coming 80 years . . . 3
1.2 The respiratory system . . . 4
1.3 The 3 phases in the pathophysiology of ARDS, (A) a normal alveolus, (B) the exudative, (C) the proliferative and (D) the fibrotic phases (Mac Sweeney and McAuley, 2016) . . . 9
1.4 Clinical modalities for the diagnosis and monitoring of the ARDS . . . 12
1.5 Some wearable devices that monitor physiological signals . . . 17
1.6 Electrocardiograph waves and RR interval . . . 19
1.7 MIMIC II Clinical database subjects . . . 22
1.8 Selection for the invasive ventilated subjects . . . 23
1.9 Selection for the non-invasive ventilated subjects . . . 24

2.1 Block diagram of decision fusion procedure . . . 32
2.2 The block diagram of weighted fusion rule . . . 34
2.3 Feature fusion strategies . . . 37

3.1 The inputs and output of the proposed prediction model . . . 49
3.2 The general block diagram of the real-time prediction model for a subject s using novelty detection . . . 50
3.3 Decision fusion model . . . 56
3.4 An example of the four signals for an ARDS subject . . . 63
3.5 An example of the four signals for a non-ARDS subject . . . 63
3.6 ROC curves for Novelty algorithm followed by data fusions based on performances in (a), and both Simple and WMV error rates functions in (b) and (c), respectively . . . 67
3.7 A comparison of the performances, presented by Youden index, between the proposed fusion functions for the different combinations of signals. In this graph, the following simplifications are done 1: HR, 2: RR, 3: SpO_2, 4: MABP and 4D is the combination of all the signals . . . 68

3.8 Early detection results for the combination of HR-RR-SpO_2-MABP using KRR fusion function. This scatter plots time of ARDS prediction versus recordings length (hours) 69
3.9 A comparison of the performances, presented by Youden index, between the proposed fusion functions for the different combinations of signals. In this graph, the following simplifications are done 1: HR, 2: RR, 3: SpO_2, 4: MABP and 4D is the combination of all the signals 71
3.10 Early prediction results for the combination of HR-RR-SpO_2-MABP using KRR fusion function for the non-invasive ventilated subjects. This scatter plots time of ARDS prediction versus recordings length (hours) 72

4.1 Block diagram of the proposed evidence-based model for ARDS prediction 80
4.2 Mass assignment of new computations in the case of the normalized standard deviation of respiratory rate signals 87
4.3 An example for computing the error rates 88
4.4 An example of identification of mild, moderate and severe phases of ARDS in heart rate signal using the PaO_2/FiO_2 signal. 92
4.5 Boxplots for the relevant parameters of HR signal 96
4.6 Boxplots for the relevant parameters of RR signal 96
4.7 Boxplots for the relevant parameters of SpO_2 signal 96
4.8 Boxplots for the relevant parameters of MABP signal 96
4.9 Ranking of parameters according to error rates. Only parameters having an error lower than 0.4 are included in the figure . 98
4.10 The parameters selection procedure using the simple procedure and the correlation extension procedure 98
4.11 ROC curve for the selection of the threshold for alert generating 99
4.12 Changes in the accuracy over the training set with windows length. Values in the figure denote (number of parameters, threshold) 100
4.13 Early-predictions of ARDS subjects using the proposed feature fusion approach 102

5.1 Block diagram of the proposed approach 113
5.2 An example of the selection procedure of parameters from the training set 126
5.3 ROC curve for the selection of the threshold for alert generating 127

5.4 Early-predictions of ARDS subjects using the proposed model 129

A.1 An example of a reconstructed phase space (a) and its recurrence plot (b) (Barbera, La Rocca, and Rizzi, 2011). 143

List of Tables

1.1 Main lung diseases grouped by the level of infection and their most common cause 7
1.2 Definitions of Acute Respiratory Distress Syndrome using American-European Consensus Conference (AECC) and the Berlin definition 8
1.3 Clinical disorders associated with the development of ARDS . 13
1.4 The definition of each used medical term 15

2.1 Different weights definitions for weighted fusion rules from the literature 35

3.1 Examples of distance functions between two vectors x and y . 52
3.2 Matrix of correlation coefficients between signals 64
3.3 Performance of the Definition 1 65
3.4 Performances of the optimal τ_o in (hours) and D_o obtained for the proposed model with each decision fusion technique . . . 66
3.5 Youden indexes over the training and testing sets for the proposed model with each decision fusion technique 66
3.6 Results on non-invasive ventilated subjects 70
3.7 Performances in % obtained with our approach compared to previous novelty detection technique and state-of-the art methods 72

4.1 The most significant parameters for each signal, parameters having a p-value < 0.05 are shown. Median $\pm$ Median absolute deviation are shown for each parameter 95
4.2 Performance of the training and testing on the example considered previously 101
4.3 Performance of the proposed model and the influence of each phase 101
4.4 Comparison to state-of-the-art classification methods 103
4.5 Accuracies of the multi-class classification for the training phase 104
4.6 Accuracies of the multi-class classification for the testing phase 105

4.7 Confusion matrix for basic discounting 105
4.8 Confusion matrix for contextual discounting 105

5.1 Statistical analysis for RQA parameters 124
5.2 Parameters having lowest error rates after optimization of η and σ 125
5.3 Performance of the training / testing set of the considered example 128
5.4 Results of the KRR-based belief functions model on the training and the test sets 128
5.5 Comparison to previous models 130

6.1 A comparison between the performances of the proposed models 136

A.1 RQA parameters (Marwan et al., 2007) 144

Chapter 1

Acute Respiratory Distress Syndrome

Contents

1.1 The elderly population 2

1.2 The respiratory system 3

1.2.1 External anatomy 4

1.2.2 Bronchi 4

1.2.3 Bronchioles 5

1.2.4 Alveoli 5

1.2.5 Respiratory conditions 6

1.3 Acute Respiratory Distress Syndrome 7

1.3.1 Pathophysiology 8

1.3.2 Epidemiology 10

1.3.3 Diagnosis and monitoring 11

1.3.4 Risk factors of ARDS 13

1.3.5 Different studies on ARDS 14

1.4 Problem statement 16

1.4.1 Objective 16

1.4.2 Medical devices and sensors 17

1.4.3 Vital Signs 18

1.5 MIMIC II database 20

1.5.1 Description of the database 20

1.5.2 Subjects Selection 21

1.6 Organization of the manuscript 25

1.7 Publications 26

1.1 The elderly population

According to the UN's 2017 Revision, the world's population is numbered approximately 7.6 billion in the mid of 2017 (United Nations, Department of Economic and Social Affairs, Population Division, 2017). More than half of the population are adults between 15 and 59 years old with a percentage of 61%. The remaining 39% of the population are divided between children under 15 years old (26 %) and older persons aged 60 or over representing 13 % of the world's inhabitants. However, the elderly population is growing faster than all younger age groups due to fertility decline and high life expectancy (Mathers et al., 2015). This phenomenon, known as population ageing, occurs throughout the world. According to the UN's Revision document, 25 % of Europe's population are aged 60 or over, which is the highest percentage of the elderly population among all the regions of the world. Figure 1.1 illustrates the population ageing in the world from 1990 until 2017 and the expectations for the coming years according to (Prince et al., 2007; Organization and International Tobacco Control, 2008; United Nations, Department of Economic and Social Affairs, Population Division, 2017). The worldwide elderly population is expected to rise from 962 million in 2017 to more than 3 billion by 2100 and exceed the number of children.

Population ageing is estimated to have a profound effect on the socio-economic structure of society in the 21st Century, especially in terms of health and social care (Institute for Health Metrics and Evaluation (IHME), 2018). Besides this, health care services become more expensive because of the increasing price of hospital care, medical instruments and drugs prescription (Patton, 2015). This phenomenon is strongly linked to the increase in the worldwide epidemic of chronic disease. This population ageing trend has increased the prevalence of different diseases, the presence of multiple pathologies together and care-demanding conditions. There are innumerable pathologies confronting the elderly as getting older. Nowadays, it is being more common that old people require continuous care in their last years. The majority of the elderly become dependent after severe conditions for the remaining years of their life.

The leading contributors to disease burden in the elderly are cardiovascular diseases (30%), malignant neoplasms (15%), respiratory diseases (10%)

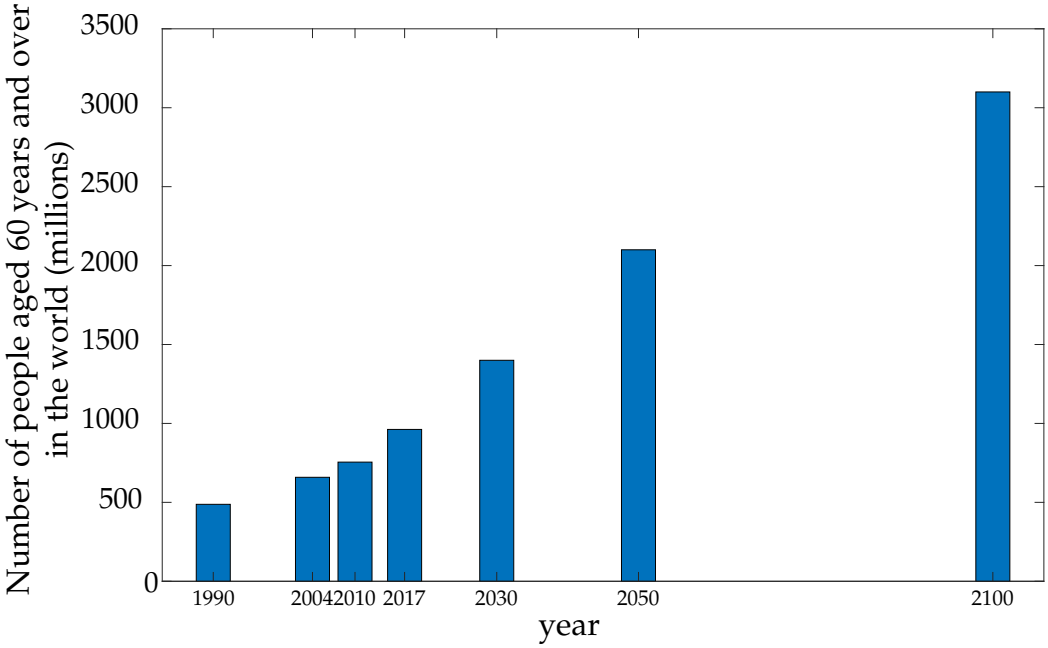


FIGURE 1.1: The growth of the worldwide elderly population in millions from 1990 to 2017 and projections for the coming 80 years

(Prince et al., 2015). Secondary contributors to disease burden are musculoskeletal diseases, neurological and mental disorders, gastrointestinal system diseases and diabetes mellitus. Respiratory diseases are widespread in the population especially with the increasing levels of pollution, alcohol abuse and other infections.

1.2 The respiratory system

The respiratory system provides the vital delivery of atmospheric air and the exchange of gases between the body and atmospheric air. The human body requires a constant supply of oxygen from the atmosphere for the cellular growth and metabolism that maintain the normal function of the body. At the same time, carbon dioxide that forms as a waste product of cellular metabolism is carried out from the blood into the lung to be exhaled out of the body before reaching toxic levels. The respiratory system consists of specific respiratory organs, blood vessels and muscles. The entire respiratory tract consists of the nose, trachea, bronchi, bronchioles and alveoli as shown in Figure 1.2 and described in the following sections.

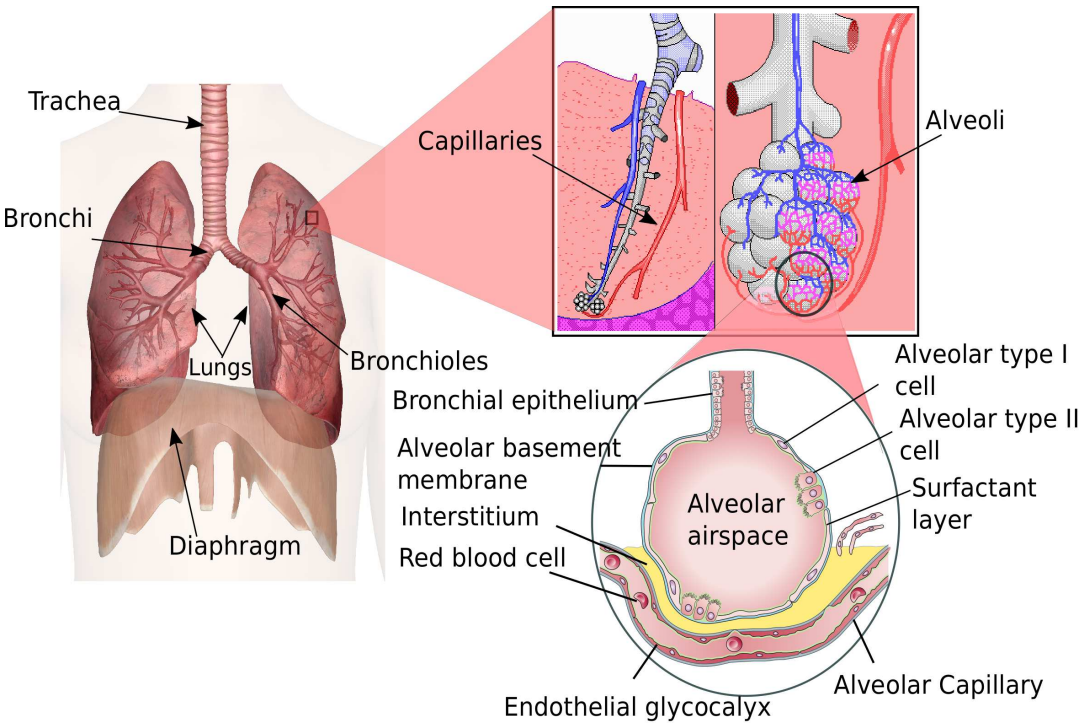


FIGURE 1.2: The respiratory system

1.2.1 External anatomy

The lungs are a pair of spongy organs lying in the chest cavity formed by the ribs. The lungs are covered by a thin tissue layer called the *pleura*. Each lung has a cone shape where the superior end of the lungs narrows to a rounded tip known as the *apex* and the inferior end rests on the diaphragm and known as the *base*. The *base* is concave to follow the shape of the diaphragm.

Each lung consists of several lobes. The right lung has 3 lobes, the superior, middle and inferior lobes; whereas the left lung has only two lobes - superior and inferior - since it is smaller than the right lung because 2/3 of the heart is located on the left side of the body.

1.2.2 Bronchi

When breathing, atmospheric air, of which 21 % is oxygen, enters the body through the mouth and the nose and arrives in the chest through the trachea or the windpipe. The trachea is about 2.6 cm in diameter and 11 cm long. It is guarded by 16-20 C-shaped cartilage rings that provide a flexible rigidity and prevent the collapsing of the trachea. At its inferior end the trachea divides into two narrower tubes called bronchi that each enters one

of the lungs.

The bronchi are made of hyaline cartilage and lined with ciliated pseudostratified epithelium. As the primary bronchi enter the lungs, they subdivide into smaller lobar bronchi that carry air to each lobe of the lung. Thus, the right main bronchus subdivides into 3 lobar bronchi while the left main bronchus branches off into 2 lobar bronchi. The lobar bronchi, also called the secondary bronchi, divide into many smaller tertiary bronchi.

1.2.3 Bronchioles

Each tertiary bronchus divides and branches into many thinner tubes called bronchioles. Bronchioles differ from bronchi both in size and in the composition of their walls. Bronchioles consist of elastin fibers and smooth muscle tissue that allow the diameter of bronchioles to change. The bronchioles dilate to allow greater airflow when the lungs require greater volumes of air, such as during exercise. They can constrict when the inhaled air contains dust or other environmental pollutants.

These air passages, known as the bronchial tree, are divided further into finer terminal bronchioles that have the same composition as bronchioles. Terminal bronchioles are the smallest air tubes in the lungs that conduct the air to the alveoli of the lungs.

1.2.4 Alveoli

Alveoli are the functional units of the lungs where gas is exchanged between the air in the lungs and the blood in the capillaries of the lungs. There are approximately 700 million alveoli in the lungs covering a total surface of about 70 m^2. Alveoli are the terminal ends of the respiratory tract, found in small grape-shaped clusters called alveolar sacs. Each alveolus is an anatomical structure that has the form of hollow cavity and is surrounded by many tiny capillaries.

The walls of the alveolus are lined with a thin epithelial layer that consists of two major types of alveolar cells. Type I cells form the structure of alveolar walls, known as squamous alveolar. They are thin and permeable in order to facilitate gas exchange with capillaries. Type II cells or the great alveolar are more common than type I cells. They maintain the elastic recoil

of the lung since these are the site of surfactant production in the lungs. The surfactant layer is a thin layer of connective tissue that underlies and supports the alveolar cells. The respiratory membrane is formed where the walls of a capillary touch the walls of an alveolus. It is a fluid space, named interstitium, made of collagen and elastin and contains no cells. At the interstitium, gases dissolve in water and diffuse freely in the capillaries.

Moreover, there are many septal cells and macrophages inside the alveoli that maintain the elasticity of the lungs, prevent the collapsing of alveolar walls and provide immune system defense of the alveoli from pathogens and other foreign materials that enter the alveoli along with inhaled air.

1.2.5 Respiratory conditions

There are many conditions or infections that can affect the respiratory system and cause a serious disease. Being the main organs of the respiratory system, the lungs are also the most critical organs. Lung illness is defined by the pulmonary response to direct injuries affecting the lungs or an indirect injury to other parts of the body. Respiratory and lung diseases can be grouped according to the level of infection inside the lungs. Table 1.1 presents the main lung illnesses grouped by units of infection and their common cause. As shown, the causes of lung illnesses are very divergent and can be a direct injury or an indirect condition. The most serious diseases are chronic obstructive pulmonary disease (COPD) that mainly happens in the bronchi, pneumonia, lung cancer and acute respiratory distress syndrome that affect the alveoli.

Acute respiratory distress syndrome (ARDS) represents the most severe condition caused commonly by a serious illness. It is a fateful lung condition occurring in critically ill patients (Ware and Matthay, 2000). ARDS is characterized by widespread inflammatory changes throughout the lung, usually accompanied by aggressive fibrosis. It results in pulmonary edema and difficulties of gas exchange with the blood. ARDS shows high incidence in population over the world. Moreover, it is associated with high mortality rates of 30% to 68% for elderly patients (Rubenfeld et al., 2005).

TABLE 1.1: Main lung diseases grouped by the level of infection and their most common cause

Level	Name	Most common cause
Pleura	Pleural effusion	Heart failure
	Pleurisy	Virus
Interstitium	Pulmonary fibrosis	Smoking
	Sarcoidosis	Inflammation
Bronchi	Asthma	Allergies
	Chronic obstructive pulmonary disease	Smoking
	Acute bronchitis	Virus
	Bronchiectasis	Infections
Alveoli	Pneumonia	Bacteria
	Lung cancer	Smoking
	Tuberculosis	Bacteria
	Acute respiratory distress syndrome	Serious illness
	Pneumoconiosis	Inhalation of substance that injures the lung

1.3 Acute Respiratory Distress Syndrome

Acute Respiratory Distress Syndrome (ARDS) is a sudden failure of the respiratory system due to alveolar injury secondary to an inflammatory process, which can be either pulmonary or systemic in origin. The term "respiratory distress syndrome" was introduced by (Ashbaugh et al., 1967) to describe the acute onset of tachypnea, hypoxemia and loss of lung compliance. Then, a formal definition of ARDS was developed during the American-European Consensus Conference (AECC) (Bernard et al., 1994). The AECC defined ARDS as the acute onset of respiratory failure, bilateral infiltrates on frontal chest radiograph, hypoxemia as defined by a ratio of the partial pressure of oxygen to the fraction of inspired oxygen PaO_2/FiO_2 $<$200 mmHg and no evidence of left atrial hypertension. The acute lung injury (ALI) was defined as well being a less severe form of ARDS.

Although this definition has advanced the knowledge of ARDS, its validity has been criticized for the lack of explicit criteria for defining ARDS. Therefore, the Berlin Definition of ARDS was published in 2012 to address the limitations of the AECC definition (Ranieri et al., 2012). The major changes in the Berlin definition are shown in Table 1.2 and include: (1) the

TABLE 1.2: Definitions of Acute Respiratory Distress Syndrome using American-European Consensus Conference (AECC) and the Berlin definition

Characteristic	**The AECC definition**	**The Berlin definition**
Onset	Acute	New or worsening within one week
Infiltrates on chest radiograph	Bilateral Infiltrates on frontal chest radiograph	Bilateral opacities radiograph
Oxygenation (mmHg)	Acute lung injury for $PaO_2/FiO_2 \leq 300$ Acute respiratory distress syndrome for $PaO_2/FiO_2 \leq 200$	Mild ARDS for $200 < PaO_2/FiO_2 \leq 300$ Moderate ARDS for $100 < PaO_2/FiO_2 \leq 200$ Severe ARDS for $PaO_2/FiO_2 \leq 100$
PEEP (cmH_2O)	Not specified	Minimum PEEP of 5 required
Heart failure	Pulmonary artery wedge pressure ≤ 17 mmHg Absence of left atrial hypertension	Left ventricular failure insufficient to solely account for clinical state

term acute lung injury was eliminated and replaced by 3 levels of ARDS based on PaO_2/FiO_2 values with at least 5 cmH_2O of applied positive-end expiratory pressure (PEEP), (2) ARDS develops within 7 days instead of being acute and (3) the pulmonary wedge pressure was eliminated. Since then, this definition has been adopted in all studies on ARDS, and there hasn't been any formal definition proposed afterwards.

1.3.1 Pathophysiology

ARDS results from inflammatory alveolar injury occurring progressively through three overlapping phases. Figure 1.3 illustrates an alveolus during all phases, starting from a normal phase (A), an inflammatory or exudative phase (B), a proliferative phase (C) and ending by a fibrotic phase (D) (Bellingan, 2002; Thille et al., 2013).

The exudative phase: The process begins with the exudative phase characterized by the destruction of the barriers of the epithelial, interstitial and endothelial membranes of the alveoli. This phase is typified by the flooding of plasma, plasma proteins and cellular content to the interstitium and the

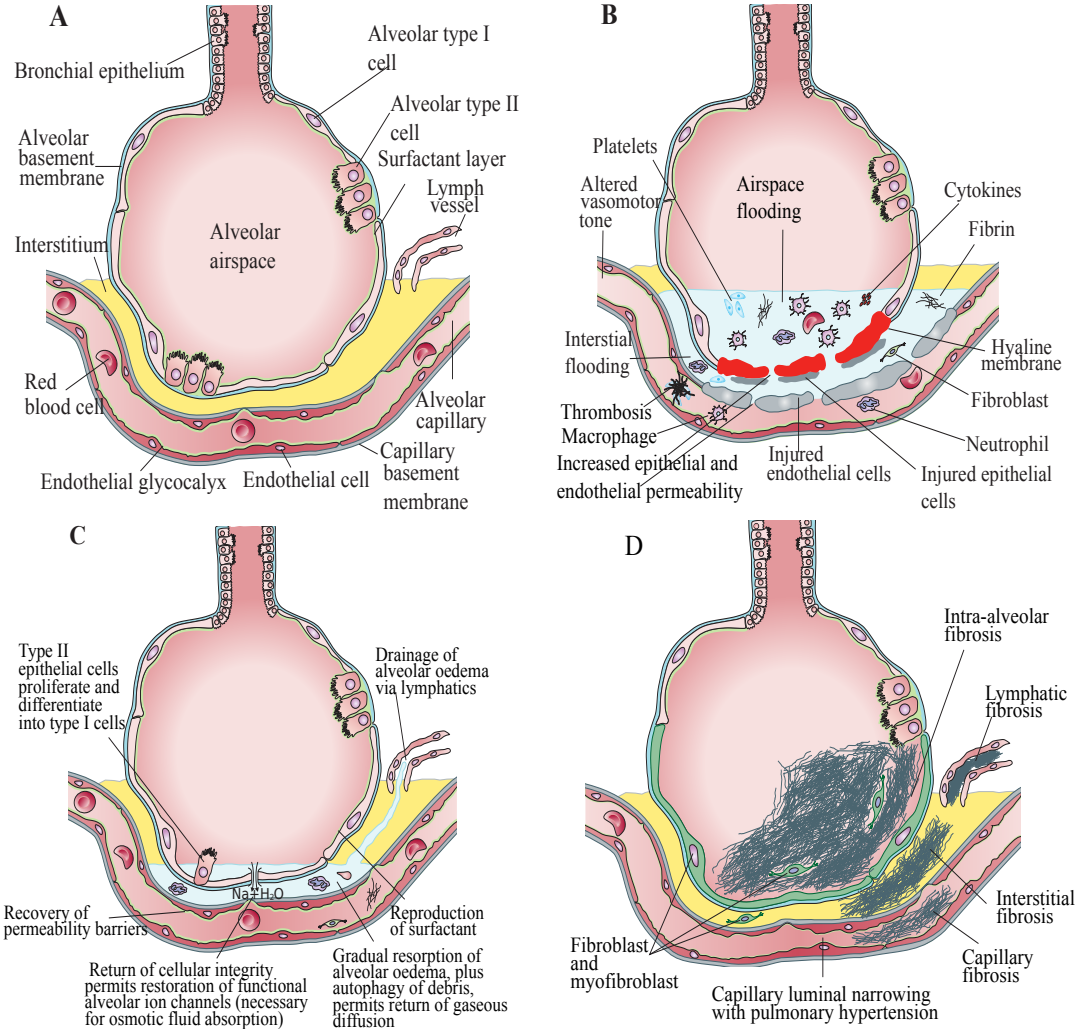


FIGURE 1.3: The 3 phases in the pathophysiology of ARDS, (A) a normal alveolus, (B) the exudative, (C) the proliferative and (D) the fibrotic phases (Mac Sweeney and McAuley, 2016)

airspace and the production of hyaline membranes as shown in Figure 1.3-B. The inflammatory mixture in the alveoli and interstitium reduces surfactant production and inactivate remaining surfactant, thereby increases alveolar permeability (Wheeler and Bernard, 2007). Neutrophils, macrophages and other cells from innate and adaptive immune system are found increasingly during the initial phases in capillaries, interstitium and progressively within airspaces (Weiland et al., 1986).

This process affects dependent portions of the lung, leading to insufficient aeration and atelectasis. The atelectasis is a complete or partial collapse of lungs that decreases lung compliance. The alterations in alveolar spaces result in microvascular occlusion, which leads to reductions in pulmonary arterial blood flow to ventilated portions of the lung, increasing the dead

space, and to pulmonary hypertension. Consequently, hypoxemia develops and the work of breathing rises, leading to dyspnea.

The proliferative phase: The exudative phase is followed by the proliferative phase, which develops 2-7 days after the initiation of lung injury and marks attempts of lung repair. This phase is characterized by the proliferation of type II alveolar cells along alveolar basement membranes. These alveolar cells synthesize new pulmonary surfactant and differentiate subsequently into type I alveolar cells. The regeneration of the epithelial membrane helps to clear the exudative fluid into the interstitium and eliminate the remaining debris. Vasomotor tone begins to return to normal, pulmonary hypertension lessens, oxygenation improves and pulmonary compliance slowly recovers.

The fibrotic phase: In some individuals, the proliferative phase progresses to a fibrotic phase that is associated with a failure of alveolar collagen removal, limited functional recovery, diminished compliance of the lung and increased pulmonary dead space. In this phase, alveolar edema and inflammatory exudates are converted to extensive alveolar duct and interstitial fibrosis. The fibroproliferation leads to progressive microvascular occlusion and pulmonary hypertension. The diffuse alveolar damage was found to be a pathognomonic sign of ARDS (Mac Sweeney and McAuley, 2016).

1.3.2 Epidemiology

The frequency of ARDS varies depending on the definition of ARDS being used and where the survey is undertaken. The population-based incidence of ARDS is estimated to 34 patients per 100 000 per year in USA (Rubenfeld et al., 2005), approximately 5-7 in Europe (Sigurdsson et al., 2013) and 15.19 in Taiwan (Chen et al., 2015). The rates of moderate or severe ARDS represent 75% of ARDS patients, while 25% of patients develop mild ARDS and one-third of them will later progress from the mild ARDS to moderate or severe disease (Rubenfeld et al., 2005; Ranieri et al., 2012). Hospital-based incidence of moderate to severe ARDS is estimated between 1.6% and 7% for intensive care unit (ICU) patients and from 8% to 19.7% for mechanically ventilated patients (Luhr et al., 1999; Brun-Buisson et al., 2004; Caser et al., 2014). There is a lack in epidemiologic studies of ARDS in low-income countries due to resource limitations (Riviello et al., 2016).

ARDS presents a mortality rate varying between 40% and 60% of patients, thus it emphasized the global burden (Singh et al., 2014; Chen et al., 2015; Bellani et al., 2016). Epidemiological studies conducted in Europe, Asia, the United States and Australia on ARDS incidence and outcomes have shown that the mortality rate increases with age from approximately 30% for young patients (age 15-19 years) to 68% for elderly patients (>85 years) (Arroliga et al., 2002; Goss et al., 2003; Hughes et al., 2003; Chen et al., 2015). Furthermore, randomized control trials have shown that 20 to 40% of patients die in 28 days after ARDS and 15-20% of patients are at high risk of death in the 12 months after first recovery (Mac Sweeney and McAuley, 2016).

Nevertheless, an important proportion of patients have a slow recovery and lungs function may return to normal over 6-12 months, but a full recovery is often limited. This is due to comorbidities rather than residual effects of ARDS. Neuromuscular weakness and neuropsychiatric problems, including cognitive impairment, anxiety, depression and post-traumatic stress disorder, are now more frequent in ARDS survivors and often delay return to normal life routine by months or permanently in some cases (Herridge et al., 2011; Fan et al., 2014).

1.3.3 Diagnosis and monitoring

The diagnosis of ARDS is clinically challenging because there is no specific clinical indicators for this disease. Various modalities are investigated for diagnosing and monitoring the clinical conditions. Figure 1.4 illustrates the different types of modalities for the diagnosis of ARDS. All patients developing ARDS require mechanical ventilation support. Thus, ventilatory parameters are shown to have a correlation with the evolution of ARDS. Pulmonary artery wedge pressure is present in the definition of the AECC and PEEP levels are required in the Berlin definition of ARDS (Bernard et al., 1994; Ranieri et al., 2012). Respiratory system compliance helps also with the monitoring of pulmonary mechanics (Mac Sweeney and McAuley, 2016).

Imaging techniques are used to differentiate ARDS from other similar conditions such as cardiogenic pulmonary edema, acute eosinophilic pneumonia, acute interstitial pneumonitis and diffuse alveolar hemorrhage.

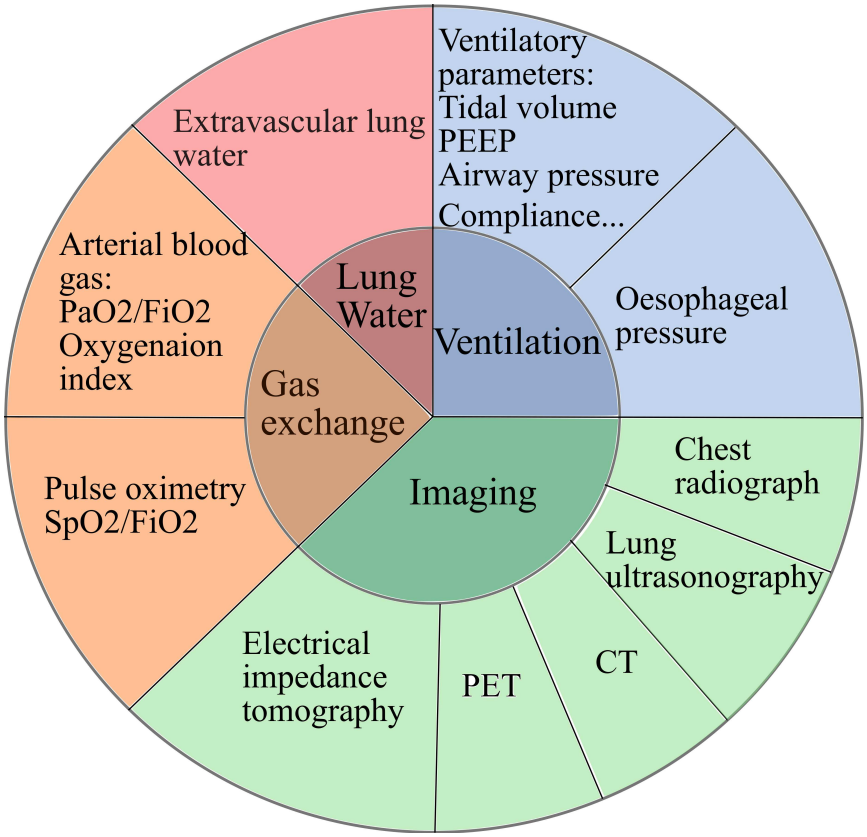


FIGURE 1.4: Clinical modalities for the diagnosis and monitoring of the ARDS

Chest radiography and computed tomographic (CT) scan provide qualitative measures of ARDS evolution; CT also provides specific quantitative measures of oedema, aeration and lung recruitability. Lung ultrasonography can be used similarly to estimate extravascular lung water (Shyamsundar et al., 2013; Enghard et al., 2015) and to differentiate ARDS from other conditions (Copetti, Soldati, and Copetti, 2008). Positron emission tomography (PET) has been increasingly used to study the pathophysiology of ARDS. It quantifies regional perfusion, aeration, lung vascular permeability, edema, metabolic activity of inflammatory cells, enzyme activity and pulmonary gene expression (Bellani et al., 2011). Electrical impedance tomography (EIT) is used to monitor regional lung ventilation and lung heterogeneities (Pesenti et al., 2016).

According to both definitions for ARDS, this latter is diagnosed mainly using the ratio of the partial pressure of oxygen (PaO_2) to the fraction of inspired oxygen (FiO_2) and classified to mild, moderate and severe conditions. The oxygenation index defined by the product of the mean airway pressure and FiO_2 divided by PaO_2 is an alternative to the previous ratio because it

TABLE 1.3: Clinical disorders associated with the development of ARDS

Direct lung injury	Indirect lung injury
Pneumonia	Sepsis
Gastric content aspiration	Pancreatitis
Severe thoracic trauma	Severe non-thoracic trauma
Near drowning	Cardiopulmonary bypass
Pulmonary contusion	Massive transfusions
Fat embolism	Drug overdose
Smoke and toxic gas inhalation	
Reperfusion pulmonary edema after lung transplantation	

includes the PEEP values with the mean airway pressure (Seeley et al., 2008). Extravascular lung water can be also monitored during ARDS to measure the degree of pulmonary oedema (Brown et al., 2013).

Different biomarkers have been investigated in the diagnosis of ARDS, but it was difficult to find a specific biomarker that reflects the responsiveness to therapy and prognosis of ARDS (Cardinal-Fernández et al., 2015). Open lung biopsy remains the gold standard for diagnosis of diffuse alveolar damage. It is usually reserved for exceptional cases in which there is a genuine diagnostic dilemma and poor response to therapy (Walkey et al., 2012).

1.3.4 Risk factors of ARDS

Several risk factors are associated with the development of ARDS (Ware and Matthay, 2000). The triggering events are grouped into two classes: direct and indirect lung injury as shown in Table 1.3. These risk factors are also grouped into primary and secondary risk factors. The sepsis syndrome appears to be the most common, but the overall risk increases with multiple factors (Fein et al., 1983). Another primary risk factors include pneumonia, trauma, aspiration of gastric contents, pancreatitis and near drowning (Pepe et al., 1982; Singh et al., 2014; Boverman and Genc, 2014). Moreover, several prospective studies have identified secondary risk factors for the onset of ARDS, like history of alcohol abuse (Moss et al., 1996), blood transfusion (Gong et al., 2005), high tidal volume (Gajic et al., 2005; Xiaoming et al., 2008), tachypnea (Iscimen et al., 2008), smoking (Thakur et al., 2009), obesity (Gong et al., 2010) and diabetes (Moss et al., 2000; Yu et al., 2013).

1.3.5 Different studies on ARDS

In addition to the aforementioned studies on the epidemiology and risk factors of ARDS, different studies were conducted to analyze the response to alveolar recruitment manoeuver. Heart rate variability and cardiovascular response were analyzed in ARDS patients during recruitment maneuver (Nielsen et al., 2005; Borghi-Silva et al., 2016). The recruitment maneuver helps to increase the transpulmonary pressure and reopen the alveolar units that are either poorly aerated or not aerated at all. There are different techniques described in the literature to apply the recruitment maneuver (Chacko and Rani, 2009). However, these techniques are applied after the diagnosis of ARDS and especially in its severe form where a high number of alveoli are infected and collapsed.

On the other hand, there have been a number of simulation models developed to capture pulmonary mechanics in ARDS patients. The first order model (FOM) describes airway pressure as a series combination of a constant resistance and constant elastance component (Bates, 2009). The viscoelastic model consists of the FOM with an additional parallel circuit of a resistance and an intrapulmonary elastance (Mount, 1955). The expiratory time constant model uses expiration data to determine a parameter that is directly proportional to passive respiratory system elastance (Van Drunen et al., 2013). Madorno's model (Madorno and O. Rodriguez, 2010) simulated the respiratory system behavior using recorded positive flow curve during inspiration and positive end expiratory pressure (PEEP) during expiration. The alveolar distension model describes the pressure–volume relationship of lung tissue using an exponential function (Salazar and Knowles, 1964). The NARX model describes the airway pressure as a function of airway flow rate, inspired volume and PEEP levels (Langdon et al., 2016).

ARDS was associated to changes in cardiovascular and hemodynamic signals. It starts by the destruction of the barriers of the alveolar membranes resulting by increasing the alveolar permeability. Then, injuries progress causing decreased pulmonary compliance that leads to pulmonary hypertension and refractory hypoxia (Mac Sweeney and McAuley, 2016; Patroniti, Isgrò, and Zanella, 2011). Moreover, ARDS has been linked to clinical features such as severe dyspnea, tachypnea and hypoxemia according to (Gajic et al., 2011). Other studies on the clinical profile of ARDS had shown that it is associated with breathlessness, hypotension, tachypnea (Chaudhury

TABLE 1.4: The definition of each used medical term

Medical term	Meaning
Tachycardia	Abnormally rapid heart rate
Dyspnea	Difficult breathing
Tachypnea	Abnormally rapid breathing
Hypoxia	Oxygen deficiency in the tissues
Hypoxemia	Low oxygen level in the blood
Hypertension	Abnormally high blood pressure
Hypotension	Abnormally low blood pressure

et al., 2017) and tachycardia (Park et al., 2016). Definitions on these terms are presented in Table 1.4.

Despite the intense research activity on ARDS characterization, there are very few predictive models for this elusive syndrome. A predictive algorithm was developed to detect ARDS in intensive care unit (ICU) patients in (Ennett et al., 2008). Clinical variables from the multi-parameter Intelligent Monitoring of Intensive Care (MIMIC)-II clinical database (Saeed et al., 2011) were used to establish single and combined rule sets such as plateau pressure, mean airway pressure, oxygen saturation and heart rate. They reached a sensitivity and a specificity of 60% and 82% respectively on a single rule basis, and 50% and 90% respectively for combined rules (Ennett et al., 2008). Other researchers have worked on mortality prediction related to ARDS. The ICU mortality of ARDS patients was shown to be related to the severity APACHE III score, the gas exchange evolution and the development of multiple organ failure (Navarrete-Navarro et al., 2000). Thin-section computed tomography (CT) was a good predictor in patients with clinically early stage of ARDS (Ichikado et al., 2006).

Recent studies have presented a more general issue like respiratory complications. A method of detecting unexpected hospital death associated with respiratory complications was presented in (Ravishankar et al., 2014). Physiological measurements such as respiratory rate and pulse oximetry signals were used to develop a Markov model based algorithm. The proposed algorithm detects various levels of relevant patterns from signals prior to the execution of the Markov model. In a similar scenario, an approach for computing alert thresholds for pulmonary disease patients was developed using three-dimensional set of data collected with an m-health application. Univariate and multivariate methodologies based on novelty detection are used

to analyze daily symptom scores, heart rate and oxygen saturation measurements (Velardo et al., 2014).

1.4 Problem statement

1.4.1 Objective

As mentioned above, ARDS is a critical disease of the respiratory system that threats the health of the elderly. Therefore, the development and implementation of new strategies and technologies are at high priority in order to provide better healthcare services at an affordable price to the ageing population while ensuring maximum comfort and independence. To our knowledge, there is no work in the literature that proposes such models for the prediction of the onset of ARDS.

The main objective of this thesis is to develop a surveillance model that monitors the ongoing health state of elderly people using only physiological signals in order to predict whether patients will develop ARDS in the near future. Although there is no clear clinical feature for ARDS, it is evident that ARDS is linked to cardiorespiratory and cardiovascular responses such as tachycardia, dyspnea, hypoxia and fluctuations in blood pressure. Therefore, four vital signs are considered in this thesis, which are the heart rate (HR), the respiratory rate (RR), the oxygen saturation (SpO_2) and the mean arterial blood pressure (MABP). These signals can be collected in real-time using non-invasive and unobtrusive wearable sensors.

The non-invasive sensors monitor in real-time physiological signals such as ECG, heart rate, oxygen saturation, blood pressure, body temperature and respiratory rate. The measurements obtained from the sensors are transmitted to a processing node using a communication protocol like Bluetooth, ZigBee or ANT (Dementyev et al., 2013). The processing node runs advanced processing, analysis and decision models and may also save and display the results to the user. It transmits the measured data over the Internet to the healthcare agent, thus functioning as the gateway to remote healthcare facilities.

FIGURE 1.5: Some wearable devices that monitor physiological signals

This thesis work is located at the processing node. Thus, processing and analysis models are proposed in the following chapters to predict the occurrence of ARDS. Before the implementation of these models in surveillance systems, they must be tested on existing labeled database to assure the reliability and performance of such models. To validate these models, the MIMIC II database is considered (Goldberger et al., 2000). It is a publicly available database, approved by the Laboratory for Computational Physiology at the Harvard - Massachusetts Institute of Technology (MIT). MIMIC II is valuable in medical education and researches since it provides well-documented case studies of clinically significant pathologies.

1.4.2 Medical devices and sensors

A variety of wearable sensors and physiological signals monitoring systems were proposed in the literature. Examples of these devices are given in Figure 1.5. They can be listed by applications:

- **Commercial products:** Biometric shirt ((by Hexoskin®, Montreal, QC, Canada), Band 2 (by Microsoft®) and fitness trackers (by Fitbit®, San Francisco, CA, USA, Jawbone®, San Francisco, CA, USA, Striiv®, Redwood city, CA, USA and Garmin®, Olathe, KS, USA)

- **Cardiovascular Monitoring Systems:** Wearable mobile ECG monitoring system (Tseng et al., 2014), wireless portable capacitive ECG sensor (Nemati, Deen, and Mondal, 2012), HR monitoring from pressure variance in ear canal (Park et al., 2015) and heart rate monitoring with pressure sensor (Shu et al., 2015).
- **Blood Oxygen Saturation (SpO2) Monitoring Systems:** Analog single-chip pulse oximeter (Tavakoli, Turicchia, and Sarpeshkar, 2010).
- **Multi-Sensor Monitoring Systems:** Ring shaped sensor (Duun et al., 2010; Sola et al., 2006; Huang et al., 2014), wristband sensor (Cai et al., 2010), sensors embedded in soft fabrics (Chen et al., 2010), wireless oximeter (Li and Warren, 2012).

1.4.3 Vital Signs

As mentioned before, four physiological signals are considered in the implementation of the prediction models, which are the HR, the RR, the SpO_2 and the MABP.

1.4.3.1 Heart rate

The heart rate (HR) is the speed of the heartbeat defined by the contractions of the heart in one minute. It is measured by beats per minute (bpm). The normal resting adult human heart rate varies between 60 and 100 bpm (Spodick et al., 1992). It changes with different activities, such as physical exercise, sleep, stress, anxiety, illness and consumption of drugs. Heart rate is not a fixed value; it changes also in response to the body's need to maintain the equilibrium between requirement and delivery of oxygen.

A simple way to determine the heart rate is to measure the pulse at any peripheral point of the body. Heart rate is computed more precisely from the electrocardiograph (ECG). ECG shows the electrical activity of the heart over a period of time. It is formed by P-QRS-T waves as shown in Figure 1.6. On the ECG, instantaneous heart rate is calculated using the R-R interval, that is, the distance between two consecutive R waves. There are different methods to compute HR from the R-R intervals, one of them is to divide 60 by the R-R interval which is expressed in seconds.

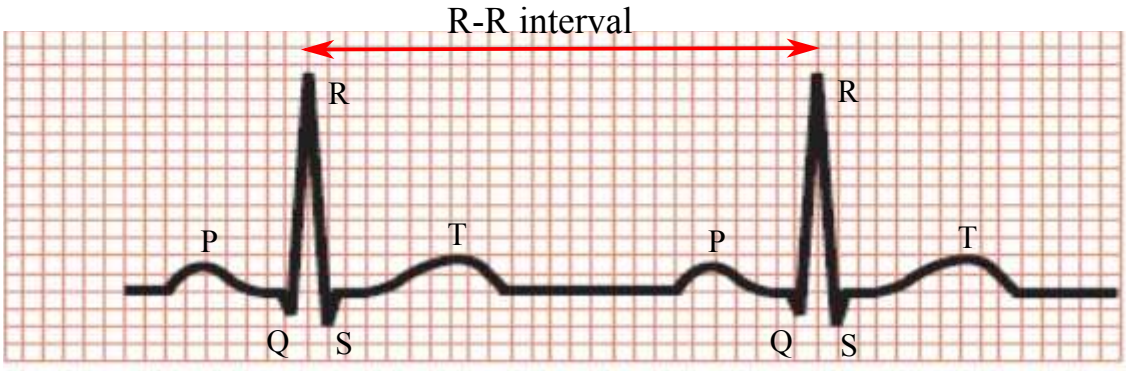


FIGURE 1.6: Electrocardiograph waves and RR interval

1.4.3.2 Respiratory rate

Respiratory rate (RR) is the number of breath cycles per minute. The normal breathing rate for an adult at rest varies in the range of 12 to 18 breaths per minute (bpm). An abnormal respiratory rate can be a sign of respiratory distress, respiratory depression or simply an adaptation of the body to a given situation, such as physical exercise, sleep, stress, anxiety and any other medical condition.

The respiratory rate can be measured simply by counting how many times the chest rises in one minute promoted by the World Health Organization (WHO) (Organization, 1990). It is measured by analyzing the respiratory sound, respiratory airflow and CO_2 emission (AL-Khalidi et al., 2011). Different studies have been conducted to estimate the respiratory rate from other signals, such as ECG signals (Chan, Ferdosi, and Narasimhan, 2013) or photoplethysmogram signals (Nilsson, Johansson, and Kalman, 2000).

1.4.3.3 Peripheral oxygen saturation

Peripheral oxygen saturation (SpO_2) is an estimate of the amount of oxygen in the blood. It presents the percentage of oxygenated hemoglobin relative to the total amount of hemoglobin in the blood. Healthy individuals have oxygen saturation values between 92% and 99%. These values change with the attitude of measurements. A low SpO_2 value reflects a poor oxygenation of the blood, also known as hypoxia.

SpO2 is a photoplethysmogram signal. It is measured by pulse oximetry, which is an indirect, non-invasive measurement method. It is placed generally on a fingertip or earlobe. It works by the emission of small beams of

light through the body part to a photodetector and measuring the amount of oxygen. It measures the changes of light absorption in oxygenated or deoxygenated blood.

1.4.3.4 Mean arterial blood pressure

The mean arterial blood pressure (MABP) is a critical hemodynamic factor. It is the average blood pressure in a body during a single cardiac cycle, that is, the pressure of the pumped blood by the heart into the arteries. Normal range of MABP is 65 to 110 mmHg. Low values of MABP can cause inadequate blood flow to organs, called hypotension. Conversely, high MABP or hypertension contributes to increasing oxygen demand by the heart.

MABP is defined as the product of the blood volume being pumped by the heart at each contraction known as cardiac output (CO) and the systemic vascular resistance (SVR), that is, the vascular resistance needed to push the flow of blood into the arteries $MABP = CO \times SVR$. MABP can be more easily measured from both systolic and diastolic pressures, it is equal to one-third of the systolic added to two-thirds of the diastolic blood pressure.

1.5 MIMIC II database

1.5.1 Description of the database

The dataset considered in this thesis is acquired from the Physio bank's Multi-parameter Intelligent Monitoring of Intensive Care (MIMIC)-II database (Goldberger et al., 2000; Saeed et al., 2011). It is a publicly available database, collected at BID Medical center over a seven-year period. The MIMIC II database directory contains thousands of physiological measurements and clinical information collected from intensive care unit (ICU) patients.

Physiological measurements were recorded using bedside monitors in the adult ICUs, and stored in the Waveform database, simultaneously with numeric time series derived from the raw signals. These data were validated by ICU nurses on an hourly basis. Waveforms normally include two leads of electrocardiographic, continuous blood pressure (BP) waveforms, pulmonary artery pressure or fingertip photoplethysmogram signals having

a sampling frequency of 125 samples/second. Numeric variables were recorded once per minute. These variables include heart rate, respiratory rate, peripheral oxygen saturation, systolic/mean/diastolic blood pressure, temperature and cardiac output.

On the other hand, the clinical database consists of data collected from different information systems in the hospital. It contains detailed clinical information recorded far less frequently than waveforms. It includes laboratory results, ventilator settings, International Classification of Diseases Ninth revision (ICD-9) codes, medications, admissions and death records. Some of the records presented in the Waveform database have been matched and time aligned with the clinical database records.

1.5.2 Subjects Selection

According to the definition of AECC, ARDS is diagnosed using the PaO_2/FiO_2 ratio, radiology readings of chest X-ray, and the pulmonary artery wedge pressure. Nevertheless, reading of chest radiograph reports is time consuming, and pulmonary artery wedge pressure is rarely recorded. Consequently, the PaO_2/FiO_2 ratio can be used alone to identify ARDS patients as mentioned in the Berlin definition. Therefore, ARDS is diagnosed if the PaO_2/FiO_2 ratio decreases and becomes less than 200. In order to identify developers and non-developers of ARDS called ARDS and non-ARDS subjects respectively, it is required to start the procedure of patients' selection from the clinical database. The clinical database can be divided to ventilated and non-ventilated patients as shown in Figure 1.7. The ventilated patients are further divided according to the type of ventilation, thus into invasive and non-invasive ventilated patients.

Two different selection procedures were handled on each type of subjects as explained in the following. The selection procedures consist of forming two groups of subjects, ARDS and non-ARDS groups. The ARDS group is formed using the definition of both moderate and severe ARDS with $PaO_2/FiO_2 \leq 200$ mmHg in addition to PEEP levels higher than 5 cmH_2O; while non-ARDS group includes subjects having mild and non-ARDS patients with PEEP <5 cmH_2O. In both types, the selection procedure starts

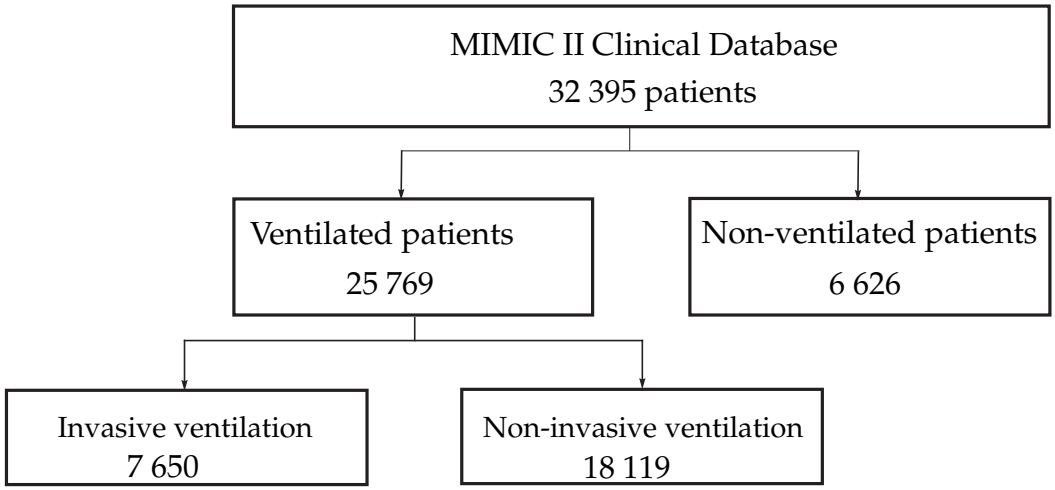


FIGURE 1.7: MIMIC II Clinical database subjects

from the clinical database to form these two groups. Than, the selected subjects were matched to the waveform database in order to extract all the subjects having physiological signals. Four time series are extracted, which are the heart rate (HR), the respiratory rate (RR), the oxygen saturation (SpO_2) and the mean arterial blood pressure (MABP). These time series have a sampling frequency of one sample/minute.

1.5.2.1 Invasive ventilated subjects

The first selection procedure was handled based on the studies of (Xiaoming et al., 2008; Ennett et al., 2008). This selection procedure consists of extracting only subjects underwent invasive ventilation for more than 48 hours and not having ARDS at the onset of ventilation. The selection of subjects is described in the block diagram of Figure 1.8.

The length of mechanical ventilation is defined as the duration of the first continuous ventilation period according to the recorded ventilator settings in the clinical database. The subjects were identified by computing the ratio of each PaO_2 value to the nearest FiO_2 value available before the corresponding blood gas value. 1 873 subjects were excluded from the study since they had PaO_2/FiO_2 values < 200 mmHg during the first 12 hours of recordings. Among 1 023 subjects that did not develop ARDS at the beginning of recordings, 502 subjects developed moderate or severe ARDS in the following, thus had worsening PaO_2/FiO_2 values and considered "ARDS subjects"; while 521 subjects maintained a PaO_2/FiO_2 ratio higher than 200 all over the signals recordings, and thus they are called the "non-ARDS

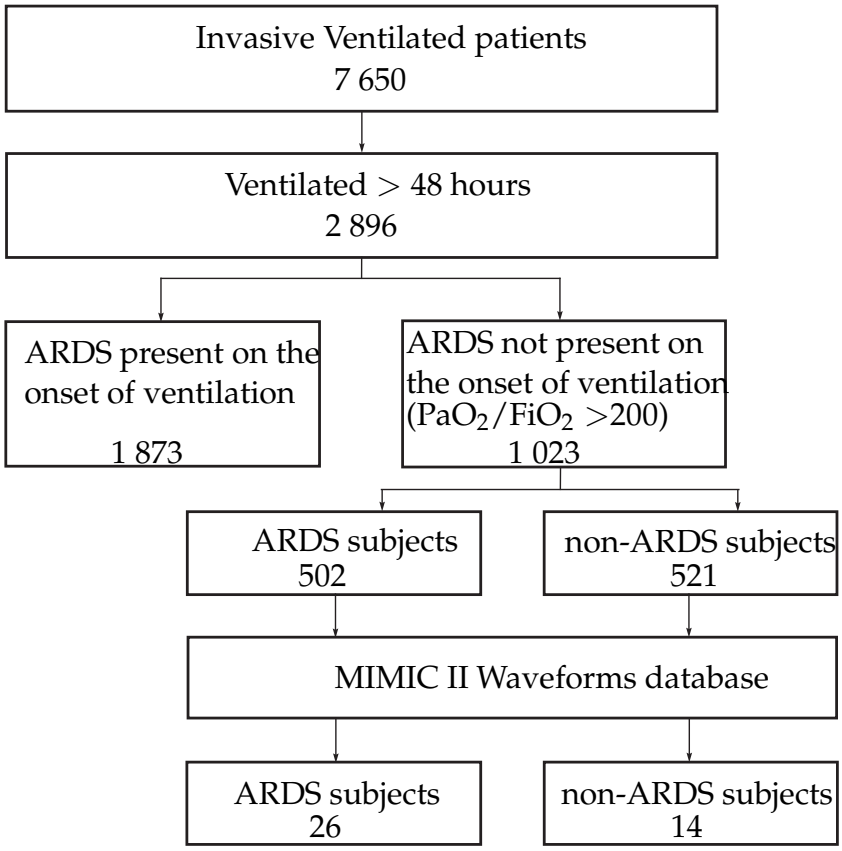


FIGURE 1.8: Selection for the invasive ventilated subjects

subjects".

After that, numeric time series of these subjects are extracted from the waveform database that are HR, RR, SpO2 and MABP. For the processing step, the first six hours of the signals are removed to rule out the effect of the ventilator onset and to ensure the stable state at the beginning of the analysis. For non-ARDS subjects, data are considered for the entire period of ventilation, whereas for ARDS subjects, data are extracted until a respiratory instability is detected, that is, $PaO_2/FiO_2 \leq 200$ mmHg. Thus, we obtained 38 ARDS subjects and 18 non-ARDS subjects. In order to validate the proposed approach, only subjects having high duration signals are considered, thus signals longer than 1440 minutes (24 hours). Finally, 26 ARDS and 14 non-ARDS subjects are obtained.

1.5.2.2 Non-Invasive ventilated subjects

As mentioned before, the main objective of this thesis is the implementation of a system for the prediction of pathologies for elderly people at home.

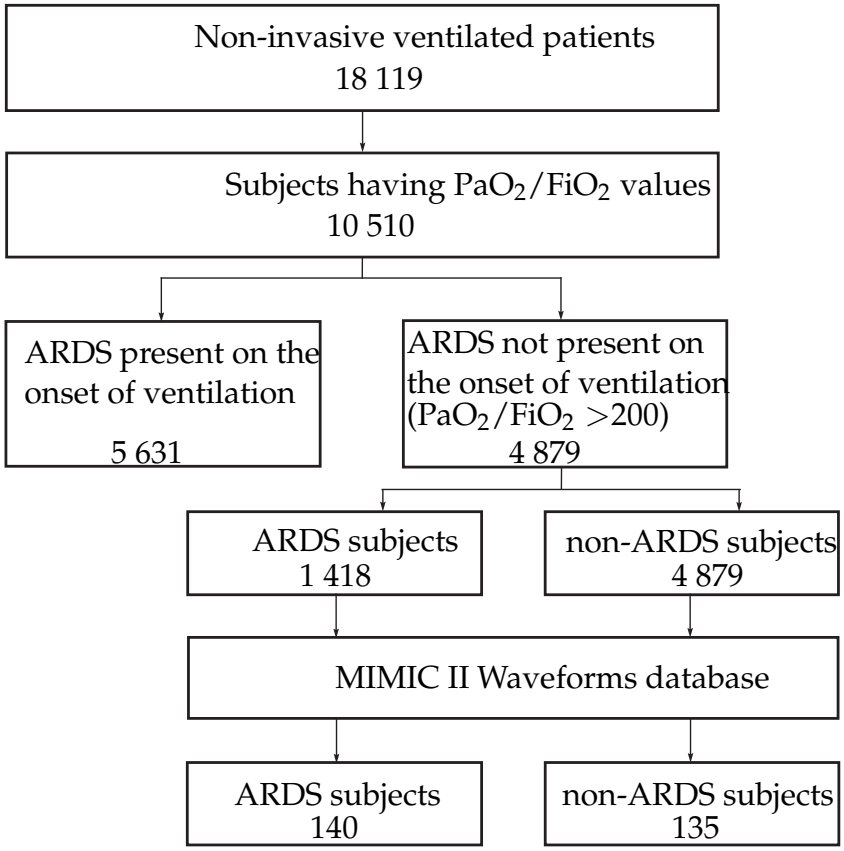


FIGURE 1.9: Selection for the non-invasive ventilated subjects

Therefore, we need data that is more similar to the data that can be acquired at home. Thus, non-invasive ventilated patients were extracted as shown in Figure 1.9. 10 510 out of 18 119 subjects were found to have values for PaO_2 and FiO_2. PaO_2/FiO_2 values were extracted for each subject and computed. Then, the same selection procedure continues as the preceding one. ARDS and non-ARDS groups of subjects are formed in the clinical database, matched to the waveform database and their time series are extracted. This leads to 140 ARDS subjects and 135 non-ARDS subjects.

Among the ARDS subjects, there are 50 subjects that started their records after ARDS diagnosis. These subjects are irrelevant for this study and thus were eliminated. Other ARDS subjects started their records before the onset of ARDS but stopped before ARDS diagnosis; these subjects, of number equal to 18, are also eliminated. Therefore, only the 72 subjects having started records before the ARDS onset and last at least until ARDS occurrence are considered. After that, a preprocessing step was required to smooth the signals and remove the noise. This leads to 72 ARDS subjects and 135 non-ARDS subjects. Then, the number of subjects is reduced because of the removal of

too short signals and the balancing of groups. This leads finally to 50 ARDS and 50 non-ARDS subjects.

1.6 Organization of the manuscript

In this chapter, we briefly described the respiratory system and its conditions. Abnormal condition in any part of the human body can provoke serious pathologies involving the main functions of the respiratory system. This chapter discusses also one such pathology, acute respiratory distress syndrome, which is the main objective of our study. We introduced and explained the main ideas behind ARDS. We observed that the management of ARDS is not an easy task especially with the lack of studies on the early prediction of this elusive syndrome. The rest of the manuscript is organized as follows.

The second chapter gives a brief overview on the different levels of information fusion techniques that can be applied to any type of data and highlights the adequate techniques to the present application. In order to predict ARDS, different time series data are considered. These data come from different types of sensors. Thus, the decision fusion level and the feature fusion level are the best to explore such signals. In addition, we introduce a description of kernel methods. All these techniques are defined on a binary concept. In this thesis, we are differentiating between two groups of subjects, an ARDS group including subjects who will develop a severe or moderate ARDS; and a non-ARDS group that consists of subjects who have mild ARDS or not developing ARDS at all.

In the third chapter, we propose a model to predict ARDS in real-time using the first type of information fusion, namely the decision level fusion. Thus, a novelty detection algorithm is proposed in order to generate individual decisions from each signal. Indeed, the algorithm's parameters are adapted for each type of signals. Then, the individual decisions are combined using one of the proposed decision fusion techniques. We propose two linear techniques based on the performance rate and the error rate and a nonlinear technique based on kernel methods. The performance of the proposed methods is illustrated using both the invasive and non-invasive ventilated subjects.

The fourth chapter proposes a feature fusion based model for the prediction of ARDS in real-time. The model consists first of the extraction of linear and nonlinear parameters from the time series data. Extracted parameters are mean, standard deviation, skewness, kurtosis, sample entropy and detrended fluctuation analysis. Then, a parameters selection procedure is performed to obtain the most pertinent parameters and reduce the input vector dimensionality. Afterwards, the belief functions theory is used to define a classifier that combines the best set of parameters. The classifier is defined from the parameters probability distributions. In this chapter, we also test the proposed model on a multi-class problem in order to differentiate between each type of ARDS subjects, namely, severe ARDS, moderate ARDS, mild ARDS and non-ARDS. The performance of the method is studied for the non-invasive ventilated subjects only.

In chapter five, we extend the model proposed in chapter 4. First, a new extraction of parameters is performed to expand the input vector. The new extracted parameters are computed from the recurrence quantification analysis. Then, the belief functions theory is also extended. A new approach of constructing mass functions is proposed in this chapter. Indeed, mass functions are constructed by a kernel based learning algorithm. A selection procedure is also defined to reduce the dimensionality of the parameters vector. The performances of this model are obtained using the non-invasive ventilated subjects and compared to the previously proposed model.

Finally, chapter six provides concluding remarks, a discussion of limitations, and an outlook on future research directions.

1.7 Publications

Peer-reviewed international journal articles (2 + 1)

- **A. Taoum**, F. Mourad-Chehade, and H. Amoud. "Early-Warning of ARDS using Novelty Detection and Data Fusion". Computers in Biology and Medicine, volume 102, pp. 191–199, November 2018.
- **A. Taoum**, F. Mourad-Chehade, and H. Amoud. "Evidence-based Model for Real-time Surveillance of ARDS". Biomedical Signal Processing and Control, volume 50, pp. 83–91, January 2019.

- **A. Taoum**, F. Mourad-Chehade, and H. Amoud. "Kernel-Evidence based Model for Real-time Surveillance of ARDS". Biomedical Signal Processing and Control. Submitted, January 2019

Peer-reviewed international conference articles (4 + 1)

- **A. Taoum**, H. Amoud, and F. Mourad-Chehade. "Multi-class surveillance for acute respiratory distress syndrome using belief functions". Accepted in 2019 Advances in Science and Engineering Technology multi-conferences (ASET), Dubai, March 2019.
- **A. Taoum**, F. Mourad-Chehade, and H. Amoud. "Features fusion using belief functions theory for ARDS prediction". In International Conference On Signal Processing (ICOSP), Paris, France, December 2018.
- **A. Taoum**, M. Sabbagh, H. Amoud and F. Mourad-Chehade. "Features extraction from vital signs to characterize acute respiratory distress syndrome". In 4th IEEE Middle East Conference on Biomedical Engineering (MECBME), Gammarth, Tunisie, March 2018.
- **A. Taoum**, F. Mourad-Chehade, H. Amoud, A. Chkeir, Z. Fawal, and J. Duchêne. "Data Fusion for Prediction of ARDS using MIMIC II Physiological Database". In 18th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom16), Munich, Germany, September 2016.
- **A. Taoum**, F. Mourad-Chehade, H. Amoud and Z. Fawal. "Predicting ARDS using MIMIC II Physiological Database". In IEEE International Multidisciplinary Conference on Engineering Technology (IMCET 16), Beirut, Lebanon, November 2016.

Peer-reviewed national conference articles (1)

- **A. Taoum**, F. Mourad-Chehade and H. Amoud. "Un modèle d'alerte précoce du SDRA par detection d'anormalité et fusion de données", in XXVIème colloque GRETSI, Juan-les-Pins, France, 5-8 Sept. 2017.

Chapter 2

Information Fusion Techniques

Contents

2.1 Introduction 30

2.2 Information fusion 30

2.3 Decision fusion techniques 31

2.3.1 Decision fusion procedure 31

2.3.2 Majority Voting 33

2.3.3 Weighted Majority voting 34

2.4 Feature fusion techniques 36

2.4.1 Probabilistic fusion 38

2.4.2 Belief functions theory 39

2.5 Machine learning and kernel methods 43

2.6 Conclusion 45

2.1 Introduction

The analysis of different types of data received from heterogeneous sensors requires fusion techniques. Multiple sensor systems were originally investigated in military applications. Then, data fusion methods have been developed for biomedical applications, especially to detect the health state of patients based on the data obtained from different medical sensors. Different types of sensors produce different sets of signals. Fusion techniques combine data obtained from several sources in order to generate a single decision for the task being handled. It consists of producing a combined result that provides the most efficient and reliable information possible.

This chapter presents a review of the different techniques and algorithms used in decision fusion and feature fusion strategies. The next section describes the different levels of information fusion. Following that, only two levels the decision fusion and feature fusion techniques are described, with the most known methods presented in the literature for fusion.

2.2 Information fusion

Fusion techniques are widely used nowadays in several fields such as biomedical applications, sensor networks, robotics, image processing and intelligent system design. The fusion of multi-source data presents important advantages over single source data. The use of various sources may increase the accuracy with which a variable can be observed. In surveillance models of biomedical applications, identifying pathologies can be done using just one signal (Klingeberg and Schilling, 2012), but most of the serious diseases occur with concurrent irregularities in multiple vital signs (Forkan and Khalil, 2017). Fusion technique is the process of combining data from different measurement sensors, information processing blocks and decision makers (McKerrow and Volk, 1996). There are mainly three types of fusion techniques (Dasarathy, 1994),

- high-level fusion: Decision fusion,
- intermediate-level fusion: Feature fusion,
- low-level fusion: Data fusion.

Data fusion combines raw data produced by multiple sources to obtain a new combined raw data that is expected to be more informative than the

inputs. Low-level fusion can be handled if the source's data are homogeneous, thus are of same type. Conversely, if the data are of different types then the data must be combined at an intermediate or high-levels fusion (Llinas and Hall, 1998). In biomedical applications where different signals are considered, like the one studied in this manuscript, low-level fusion cannot be employed.

Instead of using the raw data, various feature extraction techniques are proposed in the literature to extract features that represent the characteristics and information hidden in the source data. These features are then combined together at the intermediate-level to result in a feature with higher discrimination (Rattani et al., 2007). Feature fusion methods include the selection and combination of features to eliminate irrelevant and redundant features present in the feature space. Decision fusion aggregates decisions from several models in order to obtain a more accurate fused decision (Wilk and Wozniak, 2012). In decision fusions, one classifier is considered per data source to generate individual decisions that are combined afterwards. Some decision fusion techniques combine decisions with an uncertainty measure to generate a composite decision.

2.3 Decision fusion techniques

Decision fusion uses various rules to combine decisions from multiple decision makers into a final decision. The purpose of merging the decisions of separate data-mining algorithms operating on separate data sources is to increase the overall performance. The decision makers can have same or different type of data (Kuncheva, Bezdek, and Duin, 2001). In recent years, decision fusion techniques have been widely investigated (Kuncheva, Bezdek, and Duin, 2001; Ho, Hull, and Srihari, 1994; Xu, Krzyzak, and Suen, 1992; Kuncheva and Whitaker, 2003). Various decision fusion rules have been created and it has been demonstrated that most of them outperform single classifiers. However, it is unclear why some combination rules work better than others and which combination rule works better for a specific condition.

2.3.1 Decision fusion procedure

There are different classifiers that can be applied from the literature to make individual decisions, such as support vector machines (SVM) with

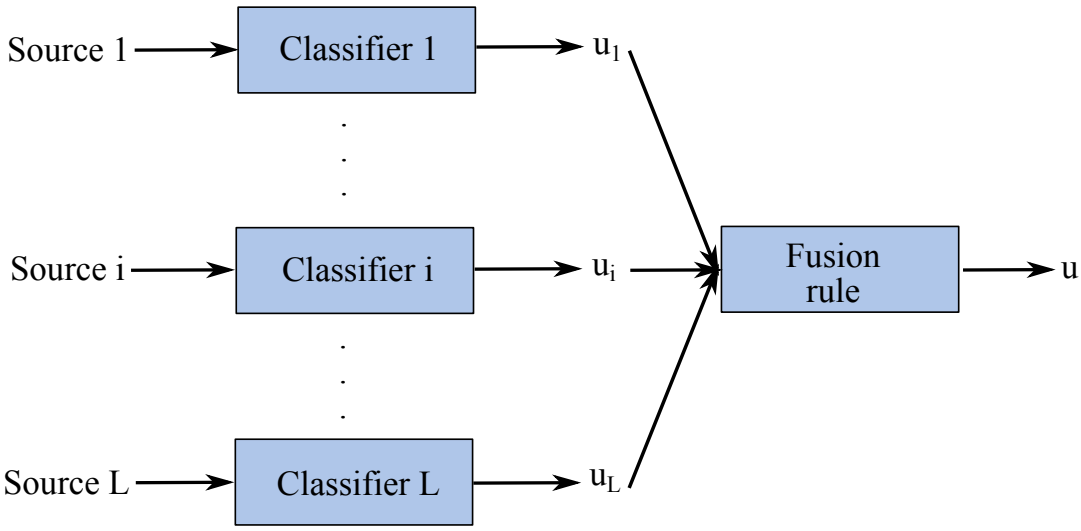


FIGURE 2.1: Block diagram of decision fusion procedure

various kernels (Ho, Hull, and Srihari, 1994), artificial neural networks (ANN) (Bishop and Bishop, 1995), K-nearest neighbors (K-NN) (Cover and Hart, 1967) and gaussian mixture models (GMM) (Dempster, Laird, and Rubin, 1977). On the other hand, new models can be proposed according to the application and the type of data in study in order to generate individual decisions.

For a given problem with L sources of information, L classifiers are defined. These classifiers can be of same or different types and of same or distinct properties. For each source, each classifier i, $i = 1, \ldots, L$, assigns a decision u_i to an input vector as shown in the block diagram of Figure 2.1. In this plot, the decision u_i of classifier i is that the input vector from source i belongs to one of the possible C classes according to its prior and current information. These decisions are then combined using a single fusion rule to obtain a final decision u as output. A review of various fusion rules is given in the following.

There are three categories of decision fusion based on the type of classifier outputs (Mangai et al., 2010): the abstract level, the rank level and the measurement level. At the abstract level, classifiers produce as output a class label. The output of rank classifiers is a subset of possible matches of classes sorted in the decreasing order of confidence. At the measurement level, the outputs of classifiers are probabilistic confidence measures obtained for each class. In this work, we are interested in the fusion of the first type of

classifiers output, the abstract level. The output of classifiers is an L×C-dimensional matrix containing the individual decisions $u_{i,c}$ of each of the C classes from each of the L classifiers. Thus, the decision obtained from classifier i corresponding to class c is defined as $u_{i,c} \in \{0,1\}$, with $i = 1,\ldots,L$, and $c = 1,\ldots,C$. Then,

$$u_{i,c} = \begin{cases} 1, & \text{if classifier } i \text{ chooses class } c; \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

The next sections describe two ways of combining the label output classifiers, namely, the majority voting and the weighted majority voting. The majority voting method relies on the labels obtained from the classifiers under the assumption of all classifiers perform equally. However, the weighted majority voting considers that there are classifiers more competent than others and have different performances. In this type of fusions, the output of each classifier is accorded with a measure of confidence based on the performance of these classifiers.

2.3.2 Majority Voting

In majority voting, a class is chosen according to the following situations, when: (1) all classifiers decide the same specific class, this rule is known as the conjunctive rule; (2) at least one classifier decides the specific class, known as the disjunctive rule; or (3) it receives the highest number of votes from the classifiers, known as the simple majority voting. In this case, the final decision is assigned according to the number of positive decisions given by the individual classifiers to each of the C classes. The final decision of this decision rule is described as follows: class k is chosen for,

$$k = \arg\max_{c=1}^{C} \sum_{i=1}^{L} u_{i,c}. \tag{2.2}$$

The majority voting is an optimal fusion rule in the following conditions: the number of classifiers is odd, the classifier outputs are independent and the classes are equiprobable, thus the probability of a classifier to give a class label is equal for any instance. It has been shown that if each of the L classifiers has an error rate less than 50%, the accuracy of the combination increases with the number of classifiers (Valdovinos and Sanchez, 2009). However, this assumption is not always satisfied. In some cases, individual classifiers

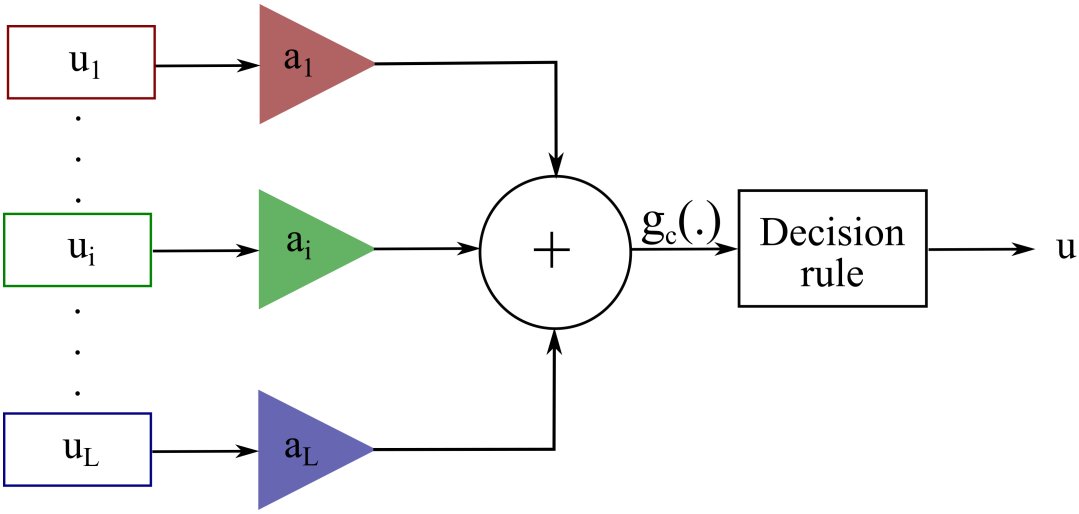


FIGURE 2.2: The block diagram of weighted fusion rule

might perform better than the simple voting fusion (Matan, 1996). Thus, weighting methods have been proposed to overcome these difficulties.

2.3.3 Weighted Majority voting

Weighted decision fusion rules have been proposed as an extension of the simple majority rules. Each classifier goes through a preliminary classification to test its performance. This performance is then used to determine a measure of confidence or weight noted a_i, $i = 1, ..., L$, that will be used jointly with the individual decision $u_{i,c}$. Then, the final decision can be made by majority, average (Kuncheva, 2001), minority (Chen and Cheng, 2001), product of votes or using other more complex methods (Kuncheva, Bezdek, and Duin, 2001). A descriptive block schema is illustrated in Figure 2.2. The intermediate function obtained through weighted voting computed for class c is defined by

$$g_c(u_c) = \sum_{i=1}^{L} a_i u_{i,c}. \tag{2.3}$$

The value of the intermediate function for each class c will be the sum of the individual decisions obtained from classifiers multiplied by the confidence measure of each classifier. Then, the class k can be chosen according to equation (2.4) as follows,

$$k = \arg\max_{c=1}^{C} \sum_{i=1}^{L} g_c(u_{i,c}) = \arg\max_{c=1}^{C} \sum_{i=1}^{L} a_i u_{i,c}. \tag{2.4}$$

TABLE 2.1: Different weights definitions for weighted fusion rules from the literature

Name	Equation	Reference
Inverse distance weight	$a_i = \frac{1}{d_i}$, if $d_i \neq 0$	(Dudani, 1976)
Average distance weight	$a_i = \frac{\sum_{\ell=1}^{L} d_\ell}{d_i}$	(Valdovinos and Sanchez, 2009)
Benediktsson weighted rule	$a_i = \frac{Acc_i - S_{min}}{S_{Diff}}$	(Meng et al., 2010)

Several weighting functions, based on the ability of classifiers to differentiate between classes, have been proposed in the literature. To apply such methods, functions of distance between class centers and functions of classifiers performances are proposed to compute the weights a_i, such as in (Meng et al., 2010; Mourão, 2014). Table 2.1 presents several weighting functions from the literature.

Dudani has proposed a voting rule for the K-NN algorithm based on the distance of labels from classes. An inverse distance weight function has been proposed to assign for each neighbor a weight according to its distance from the target class center. According to Dudani, "a neighbor with smaller distance is weighted more heavily than one with a greater distance" (Dudani, 1976). Thus, the Dudani's weight can be written as follows:

$$a_i = \frac{1}{d_i}, \text{ if } d_i \neq 0. \tag{2.5}$$

In the same scenario, another weighting function has been proposed by Valdovinos based on the average distance weight (Valdovinos and Sanchez, 2009). This weighting rule assigns the highest weight to the classifier with the nearest neighbor to the input instance. The average distance weight can be expressed as in Eq. (2.6):

$$a_i = \frac{\sum_{\ell=1}^{L} d_\ell}{d_i} \tag{2.6}$$

The most popular weighting function is the weighted majority voting or WMV proposed by (Meng et al., 2010). It is presented by Benediktsson and can be adapted to performance based weighted majority voting. The weights are calculated using the performances of the classification: all the scores are

normalized to the range 0 to 1. Here, the weighted majority voting is a normalization of Dudani's simple weight function with the criterion being the accuracy instead of distance.

$$a_i = \frac{Acc_i - S_{min}}{S_{Diff}}. \tag{2.7}$$

where Acc_i denotes the accuracy of classifier i, defined as the ratio of the number of truly identified subjects to the ratio of the total number of subjects. $S_{max} = \max_{i=1}^{L}(Acc_i)$, $S_{min} = \min_{i=1}^{L}(Acc_i)$ and $S_{Diff} = S_{max} - S_{min}$.

2.4 Feature fusion techniques

Despite the fact that decision fusion strategies have demonstrated an increase in performances against single model systems, features are the most informative indicators for raw data in classification and characterization tasks. However, noisy and overlapping set of features can distort the performance of a good classifier. Hence, the selection of a proper set of features becomes an essential step in feature fusion. It is necessary to identify the set of valuable features and exclude the ones that confuse the classifier and decrease the performance.

In addition, the dimensionality of the data participating in data mining tasks has increased explosively. Data with high dimensionality increases the time of computations without increasing in performance. Features selection is jointly used with feature fusion to overcome this dimensionality problem. Despite the effort induced on features selection algorithms, it is still not clear how to combine the features in the best possible way. Feature fusion can be divided into the following methods, namely, features or parameters extraction, features ranking, features selection and features combination. The flowchart of Figure 2.3 presents a feature fusion strategy.

For the extraction of features or parameters from raw data, several methods can be applied. Different linear and non-linear parameters exist in the literature in both temporal and frequency domains. From this initial set of features, a ranking and selection phase follows to rank the features based on their importance or to find a minimal set of features that represents the whole set. Features selection algorithms are divided into three types, namely, wrapper method, filter method and hybrid method (Liu and Yu,

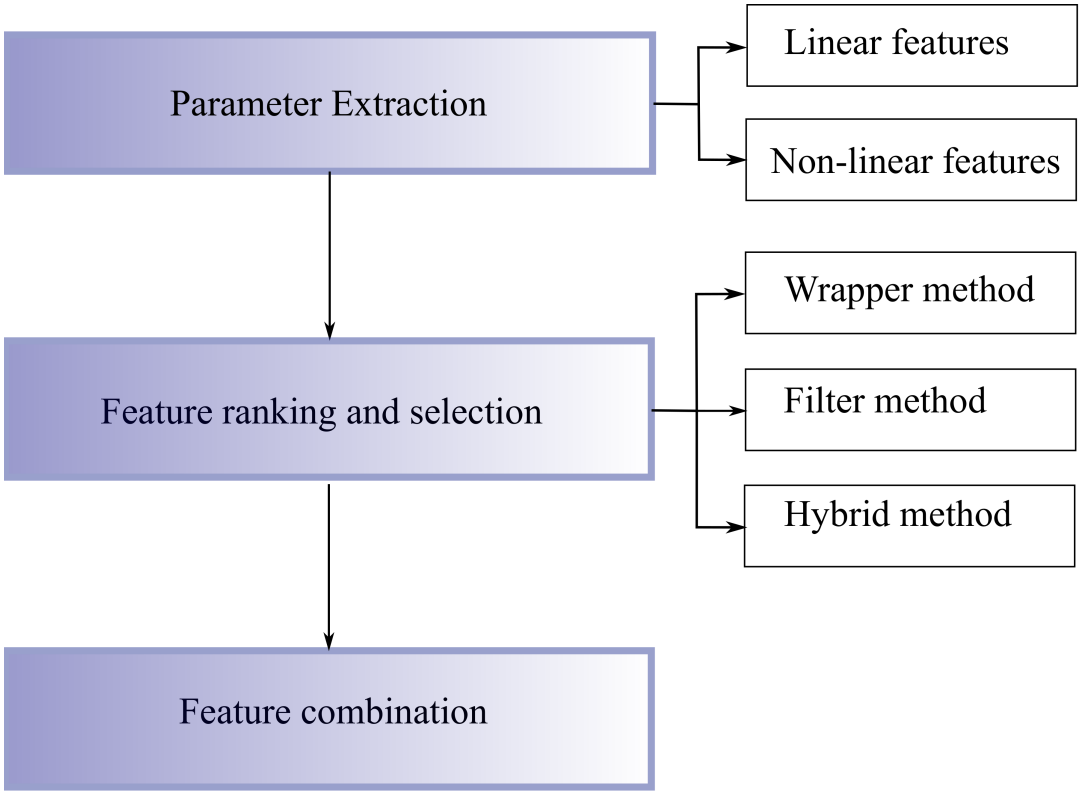


FIGURE 2.3: Feature fusion strategies

2005). The wrapper method selects the best subset based on its performance through the classifier algorithm in the evaluation phase. On the other hand, the filter method selects the optimal feature set independently of any algorithm. hybrid method is a combination of both types of methods.

Once the best feature subset is selected, a features combination is performed. The combination can be done using different methods. The features space is partitioned into distinct classes in a training phase. A new coming features vector is compared with the pre-classified features space and is attributed to one of the classes. The classification strategies are divided into parametric and nonparametric approaches (Waske and Benediktsson, 2007). Parametric methods make assumptions about the features distributions of the input data; while nonparametric methods do not require any assumption on the distribution of input data. One can uses machine learning methods to combine and classify feature sets, such as SVM, *K*-NN, neural networks and others (Keller, Gray, and Givens, 1985; Vapnik, 1999; Friedman, Hastie, and Tibshirani, 2001). In addition, probabilistic models can be implemented for the feature fusion such as the Bayesian theory or a more complex one as the belief functions theory.

2.4.1 Probabilistic fusion

The probability theory deals with almost all kinds of imperfect information like the uncertainty of the information. Data uncertainty is expressed using a rigorous mathematical framework based on the probability distribution or the density functions. It is also named the "Bayesian fusion" since it uses the Bayes estimator. The Bayesian approach consists mainly of four steps: modeling, estimation, combination and decision. In the Bayesian fusion, all classes are assumed to be independent (Mangai et al., 2010).

The Bayes estimator computes the conditional probability distributions for each variable in the feature input space (Spiegelhalter et al., 1993). The conditional probability functions employed to build the Bayesian model are used to compute the prior joint distribution of the variables in the model. The prior joint distribution is updated with new observations or likelihoods of occurrence of variables. Let us consider a final set of n features. The Bayesian fusion process determines if an object belongs to class ω_c. The conditional posterior probability of a hypothetical state or class ω_c for new observations X can be written as $p(\omega_c \mid X)$, where $X = \bigcup_i \{X_i\}$, $\forall i$. It is computed as follows

$$p(\omega_c \mid X) = \frac{p(X \mid \omega_c)p(\omega_c)}{p(X)} \tag{2.8}$$

where $p(\omega_c)$ is the prior probability distribution of the state ω_c. The denominator of Eq. (2.8) $p(X)$ is a normalizing term to ensure that the probability density function integrates to one. Finally, $p(X \mid \omega_c)$ is the likelihood function obtained from the given sensor measurement model. It is the probability of the particular set of observations X_i, $i = 1, \ldots, n$, knowing that the object belongs to class ω_c. $p(X \mid \omega_c)$ is computed as follows,

$$p(X \mid \omega_c) = p(X_1 \mid \omega_c)p(X_2 \mid \omega_c) \ldots p(X_n \mid \omega_c) \tag{2.9}$$

The representation of probability density over the state ω_c differs according to the type of the Bayesian method being used. There are Kalman filters (KF), extended Kalman filters (EKF), grid-based filters and particle-based filters (Kalman, 1960; Fox et al., 2003; Luo, Chang, and Lai, 2011).

For the last step, various decision criteria exist in the literature. The most employed rule is that of the maximum a posteriori, ω_k is chosen as follows:

$$k = \arg \max_{c \in \{1,\ldots,C\}} p(\omega_c \mid X). \tag{2.10}$$

The main interest of the probabilistic approaches is that it relies on a solid mathematical base proven by many studies for years (Cobb and Shenoy, 2003). The probabilities theory is mainly applied in the case where the knowledge was acquired during the learning phase. However, its major problem is that it essentially represents uncertainty and very poor imprecision, this often leads to a confusion between these two notions. Moreover, the probabilistic modeling deals with singletons only that represent the different decisions. These decisions are mutually exclusive and exhaustive. This means that two states cannot occur at the same time and one of them must occurs, which does not always correspond to reality.

2.4.2 Belief functions theory

The theory of belief functions was introduced by Dempster (Dempster, 1967) and Shafer (Shafer, 1976) to address the limitation of the probabilistic theory and for the analysis of imperfect information. This theory makes it possible to model well the uncertainty and the imprecision. It also takes account of ambiguities and conflicts between sources with a theoretically evidential reasoning framework. It relies on assigning belief and plausibility to an event. This modeling is realized from mass functions allowing a good representation of the knowledge.

The use of the belief functions theory for the problem of data fusion was first presented in 1981 (Garvey, Lowrance, and Fischler, 1981). Thereafter, the belief functions theory was used for information fusion in different fields (Wu et al., 2002; Wu, Siegel, and Ablay, 2003; Bossé et al., 2006; Morbee et al., 2010). In contrast to the Bayesian theory, the belief functions theory allows each source to contribute information in different levels of detail. A priori probabilities are not assigned to unknown propositions, because they are assigned only when the information is provided. In fact, it allows for explicit representation of total ignorance by assigning the entire mass to the frame of discernment.

2.4.2.1 Basic concept

The belief functions theory operates on a frame of discernment Ω which consists of a finite set containing C mutually exclusive propositions or states of the system $\Omega = \{\omega_1, \omega_2, \ldots, \omega_C\}$. Let 2^Ω be the set of all the subsets of Ω, that is, $2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, ..., \Omega\}$. Information given by a source can be represented by a *basic belief assignment* (BBA), also named a mass function, expressed by $m(\cdot)$ (Smets and Kennes, 1994). It is defined from 2^Ω to [0,1], and has the following properties:

$$\begin{cases} m(A) \rightarrow [0,1],\ \forall A \in 2^\Omega \\ m(\emptyset) = 0 \\ \sum\limits_{A \in 2^\Omega} m(A) = 1 \end{cases} \tag{2.11}$$

where the mass $m(A)$ reflects the proportion of evidence supporting the hypothesis of the state being in A, A being an element of 2^Ω. Unlike the Bayesian theory, the belief functions theory assigns evidence to subsets of Ω with no influence on the singletons (Shafer, 1976; Smets and Kennes, 1994). Thus, any object is allowed to belong to any subsets of Ω and not only singleton classes $\{\omega_c\}$. Moreover $m(A)$ is not a probability measure, but only an interpretation of the observed information. Indeed, it is possible to have $m(A) > m(B)$, even if $A \subset B$.

Two other basic functions of the belief functions theory are the belief function $Bel(\cdot)$ and the plausibility function $Pl(\cdot)$. $Bel(A)$ measures the strength of evidence supporting A and is defined as $Bel(A) = \sum_{B \subseteq A} m(B)$ where $Bel(\emptyset) = 0$, and $Bel(\Omega) = 1$. $Pl(A)$ is an upper bound of the evidence supporting A and is defined as $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$. The belief and plausibility of A provide lower and upper bounds of the probability of having the state in A as below:

$$Bel(A) \leq P(A) \leq Pl(A).$$

2.4.2.2 Discounting

The discounting operation allows taking into consideration the reliability of the information provided by a source (Shafer, 1976; Smets, 1993). Suppose that a source provides information represented by a received mass function $m(\cdot)$ and $\beta \in [0,1]$ is the degree of reliability of the source (Mercier, Lefèvre, and Delmotte, 2012; Mercier, Pichon, and Lefèvre, 2016). The mass function

$m(\cdot)$ is discounted by a discounting factor $\alpha = 1 - \beta \in [0, 1]$. α is the degree of belief that the source is not reliable. This way, the information provided by the source is controlled. Then, the discounted mass function of $m(\cdot)$ is denoted by ${}^{\alpha}m(\cdot)$ and represented by the following equation (Shafer, 1976):

$$\begin{cases} {}^{\alpha}m(A) = (1-\alpha)m(A), \forall A \subset \Omega, \\ {}^{\alpha}m(\Omega) = (1-\alpha)m(\Omega) + \alpha. \end{cases} \tag{2.12}$$

In the case when the source is not reliable at all, thus the information provided cannot be taken into account, α is equal to 1 and the mass value will be discounted to the ignorance as ${}^{\alpha}m(\Omega) = 1$. On the contrary, when the source is fully reliable, α is equal to 0 and the mass function remains the same after discounting as ${}^{\alpha}m(A) = m(A)$, thus the provided information by the source is entirely accepted.

In certain cases, the reliability of each source depends on the truth of the hypothesis of interest. In order to take into consideration the evidence on the reliability of the source, conditionally on each $\omega_c \in \Omega$, the contextual discounting operation has been proposed in (Mercier, Quost, and Denœux, 2008; Mercier, Lefèvre, and Delmotte, 2012; Mercier, Pichon, and Lefèvre, 2016). Just like the classical discounting, which is characterized by the discounting operator α, the contextual discounting is characterized by the vector $(\alpha_1, \ldots, \alpha_C)$, where α_c is the degree of non-reliability of the source, while knowing that the state is in ω_c. In this case, the computation of ${}^{\alpha}m(\cdot)$ can be generalized as follows:

$${}^{\alpha}m(A) = \sum_{\substack{B \cup B' = A \\ B, B' \in 2^{\Omega}}} \left[\prod_{\omega_v \in B'} \alpha_v \prod_{\omega_j \in \overline{B'}} \beta_j \right] m(B), \forall A, B \text{ and } B' \in 2^{\Omega}, \tag{2.13}$$

where $\beta_j = 1 - \alpha_j$ is the reliability of the source conditionally on ω_j.

2.4.2.3 Combination

Let $m_1(\cdot), \ldots, m_L(\cdot)$ be mass functions from L different sources of information defined over 2^{Ω}. They can be combined using the conjunctive rule of combination from the Dempster-Shafer's theory (Smets, 2007), denoted by

$\textcircled{\cap}$ and defined by:

$$\begin{aligned} m_{\textcircled{\cap}}(A) &= \sum_{\cap_i A^{(i)}=A} \left[\prod_{i=1}^{L} m_i(A^{(i)})\right], \forall A^{(i)} \in 2^{\Omega} \\ &= \sum_{\cap_i A^{(i)}=A} m_1(A^{(1)}) \times \cdots \times m_L(A^{(L)}). \end{aligned} \tag{2.14}$$

This rule is associative and commutative, which makes its implementation very easy. The normalized version of this rule, denoted by $\oplus$ and called Dempster combination rule, is defined by the following equation:

$$m_{\oplus}(A) = \frac{m_{\textcircled{\cap}}(A)}{1-\tau} \tag{2.15}$$

where

$$\tau = \sum_{\cap_i A^{(i)}=\emptyset} \left[\prod_{i=1}^{L} m_i(A^{(i)})\right].$$

The term τ reflects the degree of conflict between the mass functions. Aggregation of mass functions using the conjunctive rule or the Dempster rule leads to more informative ones. However, when the sources are conflicting or the definition of the masses is imprecise, these rules are not applicable.

Alternatively, the mass functions can be combined using the disjunctive rule of combination if at least one of the sources is reliable. The disjunctive rule of combination is defined by:

$$\begin{aligned} m_{\textcircled{\cup}}(A) &= \sum_{\cup_i A^{(i)}=A} \left[\prod_{i=1}^{L} m_i(A^{(i)})\right] \\ &= \sum_{\cup_i A^{(i)}=A} m_1(A^{(1)}) \times \cdots \times m_L(A^{(L)}). \end{aligned} \tag{2.16}$$

It leads to a less informative mass function by assigning higher masses to larger subsets of Ω.

2.4.2.4 Decision

In order to make a decision, the mass functions eventually obtained after combination must be transformed into a probability measure. A solution for the transformation consists of using the pignistic probability transformation,

which is defined by:

$$BetP(\{\omega_c\}) = \sum_{A \subseteq \Omega, \omega_c \in A} \frac{m(A)}{\mid A \mid (1 - m(\emptyset))}, \forall \omega_c \in \Omega. \tag{2.17}$$

$\mid A \mid$ represents the cardinality of A. The mass functions are transformed to singletons set only. Thus, the decision is assigned to the singleton $\{\omega_k\}$ that maximizes the pignistic probability, as follows

$$k = \arg \max_{1 \leq c \leq C} BetP(\{\omega_c\}).$$

Alternatively, the plausibility and belief functions can be used instead of the pignistic probability to make the decision on $\omega_c \in \Omega$,

- Maximum plausibility:

$$k = \arg \max_{1 \leq c \leq C} Pl(\{\omega_c\}),$$

- Maximum belief:

$$k = \arg \max_{1 \leq c \leq C} Bel(\{\omega_c\}).$$

2.5 Machine learning and kernel methods

Alternatively to the aforementioned information fusion techniques, there exist different pattern recognition algorithms from machine learning that can be considered in both types of fusion, decisions and feature fusion. The most known algorithms in machine learning are kernel methods, K-nearest neighbors (K-NN) and Support Vector Machine (SVM). In this thesis, K-NN and SVM algorithms are considered for comparison with the proposed models, and therefore validate the proposed ones. They are described in the following chapters. Herein, we present only kernel methods.

Kernel methods are a set of algorithms used for data analysis and fusion. They are characterized by mapping the input space into a higher dimensional feature space where the learning samples become linearly separable. Different algorithms are then applied on the new feature space to explore non-linear and hidden characteristics of the input data (Shawe-Taylor and Cristianini, 2004; Schölkopf et al., 2004). It comes out that these algorithms can be expressed as inner products between the mapped samples. Moreover,

these products can be replaced by kernel functions that are applied directly to the input space, which means that they do not need the knowledge of the mapping function to the feature space.

A kernel $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \Re$ is a symmetric similarity function of its input data as follows:

$$\kappa(x_1, x_2) = \kappa(x_2, x_1).$$

Let $x_i \in \mathcal{X}$ be a set of N entries, $i \in \{1, \ldots, N\}$. The linear kernel applied to any x_i, x_j is equal the product $x_i^T x_j$. In more general cases, each sample x is represented by a vector $\phi(x)$, using a mapping function $\phi(\cdot)$, and then the kernel applied to the inputs x_i and x_j is then defined as the inner product of their mapped samples. Then, the Gram matrix K is the matrix of size $N \times N$ having $\kappa(x_i, x_j)$ the value for $(i, j)th$ input that is computed as follows,

$$\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j). \tag{2.18}$$

The mapping function ϕ can be defined from an input space $\mathcal{X}$ to a Hilbert space $\mathcal{H}$ according to (Aronszajn, 1950) as

$$\phi : \mathcal{X} \rightarrow \mathcal{H}.$$

Then, the kernel is expressed as follows

$$\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}, \forall x_i, x_j \in \mathcal{X}, i, j \in \{1, \ldots, N\},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ represents the inner product in the Hilbert space. This product can be computed directly in the input space using the kernel function $\kappa(\cdot, \cdot)$ without any prior knowledge of the function $\phi(\cdot)$. The Hilbert space $\mathcal{H}$ is defined by the set of functions, such that

$$\mathcal{H} = \left\{ f : f(\cdot) = \sum_{i=1}^{N} \alpha_i \kappa(x_i, \cdot), x_i \in \mathcal{X}, \alpha_i \in \Re \right\}. \tag{2.19}$$

Let $f(\cdot)$, $g(\cdot)$ be two functions of the Hilbert space $\mathcal{H}$ given by:

$$f(\cdot) = \sum_{i=1}^{N} \alpha_i \kappa(x_i, \cdot) \text{ and } g(\cdot) = \sum_{j=1}^{N} \beta_j \kappa(x_j, \cdot).$$

The inner product in $\mathcal{H}$ between f and g is defined as follows:

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \beta_j \kappa(x_i, x_j) = \sum_{i=1}^{N} \alpha_i g(x_i) = \sum_{j=1}^{N} \beta_j f(x_j), \tag{2.20}$$

where $x_i, x_j \in \mathcal{X}$ and α_i, $\beta_j \in \Re$. The reproducing property of the kernel is expressed by:

$$\langle f, \kappa(x, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^{N} \alpha_i \kappa(x_i, x) = f(x),$$

where $\langle f, \kappa(x, \cdot) \rangle_{\mathcal{H}}$ is the inner product $\langle f, g \rangle_{\mathcal{H}}$ when taking $g = \kappa(x, \cdot)$. Given a function $\kappa(\cdot, \cdot)$, the corresponding Hilbert space is called its reproducing kernel Hilbert space.

There are two types of kernel functions, namely, projective kernels and radial basis functions (RBF) kernels. The projective kernel computes the inner product as a measure of similarity. The most popular projective kernel is the polynomial defined as

$$\kappa(x_i, x_j) = (\ell + x_i^T x_j)^p,$$

where $p \in \mathbb{N}_+$, and $\ell \geq 0$. The RBF kernels measure dissimilarities between features using distances. The most known RBF kernel is the Gaussian RBF kernel, computed as following

$$\kappa(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right),$$

where $\sigma > 0$ is the bandwidth of the kernel.

2.6 Conclusion

This chapter presented a brief review on the different data mining techniques for information fusion. Some of these techniques are used to propose and implement new models that fit the application and the type of data. Other techniques are applied in order to compare the performance obtained from the proposed models. As mentioned previously, the objective of this work is the surveillance of ARDS patients using different physiological signals, thus the necessity of information fusion models. Therefore, the next chapters propose models of fusion for the information provided from

each patient. Chapter 3 presents a decision fusion framework with new proposed methodologies for fusion. It proposes also a real-time decision-making model based on novelty detection. Thus a model is defined for each signal alone, then the individual decisions are combined using the proposed decision fusion methods. In chapter 4, an alternative information fusion approach is proposed on the feature level. Thus, various features are extracted from the signals and then an evidential framework is proposed to select and combine the extracted features. Finally, chapter 5 presents a new combination model that extends the previous one. In this chapter, a learning algorithm is implemented jointly with the belief functions theory in order to enhance the performance of the surveillance model. In addition, a multi-class differentiation is proposed to group the patients by severity of ARDS.

Chapter 3

ARDS Prediction using Novelty Detection

Contents

3.1 Introduction 48

3.2 Description of the approach 49

3.3 Novelty Detection Algorithm 51

3.4 Decision fusion 55

3.4.1 Linear fusion based on performance indexes 56

3.4.2 Linear fusion based on error rates 57

3.4.3 Kernel Ridge Regression-based fusion 58

3.5 Parameters estimation 60

3.5.1 Optimization of D and τ 60

3.5.2 Optimization of kernel parameters 61

3.6 Missing signals 61

3.7 Results 62

3.7.1 Materials 62

3.7.2 Invasive ventilated subjects 64

3.7.3 Non-invasive ventilated subjects 69

3.7.4 Comparison to state-of-the-art methods 71

3.8 Discussions 73

3.9 Conclusion 75

3.1 Introduction

ARDS is a critical condition that disturbs the respiratory system and may lead to death. Early identification of this syndrome is crucial for the implementation of preventive measures. The aim of this thesis is the prediction of the onset of this syndrome in real-time using physiological records of patients. We recall that ARDS is divided in the Berlin definition into three ARDS groups by severity levels according to the ratio of the partial pressure of oxygen to the fraction of inspired oxygen $PaO_2/FiO2$. Therefore, the Berlin definition defines mild ARDS having $200 < PaO_2/FiO2 \leq 300$ mmHg, moderate ARDS with $100 < PaO_2/FiO_2 \leq 200$ mmHg and severe ARDS having $PaO_2/FiO_2 \leq 100$ mmHg. In this chapter, we consider subjects developing the moderate and severe ARDS in the ARDS group; whereas subjects having mild ARDS are considered along with stable subjects in the non-ARDS group.

This chapter describes a novel approach, based on real-time analysis for early prediction of ARDS onset using only continuous physiological signals. ARDS has been linked to different clinical features such as abnormal breathing, rapid heart beat, low levels of oxygen in the blood and abnormal blood pressure. Therefore, signals related to these changes might be used for the implementation of such model, for example one could use high frequency signals such as electrocardiogram (ECG) signals and blood pressure signals, or time series data having lower frequency such as heart rate, respiratory rate, etc. Since the main objective of this thesis is the development of predictive models that can be implemented in the future in home surveillance systems, here comes the advantage of investigating time series data that might be collected from patients non-invasively at home. The proposed approach is subject-based, performing novelty detection on heart rate, respiratory rate, peripheral arterial oxygen saturation and mean airway blood pressure. This approach is a decision fusion technique, where first individual decisions are made according to the physiological signals, then these decisions are combined using fusion techniques. The subsequent paragraphs describe the proposed approach in details including the novelty detection algorithm, all the proposed decision fusion techniques and the optimization of parameters. Then, results are illustrated on both types of ventilated subjects. Finally, a discussion is made to interpret the obtained results.

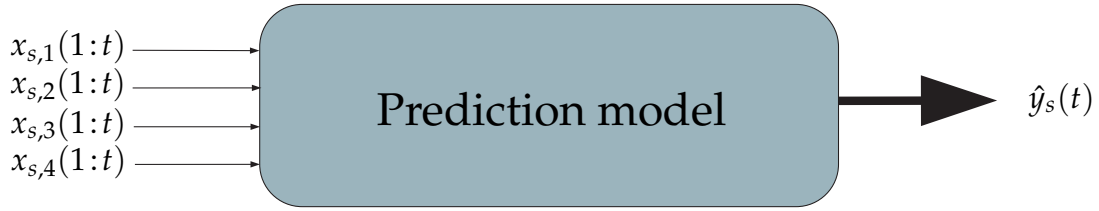


FIGURE 3.1: The inputs and output of the proposed prediction model

3.2 Description of the approach

In this chapter, we aim to predict at time t the onset of ARDS for a subject s based on its real-time recordings. Thus, we consider four types of physiological time series, which are the heart rate (HR), the respiratory rate (RR), the peripheral arterial oxygen saturation (SpO_2) and the mean airway blood pressure (MABP) having a sampling frequency of one sample per minute. Let $x_{s,i}(\ell)$, $i = 1, ..., 4$, denote respectively these time series for the subject s recorded at an instant ℓ, $\ell \in \{1, ..., t\}$ in minutes, and let $x_{s,i}(a : b) = (x_{s,i}(a), ..., x_{s,i}(b))$ denote the recorded data between time a and time b. Consider y_s is the label of subject s, that is, $y_s = +1$ if the subject develops ARDS and $y_s = -1$ otherwise. The prediction model is illustrated in Figure 3.1. It takes the four recorded time series $x_{s,i}(1:t)$ as inputs and yields $\hat{y}_s(t)$ as output with $\hat{y}_s(t) = -1$ if ARDS is not yet predicted, and $\hat{y}_s(t) = +1$ if ARDS is predicted. A warning is then generated for the subject s after a positive $\hat{y}_s(t)$.

The detailed prediction model of this chapter is given in the block diagram of Figure 3.2. The model proposes a novelty detection algorithm that considers an initial segment of each signal set as a stable state for each subject. Then, the remaining segments of the signals are analyzed to detect deviations from the stable state. A deviation detected in a given signal means that the ARDS is predicted according to this signal. This leads to real-time individual decisions that are combined afterwards using different decision fusion algorithms noted $\psi(\cdot)$, leading to a more accurate global decision. In the following, the subjects that develop ARDS, with $y_s = +1$, are called ARDS subjects, whereas non-ARDS subjects have $y_s = -1$. Next sections describe the Novelty detection algorithm and the different fusion techniques.

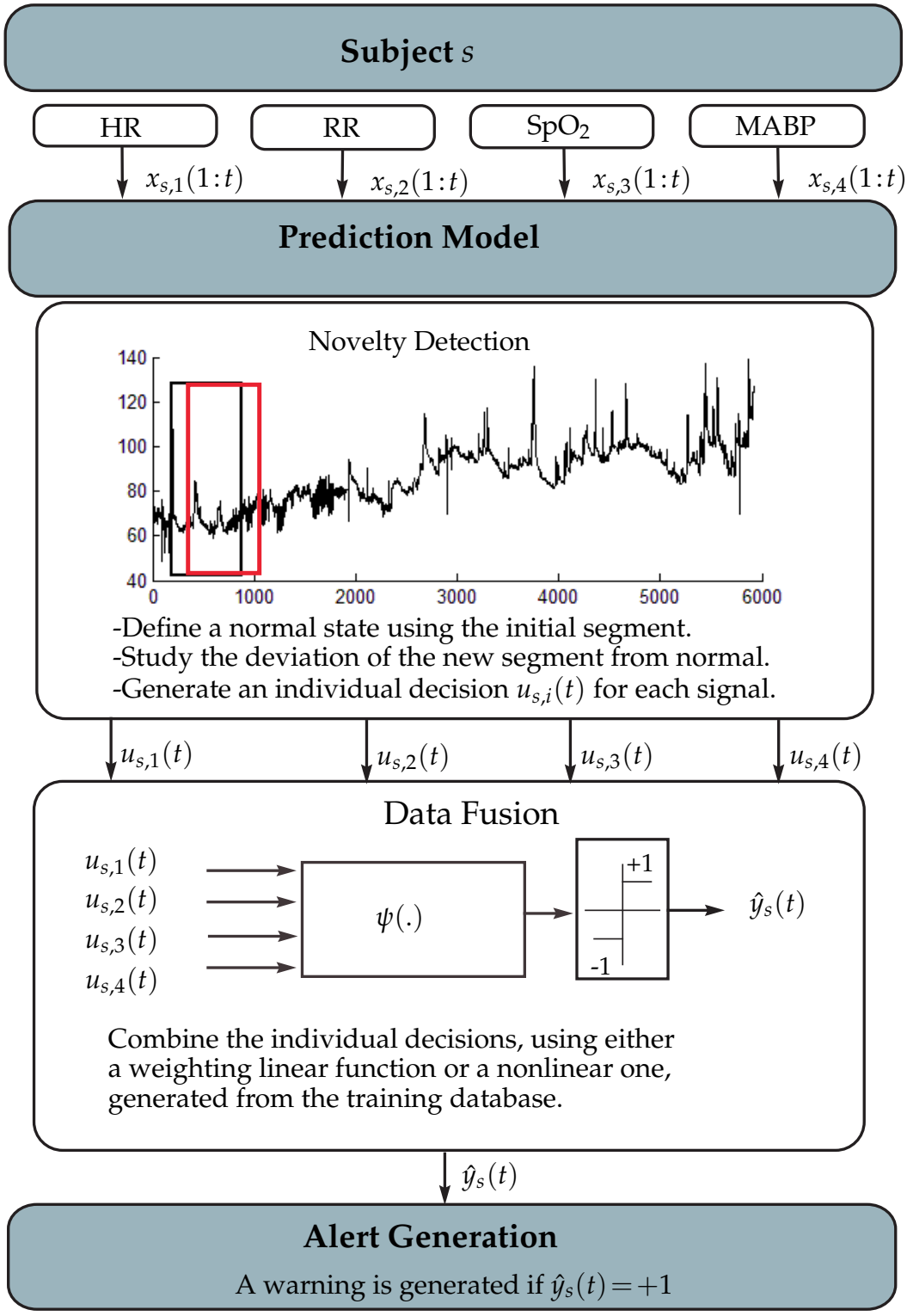


FIGURE 3.2: The general block diagram of the real-time prediction model for a subject s using novelty detection

3.3 Novelty Detection Algorithm

In order to generate decisions from each signal, a decision maker is developed based on novelty detection concept. Novelty detection or one-class classification is a challenging issue in machine learning. It consists of constructing a model of normality by deriving a classification boundary that separates this model from possible outliers (Pimentel et al., 2014). Novelty detection has been applied in different applications such as fault detection (Cheng et al., 2008), medical applications (Roberts, 2000; Clifton et al., 2009), process control (Knorr, Ng, and Tucakov, 2000; Perner, 2009) and others. Wide reviews on novelty detection can be found in (Markou and Singh, 2003a; Markou and Singh, 2003b).

We propose in this chapter a subject-based approach that identifies for each subject a normal state according to a first segment of its recorded signals and then detects in real-time the divergence from this normal state in the following recorded data. The study is conducted for each signal i for every monitored subject s, given the signals in real-time t as $x_{s,i}(1:t)$. Let $\tau < t$ be the considered duration of the first segment selected from each signal. τ could be taken equal to 6, 12 or even 24 hours, that is, 360, 720, or 1440 minutes respectively, or it can be determined by the medical team. An optimization approach is proposed in the following to set its value.

Hence, for the signal i of a subject s, $x_{s,i}(1:\tau)$=$(x_{s,i}(1), ..., x_{s,i}(\tau))$, called *first segment*, is used to define its normal state; then a sliding window of the same length τ is considered for each segment of new recorded data to detect its divergence from the first segment. In real-time, $x_{s,i}(t-\tau+1:t)$, called *latter segment*, is then used for divergence study. A first solution for defining the normal state consists of finding the distribution of $x_{s,i}(1:\tau)$, and then checking whether $x_{s,i}(t-\tau+1:t)$ follows the same distribution or not using some hypothesis statistical tests. However, this solution is worthless due to its overfitting and thus oversensitivity. Alternatively, we propose a distance-based algorithm that consists of comparing the number of outliers of the latter segment to the one of the first segment. The recorded data is called an *outlier* if it lies far away from the mean of the normal data. The outliers are identified according to the two following definitions.

Definition 1: For normally distributed data, the proportion of data lying

TABLE 3.1: Examples of distance functions between two vectors x and y

Distance Name	Distance function
Euclidean distance	$\sqrt{\sum_{j=1}^{N} \frac{(x_j - y_j)^2}{\sigma_j^2}}$
Mahalanobis distance	$\sqrt{(\overrightarrow{x} - \overrightarrow{y})^T S^{-1} (\overrightarrow{x} - \overrightarrow{y})}$
Canberra distance	$\sum_{j=1}^{N} \frac{\lvert x_j - y_j \rvert}{\lvert x_j \rvert + \lvert y_j \rvert}$
Manhattan distance	$\sum_{j=1}^{N} \lvert x_j - y_j \rvert$

three or more standard deviations from their mean is less than 0.3%. These data are called outliers. For each signal i, the mean $\mu_{s,i}$ and the standard deviation $\sigma_{s,i}$ of its first segment are computed as follows,

$$\mu_{s,i} = \frac{1}{\tau} \sum_{\ell=1}^{\tau} x_{s,i}(\ell), \quad \sigma_{s,i} = \sqrt{\frac{1}{\tau} \sum_{\ell=1}^{\tau} (x_{s,i}(\ell) - \mu_{s,i})^2}. \tag{3.1}$$

Then the outliers are detected in the latter segment by counting the number of points lying beyond $\pm 3\sigma_{s,i}$ from $\mu_{s,i}$. A subject s is considered ARDS according to the signal i if the number of the counted outliers is higher than 1% of the length τ of the latter segment, and an individual decision $u_{s,i}(t) = +1$ is released; otherwise $u_{s,i}(t) = -1$. This algorithm works well when data is normally distributed. However, in most real applications, it is unknown whether data follows a predefined distribution especially when working with biomedical signals that are known to have random distributions. Therefore, a more general definition for the distance-based model is proposed in the following.

Definition 2: This definition addresses the limitation of the Definition 1, which is assuming data are normally distributed. It defines a distance D, each point lying beyond D from the normal data is called a D-outlier. Let $f(.,.)$ be the function that computes the distance between any data point to a set of data. By applying this to our problem, a point $x_{s,i}(\ell)$ is a D-outlier if its distance $f(x_{s,i}(\ell), x_{s,i}(1:\tau))$ to the set $x_{s,i}(1:\tau)$ is higher than D. Details on the definition of D and its optimization are given later on. The distance-based novelty detection algorithm is illustrated in Algorithm 1 and described in the following.

Algorithm 1: Novelty Detection Algorithm

Input: The signal $x_{s,i}(1:t)$ of the subject s, D_i and τ
Output: $u_{s,i}(t)$

Step 1: Define the first segment $x_{s,i}(1:\tau)$ and the latter segment $x_{s,i}(t-\tau+1:t)$;

Step 2: Compute the characteristics of $x_{s,i}(1:\tau)$:
$\mu_{s,i} \longleftarrow \frac{1}{\tau}\sum_{\ell=1}^{\tau} x_{s,i}(\ell)$;
$\sigma_{s,i} \longleftarrow \sqrt{\frac{1}{\tau}\sum_{\ell=1}^{\tau}(x_{s,i}(\ell)-\mu_{s,i})^2}$;

Step 3:
for $\ell = 1, ..., \tau$ **do**
  $f\left(x_{s,i}(\ell), x_{s,i}(1:\tau)\right) \longleftarrow \frac{|x_{s,i}(\ell)-\mu_{s,i}|}{\sigma_{s,i}}$;
  if $f(x_{s,i}(\ell), x_{s,i}(1:\tau)) \geq D$ **then**
    $\mathbb{I}(x_{s,i}(\ell)) \longleftarrow 1$;
  end
end
$p_{s,i} \longleftarrow \frac{\sum_{\ell=1}^{\tau}\mathbb{I}(x_{s,i}(\ell))}{\tau}$;

Step 4:
for $\ell = t-\tau+1, ..., t$ **do**
  $f\left(x_{s,i}(\ell), x_{s,i}(1:\tau)\right) \longleftarrow \frac{|x_{s,i}(\ell)-\mu_{s,i}|}{\sigma_{s,i}}$;
  if $f(x_{s,i}(\ell), x_{s,i}(1:\tau)) \geq D$ **then**
    $\mathbb{I}(x_{s,i}(\ell)) \longleftarrow 1$;
  end
end
$r_{s,i}(t) \longleftarrow \frac{\sum_{\ell=t-\tau+1}^{t}\mathbb{I}(x_{s,i}(\ell))}{\tau}$;

Step 5: if $r_{s,i}(t) > p_{s,i}$ **then**
  $u_{s,i}(t) \longleftarrow +1$;
else
  $u_{s,i}(t) \longleftarrow -1$;
end

The algorithm aims to detect the deviation of a new segment $x_{s,i}(t-\tau+1:t)$ from the stable one $x_{s,i}(1:\tau)$. There exist different distance functions to compute this distance such as the Euclidean distance, the Mahalanobis distance and others (Duran and Odell, 1974). An example of some distance functions between two vectors x and y is provided in Table 3.1 where S is the covariance matrix and σ is the standard deviation. In this work, the Normalized Euclidean distance is used since it is the best distance function suited to this application because data are scalar for a single signal. It is defined as follows

$$f\left(x_{s,i}(\ell), x_{s,i}(1:\tau)\right) = \frac{|x_{s,i}(\ell) - \mu_{s,i}|}{\sigma_{s,i}}, \tag{3.2}$$

where $\mu_{s,i}$ and $\sigma_{s,i}$ are the mean and the standard deviation of the first segment respectively, computed as shown in Eq. (3.1). Let $\mathbb{I}(x_{s,i}(\ell))$ be an indicator yielding 1 if $x_{s,i}(\ell)$ is a D-outlier and 0 otherwise, that is,

$$\mathbb{I}\left(x_{s,i}(\ell)\right) = \begin{cases} 1, \text{ if } f(x_{s,i}(\ell), x_{s,i}(1:\tau)) \geq D; \\ 0, \text{ otherwise.} \end{cases} \tag{3.3}$$

Hence, the D-outliers within the latter segment $x_{s,i}(t-\tau+1:t)$ are counted and their ratio $r_{s,i}(t)$ over the length of the latter segment is computed,

$$r_{s,i}(t) = \frac{\sum_{\ell=t-\tau+1}^{t} \mathbb{I}(x_{s,i}(\ell))}{\tau}. \tag{3.4}$$

Let $p_{s,i}$ be the proportion of D-outliers within the first segment of signal i for subject s. In other words,

$$p_{s,i} = \frac{\sum_{\ell=1}^{\tau} \mathbb{I}(x_{s,i}(\ell))}{\tau}.$$

Then, $r_{s,i}(t)$ is compared to $p_{s,i}$ and an individual decision is generated,

$$u_{s,i}(t) = \begin{cases} +1, & \text{if } r_{s,i}(t) > p_{s,i}; \\ -1, & \text{otherwise.} \end{cases} \tag{3.5}$$

A positive $u_{s,i}(t)$ means that the considered subject will develop ARDS in the near future, according to its signal i.

The output of the algorithm of Definition 2 gives four individual decisions for each subject s. We propose, in the following section, different decision fusion functions to combine the individual decisions and obtain a global decision on the health state of the subjects.

3.4 Decision fusion

In this section we describe the decision fusion technique that helps to combine the obtained individual decisions, thus to improve the performance of the classification. In order to formulate the fusion technique, we define a fusion function $\psi(\cdot)$ that takes as input the vector of individual decisions $\boldsymbol{u}_s(t) = (u_{s,1}(t), ..., u_{s,4}(t))$ of a subject s, and gives as output a value as close as possible to the true state of the subject y_s, that is, $+1$ if the subject develops ARDS and -1 if it remains in a stable state. The final decision $\hat{y}_s(t)$ consists then of the sign of the response of ψ, as follows:

$$\hat{y}_s(t) = sign(\psi\left(\boldsymbol{u}_s(t)\right)) = \begin{cases} +1, & \text{if } \psi\left(\boldsymbol{u}_s(t)\right) > 0, \\ -1, & \text{otherwise.} \end{cases} \tag{3.6}$$

Figure 3.3 illustrates the model of decision fusion that follows the novelty detection. It takes the individual decisions, performs one of the proposed decision fusion functions and gives the final decision $\hat{y}_s(t)$. We propose linear and nonlinear decision fusion functions; linear functions are based on performance and on error rates while the nonlinear fusion function is based on kernel functions. These fusion functions consist of weighting rules to compute the measures of confidence or weights. The weights are computed conditionally on each state, thus the reliability of the decision provided is taken into consideration. For all proposed fusion algorithms, a database of signal records of ARDS and non-ARDS subjects is needed to compute the fusion parameters.

Since we are combining decisions provided from different types of signals, the correlation between the four signals is tested before applying the fusion algorithms. The aim of the correlation test is to prevent any duplication of information. Indeed, if two signals are correlated then these signals provide similar information. Therefore, one of them can be excluded from the fusion process.

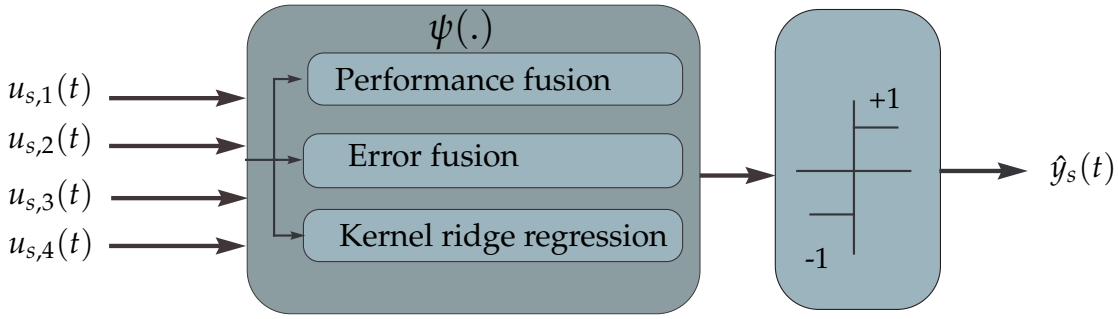


FIGURE 3.3: Decision fusion model

3.4.1 Linear fusion based on performance indexes

Each signal of the database goes through a preliminary classification using the proposed novelty detection algorithm to test its performance for detecting the pathology. An ARDS subject is supposed to be correctly identified according to a signal i if it has $u_{s,i}(\ell) = +1$ at any time ℓ of the signal; whereas a non-ARDS subject is correctly identified if $u_{s,i}(\ell) = -1$ for all times ℓ. The performance indexes consist then of the sensitivity Se, known as the true positive rate, and the specificity Sp or the true negative rate that are computed for each signal i separately as follows:

$$Se_i = \frac{\text{Number of correctly identified ARDS subjects for signal } i}{\text{Total number of ARDS subjects}},$$

$$Sp_i = \frac{\text{Number of correctly identified non-ARDS subjects for signal } i}{\text{Total number of non-ARDS subjects}}.$$

These performance indexes are then used to determine a weight noted a_i, $i = 1, ..., 4$, that will be used jointly with the decision of each signal without taking into consideration the actual label of the subject. Hence, the function $\psi(\boldsymbol{u}_s(t))$ is defined by the following equation,

$$\psi(\boldsymbol{u}_s(t)) = \sum_i a_i u_{s,i}(t). \tag{3.7}$$

Several weighting functions based on performance indexes have been proposed in the literature. To apply such methods, increasing functions of sensitivity and specificity are proposed to compute the weights a_i, such as in (Meng et al., 2010; Mourão, 2014). Different examples of such fusion techniques were presented in Chapter 2. The performance indexes are frequently

used as weighting coefficients for decision fusion. Therefore, we consider in this chapter the performance indexes. Thus, we apply it in its simplest form, which is defined as follows:

$$a_i = \begin{cases} Se_i, & \text{if } u_{s,i}(t) > 0, \\ Sp_i, & \text{otherwise.} \end{cases} \tag{3.8}$$

If the analysis of signal i of subject s leads to a positive decision, then the assigned weight is the performance index Se_i of the positive group from signal i. Whereas, if the analysis leads to a negative decision, then the assigned weight is the performance index of the negative or non-ARDS group Sp_i. Then, the final decision $\hat{y}_s(t)$ is obtained according to Eq. (3.6).

3.4.2 Linear fusion based on error rates

In this paragraph, we propose an alternative fusion function based on the error rate of the individual decisions related to each signal. Let $\epsilon_{i,+}$ and $\epsilon_{i,-}$ be respectively the error rates of positive and negative estimations. $\epsilon_{i,+}$ is defined as the percentage of subjects identified to develop ARDS with $u_{s,i}(\ell) = +1$ at any time ℓ while they will not, thus having a label $y_s = -1$; whereas $\epsilon_{i,-}$ is the percentage of the subjects identified as non-ARDS with $u_{s,i}(\ell) = -1$ for all times ℓ but they will develop ARDS and have a label $y_s = +1$. $\epsilon_{i,+}$ and $\epsilon_{i,-}$ are computed as follows:

$$\epsilon_{i,+} = \frac{\text{Number of non-ARDS subjects identified as ARDS}}{\text{Total number of subjects identified as ARDS}},$$

$$\epsilon_{i,-} = \frac{\text{Number of ARDS subjects identified as non-ARDS}}{\text{Total number of subjects identified as non-ARDS}}.$$

These quantities are inversely proportional to the reliability of an individual decision. The smaller the error rate of a decision is, the higher its confidence or weight should be. The weights a_i are then computed from the error rates, using different decreasing weighting functions $h(.)$. The goal is to give more influence to signals performing better than others in the classification step. Let ϵ_i be a variable representing the error rates as follows:

$$\epsilon_i = \begin{cases} \epsilon_{i,+}, & \text{if } u_{s,i} > 0, \\ \epsilon_{i,-}, & \text{otherwise.} \end{cases}$$

Then the weights are defined as $a_i = h(\epsilon_i)$. First, a simple function $h(\epsilon_i)$ is proposed using the error rates, as follows:

$$h(\epsilon_i) = 1 - \epsilon_i. \tag{3.9}$$

Another weighting function is proposed based on the weighted majority voting or WMV that is described in the previous chapter, presented by Benediktsson (Meng et al., 2010). It is modified to the error rate approach. Here, $h(\epsilon_i)$ performs a normalization of the error rates among all signals.

$$h(\epsilon_i) = \frac{(1 - \epsilon_i) - \epsilon_{\min}}{\epsilon_{\max} - \epsilon_{\min}}. \tag{3.10}$$

$\epsilon_{\min}$ and $\epsilon_{\max}$ are the minimum and maximum error rates among positive and negative error rates for all signals i, $\epsilon_{\min} = \min_i(\min(\epsilon_{i,+}, \epsilon_{i,-}))$ and $\epsilon_{\max} = \max_i(\max(\epsilon_{i,+}, \epsilon_{i,-}))$. Then, the fusion is performed using Eq. (3.7).

3.4.3 Kernel Ridge Regression-based fusion

In this section, we propose a fusion function based on the kernel ridge regression algorithm (Vapnik, 1998; Mahfouz et al., 2014). The kernel ridge regression consists of a non-parametric technique defining the function $\psi(\cdot)$ that assigns to each vector $\boldsymbol{u}_s(t) = (u_{s,1}(t), ..., u_{s,4}(t))$ a value as close as possible to the correct label y_s of the subject s. The constructed database of ARDS and non-ARDS subjects' signals is used to do this. Indeed, the proposed novelty detection algorithm is run for all the signals for all the subjects and individual decisions vectors are obtained at regular time steps for each subject. For ARDS subjects, the last decision vectors are selected since they are the closest to the ARDS onset; whereas for non-ARDS subjects the first decision vectors are considered since they represent at best their stability. More than one decision vector could also be selected for each subject.

A learning database of size N including N decision vectors $\boldsymbol{u}_n$ with their corresponding labels y_n, $n \in \{1, ..., N\}$ is obtained. The function $\psi(\cdot)$ can be evaluated using the reproducing kernel $\kappa(\cdot, \cdot)$ in the reproducing kernel Hilbert space $\mathcal{H}$. $\psi(\cdot)$ is obtained by minimizing the empirical error between the model's output $\psi(\boldsymbol{u}_n(k))$ and the desired outputs y_n as follows:

$$\psi = \arg\min_{\psi \in \mathcal{H}} \mathcal{L}\left((y_1, \psi(\boldsymbol{u}_1(t))), \ldots, (y_N, \psi(\boldsymbol{u}_N(t)))\right) + \eta\Omega(\|\psi\|^2_{\mathcal{H}}), \tag{3.11}$$

$\mathcal{L}(\cdot)$ is a cost function and $\Omega(\cdot)$ is strictly monotonically increasing function in $\Re^+$. The second part of the equation is a regularization term with $\|\cdot\|_{\mathcal{H}}$ is the norm in the Hilbert space and $\eta > 0$ is the smoothness parameter that controls the tradeoff between the error and the complexity of the solution.

The cost function $\mathcal{L}(\cdot)$ can take different forms and leads to different optimization problems. In the kernel ridge regression, the function $\mathcal{L}(\cdot)$ is computed using the mean quadratic error between the model's outputs $\psi(\boldsymbol{u}_n(t))$ and the desired outputs y_n:

$$\mathcal{L}\left((y_1, \psi(\boldsymbol{u}_1(t))), \ldots, (y_N, \psi(\boldsymbol{u}_N(t)))\right) = \frac{1}{N}\sum_{n=1}^{N}\left(y_n - \psi(\boldsymbol{u}_n(t))\right)^2.$$

Moreover, the regularization term $\Omega(\|\psi\|^2_{\mathcal{H}})$ is taken in its simplest form $\|\psi\|^2_{\mathcal{H}}$. The optimization problem can thus be written as follows

$$\psi = \arg\min_{\psi\in\mathcal{H}} \frac{1}{N}\sum_{n=1}^{N}\left(y_n - \psi(\boldsymbol{u}_n(t))\right)^2 + \eta\|\psi\|^2_{\mathcal{H}}. \tag{3.12}$$

Hence, the optimal function can be written according to the representer theorem (Scholkopf, Herbrich, and Smola, 2001),

$$\psi(\cdot) = \sum_{n=1}^{N}\alpha_n\kappa\left(\boldsymbol{u}_n, \cdot\right), \tag{3.13}$$

where $\kappa(\cdot,\cdot)$ is the reproducing kernel associated to the Hilbert space $\mathcal{H}$, and α_n are parameters to be estimated. By replacing $\psi(\cdot)$ in Eq. (3.12) by its expression in Eq. (3.13), we obtain a dual optimization problem:

$$\alpha = \arg\min_{\alpha}(Y - K\alpha)^T(Y - K\alpha) + \eta N\alpha^T K\alpha,$$

K is the Gram N-by-N matrix having $K(n,m) = \kappa\left(\boldsymbol{u}_n, \boldsymbol{u}_m\right)$ as its (n,m)-th element. Y is the vector of length N of the labels y_n. The solution of this quadratic regression problem is given by computing the derivative of the objective function with respect to α and setting it to zero, thus we obtain:

$$-KY + K^2\alpha + \eta NK\alpha = 0,$$

The solution $\alpha = (\alpha_1, ..., \alpha_N)$ is then given by,

$$\alpha = (K + \eta N I_N)^{-1} Y, \tag{3.14}$$

where I_N is the N-by-N identity matrix. Once the parameters α are computed, the functions $\psi(\cdot)$ can be defined using Eq. (3.13).

Several kernel functions $\kappa(\cdot, \cdot)$ exist in the literature as mentioned in Section 2.5 of Chapter 2. In this work, we consider the Gaussian kernel defined by:

$$\kappa(\boldsymbol{u}_n, \boldsymbol{u}_m) = \exp\left(\frac{-\|\boldsymbol{u}_n - \boldsymbol{u}_m\|^2}{2\sigma_\kappa^2}\right), \tag{3.15}$$

where σ_κ is the kernel bandwidth controlling with the parameter η the noise tolerance and the degree of smoothness. The estimation of the values of σ and η is presented in the following section.

3.5 Parameters estimation

In this section, we describe the optimization of all the parameters presented in the model of novelty detection such as the distance D and the size of the window τ, as well as the parameters of the kernel ridge regression such as the smoothness parameter η and the kernel bandwidth σ_κ.

3.5.1 Optimization of D and τ

As we already mentioned, the distance D and the length of the first segment τ could be set to fixed values such as 3, 3.5, 4... for D and 6, 8, 12 or 24 hours for τ, and the user chooses the values needed according to each situation. However, we propose in this work to optimize their values. Therefore, the signals database with the ARDS and non-ARDS subjects is divided into a training set and a testing set. This division could be performed using a Leave-one-out cross-validation (LOOCV) if the dataset is small or using a k-fold cross-validation when working with high number of subjects. Then, the training set is used to perform the optimization. τ and D are varied respectively within $\{60, 120, ..., 1440\}$ and $\{0.1, 0.2, ..., 0.9\}$. For each couple of values of τ and D, the signals of the training subjects are used with the proposed approach to compute the performance indexes and the error rates per signal, and the testing subjects are used with these indexes for validation. Hence, the sensitivity Se and the specificity Sp of the whole approach are

estimated for each couple of values of τ and D. The optimal τ_o and D_o are obtained by maximizing the Youden index $= Se + Sp - 100$.

3.5.2 Optimization of kernel parameters

In order to perform the aforementioned kernel decision fusion technique, it is necessary to determine the concerned parameters. Hence, the performance of the kernel ridge regression depends on the choice of the parameters η and σ_κ. Thus, an optimization of their values is also handled in this work in order to obtain the optimum performance. The values η and σ_κ are obtained by maximizing the Youden index using a grid search on $\eta = 2^r$ with $r \in \{-10, -9, ..., 0\}$ and on different ranges of values for σ_κ, $\sigma_\kappa = 10^{r'}$ and $\sigma_\kappa = -1 * r'$ with $r' \in \{-5, -4, ..., -1\}$. The cross-validation technique is also performed and the couple of values leading to the best Youden index are selected.

3.6 Missing signals

In the previous section, different data fusion functions were proposed to combine the set of the extracted time series. The aim of the fusion was to aggregate the individual decisions obtained from these time series in order to enhance the prediction's performance. In many cases, some signals may be missing for the reason of a sudden movement, the displacement of electrodes, or other logistic reasons. Moreover, it is possible that one or more signals are less discriminative than others and might compromise the prediction of ARDS. Therefore, we are interested in the fusion of different subcombinations of the signals used in this study. All of the possible subsets signals are considered, thus all the possibilities of missing signals are covered. For each of these subsets, fusion functions are performed, and then sensitivity and specificity are computed to indicate the capacity of each subset fusion to correctly identify ARDS from non-ARDS patients. Afterwards, the subset of signals giving higher performance is identified, thus the combination of these signals is more relevant to discriminate ARDS patients than other combinations.

3.7 Results

This section presents the results of the proposed D distance-based novelty detection algorithm followed by the decision fusion techniques. As a summary, we proposed one fusion function based on performance indexes, two different fusion functions based on error rates and a kernel ridge regression fusion function. A comparison between these fusion functions will be presented in the following in term of accuracy and early prediction of ARDS.

3.7.1 Materials

As mentioned in Chapter 1, four signals are extracted from the MIMIC II database to predict the onset of ARDS. These signals are the heart rate (HR), the respiratory rate (RR), the arterial oxygen saturation (SpO_2) and the mean airway blood pressure (MABP). They have a frequency of one sample per minute. Moreover, two types of subjects are extracted, namely, invasive ventilated subjects and non-invasive ventilated subjects. The first type was selected according to previous selection procedures in the literature. However, the selected subjects have a continuous monitored ventilation thus they have different situations than patients at home. Therefore, the second selection procedure was handled to obtain subjects having more similar situations of patients monitored at home.

From the invasive ventilated subjects, we obtained 38 ARDS subjects and 18 non-ARDS subjects. Among these 56 subjects, there exist some subjects with very short duration signals, thus these subjects were eliminated from the study. This leads to 40 subjects, of which 26 are ARDS and 14 are non-ARDS. Conversely, from the non-invasive ventilated subjects, we get 140 ARDS subjects and 135 non-ARDS subjects. Only 72 ARDS subjects are included in the study, since they have records before ARDS occurrence until its onset. Some subjects were removed due to noise or too short signals, and then the number of subjects was balanced. This leads finally to 50 ARDS and 50 non-ARDS subjects. Figures 3.4 and 3.5 present examples of the four extracted signals for an ARDS and a non-ARDS subjects, respectively. Therefore, the results are divided into two main sections, the first one illustrates all the results of the model on the invasive ventilated subjects and the second section presents the results of the non-invasive ventilated subjects.

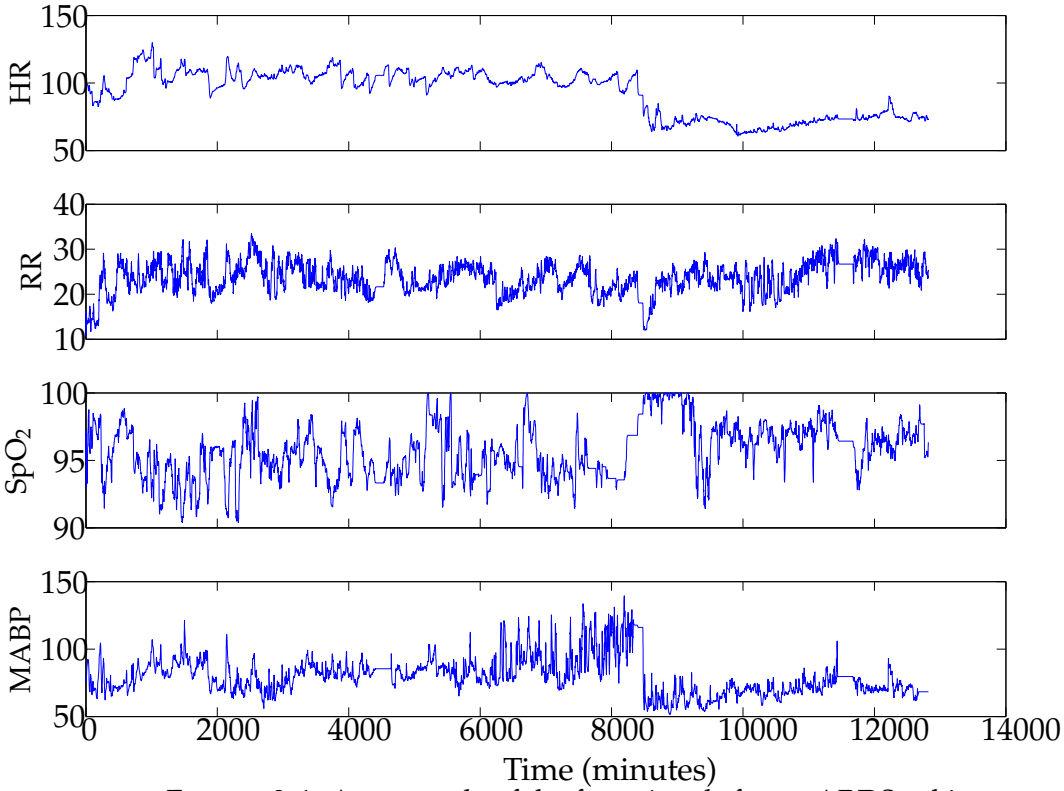


FIGURE 3.4: An example of the four signals for an ARDS subject

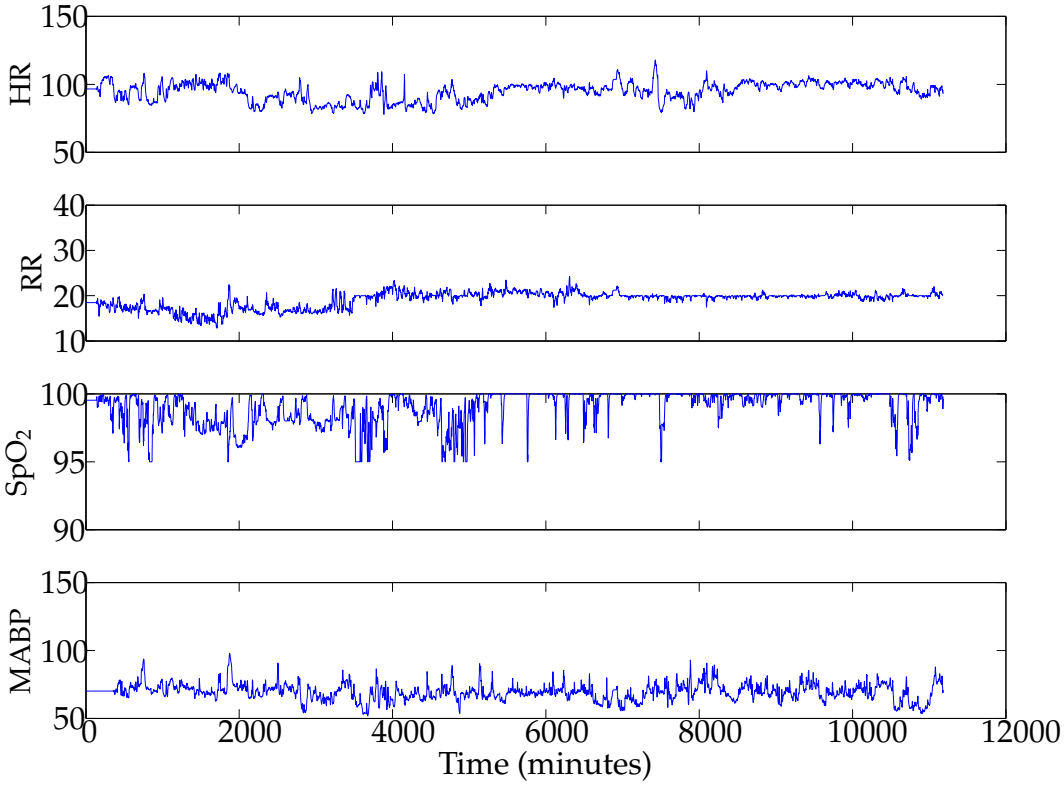


FIGURE 3.5: An example of the four signals for a non-ARDS subject

TABLE 3.2: Matrix of correlation coefficients between signals

	HR	RR	SpO_2	MABP
HR	–	0.24	-0.11	0.12
RR	0.24	–	-0.17	0.16
SpO_2	-0.11	-0.17	–	-0.01
MABP	0.12	0.16	-0.01	–

3.7.2 Invasive ventilated subjects

The proposed approach is validated at first using the invasive ventilated subjects. From the selection procedure of the invasive ventilated subjects described in Section 1.5.2.1, we obtained 40 subjects, of which 26 are ARDS and 14 are non-ARDS. To overcome the limited number of subjects and avoid overfitting, a leave-one-out cross-validation (LOOCV) procedure is considered for evaluation. Indeed, 39 are selected for the training and only one remains for evaluation. This selection is repeated for all the subjects, leading to 40 sets of data and thus an evaluation of the approach for 40 times.

First of all, the correlation between the four signals of the subjects is tested. To do this, the correlation coefficients between each pair of signals are computed and the results are presented as a matrix of correlation coefficients in Table 3.2. These coefficients take values in the interval $[-1, 1]$. The closer the coefficient is to -1 or 1, the stronger the correlation between the signals is. Table 3.2 shows that all correlation coefficients have values near 0, thus there is a weak linear relationship between signals. Therefore, the four signals are used in this work.

3.7.2.1 Results of Definition 1

The Definition 1 assumes that data are normally distributed, thus outliers in signals are at a distance higher than 3σ from the mean of the stable state. A subject is then considered as ARDS if it has outliers more than 1% of the total length of the signal. In order to evaluate the performance of the approach, we first start with an optimization of all the parameters as described in Section 3.5. Therefore, we sweep all the possible values of $\tau \in \{60, 120, \ldots, 1440\}$ minutes, as well as the parameters $\eta \in 2^{\{-10,-9,\ldots,0\}}$ and $\sigma_\kappa \in 10^{\{-5,-4,\ldots,-1\}} \cup \{1, 2, \ldots, 5\}$ in the case of the fusion based on kernel ridge regression. The performances Se and Sp of the method are computed for each couple of values on the training data to find the couple that maximizes the Youden index, the optimized parameters are then applied on

TABLE 3.3: Performance of the Definition 1

Definition1 +	Se (%)	Sp (%)	Youden Index
Performance fusion	96.15	28.57	24.72
Simple Error fusion	96.15	35.71	31.86
Error WMV fusion	96.15	21.42	17.57
KRR	77.33	58.08	35.41

the test subjects and the test performance is computed. This procedure is done separately for each fusion technique.

The algorithm of Definition 1 was considered jointly with all the decision fusion algorithms proposed earlier in this chapter. Table 3.3 presents the results obtained from each fusion technique with the decision maker based on Definition 1. Note that KRR indicates the kernel ridge regression fusion. As shown, performance and error rate based fusions gave bad specificity. However KRR based fusion applied with the model of Definition 1 gave a good performance with a sensitivity of 77.33%, specificity of 58% and a Youden index of 35.41%.

3.7.2.2 Results of the D distance-based approach

In this section, we consider the Definition 2 of the proposed novelty detection algorithm that considers a threshold distance D defining the limits of the stable state. Thus, each point lying beyond D from the initial segment of length τ is called a D-outlier. Giving individual decisions, the proposed algorithm is followed by the decision fusion functions. Hence, an optimization of the Algorithm parameters is considered. Thus, we optimize the values of the segments length τ and the distance D within $\{60, 120, \ldots, 1440\}$ minutes and $\{0.1, 0.2, \ldots, 5\}$, respectively. The parameters of the kernel ridge regression η and σ_κ are also optimized as described in the previous section.

Table 3.4 presents the optimal windows length τ_o and distance D_o that correspond to the highest Youden Index in the training phase for each fusion technique. It also presents the mean and standard deviations of the performances in the training phase through the cross validation procedure as well as the performance of the test subjects. It is shown that the training performances barely fluctuate in the performance-, simple error- and KRR-based fusion techniques with $(\Delta Se, \Delta Sp)$ values respectively (2.45, 2.17),

TABLE 3.4: Performances of the optimal τ_o in (hours) and D_o obtained for the proposed model with each decision fusion technique

Novelty Model + fusion based on	τ_o	D_o	**Training** Se $\pm\Delta$Se	Sp $\pm\Delta$Sp	**Testing** Se	Sp
Performance	10	4.4	69.23 ± 2.45	92.86 ± 2.17	69.23	78.57
Simple error rate	10	4.4	84.62 ± 1.17	71.43 ± 2.08	84.62	71.43
WMV error rate	9	7.2	74.87 ± 7.4	69.27 ± 10.2	65.38	71.42
KRR	11	4.6	65.57 ± 2.05	99.04 ± 2.57	65.38	100

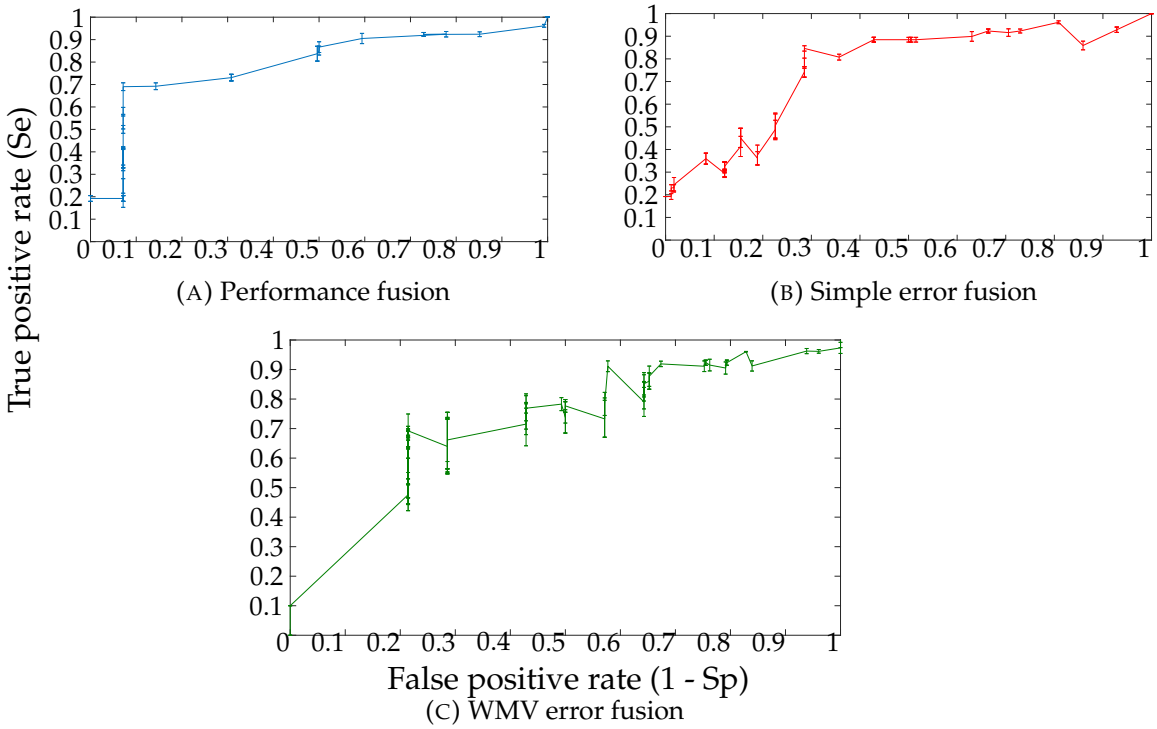
TABLE 3.5: Youden indexes over the training and testing sets for the proposed model with each decision fusion technique

Novelty Model + fusion based on	**Training** Youden index	**Testing** Youden index
Performance	62.09	47.8
Simple error rate	56.05	56.05
WMV error rate	44.14	36.8
KRR	64.61	65.38

(1.17, 2.08) and (2.05, 2.57); whereas the WMV error-based fusion presents high fluctuations of training performances with $\Delta Se = 7.4$ and $\Delta Sp = 10.2$.

The Youden indexes of the proposed model with each of the fusion algorithms over the training sets and the testing subjects are given in Table 3.5. The results show that the kernel ridge regression fusion performs better than the linear techniques over both training and testing subjects, and that the performance-based fusion remains better than the error rate-based for linear fusion in terms of Youden index over the training sets. Having a specificity of 100% with kernels in Table 3.4 over the testing subjects means that all generated alerts are worth it, without any false alarm. Note that it is also possible to increase the sensitivity of the method, at the cost of the specificity, by tuning differently the parameters.

Figure 3.6 illustrates subplots of the ROC curves of the proposed algorithm with each of the fusion functions by taking $\tau = \tau_o$ and varying D within its corresponding interval. ROC curve plots the true positive rate, known as the sensitivity against the false positive rate computed by $1 - Sp$. The subplots illustrate the curves for the linear fusion techniques based on the performances in (a), the simple and WMV functions of the error rates in (b) and (c) respectively. The areas under the curves (AUC) are equal to



(A) Performance fusion

(B) Simple error fusion

(C) WMV error fusion

FIGURE 3.6: ROC curves for Novelty algorithm followed by data fusions based on performances in (a), and both Simple and WMV error rates functions in (b) and (c), respectively

79.11%, 75.67% and 70.98% respectively. It is shown that the ROC curves in Figures 3.6 (a) and (b) have small error bars, thus the Youden indexes of fusions based on performance and simple errors vary with small fluctuations; however fusion based on WMV error highly fluctuates. Theses curves confirm the results presented in Table 3.4. The decision fusion based on kernel ridge regression is not illustrated in Figure 3.6 because of the random shape of its curve. This is due to the ability of the kernel methods to train themselves for each couple of values of the parameters. This is an additional advantage of the kernel fusion method compared to the linear methods.

3.7.2.3 Missing signals

In this section, all possible combinations of signals are considered to test the effectiveness of the proposed approach whenever a signal is missing. Figure 3.7 illustrates a comparison between the proposed fusion functions for all the possible sub-combinations of signals, starting from the combination of 4-signals, then 3 and finally 2-signals. For the simplicity of presentation, HR is presented by 1, RR by 2, SpO_2 by 3, MABP by 4 and the combination of all the signals by 4D. The fusion based on WMV error rates is not illustrated in the

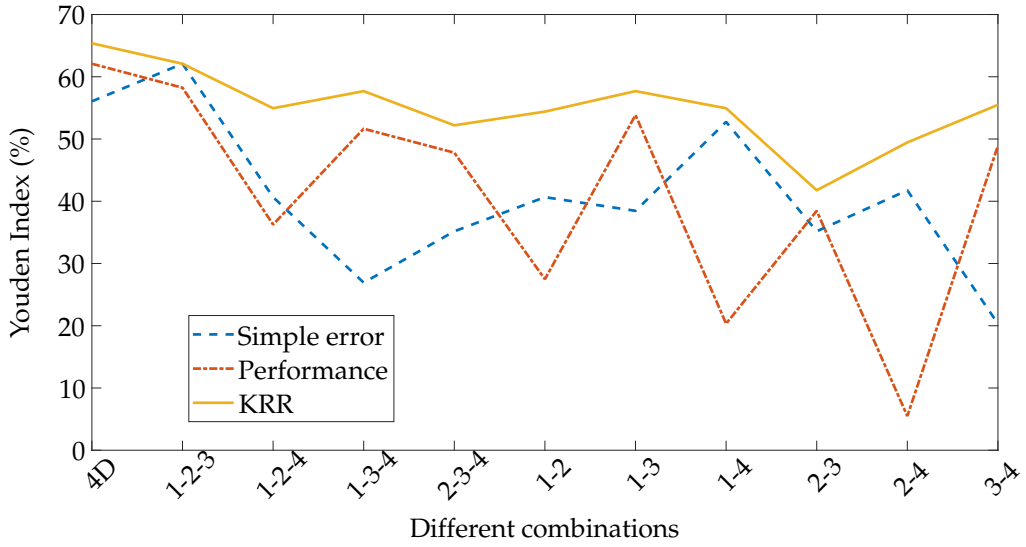


FIGURE 3.7: A comparison of the performances, presented by Youden index, between the proposed fusion functions for the different combinations of signals. In this graph, the following simplifications are done 1: HR, 2: RR, 3: SpO_2, 4: MABP and 4D is the combination of all the signals

figure because it has high fluctuations and lower performances. Figure 3.7 shows that the performances of the method with the linear fusion techniques highly fluctuate with the omission of one or two signals; whereas the kernel-based one is more stable with less fluctuation. This is due to the ability of kernel methods to be adjusted to its training data. Moreover, the KRR-based fusion function presents the best Youden indexes for all types of combinations, with the highest one of 65% approximately for the combination of the four signals HR-RR-SpO_2-MABP. For all these reasons, only the kernel-based fusion function is considered in the following illustrations.

3.7.2.4 Early predictions

The main objective of the proposed approach is to early predict ARDS in order to administer the right treatment and thus avoid its occurrence. In this section, we present the early predictions of ARDS. This is done by finding the time of ARDS detection in function of the real time of its onset. ARDS is detected at time t when a final decision $\hat{y}_s(t) = +1$ is generated. The actual time of ARDS onset is already determined by the PaO_2/FiO_2 ratio. It is used to cut the recordings so the last measurement in the signals represents the time just before ARDS occurrence. The correctly identified subjects from the combination HR-RR-SpO_2-MABP using the kernel ridge regression fusion are considered in this step.

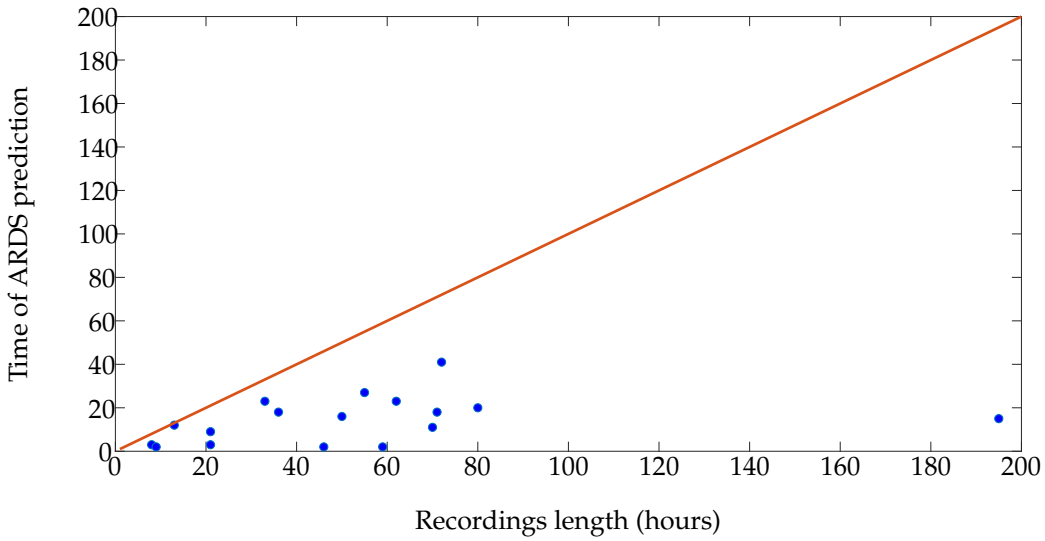


FIGURE 3.8: Early detection results for the combination of HR-RR-SpO_2-MABP using KRR fusion function. This scatter plots time of ARDS prediction versus recordings length (hours)

Figure 3.8 presents in blue a scatter plot representing the time at which ARDS is predicted against the recordings length of signals or in other words the time just before the actual onset of ARDS. The plot of Figure 3.8 shows that all the predictions are placed below the diagonal $y = x$. The nearest prediction is far from the diagonal for about 1 hour, thus approximately one hour before the actual ARDS onset. Early warnings of ARDS are then generated in a range of 1 to 180 hours before its occurrence with an average of 39 hours or approximately two days and a standard deviation of 42 hours. The figure 3.8 also shows that the warnings of ARDS are generated in the range of [1,40] hours from the beginning of the recordings, thus in the first two days of monitoring.

3.7.3 Non-invasive ventilated subjects

Since the number of subjects included in the study is small, the approach is tested on a larger database. The non-ventilated subjects were considered in this part. As described previously, the non-ventilated database includes 50 ARDS subjects and 50 non-ARDS subjects. Within the database of 100 subjects, a k-fold cross-validation technique is considered to set training databases and testing databases, with k = 5.

TABLE 3.6: Results on non-invasive ventilated subjects

Novelty detection +	Training			Testing	
	Se $\pm\Delta$Se	Sp $\pm\Delta$Sp	Acc	Se	Sp
Performance	64.26 ± 6.8	77.06 ± 7.2	70.66	60.4	72
Simple error	64.75 ± 7.2	67.33 ± 6.7	66.04	58	74
WMV error	78.77 ± 8.2	60.62 ± 8.7	69.7	54.8	71.2
KRR	83.03 ± 4.4	62.12 ± 4.3	72.48	62.89	72.67

3.7.3.1 Results of the proposed approach

As mentioned in the previous part, an optimization of the algorithm parameters is performed, $\tau \in \{60, 120, \ldots, 1440\}$ minutes, $D \in \{0.1, 0.2, \ldots, 5\}$, $\eta \in 2^{\{-10,-9,\ldots,0\}}$ and $\sigma_\kappa \in 10^{\{-5,-4,\ldots,-1\}} \cup \{1, 2, \ldots, 5\}$. The results are illustrated in the Table 3.6. As shown in the table, the novelty detection approach followed with the proposed KRR fusion method showed higher performances than the linear fusion methods with a sensitivity and specificity of 83% and 62% on the training sets and 62.89% and 72.67% on the test sets. In addition, the KRR decision fusion algorithm has shown the lower fluctuation of performances over the training sets, and the WMV error-based fusion has shown the higher fluctuation of the performances.

3.7.3.2 Missing signals

The omission of signals is also tested on the non-invasive ventilated subjects using all the decision fusion methods. The results are shown in Figure 3.9 for the linear and non-linear decision fusion techniques. As shown, the KRR fusion technique presents the higher Youden index over all the subcombinations of the four considered signals. The KRR technique has shown it advantages over the linear decision fusion techniques using both, the invasive ventilated and the non-invasive ventilated subjects.

3.7.3.3 Early predictions

It is necessary to test the predictions of ARDS in non-invasive ventilated subjects. The length of ARDS signals varies in the range of [41, 680] hours with a mean of 155.9 hours $\pm$ 120 hours. The ARDS was predicted from the beginning of recordings with an average of 24.9 hours and a standard deviation of 38 hours. The early predictions of ARDS among non-ventilated ARDS subjects are shown in Figure 3.10. The figure shows that ARDS is predicted before its development in an average of 131 hours. Among 50 ARDS subjects,

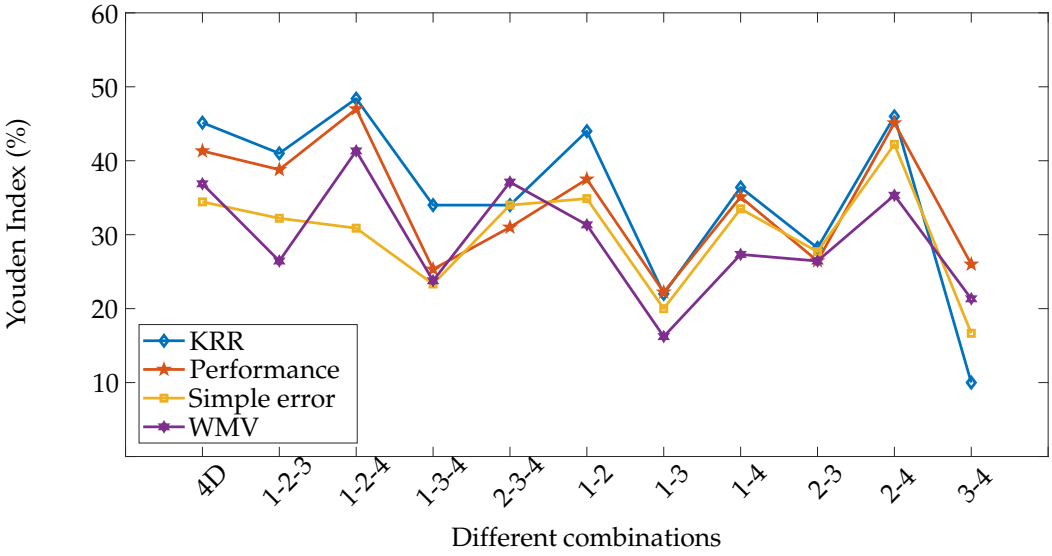


FIGURE 3.9: A comparison of the performances, presented by Youden index, between the proposed fusion functions for the different combinations of signals. In this graph, the following simplifications are done 1: HR, 2: RR, 3: SpO_2, 4: MABP and 4D is the combination of all the signals

only 2 subjects had ARDS predicted two hours before its occurrence; while ARDS was predicted at least 29 hours before its occurrence for all the other subjects.

3.7.4 Comparison to state-of-the-art methods

In this section, we compare the proposed approach to different combinations of existing methods. Table 3.7 shows the performance indexes for both, invasive ventilated subjects and non-invasive ventilated ones, in case of replacing:

1. **The fusion phase:** We propose to replace the KRR technique followed by the sign computation by SVM. In other words, one starts with the novelty detection proposed algorithm and then performs classification by SVM on the individual decisions. SVM constructs a separating boundary in a high dimensional feature space where data are assumed to be linearly separable. Then, data are classified according to this decision boundary. More details on the SVM can be found in (Vapnik, 1999; Friedman, Hastie, and Tibshirani, 2001). The results of Table 3.7 show that the proposed fusion technique with a regression followed by a comparison remains better than a direct classification.

2. **The complete proposed model:** The whole model is replaced by existing classification methods, such as SVM or a regression method based

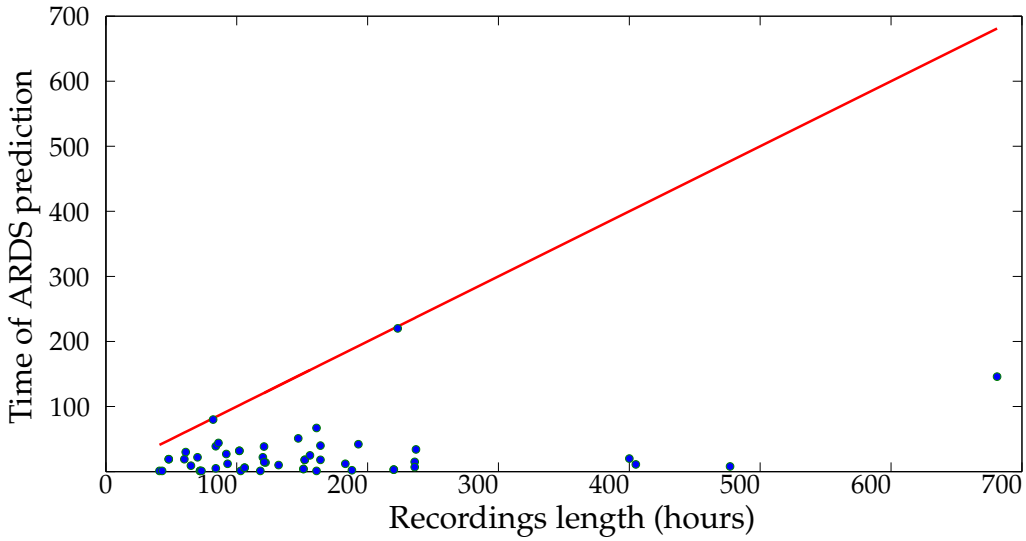


FIGURE 3.10: Early prediction results for the combination of HR-RR-SpO_2-MABP using KRR fusion function for the non-invasive ventilated subjects. This scatter plots time of ARDS prediction versus recordings length (hours)

TABLE 3.7: Performances in % obtained with our approach compared to previous novelty detection technique and state-of-the art methods

	Invasive ventilated			**Non-invasive ventilated**		
	Se	Sp	Acc	Se	Sp	Acc
Novelty detection + SVM	69.23	71.43	70	80	52	66
Global classification with SVM	73	43	62.5	50	76	63
Global classification with KRR	38	92	57.5	82	30	56
The proposed approach	**65.38**	**100**	**77.5**	**62.89**	**72.67**	**67.78**

on KRR followed by a sign computation. Thus, instead of making a decision by comparing $r_{s,i}$ to $p_{s,i}$ as in Eq. (3.5) and then insert fusion functions on decisions, we take $r_{s,i} - p_{s,i}$ as input to the aforementioned methods. Results from Table 3.7 show clearly the improvement in the performances with the newly proposed approach.

The table also shows the accuracy of all methods, that is, the proportion of correctly identified subjects over all subjects independently of their state. The proposed model outperforms the existing classification models in terms of accuracies.

Finally, results obtained in this study are compared to those of a previous one using other types of data. Indeed, Ennet *et al* developed a prediction model using clinical variables, such as plateau pressure and ventilator settings from MIMIC II clinical database (Ennett et al., 2008). They reached a sensitivity and a specificity of 60% and 82% for univariate analysis, and their

fusion gives performances of 50% and 90% for sensitivity and specificity, respectively; while higher results were obtained in this work with the 4-D combination presented by a sensitivity of 65.38% and a specificity of 100% with only non-invasive continuous physiological signals.

3.8 Discussions

An early-warning model for predicting ARDS is proposed and developed in this chapter based on decision fusion functions. The prediction model uses novelty detection and decision fusion algorithms. This detection is crucial in early ARDS phases to implement preventive measures and thereby to reduce mortality. It is also appealing in terms of identifying non-ARDS subjects, thus preventing unnecessary and expensive health care.

In order to set the model, one could use classical classifiers, such as Support Vector Machines (Vapnik, 1998; Übeyli, 2008). However, such methods do not consider the subject's normal states, but generate a global unique boundary for decision. In contrast, the proposed approach is subject-based aiming to detect in real-time by means of a sliding window for each subject whether it deviates from its initial state. It is based on novelty detection algorithm. Several models of novelty detection perform well on different data. However, the success of such models depends on the choice of the model as well as on the statistical properties of the handled data (Velardo et al., 2014). In most physiological signals, the distribution of data is unknown and do not fit any specific distribution (Chua et al., 2010a). Moreover, the data histogram varies continuously even for stable subjects.

At a first place, data can be assumed to be normally distributed as in Definition 1. In the cases that they are not, a Gaussianization phase can be applied to force the normality. The model proposed in this chapter outperforms that previous one being a data-driven model performing in real-time. Furthermore, it is obvious that the distance-based algorithm of Definition 2 performs better than that of Definition 1, and this is due to the non-normality of the data distributions. The distance-based model is associated with different decision fusion rules, linear and non-linear fusions. The results have shown the advantage of non-linear fusion technique based on kernel ridge regression for the fusion of the signals. KRR also showed a

stability in the performances whenever there are one or more missing signals.

In order to validate the proposed model, we selected two types of subjects from the MIMIC II database, namely invasive and non-invasive ventilated subjects. The invasive ventilated dataset was selected based on previous studies in the literature. However, it has a very small number of subjects. Therefore, we proposed to select the non-invasive ventilated dataset. It is worth noting that all the subjects included in this study are intensive care unit patients that developed a severe complication, such as renal failure, cardiac arrest, hypertension, sepsis, pneumonia or others. In the case of invasive ventilated subjects, the proposed model combined with the kernel ridge regression has succeeded to correctly verify all the stable subjects despite the existing health conditions.

In addition, the algorithm tested on the non-invasive ventilated subjects has demonstrated the effectiveness of the non-linear fusion method over the linear techniques. The accuracy of the model on the invasive ventilated subjects was higher than that on the non-invasive subjects. This may refer to the higher number of subjects in the non-invasive ventilated dataset. The non-invasive ventilation helps the patient to regularize their respiration in a non-continuous way. This may lead to higher fluctuations in the patients' vital signs.

As an overall performance, the proposed novelty detection algorithm followed by the KRR-based fusion technique showed the best performance using the four extracted signals. Even with the omission of signals, the performance remains the highest presenting small fluctuations. This model has predicted the onset of ARDS in the first two days of recordings and at least 2 days before its occurrence, thus earlier to its onset.

When comparing with state-of-art methods and the previous ARDS prediction model, this novelty detection - KRR fusion model - outperformed them all. The strength of this work is the use of vital sign data that can be collected from patients at home using non-invasive techniques. In addition, the analysis in real time allows to predict the onset of ARDS in its early stages. As shown, an early warning was generated for the majority of ARDS subjects earlier to onset. Finally, the model remains stable in term of performances in case of missing data.

3.9 Conclusion

In this chapter we developed a decision fusion approach to predict the onset of ARDS in real-time. This model was defined using novelty detection algorithms. It was tested on both types of subjects, since we started the selection of subjects like existing procedures in the literature, then we searched for non-invasive ventilated subjects. The next chapters propose a feature fusion function for the prediction of ARDS. In the following chapters, we test the proposed models only on non-invasive ventilated subjects because they are more numerous than the invasive ventilated subjects and they are more representative of patients at home.

Chapter 4

ARDS Prediction using the Belief Functions Theory

Contents

4.1 Introduction 78

4.2 Evidence-based approach for ARDS prediction 79

4.2.1 Description of the approach 79

4.2.2 Parameters Extraction 81

4.2.3 ARDS prediction with the belief functions theory . . 84

4.3 Evidence-based approach for four-group prediction 89

4.3.1 Identification of groups 90

4.3.2 The belief functions framework 91

4.4 Selection of parameters 93

4.5 Results 94

4.5.1 Statistical analysis 95

4.5.2 Parameters Selection 97

4.5.3 Selection of the alert generation threshold 99

4.5.4 Selection of window size 99

4.5.5 Performance Evaluation 100

4.5.6 Early predictions 102

4.5.7 Comparison to the state-of-the-art methods 102

4.5.8 Multi-class classification 103

4.6 Discussions 105

4.7 Conclusion 107

4.1 Introduction

Feature fusion techniques have demonstrated their effectiveness and advantage over decision fusion techniques. In feature fusions, parameters are extracted from raw data signals and pertinent ones are selected to enhance the fusion performance. The characterization of signals using the extraction of parameters has been widely applied in biomedical applications. For instance, an analysis of heart rate variability during syncope was conducted by extracting parameters such as the mean, the standard deviation of the RR interval and non linear parameters such as the sample entropy and the detrended fluctuation analysis (Khodor et al., 2014). Moreover, linear and non-linear parameters were extracted from heart rate variability signals for ARDS subjects to analyze their response to alveolar recruitment manoeuver (Borghi-Silva et al., 2016). The linear time domain parameters are mean, root mean square of squares of differences (RMSSD) and standard deviation of the mean of normal values (SDNN) and non-linear parameters are Poincaré plot and the detrended fluctuation analysis. Only few studies have considered other cardiovascular signals than heart rate, like blood pressure (Angelini et al., 2007; Turianikova et al., 2011; Rivera et al., 2016). A characterization of heart beat and blood pressure for diabetes patients using time domain parameters such as the mean, the standard deviation, the skewness and the kurtosis was studied in (Rivera et al., 2016). In addition, detrended fluctuation analysis was extracted from blood pressure signal to study its dynamics through surgical procedure in rats (Galhardo et al., 2009). In (Jalali, Licht, and Nataraj, 2013), a characterization of periventricular leukomalacia, a brain injury affecting infants, is proposed by extracting minimum, maximum, mean, variance, skewness, kurtosis, energy of wavelet coefficients and sample entropy from heart rate, mean arterial blood pressure and oxygen saturation. From what preceded, we can conclude that the extraction of parameters from different physiological signals might help in the characterization of physiological aspects related to several pathologies, such as ARDS.

In this chapter, we propose to extract linear and non-linear parameters from the time series data in real-time and the most informative ones are selected. After that, a prediction model based on the evidence framework is defined in order to combine information of the extracted parameters and generate a decision regarding the health state of the patient in study. The

evidence theory, known as the theory of belief functions, consists of modeling the evidence provided by different sources and combining it by considering the uncertainty and imprecision of information. In the next section, we describe all the parameters extracted from the signals. Then, we present the evidence-based method to distinguish between the previously defined ARDS and non-ARDS groups. Thereafter, we propose a multi-class classification framework to distinguish between subjects not developing ARDS and each of the three types of ARDS defined in the Berlin definition, namely, severe, moderate and mild ARDS. This framework applies the same model described for the binary classification. Finally, the results are illustrated for both approaches and interpreted.

4.2 Evidence-based approach for ARDS prediction

This section provides a detailed description on the complete feature fusion approach proposed in this chapter for the ARDS/non-ARDS prediction problem. The method described in this section has the same objective of the previous one, which is real-time monitoring of the ongoing health state of patients using their vital signs. $x_{s,i}(\ell)$, $i = 1, \ldots, 4$, denote the four time series considered in this study for a subject s, recorded at time ℓ, and t denotes the real-time. We recall that $x_{s,i}(a : b) = (x_{s,i}(a), ..., x_{s,i}(b))$ is a segment of signal i for subject s going from time a to time b. The next subsection summarizes the complete model, and then each block is described in details in the following subsections.

4.2.1 Description of the approach

The block diagram in Figure 4.1 presents the proposed model of feature fusion for ARDS/non-ARDS prediction. The aim of this work is to set a model $\Psi(\cdot)$ from a training set that distinguishes in real-time between subjects who are going to develop ARDS labeled " $+1$", and subjects who will not, labeled " -1". The model $\Psi(\cdot)$ takes as input the collected signals $x_{s,i}(1 : t)$, going from the beginning of records until the real-time t, and yields as output a decision whether the subject in study is going to develop ARDS. As described in the previous chapter, the proposed approach uses the beginning of the signals to define a normal state. It then considers a fixed window of length τ taken at the end of the signals, leading to four segments

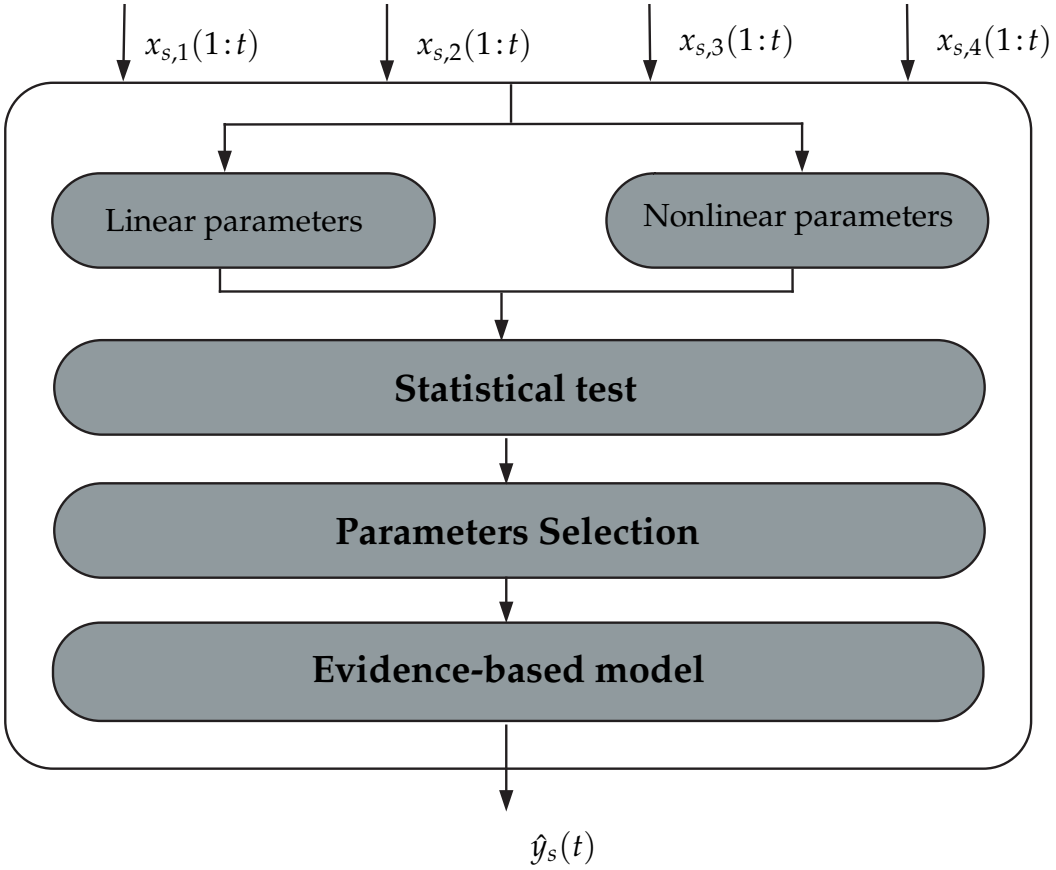


FIGURE 4.1: Block diagram of the proposed evidence-based model for ARDS prediction

$x_{s,i}(t-\tau+1 \,:\, t)$ to be analyzed at each time t. The reason of analyzing the data of a window of length τ is to have fixed length signals in computations. Moreover, an abnormality is better and faster detected within a short segment, compared to the high length signal. The value of τ is obtained by an optimization over the training sets, as shown in the following. Then, a vector of linear and nonlinear parameters is extracted from these segments at real-time t. The beginning of the signals is yet needed to normalize these parameters.

The extracted parameters go through a feature selection algorithm to choose the relevant indicators and reduce the dimensionality of the input vector. This vector of parameters is then used with the belief functions theory to assign evidence to each health state according to each parameter. The parameters are considered as sources of information, providing an evidence regarding each state of the patient at the given time. Thus, masses are assigned to each state " $+1$" and " -1" according to each parameter. The

masses are obtained by means of mass functions defined using the parameters extracted from the training set at an earlier offline phase. The masses are then combined to assign a final mass to each state. A higher mass assigned to the state " $+1$" at time t means that the subject is predicted to develop ARDS in the future, whereas a higher mass to " -1" means that the ARDS is not yet predicted at the current time. Therefore, for a new test subject given its signals in real-time, a decision will be made on its future health state at the earliest possible time.

4.2.2 Parameters Extraction

The first step of the prediction model is the extraction of parameters. This section presents the both types of the extracted parameters, namely, linear and nonlinear parameters. Having a segment $x_{s,i}(t-\tau+1:t)$ from each signal at the real-time t, several linear and non-linear parameters are extracted as follows. In the following, t^- would denote the lower bound of the considered time interval, that is, $t^- = t-\tau+1$. For linear parameters, we consider:

- **The mean** of the segment $x_{s,i}(t^-:t)$, defined as

$$\mu_{s,i}(t) = \frac{1}{\tau}\sum_{\ell=t^-}^{t} x_{s,i}(\ell).$$

- **The standard deviation** or the square root of variance of $x_{s,i}(t^-:t)$,

$$\sigma_{s,i}(t) = \sqrt{\frac{1}{\tau}\sum_{\ell=t^-}^{t}(x_{s,i}(\ell)-\mu_{s,i}(t))^2}.$$

  The variance is equal to the total power of spectral analysis. Thus, it reflects all the cyclic components responsible for the variability in the data (Camm et al., 1996).

- **The skewness** known as the third moment. It measures the symmetry of the data and it is computed using

$$Sk_{s,i}(t) = \frac{\sum_{\ell=t^-}^{t}(x_{s,i}(\ell)-\mu_{s,i}(t))^3}{\tau\times\sigma_{s,i}^3(t)}.$$

  It has been proposed that asymmetric tails to the left shown by positive skewness or to the right with negative skewness reflect, respectively,

the acceleration or deceleration capacity of the time series data, e.g. for heart rate signal, it is an approximate distinction of vagal and sympathetic effects on the cardiac modulations (Bauer et al., 2006).

- **The kurtosis** that describes the shape of the probability distribution. It is calculated using the fourth moment of the data,

$$Kt_{s,i}(t) = \frac{\sum_{\ell=t^-}^{t} (x_{s,i}(\ell) - \mu_{s,i}(t))^4}{\tau \times \sigma_{s,i}^4(t)}.$$

  The kurtosis measures the concentration of the data around the mean and reflects the rigidity of the signal (Rivera et al., 2016).

On the other hand, nonlinear parameters are also considered, such as:

- **The sample entropy** ($SampEn$) that computes the amount of complexity in segment data. It is easily applied to any type of time series, including physiological data such as heart rate variability and blood pressure variability (Moorman et al., 2011; Jalali, Licht, and Nataraj, 2013). It is computed using the negative natural logarithm of the conditional probability that a segment of length N having repeated itself for u samples within a tolerance r will also repeat itself for $u+1$ samples (Lake et al., 2002). Hence, for the segment $x_{s,i}$ of length $N = \tau$, the function $h(u,r)$ is defined being the number of pairs of patterns sequences of $x_{s,i}$ of length u whose distance is less than r. u is the length of sequences to be compared and r the tolerance of acceptance. $SampEn$ is then computed as follows,

$$SampEn_{s,i}(u,r,\tau,t) = -\ln\left(\frac{h_{s,i}(u+1,r)}{h_{s,i}(u,r)}\right).$$

- **The detrended fluctuation analysis** (DFA) that measures the scaling behavior of a segment (Francis et al., 2002). It is defined by the short and long range scaling exponents DFA_1 and DFA_2. To perform detrended fluctuation analysis for a segment $x_{s,i}(t^- : t)$, the integrated series $y(u)$, $u = 1, \ldots, \tau$, is first computed using a cumulative sum

$$y(u) = \sum_{\ell=1}^{u} [x_{s,i}(\ell + t - \tau) - \mu_{s,i}(t)].$$

Then, $y(u)$ is divided into time segments of length n, and a local trend $y_n(u)$ is obtained by a least-squares line fit that minimizes the squared error within each window. It is then subtracted from $y(u)$. The fluctuation $f(n)$ for segments of length n is calculated as follows:

$$f(n) = \sqrt{\frac{1}{\tau}\sum_{u=1}^{\tau}(y(u) - y_n(u))^2}.$$

The fluctuation measurement $f(n)$ is repeated over a range of different segment sizes n, then a graph of $\log(f(n))$ against $\log(n)$ is constructed. Finally, the scaling exponents DFA_1 and DFA_2 define, as their names denote, the slope of a straight line fit of the relation between $\log(f(n))$ and $\log(n)$ for ranges of low and high values of n using least-squares. The DFA parameters express the autocorrelation characteristics of the signal:

- DFA < 0.5: anti-correlated data,
- DFA $= 0.5$: uncorrelated data,
- DFA > 0.5: correlated data.

DFA has been used to assess the self-similar correlations of the time series (Peng et al., 1995). It has also been described as a measure of "roughness" in the signals, with higher scaling exponents representing a smoother time series (Millar et al., 2013).

Having the four measured signals, these computations lead to seven parameters per signal and thus 28 parameters for each subject s at each time t. Let $p_{s,j}(t)$, $j = 1, ..., 28$, denote these parameters. For many health conditions, the trends in vital signs differ among patients. Thus, normal is a relative state for each patient, that differs with age, weight, medical history, etc. Hence, instead of using the raw values of the parameters, a normalization phase is proposed according to each patient. Therefore, a first segment of length τ is considered from each signal, $x_{s,i}(1:\tau)$. It corresponds to the initial state of the patient that is assumed to be its stable one. Then, the different parameters are computed for the initial segments. Let $p_{s,j}^{(0)}$ denote their initial values. Then, for each real-time measurement, the obtained parameters are normalized by computing their ratio to the parameters of the first segment $p_{s,j}(t)/p_{s,j}^{(0)}$. In the following, $p_{s,j}$ would denote the normalized parameter, for simplicity.

Note that, a parameter selection would be performed in the offline phase using the training set to find the pertinent parameters and keep only informative ones. More details are given in Section 4.4. In the following, let J be the set of indices of the selected parameters, and thus $p_{s,j}(t)$, $j \in J$, are the normalized parameters extracted at real-time t from the signals of a subject s.

4.2.3 ARDS prediction with the belief functions theory

In this section, we provide a detailed description on the belief functions theory and its application in this chapter that is weighted feature fusion. The theory of belief functions, also called Dempster-Shafer theory or Evidence theory, is introduced for the analysis of data with imperfect information (Dempster, 1967; Shafer, 1976). The belief functions theory consists of illustrating the uncertainty and imprecision of information. It also takes into account ambiguities and conflicts between sources.

The belief functions theory operates on a frame of discernment Ω that consists of a finite set containing mutually exclusive propositions. In this part of the study, we consider two groups of subjects, ARDS labeled "$+1$" and non-ARDS labeled "-1". In this case, Ω is defined by $\Omega = \{+1, -1\}$. Let 2^{Ω} be the set of all possible subsets of Ω, it is then defined by:

$$2^{\Omega} = \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\}.$$

The subset $\emptyset$ denotes an empty set where the state of the subject is neither ARDS, nor non-ARDS, which means impossibility; however $\{+1, -1\}$ denotes ignorance and ambiguity where the subject has both states or, in other terms, it is hard to distinguish between the states for a given subject.

Next subsection describes the concept of belief functions that is based on mass assignment. Then, the feature fusion is introduced by combining the assigned masses. In order to consider the reliability of the provided information, a discounting operation is proposed. Furthermore, a parameters selection is proposed to select only pertinent parameters for the feature fusion process.

4.2.3.1 Basic concept

The main concept relies on the modeling of the evidence provided by the extracted parameters. Information given by a source, that is a parameter in this study, can be represented by a *basic belief assignment* (BBA), also named mass function (Smets and Kennes, 1994). Let $m_j(\cdot)$ be the mass function associated to the parameter p_j. Then, $m_j(\omega, s, t)$ reflects the strength of evidence supporting a subset $\omega \in 2^\Omega = \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\}$ according to the value of the parameter p_j at time t. In other words, $m_j(\omega, s, t)$ represents the part of evidence saying that the state of the subject s falls in ω at time t. The mass function should have the following properties $\forall s, t$:

$$\begin{cases} m_j(\emptyset, s, t) = 0, \\ m_j(\omega, s, t) \to [0, 1], \text{ for } \omega \in 2^\Omega, \\ \sum\limits_{\omega \in 2^\Omega} m_j(\omega, s, t) = 1. \end{cases} \tag{4.1}$$

In order to construct the mass functions related to each parameter, the training sets are considered. For ARDS subjects of a given training set, the last segments of their signals are used to compute their parameters $p_{s,j} = p_{s,j}(T_s)$, T_s being the length of the signals of an ARDS subject s. The reason of this choice is that the end of the signals precedes directly the occurrence of ARDS and thus contains mostly its instability. For non-ARDS subjects, the more the segments approach the beginning of the signals, the more we are sure of the stability of the subject, since it is possible that a subject labeled non-ARDS develops the ARDS after recordings. Then, the parameters of non-ARDS are computed using segments selected from the beginning of the signals, following the first segments. In addition, parameters from the initial segments are computed $p_{s,j}^{(0)}$ to normalize the extracted parameters. Let S_{+1} and S_{-1} be the sets of indices of ARDS and non-ARDS subjects of the training set respectively. Then, normalized parameters databases $p_{s,j}$, $s \in S_{+1}$, and $p_{s',j}$, $s' \in S_{-1}$, are obtained.

Afterwards, for a parameter type j, the probability density functions (pdf) are estimated in three different cases, such as

1. using only ARDS parameters values $p_{s,j}$, $s \in S_{+1}$. The estimated pdf is written $Q_{j,\{+1\}}(\cdot)$.
2. using only non-ARDS parameters values $p_{s',j}$, $s' \in S_{-1}$. The estimated pdf is written $Q_{j,\{-1\}}(\cdot)$.

3. using parameters of both classes $p_{s,j}$, $s \in S_{+1} \cup S_{-1}$. Let $Q_{j,\{+1,-1\}}(\cdot)$ denotes the concerned estimated pdf.

These distributions could be obtained by fitting the histograms of the parameters to all the classical distributions. Then, in the online phase, having a computed parameter $p_{s,j}(t)$ at the real-time t, the mass assigned to each subset of 2^{Ω} is given by

$$m_j(\omega, s, t) = \frac{Q_{j,\omega}(p_{s,j}(t))}{\sum\limits_{\omega' \in 2^{\Omega}, \omega' \neq \emptyset} Q_{j,\omega'}(p_{s,j}(t))}, \omega \in 2^{\Omega}, \omega \neq \emptyset, \tag{4.2}$$

whereas the mass of the empty set is taken equal to 0, $m_j(\emptyset, s, t) = 0$. The empty set represents that the subject's state is neither ARDS nor non-ARDS, which is impossible in our study. Herein, we suppose that all the subjects should belong to one of the Ω states that are either ARDS or non-ARDS.

Figure 4.2 illustrates an example of such a computation. The plot shows the distributions of the normalized standard deviation of respiratory rate signals parameter for ARDS subjects of a training database in straight red line, the one for non-ARDS subjects in dashed blue line and the one of all ARDS and non-ARDS subjects in straight black line. Moreover, the plot shows the mass assignment for two new parameter values $p^{(1)}$ and $p^{(2)}$. The first parameter value falls in the distribution of ARDS subjects and the second is in the middle of ARDS and non-ARDS distributions. When assigning the masses of each value from Eq. (4.2), the assigned mass of $p^{(1)}$ from $Q_{j,\{+1\}}$ is higher than those of $Q_{j,\{-1\}}$ and $Q_{j,\{+1,-1\}}$. Thus, a higher evidence is assigned to the fact that the parameter of the studied subject belongs to the state of $\{+1\}$, thus it will develop ARDS. On the other hand, having the value $p^{(2)}$, the distributions of $\{+1\}$ and $\{-1\}$ are too close, leading to similar masses from both. However, higher mass is associated in this case to the distribution of $\{+1, -1\}$, thus introducing the ambiguity of such case and no decision on the state of the subjects could be generated. The mass assigned to $\{+1, -1\}$ is higher than those of $\{+1\}$ and $\{-1\}$ in all the cases where these distributions are too close. This figure illustrates the effectiveness of the belief functions theory concept by inserting ambiguities when dealing with close data and thus avoiding erroneous decisions.

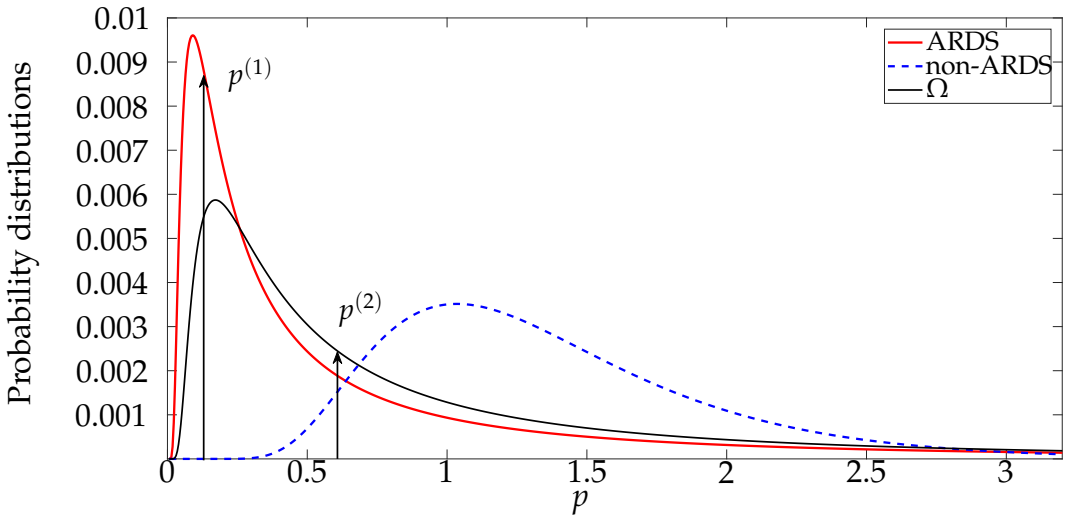


FIGURE 4.2: Mass assignment of new computations in the case of the normalized standard deviation of respiratory rate signals

4.2.3.2 Discounting

Since we are working on parameters extracted from human bodies, the information cannot be completely reliable. Hence, a discounting operation is needed to take into consideration the reliability of the information provided by each parameter (Shafer, 1976; Smets, 1993). This operation transforms each mass function to a less or more informative one, based on the degree of reliability of the source. In this chapter, we propose to estimate the reliability of each source according to each state, leading to a contextual discounting (Mercier, Lefèvre, and Delmotte, 2012; Mercier, Pichon, and Lefèvre, 2016). Let $\beta_{j,\{+1\}} = 1 - \alpha_{j,\{+1\}}$ and $\beta_{j,\{-1\}} = 1 - \alpha_{j,\{-1\}}$ be the degrees of reliability of a parameter p_j knowing that the true state is " $+1$" and " -1" respectively. Then, the discounted mass function ${}^{\alpha}m_j(\cdot)$ obtained from p_j is given by

$$\begin{cases} {}^{\alpha}m_j(\{+1\},s,t) = \beta_{j,\{-1\}} m_j(\{+1\},s,t), \\ {}^{\alpha}m_j(\{-1\},s,t) = \beta_{j,\{+1\}} m_j(\{-1\},s,t), \\ {}^{\alpha}m_j(\Omega,s,t) = m_j(\Omega,s,t) \\ \qquad\qquad + \alpha_{j,\{-1\}} m_j(\{+1\},s,t) \\ \qquad\qquad + \alpha_{j,\{+1\}} m_j(\{-1\},s,t), \\ {}^{\alpha}m_j(\emptyset,s,t) = 0. \end{cases} \tag{4.3}$$

The error rates $\alpha_{j,\{+1\}}$ and $\alpha_{j,\{-1\}}$ could be estimated using the distribution functions $Q_{j,\omega}(\cdot)$. Knowing that a subject s state is equal to " $+1$", an error is made in evidence assignment if a higher mass is given to any subset

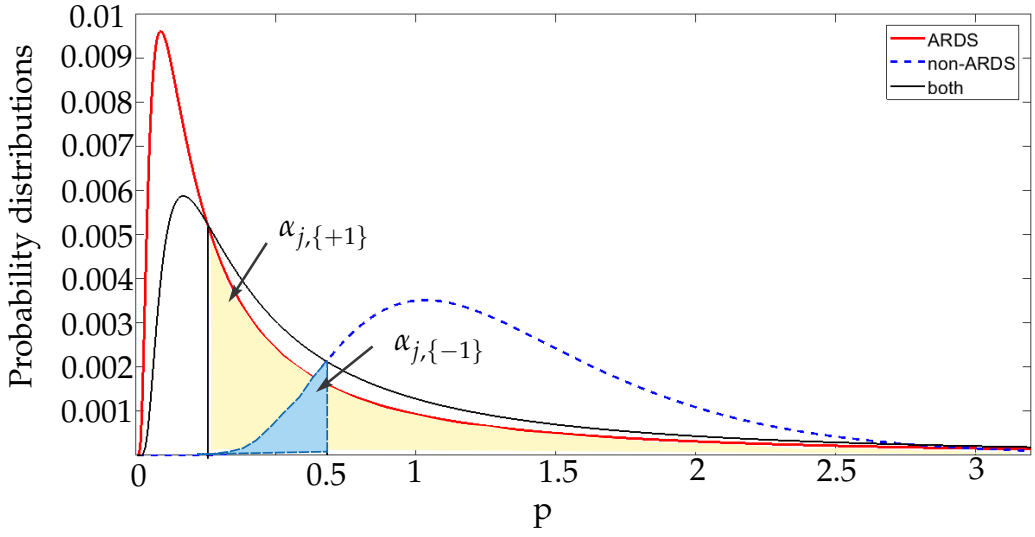


FIGURE 4.3: An example for computing the error rates

of Ω other than $\{+1\}$. It is also the same for $\{-1\}$. Then, $\alpha_{j,\omega}$, $\omega = \{+1\}$ or $\{-1\}$, is computed for all parameter values, where $Q_{j,\omega}(p)$ is less than any $Q_{j,\omega'}(p)$, for $\omega' \neq \omega \in 2^{\Omega}$. Therefore, it is defined as follows:

$$\alpha_{j,\omega} = \int_{D_{j,\omega}} Q_{j,\omega}(p)dp, \tag{4.4}$$

where $D_{j,\omega} = \{p|Q_{j,\omega}(p) < Q_{j,\omega'}(p), \forall \omega' \in 2^{\Omega}, \omega' \neq \omega\}$. Figure 4.3 illustrates an example of computing error rates $\alpha_{j,\{+1\}}$ and $\alpha_{j,\{-1\}}$ using the probability distribution functions.

4.2.3.3 Combination

Once the discounted masses of the different parameters are computed, they can be combined using the normalized conjunctive rule or the disjunctive rule of combination from the belief functions theory. We propose in this work to use the normalized conjunctive rule. By doing this computation, one obtains a more informative mass function, leading to more efficient estimation. Moreover, by discounting the masses, the obtained masses are assumed to represent reliable information. The combination of these masses is denoted by $\oplus$ and defined by the following equation:

$$m_{\oplus}(\omega, s, t) = \frac{\sum\limits_{\cap\omega'^{(j)}=\omega} \prod\limits_{j\in J} {}^{\alpha}m_j(\omega'^{(j)}, s, t)}{1 - \sum\limits_{\cap\omega'^{(j)}=\emptyset} \prod\limits_{j\in J} {}^{\alpha}m_j(\omega'^{(j)}, s, t)}. \tag{4.5}$$

4.2.3.4 Decision making

Finally, to make a decision on the health state of a subject s, we propose to use the pignistic probability. Indeed, the mass function $m_{\oplus}(\cdot)$ defined on Ω is transformed to a probability measure over the singletons with the pignistic transformation defined by:

$$BetP(\omega, s, t) = \sum_{\omega' \in 2^{\Omega}, \omega \subseteq \omega'} \frac{m_{\oplus}(\omega', s, t)}{\mid \omega' \mid}, \tag{4.6}$$

for $\omega = \{+1\}$ or $\{-1\}$. In this way, the final mass of $\{+1, -1\}$ subset is divided between $\{+1\}$ and $\{-1\}$. The state having the highest pignistic level is selected for each subject at every time t. Thus, a decision is made over the health state of subject s whether he will develop ARDS based on its recordings, and an alert is generated for a positive decision.

To make a final decision from the real-time analysis about the health state of a patient, it is also possible to wait for successive positive decisions. Here, a threshold must be defined as being the number of successive ARDS decisions needed to decide if the patient is going to develop ARDS. This is done by applying the algorithm iteratively on the total length of the signals s of the training set, then different numbers of successive positive decisions are considered and the performance indexes are computed for each number, that are the sensitivity (Se) and specificity (Sp) indexes, also called true positive rate and true negative rate, respectively. The threshold of successive ARDS decisions is obtained thereafter by maximizing the Youden index, $Youden = Se + Sp - 100$, over all the considered numbers. The highest Youden index can be also represented as the furthest point from the diagonal or chance level in the ROC curve.

4.3 Evidence-based approach for four-group prediction

The Berlin definition has identified four groups related to ARDS according to the value of the ratio of the partial pressure of oxygen to the fraction of inspired air (PaO_2/FiO_2) as following:

1. Severe ARDS: $PaO_2/FiO_2 \leq 100$ mmHg;
2. Moderate ARDS: $100 < PaO_2/FiO_2 \leq 200$ mmHg;

3. Mild ARDS: $200 < PaO_2/FiO_2 \leq 300$ mmHg;

4. Non-ARDS: $PaO_2/FiO_2 > 300$ mmHg.

In the work presented until now, we performed a two-class classification to distinguish between ARDS group consisting of severe and moderate ARDS subjects and non-ARDS group including mild ARDS and non-ARDS subjects. Mild ARDS subjects are included in the non-ARDS group because these subjects can recover without a clinical intervention or being ventilated, while moderate and severe ARDS subjects have more difficulties to recover alone. However, this work has a limitation, which is the consideration of mild ARDS state as being a healthy one; meanwhile, this state might be a preparatory phase to develop a more severe ARDS level such as moderate or the severe one.

Thus, we propose in the following to distinguish between the four groups of ARDS identified in the Berlin definition. This leads to a new frame of discernment Ω, defined by $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, where ω_k, $k = 1, \ldots, 4$, refer to severe, moderate, mild and non-ARDS states, respectively. The set 2^Ω becomes the set of all possible subsets having 2^4 elements,

$$2^\Omega = \{\emptyset, \{\omega_1\}, \ldots, \Omega\}.$$

In the following, we describe the identification of each group in the signals and the construction of the training datasets. Then, we project the four-group prediction problem in the previously described belief functions framework.

4.3.1 Identification of groups

When distinguishing the four groups of subjects, the representation of each state becomes harder than in the case of the binary classification. In the binary classification, we considered the last segments before ARDS occurrence in the modeling of the ARDS group and the first segments for the non-ARDS group. This representation cannot be applied in the case of a four-class classification, since a severe ARDS subject may develop a mild and/or moderate phases before starting the severe phase. Therefore, we propose to construct the learning dataset from the values of PaO_2/FiO_2. Indeed, we identify in each ARDS subject's signals mild segments, moderate segments and severe segments. By performing this procedure, we

increase the dimension of the training set and specify the modeling of each state. In order to model the non-ARDS group, we only consider segments extracted from the signals of non-ARDS subjects to ensure the stability.

As mentioned in Chapter 1, the variables of the MIMIC II clinical database are collected irregularly. This is the case for the collected values of PaO_2 and FiO_2 of which the ratio is computed. Thus, we linearly interpolate the PaO_2/FiO_2 vector to obtain a continuous signal having a frequency of 1 data point per minute. Hence, a comparison between the signals and the interpolated vector can be performed.

In order to expand the training dataset, we proposed in this work to select different segments of length τ from each signal, each representing an ARDS phase. In this part, length of each recorded signal is considered. Firstly, we identify in the time series the time intervals at which each of the phases occurs. Thus, we identify the three levels of ARDS for a severe subject if they exist. For a moderate ARDS subject, we identify mild and moderate phases. Mild phases are identified for a subject developing mild ARDS. Finally, non-ARDS segments are identified only from non-ARDS subjects in order to ensure the stability in the signals. Figure 4.4 illustrates an example of such identification from a heart rate signal for a severe ARDS subject. After that, we select a segment of length τ from each identified phase to compute the parameters. Parameters are then normalized using a segment from the beginning of the signals.

4.3.2 The belief functions framework

Once the training datasets are constructed, distribution functions of each subset $\omega \in 2^{\Omega}$ are estimated. It is obvious that the training data are not balanced between the groups. This may produce biased distributions especially in the case of non-singleton subsets. In order to balance the datasets, one can choose either to over-sample the minority groups or under-sample the majority groups. The over-sampling of minority groups is more suitable than the under-sampling since the original dataset is small. Thus, the over-sampling will produce more data points for minority groups from the original data of these groups. The Synthetic Minority Over-sampling Technique (SMOTE) is performed to balance the training sets (Chawla et al., 2002). It consists of considering each minority class data point and creating

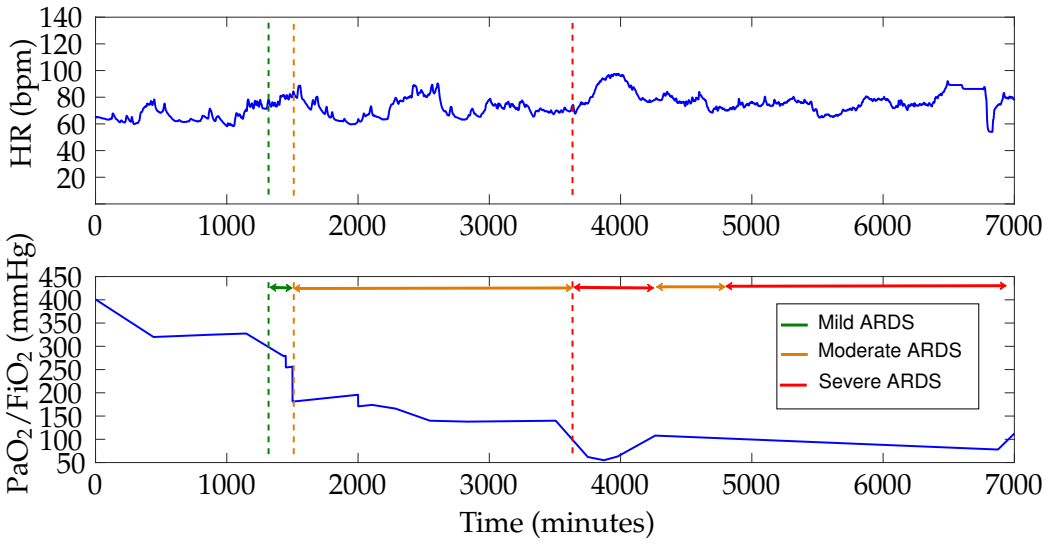


FIGURE 4.4: An example of identification of mild, moderate and severe phases of ARDS in heart rate signal using the PaO_2/FiO_2 signal.

synthetic examples along the line segments joining the K minority class nearest neighbors. The neighbors are selected according to the amount of over-sampling needed.

Afterwards, distribution functions $Q_{j,\omega}(\cdot)$ are estimated for each parameter according to subset $\omega \in 2^{\Omega}$. The mass functions $m_j(\omega, s, t)$ are then assigned for parameter p_j according to state ω using Eq. (4.2). It satisfies $\sum_{\omega \in 2^{\Omega}} m_j(\omega, s, t) = 1$. After finding the distribution functions, we compute the error rates of each parameter according to each state ω when higher mass is given to any subset of Ω other than the real state ω. The error rates are computed using Eq. (4.4). Thus, we obtain $\boldsymbol{\alpha}_{j,\omega}$ the vector of conditional discounting rates for singletons subsets $\omega \in 2^{\Omega}$. α_j would denote the average of the conditional discounting rates. Then, we define both the basic and the contextual discounting operations as follows:

$$\begin{cases} {}^{\alpha}m_j(\omega, s, t) = (1 - \alpha_j) m_j(\omega, s, t), \ \forall \omega \in 2^{\Omega}, \omega \neq \Omega, \\ {}^{\alpha}m_j(\Omega, s, t) = (1 - \alpha_j) m_j(\Omega, s, t) + \alpha_j. \end{cases} \tag{4.7}$$

$${}^{\alpha}m_j(\omega, s, t) = \sum_{\omega_k \subseteq \omega} \left(\prod_{\omega_u \in \omega \setminus \omega_k} \alpha_{j,u} \prod_{\omega_v \in \overline{\omega}} \beta_{j,v} \right) m_j(\omega_k, s, t), \forall \omega_k \subseteq \omega \in 2^{\Omega}. \tag{4.8}$$

The discounted mass functions of the different parameters are combined using Eq. (4.5) and a decision is made using the pignistic probability of Eq.

(4.6) that attributes masses to singleton sets only. The state $\omega \in 2^{\Omega}$ having the highest mass function is selected for each subject at every time t. One could follow the degradation in the state of ARDS subjects.

In order to generate an alert of ARDS, a threshold of successive ARDS decisions must be defined. Since we are working with three groups of ARDS, three different thresholds must be determined. The training dataset is considered for this study. Each of the three thresholds is determined from the corresponding training group. The threshold is defined by the average number of successions of the corresponding decision over all the training group.

4.4 Selection of parameters

Instead of considering all the extracted parameters in the model, it is relevant to select the most informative ones with less error rates. This part of the algorithm consists of ranking the parameters according to their error rates and then performing a sequential forward selection (SFS) of parameters (Devijver and Kittler, 1982; Ladha and Deepa, 2011). Hence, an error rate ϵ_j is computed for each parameter p_j, which is the average of conditional errors computed using Eq. (4.4),

$$\epsilon_j = \left(\sum_{\omega \in \Omega} \alpha_{j,\omega} \right) / |\Omega|,$$

where $|\Omega|$ is the cardinality of the frame of discernment Ω. A global ranking of parameters is performed. Parameters from all signals are ordered from the lowest error obtained to the highest one.

Once the parameters are ranked, a sequential forward selection procedure is performed. The SFS procedure begins with an empty subset. One parameter is added at a time and performance indexes of the model are computed. For each added parameter, the belief model is applied in real-time for all the subjects of the training set and a decision is generated for each of them. The accuracy of the model is computed as follows:

$$Accuracy = \frac{\text{Number of correctly identified subjects}}{\text{Total number of subjects}}.$$

Thereafter, a curve is plotted and updated at each iteration illustrating the accuracy index obtained versus the number of parameters. Whenever a new value added to the curve shows a decrease in the performance higher than a margin ε, the SFS procedure ends with the previously added parameters.

In addition, in order to take into consideration the dependence of the parameters, the correlation between the added parameter and the existing parameters is tested and compared to the selection procedure alone. If any correlation coefficient between the parameters presents a p_value ≤ 0.05 then the parameter is considered correlated with an existing one, thus it is eliminated from the selection procedure.

4.5 Results

As mentioned above, two types of parameters were extracted from each time series. Linear parameters are the mean μ, the standard deviation σ, the skewness Sk and the kurtosis Kt. Non linear parameters are the sample entropy $SampEn$ and both factors from the detrended fluctuation analysis DFA_1 and DFA_2. For sample entropy, the value of u is taken equal to 5, and r equals 0.2, as mentioned in (Fele-Žorž et al., 2008). For the detrended fluctuation analysis, DFA_1 is calculated between the range $n = 4$ and $n = 16$ and DFA_2 is calculated between $n = 16$ and $n = 64$, as shown in (Francis et al., 2002).

In this chapter, we propose to validate the proposed approach using only the non-invasive ventilated subjects because they are more similar to patients at home than the invasive ventilated subjects. The dataset consists of 100 subjects, 50 ARDS subjects and 50 non-ARDS ones. The validation procedure was carried on using a 5 fold cross validation repeated 10 times. The illustration of the approach is presented in Sections 4.5.1, 4.5.2, 4.5.3 and 4.5.4. For the illustration, only one example of training-testing sets is considered. Then, all the 10×5 folds sets are considered for the overall performance evaluation in Section 4.5.5. After that, a comparison with the state of the art methods is shown. Finally, the results of the four-group model are illustrated.

TABLE 4.1: The most significant parameters for each signal, parameters having a p-value < 0.05 are shown. Median $\pm$ Median absolute deviation are shown for each parameter

Signal	Parameters	ARDS group	non-ARDS group	p-value
HR	μ	1.0164 ± 0.1126	1.0011 ± 0.0057	0.0001*
	σ	1.0509 ± 0.5164	1.0025 ± 0.0513	0.0004*
	Sk	0.0373 ± 3.3641	0.8883 ± 1.1944	0.0001*
	Kt	1.0337 ± 0.5004	0.9911 ± 0.0879	0.0001*
	$SampEn$	1.0223 ± 0.2037	0.9987 ± 0.0205	0.0037*
	DFA_1	0.9787 ± 0.1493	1.0035 ± 0.0291	0.0108*
	DFA_2	1.0024 ± 0.0191	0.9991 ± 0.0069	0.0431
RR	μ	1.1279 ± 0.2923	1.0105 ± 0.0148	0.0001*
	σ	0.9656 ± 0.6195	0.9933 ± 0.0404	0.0012*
	Sk	0.1917 ± 8.3018	0.9389 ± 0.8991	0.0001*
	Kt	0.9067 ± 0.5324	0.9985 ± 0.0578	0.0011*
	$SampEn$	1.2373 ± 0.5201	0.9994 ± 0.0148	0.00001*
	DFA_1	0.9205 ± 0.1644	0.9904 ± 0.0458	0.0013*
SpO_2	μ	0.9912 ± 0.0131	0.9991 ± 0.0009	0.00001*
	σ	0.9694 ± 0.5643	0.9961 ± 0.0596	0.0001*
	Sk	0.4974 ± 4.606	0.9568 ± 0.5865	0.00006*
	Kt	0.8523 ± 0.6541	0.9829 ± 0.1407	0.0467
	$SampEn$	1.0532 ± 0.3084	1.0146 ± 0.0394	0.0009*
	DFA_1	1.0018 ± 0.0197	0.9992 ± 0.0217	0.0009*
MABP	μ	1.0187 ± 0.1217	0.9982 ± 0.0066	0.0045*
	σ	0.8948 ± 0.2925	0.9885 ± 0.0393	0.0169*
	Sk	0.3135 ± 0.8808	0.9086 ± 1.2939	0.0038*
	Kt	0.8517 ± 0.3024	0.9697 ± 0.0696	0.0366
	$SampEn$	1.0635 ± 0.1708	1.0055 ± 0.0088	0.004*

4.5.1 Statistical analysis

In this section, a statistical analysis is performed to identify parameters that show a significant difference between ARDS and non-ARDS groups using existing statistical tests. For each parameter, values were compared between both groups using the two-sample Kolmogorov Smirnov test for one training set. Results are presented in Table 4.1. A p-value < 0.05 was considered as significant. Parameters having p-values < 0.02 are indicated by a "*". Different linear and non-linear parameters for each signal have shown significant difference between groups. Moreover, we illustrated some examples of boxplots for relevant parameters having p-values < 0.02 for both ARDS and non-ARDS groups in Figures 4.5, 4.6, 4.7 and 4.8 for HR, RR, SpO_2 and MABP respectively.

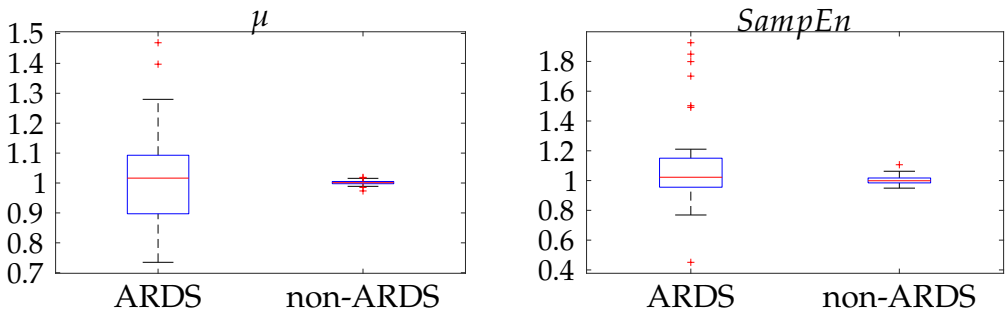


FIGURE 4.5: Boxplots for the relevant parameters of HR signal

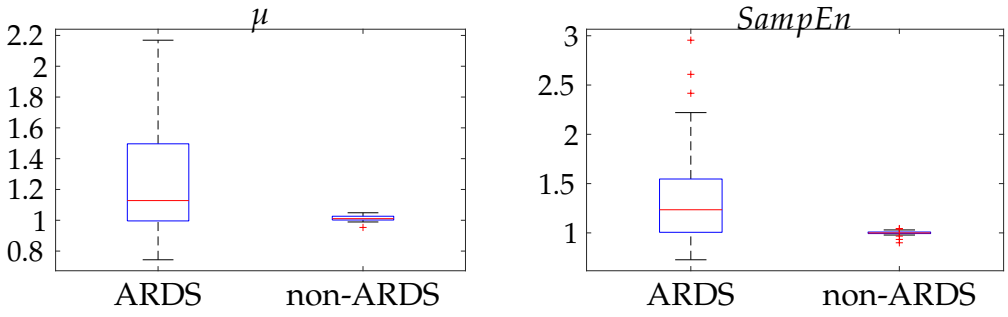


FIGURE 4.6: Boxplots for the relevant parameters of RR signal

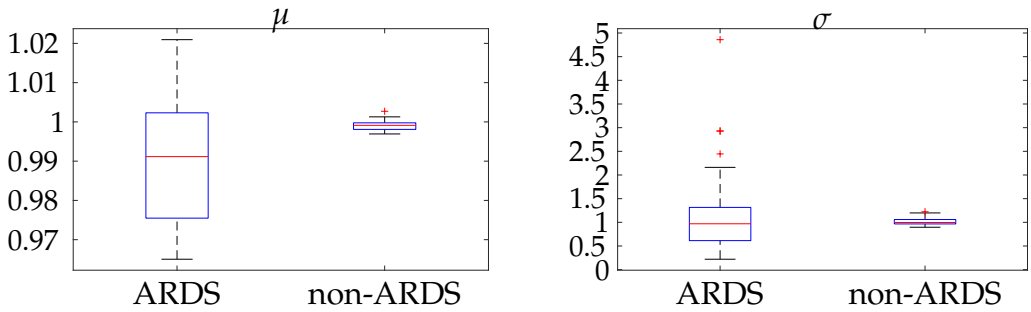


FIGURE 4.7: Boxplots for the relevant parameters of SpO_2 signal

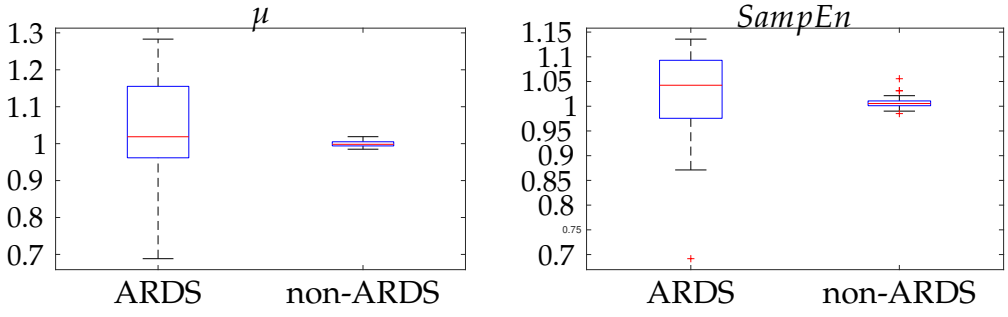


FIGURE 4.8: Boxplots for the relevant parameters of MABP signal

It can be shown that the mean of ARDS group is higher than that of non-ARDS group for HR, RR and MABP signals; whereas it is lower in non-ARDS group for SpO_2. For ARDS subjects, HR, RR and MABP increase with the development of ARDS, while SpO_2 decreases. σ of HR is greater in ARDS group than in non-ARDS group and lower for RR, SpO_2 and MABP. HR's distribution is then widely spread in ARDS group. The Skewness of non-ARDS group is higher than that of ARDS group for all the signals. Therefore, distributions are more asymmetric in non-ARDS groups. In addition to being asymmetric, kurtosis shows higher values in non-ARDS group for RR, SpO_2 and MABP, thus their distributions have heavy tails. Moreover, For ARDS group, sample entropy is higher in all signals, thus more complexity is shown in ARDS signals. Detrended fluctuation analysis factors are higher in non-ARDS group for HR and RR; while they are higher in ARDS group for SpO_2.

4.5.2 Parameters Selection

Since there is a high number of parameters, we propose a selection procedure. We first rank all the parameters based on the error rates, than we apply the selection procedure. The ranking is illustrated in Figure 4.9. Only parameters having error rates lower than 0.4 are presented in the figure. The results obtained from the error rate ranking in the figure are confirmed from the statistical analysis shown in Table 4.1. As shown, most of the parameters that have low errors in Figure 4.9 also showed a significant difference in the table. This figure shows that the sample entropy extracted from RR signals presented the lowest error rate among all parameters extracted from the four signals, thus ranked at the top.

From these ranked parameters, a sequential feature selection is considered to reduce the input vector of the belief computations. Figure 4.10 illustrates the selection procedure on the training set for both cases with and without correlations. The plot shows the accuracy as a function of the indicators of the ranked parameters. The accuracy is defined by the proportion of correctly identified subjects among the total number of subjects. A marker is placed on the curves when a parameter is selected.

Starting with *SampEn* from RR, the parameters presented in Figure 4.9 are added sequentially until the accuracy decreases more than $\epsilon > 5\%$. For the correlation study, if a parameter shows a correlation with any existing

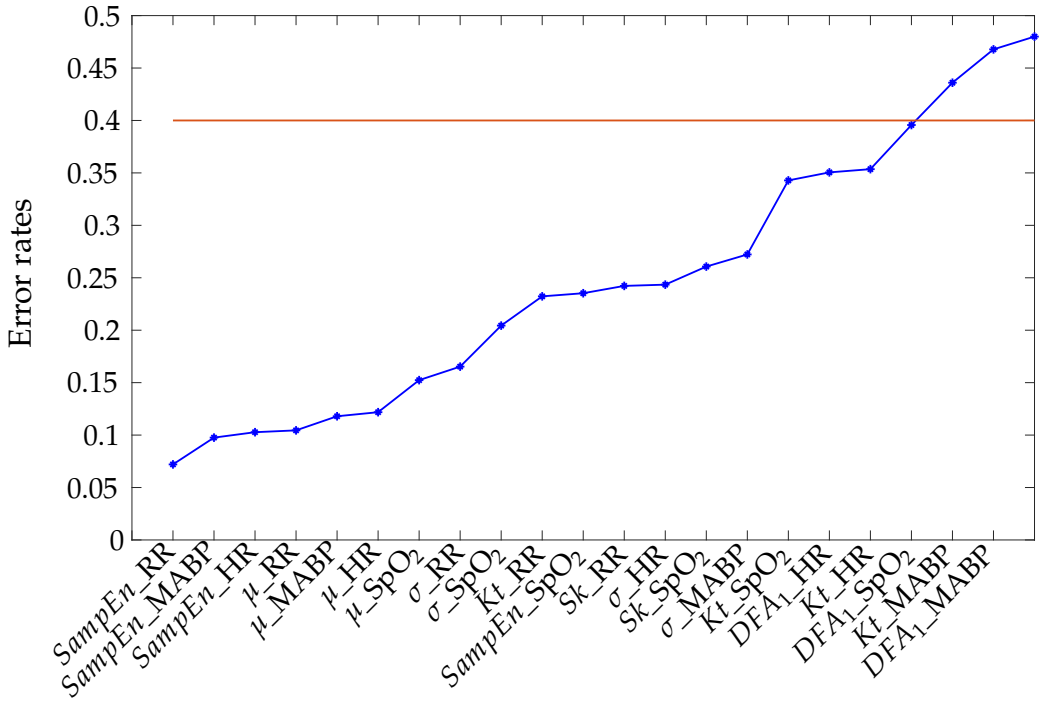


FIGURE 4.9: Ranking of parameters according to error rates. Only parameters having an error lower than 0.4 are included in the figure

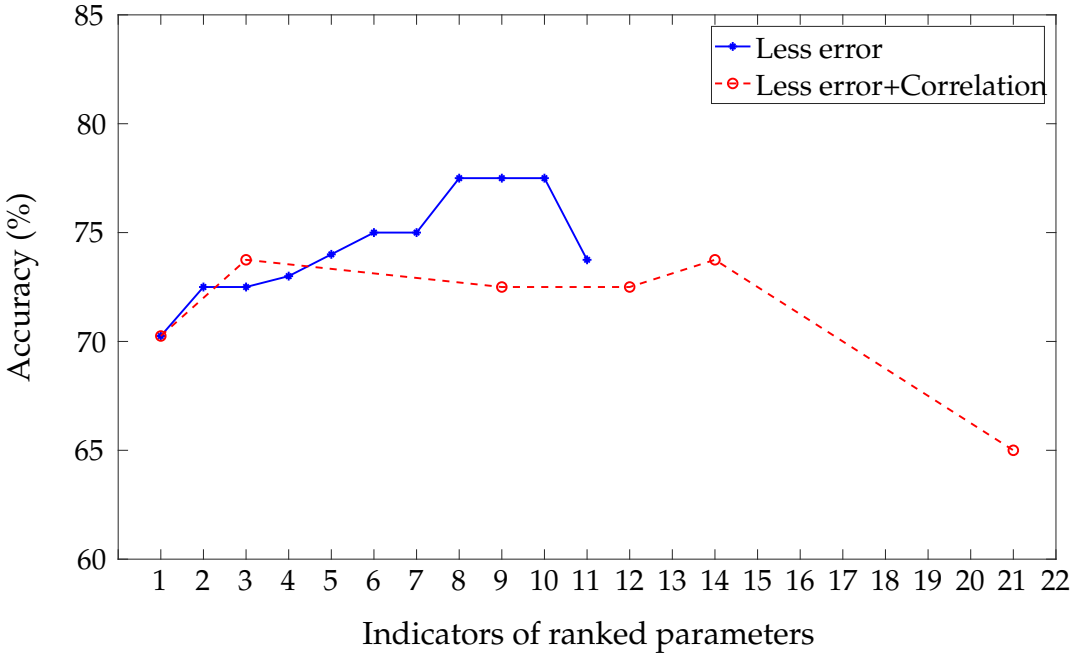


FIGURE 4.10: The parameters selection procedure using the simple procedure and the correlation extension procedure

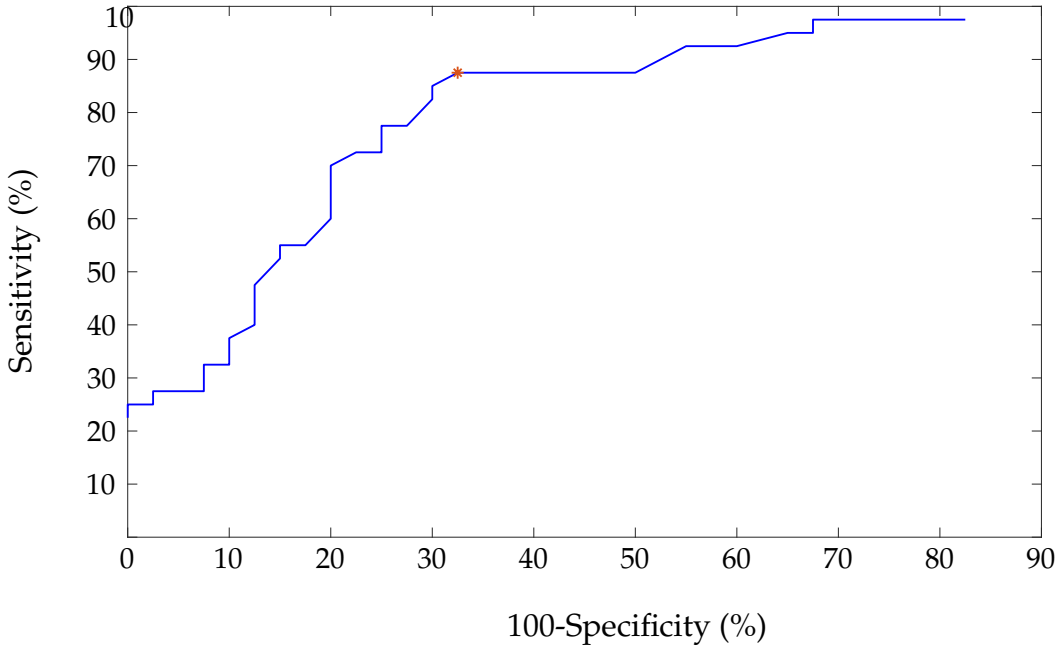


FIGURE 4.11: ROC curve for the selection of the threshold for alert generating

parameter with a p_value lower than 0.05, it is excluded from the selection. The simple selection procedure finished on the insertion of the eleventh parameter, that is, *SampEn* from SpO_2. Thus, ten parameters are selected in the simple SFS procedure which are μ and *SampEn* from both HR and MABP, μ, σ, *Kt* and *SampEn* from RR and μ and σ from SpO_2. However, the correlation study leads to five parameters presenting an accuracy of 73% which is lower than that obtained from the simple selection.

4.5.3 Selection of the alert generation threshold

In order to make a decision on the health state of a new test subject, a threshold must be defined from the training set. Therefore, we sweep the value of successive decisions on the interval $[1, ..., 30]$ hours and the performances *Se* and *Sp* are computed. The threshold that maximizes the Youden index is then considered. Figure 4.11 illustrates the ROC curve of the proposed algorithm by taking the fusion of the first ten parameters obtained in the previous section. The plot shows that the optimal cut-off value is obtained for threshold $= 21$ hours with $Se = 87.5\%$ and $Sp = 67.5\%$.

4.5.4 Selection of window size

As described in the real-time model we have used a fixed-length window for extracting the parameters vector for all subjects. Thus, the selection of

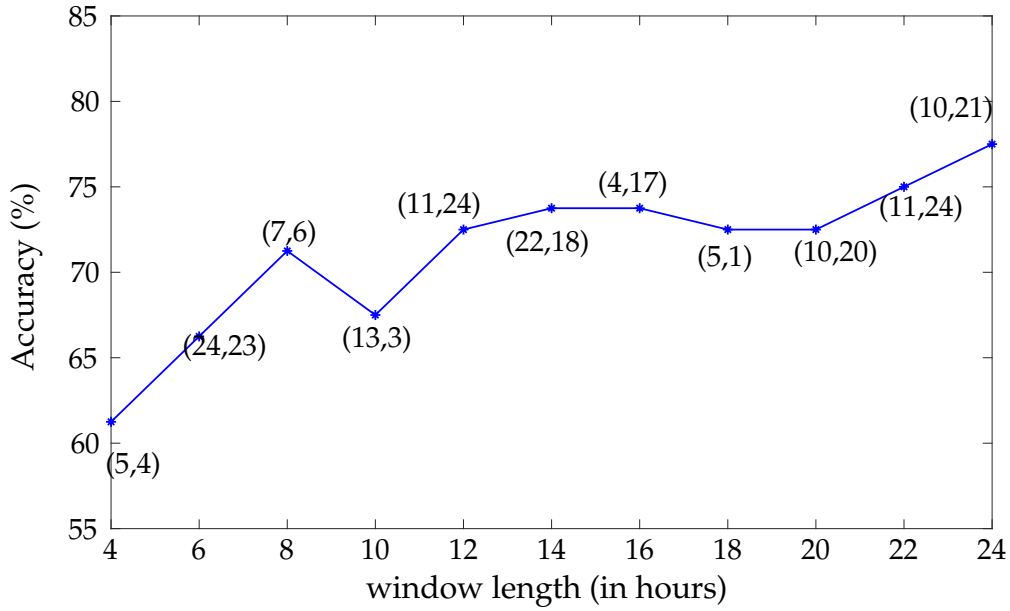


FIGURE 4.12: Changes in the accuracy over the training set with windows length. Values in the figure denote (number of parameters, threshold)

the size of the window is important in this study. A short window may not contain enough information and a long window may eliminate an abnormal change in the signal. Different window sizes are evaluated in this study, values are considered in the range $[4, 6, \ldots, 24]$ hours. Figure 4.12 illustrates the accuracy for each window size (in hours) as well as the number of selected parameters and the threshold shown as a couple (number of parameters, threshold). As shown in the figure, window size $= 24$ hours presents the highest accuracy of 77.5% on the considered training set.

4.5.5 Performance Evaluation

After the illustration of the extraction and selection of the parameters, the model is tested on the test set. Table 4.2 presents an example of the performances of the training and testing sets obtained after the selection of parameters and the optimal threshold and window size. We obtained with the considered training set an accuracy of 77.5% and on the testing set an accuracy of 75%.

In this paragraph, we test the overall performance of the model and we compare the influence of each added phase to arrive to the complete model. In order to ensure the stability of the results, we performed a 5-fold cross validation repeated 10 times. Then, the final performance is obtained

TABLE 4.2: Performance of the training and testing on the example considered previously

Performance	Training	Testing
Se (%)	87.5	70
Sp (%)	67.5	80
Acc (%)	77.5	75

TABLE 4.3: Performance of the proposed model and the influence of each phase

Method	training set		test set	
	Se(%)	Sp(%)	Se(%)	Sp(%)
BF without discounting, nor selection	70.23	71.74	70.16	67.89
BF without discounting	78.10	66.91	72.59	61.16
complete BF model	81.65	74.72	77.24	71.25
complete BF model+corr	80.18	69.29	74.10	61.12

by computing the average of the obtained performances over the cross validation procedure. Table 4.3 presents the performance of each added phase to obtain at the end the performance of the complete proposed model. The modeling of the data by associating and combining masses from all the parameters leads to a sensitivity of 70.23% and a specificity of 71.74% over the training sets and Se = 70.16% and Sp = 67.89% over the test sets. However, when the parameters selection phase was performed, an enhancement in the sensitivity was noted over the training sets (Se = 78.10%, Sp = 66.91%) and the test sets (Se = 72.59%, Sp = 61.16%). In addition, discounting the mass functions using the error rates improved the performance to 81.65% of sensitivity and 74.72% of specificity over the training sets and 77.24% and 71.25% for sensitivity and specificity respectively over the test sets. However, when the correlation was considered in the selection phase the performances decreased with Se = 80.18% and Sp = 69.29% over the training sets and 74.10%, 61.12% over the test sets.

We recall that the previously proposed decision fusion model has achieved an accuracy of 72.48% over training sets and 67.78% over testing sets using the non-invasive ventilated subjects. This feature fusion evidence-based model has achieved an accuracy of 78.18% over training sets and 74.25% over test sets, thus higher than the previous proposed decision fusion model.

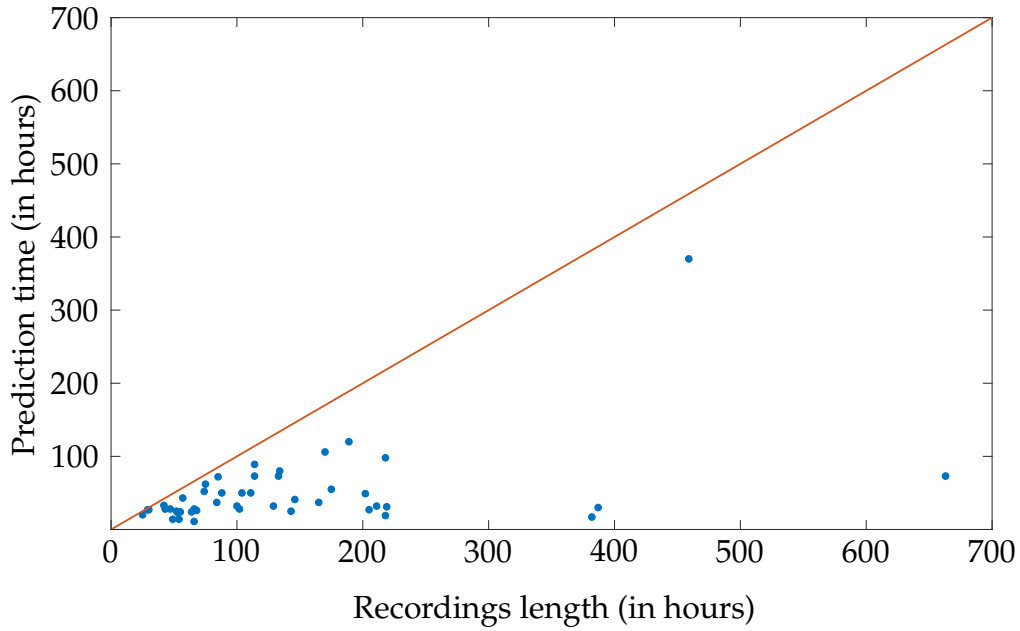


FIGURE 4.13: Early-predictions of ARDS subjects using the proposed feature fusion approach

4.5.6 Early predictions

In this section, we test the early prediction of the proposed model for the binary problem. Figure 4.13 illustrates the prediction time of ARDS versus the recordings length or the actual time of ARDS occurrence. As shown in the figure, the latest prediction of ARDS was 2 hours of its occurrence. Alerts were generated on an average of 89 hours before the occurrence of ARDS.

4.5.7 Comparison to the state-of-the-art methods

The proposed binary model is compared to some classification techniques such as Bayesian hypothesis test, K-nearest neighbors (K-NN) and SVM. The Bayesian hypothesis test is based on Bayes theorem. It is a probability model that consists of factorizing a joint probability distribution into a set of conditional distributions for each variable. The K-nearest neighbors is one of the basic algorithms in machine learning. It classifies a new observation to a class by choosing the majority class among the K nearest neighbors in the training set. There exist different distance functions to choose the closest neighbors, such as the Euclidean, the Manhattan or the Minkowski distance function. The value K can be optimized from the training set as being the minimum number of nearest neighbors that gives the optimal features set. In this chapter, a ten-fold cross validation was performed to obtain K that minimizes the test loss. The optimal K was found to be 8. SVM, as described

TABLE 4.4: Comparison to state-of-the-art classification methods

Method	Sensitivity (%)	Specificity (%)
K-NN	59.20	79.75
Bayesian test	67.74	65
SVM	58.67	74.06
The proposed Model	**77.24**	**71.25**

in Chapter 3, has proven to be a good model to classify data. It can be used to differentiate the data using a decision boundary that produces the optimal separation of classes.

Table 4.4 shows the performance indexes of the proposed model compared to these described techniques. As shown, the proposed model outperforms the *K*-NN, the SVM as well as the Bayesian tests, which are a simplest form of the belief functions theory where the ambiguities between the sets are not considered. These results show the importance of the proposed model.

4.5.8 Multi-class classification

In this section, we present the results of the proposed method in detecting each of the ARDS phases. Thus, new groups of subjects are defined using the PaO_2/FiO_2 values. This leads to 14 subjects in the severe group, 36 subjects in the moderate group, 15 mild ARDS subjects and 35 non-ARDS subjects. In order to validate the model, a holdout validation procedure is considered by taking 5 subjects from each group for testing and the remaining subjects for the training step. Thus, we obtain a total of 80 subjects for training and 20 subjects for testing. The validation procedure is repeated 10 times.

Once the training data groups are constructed and balanced, all the distribution functions of the subsets of ω for each parameter are estimated. Then, the model of the belief functions is defined as described in Section 4.3. Hence, given the signals in real-time for a new test subject, the selected parameters are extracted from the signals. Then, mass functions are computed for each extracted parameter, discounted using the training discounting rate, combined and transformed using the pignistic transformation. The results of the proposed approach are presented in Table 4.5. The first row refers to the belief functions model without the discounting operation. The second represents the results of the model with the basic discounting and the last

TABLE 4.5: Accuracies of the multi-class classification for the training phase

BF Model	Average (%)	Severe (%)	Moderate (%)	Mild (%)	non-ARDS (%)
	45.62	88	22.5	14.5	57.5
discounting	46.96	82.85	24.28	14.28	66.43
contextual	39	19	55.5	58	23.5

row goes for the contextual discounting.

As shown in the table, the belief functions model gave good accuracies for the severe and non-ARDS groups. This could be caused by the distribution functions of each group. Severe and non-ARDS groups have dominant distributions over moderate and mild distributions. Moreover, the basic discounting represents the best average accuracy over the training sets. This type of discounting has mainly increased the accuracy of the non-ARDS group with bare changes in ARDS groups accuracies. However, the contextual discounting has increased the accuracies of the moderate and mild ARDS with sharply decreasing the accuracies of severe and non-ARDS groups. It is worth noting that contextual discounting attributes high confidence measure to minor mass functions.

In the training set, a subject is correctly identified if the generated decision is equal to the actual label of the subject. However, since the analysis is performed in real-time for the testing phase, decisions are generated continuously with time. Indeed, a subject might develop a less severe form of ARDS before it develops the actual condition. Thus, if the generated decision belongs to the actual group or to a less severe one, the subject will be considered correctly identified. Table 4.6 illustrates the accuracies over the test sets for each of the belief model phases. As shown, the contextual discounting presents the highest average accuracy of 64%, with high accuracies for ARDS groups and the worst accuracy for the non-ARDS group. These results are evident because of the decision rule and the higher importance that is given to the moderate and the mild groups as shown in the training results. The basic discounting rule has shown high performances for severe and non-ARDS groups, the accuracy of the moderate group has also increased. However, mild group shows low accuracy.

TABLE 4.6: Accuracies of the multi-class classification for the testing phase

BF Model	Average (%)	Severe (%)	Moderate (%)	Mild (%)	non-ARDS (%)
	41.15	94	30	2	40
discounting	54.5	86	44	14	74
contextual	64	100	82	56	18

TABLE 4.7: Confusion matrix for basic discounting

	SEV	MOD	Mild	$\overline{\text{ARDS}}$
SEV	18	21	4	7
MOD	20	18	4	8
Mild	9	11	7	23
$\overline{\text{ARDS}}$	6	4	3	37

TABLE 4.8: Confusion matrix for contextual discounting

	SEV	MOD	Mild	$\overline{\text{ARDS}}$
SEV	11	15	24	0
MOD	8	14	27	1
Mild	3	9	28	10
$\overline{\text{ARDS}}$	9	14	18	9

Tables 4.7 and 4.8 present confusion matrices for the basic and the contextual discounting, respectively. The confusion matrices represent the number of test subjects identified for each group. We recall that the validation procedure considers 5 subjects from each group in the testing set. This validation is repeated 10 times to obtain finally 50 subjects for each group. It is shown from the Tables that the basic discounting gives more importance to severe and non-ARDS groups; whereas the contextual discounting gives priority to moderate and mild groups, with mild group having higher priority than moderate one.

4.6 Discussions

This chapter proposes a feature fusion model for predicting ARDS in real-time prior to its onset. The model consists of extracting linear and non-linear parameters from the time series data and combining them using the belief functions theory. In order to validate the proposed model, we considered only the non-invasive ventilated subjects from the MIMIC II database. Two concepts of predictions were presented, which are the binary prediction and the four-group prediction concepts.

Despite the intense research on ARDS, there is a lack in the literature for the characterization of ARDS using features extraction. Previous studies were conducted on the prediction of ARDS or the mortality related

to ARDS using computed tomographic images (Ichikado et al., 2006) or clinical variables (Navarrete-Navarro et al., 2000; Ennett et al., 2008), which necessitate clinical exams or the presence of subjects in the hospitals for long periods. However, the proposed approach uses only vital signs that can be collected using non-invasive techniques, thus without disturbing the subjects. Moreover, the previous studies present general statistical models to detect ARDS assuming that all subjects will present similar changes; while this approach is a subject-based one that considers an initial stable state for each subject and computes the ongoing parameters with respect to the stable state. In addition, an information fusion was handled using the belief functions theory to consider the ambiguity and unreliability of the sources.

Hence, in this chapter, linear and non-linear parameters were extracted from the time series as mean, standard deviation, skewness, kurtosis, sample entropy and detrended fluctuation analysis. Linear parameters provide information regarding the distributions of the data. Distributions were shown to be asymmetric, having wide tails and being widely spread. This verifies that biomedical data are not Gaussian (Chua et al., 2010b). Conversely, it was found that all the signals had higher sample entropy in ARDS group, thus increased complexity (Richman and Moorman, 2000). From DFA_1 values, we can notice that SpO_2 had higher short-term correlations in ARDS subjects, while HR and RR had lower short-term correlations (Peng et al., 1995). The HR had higher long term correlations in ARDS subjects.

From the results, the reduction of the number of parameters, by performing a selection procedure, improved the performance of the model and reduced the complexity of the computations as well. This way only parameters having low error rates are included in the model and the combination that presents the best local accuracy is considered. Then, the inclusion of the conditional reliability of each source has led to an enhancement in the accuracy of the model over the training and the test sets. Moreover, the model has successfully generated early alerts of ARDS at least two hours before its occurrence and on average of 89 hours, thus more than 3 days.

The approach proposed in this chapter outperformed the model based on decision fusions proposed in the previous chapter. Moreover, it is shown from the results that this model outperformed the classical classifiers tested similarly on the data. It is worth noting that the Bayesian tests perform

only on singletons and the state membership is determined by computing the likelihood ratio. However, the added ambiguity set in the proposed evidence-based model may affect the final decision since it is considered in both discounting and combination phases. Thus, the advantage of such models over the classical classifiers, especially the Bayesian tests, is the inclusion of higher order sets that illustrate the ambiguity intervals.

In addition, the study of the four-group classification has lead to good average accuracies for the basic discounting rule over the training sets and for both basic and contextual discounting rules over the test sets. The basic discounting has enhanced the accuracy of the non-ARDS group while maintaining a high accuracy for the severe ARDS group. However, the contextual discounting has increased the accuracies for all the ARDS groups but it decreased the accuracy of the non-ARDS group. Thereafter, we obtain two different models, the first one is a high sensitive model for ARDS degradations and the second one is a sensitive model for severe ARDS and at the same time specific for non-ARDS subjects.

4.7 Conclusion

This chapter proposes a feature fusion model for the prediction of ARDS based on the belief functions theory. This model has achieved high performances in the binary identification as well as in the four-class identification. In addition, it outperforms state of the art techniques. The mass functions were computed by a distribution fitting method. The next chapter proposes a novel mass assignment method for the evidence-based feature fusion technique.

Chapter 5

ARDS Prediction using Belief Functions and Kernel Methods

Contents

5.1 Introduction 110

5.2 Literature review on masses definition 110

5.3 Evidence-based approach using kernel methods 112

5.3.1 Description of the approach 112

5.3.2 The recurrence quantification analysis 113

5.3.3 Belief functions approach using kernels 115

5.3.4 Selection of parameters 122

5.4 Results 123

5.4.1 Statistical analysis 124

5.4.2 Optimization of the KRR parameters 124

5.4.3 Selection of parameters 125

5.4.4 Selection of the alert generation threshold 126

5.4.5 Performance evaluation 127

5.4.6 Early predictions 128

5.4.7 Comparison with previous models 129

5.5 Discussion 130

5.6 Conclusion 131

5.1 Introduction

In the previous chapter, we presented an approach for the prediction of ARDS based on feature fusion concept. Different linear and non-linear parameters were extracted from the signals. The approach was based on belief functions theory using distributions of parameters and then comparison of probabilities to identify mass functions.

This chapter proposes a novel approach for predicting ARDS that constructs the mass functions associated to the parameters using a machine learning algorithm. Indeed, the mass functions are computed for each parameter using a kernel-based regression technique. They are then combined by the conjunctive fusion function from the belief functions theory. In addition, a new set of parameters is extracted from the signals based on the recurrence quantification analysis. In this chapter, we consider the binary prediction problem described previously. The next section presents a literature review on the definition of mass functions and, in the following section, the proposed model is described.

5.2 Literature review on masses definition

The belief functions theory has attracted extensive attention for its advantages in combining information from distinct sources (Sun, Huang, and Miao, 2008; Khaleghi et al., 2013). It models all possible classes of an object, i.e., it considers the unions of objects and not only single objects. There exist various studies that propose extensions to belief functions especially for discounting and combination. Deng has proposed a generalized evidence theory (GET) that addresses the conflict management in an open world environment (Deng, 2015). In addition, the belief functions theory has been extended to handle various types of uncertainty using fuzzy set theories (Chou et al., 2011; Zhang et al., 2013; Kahraman, Onar, and Oztaysi, 2015), genetic algorithm (Deng, Liu, and Zhou, 2015) and many others. A modified average method to combine belief functions based on distance measures of evidence has been proposed in (Yong et al., 2004). An evidence reasoning rule has been proposed by Xu *et al* to improve the Dempster's combination and add the reliability and weight of evidence (Xu et al., 2017).

Even though there is an extensive research in this area, research neglects a very important problem. The belief functions theory relies on mass functions to model the data. The definition of the belief measure and the modeling of the data is one of the most important problems in evidence theory. This measure could be a measure of similarity, satisfiability, resemblance or inclusion. It is based on the characteristics of the data to compare and the purpose of the comparison (Diaz, Rifqi, and Bouchon-Meunier, 2006). The mass function refers to the basic probability assignment (BPA) where masses are assigned using the probabilities of each class using the training objects that belong to this class (Dempster, 1967; Xiao, 2019). A very practical BPA construction consists of deriving BPA from confusion matrix (Deng et al., 2016). Yager has introduced a whole class of fuzzy measures that can be associated with a belief function structure (Yager, 1999). Xu *et al* have proposed a method to obtain mass functions by building a normal distribution-based model for each attribute of the data (Xu et al., 2013). On the other hand, Denoeux has proposed a mass function construction based on neural network classifier (Denoeux, 2000). In Denoeux's work, the determination of masses is implemented in a multilayer neural network having one input layer and two hidden layers. The output vector of the neural network is then used to determine the masses based on the distance of pattern to its *K*-nearest neighbors (*K*-NN) prototypes. A similar work was done by Hu *et al* to compute mass functions from the output of neural network model (Hu et al., 2003). Zhang *et al* have proposed to determine masses from core samples (Zhang et al., 2014). Core samples are generated for each attribute model; masses are then assigned based on the distance between test data and the core samples. Another construction method, proposed in (Zhang et al., 2018), relies on the distance between the objects and the training classes. Then, the distances are converted to masses by a mapping function, such as the exponential function. There are many types of functions in the literature satisfying the principles of mapping functions.

In the previous chapter, we have proposed a feature fusion model to predict ARDS. The model uses belief functions theory to combine mass functions obtained from different parameters. The mass functions were obtained by a fitting of distribution functions from a training dataset. The current chapter proposes an extension of the previous one. It presents a novel approach for the construction of mass functions using a kernel-based machine learning

algorithm. Then, mass functions are combined using the normalized conjunctive rule of combination. Details on the construction of masses are given later on. This is, to our knowledge, the first implementation of kernel methods into belief functions.

5.3 Evidence-based approach using kernel methods

This chapter proposes a novel approach for the construction of mass functions in the belief functions theory. We reconsider in this chapter the problem of binary prediction, where two groups of subjects are considered, the ARDS group that includes severe and moderate ARDS subjects and the non-ARDS group including mild and non-ARDS subjects. Mass functions are constructed using a kernel-based regression model. We also propose in this chapter to extend the input vector to a higher dimensional vector by adding new non-linear parameters using the recurrence quantification analysis.

5.3.1 Description of the approach

The block diagram of Figure 5.1 presents the model proposed in this chapter that extends the one developed in the previous chapter. The main objective of this thesis is to develop a model for the prediction of ARDS in real-time using only vital signs. Given the four extracted signals $x_{s,i}(1:t)$ in real-time for a subject s, a new extraction of non-linear parameters is proposed. Thus, each signal i goes through a feature extraction phase to extract the previously explained parameters in Chapter 4 and a new set of parameters from the recurrence quantification analysis. Then, the total set of parameters is used in a kernel-based regression phase in order to assign masses to each state $\{+1\}$, $\{-1\}$ and $\{+1,-1\}$ according to each parameter.

Afterwards, in the belief framework, the masses are discounted using the contextual discounting and combined using the normalized conjunctive rule of combination in order to assign a final mass to each state. The state having the highest mass will represent the output $\hat{y}_s(t)$ of the model, that is, the selected state. With the addition of new parameters, we obtain an input space with higher dimensionality. Thus, we propose in this chapter a selection procedure based on the fusion performance. More details are provided in the following.

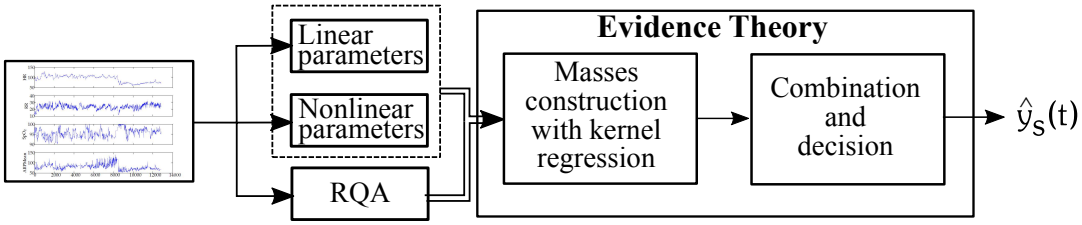


FIGURE 5.1: Block diagram of the proposed approach

5.3.2 The recurrence quantification analysis

The recurrence quantification analysis (RQA) quantifies the number and duration of recurrences of a dynamical system presented by its phase space trajectory. The reconstructed phase space is a space that represents all possible states of a system. It consists of transforming time-lagged copies of the initial signal onto axes of a new high dimensional space (Takens, 1981). Each axis in the phase space represents a dynamic variable in the data. It is constructed using a time delay and an embedding dimension. The choice of the appropriate time delay t_d and embedding dimension M is very important to ensure the best representation of the initial data (Buzug and Pfister, 1992; Cellucci, Albano, and Rapp, 2003). The vector $x_{s,i}(\ell)$ at time ℓ is defined by $X_{s,i}(\ell)$ in the reconstructed phase space and can be written as follows:

$$X_{s,i}(\ell) = [x_{s,i}(\ell), x_{s,i}(\ell + t_d), x_{s,i}(\ell + 2t_d), \ldots, x_{s,i}(\ell + (M-1)t_d))].$$

A correct choice of the time delay and the embedding dimension results in a good representation of the data. The reconstructed phase space trajectories will also represent at best the data (Takens, 1981). The time delay t_d is generally chosen by minimizing the average mutual information (Fraser and Swinney, 1986). However, the embedding dimension M is estimated using the algorithm of the false nearest neighbors (Kennel, Brown, and Abarbanel, 1992).

The recurrence plot proposed by Eckmann *et al.* projects the M-dimensional phase space trajectory through 2-dimensional illustration of its recurrences (Eckmann, Kamphorst, and Ruelle, 1987). The recurrence plot can be shown as a square matrix with elements corresponding to the times at which a state is repeated. It consists of two types of lines, namely, diagonal and vertical lines. Diagonal lines correspond to periods where phase space segments run parallel, i.e., the trajectory visits the same region of the phase space at different times and vertical lines represent periods when segments

remain in the same phase space region or when the state does not change or changes very slowly.

Different parameters can be extracted from the recurrence plot to characterize the diagonal and vertical lines. Parameters extracted from the diagonal lines are:

- The determinism (DET) - the percentage of recurrence points that form diagonal lines;
- The average diagonal line length ($\langle L \rangle$) - the average length of the diagonal lines;
- The longest diagonal line (L_{max})- the length of the longest diagonal line;
- The entropy (ENTR) - the Shannon entropy of the probability distribution of the diagonal line lengths;

Whereas parameters extracted from vertical lines are:

- The laminarity (LAM) - the percentage of recurrence points that form vertical lines;
- The trapping time (TT) - the average length of the vertical lines;
- The longest vertical line (V_{max}) - the length of the longest vertical line;

Finally, parameters extracted from the entire recurrence plot are:

- The recurrence rate R_eR_a - the percentage of recurrence points in an RP;
- The recurrence period density entropy (RPDE) - the normalization of the entropy by the longest diagonal length.

More details on the reconstructed phase space and the recurrence quantification analysis are provided in Appendix A.

Hereafter, we extract from each signal the linear parameters, which are the mean (μ), the standard deviation (σ), the skewness (Sk) and the kurtosis (Kt), as well as the non-linear parameters, which are the sample entropy ($SampEn$), both factors from the detrended fluctuation analysis (DFA_1 and DFA_2) and the already explained parameters from the recurrence quantification analysis. This parameter extraction leads to 16 parameters per signal and thus 64 parameters for each subject s at each time t. The parameters are

defined by $p_{s,j}(t), j = 1, \ldots, 64$. As explained previously, an initial segment $x_{s,i}(1:\tau)$ is considered from each signal representing the stable state of the subject s. $p_{s,j}^{(0)}$ is the j^{th} parameter extracted from the initial segment of the signals for a subjects s. Since each subject has its own normal state, we propose to normalize the extracted parameters in real-time using the parameters of the initial stable segment. Thus, $p_{s,j}$ denote the normalized parameter and is defined by $p_{s,j} = p_{s,j}(t)/p_{s,j}^{(0)}$.

5.3.3 Belief functions approach using kernels

In this chapter, we present a novel technique to assign mass functions in the belief functions theory using kernel methods from the machine learning techniques. Given the model of belief functions that operates on the frame of discernment $\Omega = \{+1, -1\}$, our aim is to find the mass functions $m_j(\cdot)$ for each subset $\omega \in 2^\Omega$ associated to each parameter $j \in J$. We recall that $2^\Omega = \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\}$ is the set of all possible subsets of Ω, with $\sum_{\omega \in 2^\Omega} m_j(\omega, s, t) = 1$ and $m_j(\emptyset, s, t) = 0$.

The training sets are considered to construct the mass functions related to each parameter. As described in the previous chapters, we consider the total length of recordings for non-ARDS subjects; whereas for ARDS subjects, recordings are cut at the last measurement before the diagnosis of ARDS. Then, parameters are extracted from the last segments of the trimmed recordings of ARDS subjects followed by a normalization, since the instability related to ARDS is mostly present at the end of the signals preceding the occurrence of ARDS. In parallel, parameters are extracted from the beginning of the signals for non-ARDS subjects. Couples of parameters extracted from training signals with their associated labels $(p_{n,j}, y_n)$, $n \in \{1, \ldots, N\}$, are obtained, with $y_n = +1$ for ARDS or $y_n = -1$ for non-ARDS. Therefore, the computation of mass functions is done using these couples as follows.

5.3.3.1 Mass assignment by kernel methods

In this section, we present the construction of mass functions using kernel-based regression analysis. Regression analysis consists of predicting a dependent variable from given input variables. Here, the objective is to find a set of functions $g_j(\cdot)$ that take the values of the parameters j and the subsets $\omega \in 2^\Omega$ as inputs and yield the masses associated to ω as outputs.

Since we are working on the binary problem, a first intuitive proposition to construct mass functions is to compute a scalar function $\psi_j(\cdot)$ for each parameter j using its corresponding training dataset. Indeed, the function $\psi_j(\cdot)$ takes the parameters values $p_{n,j}$, $n = \{1, ..., N\}$, as inputs and yields the labels $y_n = +1$ for ARDS subjects and $y_n = -1$ for non-ARDS subjects. Details on the definition of $\psi_j(\cdot)$ are given in the next paragraph. It is obvious that the functions $\psi_j(\cdot)$ are not masses since they can yield negative outputs. In order to define the mass functions on the frame Ω while having a new parameter $p_{s,j}(t)$, we propose to compute $\psi_j(p_{s,j}(t))$, then

- if $sign(\psi_j(p_{s,j}(t))) > 0$,

$$\begin{cases} m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = \psi_j(p_{s,j}(t)), \\ m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = 0, \\ m_j(\{+1, -1\}, s, t) = g_j(\{+1, -1\}, p_{s,j}(t)) = 1 - \psi_j(p_{s,j}(t)); \end{cases} \tag{5.1}$$

- if $sign(\psi_j(p_{s,j}(t));) \leq 0$,

$$\begin{cases} m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = 0, \\ m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = |\psi_j(p_{s,j}(t))|, \\ m_j(\{+1, -1\}, s, t) = g_j(\{+1, -1\}, p_{s,j}(t)) = 1 - |\psi_j(p_{s,j}(t))|; \end{cases} \tag{5.2}$$

where $|\cdot|$ is the absolute value of its argument. This definition can be rewritten as follows,

$$\begin{cases} m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = \max(\psi(p_{s,j}(t)), 0), \\ m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = |\min(\psi(p_{s,j}(t)), 0)|, \\ m_j(\{+1, -1\}, s, t) = g_j(\{+1, -1\}, p_{s,j}(t)) = 1 - |\psi(p_{s,j}(t))|. \end{cases} \tag{5.3}$$

This mass computation assigns a null mass to either one of the singletons subsets of Ω, which might penalize the masses combination. Alternatively, we propose a different way of constructing mass functions by considering a label vector Y_n instead of the scalar label for each subject n of the training dataset. Indeed, an ARDS subject gets a label [1 0] instead of "+1"; whereas a non-ARDS subject gets a label of [0 1] instead of "-1". After that, the model $\boldsymbol{\psi_j}(\cdot)$ is constructed and optimized using the training dataset $(p_{n,j}, Y_n)$, $n \in \{1, \ldots, N\}$. Then, for a new given parameter's value in real-time $p_{s,j}(t)$, $\boldsymbol{\psi_j}(p_{s,j}(t))$ is computed. It gives a bi-dimensional vector output, where the first element refers to ARDS state and the second one

refers to non-ARDS state. The mass attributed to the subset $\{+1\}$ is thus equal to the output of the $\boldsymbol{\psi_j}(\cdot)$ related to the ARDS state, that is, the first element of the vector, $m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = \psi_{j,1}(p_{s,j}(t))$, the mass attributed to the subset $\{-1\}$ is equal to that of the non-ARDS state $m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = \psi_{j,2}(p_{s,j}(t))$, and the mass attributed to the total set is equal to $m_j(\{+1, -1\}, s, t) = g_j(\{+1, -1\}, p_{s,j}(t)) = 1 - \psi_{j,1}(p_{s,j}(t)) - \psi_{j,2}(p_{s,j}(t))$.

5.3.3.2 Kernel-based regression methods

In this paragraph, we describe the different kernel techniques that can be applied to construct the models $\psi_j(\cdot)$ or $\boldsymbol{\psi_j}(\cdot)$ while having training datasets $(p_{n,j}, y_n)$ or (p_{n_j}, Y_n), $n = 1, \ldots, N$, respectively. While the first definition considers scalar models $\psi_j(\cdot)$ and thus scalar labels y_n, the second definition introduces vectorial models $\boldsymbol{\psi_j}(.) = [\psi_{j,1}(\cdot)\ \psi_{j,2}(\cdot)]$ with vectorial labels Y_n. In the following, the notations $\psi_j(\cdot)$ and y_n will represent respectively the model and the scalar labels for the first definition or one element of the vectorial function and its corresponding label's element for the second definition for simplicity. Kernel methods map the input p into its label output y: $p \rightarrow y = \psi_j(p)$ which is induced by a reproducing kernel $\kappa(\cdot, \cdot)$ without the need to access the feature vectors (Aronszajn, 1950). The mapping function $\psi_j(\cdot)$ is expressed in the Hilbert space $\mathcal{H}$ by an inner product:

$$\psi_j(p) = \langle \psi_j(\cdot), \kappa(\cdot, p) \rangle_{\mathcal{H}}. \tag{5.4}$$

It can be obtained by minimizing the error between the model's output $\psi_j(p_{n,j})$ and the desired output y_n defined using the training dataset as follows

$$\mathcal{L}\left((y_1, \psi_j(p_{1,j})), \ldots, (y_N, \psi_j(p_{N,j}))\right) + \eta \mathcal{R}(\|\psi_j\|^2_{\mathcal{H}}). \tag{5.5}$$

The first term of the expression represents an arbitrary cost function, the second one is a strictly monotonically increasing function on $[0, +\infty[$ that serves as a regularization term with η is a positive tunable parameter and $\|\cdot\|_{\mathcal{H}}$ is the norm in the Hilbert space.

Kernel methods rely on the representer theorem that maps the features into a space whose dimension does not exceed the dimension of the training set. According to the representer theorem, the function $\psi_j(\cdot)$ is described by a

simple linear combination of the kernels $\kappa(p_{n,j}, \cdot)$, having the following form:

$$\psi_j(\cdot) = \sum_{n=1}^{N} \lambda_{n,j} \kappa(p_{n,j}, \cdot). \tag{5.6}$$

The optimization problem is then described by a simpler form such as to find the parameters $\lambda_{n,j}$. Thus, the terms of the optimization problem in the expression (5.5) can take different forms leading to different optimization problems and different solutions. According to the considered loss function, one can obtain a kernel ridge regression (KRR) problem (Saunders, Gammerman, and Vovk, 1998), a support vector regression (SVR) problem (Vapnik, 1995) or a vector output-regularized least squares (vo-RLS) problem (Honeine, Noumir, and Richard, 2013).

The kernel ridge regression: The first optimization problem is defined using the kernel ridge regression or the least squares support vectors machine (Saunders, Gammerman, and Vovk, 1998). In this case, the first term of Eq. (5.5) is defined using the mean squared loss,

$$\mathcal{L}\left((y_1, \psi_j(p_{1,j})), \ldots, (y_N, \psi_j(p_{N,j}))\right) = \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \psi_j(p_{n,j})\right)^2.$$

The function $\psi_j(\cdot)$ is then determined by minimizing the mean quadratic cost. The regularization term is taken in its simplest form as $\mathcal{R}(\|\psi_j\|^2_{\mathcal{H}}) = \|\psi_j\|^2_{\mathcal{H}}$. The optimization problem can be written as follows:

$$\psi_j(\cdot) = \arg\min_{\psi_j \in \mathcal{H}} \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \psi_j(p_{n,j})\right)^2 + \eta \|\psi_j\|^2_{\mathcal{H}}. \tag{5.7}$$

By injecting Eq. (5.6) in Eq. (5.7), we get the following solution:

$$\boldsymbol{\lambda}_j = (K_j + \eta N I_N)^{-1} \mathbb{Y}. \tag{5.8}$$

$\boldsymbol{\lambda}_j$ is the $N \times 1$ coefficients vector having $\lambda_{n,j}$ as its n-th element, $\mathbb{Y}$ is the $N \times 1$ vector of labels having y_n as its n-th element and I_N is the $N \times N$ identity matrix. K_j is the $N \times N$ Gram matrix whose (n_1, n_2)-th element is $\kappa(p_{n_1,j}, p_{n_2,j})$.

The support vector regression: In the support vector regression (SVR), $\psi_j(\cdot)$ is determined over a margin with an acceptance of errors that are less

than ϵ (Vapnik, 1995). The cost function is given by:

$$\mathcal{L}\left((y_1, \psi_j(p_{1,j})), \ldots, (y_N, \psi_j(p_{N,j}))\right) = \frac{1}{2N} \sum_{n=1}^{N} \max\left(0, |y_n - \psi_j(p_{n,j})| - \epsilon\right),$$

where the regularization term is taken $\|\psi_j\|^2_{\mathcal{H}}$ and ϵ is a positive tunable parameter.

The optimization problem can be obtained using a dual formulation according to (McCormick, 1983; Smola and Schölkopf, 2004):

$$\begin{aligned}\psi_j(\cdot) = \arg\max_{\lambda,\tilde{\lambda}} & \sum_{n=1}^{N} \tilde{\lambda}_{n,j}(y_n - \epsilon) - \lambda_{n,j}(y_n + \epsilon) \\ & - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} (\tilde{\lambda}_{n,j} - \lambda_{n,j})(\tilde{\lambda}_{m,j} - \lambda_{m,j}) \kappa(p_{n,j}, p_{m,j}),\end{aligned} \tag{5.9}$$

with the following constraints:

$$\begin{cases} 0 \leq \lambda_{n,j}, \tilde{\lambda}_{n,j} \leq \frac{1}{2\eta N}, n \in \{1, \ldots, N\}, \\ \sum_{n=1}^{N} (\lambda_{n,j} - \tilde{\lambda}_{n,j}) = 0. \end{cases} \tag{5.10}$$

The function $\psi_j(\cdot)$ is then determined by solving the optimization problem with the constraints as follows:

$$\psi_j(\cdot) = \sum_{n=1}^{N} (\lambda_{n,j} - \tilde{\lambda}_{n,j}) \kappa(p_{n,j}, \cdot) + b, \tag{5.11}$$

where b is an offset term obtained by applying the Karush-Kuhn-Tucker conditions (Karush, 2014; Kuhn, 2014). Therefore, we obtain a quadratic programming problem that needs an off-the-shell optimization technique to be solved.

The vector-output regularized least square: In the vo-RLS technique, a multi-task learning is performed to estimate simultaneously all elements of a vectorial output (Honeine, Noumir, and Richard, 2013). This computation is interesting for the second vectorial definition for mass assignment. Herein, the vectorial function $\psi_j(\cdot)$ with its two elements is computed at once instead of determining separately the scalar functions $\psi_{j,1}(\cdot)$ or $\psi_{j,2}(\cdot)$ according to

each label's element. The global function $\psi_j(\cdot)$ is written as follows,

$$\psi_j(\cdot) = \sum_{n=1}^{N} \lambda_{n,j} Y_n \kappa(p_{n,j}, \cdot), \tag{5.12}$$

where $\boldsymbol{p}_n$ is the vector of the J parameters for a given n training sample, $\boldsymbol{p}_n = \{p_{n,1}, \ldots, p_{n,J}\}$. $\lambda_{n,j}$, $n \in \{1, \ldots, N\}$, are parameters to be defined. Then, the optimization problem can be rewritten in the following form

$$\psi_j(\cdot) = \arg\min_{\psi_j} \frac{1}{N} \sum_{n=1}^{N} \|Y_n - \psi_j(p_{n,j})\|^2 + \eta \|\boldsymbol{\lambda}_j\|^2. \tag{5.13}$$

$\boldsymbol{\lambda}_j$ is the $N \times 1$ vector whose n-th element is $\lambda_{n,j}$ We substitute the expression of $\psi_j(\cdot)$ given in Eq. (5.12) in the optimization problem of Eq. (5.13). This leads to the following problem:

$$\boldsymbol{\lambda}_j = \arg\min_{\boldsymbol{\lambda}} tr(\boldsymbol{Y}\boldsymbol{Y}^T) - 2d^T\boldsymbol{\lambda} + \boldsymbol{\lambda}^T G\boldsymbol{\lambda} + \eta N \boldsymbol{\lambda}^T\boldsymbol{\lambda}, \tag{5.14}$$

$tr(\cdot)$ is the matrix trace operator, G is the $N \times N$ matrix whose (u, v)-th entry is

$$G(u, v) = Y_u Y_v^T \sum_{n=1}^{N} \kappa(p_{u,j}, p_{n,j}) \kappa(p_{v,j}, p_{n,j}),$$

and d is the $N \times 1$ vector whose u-th entry is

$$d(u) = \sum_{n=1}^{N} Y_u Y_n^T \kappa(p_{u,j}, p_{n,j}).$$

The final solution is obtained by taking the gradient of the Eq. (5.14) with respect to $\boldsymbol{\lambda}$, which is $-v + G\boldsymbol{\lambda} + \eta N \boldsymbol{\lambda}$ and setting it to zero,

$$\boldsymbol{\lambda}_j = (G + \eta N \boldsymbol{I}_N)^{-1} \boldsymbol{d}.$$

The model $\psi(\cdot)$ can be determined after the calculation of the vector $\boldsymbol{\lambda}$.

According to a comparison study between these kernel-based techniques, it comes out that the ridge regression is more interesting than the SVR and the vo-RLS in terms of time complexity and performance. In addition, the SVR has an enormous time complexity comparing to the other techniques (Mahfouz et al., 2015). Even though the vo-RLS technique is a vectorial technique, however both elements of $\boldsymbol{\psi}(\cdot)$ will be affected by the same optimized coefficients $\boldsymbol{\lambda}_j$. Conversely, in the case of the two first techniques,

unique optimized $\boldsymbol{\lambda}_j$ will affect each of the elements of $\boldsymbol{\psi}(\cdot)$. Therefore, in this chapter we propose to illustrate the construction of mass functions using the kernel ridge regression technique.

In order to define the kernel function, a Gaussian kernel is considered as given in the following equation:

$$\kappa(p_n, p_m) = \exp\left(\frac{-(p_n - p_m)^2}{2\sigma^2}\right), \tag{5.15}$$

where σ is the kernel bandwidth to be determined.

5.3.3.3 Discounting

Once the mass functions are computed, a contextual discounting operation is handled to take into consideration the reliability of the extracted parameters according to each state, namely, ARDS and non-ARDS states. $\beta_{j,\{+1\}} = 1 - \alpha_{j,\{+1\}}$ and $\beta_{j,\{-1\}} = 1 - \alpha_{j,\{-1\}}$ are the measures of reliability of a parameter j knowing that the true state is " $+1$" and " -1" respectively. Then the discounted mass function ${}^{\alpha}m_j$ assigned to parameter j is given by

$$\begin{cases} {}^{\alpha}m_j(\{+1\}, s, t) = \beta_{j,\{-1\}} m_j(\{+1\}, s, t), \\ {}^{\alpha}m_j(\{-1\}, s, t) = \beta_{j,\{+1\}} m_j(\{-1\}, s, t), \\ {}^{\alpha}m_j(\Omega, s, t) = m_j(\Omega, s, t) \\ \qquad\qquad + \alpha_{j,\{-1\}} m_j(\{+1\}, s, t) \\ \qquad\qquad + \alpha_{j,\{+1\}} m_j(\{-1\}, s, t), \\ {}^{\alpha}m_j(\emptyset, s, t) = 0. \end{cases} \tag{5.16}$$

The discounting factors $\alpha_{j,\cdot}$ are obtained from the performance of the parameter j to distinguish between ARDS and non-ARDS subjects. Thus, mass functions are computed for each parameter j, and then decisions are generated for each subject s. Then, the performance indexes such as sensitivity and specificity are calculated for each parameter j. The sensitivity Se_j is the proportion of ARDS subjects that are correctly identified and the specificity Sp is the proportion of non-ARDS subjects that are correctly identified according to the parameter j. Thus, $\alpha_{j,\{+1\}} = 1 - Se_j$ and $\alpha_{j,\{-1\}} = 1 - Sp_j$.

5.3.3.4 Combination

The discounted mass functions are combined using the normalized conjunctive rule $\oplus$, defined by

$$m_{\oplus}(\omega, s, t) = \frac{\sum\limits_{\cap\omega'^{(j)}=\omega} \prod\limits_{j\in J} {}^{(\alpha)}m_j(\omega'^{(j)}, s, t)}{1 - \sum\limits_{\cap\omega'^{(j)}=\emptyset} \prod\limits_{j\in J} {}^{(\alpha)}m_j(\omega'^{(j)}, s, t)}. \tag{5.17}$$

5.3.3.5 Decision making

The mass functions are transformed to probability measures over the singletons $\omega = \{+1\}$ or $\{-1\}$ in order to make a decision on the health state of a subject s at time t. This transformation is done using the pignistic probability. The state ω showing the highest pignistic probability is selected for the subject s at time t as follows

$$\omega = \arg\max_{\omega'} \sum_{\omega'\in 2^{\Omega}, \omega\subseteq\omega'} \frac{m_{\oplus}(\omega', s, t)}{\mid \omega' \mid}, \forall\omega \in \Omega. \tag{5.18}$$

The algorithm is performed in real-time whenever a new data point is added to the recording. Thus, one can choose to have a very sensitive model that generates an alarm at the first positive decision or to be more sure about the alert generated and wait for a succession of positive decisions. Therefore, we also propose in this chapter to optimize the threshold of successive positive decisions. Thus, the algorithm is applied on the parameters extracted from the subjects of the training set. Then, we sweep the value of the threshold in the range of [1, 30] hours. For each value of the threshold, final decisions are generated for each subject s of the training set and the error rate of the classification is computed. The threshold of successive ARDS decisions is then obtained by minimizing the error rate over all the thresholds considered.

5.3.4 Selection of parameters

As mentioned previously, we extended the parameters extraction phase to include more parameters from the recurrence quantification analysis. We obtained in this case 64 parameters for each subject s. These parameters represent the input to the belief functions model having the KRR algorithm

as mass functions calculator. Thus, mass functions are assigned for each type of parameter p_j. Then, a discounting operation and a fusion of these mass functions using the normalized conjunctive fusion function are performed. Even though the extraction of higher number of parameters leads to a better representation of the data, it increases the complexity of the computations and may leads, in some cases, to discrepancy of information. Therefore, we perform a selection procedure in order to select the most pertinent parameters for the evidence model and reduce thereby the dimensionality of the input vector and the computations complexity.

In the previous chapter, we performed a ranking and a selection of the parameters according to the error rates computed from the fitted distributions of the groups since these distributions are used to define the mass functions. Hence, in this chapter the mass functions for each parameter are constructed via a learning algorithm. Thus, we propose to select the most pertinent parameters in a different way. The selection procedure is performed using a modified version of the sequential forward selection procedure. The training dataset is considered to do this. First, we start with the fusion of all the sub-combinations of two parameters and we compute the error rates of each of these fusions. The fusion having the lowest error rate is selected as a start point for the selection procedure. Then, we test the selected parameters with the remaining parameters at a time and we select the new vector of parameters that represent the lowest error rate. This procedure is repeated until the new parameters vector shows a decrease in the error rate.

5.4 Results

This section presents the results of the prediction model with the newly proposed mass assignment method that is based on kernel ridge regression. In addition, the extracted parameters from the recurrence quantification analysis are considered in the new model. The validation of this model is performed on the non-invasive ventilated subjects using a 5 fold cross validation repeated 10 times. The results are obtained using a fixed length window $\tau = 24$ hours and mass functions are constructed using the vector label configuration. First, a statistical test is performed to identify from the newly extracted parameters the ones that show a significant difference between ARDS and non-ARDS groups. Then, an illustration of each step of the method is

TABLE 5.1: Statistical analysis for RQA parameters

Signal	Parameters
HR	R_eR_a, L_{Max}, ENT, LAM, V_{Max}
RR	R_eR_a, DET, L_{Max}, ENT, LAM, TT, V_{Max}, LL
SpO$_2$	R_eR_a, L_{Max}, LAM, TT, V_{Max}, LL, $RPDE$
MABP	R_eR_a, L_{Max}, LAM, TT, V_{Max}, LL, $RPDE$

presented considering an example of a train / test set. Finally, the global performance is illustrated using the 5-fold cross validation.

5.4.1 Statistical analysis

The newly extracted parameters from RQA are tested statistically using the two-sample Kolmogorov-Smirnov test to identify parameters that show a statistical significant difference between ARDS and non-ARDS groups. The results in Table 5.1 show the significant parameters for each type of signals having a p-value < 0.02. Different RQA parameters have shown significant difference between both groups.

5.4.2 Optimization of the KRR parameters

As described previously, we construct the mass functions using the kernel ridge regression according to each parameter of type j. The KRR technique relies on the choice of the kernel bandwidth σ and the regularization parameter η. The values of σ and η may affect the performance of the kernel ridge regression algorithm since they control the noise tolerance and the smoothness degree. We propose to optimize their values by minimizing the error rate of the KRR-evidence algorithm on the subjects of the training set according to each parameter. The optimization is performed using a grid search on $\eta = 2^r$ with $r \in \{-10, -9, \ldots, 0\}$ and $\sigma = [10^{r'}, -1 \times r']$ with $r' \in \{-5, -4, \ldots, -1\}$. The couple of values (η, σ) that minimizes the error rate of the model is selected. The error rate is defined by the proportion of subjects that are mistakenly identified.

Table 5.2 presents an example of the lowest error rates obtained in the optimization procedure on the training set and their corresponding parameters.

TABLE 5.2: Parameters having lowest error rates after optimization of η and σ

Rank	Parameter	Error rate
1	$SampEn\_RR$	0.2
2	σ_RR	0.21
3	TT_RR	0.21
4	$R_eR_a\_HR$	0.22
5	LAM_RR	0.22
6	V_{Max}_RR	0.22
7	L_{Max}_RR	0.24
8	Sk_RR	0.24
9	$R_eR_a\_RR$	0.24
10	DET_RR	0.24
11	$L_{Max}\_SpO_2$	0.24
12	LAM_SpO_2	0.25

It is shown that low errors are obtained for the belief functions applied on individual parameters. It is worth noting that the lowest errors are obtained for parameters extracted from heart rate and respiratory rate signals. Moreover, the newly extracted parameters present the majority of the parameters giving low errors.

5.4.3 Selection of parameters

Since the dimensionality of the input vector is very high, it is necessary to perform a selection procedure to reduce the dimensionality of the input vector and thereby reduce the complexity of computations. We could perform a similar selection to that described in the previous chapter, thus considering the ranked parameters. Starting from the top ranked parameter, then adding one parameter at a time and analyzing the performance of the combined parameters. Since we could obtain ranked parameters from just one signal as shown in the previous paragraph, we propose in this chapter a more global selection procedure.

We start by performing a global search to find the best combination of two parameters from the whole set of parameters, giving the minimum error rate of the belief functions fusion. The best combination is considered as a starting set. Then, we test the remaining parameters with the selected set and find the best combination giving the lowest error rate. This is repeated until an increase in the lowest error rate occurs. Figure 5.2 represents an illustrative example of the selection procedure for a training dataset. Starting

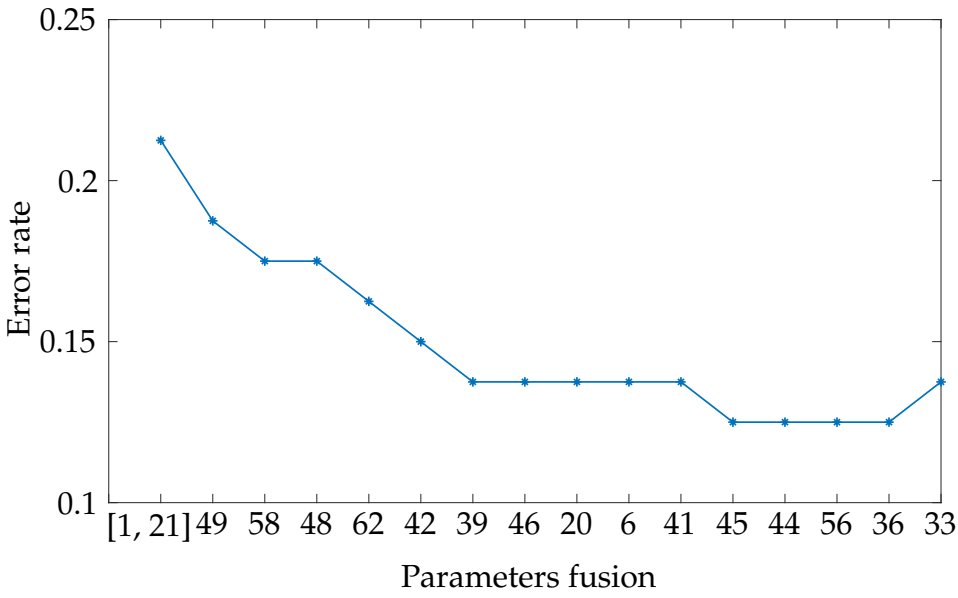


FIGURE 5.2: An example of the selection procedure of parameters from the training set

from the first set of parameters $[j = 1, j = 21]$, a parameter is added at a time and the error rate is computed. The procedure is repeated until parameter number $j = 33$ is added giving a set of 17 parameters, where the error rate increased. As shown, the number of selected parameters is equal to 16 parameters for this selection criterion, which remains a high number of parameters. In addition, we can notice that there is some added parameters that didn't change the error rate. Therefore, we propose for the following computations to change the previous criterion to a sharper one. The selection procedure ends if there is no decrease in the error rate.

5.4.4 Selection of the alert generation threshold

Whenever we test the model to choose either the optimized parameters or the best combinations in the selection procedure, we are optimizing the alert generation threshold. It is optimized on the range of $[1, \ldots, 30]$ hours. The optimal threshold gives the lowest error rate for the considered problem. We present the ROC curve of the selected parameters fusion of the previous paragraph in Figure 5.3. Thus, the optimal threshold gives the higher couple of (100-Sp, Se) in the ROC curve. In this example, it is given for threshold = 19 hours. The obtained performances on the considered training set are 85% of sensitivity and 80% of specificity.

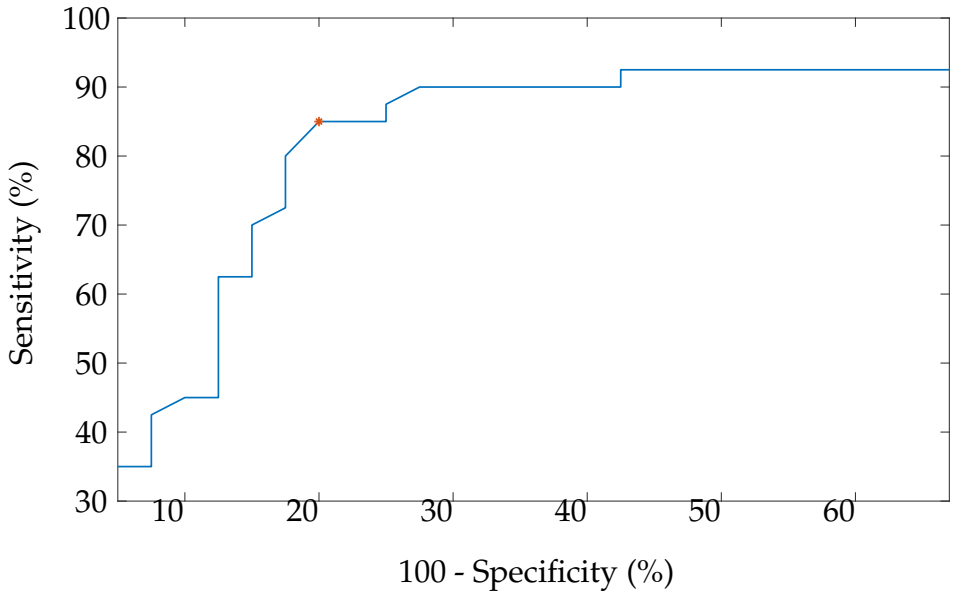


FIGURE 5.3: ROC curve for the selection of the threshold for alert generating

5.4.5 Performance evaluation

Once all the optimizations are performed on the training set, the model is tested on the test set with the optimized parameters. The results of the proposed model are presented in Table 5.3. This table presents a comparison between both configurations of labels. For the vector label, we obtained with the considered training set an accuracy of 82.5 % and on the testing set an accuracy of 85 %, with a sensitivity of 80 % and a specificity of 90%. However, the scalar label has achieved lower accuracies over both training and testing sets.

Thereafter, the overall performance of the model is obtained by the 5-fold cross validation repeated 10 times. The final performance is computed as being the average of the obtained performances over the cross validation. The overall performance is presented in Table 5.4. In this table, we also illustrate the performance of the different phases of the belief functions model. We first present the performance of the belief function model without the selection of parameters, nor the discounting. Thus, we perform a J fusion for all the parameters in the input vector and consider the initial mass functions. Then, we perform a selection of the pertinent parameters for the fusion while using the initial mass functions. Finally, we add the phase of the contextual discounting of the mass functions using the partial error rates obtained in the optimization of the KRR parameters.

TABLE 5.3: Performance of the training / testing set of the considered example

Mass assignment using	Training			Testing		
	Se(%)	Sp (%)	Acc (%)	Se(%)	Sp (%)	Acc (%)
scalar label	77.5	80	78.75	60	80	70
vector label	85	80	82.5	80	90	85

TABLE 5.4: Results of the KRR-based belief functions model on the training and the test sets

Model	Training set		Test set	
	Se(%)	Sp(%)	Se(%)	Sp(%)
BF without discounting, nor selection	74.3	71.9	69.6	62
BF without discounting	82.5	80	79.71	64
complete BF model + criterion 1	87.75	86.85	67.4	69.2
complete BF model + criterion 2	87.75	83.5	79.2	74

Table 5.4 presents the average performance over the training sets and the test sets. It is shown that the combination of the initial masses for all the parameters leads to sensitivity and specificity indexes of, respectively, 74% and 71% over the training sets and 69%, 62% over the test sets. The selection procedure enhances the performance over both training and testing sets with (Se = 82.5%, Sp = 80%) over the training sets and (Se = 79.71%, Sp = 64%) over the test sets. Finally, the incursion of the contextual discounting into the belief model is tested using both criteria for the selection procedure. The first criterion is having an increase in the error rate, while the second one is not having a decrease in the error rate. As shown, the first criterion gives the highest performances over the training sets, however the performances over the test sets are not good enough. The second criterion gives lower performances than the first criterion over training sets with a sensitivity of 87.75 % and a specificity of 83.5 % over the training sets, however it shows the best performances over the test sets having 79.2 % of sensitivity and 74% of specificity and an overall accuracy of 76.68%.

5.4.6 Early predictions

In this section, we test the effectiveness of the proposed model in early predicting the ARDS. Figure 5.4 illustrates the prediction time of ARDS, in the y-axis, versus the recordings length of parameters, in the x-axis, expressed in hours. The red line represents the limit of prediction. All points should be placed below this line to consider them as predictions. As shown

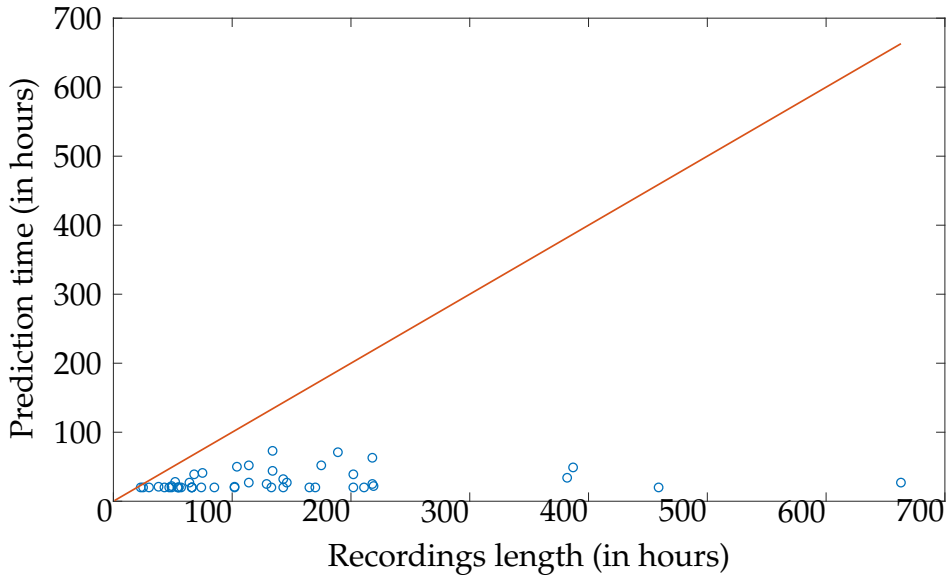


FIGURE 5.4: Early-predictions of ARDS subjects using the proposed model

in the figure, ARDS was predicted in the early segments of recordings and far away from the red line. They were predicted on average in the first 29 hours of recordings with a standard deviation of 14 hours. Moreover, the most important finding is that ARDS was predicted at least 24 hours before its occurrence for subjects having records higher than 40 hours.

5.4.7 Comparison with previous models

In this section, we provide a comparison of the latest proposed model to the models based on feature fusion proposed in the previous chapters. It is worth noting that the example considered in this chapter is the same of the previous one. The performance of the current model is better on both training and test set than that obtained from the previous model. Moreover, we present in Table 5.5 a comparison of the overall performance. The newly proposed model outperforms the model proposed in the last chapter over both training and test sets. In the newly proposed model, a new set of parameters is extracted as well as the implementation of the kernel-based technique to assign mass functions. Then, we tested the performances of the present model of belief functions theory without the new added parameters. Thus, the extracted parameters of Chapter 4 were tested along with the newly proposed model. This leads to an accuracy of 84.58% on the training set which is higher than that obtained with the previous model and 71.33% on the test

TABLE 5.5: Comparison to previous models

Model	**Training accuracy** (%)	**Test accuracy** (%)
Model of Chapter 4	78.18	74.24
This model	85.63	76.68
// + old set of parameters	84.58	71.33

set. These findings show the importance of the KRR phase for the construction of mass functions over the probability distributions, as well as the RQA parameters in the modeling of the data.

5.5 Discussion

In the present chapter, we proposed a novel incursion of the kernel methods into the belief functions model for the application of the prediction of ARDS in real-time. In addition, a new parameters extraction procedure is proposed based on the recurrence quantification analysis in order to extend the previously described parameters extraction of Chapter 4. The considered signals are the heart rate, the respiratory rate, the peripheral arterial oxygen saturation and the mean arterial blood pressure. This model is tested using the non-invasive ventilated subjects of the MIMIC II database.

The reconstructed phase space and the recurrence quantification analysis present the dynamic behavior of time series data. It has been found that the RQA parameters illustrate at best the difference between ARDS and non-ARDS groups over the signals, since they gave the lowest error rates of the belief model on the individual parameters. Thus, the RQA combined with the kernel- based belief functions demonstrates the interest to take into consideration the dynamics of the vital signs and their capability to predict ARDS in real-time.

To our knowledge, this is the first model that combines the concept of kernel methods with the belief functions theory, especially in the construction of mass functions. Thus, mass functions are assigned by a kernel-based learning phase instead of estimating the probability distributions of each parameter. As shown, this novel proposition of constructing mass functions outperformed the basic probability assignment of the previous chapter.

Moreover, it has been shown that the present model outperformed the previous one tested with the same set of parameters. Therefore, this highlights the advantage of the proposed construction of mass functions.

Moreover, the proposed model has shown high performances over both training and testing sets. It has also allowed early predictions of ARDS, where ARDS subjects were identified at least 24 hours before its onset with an average of 109 hours. The results presented in this chapter were obtained for a fixed length of window $\tau = 24$ hours. This window length can be optimized as performed in the previous chapters. Thus, we may obtain better results on both training and testing sets. An optimization phase might be considered in future work to enhance the performance of the latest proposed model.

5.6 Conclusion

This chapter presents a feature fusion model based on parameters extraction and belief functions theory. The proposed model extends that of chapter 4 from the point of parameters extraction and mass constructions. Thus, masses were constructed via a learning phase based on the kernel ridge regression. The proposed model has shown the best performances over all the previously proposed models, which are the novelty detection and decision fusion model as well as the belief functions using probability distributions model. Future work will handle further improvements of this method, such as to optimize the window length and generalize the mass construction phase to handle the multi-class problem. In addition, it might be interesting to test the other existing kernel-based techniques with this prediction model.

Chapter 6

Conclusion and Perspectives

Contents

6.1 Summary of Contributions 134

6.2 Perspectives 136

This manuscript has addressed an important health condition in the elderly population, that is, the acute respiratory distress syndrome. As mentioned in Chapter 1, the ARDS is a serious pulmonary complication that disturbs the health state of patients and may lead to death. Despite the intensive research on understanding ARDS, there remains a lack in the research for predicting this elusive syndrome. This thesis proposes different models for the prediction of this syndrome in real-time based on different processing techniques. It has multiple contributions in manipulating ARDS, as well as in signal processing. In addition, there exist different undiscovered areas for this problem that may be the subject of future studies. In this chapter, we summarize the proposed models presented in the manuscript, and the added contributions of each model. Then, we propose some perspectives and future work.

6.1 Summary of Contributions

In this section, we briefly remind each of the proposed models and highlight their performances. We perform a comparison between the proposed models in terms of concept of application and performances.

The main contribution of this thesis, despite the model being proposed, is the application itself. Indeed, ARDS is not yet well studied in the detection of its characteristics, nor the prediction of its occurrence since all the existing studies on ARDS are conducted on the physiology of this complication and its incidence and outcomes. This thesis aims to predict the occurrence of ARDS in real-time using only vital signs, thus without referring to generally used clinical data. The choice of the considered physiological signals is based on the physiological changes that were associated with ARDS in the literature and on the main objective of considering signals that can be collected at home using non-invasive sensors.

In order to predict the occurrence of ARDS, two basic models based on information fusion were proposed in this thesis. The first one is based on novelty detection and decision fusion techniques; whereas the second one is based on feature fusion techniques using the belief functions theory. In addition, a kernel based regression model is proposed jointly with the belief functions theory as a novel approach to construct mass functions. Chapter 2 has presented a brief description on the different steps of each type of

information fusion techniques, as well as an introduction to kernel methods.

In Chapter 3, we developed a real-time prediction model for ARDS using distance-based novelty detection and decision fusion. The distance-based novelty detection allows the detection in real-time of the deviations from a predefined stable state. It is performed on each signal separately in order to obtain individual decisions. Then, these decisions were combined using one of the proposed linear/nonlinear decision fusion techniques. Hence, linear decision fusion techniques were based on performances and on error rates, and the nonlinear fusion was based on the kernel ridge regression. The model was tested on both invasive and non-invasive ventilated subjects. ARDS was detected in the early stages of its development by considering the combination of the four signals using the kernel ridge regression on both types of ventilated subjects.

Then, Chapter 4 proposes a feature fusion model based on the belief functions theory. Thus, several linear and nonlinear parameters were extracted from the signals and the most pertinent parameters were selected using a forward sequential feature selection. Then, the information provided from the parameters is modeled by a mass function using the evidence framework. In this chapter, the mass functions are constructed using an estimation of the probability distributions of each group. After that, the mass functions were discounted and combined to obtain a final mass function that will represent the state of a given subject. This model was tested on two different approaches using the non-invasive ventilated subjects. The first one consists in distinguishing between two groups, namely, ARDS and non-ARDS. The ARDS group includes subjects developing severe and moderate ARDS, while the non-ARDS group includes subjects developers of mild ARDS or non developers of ARDS. However, the second approach distinguishes between each of the groups of ARDS, thus it is a four-class classification approach. The performance of the first approach has been enhanced from that of Chapter 3. The second approach has been biased to one of two options, that are being sensitive on each level of severity or being sensitive on severe ARDS and non-ARDS states.

Finally, Chapter 5 expands the previous model and proposes a novel method to construct the mass functions of each parameter. Mass functions

TABLE 6.1: A comparison between the performances of the proposed models

Model	Training accuracy (%)	Test accuracy (%)
Chapter 3	72.48	67.78
Chapter 4	78.18	74.25
Chapter 5	**85.63**	**76.68**

are constructed in this chapter via a learning phase using the kernel ridge regression. In addition, we proposed the extraction of a new set of parameters from the recurrence quantification analysis. This model was proposed on the binary classification problem. This model has increased the performances of the early-identification of ARDS. Table 6.1 illustrates a summary of the obtained accuracies for each proposed model on the training and testing sets. As shown, the accuracy has enhanced with each proposed model to finally achieve accuracy of 86% and 77% over, respectively, the training and testing sets for the last proposed model. In addition, all the proposed models have achieved early generations of ARDS alerts with at least 24 hours before its occurrence.

6.2 Perspectives

This thesis presented several important contributions for the prediction of ARDS using physiological signals. As part of future work, we would like to investigate the following areas concerning the improvements of our proposed methods.

- *Modeling of ARDS state*
  In the case of the binary classification, the total length of signals for non-ARDS subjects were considered, while signals of ARDS subjects were cut at the actual onset of ARDS, i.e., on the time where PaO_2/FiO_2 levels were lower than 200 mmHg. Then, the modeling of the ARDS state was done using the last segments of signals just before the ARDS occurrence. As shown in Chapter 4, a subject might have a mild ARDS form before ARDS occurrence but also it might have a stable state. Therefore, it is important to identify the time intervals where ARDS is developed and model the ARDS state using these segments.

- *Selection of pertinent parameters*
  In Chapter 4, we performed a sequential feature selection starting from

the parameter that gave the lower error rate between distributions. Future works will include a more global selection procedures like the one performed in Chapter 5.

- *Four-class classification*
  The problem of detecting the four groups of ARDS was only presented in Chapter 4. In addition, results were not balanced between the groups, each technique has enhanced the accuracies of groups and decreased the accuracies of others. Thus, more work must be done to efficiently perform the discounting rules. Moreover, the work will be tested on the model of belief functions using the kernel regression for mass construction. As shown in Chapter 4, a subject might develop a less severe form of ARDS before it develops the actual condition. Thus, it would be interesting to study the evolution of states to detect the degradation in signals using the Hidden Markov Models.

The already cited perspectives are related directly to the work presented in this thesis. In addition, we would like to enlarge the case study by the following propositions.

- *Extend to multi-pathology*
  The presented work considers one pathology, that is, the ARDS. Future projects will have the objective of extending the current one by performing a multi-pathology identification and surveillance. In this case, the binary classification model will turn into a multi-class classification problem or a multi-label problem since a subject might develop different pathologies at a time.

- *Deep learning*
  The human body is a very complex system where all the sub-systems are related to each other and are affected by any sudden condition to any part of it. In addition, a huge amount of data can be collected from our body to monitor its different physiological aspects. Thus, it would be interesting to exhibit deep learning methods for the prediction of pathologies.

- *Data acquisition*
  In this work, we have used the MIMIC II public database from the Physionet. It contains physiological records for patients admitted to intensive care units. We would like to perform a real-time data acquisition

for physiological signals from elderly patients at home in order to obtain a more realistic cases.

- *Implementation in home surveillance system*
  Finally, the main goal of such studies is the implementation of a predictive model for ARDS in home surveillance systems in order to alert the patient and the medical staff of a possible danger of the occurrence of ARDS in a near future so they can quickly intervene and prevent its severity or even its occurrence.

Appendix A

Phase Space and Recurrence Quantification Analysis

Contents

A.1 Reconstructed phase space 140
A.1.1 Estimation of the time delay 141
A.1.2 Choosing the embedding dimension 141
A.2 Recurrence quantification analysis 142
A.2.1 Recurrence plot 142
A.2.2 Quantitative analysis of the RP 143

In this section, a brief introduction to the reconstructed phase space and recurrence quantification analysis is presented. For more details, please refer to the following references, (Takens, 1981; Buzug and Pfister, 1992; Cellucci, Albano, and Rapp, 2003) for reconstructed phase space and (Eckmann, Kamphorst, and Ruelle, 1987; Marwan et al., 2007) for recurrence quantification analysis.

A.1 Reconstructed phase space

The phase space is an abstract mathematical space that visualizes the changes in the dynamic variables of a system. Each of these variables is represented by an axis of the phase space and each possible state of the system is represented by a point in this space. The succession of these points represents the dynamic evolution of the system and the resulting shape allows to extract dynamic properties. The reconstructed phase space is created using a time delay t_d and an embedding dimension M.
Given a time series $x(\ell)$ with $\ell = 1, \ldots, N$, N being the number of data points, the vector in reconstructed phase space at time ℓ is written as following:

$$X(\ell) = [x(\ell), x(\ell + t_d), x(\ell + 2t_d), \ldots, x(\ell + (M-1)t_d)]. \tag{A.1}$$

Takens theorem specifies that a correct choice of these variables (t_d, M) and in case of non-noisy observations will result in a good representation of real data and integrated reconstructed phase space trajectories will have the same properties of real attractors (Takens, 1981). The choice of t_d and M are very critical (Buzug and Pfister, 1992; Cellucci, Albano, and Rapp, 2003). An overestimation of the embedding dimension M will complicate in a useless way the exploitation of the trajectory while an underestimate will create a folded trajectory. For the choice of the time delay, a very small value of t_d will result in high correlations between points of different axes and the really important correlations will be masked, while too large t_d value will give randomly distributed points in the reconstructed phase space. The time delay is frequently chosen by minimizing the mutual information (Fraser and Swinney, 1986; Kraskov, Stögbauer, and Grassberger, 2004) while the algorithm of the false nearest neighbors (Kennel, Brown, and Abarbanel, 1992) estimates iteratively the sufficient embedding dimension.

A.1.1 Estimation of the time delay

The best time delay is determined using the average mutual information (AMI) (Fraser and Swinney, 1986). Given two time series U and V, the AMI between these two time series is defined by (Abarbanel, 2012):

$$I_{UV}(u_i, v_i) = \sum_{u_i, v_i} P_{UV}(u_i, v_i) \log_2 \left[\frac{P_{UV}(u_i, v_i)}{P_U(u_i) P_V(v_i)} \right] \tag{A.2}$$

where $P_U(u_i)$, $P_V(v_i)$ are the probabilities of occurrence of u_i and v_i in sets U and V respectively. $P_{UV}(u_i, v_i)$ is the probability of co-occurrence of u_i in set U and v_i in set V. When replacing U by $x(1), x(2), \ldots, x(\ell)$ and V by $x(1+t_d), x(2+t_d), \ldots, x(\ell+t_d)$, the AMI of estimating $x(t+t_d)$ from $x(t)$ becomes:

$$I(t_d) = \sum_{\ell, \ell+t_d} P(x(\ell), x(\ell+t_d)) \log_2 \left[\frac{P(x(\ell), x(\ell+t_d))}{P(x(\ell)) P(x(\ell+t_d))} \right] \tag{A.3}$$

The suitable value of t_d can be then determined being the first local minimum of $I(t_d)$ (Fraser and Swinney, 1986).

A.1.2 Choosing the embedding dimension

The embedding dimension M is estimated using the false nearest neighbor (FNN) (Kennel, Brown, and Abarbanel, 1992). When the dimension is insufficient to unfold the trajectory, false nearest neighbors appear. The number of FNN and trajectory self-crossings decrease with increasing the dimension. Let $x^n(\ell)$ be the neighbor of the point $x(\ell)$ in the dimension M, the square of the Euclidean distance between these two points is:

$$R_M^2(t) = \sum_{m=0}^{M-1} [x(\ell + m t_d) - x^n(\ell + m t_d)]^2. \tag{A.4}$$

When passing to a higher dimension $M+1$, the distance becomes:

$$R_{M+1}^2(\ell) = R_M^2(\ell) + [x(\ell + M t_d) - x^n(\ell + M t_d)]^2 \tag{A.5}$$

$x^n(\ell)$ is defined as false neighbor of the point $x(\ell)$, if the distance between $x^n(\ell)$ and $x(\ell)$ increases when passing from dimension M to dimension $M+$

1. Two criteria are used to verify FNN:

$$R_a(\ell) = \frac{|x(\ell + Mt_d) - x^N(\ell + Mt_d)|}{R_M(\ell)} \geq R_{th} \tag{A.6}$$

$$\frac{|x(\ell + Mt_d) - x^n(\ell + Mt_d)|}{R_A} \geq A_T \tag{A.7}$$

R_{th} is a threshold to be determined, A_T is around 2 with R_A is the standard deviation of x (Abarbanel, 2012).

A.2 Recurrence quantification analysis

A.2.1 Recurrence plot

The recurrence of states means that states are arbitrary near to each other after some time (Monk and Compton, 1939; Argyris, Haase, and Faust, 1994; Ott, 2002). It is visualized using the recurrence plot (RP) as presented in (Eckmann, Kamphorst, and Ruelle, 1987). The recurrence plot projects the M-dimensional phase space into a 2-dimensional sub-space of its recurrences. It is employed because the phase space cannot be plotted when M is higher than three. It shows the times when the trajectory of the phase space passes approximately in the same area. The RP consists of a square matrix with elements corresponding to the times at which a state at time ℓ recurs at a different time ℓ'. The matrix at position (ℓ, ℓ') is equal to one if the state is repeated and zero otherwise. This is marked as black and white dots in a plot, this plot having thus two time axes. The matrix of RP can be then computed using:

$$R_{\ell,\ell'} = \Theta\left(\varepsilon - \|X(\ell) - X(\ell')\|\right); i, j = 1, \ldots, N \tag{A.8}$$

with N being the number of states, ε is a threshold distance equal to 0.2, $\|\cdot\|$ is a norm such as the Euclidean norm and $\Theta(\cdot)$ is the Heaviside function computed as:

$$\Theta(\alpha) = \begin{cases} 0, & \text{if } \alpha < 0, \\ 1, & \text{if } \alpha \geq 0. \end{cases} \tag{A.9}$$

Figure A.1 presents a typical example of a reconstructed phase space and its recurrence plot. The RP displays characteristics caused by the dynamical behavior of the system (Eckmann, Kamphorst, and Ruelle, 1987; Marwan et al., 2007), such as:

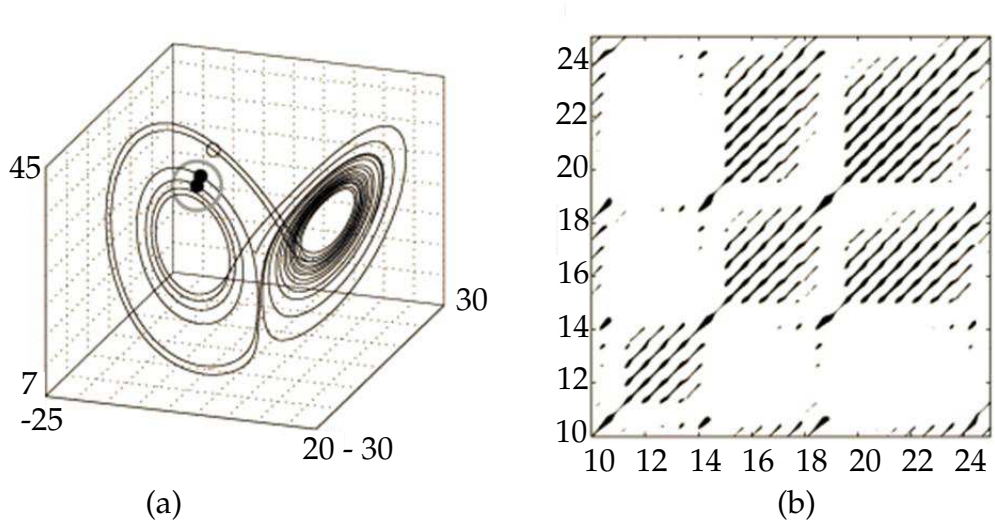


FIGURE A.1: An example of a reconstructed phase space (a) and its recurrence plot (b) (Barbera, La Rocca, and Rizzi, 2011).

- Diagonal line: corresponds to periods where segments of phase space trajectory run parallel, the length of this diagonal line is determined by the duration of such similar local evolution of this trajectory segment.
- Vertical line: designates a time length in which a state does not change or changes slowly, it is an indication of laminar states.

A.2.2 Quantitative analysis of the RP

The recurrence quantification analysis (RQA) is a method to quantify the number and duration of state recurrences of a trajectory in a phase space. The main advantage of RQA is that it provides useful information, even for short and nonstationary data. Different parameters of RQA have been proposed in the literature to characterize diagonal and vertical structures (Gao and Cai, 2000; Marwan et al., 2007). These parameters are described in Table A.1.

TABLE A.1: RQA parameters (Marwan et al., 2007)

Parameter name	Definition
Recurrence Rate R_eR_a	The percentage of recurrence points in an RP, it corresponds to the correlation sum: $$R_eR_a = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}$$
Determinism DET	The percentage of recurrence points which form diagonal lines: $$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}}$$ $P(l)$ is the histogram of the lengths l of the diagonal lines.
Laminarity LAM	The percentage of recurrence points which form vertical lines: $$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}$$ $P(v)$ is the histogram of the lengths v of the vertical lines
Average diagonal line length $\langle L \rangle$	The average length of the diagonal lines: $$L = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}$$

To be continued

Table A.1 (continued)

Parameter name	Definition
Trapping Time TT	The average length of the vertical lines: $$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)}$$
Longest diagonal line L_{max}	The length of the longest diagonal line: $$L_{max} = \max(\{l_i; i = 1, \ldots, N_l\})$$ N_l: number of diagonal lines in the RP
Longest vertical line V_{max}	The length of the longest vertical line: $$V_{max} = \max(\{v_i; i = 1, \ldots, N_v\})$$ N_v: number of vertical lines in the RP
Entropy ENTR	The Shannon entropy of the probability distribution of the diagonal line lengths $p(l)$: $$ENTR = -\sum_{l=l_{min}}^{N} p(l) \ln p(l)$$
Recurrence period density entropy RPDE	$$RPDE = -(\ln L_{max})^{-1} \sum_{l=l_{min}^{N}} p(l) \ln p(l)$$

Appendix B

Résumé en Francais

Contents

B.1 Le syndrome de détresse respiratoire aigue 149
B.1.1 Introduction 149
B.1.2 Le système respiratoire 149
B.1.3 Le syndrome de détresse respiratoire aigüe 150
B.1.4 Objectif 152
B.1.5 La base de données MIMIC II 152
B.1.6 Contributions 153
B.2 La prévention du SDRA par la détection d'anomalies . . . 154
B.2.1 Introduction 154
B.2.2 Approche proposée 155
B.2.3 Résultats 160
B.2.4 Conclusion 162
B.3 La prévention du SDRA par les fonctions de croyance . . . 162
B.3.1 Introduction 162
B.3.2 Prévention binaire 163
B.3.3 Classification multi-classe 166
B.3.4 Identification des groupes 166
B.3.5 Modèle des fonctions de croyance 167
B.3.6 Résultats 168
B.3.7 Conclusion 171
B.4 La prévention du SDRA par les croyances et les noyaux . . 171
B.4.1 Introduction 171
B.4.2 Approche proposée 172
B.4.3 Résultats 177

B.4.4 Conclusion 179
B.5 Conclusion et perspectives 179
B.5.1 Contributions principales 180
B.5.2 Perspectives 181

B.1 Le syndrome de détresse respiratoire aigue

B.1.1 Introduction

La population âgée de 60 ans ou plus est prévue de se multiplier plus rapidement que les autres gammes d'âges (Prince et al., 2007; Organization and International Tobacco Control, 2008; Mathers et al., 2015). Ce phénomène, connu sous le nom de vieillissement de la population, se répand dans le monde entier. L'Europe sera la région la plus affectée par cette transformation sociale, ayant le pourcentage de population âgée le plus élevé (United Nations, Department of Economic and Social Affairs, Population Division, 2017). Le vieillissement de la population est accompagné d'un profond impact sur la structure socio-économique de la société, notamment dans les domaines de santé et de sécurité sociale (Institute for Health Metrics and Evaluation (IHME), 2018; Patton, 2015). Cette tendance de vieillissement a augmenté la fréquence des maladies et la présence de multiples pathologies ensemble entrainant la dépendance des personnes âgées et par la suite l'exigence des soins continus. Il existe d'innombrables pathologies confrontant les personnes âgées. On cite entre autres les maladies cardio-vasculaires, le cancer et les maladies respiratoires (Prince et al., 2015). L'occurrence de ces dernières augmente chez la population âgée avec la croissance des niveaux de pollution, d'abus d'alcool et d'autres infections.

B.1.2 Le système respiratoire

Le système respiratoire est responsable de transmettre l'apport essentiel d'air atmosphérique et de l'échange de gaz entre le corps et l'air transmis. Durant la respiration, le corps retire l'oxygène de l'atmosphère pour la croissance cellulaire et le métabolisme, et rejette le dioxyde de carbone qui se forme comme un déchet du métabolisme cellulaire. Le système respiratoire comprend le nez, la trachée, les bronches, les bronchioles et les alvéoles, comme indiqué dans la Figure 1.2. L'air entre dans le corps à travers la bouche et le nez et arrive dans les poumons par la trachée qui a la forme d'un tube de 2.6 cm de diamètre. Elle se divise en deux tubes plus étroits, appelés bronches, qui pénètrent chacun dans l'un des poumons. Ces deux bronches primaires se subdivisent en des lobes secondaires et tertiaires à l'intérieur des poumons. Les bronches tertiaires se ramifient en de nombreux tubes plus minces, appelés bronchioles, qui sont divisés eux-mêmes en des plus petites

bronchioles jusqu'à l'arrivée aux alvéoles où se déroule l'échange d'oxygène et de dioxyde de carbone.

B.1.3 Le syndrome de détresse respiratoire aigüe

Le Syndrome de Détresse Respiratoire Aigüe (SDRA) est une forme très sévère de défaillance intense du système respiratoire suite à une altération de la perméabilité capillaire. Le terme "syndrome de détresse respiratoire" a été introduit par (Ashbaugh et al., 1967) pour décrire le développement aigu d'une tachypnée, d'hypoxémie et de perte de compliance pulmonaire. La définition du SDRA a largement évolué pour arriver à la définition de Berlin en 2012 (Ranieri et al., 2012), qui avait pour but d'aborder les limitations des définitions précédentes (Ashbaugh et al., 1967; Bernard et al., 1994). La définition de Berlin a présenté trois sous-groupes de SDRA, léger, modéré et sévère, selon le niveau d'oxygénation.

Le SDRA possède un taux d'incidence remarquable, il est mesuré sur 100 000 habitants par an. L'incidence est estimée à 34 patients aux Etats-Unis (Rubenfeld et al., 2005), environ 5-7 en Europe (Sigurdsson et al., 2013) et 15,19 en Taiwan (Chen et al., 2015). Les taux du SDRA modéré ou sévère représentent 75% des patients atteints du SDRA, les 25% restants développent un SDRA léger dont un tiers progressera par la suite en un stade modéré ou sévère (Rubenfeld et al., 2005; Ranieri et al., 2012). Le SDRA est une affection parfois mortelle possédant un taux de mortalité qui varie entre 40 et 60% dans la plupart des patients (Singh et al., 2014; Chen et al., 2015; Bellani et al., 2016). En outre, il augmente avec l'âge (Arroliga et al., 2002; Goss et al., 2003; Hughes et al., 2003; Chen et al., 2015). Par ailleurs, peu de personnes survivant au SDRA s'en remettent complètement, la majorité aura des séquelles pulmonaires à vie.

Le diagnostic du SDRA est cliniquement difficile car il n'y a pas d'indicateurs cliniques précis pour cette maladie. Diverses modalités sont étudiées pour diagnostiquer et surveiller les conditions cliniques. On trouve principalement les paramètres ventilatoires (Bernard et al., 1994; Ranieri et al., 2012; Mac Sweeney and McAuley, 2016), les techniques d'imagerie comme la radiographie thoracique, l'échographie pulmonaire (Copetti, Soldati, and Copetti, 2008) et la tomographie par émission de positons (PET) (Bellani et al., 2011), et finalement le rapport PaO_2/FiO_2 entre la

pression partielle d'oxygène (PaO_2) et la fraction d'oxygène inspiré (FiO_2). Le SDRA peut être stimulé par divers facteurs : le sepsis, la pneumonie, les traumatismes, le pancréatite, l'abus d'alcool, le diabète, l'obésité, etc (Ware and Matthay, 2000; Gong et al., 2010; Boverman and Genc, 2014).

En outre, il existe différentes études menées pour analyser la réponse des patients aux manœuvres de recrutement alvéolaire qui permettent de rouvrir les alvéoles mal aérées ou non aérées (Nielsen et al., 2005; Borghi-Silva et al., 2016). Cependant, ces techniques sont appliquées après le diagnostic du SDRA et surtout dans sa forme sévère où un nombre élevé d'alvéoles est infecté et s'est effondré. D'un autre côté, un certain nombre de modèles de simulation a été développé pour capturer la mécanique pulmonaire chez les patients développant le SDRA (Bates, 2009; Madorno and O. Rodriguez, 2010; Van Drunen et al., 2013; Langdon et al., 2016). Le SDRA était associé à des changements cardiovasculaires et hémodynamiques. Il commence par la destruction des barrières des membranes alvéolaires résultant de l'augmentation de la perméabilité alvéolaire. Ensuite, les lésions progressent entraînant une diminution de la compliance pulmonaire qui conduit à une hypertension artérielle et à une hypoxie (Mac Sweeney and McAuley, 2016; Patroniti, Isgrò, and Zanella, 2011). De plus, le SDRA a été associé à des signes cliniques tels qu'une dyspnée grave, une tachypnée et une hypoxémie selon (Gajic et al., 2011). D'autres études sur le profil clinique du SDRA ont montré qu'il est associé à l'hypotension, la tachypnée (Chaudhury et al., 2017) et la tachycardie (Park et al., 2016).

En outre des recherches menées sur la caractérisation du SDRA, Il est important de prédire son occurrence en raison de sa sévérité et de ses profondes répercussions économiques. Ennett *et al* ont développé un algorithme pour la détection du SDRA en utilisant des variables cliniques de la base de données cliniques MIMIC II (Goldberger et al., 2000). Ils ont établi un ensemble de règles conjonctives simples et combinées en utilisant des variables cliniques comme la pression de plateau et la pression moyenne (Ennett et al., 2008). D'autre part, différentes études ont été induites sur la prédiction de la mortalité liée au SDRA (Navarrete-Navarro et al., 2000), ou sur la détection des complications et insuffisances respiratoires (Madorno and O. Rodriguez, 2010; Ravishankar et al., 2014; Velardo et al., 2014).

B.1.4 Objectif

L'objectif principal de cette thèse est de développer un modèle de surveillance continu de l'état de santé des personnes âgées en utilisant uniquement des signaux physiologiques afin de prédire si les patients développeront un SDRA dans un futur proche. Il est évident que le SDRA est lié à des changements cardiorespiratoires et cardiovasculaires telles que la tachycardie, la dyspnée, l'hypoxie et les fluctuations de la tension artérielle. Par conséquent, nous considérons dans cette thèse quatre signaux vitaux tels que le rythme cardiaque, le rythme respiratoire, la saturation d'oxygène dans le sang et la pression artérielle moyenne. Ces signaux peuvent être collectés en temps réel à l'aide des capteurs portables non invasifs et discrets. Nous proposons dans la suite des modèles de traitement et d'analyse de données pour prédire l'occurrence du SDRA. La base de données MIMIC II est considérée afin de valider ces modèles. Il s'agit d'une base de données en ligne accessible au public.

B.1.5 La base de données MIMIC II

MIMIC-II (Multi Parameter Intelligent Monitoring of Intensive Care) est une grande base de données contenant des enregistrements de soins intensifs (Goldberger et al., 2000; Saeed et al., 2011). La base de données MIMIC-II est composée de deux jeux de données, l'un contenant des signaux physiologiques et l'autre contenant les données cliniques. Les données physiologiques consistent en des informations physiologiques mesurées par les moniteurs de chevet dans l'unité de soin intensif à des fréquences bien précises. Ces données telles que le rythme cardiaque, la pression artérielle ou autres ont été continuellement validées par les infirmiers de l'unité. Le jeu de données cliniques contient des enregistrements cliniques concernant les cellules sanguines, les valeurs de laboratoire, les notes des infirmiers, les médicaments et d'autres variables cliniques. Ces variables sont enregistrées beaucoup moins fréquemment que les signaux physiologiques. Une correspondance entre ces deux jeux de données est faite sur une partie de données.

Dans cette thèse, nous avons considéré deux groupes de sujets, "SDRA" et "non-SDRA". Le groupe SDRA inclut les sujets qui développent le SDRA modéré et sévère, tandis que le groupe non-SDRA regroupe les sujets ayant le SDRA léger et les sujets n'ayant pas le SDRA. Le groupe SDRA est alors identifié lorsque le rapport PaO_2/FiO_2 diminue et devient inférieur à 200

mm Hg. Afin d'identifier les deux groupes de sujets, la procédure de sélection des sujets commence à partir de la base de données cliniques pour obtenir les sujets ventilés présentant des mesures du rapport PaO_2/FiO_2. Parmi ces sujets, ceux commençant par un $PaO_2/FiO_2 > 200$ mm Hg puis descendant en-dessous de ce seuil pour une durée minimale de 12 heures définissent le groupe SDRA; tandis que le groupe non-SDRA comprend des sujets ayant des rapports $PaO_2/FiO_2 > 200$ mm Hg pendant toute la durée d'enregistrement. Ensuite, les sujets sélectionnés sont identifiés dans la base de données physiologiques afin d'extraire leurs signaux physiologiques. A partir de cette base, nous extrayons quatre séries temporelles : le rythme cardiaque (HR), le rythme respiratoire (RR), la saturation d'oxygène dans le sang (SpO_2) et la pression artérielle moyenne (MABP). Ces signaux vitaux ont une fréquence d'échantillonnage d'un échantillon/minute.

Les sujets identifiés sont soumis à l'une ou l'autre des deux conditions de ventilation suivante, la première invasive et la deuxième non-invasive. Les procédures de sélection sont identiques pour les 2 types de sujets, elles sont présentées dans les Figures 1.8 et 1.9. Nous obtenons au final 26 sujets SDRA et 14 sujets non-SDRA pour la ventilation invasive et 50 sujets SDRA et 50 sujets non-SDRA pour la ventilation non-invasive.

B.1.6 Contributions

Dans le cadre de cette thèse, nous nous intéressons au problème de prévention du SDRA à l'aide de signaux physiologiques. Ce problème consiste en un problème de fusion de données, dont le but est de combiner les informations collectées pour remonter à une décision concernant l'état des personnes. La première contribution dans ce sens est le développement d'une approche de détection d'anomalies centrée sur la personne. Ayant les signaux physiologiques de la personne, nous définissons un premier segment de stabilité de la personne, puis nous détectons les anomalies comparées à cet état initial, dit stable. Une fusion des décisions individuelles par signal est ensuite effectuée pour améliorer les performances de l'algorithme.

Ensuite, nous proposons un modèle basé sur la fusion de caractéristiques pour la prédiction du SDRA en temps réel. Le modèle consiste d'abord à extraire des paramètres linéaires et non linéaires des signaux. Ensuite, une procédure de sélection de paramètres est effectuée pour obtenir les

paramètres les plus pertinents et réduire la dimensionnalité du vecteur d'entrée. La théorie des fonctions de croyance est utilisée pour combiner le meilleur ensemble des paramètres. Le modèle est défini à partir des distributions de probabilité des paramètres. Nous testons également le modèle proposé sur un problème multi-classes afin de différencier entre quatre groupes de sujets, les sujets à différents niveaux du SDRA sévère, modéré et léger et les sujets non-SDRA.

Finalement, nous améliorons le modèle proposé précédemment en introduisant les notions d'apprentissage à noyau dans le calcul des fonctions de croyance. En outre, une nouvelle extraction de paramètres est effectuée pour augmenter le choix du vecteur d'entrée. Les nouveaux paramètres extraits sont calculés à partir de l'analyse de quantification de récurrence. Une procédure de sélection est également définie pour réduire la dimensionnalité du vecteur de paramètres.

B.2 La prévention du SDRA par la détection d'anomalies

B.2.1 Introduction

Les techniques de fusion de données sont largement utilisées dans plusieurs domaines tels que les applications biomédicales, la robotique, les applications industrielles et environnementales (Ho, Hull, and Srihari, 1994; Kuncheva, Bezdek, and Duin, 2001; Kuncheva and Whitaker, 2003). La fusion de données provenant de sources multiples présente des avantages par rapport aux données provenant d'une source unique. L'utilisation de différentes sources garantit une meilleure observation d'un évènement donné. Dans la plupart des modèles de surveillance des applications biomédicales, l'identification des pathologies peut être effectuée en utilisant un seul signal (Klingeberg and Schilling, 2012), mais la plupart des maladies graves surviennent avec des irrégularités dans plusieurs signaux vitaux (Forkan and Khalil, 2017).

Il existe principalement trois types de fusion d'informations, la fusion de décisions, la fusion de caractéristiques et la fusion des données brutes (Dasarathy, 1994). Dans cette partie, nous nous intéressons à la fusion de

décisions. Alors, nous proposons une nouvelle approche pour la prévention précoce de l'occurrence du SDRA, basée sur l'analyse en temps réel des signaux physiologiques continus. L'approche proposée effectue une détection d'anomalies en identifiant un état initial pour chaque sujet. La déviation de l'état initial dans chaque signal sera étudiée et une décision individuelle sera générée. Les décisions individuelles obtenues sont alors combinées par des méthodes de fusion de décisions pondérée. Dans la suite, nous décrivons l'approche proposée, y compris l'algorithme de détection d'anomalies, les techniques de fusion de décisions proposées et l'optimisation des paramètres. Finalement, les résultats sont illustrés sur les deux types de sujets ventilés.

B.2.2 Approche proposée

L'objectif de cette approche est de développer un modèle d'alerte précoce pour le SDRA, incluant les niveaux modéré et sévère, en utilisant des signaux vitaux. Ainsi, nous considérons quatre types de signaux physiologiques, qui sont le rythme cardiaque (HR), le rythme respiratoire (RR), la saturation d'oxygène dans le sang (SpO_2) et la pression artérielle moyenne (MABP). Soit $x_{s,i}(\ell)$, $i = 1,\ldots,4$, la valeur du signal i pour un sujet s enregistrée à un instant ℓ, et soit t le temps courant où les calculs sont effectués. $x_{s,i}(1:t) = (x_{s,i}(1),...,x_{s,i}(t))$ représente le vecteur des données enregistrées du début jusqu'au temps t en minutes. On considère un label y_s pour le sujet s concernant son état de santé, c'est-à-dire $y_s = +1$ si le sujet développe le SDRA et $y_s = -1$ sinon.

Le modèle de prévention est illustré dans les Figures 3.1 et 3.2. Il prend en entrée les quatre signaux enregistrés $x_{s,i}(1:t)$ et donne en sortie $\hat{y}_s(t)$ avec $\hat{y}_s(t) = -1$ si le SDRA n'est pas encore prévu, et $\hat{y}_s(t) = +1$ si le SDRA est prévu. L'approche proposée est individuelle, basée sur la détection d'anomalies, permettant de détecter pour chaque sujet s'il s'écarte de son état initial. En effet, un segment initial de longueur τ de chaque signal est considéré pour définir son état stable, puis les segments restants des signaux sont analysés pour détecter s'ils divergent de ces segments initiaux. Chaque signal conduit à une décision individuelle en temps réel, notée $u_{s,i}(t)$. Une approche de fusion de données pondérée notée $\psi(\cdot)$ est ensuite proposée, pour combiner ces décisions, conduisant à une décision globale plus précise.

Dans ce qui suit, les sujets qui développent un SDRA avec $y_s = +1$ sont appelés sujets SDRA, alors que les sujets non-SDRA ont $y_s = -1$.

B.2.2.1 Détection d'anomalies

La détection d'anomalies a pour objectif de détecter la présence des anomalies, à partir d'un modèle initial défini comme normal. Elle est appliquée dans différents domaines, tels que la détection des défauts (Cheng et al., 2008), les applications médicales (Roberts, 2000; Clifton et al., 2009), le contrôle de processus (Knorr, Ng, and Tucakov, 2000; Perner, 2009) et bien d'autres.

L'approche proposée identifie pour chaque sujet un état stable en fonction d'un premier segment de ses signaux enregistrés, puis détecte en temps réel la divergence par rapport à cet état stable dans les données enregistrées suivantes. Si l'on considère les signaux vitaux d'un certain sujet s, un premier segment de chacun des signaux est d'abord identifié. Soit $\tau < t$ la durée du premier segment sélectionné. τ pourrait être égale à 6, 12, 18 ou même 24 heures, soit 360, 720, 1080 ou 1440 minutes respectivement. Une approche d'optimisation est proposée dans la suite pour définir sa valeur.

Ainsi pour le signal i d'un sujet considéré s, $x_{s,i}(1:\tau)$, appelé premier segment, est utilisé pour définir l'état initial; alors que $x_{s,i}(t-\tau+1:t)$, appelé deuxième segment, est ensuite utilisé pour l'étude de divergence. Nous proposons un algorithme de détection d'anomalies basé sur la distance qui consiste à comparer le nombre de valeurs aberrantes de ce deuxième segment à celui du segment stable. Les anomalies sont identifiées en fonction des deux définitions suivantes :

Définition 1 : Pour les données normalement distribuées, la proportion de données se situant à 3σ ou plus de leur moyenne doit être égale à 1%. Ces données sont appelées des valeurs aberrantes. Pour chaque signal i, la moyenne $\mu_{s,i}$ et l'écart type $\sigma_{s,i}$ du segment stable sont calculés. Ensuite, les valeurs aberrantes sont détectées dans le deuxième segment en comptant le nombre de points situés au-delà de $\pm 3\sigma_{s,i}$ à partir de $\mu_{s,i}$. Un sujet s est considéré SDRA selon le signal i avec $u_{s,i}(t) = +1$ si le nombre des valeurs aberrantes comptées est supérieur à 1% de la longueur de ce dernier segment; sinon $u_{s,i}(t) = -1$. Une définition plus générale du modèle basé sur la distance est proposée ci-après.

Définition 2 : Étant donné un seuil D, chaque point $x_{s,i}(\ell)$, $\ell \in [\tau+1, t]$, est défini comme anormal s'il est situé à une distance de $x_{s,i}(1:\tau)$ supérieure à D. La distance euclidienne est utilisée pour calculer cette distance comme suit :

$$f(x_{s,i}(\ell), x_{s,i}(1:\tau)) = \frac{|x_{s,i}(\ell) - \mu_{s,i}|}{\sigma_{s,i}}, \tag{B.1}$$

Où $\mu_{s,i}$ et $\sigma_{s,i}$ sont respectivement la moyenne et l'écart type du premier segment du signal i d'un sujet s. Soit $\mathbb{I}(x_{s,i}(\ell))$ l'indicateur donnant 1 si $x_{s,i}(\ell)$ est anormal et 0 sinon,

$$\mathbb{I}\left(x_{s,i}(\ell)\right) = \begin{cases} 1, & \text{si } f(x_{s,i}(\ell), x_{s,i}(1:\tau)) > D; \\ 0, & \text{sinon.} \end{cases} \tag{B.2}$$

Soit $p_{s,i}$ la proportion des anomalies contenues dans le premier segment du signal i pour le sujet s. En d'autres termes, $p_{s,i} = \frac{\sum_{\ell=1}^{\ell=\tau} \mathbb{I}(x_{s,i}(\ell))}{\tau}$. Par conséquence, les anomalies dans le deuxième segment $x_{s,i}(t-\tau+1:t)$ sont comptées et leur rapport $r_{s,i}(t)$ sur la longueur du dernier segment est calculé, $r_{s,i}(t) = \frac{\sum_{\ell=t-\tau+1}^{\ell=t} \mathbb{I}(x_{s,i}(\ell))}{\tau}$. Ensuite, $r_{s,i}(t)$ est comparé à $p_{s,i}$ et une décision partielle est générée :

$$u_{s,i}(t) = \begin{cases} +1, & \text{si } r_{s,i}(t) > p_{s,i}; \\ -1, & \text{sinon.} \end{cases} \tag{B.3}$$

B.2.2.2 Fusion des décisions individuelles

Dans cette section, nous décrivons la technique de fusion de décisions qui permet de combiner les décisions individuelles obtenues afin d'améliorer les performances de la classification. Nous définissons une fonction de fusion $\psi(\cdot)$ qui prend en entrée le vecteur des décisions individuelles $\boldsymbol{u}_s(t) = (u_{s,1}(t), \ldots, u_{s,4}(t))$ d'un sujet s et qui donne en sortie une valeur aussi proche que possible de l'état réel du sujet y_s qui est $+1$ si le sujet développe le SDRA et -1 s'il reste dans un état stable. La décision finale $\hat{y}_s(t)$ est obtenue par le signe de la réponse de $\psi(\cdot)$, comme suit:

$$\hat{y}_s(t) = sign(\psi\left(\boldsymbol{u}_s(t)\right)) = \begin{cases} +1, & \text{si } \psi\left(\boldsymbol{u}_s(t)\right) > 0; \\ -1, & \text{sinon.} \end{cases} \tag{B.4}$$

Nous proposons trois types de fusion différant par le calcul de la fonction $\psi(\cdot)$, la première basée sur les performances, la deuxième basée sur les taux d'erreur et la troisième non linéaire basée sur la régression à noyau. Ces

règles de fusion consistent essentiellement à intégrer des décisions multiples dans le but de réduire l'incertitude sur la décision résultante.

Fusion à base de performances : Dans ce type de fusion, les performances sont obtenues pour chaque signal à partir de la classification préliminaire. Un sujet SDRA est supposé d'être correctement identifié par son signal i s'il donne une décision $u_{s,i}(\ell) = +1$ à un moment ℓ du signal; alors qu'un sujet non-SDRA est correctement identifié si $u_{s,i}(\ell) = -1$ pour tous les temps ℓ. Les performances sont définies par la sensibilité Se et la spécificité Sp qui représentent les pourcentages des sujets malades correctement identifiés, et des sujets non-malades correctement identifiés respectivement. Ces performances sont ensuite utilisées pour déterminer un degré de confiance ou un poids noté a_i, $i = 1, \ldots, 4$, qui sera utilisé conjointement avec la décision partielle de chaque signal. Les poids a_i sont calculés comme suit :

$$a_i = \begin{cases} Se_i, & \text{si } u_{s,i} > 0; \\ Sp_i, & \text{sinon.} \end{cases} \tag{B.5}$$

La fonction $\psi\left(\boldsymbol{u}_s(t)\right)$ sera définie selon l'équation suivante :

$$\psi\left(\boldsymbol{u}_s(t)\right) = \sum_i a_i u_{s,i}(t). \tag{B.6}$$

Fusion à base de taux d'erreur : Dans ce paragraphe, nous proposons une nouvelle fonction de fusion basée sur les taux d'erreur. Soient $\epsilon_{i,+}$ et $\epsilon_{i,-}$ les taux de faux positifs et faux négatifs. En d'autres termes, $\epsilon_{i,+}$ est défini comme étant le pourcentage des sujets estimés à développer le SDRA avec $u_{s,i} = +1$, alors que ce n'est pas le cas; et $\epsilon_{i,-}$ est le pourcentage de sujets estimés stables alors qu'ils développeront le SDRA. Les poids a_i sont calculés à l'aide d'une fonction décroissante du taux d'erreur. Deux fonctions sont proposées dans cette thèse comme suit :

$$a_i = \begin{cases} 1 - \epsilon_{i,+}, & \text{si } u_{s,i} > 0; \\ 1 - \epsilon_{i,-}, & \text{sinon.} \end{cases} \tag{B.7}$$

$$a_i = \begin{cases} \frac{(1-\epsilon_{i,+})-\epsilon_{\min}}{\epsilon_{\max}-\epsilon_{\min}}, & \text{si } u_{s,i} > 0; \\ \frac{(1-\epsilon_{i,-})-\epsilon_{\min}}{\epsilon_{\max}-\epsilon_{\min}}, & \text{sinon.} \end{cases} \tag{B.8}$$

$\epsilon_{\min}$ and $\epsilon_{\max}$ sont les taux d'erreur minimum et maximum parmi les taux d'erreur positifs et négatifs pour tous les signaux i, $\epsilon_{\min} = \min_i(\min(\epsilon_{i,+}, \epsilon_{i,-}))$ et $\epsilon_{\max} = \max_i(\max(\epsilon_{i,+}, \epsilon_{i,-}))$. Ensuite, la fusion est effectuée en utilisant l'Eq. (B.6).

Fusion à base de régression à noyau : Dans cette section, nous proposons une fonction de fusion basée sur la régression à noyau (Vapnik, 1998; Mahfouz et al., 2014). La méthode à noyau est une technique non paramétrique définissant la fonction $\psi(\cdot)$ qui associe à chaque vecteur d'entrée $\boldsymbol{u}_s(t) = (u_{s,1}(t), ..., u_{s,4}(t))$ une valeur de sortie aussi proche que possible du label y_s du sujet s. En effet, l'algorithme de détection d'anomalies est exécuté pour obtenir les vecteurs de décision individuels. Pour les sujets SDRA, les derniers vecteurs de décision sont sélectionnés car ils sont les plus proches du début du SDRA; alors que pour les sujets non-SDRA, les premiers vecteurs de décision sont considérés car ils représentent au mieux leur stabilité. Plus d'un vecteur de décision pourrait également être sélectionnés pour chaque sujet.

Nous obtenons alors un ensemble d'apprentissage de taille N comprenant N vecteurs de décision $\boldsymbol{u}_n$ avec leurs labels correspondants $y_n, n \in \{1, ..., N\}$. Pour définir $\psi(\cdot)$, nous considérons la méthode des moindres carrés à noyaux (kernel ridge regression). La fonction $\psi(\cdot)$ est estimée à l'aide de l'ensemble d'apprentissage en minimisant l'erreur quadratique moyenne entre les sorties du modèle $\psi(\boldsymbol{u}_s(t))$ et les sorties souhaitées y_n.

$$\min_{\psi \in \mathcal{H}} \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \psi(\boldsymbol{u}_s(t))\right)^2 + \eta \|\psi\|_{\mathcal{H}}^2, \tag{B.9}$$

η est un paramètre de régularisation qui contrôle le compromis entre l'erreur et la complexité de la solution, $\mathcal{H}$ est l'espace de Hilbert fonctionnel et $\|\cdot\|_{\mathcal{H}}$ est la norme dans l'espace de Hilbert. Par conséquent, la fonction optimale peut être écrite selon le théorème du représentation (Scholkopf, Herbrich, and Smola, 2001) comme suit :

$$\psi(\cdot) = \sum_{n=1}^{N} \alpha_n \kappa\left(\boldsymbol{u}_n, \cdot\right), \tag{B.10}$$

Où $\kappa(\cdot, \cdot)$ est un noyau de reproduction et α_n sont des paramètres à estimer. En injectant l'expression de $\psi(\cdot)$ de l'Eq. (B.10) dans l'Eq. (B.9), la solution

$\alpha = (\alpha_1, ..., \alpha_N)$ sera donnée par $\alpha = (K + \eta N I_N)^{-1} Y$, où I_N est la matrice identité de taille $N \times N$, K est la matrice de Gram de taille $N \times N$ ayant $K(n,m) = \kappa(\boldsymbol{u}_n, \boldsymbol{u}_m)$ comme son (n,m)-ème élément. Y est le vecteur de longueur N des labels y_n. Plusieurs fonctions à noyau $\kappa(\cdot)$ existent dans la littérature. Dans ce travail, nous considérons le noyau gaussien défini par:

$$\kappa(\boldsymbol{u}_n, \boldsymbol{u}_m) = \exp\left(\frac{-\|\boldsymbol{u}_n - \boldsymbol{u}_m\|^2}{2\sigma_\kappa^2}\right), \tag{B.11}$$

Où σ_κ est la bande passante du noyau contrôlant avec le paramètre η la tolérance au bruit et le degré de lissage. L'estimation des valeurs de σ_κ et η est présentée dans le paragraphe suivant.

B.2.2.3 Optimisation des paramètres

Dans cette étude, nous proposons d'optimiser les valeurs du seuil D à considérer et la longueur du premier segment τ, ainsi que les paramètres de la fusion de décisions à noyau. Cela nécessite une base d'apprentissage. Un balayage sur un ensemble de valeurs possibles de τ sur {60,120, ..., 1440} minutes, est d'abord effectué. Pour chacune des valeurs de τ, un balayage sur les valeurs possibles de $D \in \{0.1, 0.2, ..., 0.9\}$ est effectué. Pour chaque couple de valeurs, la méthode est appliquée sur les signaux de la base d'apprentissage. Les décisions partielles sont comparées aux vrais états des sujets et les performances sont calculées. Les valeurs optimales τ_o et D_o sont obtenues en maximisant l'indice de Youden $= Se + Sp - 100$. Pour effectuer la technique de fusion de données à noyau, les valeurs η et σ_κ sont obtenues en maximisant l'indice de Youden en variant η et σ_κ sur $\eta = 2^r$ avec $r \in \{-10, -9, ..., 0\}$ et $\sigma_\kappa = 10^{r'}$ et $\sigma_\kappa = -1 * r'$ avec $r' \in \{-5, -4, ..., -1\}$.

B.2.3 Résultats

Nous avons appliqué la méthode proposée sur les deux types de sujets ventilés, le premier ensemble des sujets ayant une ventilation invasive continue et le second type de sujets ayant une ventilation non-invasive. Par conséquent, les résultats sont divisés en deux parties principales, la première illustre les résultats du modèle sur le premier type des sujets et la seconde présente les résultats des sujets ventilés non invasifs.

B.2.3.1 Sujets ventilés invasifs

Nous avons obtenu 40 sujets ayant une ventilation invasive, dont 26 ont développé le SDRA et 14 sont non-SDRA. Une procédure de validation croisée (LOOCV) est envisagée pour évaluer le modèle. Tout d'abord, la corrélation entre les quatre signaux est testée. Les coefficients de corrélation entre chaque paire de signaux sont calculés. Le tableau 3.2 du Chapitre 3 montre que tous les coefficients de corrélation ont des valeurs proches de 0, il existe donc une relation linéaire très faible entre les signaux. Par conséquent, les quatre signaux sont utilisés dans ce travail.

Ensuite, nous obtenons les performances du modèle en appliquant les deux définitions de détection d'anomalies tout en optimisant les paramètres. Nous avons montré que la détection basée sur la définition 2 est plus performante que celle basée sur la définition 1. En plus, la méthode de fusion à noyau a donné les meilleures performances sur les deux ensembles d'apprentissage et de test. La technique de régression à noyau a montré une stabilité de performance même s'il existe des trous d'enregistrement dans les signaux. Comme montré dans la section 3.7.2.4, nous avons réussi à détecter le SDRA dans les deux premiers jours d'enregistrement et en moyenne deux jours avant son occurrence.

B.2.3.2 Sujets ventilés non-invasifs

Dans ce paragraphe, nous considérons les sujets ventilés de façon non-invasive, parce qu'ils représentent le mieux les patients à domicile qui peuvent développer le SDRA. Nous obtenons par cette sélection 50 sujets SDRA et 50 sujets non-SDRA. Cet ensemble de sujets est divisé en ensemble d'apprentissage et ensemble de test en utilisant la validation croisée (k-fold cross validation). Les résultats du modèle sur les sujets ayant une ventilation non-invasive sont cohérents. Le modèle de détection d'anomalies suivi par la fusion par KRR donne les meilleurs résultats sur les quatre signaux. En outre, il présente moins de fluctuation en fonction des signaux utilisés. En termes de prévention précoce, notre modèle a réussi à détecter les sujets SDRA dans le premier jour d'enregistrement et au moins 2 jours avant son occurrence.

B.2.3.3 Comparaison à d'autres méthodes de classification

L'objectif de ce paragraphe est de fournir une comparaison de l'approche proposée à des approches de l'état de l'art. Tout d'abord la méthode de fusion de données est remplacée par la technique de machine à vecteurs de support (SVM), les décisions individuelles sont donc combinées à l'aide des SVM. Les SVM sont un ensemble de techniques d'apprentissage supervisé. Les SVM sont utilisées pour différencier les données en utilisant une marge de décision qui produit la séparation optimale entre les classes. Ensuite, les parties de prise de décision et de fusion sont remplacées par des modèles complets de classification tels que les SVM ou les méthodes à noyau. Par conséquent, le pourcentage d'anomalies est considéré comme un paramètre d'entrée. Le tableau 3.7 montre les performances des méthodes de la littérature et de l'approche proposée. Les résultats montrent que la méthode proposée est plus précise que les autres méthodes.

B.2.4 Conclusion

Dans cette partie, nous avons proposé un modèle basé sur la détection d'anomalies et la fusion des décisions pour prédire le développement du SDRA en temps-réel. Ce modèle a été testé sur les deux types des sujets ventilés. Nous avons obtenu de bonnes performances sur les deux types de sujets. Dans les parties suivantes, nous proposons des modèles basés sur la fusion des caractéristiques. Dans la suite, nous ne considérons que les sujets ventilés non-invasifs.

B.3 La prévention du SDRA par les fonctions de croyance

B.3.1 Introduction

Dans cette partie, nous proposons un modèle de prévention du SDRA basé sur les techniques de fusion de caractéristiques. Ces techniques ont démontré leurs avantage et efficacité par rapport aux techniques de fusion de décisions. Elles consistent en l'extraction de paramètres des signaux, puis la sélection des paramètres les plus pertinents pour la fusion. Par conséquent, nous avons extrait des paramètres linéaires et non linéaires en temps réel, et nous avons sélectionné les paramètres les plus pertinents. Ensuite, un modèle de prévention basé sur la théorie des fonctions de croyance est défini afin

de combiner les informations des paramètres extraits et de générer une décision concernant l'état de santé du sujet surveillé. Cette méthode est testée sur deux approches, la première consiste en la détection binaire déjà décrite dans le section précédente et la deuxième est une prédiction multi-classes pour distinguer entre les quatre groupes, SDRA sévère, modéré, léger et non-SDRA.

B.3.2 Prévention binaire

Ce paragraphe fournit une description détaillée de l'approche de fusion proposée. Le but de ce travail est de définir un modèle à partir d'un ensemble d'apprentissage qui distingue en temps réel les sujets qui vont développer le SDRA et ceux qui ne l'auront pas. Le modèle prend en entrée les signaux collectés $x_{s,i}(1:t), i = 1, \ldots, 4$, allant du début des enregistrements jusqu'au temps réel t, et produit en sortie une décision indiquant si le sujet surveillé va développer le SDRA. L'approche proposée considère une fenêtre fixe de longueur τ prise à la fin des signaux, conduisant à quatre segments $x_{s,i}(t-\tau+1:t)$ à analyser pour chaque instant t. La valeur de τ est obtenue par une optimisation sur les ensembles d'apprentissage, comme indiqué ci-après. Un vecteur de paramètres linéaires et non linéaires est ensuite extrait de ces segments en temps réel t. Nous effectuons une procédure de sélection des paramètres pertinents pour réduire la dimensionnalité du vecteur d'entrée. Ce dernier vecteur est utilisé comme un ensemble de sources d'information pour le modèle des fonctions de croyance. Chaque paramètre permet ainsi la formulation d'une fonction de masses. Les masses sont ensuite combinées pour attribuer une masse finale aux états possibles du sujet surveillé et par la suite obtenir une décision finale pour ce sujet.

B.3.2.1 Extraction de paramètres

Dans ce paragraphe, nous décrivons tous les paramètres extraits. Ayant un segment $x_{s,i}(t-\tau+1:t)$ de chaque signal en temps réel t, plusieurs paramètres linéaires et non linéaires sont extraits. Dans ce qui suit, t^- serait la limite inférieure de l'intervalle de temps considéré, c'est-à-dire $t^- = t - \tau + 1$. Pour les paramètres linéaires, nous considérons la moyenne, l'écart-type, le skewness qui mesure la symétrie dans les données et le kurtosis qui représente la forme de la distribution des données. Les paramètres non linéaires sont l'entropie d'échantillon (sample entropy) mesurant la complexité dans le segment de données (Lake et al., 2002), les facteurs d'analyse

des fluctuations redressées (detrended fluctuation analysis) qui extraient les fluctuations du rythme pour montrer la présence des corrélations à court et long termes (Francis et al., 2002). Soit $p_{s,j}(t), j \in J$ les paramètres obtenus à partir des 4 signaux, J est l'ensemble total des indices des paramètres. Nous effectuons une normalisation de ces paramètres au lieu d'utiliser la valeur brute obtenue. Par conséquent, un premier segment $x_{s,i}(1:\tau)$ de longueur τ est considéré correspondant à l'état initial, supposé stable, de chaque sujet. Nous calculons les paramètres des segments initiaux notés $p_{s,j}^{(0)}$. Ensuite, pour chaque mesure en temps réel, les paramètres obtenus sont normalisés, en calculant leur rapport aux paramètres du premier segment $p_{s,j}(t)/p_{s,j}^{(0)}$. Dans la suite, nous désignons par $p_{s,j}$ le paramètre normalisé.

B.3.2.2 Fonctions de croyance

La théorie des fonctions de croyance, également appelée théorie de Dempster-Shafer, repose sur la modélisation de la croyance en un évènement (Dempster, 1967; Shafer, 1976). Elle est définie sur un cadre de discernement Ω, ensemble de toutes les décisions possibles. Dans cette étude, il existe deux groupes de sujets, le SDRA marqué "+1" et le non-SDRA, marqué "-1". Alors Ω sera défini sur $\{+1, -1\}$. Soit 2^Ω l'ensemble de tous les sous-ensembles possibles de Ω, il est alors défini par $2^\Omega = \{\emptyset, \{+1\}, \{-1\}, \{+1, -1\}\}$.

Les informations fournies par une source peuvent être représentées par une fonction de masse (Smets and Kennes, 1994). Soit $m_j(\cdot)$ la fonction de masse associée au paramètre j. Alors, $m_j(\omega, s, t)$ représente une opinion pondérée sur un sous-ensemble $\omega \in 2^\Omega$ selon la valeur du paramètre j extrait à l'instant t pour le sujet s. La fonction de masse doit avoir les propriétés suivantes $\forall s, t$:

$$\begin{cases} m_j(\emptyset, s, t) = 0, \\ m_j(\omega, s, t) \to [0, 1], \text{ pour } \omega \in 2^\Omega, \\ \sum\limits_{\omega \in 2^\Omega} m_j(\omega, s, t) = 1. \end{cases} \tag{B.12}$$

Afin de construire les fonctions de masse liées à chaque paramètre, nous utilisons les derniers segments des signaux pour un sujet SDRA et les premiers segments pour un sujet non-SDRA. La raison de ce choix est que la fin des signaux précède directement l'occurrence du SDRA et contient donc principalement son instabilité. Pour les sujets non-SDRA, plus les segments se rapprochent du début des signaux, plus nous sommes sûrs de la stabilité du sujet. Soit S_{+1} et S_{-1} les ensembles d'indices des sujets SDRA et

non-SDRA de l'ensemble d'apprentissage, respectivement. On obtient alors des ensembles de paramètres normalisés $p_{s,j}, s \in S_{+1}$ et $p_{s',j}, s' \in S_{-1}$. Les fonctions de masses sont ensuite définies par les distributions de probabilité $Q_{j,\{+1\}}(\cdot)$, $Q_{j,\{-1\}}(\cdot)$ et $Q_{j,\{+1,-1\}}(\cdot)$ pour chaque sous-ensemble de décision avec $m_j(\emptyset, s, t) = 0$, comme suit :

$$m_j(\omega, s, t) = \frac{Q_{j,\omega}(p_{s,j}(t))}{\sum\limits_{\omega' \in 2^\Omega, \omega' \neq \emptyset} Q_{j,\omega'}(p_{s,j}(t))}, \omega \in 2^\Omega, \omega \neq \emptyset. \tag{B.13}$$

Nous effectuons ensuite une opération d'affaiblissement conditionnel des fonctions de masse qui permet de modéliser la fiabilité des paramètres en fonction de chaque état en introduisant un coefficient $\alpha_{j,\omega}$ (Mercier, Lefèvre, and Delmotte, 2012; Mercier, Pichon, and Lefèvre, 2016). Soit $\beta_{j,\{+1\}} = 1 - \alpha_{j,\{+1\}}$ et $\beta_{j,\{-1\}} = 1 - \alpha_{j,\{-1\}}$ les degrés de fiabilité d'un paramètre j sachant que les vrais états sont "+ 1" et "- 1" respectivement. Ensuite, la fonction de masse affaiblie ${}^{\alpha}m_j(\cdot)$ attribuée au paramètre j est donnée par

$$\begin{cases} {}^{\alpha}m_j(\{+1\}, s, t) = \beta_{j,\{-1\}} m_j(\{+1\}, s, t), \\ {}^{\alpha}m_j(\{-1\}, s, t) = \beta_{j,\{+1\}} m_j(\{-1\}, s, t), \\ {}^{\alpha}m_j(\Omega, s, t) = m_j(\Omega, s, t) \\ \qquad\qquad + \alpha_{j,\{-1\}} m_j(\{+1\}, s, t) \\ \qquad\qquad + \alpha_{j,\{+1\}} m_j(\{-1\}, s, t), \\ {}^{\alpha}m_j(\emptyset, s, t) = 0. \end{cases} \tag{B.14}$$

Les taux d'erreur $\alpha_{j,\{+1\}}$ et $\alpha_{j,\{-1\}}$ sont estimés à l'aide des fonctions de distribution $Q_{j,\omega}(\cdot)$. Le taux d'erreur $\alpha_{j,\omega}, \omega = \{+1\}$ ou $\{-1\}$, est calculé pour toutes les valeurs du paramètre j, où $Q_{j,\omega}(p)$ est inférieur à tout $Q_{j,\omega'}(p)$, pour $\omega' = \omega \in 2^\Omega$. Par conséquent, il est défini sur $D_{j,\omega} = \{p | Q_{j,\omega}(p) < Q_{j,\omega'}(p), \forall \omega' \in 2^\Omega, \omega' \neq \omega\}$, comme suit :

$$\alpha_{j,\omega} = \epsilon_{j,\omega} = \int_{D_{j,\omega}} Q_{j,\omega}(p) dp. \tag{B.15}$$

Pour appliquer la fusion des caractéristiques, nous avons combiné les fonctions de masses en utilisant la règle conjonctive normalisée de combinaison de la théorie de Dempster-Shafer. Elle est notée par $\oplus$ et définie par

l'équation suivante :

$$m_{\oplus}(\omega, s, t) = \frac{\sum\limits_{\cap\omega'^{(j)}=\omega} \prod\limits_{j\in J} {}^{(\alpha)}m_j(\omega'^{(j)}, s, t)}{1 - \sum\limits_{\cap\omega'^{(j)}=\emptyset} \prod\limits_{j\in J} {}^{(\alpha)}m_j(\omega'^{(j)}, s, t)}. \tag{B.16}$$

Finalement, une décision est générée selon la transformation pignistique suivante :

$$BetP(\omega, s, t) = \sum_{\omega'\in 2^{\Omega}, \omega\subseteq\omega'} \frac{m_{\oplus}(\omega', s, t)}{\mid \omega' \mid}, \forall\omega \in \Omega. \tag{B.17}$$

La masse finale du sous-ensemble {+ 1, -1} est divisée entre {+1} et {-1}. L'état ayant le plus haut niveau pignistique est sélectionné pour chaque sujet à chaque instant t. Ainsi, une alerte de SDRA est générée s'il y aura une décision positive.

B.3.3 Classification multi-classe

Dans le travail déjà présenté, nous avons effectué une classification binaire afin de distinguer entre le groupe SDRA composé de sujets atteints du SDRA sévère ou modéré et le groupe non-SDRA comprenant les sujets présentant un SDRA léger et les sujets n'ayant pas le SDRA. Dans ce qui suit, nous proposons d'effectuer une classification multi-classes pour distinguer entre les quatre états des sujets : l'état stable et les états du SDRA léger, modéré et sévère. D'où, on aura le cadre de discernement défini sur $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, avec ω_k, $k = 1, \ldots, 4$.

B.3.4 Identification des groupes

Dans cette partie, nous proposons d'identifier chaque état du SDRA dans les signaux des sujets à partir des valeurs de PaO_2/FiO_2. En effet, nous identifions dans les signaux de chaque sujet SDRA des segments légers, des segments modérés et des segments sévères. En effectuant cette procédure, nous augmentons la dimension de l'ensemble d'apprentissage et nous spécifions la modélisation de chaque état. Afin de modéliser le groupe non-SDRA, nous ne considérons que les segments extraits des signaux des sujets non-SDRA afin d'assurer la stabilité dans les signaux.

Afin d'effectuer cette identification, la longueur totale des signaux enregistrés est considérée. Premièrement, nous identifions dans les séries temporelles les intervalles de temps auxquels chacune des phases se produit. Ainsi, nous identifions les trois niveaux de SDRA chez un sujet sévère s'ils existent. Pour un sujet SDRA modéré, nous identifions les phases légère et modérée. Les phases légères sont identifiées pour un sujet développant un SDRA léger. Enfin, les segments non-SDRA sont identifiés uniquement à partir des sujets non-SDRA. Ensuite, nous sélectionnons un segment de longueur τ dans chaque phase identifiée pour calculer les paramètres. Les paramètres sont ensuite normalisés en utilisant un segment du début des signaux.

B.3.5 Modèle des fonctions de croyance

Une fois l'ensemble d'apprentissage est construit, les distributions de probabilité de chaque sous-ensemble $\omega \in 2^{\Omega}$ sont estimées. Le fait que les données d'apprentissage ne sont pas équilibrées entre les groupes, produira des distributions biaisées. Nous proposons d'effectuer une technique de suréchantillonnage des groupes minoritaires puisque l'ensemble d'apprentissage est petit. Ainsi, la technique SMOTE est utilisée pour équilibrer les ensembles d'apprentissage (Chawla et al., 2002).

Ensuite, les fonctions de distribution $Q_{j,\omega}(\cdot)$ sont estimées pour chaque paramètre en fonction du sous-ensemble $\omega \in 2^{\Omega}$. Les fonctions de masse $m_j(\omega, s, t)$ sont ensuite définies pour chaque paramètre j selon l'état ω en utilisant l'Eq. (4.2). Les fonctions de masse seront affaiblies selon l'un des deux types d'affaiblissement, de base ou contextuel définis dans l'Eq. (4.8). Ensuite, ces masses seront combinées par la fonction conjonctive. Finalement, une décision finale sera générée pour le sujet s en temps t en utilisant la transformation pignistique.

Afin de générer une alerte du SDRA, des seuils de décisions successives doivent être définis pour chacun des trois sous-groupes du SDRA en utilisant l'ensemble d'apprentissage. Chacun des seuils est déterminé à partir du groupe d'apprentissage correspondant. Le seuil est défini par le nombre moyen de successions de la décision correspondante sur tout le groupe d'apprentissage considéré.

B.3.6 Résultats

Dans cette partie, nous présentons les résultats de l'approche de fusion de caractéristiques proposée pour la prévention du SDRA binaire et multi-classes. L'approche est testée sur les sujets ventilés non invasifs. L'ensemble de données est formé de 100 sujets, dont 50 sujets SDRA et 50 non-SDRA. La procédure de validation a été réalisée en utilisant une validation croisée 5-fold répétée 10 fois. Nous présentons en premier lieu le modèle en considérant le problème binaire, ensuite nous présentons les résultats du problème multi-classes.

B.3.6.1 Exemple sur un ensemble d'apprentissage

Dans ce paragraphe, nous avons considéré un exemple d'ensemble d'apprentissage et son ensemble de test pour illustrer chaque étape de l'approche. Une analyse statistique est réalisée tout d'abord afin de visualiser les différences statistiques entre les 2 groupes de sujets. Les résultats présentés dans le tableau 4.1 et les figures 4.5, 4.6, 4.7 et 4.8 montrent qu'il existe pour chaque signal différents paramètres qui présentent une différence significative entre les groupes de sujet. Ensuite, nous avons effectué un classement et une sélection de paramètres. Les paramètres sont classés selon le taux d'erreur d'après les distributions estimées comme montre la figure 4.9. Cette figure montre la cohérence entre cette méthode de classement et les résultats du test statistique. Ainsi, nous effectuons la sélection par recherche séquentielle, pour chaque itération un paramètre est ajouté. Pour chaque paramètre ajouté, le modèle de prévention est appliqué en temps réel sur l'ensemble d'apprentissage et la précision du modèle est calculée comme suit :

$$\text{Précision} = \frac{\text{Nombre de sujets correctement identifiés}}{\text{Nombre total des sujets d'apprentissage}}.$$

Une fois la précision diminue d'un facteur de plus de 5%, la sélection s'arrête en obtenant les paramètres déjà insérés dans l'algorithme. Nous avons obtenu pour cet ensemble d'apprentissage 10 paramètres sélectionnés.

Afin de générer une décision finale à partir de l'analyse en temps réel, il est possible de générer l'alerte dès la première apparition d'une décision positive ou d'attendre une succession de décisions positives. Nous avons proposé dans cette approche d'optimiser un seuil de succession de décisions

positives à partir de l'ensemble d'apprentissage. Par conséquent, nous balayons la valeur de succession de décisions sur l'intervalle [1, ..., 30] heures et les performances Se et Sp sont calculées. Le seuil qui maximise l'indice de Youden est alors pris en compte. La figure 4.11 illustre la courbe COR de l'algorithme proposé en prenant la fusion des dix premiers paramètres obtenus dans la section précédente.

Comme décrit précédemment, nous avons utilisé une fenêtre de longueur fixe τ pour extraire le vecteur de paramètres pour tous les sujets. De même, nous proposons d'optimiser sa valeur. Différentes tailles de fenêtre sont évaluées dans cette étude, les valeurs sont considérées sur l'intervalle [4,6, ..., 24] heures. La figure 4.12 illustre la précision de chaque taille de fenêtre (en heures) ainsi que le nombre de paramètres sélectionnés et le seuil affiché sous forme de couple (nombre de paramètres, seuil). Comme le montre la figure, la taille de la fenêtre = 24 heures présente la précision la plus élevée de 77,5% sur l'ensemble d'apprentissage considéré. Comme résultat final, cette division d'ensembles a donné une sensibilité de 87,5% et une spécificité de 67,5% sur l'ensemble d'apprentissage et Se = 70% et Sp = 80% sur l'ensemble de test pour les dix paramètres sélectionnés, un seuil de succession = 21 et une longueur de segment = 24 heures.

B.3.6.2 Evaluation des performances

Dans ce paragraphe, nous présentons les performances globales du modèle et comparons l'influence de chaque phase ajoutée pour arriver au modèle complet. Afin d'assurer la stabilité des résultats, nous avons effectué une validation croisée 5-fold répétée 10 fois. Ensuite, la performance finale est obtenue en calculant la moyenne des performances obtenues de la procédure de validation croisée. Le tableau 4.3 présente les performances de chaque phase ajoutée pour obtenir à la fin les performances du modèle proposé. Le tableau montre que le modèle proposé donne les meilleures performances sur les ensembles d'apprentissage et de test. En plus, nous avons pu détecter le SDRA au moins 2 heures avant son occurrence.

B.3.6.3 Comparaison à d'autres méthodes

Le modèle proposé est comparé à certaines techniques de classification telles que les tests d'hypothèses bayésiennes, les *K*-plus proches voisins (*K*-NN) et les SVM. Les tests d'hypothèses bayésiennes sont basés sur le

théorème de Bayes. C'est un modèle de probabilité qui consiste à factoriser une distribution de probabilité conjointe en un ensemble de distributions conditionnelles pour chaque variable. Ils constituent la forme la plus simple de la théorie des croyances où les ambiguïtés entre les sources ne sont pas prises en compte. La méthode des K-plus proches voisins est l'un des algorithmes de base de l'intelligence artificielle. Il attribue une nouvelle observation à une classe en choisissant la classe ayant les K voisins d'apprentissage les plus proches à l'observation. Une validation croisée a été réalisée pour obtenir la valeur de K qui minimise l'erreur de test. La valeur optimale de K est égale à 8. Le tableau 4.4 montre que le modèle proposé surpasse en performance le K-NN, le SVM ainsi que les tests bayésiens.

B.3.6.4 Classification multi-classes

Dans cette section, nous présentons les résultats de la méthode proposée pour détecter chacune des sous-groupes du SDRA. Ainsi, de nouveaux groupes de sujets sont définis à l'aide des valeurs du rapport PaO_2/FiO_2. Cela conduit à 14 sujets du groupe sévère, 36 sujets du groupe modéré, 15 sujets du groupe SDRA léger et 35 sujets non-SDRA. Afin de valider le modèle, une procédure de validation Hold-out est considérée en prenant 5 sujets de chaque groupe pour l'ensemble de test et les sujets restants comme ensemble d'apprentissage. Ainsi, nous obtenons un total de 80 sujets dans la base d'apprentissage et 20 sujets pour la validation. La procédure de validation est répétée 10 fois.

Le modèle proposé pour la prédiction multi-classes a donné des résultats intéressants. Nous avons obtenus des bonnes précisions pour les groupes sévère et non-SDRA en utilisant le modèle des fonctions de croyance simple. L'affaiblissement basique a montré des résultats similaires, mais en améliorant la précision du groupe non-SDRA. En revanche, l'affaiblissement contextuel a basculé les résultats, en améliorant les précisions des groupes modéré et léger et diminuant considérablement la précision des groupes sévère et non-SDRA sur l'ensemble d'apprentissage. Nous avons obtenu des résultats cohérents en utilisant l'ensemble de test. Ainsi, nous avons obtenu une précision élevée pour les groupes sévère et non-SDRA et une précision moyenne pour le groupe modéré avec l'affaiblissement basique. L'affaiblissement contextuel a aboutit à une précision élevé pour les trois groupes de SDRA.

B.3.7 Conclusion

Nous avons proposé un modèle de fusion de caractéristiques pour la prédiction du SDRA basé sur la théorie des fonctions de croyance. Nous avons testé le modèle sur deux problèmes de prévention, un problème binaire et un problème multi-classes. Nous avons obtenu de hautes performances dans l'identification binaire ainsi que dans l'identification à quatre classes. De plus, les performances sont améliorées par rapport au modèle précèdent de fusion de décisions. Les fonctions de masse ont été calculées à partir des distributions de probabilité. La partie suivante propose une nouvelle méthode de construction des fonctions de masse.

B.4 La prévention du SDRA par les croyances et les noyaux

B.4.1 Introduction

La théorie des fonctions de croyance se caractérise principalement par sa capacité à gérer des informations provenant de sources distinctes et les combiner (Khaleghi et al., 2013). Son avantage est la modélisation de tous les sous-ensembles qui présentent les classes au lieu de modéliser les singletons seuls. Il existe diverses études proposant des extensions aux fonctions de croyance, notamment pour l'affaiblissement (Chou et al., 2011; Zhang et al., 2013; Kahraman, Onar, and Oztaysi, 2015) et la combinaison (Yong et al., 2004; Xu et al., 2017).

Malgré l'effort induit dans ce domaine, il n'existe pas assez d'études sur la construction des fonctions de masse, qui présente le problème le plus important de la théorie des fonctions de croyance. Cette mesure pourrait être une mesure de similarité, de satisfiabilité, de ressemblance ou d'inclusion selon les caractéristiques des données à comparer (Diaz, Rifqi, and Bouchon-Meunier, 2006). Une méthode de base pour la construction des fonctions de masse est celle basée sur les distributions de probabilité comme effectué dans la partie précédente de notre travail (Dempster, 1967; Xiao, 2019; Deng et al., 2016; Xu et al., 2013). En outre, une nouvelle construction de fonctions de masse a été proposée par (Denoeux, 2000; Hu et al., 2003), elle est basée sur les réseaux de neurones multicouches. Le vecteur de sortie du réseau de neurones est utilisé pour déterminer les fonctions de masses. Les fonctions

de masse sont attribuées en fonction de la distance entre les données et des échantillons de base, qui sont extraits des données d'apprentissage (Zhang et al., 2014). Une autre méthode de construction repose sur la distance entre les données de test et les classes formées par les données d'apprentissage (Zhang et al., 2018).

Le modèle présenté précédemment est basé sur la théorie des fonctions de croyance pour combiner des fonctions de masse obtenues de différents paramètres. Les fonctions de masse ont été attribuées en trouvant les distributions de probabilité des données de la base d'apprentissage. Dans cette partie, nous proposons une extension du modèle précédent, en particulier sur la phase de construction des fonctions de masse. Nous proposons une approche originale pour la construction de ces fonctions, elle est basée sur un algorithme d'apprentissage automatique à noyau. En outre, nous proposons une nouvelle extraction de paramètres non-linéaires pour élargir le vecteur d'entrée et par la suite avoir plus d'informations sur les caractéristiques des signaux.

B.4.2 Approche proposée

Nous reconsidérons le problème de détection du SDRA en deux groupes décrit dans les parties précédentes. Etant donné les quatre signaux extraits $x_{s,i}(1 : t)$ en temps réel pour un sujet s, nous extrayons les paramètres déjà décrits dans la partie B.2 et nous ajoutons de nouveaux paramètres issus de l'analyse de quantification de récurrences. Ensuite, l'ensemble des paramètres est utilisé dans l'apprentissage à noyau afin d'attribuer des fonctions de masse à chaque sous-ensemble $\{+1\}$, $\{-1\}$ et $\{+1, -1\}$ en fonction de chaque paramètre. Les fonctions de masse attribuées sont considérées comme des entrées à la combinaison des croyances. Les masses seront affaiblies et combinées comme décrit dans la partie précédente.

B.4.2.1 Analyse de quantification de récurrence

L'analyse de quantification de récurrence (AQR) quantifie le nombre et la durée des récurrences d'un système dynamique représenté par sa trajectoire dans l'espace des phases (Takens, 1981). Ce dernier est un espace multidimensionel qui représente tous les états possibles d'un système. Cette représentation s'effectue par la méthode des retards. Elle est caractérisée par sa dimension M et un retard temporel t_d (Buzug and Pfister, 1992; Cellucci,

Albano, and Rapp, 2003). Soit un vecteur $x_{s,i}(\ell)$ à l'instant ℓ défini par $X_{s,i}(\ell)$ dans l'espace des phases, comme suit:

$$X_{s,i}(\ell) = [x_{s,i}(\ell), x_{s,i}(\ell + t_d), x_{s,i}(\ell + 2t_d, \ldots, x_{s,i}(\ell + (M-1)t_d))].$$

Le diagramme des récurrences projette la trajectoire de l'espace des phases de dimension M en une illustration bidimensionnelle des récurrences (Eckmann, Kamphorst, and Ruelle, 1987). Le diagramme des récurrences est présenté sous forme de matrice carrée ayant des éléments correspondant aux instants des récurrences. Il est composé de deux types de lignes, telles que les lignes diagonales et verticales. Les lignes diagonales correspondent aux périodes où la trajectoire passe par les mêmes régions de l'espace de phase pour des temps différents et les lignes verticales représentent les périodes où les segments restent dans la même région d'espace de phase. Ensuite nous extrayons différents paramètres du diagramme des récurrences, tels que :

- à partir des lignes diagonales - le déterminisme (DET), la longueur moyenne des lignes diagonales ($\langle L \rangle$), La plus longue diagonale (L_{max}) et l'entropie ($ENTR$);
- à partir des lignes verticales - la laminarité (LAM), la longueur moyenne des lignes verticales (TT) et La plus longue verticale (V_{max});
- à partir du diagramme - le taux des récurrences ($R_e R_a$) et l'entropie de densité de période de récurrence ($RPDE$).

Par conséquent, nous extrayons de chaque signal les paramètres linéaires qui sont la moyenne (μ), l'écart type (σ), le skewness (Sk) et le kurtosis (Kt), ainsi que les paramètres non linéaires constituant l'entropie de l'échantillon ($SampEn$), les facteurs d'analyse des fluctuations redressées (DFA_1 et DFA_2) ainsi que les paramètres déjà expliqués de l'analyse de quantification de récurrence. Nous obtenons de cette extraction 16 paramètres par signal et donc 64 paramètres pour chaque sujet s à chaque instant t. Ils sont définis par $p_{s,j}(t)$, $j = 1, \ldots, 64$. Comme expliqué précédemment, nous extrayons un paramètre $p_{s,j}^{(0)}$ du segment initial des signaux pour un sujet s de son paramètre de type j. Chaque sujet ayant son propre état normal, nous normalisons les paramètres extraits en temps réel.

B.4.2.2 La théorie des fonctions de croyance

Nous proposons, dans cette partie, une nouvelle technique pour l'attribution des fonctions de masses. Elle est basée sur l'apprentissage

automatique à noyau. Étant donné le modèle des fonctions de croyance, défini dans la partie précédente, et qui opère sur le cadre de discernement $\Omega = \{+1, -1\}$, notre objectif est de trouver, à partir des bases d'apprentissage, la fonction de masse $m_j(\cdot)$ pour chaque sous-ensemble $\omega \in 2^\Omega$ associée à chaque paramètre $j \in J$.

Comme décrit dans les chapitres précédents, les paramètres sont extraits des derniers segments des signaux pour les sujets SDRA et du début des signaux pour les sujets non-SDRA. Ainsi, nous obtenons un couple de paramètres extraits de signaux d'apprentissage avec leurs labels correspondants $(p_{n,j}, y_n)$, $n \in 1, \ldots, N$, avec $y_n = +1$ pour un sujet SDRA et $y_n = -1$ pour les sujets non-SDRA. Par conséquent, le calcul des fonctions de masse est effectué à l'aide de ces couples comme suit.

Attribution de fonctions de masse par régression
La régression est généralement utilisée pour prédire une variable dépendante à partir d'une entrée. L'objectif de ce modèle est de trouver un ensemble de fonctions $g_j(\cdot)$ qui prennent en entrée les valeurs des paramètres j et les sous-ensembles $\omega \in 2^\Omega$ et produisent en sortie les masses associées à ω. Une première proposition pour construire ces fonctions de masses consiste à définir une fonction scalaire $\psi_j(\cdot)$ pour chaque paramètre j. En effet, la fonction $\psi_j(\cdot)$ prend en entrée les valeurs des paramètres $p_{n,j}$, $n = \{1, \ldots, N\}$ de l'ensemble d'apprentissage et donne un label $y_n = +1$ pour les sujets SDRA et $y_n = -1$ pour les sujets non-SDRA. Afin de trouver les fonctions de masse pour un paramètre donné $p_{s,j}(t)$ à partir des fonction $\psi_j(p_{s,j}(t))$, nous définissons la fonction $g_j(\cdot)$ comme suit :

$$\begin{cases} m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = \max(\psi(p_{s,j}(t)), 0), \\ m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = |\min(\psi(p_{s,j}(t)), 0)|, \\ m_j(\{+1, -1\}, s, t) = g_j(\{+1, -1\}, p_{s,j}(t)) = 1 - |\psi(p_{s,j}(t))|. \end{cases} \tag{B.18}$$

où $|\psi(\cdot)|$ est la valeur absolue de $|\psi(\cdot)|$.

Ce calcul de masse attribue une masse nulle à l'un des sous-ensembles de Ω, ce qui pourrait pénaliser la combinaison de masses. Alternativement, nous proposons une nouvelle construction des fonctions de masse en considérant un vecteur de labels Y_n au lieu du label scalaire pour chaque sujet n de l'ensemble d'apprentissage. En effet, un sujet SDRA reçoit un label $[1\ 0]$ au

lieu de "+1", alors qu'un sujet non-SDRA reçoit un label de [0 1] au lieu de "-1". Ensuite, le modèle $\boldsymbol{\psi_j}(\cdot)$ est construit et optimisé à l'aide de l'ensemble d'apprentissage $(p_{n,j}, Y_n)$, $n \in 1, \ldots, N$. Ainsi, nous calculons la fonction $\boldsymbol{\psi_j}(p_{s,j}(t))$ appliquée à une nouvelle valeur du paramètre j en temps réel $p_{s,j}(t)$. Celle-ci aboutit à une sortie vectorielle bidimensionnelle. Son premier élément représente l'état SDRA et le second l'état non-SDRA. La fonction de masse est alors définie comme suit :

$$\begin{cases} m_j(\{+1\}, s, t) = g_j(\{+1\}, p_{s,j}(t)) = \psi_{j,1}(p_{s,j}(t)); \\ m_j(\{-1\}, s, t) = g_j(\{-1\}, p_{s,j}(t)) = \psi_{j,2}(p_{s,j}(t)); \\ m_j(\{+1,-1\}, s, t) = g_j(\{+1,-1\}, p_{s,j}(t)) = 1 - \psi_{j,1}(p_{s,j}(t)) - \psi_{j,2}(p_{s,j}(t)). \end{cases} \tag{B.19}$$

Méthodes de régression à noyau

Dans ce paragraphe, nous décrivons les différentes techniques de régression à noyau qui peuvent être appliquées pour construire les modèles $\psi_j(\cdot)$ ayant l'ensemble d'apprentissage $(p_{n,j}, y_n)$ pour un modèle scalaire ou les modèles $\boldsymbol{\psi_j}(\cdot)$ avec l'ensemble d'apprentissage $(p_{n,j}, Y_n)$ pour un modèle vectoriel, $n = 1, \ldots, N$. Pour plus de simplicité, nous utilisons les notations $\psi_j(\cdot)$ et y_n pour le modèle et les labels correspondants. Ceux-là représenteront un élément du modèle vectoriel et de ses labels. Les méthodes à noyau projettent l'entrée p en un label de sortie y en se servant du noyau reproduisant $\kappa(\cdot, \cdot)$. La fonction $\psi_j(\cdot)$ est estimée en minimisant l'erreur entre la sortie du modèle $\psi_j(p_{n,j})$ est la sortie désirée y_n, comme suit

$$\mathcal{L}\left((y_1, \psi_j(p_{1,j})), \ldots, (y_N, \psi_j(p_{N,j}))\right) + \eta \mathcal{R}(\|\psi_j\|^2_{\mathcal{H}}), \tag{B.20}$$

où $\mathcal{L}(\cdot)$ et $\mathcal{R}(\cdot)$ sont respectivement une fonction coût et une fonction croissante sur $[0, +\infty[$ représentant un terme de régularisation avec η le paramètre qui maintient le compromis entre la finesse de la solution et sa complexité, et $\|\cdot\|_{\mathcal{H}}$ est la norme dans l'espace de Hilbert.

Les méthodes à noyau reposent sur le théorème de représentation qui permet de définir la fonction $\psi_j(\cdot)$ par une simple combinaison linéaire des noyau $\kappa(p_{n,j}, \cdot)$, comme suit:

$$\psi_j(\cdot) = \sum_{n=1}^{N} \lambda_{n,j} \kappa(p_{n,j}, \cdot). \tag{B.21}$$

$\psi_j(\cdot)$ sera définie en optimisant les paramètres $\lambda_{n,j}$ selon le problème de l'Eq. (B.20). Le premier terme de la fonction pourra prendre plusieurs formes pour aboutir à différents techniques de régression. Il existe différentes méthodes d'apprentissage à noyau telles que la technique de kernel ridge regression (KRR) présentée ci-dessous (Saunders, Gammerman, and Vovk, 1998).

Kernel ridge regression :
Selon la technique de kernel ridge regression (Saunders, Gammerman, and Vovk, 1998), la fonction coût est la fonction d'erreur quadratique moyenne et le terme de régularisation est $\mathcal{R}(\|\psi_j\|^2_{\mathcal{H}}) = \|\psi_j\|^2_{\mathcal{H}}$. Nous obtenons le problème d'optimisation suivant :

$$\psi_j(\cdot) = \arg\min_{\psi_j \in \mathcal{H}} \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \psi_j(p_{n,j})\right)^2 + \eta \|\psi_j\|^2_{\mathcal{H}}. \tag{B.22}$$

En injectant l'Eq. (B.21) dans l'Eq. (B.22), nous obtenons la solution au problème d'optimisation suivante :

$$\boldsymbol{\lambda}_j = (K_j + \eta N I_N)^{-1} \mathbb{Y}. \tag{B.23}$$

Où $\boldsymbol{\lambda}_j$ est un vecteur de coefficients de taille $N \times 1$ ayant $\lambda_{n,j}$ son n-ième élément, $\mathbb{Y}$ est le vecteur de label pour chaque donnée d'apprentissage de taille $N \times 1$, I_N est la matrice identité de taille $N \times N$ et K_j est la matrice de Gram de taille $N \times N$.

Une fois nous avons attribué les fonctions de masse à chaque paramètre j, nous pouvons appliquer la procédure d'affaiblissement de ces fonctions selon l'Eq. (B.14). Les taux d'affaiblissement α_j sont calculés à partir des performances calculées pour chaque paramètre j. D'où les facteurs seront définies par $\alpha_{j,\{+1\}} = 1 - Se_j$ et $\alpha_{j,\{-1\}} = 1 - Sp_j$. Pour plus de détails, se référer à la Section 5.3.3.3. Ensuite, les fonctions de masse sont combinées suivant la fusion conjonctive dans l'Eq. (B.16) et une décision finale sera obtenue par la transformation pignistique de l'Eq. (B.17).

L'algorithme étant exécuté en temps réel, nous obtenons une décision à chaque fois un nouvel enregistrement est ajouté. Ainsi, nous pouvons choisir

d'avoir un modèle très sensible qui génère une alarme dès la première décision positive ou d'être plus sûr de l'alerte générée en attendant une succession de décisions positives. Nous proposons dans notre travail d'optimiser le seuil de succession de décisions positives. L'optimisation est ainsi appliquée sur l'ensemble d'apprentissage en balayant la valeur du seuil sur [1,30] heures. Le seuil de succession de décisions est alors obtenu en minimisant le taux d'erreur sur la base d'apprentissage.

B.4.2.3 Sélection des paramètres

Dans ce modèle proposé, nous avons étendu la phase d'extraction de paramètres pour inclure davantage des paramètres issus de l'analyse de quantification de récurrence ce qui a abouti à 64 paramètres pour chaque sujet s. L'augmentation de la dimensionnalité du vecteur d'entrée pourra conduire à une meilleure représentation des données, mais elle accroît très souvent la complexité des calculs et entraîne à une contradiction dans les informations fournies. Par conséquent, nous effectuons une procédure de sélection afin de garder les paramètres les plus pertinents, de réduire ainsi la dimensionnalité du vecteur d'entrée et la complexité du calcul.

Dans cette partie, nous proposons une nouvelle méthode de sélection de paramètres. Ainsi, nous commençons par la sélection de la sous-combinaison de deux paramètres minimisant le taux d'erreur du modèle. Ensuite, nous considérons le second niveau de sous-combinaisons, et nous testons les paramètres déjà sélectionnés avec les paramètres restants chacun à la fois. Nous sélectionnons le nouveau vecteur de paramètres représentant le taux d'erreur le plus faible. Cette procédure sera répétée jusqu'à ce que le nouveau vecteur de paramètres montre une augmentation du taux d'erreur.

B.4.3 Résultats

Dans cette section, nous présentons les résultats du modèle de prévention avec la nouvelle méthode d'attribution des fonctions de masse proposée, qui est basée sur la régression à noyau. La validation de ce modèle est effectuée pour les sujets soumis à une ventilation non-invasive. Les résultats sont obtenus en considérant une fenêtre de longueur fixe $\tau = 24$ heures. Tout d'abord, un test statistique est effectué pour identifier parmi les nouveaux paramètres extraits ceux qui montrent une différence significative entre le groupe SDRA et le groupe non-SDRA. Ensuite, une illustration de chaque

étape de la méthode est présentée en considérant un exemple d'ensemble d'apprentissage. Enfin, la performance globale est illustrée à l'aide de la validation croisée 5-fold répétée 10 fois.

B.4.3.1 Exemple sur un ensemble d'apprentissage

Nous considérons, dans cette partie le même exemple d'ensemble d'apprentissage illustré dans la partie B.3.6.1. Pour cet ensemble, nous avons testé statistiquement les nouveaux paramètres extraits afin de trouver les paramètres qui montrent une différence statistique significative entre les groupes SDRA et les groupes non SDRA. Différents paramètres issus de la technique AQR ont montré une différence significative entre les deux groupes comme le montre le tableau 5.1.

Les performances de KRR dépendent du choix de la bande passante du noyau σ_κ et du paramètre de régularisation η. Ces paramètres contrôlent la tolérance au bruit et le degré de fluidité, ainsi leurs valeurs affectent les performances de l'algorithme de KRR. Nous proposons d'optimiser leurs valeurs, pour chaque paramètre, en minimisant le taux d'erreur de l'algorithme complet sur les données d'apprentissage. Pour plus de détails, se référer à la section 5.4.2. Les paramètres donnant les plus bas taux d'erreur sont donnés dans le tableau 5.2. Les paramètres extraits du rythme cardiaque et du rythme respiratoire présentent les plus faibles erreurs comparées aux autres signaux.

La dimensionnalité du vecteur d'entrée étant très élevée, il est nécessaire d'effectuer une procédure de sélection pour réduire cette dimensionnalité et ainsi la complexité des calculs. Nous proposons une procédure de sélection plus globale que celle proposée dans la partie précédente. Cette procédure est illustrée dans la figure 5.2. À partir de ces résultats, nous avons proposé un nouveau critère pour arrêter la procédure de sélection. Ainsi, la procédure de sélection se termine si le nouvel ensemble de paramètres ne présente pas une diminution de l'erreur.

Nous avons décrit précédemment qu'une optimisation du seuil de succession de décisions positives est envisagée. Ainsi, l'optimisation du seuil est appliquée sur la plage de [1,..., 30] heures sur l'ensemble d'apprentissage. À partir de la courbe COR de la figure 5.3, nous avons obtenu un seuil = 19

pour l'exemple considéré.

Cet ensemble d'apprentissage a abouti à une précision de 82% en utilisant les labels vectoriels et 78.75% pour les labels scalaires. En plus, nous avons obtenu pour l'ensemble de test des précisions de 85% et 70% pour les labels vectoriels et scalaires respectivement. Ainsi, nous ne considérons que les labels vectoriels pour la suite.

B.4.3.2 Evaluation des performances

La performance globale du modèle est obtenue par la moyenne des performances obtenues par la validation croisée. Le modèle brut des fonctions de croyance c.-à-d. sans sélection de paramètres ni affaiblissement a donné des bonnes performances sur l'ensemble d'apprentissage. Ces performances sont améliorées avec l'affaiblissement et la sélection selon les deux critères présentés. Nous avons obtenu finalement une précision de 85.6% sur les ensembles d'apprentissage et 76.68% sur les ensembles de test. En outre, nous avons pu détecter le SDRA en moyenne dans les premières 29 heures d'enregistrements, et au moins 24 heures avant son occurrence. Finalement, en comparant ce modèle au modèle proposé dans la partie précédente, ce nouveau modèle surpasse le modèle proposé précédemment en termes de précision sur les bases d'apprentissage et de test.

B.4.4 Conclusion

Cette dernière partie présente un modèle de fusion de caractéristiques basé sur l'extraction de paramètres et la théorie des fonctions de croyance. Le modèle proposé est une extension du modèle précèdent du point de vue de l'extraction des paramètres et de la construction des masses. Ainsi, les masses ont été construites via une phase d'apprentissage basée sur la régression à noyau. Le modèle proposé a montré de meilleures performances par rapport à tous les modèles proposés précédemment.

B.5 Conclusion et perspectives

Ce manuscrit aborde un problème de santé important chez les personnes âgées qui est le syndrome de détresse respiratoire aigue. Cette thèse propose différents modèles de prédiction de ce syndrome en temps réel basés sur différentes techniques de fusion d'informations. Dans cette section, nous

résumons les modèles proposés présentés dans le manuscrit, ainsi que les contributions ajoutées de chaque modèle. Ensuite, nous proposons des perspectives futures de recherche.

B.5.1 Contributions principales

La contribution principale de cette thèse est l'application elle-même. En effet, le SDRA n'est pas encore bien étudié dans la détection de ses caractéristiques ni dans la prévention de son apparition. Cette thèse a pour but de prédire l'occurrence du SDRA en temps réel en utilisant uniquement des signaux physiologiques, sans se référer aux données cliniques généralement utilisées. Afin de prédire l'apparition du SDRA, deux modèles de base basés sur la fusion d'informations ont été proposés dans cette thèse. Le premier modèle est basé sur la détection d'anomalies qui permet la détection en temps réel des déviations par rapport à un état stable prédéfini et la génération de décisions individuelles à partir de chaque signal. Ensuite, ces décisions sont combinées en utilisant l'une des techniques de fusion de décisions linéaires/non linéaires proposées. Le modèle a été testé sur deux types des sujets ventilés invasifs et non invasifs. Le SDRA a été détecté au début de son développement en considérant la combinaison des quatre signaux par la fusion basée sur la régression à noyau sur les deux types de sujets ventilés.

Ensuite, la section B.3 propose un modèle de fusion de caractéristiques basé sur la théorie des fonctions de croyance pour un problème de prévention binaire ainsi que multi-classes. Ainsi, plusieurs paramètres linéaires et non linéaires ont été extraits des signaux et les paramètres les plus pertinents ont été sélectionnés. Ensuite, les informations fournies à partir des paramètres sont modélisées par une fonction de masse en utilisant les distributions de probabilité de chaque paramètre selon chaque état. Ces fonctions de masse sont ainsi affaiblies et combinées. La performance de la prévention binaire a été améliorée par rapport à celle de la section B.2. Dans le problème de prévention multi-classes, le modèle a privilégié l'une des deux options suivantes: soit avoir une sensibilité pour tous les groupes du SDRA, soit une sensibilité pour les groupes de SDRA sévères et non-SDRA.

La section B.4 a été proposée pour améliorer le modèle binaire précèdent ainsi que les résultats. Les fonctions de masse sont construites à partir d'une

phase d'apprentissage par régression à noyau. De plus, nous avons proposé l'extraction d'un nouvel ensemble de paramètres à partir de l'analyse de quantification de récurrence. Ce modèle a amélioré les performances de l'identification précoce du SDRA.

B.5.2 Perspectives

Cette thèse a fourni plusieurs contributions importantes pour la prédiction du SDRA en utilisant des signaux physiologiques. Dans le cadre des futures recherches, nous tenons à étudier les aspects suivants concernant l'amélioration des méthodes proposées.

- *Modélisation des états SDRA*
  Dans le cas de la classification binaire, nous avons considéré la totalité des enregistrements pour les signaux non-SDRA, tandis que nous avons coupé les signaux des sujets SDRA au moment du vrai développement du SDRA. Nous avons ainsi considéré les derniers segments des signaux qui précèdent l'occurrence du SDRA pour la génération du modèle. Cependant, un sujet peut développer un état modéré ou léger du SDRA avant son apparition, mais également peut avoir un état stable. Par conséquent, il est important d'identifier les intervalles de temps où le SDRA est développé et de modéliser l'état du SDRA à l'aide de ces segments comme effectué dans le problème multi-classes.

- *Sélection de paramètres*
  Dans la section B.3, nous avons effectué une sélection séquentielle de paramètres à partir du paramètre ayant le taux d'erreur le plus faible. Les futurs travaux incluront une procédure de sélection plus globale comme celle décrite dans la section B.4.

- *Classification quatre-classes*
  Le problème de la prédiction sur les quatre-classes n'a été présenté qu'à la section B.3. Ce travail nécessite encore des améliorations au niveau de la procédure d'affaiblissement. De plus, le travail sera testé sur le modèle proposé dans la section B.4. Comme indiqué précédemment, un sujet peut développer une forme moins sévère de SDRA avant de développer l'état actuel. Ainsi, il serait intéressant d'étudier l'évolution des états pour détecter la dégradation des signaux à l'aide des modèles de Markov cachés.

Les perspectives déjà citées sont liées directement au travail présenté dans cette thèse. De plus, nous souhaitons élargir notre étude par les propositions suivantes.

- *Extension pour un problème multi-pathologies*
  Le travail présenté considère une seule pathologie qui est le SDRA. Les futurs projets auront pour objectif d'étendre l'actuel travail en effectuant une identification et une surveillance multi-pathologiques. Dans ce cas, le modèle de classification binaire deviendra un problème de classification multi-classes ou un problème multi-label puisqu'un sujet peut développer différentes pathologies à la fois.

- *Deep learning*
  Le corps humain est un système très complexe dans lequel tous les sous-systèmes sont liés les uns aux autres et sont affectés par toute condition soudaine. En outre, Il existe plusieurs signaux qui peuvent être collectés auprès de notre corps pour surveiller les différents aspects physiologiques. Ainsi, il serait intéressant de développer des méthodes de "deep learning" pour la prédiction de pathologies.

- *Acquisition de données*
  Dans ce travail, nous avons utilisé la base de données publique MIMIC II de Physionet. Elle contient des informations physiologiques sur les patients admis à des unités de soins intensifs. Nous souhaitons effectuer une acquisition de données en temps réel pour les signaux physiologiques des patients âgés à domicile afin d'obtenir des cas plus réalistes.

- *Implémentation dans un système de surveillance à domicile*
  Enfin, l'objectif principal de ces études est la mise en œuvre d'un modèle prédictif du SDRA dans les systèmes de surveillance à domicile afin d'alerter le patient et le personnel médical d'un danger possible de développement du SDRA dans un futur proche afin qu'ils puissent intervenir rapidement et efficacement pour empêcher sa sévérité ou même son occurrence.

Bibliography

Abarbanel, Henry (2012). *Analysis of observed chaotic data*. Springer Science & Business Media.

AL-Khalidi, Farah Q et al. (2011). "Respiration rate monitoring methods: A review". In: *Pediatric pulmonology* 46.6, pp. 523–529.

Angelini, Leonardo et al. (2007). "Multiscale analysis of short term heart beat interval, arterial blood pressure, and instantaneous lung volume time series". In: *Artificial intelligence in medicine* 41.3, pp. 237–250.

Argyris, John, Maria Haase, and Gunter Faust (1994). *An exploration of chaos: An Introduction for Natural Scientists and Engineers*. North Holland, Amsterdam.

Aronszajn, Nachman (1950). "Theory of reproducing kernels". In: *Transactions of the American mathematical society* 68.3, pp. 337–404.

Arroliga, Alejandro C et al. (2002). "Incidence of ARDS in an adult population of northeast Ohio." In: *Chest* 121.6, pp. 1972–1976. ISSN: 0012-3692.

Ashbaugh, David G. et al. (1967). "Acute Respiratory Distress in Adults". In: *The Lancet* 290.7511, pp. 319–323. ISSN: 01406736.

Barbera, Roberto, Giuseppe La Rocca, and Massimo Rizzi (2011). "Grid computing technology and the recurrence quantification analysis to predict seizure occurrence in patients affected by drug-resistant epilepsy". In: *Data Driven e-Science*. Springer, pp. 493–506.

Bates, Jason HT (2009). *Lung mechanics: an inverse modeling approach*. Cambridge University Press.

Bauer, Axel et al. (2006). "Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study". In: *The lancet* 367.9523, pp. 1674–1681.

Bellani, Giacomo et al. (2011). "Lung regional metabolic activity and gas volume changes induced by tidal ventilation in patients with acute lung injury". In: *American journal of respiratory and critical care medicine* 183.9, pp. 1193–1199.

Bellani, Giacomo et al. (2016). "Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries". In: *Jama* 315.8, pp. 788–800.

Bellingan, G J (2002). "The pulmonary physician in critical care • 6: The pathogenesis of ALI/ARDS". In: *Thorax* 57.6, pp. 540–546. ISSN: 0040-6376.

Bernard, G R et al. (1994). "The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination." In: *AJRCCM* 149, pp. 818–824. ISSN: 1073-449X.

Übeyli, Elif Derya (2008). "Support vector machines for detection of electrocardiographic changes in partial epileptic patients". In: *Engineering Applications of Artificial Intelligence* 21.8, pp. 1196 –1203. ISSN: 0952-1976.

Bishop, Chris, Christopher M Bishop, et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Borghi-Silva, Audrey et al. (2016). "Analysis of Heart Rate Variability and Cardiovascular Response in the Alveolar Recruitment Manoeuver in Acute Respiratory Distress Syndrome". In: *Journal of Respiratory and CardioVascular Physical Therapy* 3.2, pp. 30–39.

Bossé, É. et al. (2006). "Exploitation of a priori knowledge for information fusion". In: *Information Fusion* 7.2, pp. 161 –175.

Boverman, Gregory and Sahika Genc (2014). "Prediction of Mortality from Respiratory Distress Among Long-Term Mechanically Ventilated Patients". In: *36th Int. Conf. of the IEEE EMBS*. Chicago, Illinois, USA: IEEE, pp. 3464–3467. ISBN: 9781424479290.

Brown, Lisa M et al. (2013). "Comparison of thermodilution measured extravascular lung water with chest radiographic assessment of pulmonary oedema in patients with acute lung injury". In: *Annals of intensive care* 3.1, p. 25.

Brun-Buisson, Christian et al. (2004). "Epidemiology and outcome of acute lung injury in European intensive care units". In: *Intensive care medicine* 30.1, pp. 51–61.

Buzug, Th and G Pfister (1992). "Comparison of algorithms calculating optimal embedding parameters for delay time coordinates". In: *Physica D: Nonlinear Phenomena* 58.1-4, pp. 127–137.

Cai, Qing et al. (2010). "Implementation of a wireless pulse oximeter based on wrist band sensor". In: *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*. Vol. 5. IEEE, pp. 1897–1900.

Camm, AJMM 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.

Cardinal-Fernández, Pablo et al. (2015). "ARDS: lessons learned from the heart". In: *Chest* 147.1, pp. 7–8.

Caser, Eliana B et al. (2014). "Impact of distinct definitions of acute lung injury on its incidence and outcomes in Brazilian ICUs: prospective evaluation of 7,133 patients". In: *Critical care medicine* 42.3, pp. 574–582.

Cellucci, Christopher J, Alfonso M Albano, and PE Rapp (2003). "Comparative study of embedding methods". In: *Physical Review E* 67.6, p. 066210.

Chacko, Jose and Usha Rani (2009). "Alveolar recruitment maneuvers in acute lung injury/acute respiratory distress syndrome". In: *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 13.1, p. 1.

Chan, Alexander M, Nima Ferdosi, and Ravi Narasimhan (2013). "Ambulatory respiratory rate detection using ECG and a triaxial accelerometer". In: *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, pp. 4058–4061.

Chaudhury, Deepak et al. (2017). "A study on clinical profile and outcome of patients with acute respiratory distress syndrome in a tertiary care hospital in North East India". In: *Sepsis* 13, pp. 29–5.

Chawla, Nitesh V et al. (2002). "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16, pp. 321–357.

Chen, Dechang and Xiuzhen Cheng (2001). "An asymptotic analysis of some expert fusion methods". In: *Pattern Recognition Letters* 22.8, pp. 901–904.

Chen, Wei et al. (2010). "Non-invasive blood oxygen saturation monitoring for neonates using reflectance pulse oximeter". In: *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010*. IEEE, pp. 1530–1535.

Chen, Wei et al. (2015). "Incidence and Outcomes of Acute Respiratory Distress Syndrome". In: *Medicine* 94.43, pp. 1–7. ISSN: 0025-7974.

Cheng, Qi et al. (2008). "Fault Detection in Dynamic Systems via Decision Fusion". In: *IEEE Trans. Aerosp. Electron. Sysyt.* 44.1, pp. 227–242. ISSN: 0018-9251.

Chou, Chien-Chang et al. (2011). "An evaluation of airline service quality using the fuzzy weighted SERVQUAL method". In: *Applied Soft Computing* 11.2, pp. 2117–2128.

Chua, Kuang Chua et al. (2010a). "Application of higher order statistics/spectra in biomedical signals — A review". In: *Med. Eng. Phys.* 32.7, pp. 679 –689. ISSN: 1350-4533.

— (2010b). "Application of higher order statistics/spectra in biomedical signals—A review". In: *Medical engineering & physics* 32.7, pp. 679–689.

Clifton, D et al. (2009). "Patient-Specific Biomedical Condition Monitoring for Post-operative Cancer Patients". In: *6th Int. Conf. on Condition Monitoring and Machinery Failure Prevention Technologies*, pp. 424–433. ISBN: 9781618390097.

Cobb, Barry R and Prakash P Shenoy (2003). "A comparison of Bayesian and belief function reasoning". In: *Information Systems Frontiers* 5.4, pp. 345–358.

Copetti, Roberto, Gino Soldati, and Paolo Copetti (2008). "Chest sonography: a useful tool to differentiate acute cardiogenic pulmonary edema from acute respiratory distress syndrome". In: *Cardiovascular Ultrasound* 6.1, p. 16.

Cover, Thomas and Peter Hart (1967). "Nearest neighbor pattern classification". In: *IEEE transactions on information theory* 13.1, pp. 21–27.

Dasarathy, Belur V (1994). *Decision fusion*. Vol. 1994. IEEE Computer Society Press Los Alamitos, CA.

Dementyev, Artem et al. (2013). "Power consumption analysis of Bluetooth Low Energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario". In: *Wireless Symposium (IWS), 2013 IEEE International*. IEEE, pp. 1–4.

Dempster, Arthur P (1967). "Upper and lower probabilities induced by a multivalued mapping". In: *The annals of mathematical statistics* 83, pp. 325–339.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.

Deng, Xinyang et al. (2016). "An improved method to construct basic probability assignment based on the confusion matrix for classification problem". In: *Information Sciences* 340, pp. 250–261.

Deng, Yong (2015). "Generalized evidence theory". In: *Applied Intelligence* 43.3, pp. 530–543.

Deng, Yong, Yang Liu, and Deyun Zhou (2015). "An improved genetic algorithm with initial population strategy for symmetric TSP". In: *Mathematical Problems in Engineering* 2015.

Denoeux, Thierry (2000). "A neural network classifier based on Dempster-Shafer theory". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 30.2, pp. 131–150.

Devijver, Pierre A and Josef Kittler (1982). *Pattern recognition: A statistical approach*. Prentice hall.

Diaz, Javier, Maria Rifqi, and Bernadette Bouchon-Meunier (2006). "A similarity measure between basic belief assignments". In: *Information Fusion, 2006 9th International Conference on*. IEEE, pp. 1–6.

Dudani, Sahibsingh A (1976). "The distance-weighted k-nearest-neighbor rule". In: *IEEE Transactions on Systems, Man, and Cybernetics* 4, pp. 325–327.

Duran, B. S. and P. L. Odell (1974). *Cluster Analysis: A Survey*. Vol. 100. Springer-Verlag Berlin Heidelberg, pp. VI, 140.

Duun, Sune Bro et al. (2010). "A ring-shaped photodiode designed for use in a reflectance pulse oximetry sensor in wireless health monitoring applications". In: *IEEE Sensors Journal* 2.10, pp. 261–268.

Eckmann, J-P, S Oliffson Kamphorst, and David Ruelle (1987). "Recurrence plots of dynamical systems". In: *EPL (Europhysics Letters)* 4.9, p. 973.

Enghard, Philipp et al. (2015). "Simplified lung ultrasound protocol shows excellent prediction of extravascular lung water in ventilated intensive care patients". In: *Critical Care* 19.1, p. 36.

Ennett, Colleen M et al. (2008). "Predicting respiratory instability in the ICU". In: *30th Int. Conf. of the IEEE EMBS*. Vancouver, British Columbia, Canada: IEEE, pp. 2848–2851. ISBN: 9781424418152.

Fan, Eddy et al. (2014). "Physical complications in acute lung injury survivors: a 2-year longitudinal prospective study". In: *Critical care medicine* 42.4, p. 849.

Fein, A. M. et al. (1983). "The risk factors, incidence, and prognosis of ARDS following septicemia". In: *Chest* 83.1, pp. 40–42. ISSN: 00123692.

Fele-Žorž, Gašper et al. (2008). "A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups". In: *Medical & biological engineering & computing* 46.9, pp. 911–922.

Forkan, Abdur Rahim Mohammad and Ibrahim Khalil (2017). "A clinical decision-making mechanism for context-aware and patient-specific remote monitoring systems using the correlations of multiple vital signs". In: *Computer methods and programs in biomedicine* 139, pp. 1–16.

Fox, Dieter et al. (2003). "Bayesian filtering for location estimation". In: *IEEE pervasive computing* 3, pp. 24–33.

Francis, D P et al. (2002). "Physiological basis of fractal complexity properties of heart rate variability". In: *man J. Physiol.* 542, pp. 619–629.

Fraser, Andrew M and Harry L Swinney (1986). "Independent coordinates for strange attractors from mutual information". In: *Physical review A* 33.2, p. 1134.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2001). *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.

Gajic, Ognjen et al. (2005). "Ventilator settings as a risk factor for acute respiratory distress syndrome in mechanically ventilated patients." In: *ICM Journal* 31.7, pp. 922–926. ISSN: 0342-4642.

Gajic, Ognjen et al. (2011). "Early identification of patients at risk of acute lung injury: evaluation of lung injury prediction score in a multicenter cohort study". In: *American journal of respiratory and critical care medicine* 183.4, pp. 462–470.

Galhardo, CEC et al. (2009). "Detrended fluctuation analysis of a systolic blood pressure control loop". In: *New Journal of Physics* 11.10, p. 103005.

Gao, Jianbo and Huaqing Cai (2000). "On the Structures and Quantification of Recurrence Plots". In: *Physics Letters A* 270.1–2, pp. 75–87.

Garvey, Thomas D., John D. Lowrance, and Martin A. Fischler (1981). "An Inference Technique for Integrating Knowledge from Disparate Sources". In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 1*. IJCAI'81. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc., pp. 319–325.

Goldberger, Ary L. et al. (2000). "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals". In: *Circulation* 101.23, e215–e220. ISSN: 0009-7322.

Gong, M N et al. (2010). "Body mass index is associated with the development of acute respiratory distress syndrome." In: *Thorax* 65.1, pp. 44–50. ISSN: 1468-3296.

Gong, Michelle Ng et al. (2005). "Clinical predictors of and mortality in acute respiratory distress syndrome: potential role of red cell transfusion." In: *CCM Journal* 33.6, pp. 1191–1198. ISSN: 0090-3493.

Goss, Christopher H et al. (2003). "Incidence of acute lung injury in the United States." In: *CCM Journal* 31.6, pp. 1607–1611. ISSN: 0090-3493.

Herridge, Margaret S et al. (2011). "Functional disability 5 years after acute respiratory distress syndrome". In: *New England Journal of Medicine* 364.14, pp. 1293–1304.

Ho, Tin Kam, Jonathan J. Hull, and Sargur N. Srihari (1994). "Decision combination in multiple classifier systems". In: *IEEE transactions on pattern analysis and machine intelligence* 16.1, pp. 66–75.

Honeine, Paul, Zineb Noumir, and Cédric Richard (2013). "Multiclass classification machines with the complexity of a single binary classifier". In: *Signal Processing* 93.5, pp. 1013–1026.

Hu, Yong et al. (2003). "A new method of determining the basic belief assignment in DS evidence theory". In: *Machine Learning and Cybernetics, 2003 International Conference on*. Vol. 5. IEEE, pp. 3208–3211.

Huang, Cheng-Yang et al. (2014). "Novel wearable and wireless ring-type pulse oximeter with multi-detectors". In: *Sensors* 14.9, pp. 17586–17599.

Hughes, M et al. (2003). "Acute respiratory distress syndrome: an audit of incidence and outcome in Scottish intensive care units." In: *Anaesthesia* 58.9, pp. 838–845. ISSN: 0003-2409.

Ichikado, Kazuya et al. (2006). "Prediction of prognosis for acute respiratory distress syndrome with thin-section CT: validation in 44 cases". In: *Radiology* 238.1, pp. 321–329.

Institute for Health Metrics and Evaluation (IHME) (2018). *Financing Global Health 2017: Funding Universal Health Coverage and the Unfinished HIV/AIDS Agenda*. Seattle, WA:IHME.

Iscimen, Remzi et al. (2008). "Risk factors for the development of acute lung injury in patients with septic shock: an observational cohort study." In: *CCM Journal* 36.5, pp. 1518–1522. ISSN: 1530-0293.

Jalali, Ali, Daniel J Licht, and Chandrasekhar Nataraj (2013). "Discovering hidden relationships in physiological signals for prediction of Periventricular Leukomalacia". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 7080–7083.

Kahraman, Cengiz, Sezi Cevik Onar, and Basar Oztaysi (2015). "Fuzzy multicriteria decision-making: a literature review". In: *International Journal of Computational Intelligence Systems* 8.4, pp. 637–666.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems". In: *Journal of basic Engineering* 82.1, pp. 35–45.

Karush, William (2014). "Minima of functions of several variables with inequalities as side conditions". In: *Traces and Emergence of Nonlinear Programming*. Springer, pp. 217–245.

Keller, James M, Michael R Gray, and James A Givens (1985). "A fuzzy k-nearest neighbor algorithm". In: *IEEE transactions on systems, man, and cybernetics* 4, pp. 580–585.

Kennel, Matthew B, Reggie Brown, and Henry DI Abarbanel (1992). "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical review A* 45.6, p. 3403.

Khaleghi, Bahador et al. (2013). "Multisensor data fusion: A review of the state-of-the-art". In: *Information fusion* 14.1, pp. 28–44.

Khodor, N et al. (2014). "Kernel based support vector machine for the early detection of syncope during head-up tilt test". In: *Physiological measurement* 35.10, p. 2119.

Klingeberg, T and Marcel Schilling (2012). "Mobile wearable device for long term monitoring of vital signs". In: *Computer methods and programs in biomedicine* 106.2, pp. 89–96.

Knorr, Edwin M, Raymond T Ng, and Vladimir Tucakov (2000). "Distance-based outliers : algorithms and applications". In: *VLDB Journal* 8.3-4, pp. 237–253.

Kraskov, Alexander, Harald Stögbauer, and Peter Grassberger (2004). "Estimating mutual information". In: *Physical review E* 69.6, p. 066138.

Kuhn, Harold W (2014). "Nonlinear programming: a historical view". In: *Traces and Emergence of Nonlinear Programming*. Springer, pp. 393–414.

Kuncheva, Ludmila I (2001). "Using measures of similarity and inclusion for multiple classifier fusion by decision templates". In: *Fuzzy sets and systems* 122.3, pp. 401–407.

Kuncheva, Ludmila I, James C Bezdek, and Robert PW Duin (2001). "Decision templates for multiple classifier fusion: an experimental comparison". In: *Pattern recognition* 34.2, pp. 299–314.

Kuncheva, Ludmila I and Christopher J Whitaker (2003). "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy". In: *Machine learning* 51.2, pp. 181–207.

Ladha, L and T Deepa (2011). "Feature selection methods and algorithms". In: *International journal on computer science and engineering* 3.5, pp. 1787–1797.

Lake, D E et al. (2002). "Sample entropy analysis of neonatal heart rate variability". In: *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 283, pp. 789–797.

Langdon, Ruby et al. (2016). "Use of basis functions within a non-linear autoregressive model of pulmonary mechanics". In: *Biomedical Signal Processing and Control* 27, pp. 44–50.

Li, Kejia and Steve Warren (2012). "A wireless reflectance pulse oximeter with digital baseline control for unfiltered photoplethysmograms". In: *IEEE transactions on biomedical circuits and systems* 6.3, pp. 269–278.

Liu, Huan and Lei Yu (2005). "Toward integrating feature selection algorithms for classification and clustering". In: *IEEE Transactions on knowledge and data engineering* 17.4, pp. 491–502.

Llinas, James and David L Hall (1998). "An introduction to multi-sensor data fusion". In: *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, 1998. ISCAS'98.* Vol. 6. IEEE, pp. 537–540.

Luhr, O R et al. (1999). "Incidence and mortality after acute respiratory failure and acute respiratory distress syndrome in Sweden, Denmark, and Iceland. The ARF Study Group." In: *AJRCCM* 159.6, pp. 1849–1861. ISSN: 1073-449X.

Luo, Ren C, Chih Chia Chang, and Chun Chi Lai (2011). "Multisensor fusion and integration: Theories, applications, and its perspectives". In: *IEEE Sensors Journal* 11.12, pp. 3122–3138.

Mac Sweeney, Rob and Daniel F McAuley (2016). "Acute respiratory distress syndrome". In: *The Lancet* 388.10058, pp. 2416–2430.

Madorno, Matias and Pablo O. Rodriguez (2010). "Non linear respiratory sysytems mechanics simulation of acute respiratory distress syndrome during mechanical ventilation". In: *32th Int. Conf. of the IEEE EMBS*. 1. Buenos Aires, Argentina: IEEE, pp. 232–234. ISBN: 9781424441242.

Mahfouz, Sandy et al. (2014). "Target tracking using machine learning and Kalman filter in wireless sensor networks". In: *IEEE Sensors Journal* 14.10, pp. 3715–3725.

— (2015). "Kernel-based machine learning using radio-fingerprints for localization in wsns". In: *IEEE Transactions on Aerospace and Electronic Systems* 51.2, pp. 1324–1336.

Mangai, Utthara Gosa et al. (2010). "A survey of decision fusion and feature fusion strategies for pattern classification". In: *IETE Technical review* 27.4, pp. 293–307.

Markou, Markos and Sameer Singh (2003a). "Novelty detection: A review - Part 1: Statistical approaches". In: *Signal Processing* 83.12, pp. 2481–2497. ISSN: 01651684.

— (2003b). "Novelty detection: A review - Part 2: Neural network based approaches". In: *Signal Processing* 83.12, pp. 2499–2521. ISSN: 01651684.

Marwan, Norbert et al. (2007). "Recurrence plots for the analysis of complex systems". In: *Physics reports* 438.5–6, pp. 237–329.

Matan, Ofer (1996). "On voting ensembles of classifiers". In: *Proceedings of AAAI-96 workshop on integrating multiple learned models*. Citeseer, pp. 84–88.

Mathers, Colin D et al. (2015). "Causes of international increases in older age life expectancy". In: *The Lancet* 385.9967, pp. 540–548.

McCormick, Garth P (1983). *Nonlinear programming; theory, algorithms, and applications*. Tech. rep.

McKerrow, PJ and SJ Volk (1996). "A systems approach to data fusion". In: *Data Fusion Symposium, 1996. ADFS'96., First Australian*. IEEE, pp. 217–222.

Meng, Tao et al. (2010). "Histology Image Classification Using Supervised Classification and Multimodal Fusion". In: *IEEE ISM*. IEEE, pp. 145–152. ISBN: 978-1-4244-8672-4.

Mercier, David, Éric Lefèvre, and Francois Delmotte (2012). "Belief functions contextual discounting and canonical decompositions". In: *International Journal of Approximate Reasoning* 53.2, pp. 146 –158.

Mercier, David, Frédéric Pichon, and Éric Lefèvre (2016). "Corrigendum to Belief functions contextual discounting and canonical decompositions [International Journal of Approximate Reasoning 53 (2012) 146—158]". In: *International Journal of Approximate Reasoning* 70, pp. 137 –139.

Mercier, David, Benjamin Quost, and Thierry Denœux (2008). "Refined modeling of sensor reliability in the belief function framework using contextual discounting". In: *Information Fusion* 9.2, pp. 246–258.

Millar, PJ et al. (2013). "Isometric handgrip training lowers blood pressure and increases heart rate complexity in medicated hypertensive patients". In: *Scandinavian journal of medicine & science in sports* 23.5, pp. 620–626.

Monk, A. T. and A. H. Compton (1939). "Recurrence phenomena in cosmic-ray intensity". In: *Reviews of Modern Physics* 11.3–4, pp. 173–179.

Moorman, J Randall et al. (2011). "Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring". In: *Physiological measurement* 32.11, p. 1821.

Morbee, M. et al. (2010). "Dempster-Shafer based multi-view occupancy maps". In: *Electronics Letters* 46.5, pp. 341 –343.

Moss, Marc et al. (1996). "The role of chronic alcohol abuse in the development of acute respiratory distress syndrome in adults." In: *JAMA* 275.1, pp. 50–54. ISSN: 0098-7484.

Moss, Marc et al. (2000). "Diabetic patients have a decreased incidence of acute respiratory distress syndrome." In: *CCM Journal* 28.7, pp. 2187–2192. ISSN: 0090-3493.

Mount, LE (1955). "The ventilation flow-resistance and compliance of rat lungs". In: *The Journal of physiology* 127.1, pp. 157–167.

Mourão, André et al. (2014). "Inverse square rank fusion for multimodal search". In: *12th Int. Workshop CBMI*. IEEE, pp. 1–6. ISBN: 9781479939909.

Navarrete-Navarro, Pedro et al. (2000). "Acute respiratory distress syndrome in trauma patients: ICU mortality and prediction factors". In: *Intensive care medicine* 26.11, pp. 1624–1629.

Nemati, Ebrahim, M Jamal Deen, and Tapas Mondal (2012). "A wireless wearable ECG sensor for long-term applications". In: *IEEE Communications Magazine* 50.1.

Nielsen, Jonas et al. (2005). "Lung recruitment maneuver depresses central hemodynamics in patients following cardiac surgery". In: *Intensive care medicine* 31.9, pp. 1189–1194.

Nilsson, Lena, Anders Johansson, and Sigga Kalman (2000). "Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique". In: *Journal of clinical monitoring and computing* 16.4, pp. 309–315.

Organization, World Health and Research for International Tobacco Control (2008). *WHO report on the global tobacco epidemic, 2008: the MPOWER package*. World Health Organization.

Organization, World Health et al. (1990). "Acute respiratory infections in children: case management in small hospitals in developing countries, a manual for doctors and other senior health workers". In:

Ott, Edward (2002). *Chaos in Dynamical Systems*. Cambridge University Press.

Park, Jang-Ho et al. (2015). "Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor". In: *Sensors* 15.9, pp. 23402–23417.

Park, Pauline K et al. (2016). "Incidence, risk factors, and mortality associated with acute respiratory distress syndrome in combat casualty care". In: *Journal of Trauma and Acute Care Surgery* 81.5, S150–S156.

Patroniti, Nicolò, Stefano Isgrò, and Alberto Zanella (2011). "Clinical management of severely hypoxemic patients". In: *Current opinion in critical care* 17.1, pp. 50–56.

Patton, Mike (2015). "US health care costs rise faster than inflation". In: *Forbes*.

Peng, C-K et al. (1995). "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 5.1, pp. 82–87.

Pepe, Paul E et al. (1982). "Clinical predictors of the adult respiratory distress syndrome." In: *AmJ Surgery* 144.1, pp. 124–130. ISSN: 0002-9610.

Perner, Petra (2009). "Concepts for novelty detection and handling based on a case-based reasoning process scheme". In: *Engineering Applications of Artificial Intelligence* 22.1, pp. 86 –91. ISSN: 0952-1976.

Pesenti, Antonio et al. (2016). "Imaging in acute respiratory distress syndrome". In: *Intensive care medicine* 42.5, pp. 686–698.

Pimentel, Marco AF et al. (2014). "A review of novelty detection". In: *Signal Processing* 99, pp. 215–249. ISSN: 01651684.

Prince, Martin et al. (2007). "The protocols for the 10/66 dementia research group population-based research programme". In: *BMC public health* 7.1, p. 165.

Prince, Martin J et al. (2015). "The burden of disease in older people and implications for health policy and practice". In: *The Lancet* 385.9967, pp. 549–562.

Ranieri, V Marco et al. (2012). "Acute Respiratory Distress Syndrome: The Berlin Definition". In: *Jama* 307.23, pp. 2526–2533. ISSN: 0098-7484.

Rattani, A et al. (2007). "Feature level fusion of face and fingerprint biometrics". In: *Proceedings of IEEE Int. Conf. Biometrics: Theory, Appl., Syst.* Pp. 1–6.

Ravishankar, Hariharan et al. (2014). "An early respiratory distress detection method with Markov models". In: *36th Int. Conf. of the IEEE EMBS*. Chicago, Illinois, USA: IEEE, pp. 3438–3441. ISBN: 9781424479290.

Richman, Joshua S and J Randall Moorman (2000). "Physiological time-series analysis using approximate entropy and sample entropy". In: *American Journal of Physiology-Heart and Circulatory Physiology* 278.6, H2039–H2049.

Rivera, Ana Leonor et al. (2016). "Heart rate and systolic blood pressure variability in the time domain in patients with recent and long-standing diabetes mellitus". In: *PloS one* 11.2, e0148378.

Riviello, Elisabeth D et al. (2016). "Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition". In: *American journal of respiratory and critical care medicine* 193.1, pp. 52–59.

Roberts, S.J. (2000). "Extreme value statistics for novelty detection in biomedical data processing". In: *IEE Proceedings - Science, Measurement and Technology*. Vol. 147, pp. 363–367.

Rubenfeld, Gordon D. et al. (2005). "Incidence and Outcomes of Acute Lung Injury". In: *NEJM* 353.16, pp. 1685–1693. ISSN: 0028-4793.

Saeed, Mohammed et al. (2011). "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC- II): A public-access intensive care unit database". In: *CCM Journal* 39.5, pp. 952–960. ISSN: 1530-0293.

Salazar, Eduardo and John H Knowles (1964). "An analysis of pressure-volume characteristics of the lungs". In: *Journal of applied physiology* 19.1, pp. 97–104.

Saunders, Craig, Alexander Gammerman, and Volodya Vovk (1998). "Ridge regression learning algorithm in dual variables". In:

Scholkopf, Bernhard, Ralf Herbrich, and Alex J Smola (2001). "A generalized representer theorem". In: *International conference on computational learning theory*. Springer, pp. 416–426.

Schölkopf, Bernhard et al. (2004). *Kernel methods in computational biology*. MIT press.

Seeley, Eric et al. (2008). "Predictors of mortality in acute lung injury during the era of lung protective ventilation". In: *Thorax* 63.11, pp. 994–998.

Shafer, Glenn (1976). *A mathematical theory of evidence*. Vol. 1. Princeton university press, Princeton.

Shawe-Taylor, John, Nello Cristianini, et al. (2004). *Kernel methods for pattern analysis*. Cambridge university press.

Shu, Yi et al. (2015). "A Pressure sensing system for heart rate monitoring with polymer-based pressure sensors and an anti-interference post processing circuit". In: *Sensors* 15.2, pp. 3224–3235.

Shyamsundar, Murali et al. (2013). "Clinical review: the role of ultrasound in estimating extra-vascular lung water". In: *Critical Care* 17.5, p. 237.

Sigurdsson, MI et al. (2013). "Acute respiratory distress syndrome: nationwide changes in incidence, treatment and mortality over 23 years". In: *Acta anaesthesiologica Scandinavica* 57.1, pp. 37–45.

Singh, Georgene et al. (2014). "Incidence and outcome of acute lung injury and acute respiratory distress syndrome in the surgical intensive care unit." In: *IJCCM* 18.10, pp. 659–665. ISSN: 0972-5229.

Smets, Philippe (1993). "Belief function: the disjunctive rule of combination and the generalized Bayesian theorem". In: *International Journal of Approximate Reasoning* 9.1, pp. 1 –35.

— (2007). "Analyzing the combination of conflicting belief functions". In: *Information fusion* 8.4, pp. 387–412.

Smets, Philippe and Robert Kennes (1994). "The transferable belief model". In: *Artificial Intelligence* 66.2, pp. 191 –234. ISSN: 0004-3702.

Smola, Alex J and Bernhard Schölkopf (2004). "A tutorial on support vector regression". In: *Statistics and computing* 14.3, pp. 199–222.

Sola, J et al. (2006). "SpO2 Sensor Embedded in a Finger Ring: design and implementation". In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, pp. 4295–4298.

Spiegelhalter, David J et al. (1993). "Bayesian analysis in expert systems". In: *Statistical science*, pp. 219–247.

Spodick, David H. et al. (1992). "Operational definition of normal sinus heart rate". In: *The American Journal of Cardiology* 69.14, pp. 1245 –1246. ISSN: 0002-9149.

Sun, Rui, Hong-Zhong Huang, and Qiang Miao (2008). "Improved information fusion approach based on DS evidence theory". In: *Journal of mechanical science and technology* 22.12, pp. 2417–2425.

Takens, Floris (1981). "Detecting strange attractors in turbulence". In: *Dynamical systems and turbulence, Warwick 1980*. Springer, pp. 366–381.

Tavakoli, Maziar, Lorenzo Turicchia, and Rahul Sarpeshkar (2010). "An ultra-low-power pulse oximeter implemented with an energy-efficient transimpedance amplifier". In: *IEEE transactions on biomedical circuits and systems* 4.1, pp. 27–38.

Thakur, Lokendra et al. (2009). "Alcohol consumption and development of acute respiratory distress syndrome: a population-based study." In: *IJERPH* 6.9, pp. 2426–2435. ISSN: 1660-4601.

Thille, Arnaud W et al. (2013). "Chronology of histological lesions in acute respiratory distress syndrome with diffuse alveolar damage: a prospective cohort study of clinical autopsies". In: *The lancet Respiratory medicine* 1.5, pp. 395–401.

Tseng, Kevin C et al. (2014). "Development of a wearable mobile electrocardiogram monitoring system by using novel dry foam electrodes". In: *IEEE systems Journal* 8.3, pp. 900–906.

Turianikova, Zuzana et al. (2011). "The effect of orthostatic stress on multiscale entropy of heart rate and blood pressure". In: *Physiological measurement* 32.9, p. 1425.

United Nations, Department of Economic and Social Affairs, Population Division (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. Working Paper No. ESA/P/WP/248.

Valdovinos, Rosa Maria and Jose Salvador Sanchez (2009). "Combining multiple classifiers with dynamic weighted voting". In: *International Conference on Hybrid Artificial Intelligence Systems*. Springer, pp. 510–516.

Van Drunen, Erwin J et al. (2013). "Expiratory model-based method to monitor ARDS disease state". In: *Biomedical engineering online* 12.1, p. 57.

Vapnik, V (1995). *The Nature of Statitiscal Learning TheorylMj*.

Vapnik, Vladimir (1998). *Statistical learning theory. 1998*. Wiley, New York.

Vapnik, Vladimir Naumovich (1999). "An overview of statistical learning theory". In: *IEEE transactions on neural networks* 10.5, pp. 988–999.

Velardo, Carmelo et al. (2014). "Automatic generation of personalised alert thresholds for patients with COPD". In: *2nd EUSIPCO*. Lisbon, Portugal: IEEE, pp. 1990–1994. ISBN: 9780992862619.

Walkey, Allan J et al. (2012). "Acute respiratory distress syndrome: epidemiology and management approaches". In: *Clinical epidemiology* 4, p. 159.

Ware, LB and MA Matthay (2000). "The acute respiratory distress syndrome". In: *NEJM* 342.18, pp. 1334–1349. ISSN: 0028-4793.

Waske, B. and J. A. Benediktsson (2007). "Fusion of support vector machines for classification of multisensor data". In: *IEEE Trans. Geosci. Remote Sensing* 45.12, pp. 3858–3866.

Weiland, Jeffrey E et al. (1986). "Lung neutrophils in the adult respiratory distress syndrome: clinical and pathophysiologic significance". In: *American Review of Respiratory Disease* 133.2, pp. 218–225.

Wheeler, Arthur P and Gordon R Bernard (2007). "Acute lung injury and the acute respiratory distress syndrome: a clinical review". In: *The Lancet* 369.9572, pp. 1553–1564.

Wilk, Tomasz and Michal Wozniak (2012). "Soft computing methods applied to combination of one-class classifiers". In: *Neurocomputing* 75.1, pp. 185–193.

Wu, H., M. Siegel, and S. Ablay (2003). "Sensor fusion using dempster-shafer theory II: static weighting and Kalman filter-like dynamic weighting". In: *Proceedings of the 20th IEEE Information and Measurement Technology Conference*. TMTC'03, pp. 907–912.

Wu, H. et al. (2002). "Sensor fusion using Dempster-Shafer theory". In: *Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference*. TMTC'02, pp. 7–11.

Xiao, Fuyuan (2019). "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy". In: *Information Fusion* 46, pp. 23–32.

Xiaoming, Jia et al. (2008). "Risk Factors for Acute Respiratory Distress Syndrome in Patients Mechanically Ventilated for Greater Than 48 Hours". In: *Chest* 133.4, pp. 853–861. ISSN: 00123692.

Xu, Lei, Adam Krzyzak, and Ching Y Suen (1992). "Methods of combining multiple classifiers and their applications to handwriting recognition". In: *IEEE transactions on systems, man, and cybernetics* 22.3, pp. 418–435.

Xu, Peida et al. (2013). "A new method to determine basic probability assignment from training data". In: *Knowledge-Based Systems* 46, pp. 69–80.

Xu, Xiaobin et al. (2017). "Data classification using evidence reasoning rule". In: *Knowledge-Based Systems* 116, pp. 144–151.

Yager, Ronald R (1999). "A class of fuzzy measures generated from a Dempster–Shafer belief structure". In: *International Journal of Intelligent Systems* 14.12, pp. 1239–1247.

Yong, Deng et al. (2004). "Combining belief functions based on distance of evidence". In: *Decision support systems* 38.3, pp. 489–493.

Yu, Shun et al. (2013). "Role of diabetes in the development of acute respiratory distress syndrome." In: *CCM Journal* 41.12, pp. 2720–2732. ISSN: 1530-0293.

Zhang, Chenwei et al. (2014). "A new method to determine basic probability assignment using core samples". In: *Knowledge-based systems* 69, pp. 140–149.

Zhang, Xiaoge et al. (2013). "IFSJSP: a novel methodology for the job-shop scheduling problem based on intuitionistic fuzzy sets". In: *International Journal of Production Research* 51.17, pp. 5100–5119.

Zhang, Yang et al. (2018). "Classification of Incomplete Data Based on Evidence Theory and an Extreme Learning Machine in Wireless Sensor Networks". In: *Sensors* 18.4, p. 1046.

Aline TAOUM

Doctorat : Optimisation et Sûreté des Systèmes

Année 2019

Fusion de données multi-capteurs : application à la prévention du SDRA

La population âgée représente actuellement 13% de la population mondiale. Ce taux est estimé d'augmenter très rapidement entrainant le phénomène de vieillissement de la population. Ce phénomène augmente la fréquence d'occurrence de diverses maladies, telles que les pathologies cardiopulmonaires. Dans cette thèse, nous nous intéressons à la pathologie connue sous le nom de syndrome de détresse respiratoire aigüe (SDRA). Il s'agit d'une forme très sévère de défaillance pulmonaire aiguë, consécutive à une altération de la perméabilité capillaire. Le SDRA représente un taux de mortalité élevé. Cette thèse a pour objectif de développer des méthodes de prévention du SDRA en temps réel en utilisant des signaux physiologiques, telles que le rythme cardiaque, le rythme respiratoire, le taux d'oxygène dans le sang et la moyenne de la pression artérielle. Au cours de cette thèse, nous avons développé trois méthodes originales pour la prévention du SDRA. La première méthode est basée sur la détection d'anormalité dans chaque signal physiologique, suivie d'une fusion de décisions, selon différentes approches. La deuxième méthode consiste à extraire des caractéristiques des signaux, à sélectionner les paramètres les plus pertinents et à les fusionner à l'aide de la théorie des fonctions de croyance. Finalement, nous avons proposé un modèle combinant à la fois les méthodes à noyaux et la théorie des fonctions de croyance pour la prévention du SDRA. Nous avons également proposé l'extraction de nouveaux paramètres.

Mots clés : fusion multicapteurs – théorie de Dempster-Shafer – Ridge régression (statistique) – syndrome de détresse respiratoire aiguë de l'adulte.

Multi-sensor Data Fusion: Application to the Prediction of ARDS

Nowadays, the elderly population represents 13% of the world's population. It is estimated that this rate will increase very rapidly leading to the phenomenon of the population ageing. This phenomenon increases the incidence of various diseases, such as cardiopulmonary pathologies. In this thesis, we are interested in the pathology known as Acute Respiratory Distress Syndrome (ARDS). ARDS is a fatal lung condition occurring in critically ill patients representing a high mortality rate. The aim of this thesis is to develop methods for the prediction of ARDS in real-time using physiological signals, such as heart rate, respiratory rate, peripheral arterial oxygen saturation and mean arterial blood pressure. During the course of this thesis, we developed three original methods for the prediction of ARDS. The first method is based on the novelty detection on each physiological signal, followed by different techniques of decision fusion. The second method consists of the extraction of parameters from the signals, selecting the most relevant ones and fusing them using the theory of belief functions. Finally, we proposed a model that combines both kernel methods and belief function theory for the prediction of ARDS. We also proposed the extraction of new parameters to complete the list used in the previous method.

Keywords: multisensor data fusion – Dempster-Shafer theory – Ridge regression (statistics) – respiratory distress syndrome, adult.

Thèse réalisée en partenariat entre :

Université Libanaise

utt UNIVERSITÉ DE TECHNOLOGIE TROYES

Ecole Doctorale "Sciences pour l'Ingénieur"