



**HAL**  
open science

# Classification évidentielle mono- et multi-label : application à la détection de maladies cardio-vasculaires

Mohamed Mroueh Dit Injibar

## ► To cite this version:

Mohamed Mroueh Dit Injibar. Classification évidentielle mono- et multi-label : application à la détection de maladies cardio-vasculaires. Intelligence artificielle [cs.AI]. Université de Technologie de Troyes; Université Libanaise, 2021. Français. NNT : 2021TROY0011 . tel-03808752

**HAL Id: tel-03808752**

**<https://theses.hal.science/tel-03808752v1>**

Submitted on 10 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse  
de doctorat  
de l'UTT

**Mohamed MROUEH DIT INJIBAR**

**Classification évidentielle  
mono- et multi-label :  
application à la détection  
de maladies cardio-vasculaires**

**Champ disciplinaire :  
Sciences pour l'Ingénieur**

2021TROY0011

Année 2021

**Thèse en cotutelle avec l'Université Libanaise – Beyrouth - Liban**



---

---

# 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*

**Mohamed MROUEH DIT INJIBAR**

*le 15 mars 2021*

---

---

**Classification évidentielle mono- et multi-label :  
application à la détection de maladies cardio-vasculaires**

---

---

## JURY

M. P. HONEINE

Mme R. LE BOUQUIN JEANNÈS

Mme S. RIHANA

M. B. QUOST

M. F. ABDALLAH

Mme F. CHEHADE

PROFESSEUR DES UNIVERSITES

PROFESSEURE DES UNIVERSITES

PROFESSEURE

MAITRE DE CONFERENCES - HDR

PROFESSEUR - HDR

MAITRE DE CONFERENCES - HDR

Président

Rapporteur

Rapporteur

Examineur

Directeur de thèse

Directrice de thèse

# Remerciement

Je tiens tout d'abord à remercier chaleureusement mes directeurs de thèse Mme CHEHADE Farah et M. ABDALLAH Fahed pour leur encadrement de mes travaux de recherches. J'exprime particulièrement ma gratitude pour la disponibilité de tous les instants ainsi que le soutien scientifique et moral dont j'ai bénéficié, la confiance qui m'a été accordée et l'environnement de travail confortable qui m'a été fourni.

J'adresse mes remerciements à Mme LE BOUQUIN JEANNES Régine, Professeur des Universités à l'Université de Rennes et à Mme RIHANA Sandy, Professeur à l'Université Saint-Esprit de Kaslik de m'avoir fait l'honneur d'être mes rapporteurs de thèse.

J'exprime ma gratitude aussi à M. QUOST Benjamin, Maître de Conférences à l'Université de Technologie de Compiègne et à M. HONEINE Paul, Professeur des Universités à l'Université de Rouen Normandie, de m'avoir fait l'honneur d'examiner mon travail.

Je remercie l'Université Libanaise (UL) et le Fonds européen de développement régional (FEDER) pour avoir financé mes travaux de recherche.

J'exprime ma gratitude au directeur et aux chargés des affaires administratives de l'école doctorale à l'Université de Technologie de Troyes : Monsieur SAANOUNI Khemais, Mme LECLERCQ Isabelle, Mme DENIS Pascale et Mme KAZARIAN Thérèse. Merci également à Mme ANDRE Bernadette et Mme BANSE Véronique pour l'assistance technique et administrative qu'elles m'ont apportée.

J'exprime ma reconnaissance à mes collègues, mes amis et tous ceux qui m'ont accompagné durant ces trois ans de thèse.

Enfin, je remercie mes chères parents pour leur soutien continue et leur appuie sans qui j'aurais sans doute jamais franchi la première marche vers cet accomplissement.

---

# Table des matières

<b>Table des matières</b>	<b>iii</b>
<b>Liste des figures</b>	<b>vii</b>
<b>Liste des tableaux</b>	<b>ix</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Maladies cardio-vasculaires</b>	<b>3</b>
1.1 Introduction	4
1.2 Maladies cardio-vasculaires	4
1.2.1 Fibrillation Atriale	5
1.2.2 Autres maladies cardiaques	14
1.3 Analyse épidémiologique	19
1.3.1 Facteurs de risque	21
1.3.2 Mauvaises habitudes liées	22
1.3.3 Analyse de morbidité	24
1.4 Présentation de la problématique	30
1.4.1 E-santé	30
1.4.2 Problématique	31
1.4.3 Medical Information Mart for Intensive Care III	32
1.5 Organisation du manuscrit et contributions	32
<b>2 Théorie des fonctions de croyance</b>	<b>35</b>
2.1 Introduction	36
2.2 Utilité de la théorie et définitions	36
2.2.1 Approche probabiliste et ses limitations	36
2.2.2 Approche évidentielle et fonctions de masse	37
2.3 Fonctions indicatives	39
2.3.1 Fonction de croyance	39
2.3.2 Fonction de plausibilité	40
2.3.3 Fonction de communalité	41
2.3.4 Valeurs des fonctions pour les exemples	41
2.4 Méthodes de combinaison de sources	42
2.4.1 Méthode conjonctive	43
2.4.2 Méthode disjonctive	44
2.4.3 Méthode de Dempster	44
2.4.4 Méthode de Yager	45
2.5 Affaiblissement	45
2.6 Prise de décision	47
2.7 Conclusion	48

<b>3</b>	<b>Classification évidentielle pour la prédiction de la FA</b>	<b>49</b>
3.1	Introduction	50
3.2	Base de données médicale MIMIC III	50
3.3	Sélection des sujets et des signaux	52
3.4	Présentation de l'approche	55
3.5	Extraction des caractéristiques	55
3.6	Modélisation des fonctions de masse	58
3.6.1	Introduction à la classification	58
3.6.2	Division de la base de données	60
3.6.3	Attribution des masses	61
3.6.4	Sélection des caractéristiques	62
3.6.5	Combinaison et prise de décision	63
3.7	Résultats et discussions	66
3.7.1	Illustration des performances	66
3.7.2	Impact des paramètres	68
3.7.3	Comparaison aux méthodes de la littérature	70
3.8	Conclusion	71
<b>4</b>	<b>Classification évidentielle avec rejet pour la prédiction de la FA</b>	<b>73</b>
4.1	Introduction	74
4.2	Préambule et notations	74
4.3	Modélisation avancée des fonctions de masse	75
4.4	Méthode de combinaison associative	78
4.4.1	Présentation de la méthode proposée	78
4.4.2	Transformation et décision	79
4.5	Option de rejet	81
4.6	Résultats et discussions	82
4.6.1	Méthode d'attribution des masses	82
4.6.2	Méthode associative de combinaison des masses	83
4.6.3	Option de rejet	84
4.6.4	Validation de l'approche globale	85
4.7	Conclusion	86
<b>5</b>	<b>Classification multi-label évidentielle pour le suivi médical</b>	<b>89</b>
5.1	Introduction	90
5.2	Définitions et formalisation mathématique	90
5.2.1	Classification multi-label	90
5.2.2	Formalisation mathématique	90
5.2.3	Caractéristiques des données multi-label	92
5.2.4	Réglage du seuil de classification	93
5.3	Métriques d'évaluation	93
5.3.1	Les métriques basées sur les instances	93
5.3.2	Les métriques basées sur les labels	95
5.4	Méthodes de classification multi-label	96
5.4.1	Méthodes de transformation	97
5.4.2	Méthodes d'adaptation	102
5.5	Classification multi-label de maladies cardio-vasculaires	105
5.5.1	Construction des données multi-label	105
5.5.2	Fonctions de masse et règles d'association	108
5.6	Résultats et discussion	110
5.7	Conclusion	113

<b>6 Classification multi-label par ensemble pour le suivi médical</b>	<b>115</b>
6.1 Introduction . . . . .	116
6.2 Approche proposée : COLEX . . . . .	116
6.2.1 La méthode RAKEL . . . . .	117
6.2.2 Le coefficient de corrélation binaire $\phi$ . . . . .	117
6.2.3 Facteur de corrélation globale . . . . .	118
6.2.4 Stratégie d'échange : Méthode COLEX . . . . .	119
6.3 Résultats et discussions . . . . .	122
6.3.1 Application de la méthode CoLEx à la base de données médicale . . . . .	122
6.3.2 Illustration des performances de la méthode proposée sur des bases de données réelles multi-label . . . . .	123
6.3.3 Efficacité de la stratégie d'échange . . . . .	126
6.3.4 Impact des paramètres $k$ et $n$ . . . . .	127
6.3.5 Comparaison avec d'autres méthodes . . . . .	128
6.4 Conclusion . . . . .	132
<b>Conclusion générale et perspectives</b>	<b>135</b>
<b>Bibliographie</b>	<b>148</b>



# Liste des figures

1.1	Différence entre une artère normale et une artère rétrécie [1]. . . . .	4
1.2	Taux d'incidence moyen de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans les pays à haut revenu et à revenu intermédiaire [2]. . . . .	6
1.3	Taux d'incidence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu [2]. . . . .	8
1.4	Taux d'incidence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu avec répartition des sexes [2]. . . . .	8
1.5	Taux d'incidence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à revenu intermédiaire [2]. . . . .	9
1.6	Taux d'incidence standardisé selon l'âge de la fibrillation atriale de 1990 à 2017 dans quelques pays à revenu intermédiaire avec répartition des sexe [2]. . . . .	10
1.7	Taux de prévalence moyen standardisé selon l'âge de la fibrillation atriale de 1990 à 2017 dans les pays à haut revenu et à revenu intermédiaire [2]. . . . .	10
1.8	Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu [2]. . . . .	12
1.9	Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu avec répartition des sexe [2]. . . . .	12
1.10	Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à revenu intermédiaire [2]. . . . .	13
1.11	Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à revenu intermédiaire avec répartition des sexes [2]. . . . .	14
1.12	Répartition des décès par maladies cardio-vasculaires standardisée selon l'âge par 100000 habitants en 2000. Source : "www.ouronlinedata.org" . . . . .	20
1.13	Répartition des décès par maladies cardio-vasculaires standardisée selon l'âge par 100000 habitants en 2017. Source : "www.ouronlinedata.org" . . . . .	20
1.14	Taux de mortalité standardisé selon l'âge des maladies cardio-vasculaires dans le monde de 1990 à 2017 pour 100000 habitants. Source de données : [3] . . . . .	21
1.15	Pourcentage de fumeurs quotidiens réguliers dans la population de 15 ans et plus en 2017. Source : Site officiel de l'Organisation Mondiale de la Santé (OMS). . . . .	23
1.16	Prévalence de l'activité physique insuffisante chez les adultes en Europe en 2016. Source : Site officiel de l'Organisation Mondiale de la Santé (OMS). . . . .	24
1.17	Incidence des maladies cardio-vasculaire en 2017 [2]. . . . .	25
1.18	Prévalence des maladies cardio-vasculaire en 2017 [2]. . . . .	25
1.19	Incidence standardisée selon l'âge de la cardiopathie ischémique en France de 1990 à 2017 [2]. . . . .	26
1.20	Incidence standardisée selon l'âge de la cardiopathie ischémique au Liban de 1990 à 2017 [2]. . . . .	27
1.21	Incidence de l'accident vasculaire cérébral en 2017 [2]. . . . .	27
1.22	Incidence standardisée selon l'âge de l'accident vasculaire cérébral en France de 1990 à 2017 [2]. . . . .	28
1.23	Incidence standardisée selon l'âge de l'accident vasculaire cérébral au Liban de 1990 à 2017 [2]. . . . .	28

1.24	Incidence des maladies vasculaires périphériques en 2017 [2]. . . . .	29
1.25	Incidence standardisée selon l'âge des maladies vasculaires périphériques en France de 1990 à 2017 [2]. . . . .	29
1.26	Incidence standardisée selon l'âge des maladies vasculaires périphériques au Liban de 1990 à 2017 [2]. . . . .	30
2.1	Arbre de probabilité du jeu de l'exemple 2.2.1. . . . .	37
2.2	Organigramme de l'héritage génétique (a) et arbre de probabilité (b) de l'exemple 2.2.2. . . . .	38
3.1	Extrait de la table "PATIENTS" de la base de données clinique. . . . .	51
3.2	Extrait de la table "ADMISSIONS" de la base de données clinique. . . . .	52
3.3	Extrait de signaux physiologiques de la base de données cartographique. . . . .	52
3.4	Histogrammes des durées de séjour pour les patients avec et sans FA. . . . .	53
3.5	Statistiques de disponibilité des signaux physiologiques. . . . .	54
3.6	Processus de prédiction de la FA. . . . .	55
3.7	Répartition des fenêtres selon les bases d'apprentissage et de test. . . . .	60
3.8	Exemple de distribution des masses. . . . .	63
3.9	Illustration d'une distribution de masse. . . . .	63
3.10	Illustration de la combinaison de YAGER. . . . .	65
3.11	Illustration de la méthode de combinaison de YAGER. . . . .	66
3.12	Précision moyenne sur les 10 blocs de la base d'apprentissage en fonction de $\gamma$ pour $\beta = 1$ and $N = 60$ min. . . . .	69
4.1	Illustration de la méthode de transformation proposée. . . . .	80
4.2	Solutions obtenues sur la base d'apprentissage en faisant varier $\epsilon$ . . . . .	85
4.3	Front de Pareto pour $\epsilon$ variable de 0 à 1. . . . .	85
5.1	Exemple de données multi-label. . . . .	91
5.2	Illustration de la méthode "Binary Relevance". . . . .	97
5.3	Illustration de la méthode des chaînes de classifieurs. . . . .	99
5.4	Tableau regroupant les labels accumulés par chaque maladie. . . . .	107
5.5	Valeurs moyennes du "Subset Accuracy" pour $0 \leq \text{Supp}_{\min} \leq 0.1$ en fonction de $\gamma$ . . . . .	111
6.1	Graphe des proportions d'instances et de caractéristiques restantes en variant $\xi$ . . . . .	123
6.2	Comparaison entre RAKEL et COLEX pour différentes valeurs de $n$ avec $k = 3$ pour la base de données Genbase. . . . .	127
6.3	Comparaison entre RAKEL et COLEX pour différentes valeurs de $n$ avec $k = 10$ pour la base de données Genbase. . . . .	128
6.4	Comparaison entre RAKEL et COLEX pour différentes valeurs de $n$ avec $k = 2$ pour la base de données Yeast. . . . .	129
6.5	Comparaison entre RAKEL et COLEX pour différentes valeurs de $k$ avec $n = 7$ pour la base de données Yeast. . . . .	130

# Liste des tableaux

1.1	Incidence brute de la fibrillation atriale en 2017 dans les pays à haut revenu [2]. . . . .	7
1.2	Incidence brute de la fibrillation atriale en 2017 dans les pays à revenu intermédiaire [2]. . . . .	9
1.3	Prévalence brute de la fibrillation atriale en 2017 dans les pays à haut revenu [2]. . . . .	11
1.4	Prévalence brute de la fibrillation atriale en 2017 dans les pays à revenu intermédiaire [2]. . . . .	13
2.1	Tableau récapitulatif des valeurs des fonctions indicatives pour l'exemple 2.2.1. . . . .	42
2.2	Tableau récapitulatif des valeurs des fonctions indicatives pour l'exemple 2.2.2. . . . .	42
2.3	Modélisation évidentielle de l'exemple. . . . .	43
2.4	Fonctions de masse issues de différentes méthodes de combinaison. . . . .	43
3.1	Performances de la méthode pour des valeurs autour $\gamma^* = 0.91$ sur les bases de test et d'apprentissage. . . . .	67
3.2	AUC pour différentes combinaisons de $\alpha$ et $N$ . . . . .	68
3.3	Performance de la méthode proposée sur la base de test pour $N=60$ min. . . . .	69
3.4	Performances de la méthode proposée pour différentes combinaisons de $\alpha$ et $\beta$ sur la base d'apprentissage pour $N=60$ min. . . . .	70
3.5	Comparaison aux méthodes classiques de classification. . . . .	71
3.6	Comparaison à des méthodes de prédiction basées sur l'ECG dans la littérature. . . . .	71
4.1	Précision moyenne sur les 10 blocs sur la base d'apprentissage en fonction de $v$ pour $\phi = +\infty$ et $\rho = 0$ . . . . .	82
4.2	Précision moyenne sur les 10 blocs sur la base d'apprentissage en fonction de $\phi$ et $\rho$ pour $v = v^* = 100$ . . . . .	83
4.3	Performances sur la base de test pour les valeurs optimales des paramètres obtenues à partir de la base d'apprentissage. . . . .	83
4.4	Impact de l'option pour plusieurs valeurs de $\epsilon$ sur la base de test. . . . .	84
4.5	Évaluation des contributions du chapitre sur la base de test . . . . .	86
4.6	Comparaison à des méthodes de prédiction basées sur l'ECG dans la littérature. . . . .	86
5.1	Récapitulatif des notations. . . . .	92
5.2	Numérotation des labels. . . . .	106
5.3	Statistiques de la base de données multi-label issue de MIMIC III. . . . .	111
5.4	Métriques de performances pour les valeurs optimales de $\gamma$ et $\text{Supp}_{\min}$ . . . . .	112
5.5	Comparaison des méthodes classiques de classification multi-label avec la méthode proposée. . . . .	112
5.6	Comparaison des méthodes classiques de classification multi-label avec la méthode proposée (suite). . . . .	113
6.1	Tableau comparatif des améliorations apportées par la méthode CoLEX sur RAKEL. . . . .	124
6.2	Caractéristiques des bases de données multi-label du monde réel (Voir Section 5.2.3). . . . .	125
6.3	Comparaison entre RAKEL et COLEX pour la base de données Genbase $k = 3$ et $n = 9$ . . . . .	125

---

6.4	Comparaison entre RAKEL et COLEX pour la base de données Genbase $k = 3$ et $n = 18$ .	125
6.5	Comparaison entre RAKEL et COLEX pour la base de données Yeast $k = 2$ et $n = 7$ . . .	126
6.6	Comparaison entre RAKEL et COLEX pour la base de données Yeast $k = 5$ et $n = 3$ . . .	126
6.7	Comparaison de la méthode exacte à la méthode proposée avec la stratégie d'échange.	126
6.8	Méthodes multi-label utilisées dans la partie expérimentale. . . . .	130
6.9	Comparaison des performances entre les méthodes EME et CoLEx. . . . .	131
6.10	Comparaison entre CoLEx et d'autres méthodes de l'état de l'art (1). . . . .	131
6.11	Comparaison entre CoLEx et d'autres méthodes de l'état de l'art (2). . . . .	132

# Introduction générale

De nos jours, l'apprentissage automatique s'impose dans une grande variété de domaines d'application. Avec la généralisation du "Big Data", son utilisation s'est tellement répandue qu'il est aujourd'hui implémenté dans la quasi-totalité des infrastructures urbaines de la vie quotidienne. Le domaine médical ne faisant pas exception, le concept de la "E-santé" est devenu de plus en plus performant et repose toujours plus sur des systèmes de santé tirant partie des prouesses de l'apprentissage automatique. Alors, de plus en plus d'appareils médicaux portables, comme les capteurs mesurant les signaux physiologiques, sont développés et ces signaux deviennent donc plus facilement mesurables dans un environnement quotidien.

Dans cette thèse, on s'intéresse au risque de malaise cardiaque lié à des maladies cardiovasculaires (MCV). En effet, les MCV sont la première cause de mortalité dans le monde. Chaque année, ce sont environ dix-huit millions de personnes qui décèdent dans le monde, suite à des MCV. Pour cette raison, les recherches sur la prévention, la prédiction, le diagnostic et le traitement des complications liées à des MCV restent des thèmes d'importance majeure dans le domaine de la recherche. Cette thèse s'intéresse particulièrement au développement méthodologique d'un système de suivi médical automatique qui a la capacité de surveiller et d'anticiper la progression des MCV chez des patients en dehors du cadre hospitalier ou clinique sans pour autant leur imposer des actions recommandées ou une invasivité potentiellement gênante pour eux. Ce système repose entièrement sur la mesure de signaux physiologiques en temps réel et leur analyse. Le problème de suivi et d'anticipation est défini comme un problème de classification prenant les signaux physiologiques en entrée et rendant une décision sur l'état du patient en sortie. L'objectif de cette thèse est ainsi de proposer des approches originales de classification adaptées au contexte médical, avec des données manquantes, hétérogènes et imprécises. Dans la première partie de cette thèse, on développe des approches de classification mono-label permettant la prédiction d'épisode de fibrillation atriale (FA) qui est une forme d'arythmie cardiaque. Ces approches sont basées sur la théorie des fonctions de croyance qui débouche sur une modélisation évidentielle permettant de faire de l'apprentissage automatique d'un classifieur. Dans la deuxième partie de cette thèse, on adopte une approche de diagnostic multi-pathologique en modélisant le système par un problème de classification multi-label.

Le premier chapitre présente une analyse statistique et épidémiologique des MCV, tout en introduisant la base de données médicale qui sera utilisée dans les travaux de recherche de cette thèse. Le deuxième chapitre introduit la théorie des fonctions de croyance appelée aussi théorie de Dempster-Shafer. Le troisième chapitre aborde la première partie de la thèse en présentant une méthode évidentielle proposée pour la prédiction d'un épisode de la FA. Le quatrième chapitre apporte des contributions au niveau de cette méthode évidentielle en proposant une nouvelle modélisation de l'information, une nouvelle approche de la fusion des données et suggère une option de rejet. Ensuite, le cinquième chapitre aborde la seconde partie de la thèse en introduisant la notion de classification multi-label et en l'appliquant pour le diagnostic multi-pathologique. Le dernier chapitre présente une contribution théorique pour la classification multi-label en exploitant les corrélations entre les différentes pathologies. La méthode développée sera ensuite utilisée pour notre problème de diagnostic.



# Chapitre 1

## Maladies cardio-vasculaires

### Sommaire

---

<b>1.1 Introduction</b> . . . . .	<b>4</b>
<b>1.2 Maladies cardio-vasculaires</b> . . . . .	<b>4</b>
1.2.1 Fibrillation Atriale . . . . .	5
1.2.2 Autres maladies cardiaques . . . . .	14
<b>1.3 Analyse épidémiologique</b> . . . . .	<b>19</b>
1.3.1 Facteurs de risque . . . . .	21
1.3.2 Mauvaises habitudes liées . . . . .	22
1.3.3 Analyse de morbidité . . . . .	24
<b>1.4 Présentation de la problématique</b> . . . . .	<b>30</b>
1.4.1 E-santé . . . . .	30
1.4.2 Problématique . . . . .	31
1.4.3 Medical Information Mart for Intensive Care III . . . . .	32
<b>1.5 Organisation du manuscrit et contributions</b> . . . . .	<b>32</b>

---

## 1.1 Introduction

Ce chapitre aborde une introduction générale aux maladies cardio-vasculaires qui sont au coeur de la recherche menée par les différentes contributions apportées dans le cadre de ce projet de thèse. L'objectif principal du projet autour duquel s'articule la thèse est le développement d'un système de surveillance médicale non-invasif pour les personnes atteintes de complications cardiaques. Les travaux menés au cours de cette thèse se focalisent essentiellement sur la prédiction et le diagnostic des maladies cardio-vasculaires en se basant uniquement sur des signaux physiologiques mesurables de manière non-invasive. Ainsi, le chapitre commence par une définition des maladies cardio-vasculaires. Il s'en suit une présentation de plusieurs types de maladies cardio-vasculaires en détaillant le mécanisme, les facteurs de risques et les conséquences de chacun d'eux. Ensuite, une analyse épidémiologique est présentée sur les maladies cardio-vasculaires. La problématique de la thèse sera ensuite détaillée en s'appuyant sur les ressources utilisées au cours des travaux de recherche. Une base de données publique et largement accessible y sera notamment décrite. Enfin, l'organisation du manuscrit sera exposée avec la liste des contributions apportées dans le cadre de cette thèse.

## 1.2 Maladies cardio-vasculaires

Les maladies cardio-vasculaires (MCV) sont généralement décrites comme des conditions qui impliquent un rétrécissement ou un blocage (thrombose) des vaisseaux sanguins pouvant entraîner une cardiopathie ischémique (CI) (infarctus du myocarde, angine de poitrine, etc.) ou un accident vasculaire cérébral (AVC). Cela empêche le sang d'atteindre le cœur ou le cerveau. La raison la plus courante est une accumulation de dépôts graisseux sur les parois internes des vaisseaux sanguins qui forment le caillot bloquant les artères. Les AVC peuvent être causés par des caillots sanguins ou des saignements d'un vaisseau sanguin dans le cerveau [4]. La FIGURE 1.1 illustre la différence entre une artère normale et une artère rétrécie. Dans cette section, les maladies cardio-vasculaires étudiées dans le cadre de la thèse sont brièvement présentées. Une maladie en particulier fût le centre d'attention d'une plus grande partie de mes travaux de recherche. Il s'agit de la fibrillation atriale qui sera présentée en premier lieu, ainsi que les résultats d'une analyse épidémiologique poussée. Les autres maladies cardiaques seront présentées sous une forme plus abrégée.

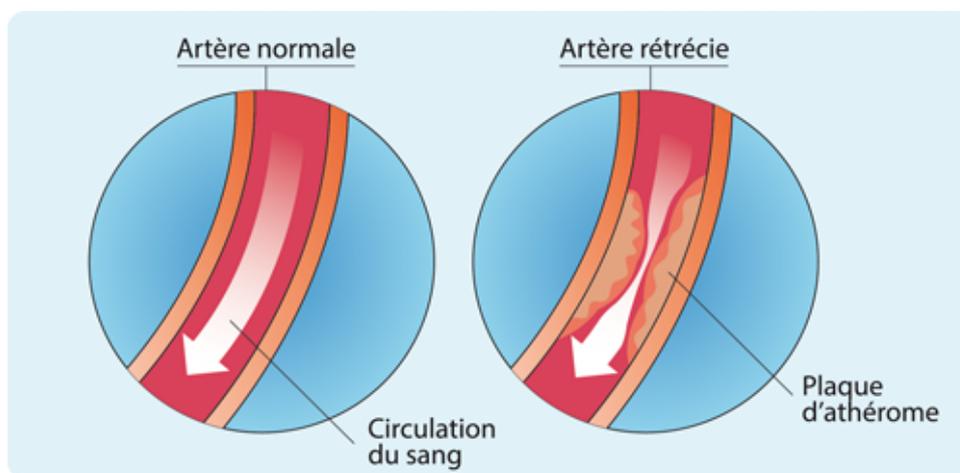


FIGURE 1.1 – Différence entre une artère normale et une artère rétrécie [1].

### 1.2.1 Fibrillation Atriale

#### Mécanisme

La fibrillation atriale (FA) est une arythmie supra-ventriculaire caractérisée par une activité électrique désordonnée avec pour principale conséquence un dysfonctionnement de la mécanique des oreillettes. Sur l'électrocardiogramme (ECG), la FA est caractérisée par le remplacement de l'onde P par des fluctuations électriques rapides ou ondes de fibrillations de différentes amplitudes et formes. Il s'en suit une irrégularité dans l'activité ventriculaire, généralement rapide, en présence d'une conduction auriculoventriculaire normale. La réponse ventriculaire dépend des propriétés électrophysiologiques du nœud auriculoventriculaire, mais aussi du tonus sympathique et parasympathique, ainsi que de la présence d'une voie accessoire ou de l'effet de médicaments potentiellement utilisés. L'irrégularité du cycle n'est pas garantie dans le cas d'une FA. En effet, il est possible d'observer, même en présence d'un bloc auriculoventriculaire, des cycles réguliers mais le cycle ventriculaire sera plus lent. Il est aussi possible d'observer un rythme rapide et irrégulier avec des complexes larges suggérant un bloc de branche fonctionnel ou, dans des cas plus rares, une pré-excitation ventriculaire [5]. La FA survient donc quand les cavités supérieures du cœur, aussi appelées oreillettes ou atriums, perdent la synchronisation des contractions et frémissent. Ainsi, le sang n'est plus pompé efficacement dans le reste de l'organisme, ce qui conduit à une sensation de faiblesse ou de fatigue ou encore des inconforts cardiaques comme des palpitations. Du coup, le sang n'étant plus pompé efficacement dans la cavité, il y reste plus longtemps et on peut observer la formation de caillots. Ces derniers peuvent potentiellement être emportés dans la circulation et si, par malheur, un tel caillot atteint une artère du cerveau, il peut en découler un AVC. Il existe différents types de fibrillation atriale :

- Paroxystique (occasionnelle)  
Une FA qui peut varier de quelques secondes à plusieurs jours, et qui par la suite disparaît d'elle-même.
- Persistante  
Une FA qui dure plus de 7 jours et nécessite l'administration d'un traitement afin d'aider à rétablir le rythme normal du cœur.
- Permanente  
Une FA présente en permanence et il n'est pas possible de retourner à un rythme cardiaque régulier.

La FA passe normalement d'épisodes paroxystiques occasionnels vers des épisodes plus fréquents, puis vers la FA persistante et la FA permanente.

#### Causes

La FA peut être due à plusieurs causes pas toujours identifiables. Elle peut être provoquée par une anomalie innée, qui relève donc de l'ordre de la génétique, ou par une lésion laissée par une crise cardiaque à la structure du cœur. Elle peut être aussi due à un trouble valvulaire. Toutefois, il est possible que la FA survienne spontanément chez des personnes sans antécédents particuliers. En général, indépendamment de la cause, le risque de développement de la FA augmente avec l'âge. Plusieurs facteurs de risques peuvent être à la source d'une FA. Certains sont évitables, d'ordre comportemental, et d'autres ne le sont pas et sont d'ordre génétique. On distingue parmi les facteurs de risque de la FA, l'hypertension artérielle, la cardiopathie, la coronaropathie, les suites d'une chirurgie de pontage coronarien, la cardiomyopathie, le diabète, la myocardite ou la péricardite qui est une inflammation de zones du cœur, soit du myocarde ou du péricarde, qui peut résulter d'une infection, les maladies qui endommagent les valvules cardiaques, l'hyperthyroïdie qui est une hyperactivité de la glande thyroïde, l'embolie pulmonaire résultant de la formation d'un caillot sanguin dans les poumons, la communication interauriculaire originaire d'une malformation septale corrigée durant l'enfance, le tabagisme, l'obésité, la consommation

excessive de caféine, la consommation excessive d'alcool, la prise de drogues, la consommation de certains médicaments, l'apnée du sommeil, le vieillissement et les facteurs génétiques [6].

### Conséquences

Sans traitement, la FA augmente considérablement le risque de subir un AVC ou de développer une insuffisance cardiaque. Les personnes aux prises avec cette maladie courent de trois à cinq fois plus de risque d'avoir un AVC ischémique. Le risque de développer un AVC ischémique est alors décuplé de 3 à 5 fois. Les contractions chaotiques dues à la FA conduisent à une agitation des oreillettes qui ne parviennent plus à faire circuler le sang adéquatement qui finit par se loger dans les petits plis recouvrant le tissu cardiaque. Cette accumulation de sang stagnant peut conduire à la formation de petits caillots qui, une fois expulsés vers le cerveau, peuvent interrompre l'irrigation de celui-ci par le sang et provoquer un AVC ischémique. Statistiquement, après l'âge de 40 ans, un AVC sur quatre est causé par une fibrillation atriale. La FA peut être aussi à la source d'une insuffisance cardiaque qui est une maladie caractérisée par l'incapacité du cœur à pomper assez de sang dans l'ensemble de l'organisme ne répondant plus totalement à ses besoins.

### Incidence

L'incidence de la fibrillation atriale au sein des pays membres de la société européenne de cardiologie ou ESC pour *European Society of Cardiology* ne varie pas significativement dans sa moyenne de 1990 à 2017. La FIGURE 1.2 montre l'évolution du taux d'incidence moyen standardisé selon l'âge dans les pays à haut revenu et à revenu intermédiaire. On remarque que le taux d'incidence dans les pays à haut revenu est plus élevé que dans les pays à revenu intermédiaire. Ceci s'explique, en partie, par l'espérance de vie plus élevée en général dans les pays à haut revenu, ce qui indique une population plus âgée et donc plus à risque aux pathologies cardiaques. La FIGURE 1.2 montre aussi la répartition selon le sexe du taux d'incidence où il est clair que les hommes possèdent un taux d'incidence plus élevé que les femmes que ce soit dans les pays à haut revenu ou dans les pays à revenu intermédiaire.

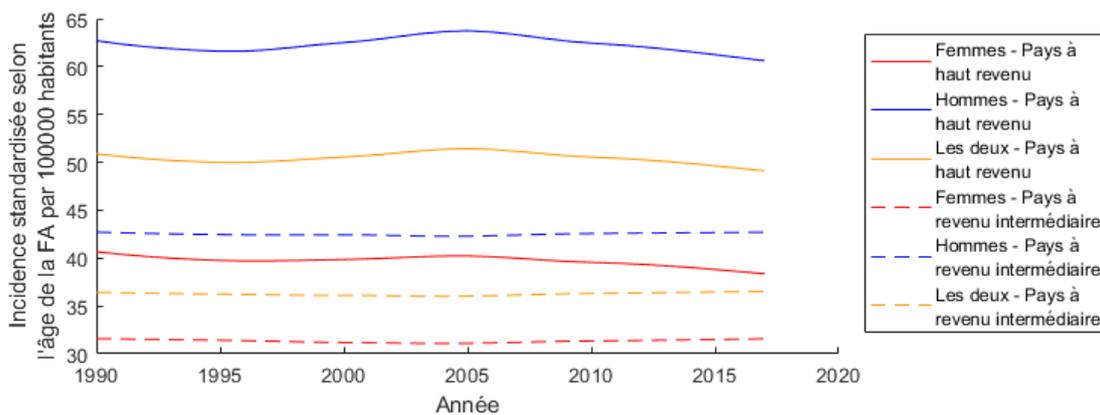


FIGURE 1.2 – Taux d'incidence moyen de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans les pays à haut revenu et à revenu intermédiaire [2].

Le TABLEAU 1.1 montre l'incidence brute (nombre de nouveaux cas) en 2017 dans les différents pays à haut revenu membres de l'ESC ainsi que la répartition selon le sexe des cas. Il s'agit du nombre global d'incidence, sans standardisation en fonction de l'âge. L'Allemagne est le pays qui enregistre le plus haut nombre d'incidence avec près de 91 milles nouveaux cas dont 58% environ d'hommes. Ensuite, nous avons le Royaume-Uni, l'Italie et la France qui enregistrent des incidences similaires entre 60 et 70 milles cas, toujours avec une supériorité des cas pour les hommes. La Chypre, le Luxembourg, la République de Malte et l'Islande sont les pays les moins touchés reportant moins de mille cas. On remarque que, dans sa globalité, les hommes sont plus atteints par

la maladie que les femmes dans les pays à haut revenu avec 55% de cas chez les hommes.

TABLEAU 1.1 – Incidence brute de la fibrillation atriale en 2017 dans les pays à haut revenu [2].

Pays	Incidence	Hommes	Femmes
Allemagne	90801	52351(57.65%)	38450(42.35%)
Royaume-Uni	69562	39620(56.96%)	29942(43.04%)
Italie	62769	35847(57.11%)	26921(42.89%)
France	60203	32458(53.91%)	27745(46.09%)
Espagne	42980	21810(50.74%)	21170(49.26%)
Pologne	35974	19357(53.81%)	16618(46.19%)
Pays-Bas	15574	9252(59.41%)	6322(40.59%)
Suède	12672	7411(58.48%)	5261(41.52%)
Belgique	10206	5482(53.71%)	4724(46.29%)
L'Autriche	10101	5196(51.44%)	4905(48.56%)
Grèce	10019	5531(55.21%)	4489(44.80%)
République Tchèque	9559	5400(56.49%)	4159(43.51%)
Hongrie	8879	4519(50.90%)	4360(49.10%)
le Portugal	7883	4232(53.69%)	3651(46.31%)
Finlande	6467	3611(55.84%)	2855(44.15%)
Suisse	6165	3721(60.36%)	2444(39.64%)
Norvège	5857	3487(59.54%)	2370(40.46%)
Danemark	5839	3305(56.60%)	2534(43.40%)
Israël	4831	2725(56.41%)	2106(43.59%)
Slovaquie	4606	2368(51.41%)	2239(48.61%)
Irlande	3371	2037(60.43%)	1334(39.57%)
Croatie	3070	1789(58.27%)	1281(41.73%)
Lituanie	2493	1206(48.38%)	1287(51.62%)
Slovénie	2180	1233(56.56%)	948(43.49%)
Lettonie	1760	861(48.92%)	899(51.08%)
Estonie	1122	548(48.84%)	573(51.07%)
Chypre	823	468(56.87%)	355(43.13%)
Luxembourg	589	320(54.33%)	269(45.67%)
Malte	452	241(53.32%)	211(46.68%)
Islande	262	160(61.07%)	102(38.93%)
TOTAL	497069	276546(55.64%)	220524(44.36%)

La FIGURE 1.3 montre l'évolution du taux d'incidence de la FA standardisé selon l'âge de 1990 à 2017 pour quelques pays à haut revenu. Le pays le plus touché par la FA est la Suède avec un taux d'incidence qui fluctue faiblement entre 63.13 en 1990 et 63.4 en 1995, puis entame une montée concave qui atteint son pic à 75.89 en 2007 avant de redescendre aussi brutalement et atteindre 63.85 en 2017. La France connaît un décroissement monotone du taux d'incidence de la FA de 55.73 en 1990 à 46.97 en 2017. L'Italie suit une courbe similaire à celle de la France à l'exception d'une remontée brusque de 53.22 en 1995 qui atteint son apogée en 2000 avec une valeur de 57.56. La FIGURE 1.4 montre l'évolution du taux d'incidence de la FA standardisé selon l'âge de 1990 à 2017 pour les mêmes pays à haut revenu en détaillant la répartition du taux de chaque sexe. Les schémas montrent clairement que le taux d'incidence chez les hommes est plus élevé que chez les femmes.

Le TABLEAU 1.2 montre l'incidence brute en 2017 dans les différents pays à revenu intermédiaire membres de l'ESC ainsi que la répartition selon le sexe des cas. La Russie enregistre le plus grand nombre de nouveaux cas parmi les pays à revenu intermédiaire. Ceci peut être expliqué par la taille de la population. Une légère dominance des cas féminins de FA est remarquable dans les 3 pays qui reportent le plus de cas d'incidence, à savoir, la Russie, l'Ukraine et la Turquie. Le Monténégro possède le plus petit nombre d'incidence, environ 450, mais ceci s'explique encore par la taille de la population entière. Malgré le fait que dans la plupart des pays à revenu intermédiaire, on a une légère dominance des cas masculins, sur l'ensemble des incidences, qui représente environ 253 mille cas, une légère majorité est présente chez les femmes, les pays avec le plus de cas ayant complètement dominé la répartition.

La FIGURE 1.5 montre l'évolution du taux d'incidence de la FA standardisé selon l'âge de 1990 à 2017 pour quelques pays à revenu intermédiaire. Le Monténégro est l'un des pays les plus touchés par la FA avec un taux d'incidence standardisé selon l'âge de 44.52 en 2017. Le plus bas taux d'incidence enregistré de 1990 à 2017 est de 43.33 en 2000. Le taux d'incidence au Monténégro

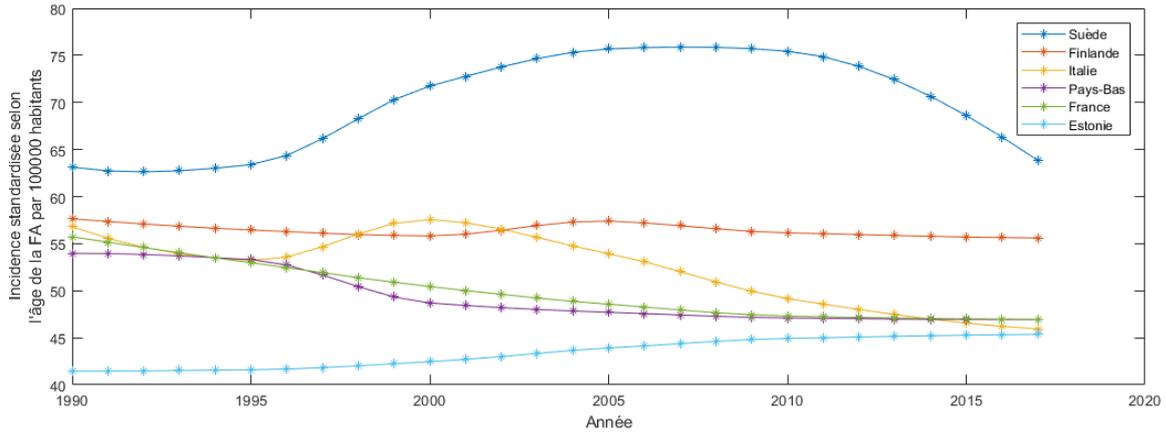


FIGURE 1.3 – Taux d’incidence de la fibrillation atriale standardisé selon l’âge de 1990 à 2017 dans quelques pays à haut revenu [2].

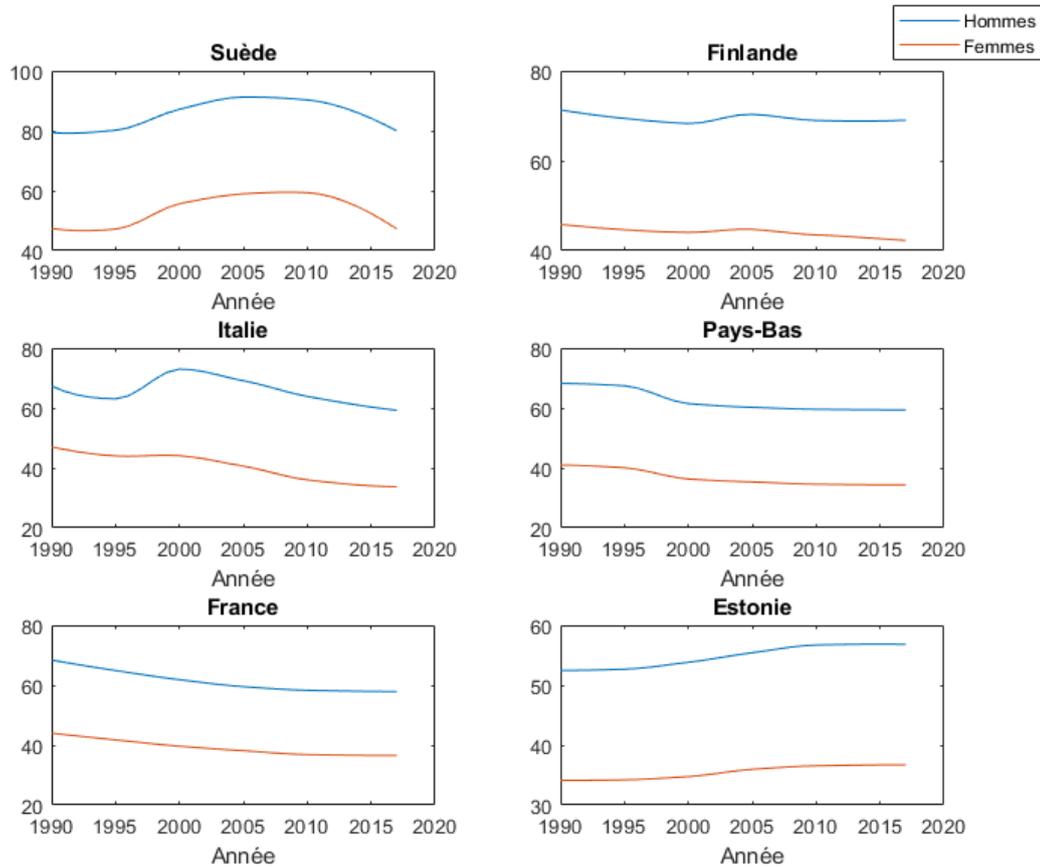


FIGURE 1.4 – Taux d’incidence de la fibrillation atriale standardisé selon l’âge de 1990 à 2017 dans quelques pays à haut revenu avec répartition des sexes [2].

est pratiquement stable. La Roumanie affiche un taux d’incidence de 44.35 en 1990 et connaît une baisse significative entre 2000 et 2005 ce qui la ramène à 39.02 en 2005 puis reste stable jusqu’en 2017. Le Liban et l’Algérie font partie des pays les moins touchés avec des taux pratiquement stables ne dépassant pas 28.

La FIGURE 1.6 montre l’évolution du taux d’incidence de la FA standardisé selon l’âge de 1990 à 2017 pour les mêmes pays à revenu intermédiaire en détaillant la répartition du taux de chaque sexe. On remarque que sur la totalité des pays représentés dans la FIGURE 1.6, les hommes sont

TABEAU 1.2 – Incidence brute de la fibrillation atriale en 2017 dans les pays à revenu intermédiaire [2].

Pays	Incidence	Hommes	Femmes
Russie	100825	46823(46.44%)	54002(53.56%)
Ukraine	34920	15972(45.74%)	18948(54.26%)
Turquie	18981	8481(44.68%)	10500(55.32%)
Roumanie	14585	7553(51.79%)	7032(48.21%)
Egypte	14017	7408(52.85%)	6609(47.15%)
Maroc	8213	4132(50.31%)	4081(49.69%)
Algérie	7991	4109(51.42%)	3882(48.58%)
Kazakhstan	6948	3528(50.78%)	3420(49.22%)
Biélorussie	6925	3396(49.04%)	3529(50.96%)
Bulgarie	6223	3277(52.66%)	2946(47.34%)
Serbie	5875	3177(54.08%)	2698(45.92%)
Azerbaïdjan	3743	2046(54.66%)	1698(45.36%)
Syrie	3251	1706(52.48%)	1545(47.52%)
Tunisie	3050	1505(49.34%)	1545(50.66%)
Bosnie Herzégovine	2724	1495(54.88%)	1228(45.08%)
Géorgie	2465	1234(50.06%)	1231(49.94%)
Moldova	2428	1246(51.32%)	1182(48.68%)
Albanie	1821	1054(57.88%)	767(42.12%)
Arménie	1779	928(52.16%)	851(47.84%)
Kirghizistan	1664	890(53.49%)	774(46.51%)
Liban	1565	699(44.66%)	866(55.34%)
Macédoine du Nord	1421	809(56.93%)	611(43.00%)
Libye	1159	595(51.34%)	563(48.58%)
Monténégro	451	250(55.43%)	201(44.57%)
TOTAL	253024	122313(48.34%)	130709(51.66%)

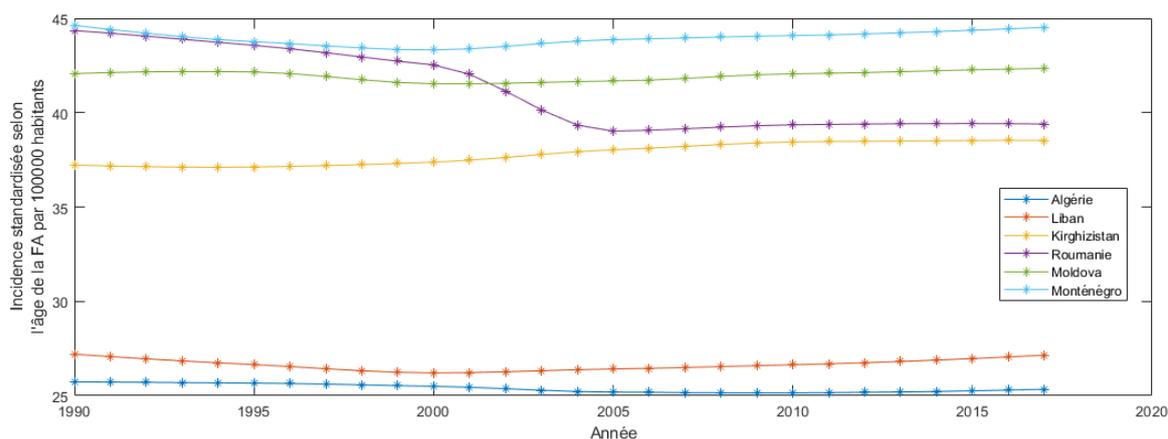


FIGURE 1.5 – Taux d’incidence de la fibrillation atriale standardisé selon l’âge de 1990 à 2017 dans quelques pays à revenu intermédiaire [2].

plus touchés par la FA que les femmes. Cette différence est moins accentuée dans les pays comme le Liban ou l’Algérie où la différence relative entre les taux qui s’obtient en calculant le rapport entre la valeur absolue de la différence des taux et leur moyenne, se situe en moyenne sur les années de 1990 à 2017, au alentour des 3%, alors que pour les autres pays de la figure, elle se situe entre 37% et 44%.

Sur la totalité des pays membre de l’ESC, 750093 cas ont été reportés en 2017 dont 351233 cas chez les femmes et 398859 cas chez les hommes ce qui représente respectivement environ 47% et 53% des cas.

### Prévalence

La FIGURE 1.7 montre l’évolution du taux de prévalence moyen standardisé selon l’âge dans les pays à haut revenu et à revenu intermédiaire. Il est à noter que le taux de prévalence est en général plus élevé dans les pays à haut revenu que ceux à revenu intermédiaire. De plus, les hommes semblent être plus impactés que les femmes dans tous les pays membres de l’ESC dont les don-

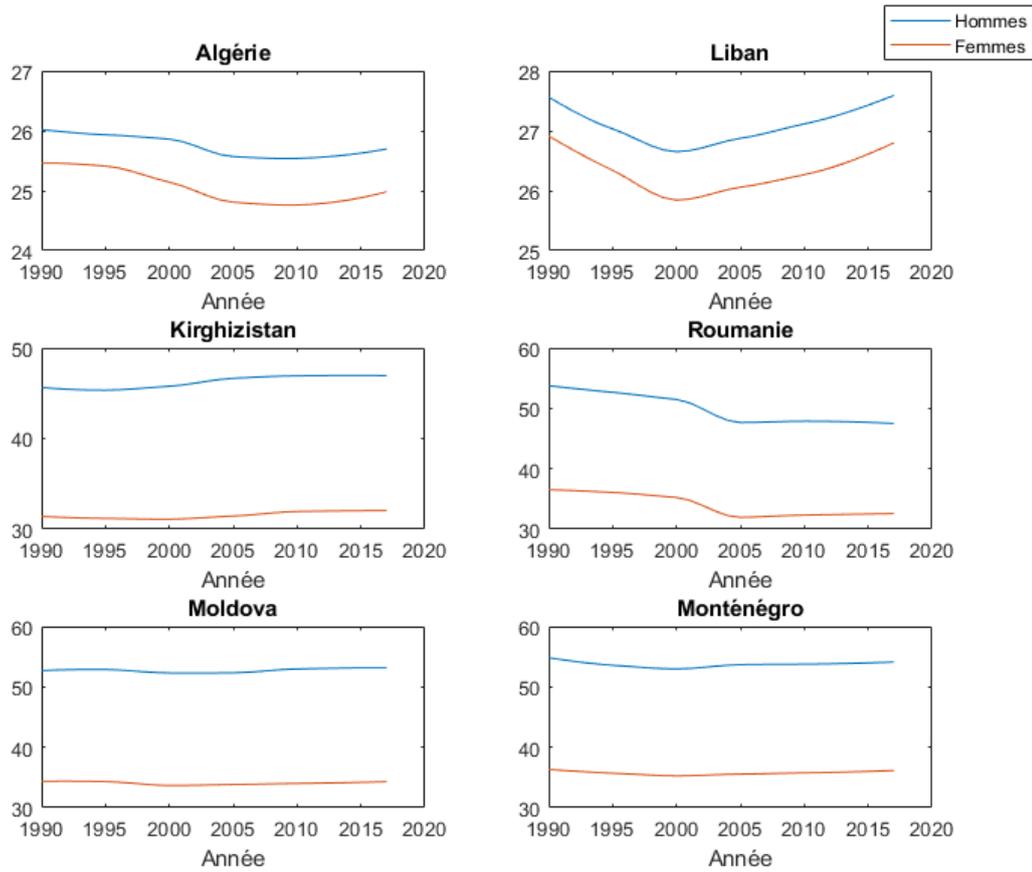


FIGURE 1.6 – Taux d’incidence standardisé selon l’âge de la fibrillation atriale de 1990 à 2017 dans quelques pays à revenu intermédiaire avec répartition des sexe [2].

nées ont été rendues disponibles. D’après cette figure, en moyenne, 618 personnes sur 100000 vivent avec la FA en 2017. Ce taux n’a pas beaucoup changé durant les deux dernières décennies, même si on remarque une hausse légère en 2005 qui affiche un taux de prévalence de 642 sur 100000.

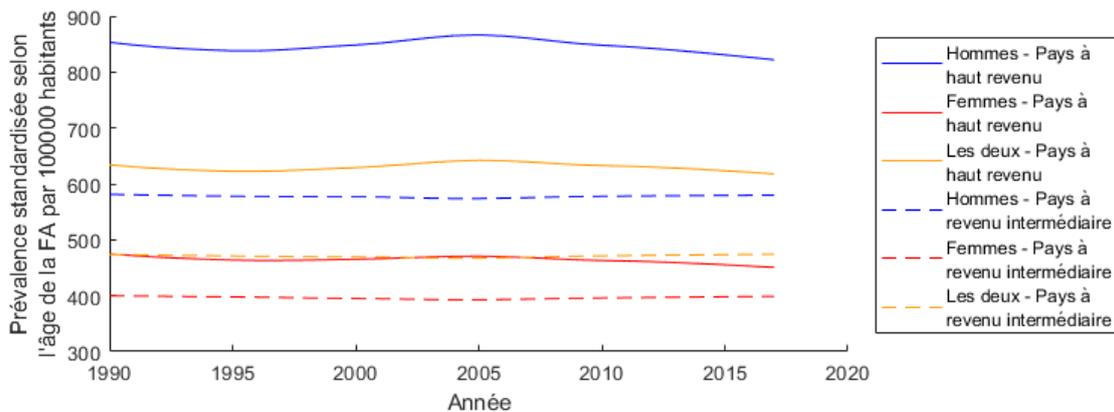


FIGURE 1.7 – Taux de prévalence moyen standardisé selon l’âge de la fibrillation atriale de 1990 à 2017 dans les pays à haut revenu et à revenu intermédiaire [2].

Le TABLEAU 1.3 montre la prévalence brute (nombre de cas actifs) en 2017 dans les différents pays à haut revenu membres de l’ESC ainsi que la répartition selon le sexe des cas. L’Allemagne est le pays qui compte le plus grand nombre de personnes atteintes de FA avec à son compte près

de 1.2 millions d'atteints dont une majorité masculine notable est à relever avec un rapport de 61%. Le Royaume-Uni, l'Italie et la France ont des prévalences supérieures à 800 milles malades avec toujours la même majorité d'hommes. Les prévalences brutes sont directement impactées par la taille de la population et son espérance de vie. Raison pour laquelle le taux de prévalence standardisé selon l'âge est considéré dans ce genre d'analyse. Dans les pays à haut revenu, environ 6.7 millions de cas de FA sont reportés dont une majorité de 57% de patients masculins.

TABEAU 1.3 – Prévalence brute de la fibrillation atriale en 2017 dans les pays à haut revenu [2].

Pays	Incidence	Hommes	Femmes
Allemagne	1201402	733080(61.02%)	468322(38.98%)
Royaume-Uni	934851	565844(60.53%)	369007(39.47%)
Italie	847845	495892(58.49%)	351952(41.51%)
France	820312	461388(56.25%)	358924(43.75%)
Espagne	588301	312625(53.14%)	275676(46.86%)
Pologne	495314	260378(52.57%)	234936(47.43%)
Pays-Bas	197572	122645(62.08%)	74926(37.92%)
Suède	174243	108878(62.49%)	65366(37.51%)
L'Autriche	142603	73975(51.87%)	68628(48.13%)
Grèce	139047	82580(59.39%)	56467(40.61%)
Belgique	137448	77188(56.16%)	60261(43.84%)
République Tchèque	135868	75863(55.84%)	60005(44.16%)
Hongrie	125060	62010(49.58%)	63050(50.42%)
Portugal	102488	58002(56.59%)	44485(43.41%)
Finlande	87422	51077(58.43%)	36345(41.57%)
Suisse	80875	50643(62.62%)	30232(37.38%)
Danemark	74597	44516(59.68%)	30081(40.32%)
Norvège	69017	41803(60.57%)	27214(39.43%)
Israël	61750	36867(59.70%)	24883(40.30%)
Slovaquie	61154	30987(50.67%)	30167(49.33%)
Croatie	43670	24843(56.89%)	18827(43.11%)
Irlande	42228	26695(63.22%)	15533(36.78%)
Lituanie	36884	17321(46.96%)	19563(53.04%)
Slovénie	31271	17084(54.63%)	14187(45.37%)
Lettonie	26232	12434(47.40%)	13798(52.60%)
Estonie	16619	7827(47.10%)	8792(52.90%)
Chypre	10161	5877(57.84%)	4283(42.16%)
Luxembourg	7582	4225(55.73%)	3357(44.27%)
Malte	5583	3068(54.96%)	2514(45.04%)
Islande	3399	2177(64.04%)	1222(35.96%)
TOTAL	6700798	3867792(57.72%)	2833004(42.28%)

La FIGURE 1.8 montre l'évolution du taux de prévalence standardisé selon l'âge de la FA de 1990 à 2017 pour quelques pays à haut revenu. Les pays qui affichent les plus hauts niveaux de prévalence de la FA sont la Suède et le Danemark. Les taux de ces deux pays ont chacun connu une apogée de 2005 à 2010 avec une valeur de 936 en 2008 pour la Suède et de 2002 à 2006 avec une valeur de 747.8 en 2005 pour le Danemark. La France connaît une décroissance continue du taux de prévalence de la FA sur la période qui s'étend de 1990 à 2017 avec une valeur de 676 en 1990 et de 577 en 2017. Le Portugal, en plus d'une décroissance légère continue, a connu une période de forte chute du taux de prévalence de 2005 à 2010 avec une valeur de 526 en 2005 et de 410 en 2010.

La FIGURE 1.9 montre l'évolution du taux de prévalence de la FA standardisé selon l'âge de 1990 à 2017 pour les mêmes pays à haut revenu en détaillant la répartition du taux de chaque sexe. Les pays qui affichent les plus grandes différenciations entre les taux masculins et féminins sont l'Allemagne et la Suède. Les différences sont moins accentuées en Belgique que dans le reste des pays présentés dans la FIGURE 1.9.

Le TABLEAU 1.4 montre la prévalence brute en 2017 dans les différents pays à revenu intermédiaire membres de l'ESC ainsi que la répartition selon le sexe des cas. Avec 1.3 millions de cas de FA en Russie, elle enregistre le plus grand nombre de cas actifs. Contrairement aux pays à haut revenu, les pays à revenu intermédiaire affiche une légère supériorité d'environ 52% pour le genre féminin. C'est notamment le cas au Liban avec une majorité de 56% de cas féminins pour un total d'environ 19 mille personnes atteintes. Dans les pays à revenu intermédiaire, ce sont environ 3.3

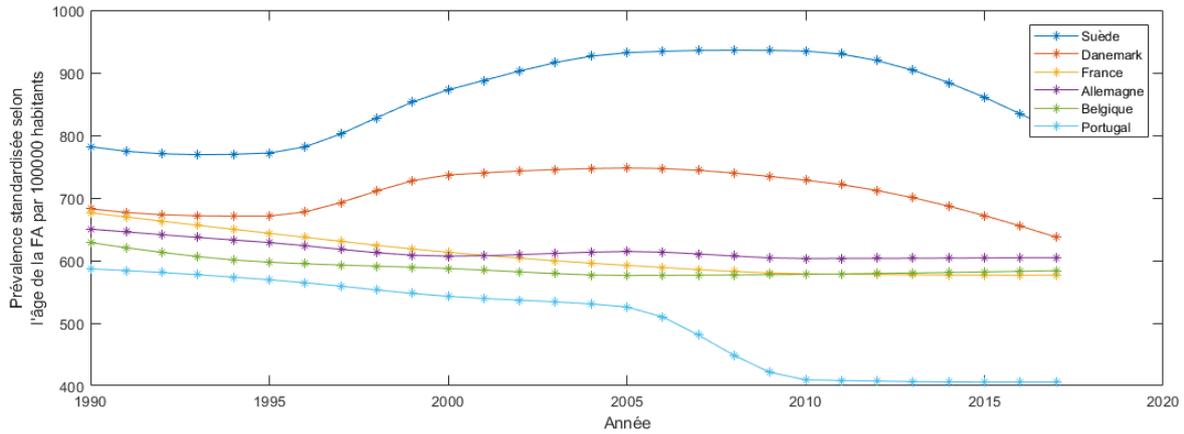


FIGURE 1.8 – Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu [2].

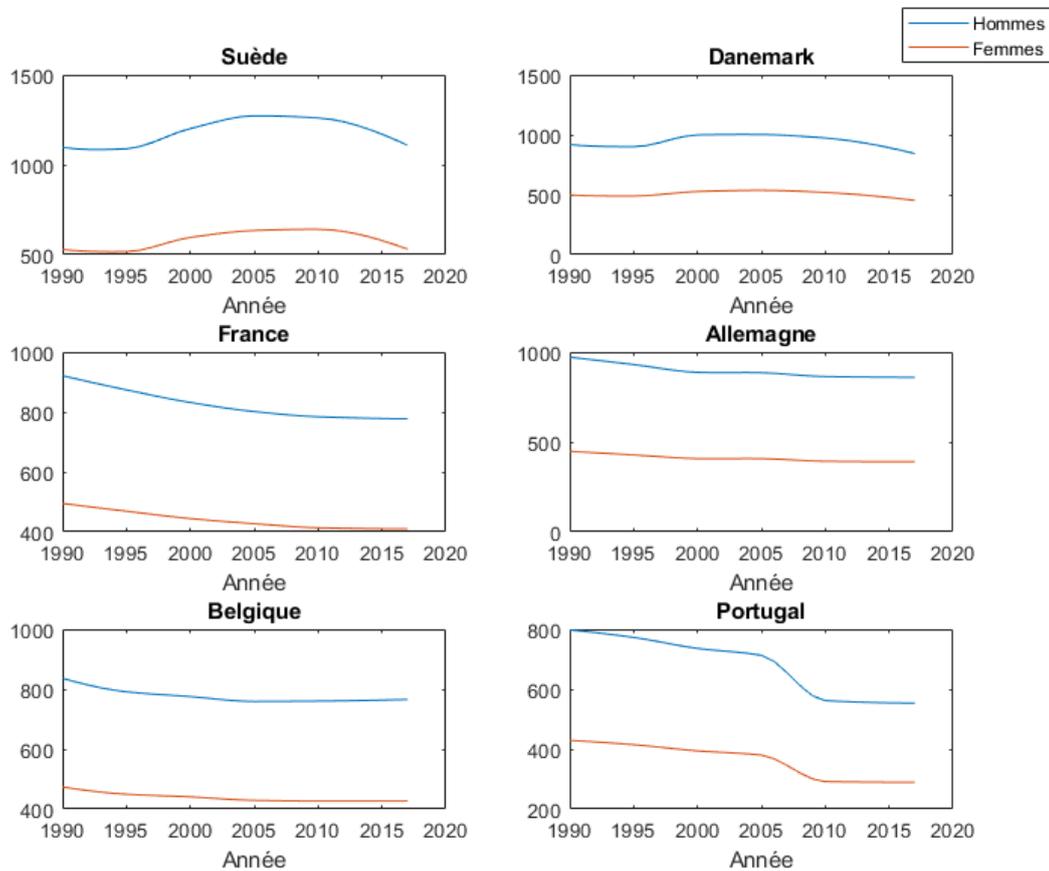


FIGURE 1.9 – Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à haut revenu avec répartition des sexe [2].

millions de personnes qui vivent avec la FA.

La FIGURE 1.10 montre l'évolution du taux de prévalence de la FA standardisé selon l'âge de 1990 à 2017 pour quelques pays à revenu intermédiaire. Sur cette figure, le Liban et l'Algérie sont bien moins concernés par la FA et présentent des courbes de prévalence quasi-stable de 1990 à 2017 avec une valeur moyenne d'environ 326 et 311 respectivement. Le Monténégro et l'Albanie présentent les plus hauts taux de prévalence sans variations significative de 1990 à 2017 avec une valeur moyenne de 586 et de 558 respectivement. La Roumanie et la Serbie quant à elles

TABLEAU 1.4 – Prévalence brute de la fibrillation atriale en 2017 dans les pays à revenu intermédiaire [2].

Pays	Incidence	Hommes	Femmes
Russie	1351142	612871(45.36%)	738271(54.64%)
Ukraine	482686	216550(44.86%)	266136(55.14%)
Turquie	227610	98974(43.48%)	128635(56.52%)
Roumanie	196967	102352(51.96%)	94615(48.04%)
Egypte	153059	79974(52.25%)	73085(47.75%)
Maroc	96434	48379(50.17%)	48055(49.83%)
Algérie	95426	49326(51.69%)	46100(48.31%)
Biélorussie	94228	44613(47.35%)	49615(52.65%)
Bulgarie	88079	46054(52.29%)	42025(47.71%)
Kazakhstan	83664	41975(50.17%)	41689(49.83%)
Serbie	77900	41776(53.63%)	36124(46.37%)
Azerbaïdjan	43656	23760(54.43%)	19896(45.57%)
Syrie	37437	19523(52.15%)	17914(47.85%)
Tunisie	36203	17797(49.16%)	18406(50.84%)
Bosnie Herzégovine	35987	19812(55.05%)	16175(44.95%)
Géorgie	32972	16399(49.74%)	16572(50.26%)
Moldova	32298	16489(51.05%)	15809(48.95%)
Albanie	24207	14175(58.56%)	10032(41.44%)
Arménie	23201	12227(52.70%)	10974(47.30%)
Kirghizistan	19981	10599(53.05%)	9381(46.95%)
Liban	18818	8301(44.11%)	10517(55.89%)
Macédoine du Nord	18637	10580(56.77%)	8057(43.23%)
Libye	13715	7087(51.67%)	6628(48.33%)
Monténégro	6020	3350(55.66%)	2669(44.34%)
TOTAL	3290327	1562944(47.50%)	1727382(52.50%)

connaissent une décroissance brute de 2000 à 2005, ce qui fait passer le taux de 567 à 502 pour la Roumanie et de 516 à 463 pour la Serbie.

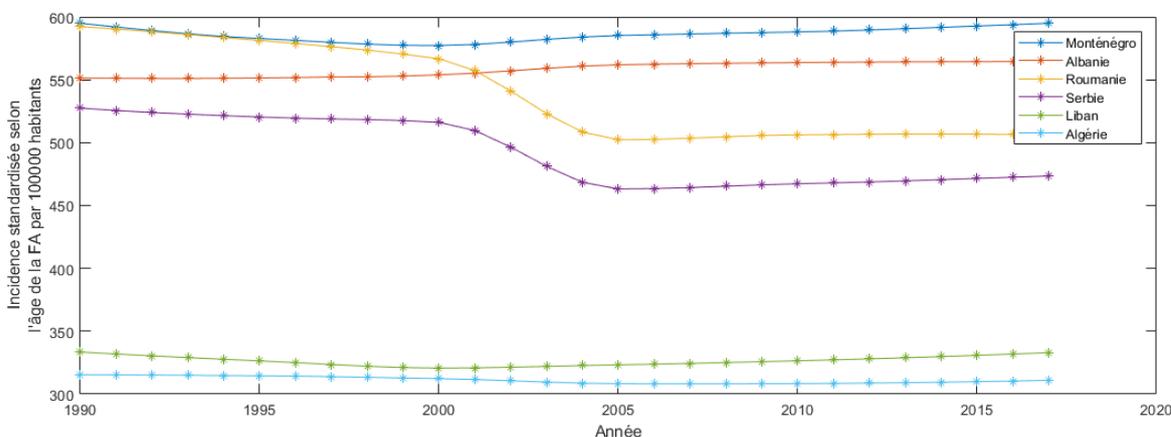


FIGURE 1.10 – Taux de prévalence de la fibrillation atriale standardisé selon l’âge de 1990 à 2017 dans quelques pays à revenu intermédiaire [2].

La FIGURE 1.11 montre l’évolution du taux de prévalence de la FA standardisé selon l’âge de 1990 à 2017 pour les mêmes pays à revenu intermédiaire en détaillant la répartition du taux de chaque sexe. Sur l’ensemble des pays dont les taux de prévalence de la FA sont tracés sur la FIGURE 1.11, les hommes observent des taux plus élevés que chez les femmes. Néanmoins, les différences sont beaucoup moins prononcées dans le cas du Liban et de l’Algérie.

Sur la totalité des pays membres de l’ESC, près de 10 millions d’habitants vivent avec la FA. Ceci fait en moyenne 185 mille habitants par pays. Il est à noter que près de 54% de ces habitants sont des hommes.

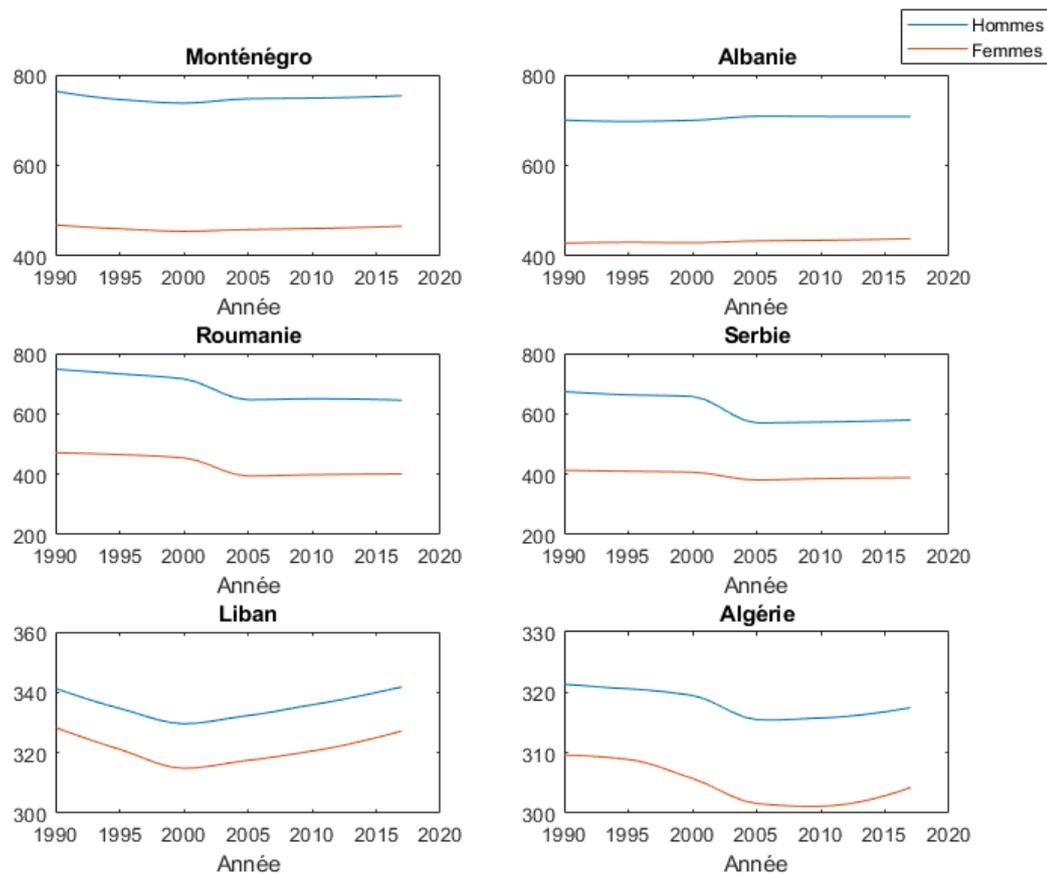


FIGURE 1.11 – Taux de prévalence de la fibrillation atriale standardisé selon l'âge de 1990 à 2017 dans quelques pays à revenu intermédiaire avec répartition des sexes [2].

## 1.2.2 Autres maladies cardiaques

### Cardiopathie hypertensive

Les maladies cardiaques hypertensives se réfèrent aux maladies cardiaques causées par une pression artérielle élevée. Le cœur travaillant sous une pression accrue provoque différents troubles cardiaques. Les maladies cardiaques hypertensives comprennent l'épaississement du muscle cardiaque, la maladie coronarienne et d'autres complications. Une cardiopathie hypertensive peut entraîner de graves problèmes de santé. C'est la principale cause de décès par hypertension. En général, les problèmes cardiaques associés à l'hypertension artérielle sont liés aux artères et aux muscles du cœur. On distingue deux types de cardiopathie hypertensive. Le premier est le rétrécissement des artères. Les artères coronaires transportent le sang vers le muscle cardiaque. Lorsque l'hypertension artérielle fait rétrécir les vaisseaux sanguins, le flux sanguin vers le cœur peut ralentir ou s'arrêter. Cette condition est connue sous le nom de maladie coronarienne. Cette dernière rend difficile le fonctionnement du cœur et l'approvisionnement en sang des autres organes. Cela peut exposer l'individu à un risque de crise cardiaque à cause d'un caillot de sang qui se coince dans l'une des artères rétrécies et coupe le flux sanguin vers le cœur. Le second type de cardiopathie hypertensive est l'épaississement et l'élargissement du cœur. L'hypertension artérielle empêche le cœur de pomper correctement le sang dans l'organisme. Comme les autres muscles du corps, un travail acharné régulier entraîne l'épaississement et la croissance des muscles cardiaques. Cela modifie le fonctionnement du cœur. Ces changements se produisent généralement dans la chambre de pompage principale du cœur, le ventricule gauche. La condition est connue comme l'hypertrophie ventriculaire gauche. Les maladies coronariennes peuvent provoquer l'hypertrophie ventriculaire gauche et vice versa. Lorsqu'un individu est atteint d'une ma-

l'infarctus coronarien, son cœur doit travailler plus fort. Si l'hypertrophie ventriculaire gauche agrandit le cœur, il peut comprimer les artères coronaires. Les maladies coronariennes et l'hypertrophie ventriculaire gauche peuvent conduire à une insuffisance cardiaque caractérisée par l'incapacité du cœur à pomper suffisamment de sang vers le reste de votre corps. Ils peuvent aussi provoquer une arythmie où le cœur bat anormalement, une cardiopathie ischémique qui se caractérise par le cœur qui ne reçoit pas assez d'oxygène, une crise cardiaque qui consiste en une interruption du flux sanguin vers le cœur conduisant au décès du muscle cardiaque par manque d'oxygène. D'autres complications peuvent en résulter comme l'arrêt cardiaque soudain où le cœur cesse soudainement de fonctionner, la respiration s'arrête et s'en suit une perte de connaissance [7].

### **Infarctus aigu du myocarde**

L'infarctus aigu du myocarde est le nom médical d'une crise cardiaque. Une crise cardiaque est une maladie mortelle qui survient lorsque le flux sanguin vers le muscle cardiaque est brusquement interrompu, causant des dommages aux tissus. Ceci est généralement le résultat d'un blocage dans une ou plusieurs des artères coronaires. Un blocage peut se développer en raison d'une accumulation de plaque, une substance principalement constituée de matières grasses, de cholestérol et de déchets cellulaires (voir la FIGURE 1.1). Bien que les symptômes classiques d'une crise cardiaque soient des douleurs thoraciques et un essoufflement, les symptômes peuvent être assez variés. Les symptômes les plus courants d'une crise cardiaque comprennent une pression ou oppression dans la poitrine, une douleur dans la poitrine, le dos, la mâchoire et d'autres parties du haut du corps qui dure plus de quelques minutes ou qui disparaît et revient, un essoufflement, la transpiration, un sentiment de nausée, des vomissements, de l'anxiété, de la toux, des vertiges et une fréquence cardiaque rapide. Toutes les personnes qui ont une crise cardiaque ne présentent pas les mêmes symptômes ni la même gravité des symptômes. La douleur thoracique est le symptôme le plus souvent signalé chez les femmes et les hommes. Cependant, les femmes sont plus susceptibles que les hommes de ressentir un essoufflement, une douleur à la mâchoire ou dans le haut du dos, des étourdissements, de la nausée et des vomissements. Le cœur est le principal organe du système cardio-vasculaire, qui comprend également différents types de vaisseaux sanguins. Certains des vaisseaux les plus importants sont les artères. Ils transportent du sang riche en oxygène vers le corps et tous les organes. Les artères coronaires transportent le sang riche en oxygène spécifiquement vers le muscle cardiaque. Lorsque ces artères deviennent bloquées ou rétrécies en raison d'une accumulation de plaque, le flux sanguin vers le cœur peut diminuer considérablement ou s'arrêter complètement. Cela peut provoquer une crise cardiaque. Plusieurs facteurs peuvent conduire à un blocage des artères coronaires. D'abord le mauvais cholestérol, également appelé lipoprotéine de basse densité, est l'une des principales causes d'un blocage dans les artères. Le cholestérol est une substance incolore présente dans les aliments comestibles au quotidien. Le corps le produit aussi naturellement. Tout le cholestérol n'est pas mauvais, mais la lipoprotéine de basse densité peut coller aux parois des artères et produire de la plaque. La plaque est une substance dure qui bloque la circulation sanguine dans les artères. Les plaquettes sanguines, qui aident le sang à coaguler, peuvent coller à la plaque et s'accumuler avec le temps. En second lieu, les graisses saturées qui peuvent également contribuer à l'accumulation de plaque dans les artères coronaires. Les graisses saturées se trouvent principalement dans la viande et les produits laitiers, y compris le bœuf, le beurre et le fromage. Ces graisses peuvent entraîner un blocage artériel en augmentant la quantité de mauvais cholestérol dans le système sanguin et en réduisant la quantité du bon cholestérol. Enfin, un autre type de graisse qui contribue aux artères obstruées est la graisse trans ou la graisse hydrogénée. Les gras trans sont généralement produits artificiellement et peuvent être trouvés dans une variété d'aliments transformés. Les gras trans sont généralement inscrits sur les étiquettes des aliments sous forme d'huile hydrogénée ou d'huile partiellement hydrogénée [8].

## Angine de poitrine

L'angine de poitrine est un type de douleur thoracique qui résulte d'une diminution du flux sanguin vers le cœur. Un manque de circulation sanguine signifie que le muscle cardiaque ne reçoit pas suffisamment d'oxygène. La douleur est souvent déclenchée par l'activité physique ou le stress émotionnel. L'angine de poitrine stable est le type d'angine de poitrine le plus courant. Elle constitue un schéma prévisible de douleur thoracique. L'angine de poitrine instable est une autre forme d'angine de poitrine. Elle survient soudainement et s'aggrave avec le temps. Cela peut éventuellement conduire à une crise cardiaque. Bien que l'angine de poitrine stable soit moins grave que l'angine instable, elle peut être douloureuse et inconfortable. Les deux types d'angine de poitrine sont généralement des signes d'une affection cardiaque sous-jacente. L'angine de poitrine stable se produit lorsque le muscle cardiaque ne reçoit pas l'oxygène dont il a besoin pour fonctionner correctement. Le cœur travaille alors plus fort lors d'une activité physique ou d'un stress émotionnel. Certains facteurs, tels que le rétrécissement des artères (athérosclérose), peuvent empêcher le cœur de recevoir plus d'oxygène. Les artères peuvent devenir étroites et dures lorsque la plaque (une substance constituée de graisse, de cholestérol, de calcium et d'autres substances) s'accumule à l'intérieur des parois des artères. Les caillots sanguins peuvent également bloquer les artères et réduire le flux de sang riche en oxygène vers le cœur. La sensation douloureuse qui se produit lors d'un épisode d'angor stable est souvent décrite comme une pression ou une plénitude au centre de la poitrine. La douleur peut se sentir comme un étau serrant votre poitrine ou comme un poids lourd reposant sur votre poitrine. Cette douleur peut se propager de votre poitrine à votre cou, vos bras et vos épaules. Lors d'un épisode d'angine stable, certains symptômes reviennent régulièrement tel que l'essoufflement, la nausée, la fatigue, les vertiges, la transpiration abondante et l'anxiété. Une angine stable se produit généralement après une activité physique intense. Les symptômes ont tendance à être temporaires, pouvant durer jusqu'à 15 minutes dans la plupart des cas. Ceci se distingue de l'angine instable, par la continuité et la sévérité de la douleur. Il est possible d'avoir un épisode d'angine de poitrine stable à tout moment de la journée. Cependant, il est plus probable de ressentir des symptômes le matin. La sensation douloureuse qui se produit lors d'un épisode d'angine stable est souvent décrite comme une pression ou une plénitude au centre de la poitrine. La douleur peut se sentir comme un étau serrant votre poitrine ou comme un poids lourd reposant sur votre poitrine. Cette douleur peut se propager de votre poitrine à votre cou, vos bras et vos épaules. Les facteurs de risque de l'angine de poitrine stable sont nombreux. On en distingue surtout l'obésité, les antécédents de maladies cardiaques, un taux de cholestérol élevé ou une pression artérielle élevée, le diabète, le tabagisme, l'inactivité physique. Les repas copieux, des entraînements physiques vigoureux et un temps extrêmement chaud ou froid peuvent également déclencher une angine stable dans certains cas [9].

## Arythmies cardiaques

Une arythmie est un trouble du cœur qui affecte le rythme auquel le cœur bat. Une arythmie se produit lorsque les impulsions électriques, qui dirigent et régulent les battements cardiaques, ne fonctionnent pas correctement. Cela fait battre le cœur soit trop vite (tachycardie), soit trop lentement (bradycardie), soit trop tôt (contraction prématurée) ou soit de façon erratique (fibrillation). Presque tout le monde connaîtra au moins une fois un rythme cardiaque anormal. Les arythmies sont courantes et généralement inoffensives, mais certaines sont problématiques. Lorsqu'une arythmie interfère avec le flux sanguin vers le corps, elle peut endommager le cerveau, les poumons et d'autres organes vitaux. Si elles ne sont pas traitées, ces arythmies peuvent être mortelles. Le cœur est divisé en quatre chambres. Chaque moitié du cœur se compose d'une chambre supérieure (l'oreillette) et d'une chambre inférieure (le ventricule). Les deux moitiés créent deux pompes, une de chaque côté du cœur. Dans un cœur qui bat correctement, les impulsions électriques suivent des voies précises à travers le cœur jusqu'à chaque pompe. Ces signaux coordonnent l'activité du muscle cardiaque afin que le sang pénètre dans le cœur et en sort. Toute interruption de ces voies ou impulsions peut provoquer un rythme cardiaque anormal. Le sang

pénètre dans le cœur et arrive en premier dans les oreillettes. Ensuite, un seul battement de cœur implique plusieurs étapes. D'abord le nœud sinusal (un groupe de cellules dans l'oreillette droite) envoie une impulsion électrique aux oreillettes droite et gauche, leur disant de se contracter. Cette contraction permet aux ventricules de se remplir de sang. Ensuite, pendant que les ventricules se remplissent, l'impulsion électrique se déplace vers le centre du cœur jusqu'au nœud auriculo-ventriculaire (un groupe de cellules entre les oreillettes et les ventricules). L'impulsion sort alors du nœud et se rend dans les ventricules remplis de sang, leur ordonnant de se contracter. Cette contraction pousse le sang hors du cœur et dans le corps pour la circulation. C'est un battement de cœur. Puis, le processus recommence dans le cadre d'un cycle cardiaque. Dans des conditions normales, les côtés gauche et droit du cœur battent l'un après l'autre. Cela permet au flux sanguin de se déplacer dans une direction de façon continue. Un cœur normal répétera ce processus environ 100 000 fois par jour. Cela représente 60 à 100 battements par minute pour une personne en bonne santé moyenne au repos.

Les arythmies sont nommées et classées en fonction de trois caractéristiques, le rythme (que ce soit trop lent ou trop rapide), l'origine (que ce soit dans les ventricules ou les oreillettes) et la régularité. La bradycardie est un rythme cardiaque lent. Elle est définie comme une fréquence cardiaque au repos de moins de 60 battements par minute. Toutes les bradycardies ne sont pas un problème. Les athlètes et les personnes en bonne forme physique souffrent souvent de bradycardies. Leur fréquence cardiaque au repos peut être inférieure à 60 battements par minute car leur cœur est plus efficace et peut pomper suffisamment de sang avec moins de battements. Cependant, un rythme cardiaque lent peut signifier que votre cœur ne bat pas assez fréquemment pour assurer un flux sanguin adéquat dans tout votre corps. On distingue plusieurs types de bradycardies. D'abord, le sinus malade dont le nœud est chargé de définir le rythme du cœur. S'il n'envoie pas correctement les impulsions électriques, le cœur peut pomper trop lentement ou irrégulièrement. Les cicatrices à proximité du nœud sinusal d'une maladie cardiaque ou d'une crise cardiaque peuvent également ralentir ou bloquer les impulsions électriques lorsqu'elles tentent de traverser le cœur. Suite au blocage de conduction qui survient lorsque les voies électriques du cœur sont bloquées, les cavités du cœur se contractent lentement ou pas du tout. Un blocage peut se produire n'importe où le long des voies électriques du cœur entre le nœud sinusal et le nœud auriculoventriculaire, ou entre le nœud auriculoventriculaire et les ventricules. Il ne peut y avoir aucun signe de ces blocages autre que des battements cardiaques sautés ou ralentis.

La tachycardie est un rythme cardiaque rapide. Elle est définie comme une fréquence cardiaque au repos de plus de 100 battements par minute. Les deux types de tachycardie les plus courants sont la tachycardie supraventriculaire et la tachycardie ventriculaire. La tachycardie supraventriculaire est toute arythmie qui commence au-dessus du ventricule. Elles sont généralement identifiées par une explosion de battements cardiaques rapides qui peuvent être chroniques ou commencer et se terminer soudainement. Ces salves peuvent durer quelques secondes ou plusieurs heures et provoquer des battements cardiaques plus de 160 fois par minute. Les tachycardies supraventriculaires les plus courantes incluent la fibrillation atriale et le flutter auriculaire. Dans une fibrillation atriale, l'oreillette bat très rapidement (de 240 à 350 battements par minute). Les oreillettes se déplacent si rapidement qu'elles ne sont pas capables de se contracter complètement. Au lieu de cela, elles tremblent ou se fibrillent. Cela peut provoquer une gêne, mais pas un pouls rapide. Cependant, certains de ces battements auriculaires peuvent se transférer vers les ventricules et provoquer une fréquence cardiaque élevée. La fibrillation atriale affecte principalement les personnes âgées. Le risque de développer cette arythmie augmente au-delà de 60 ans, principalement en raison de l'usure d'une expérience cardiaque plus ancienne. Les chances de développer une fibrillation atriale sont également élevées en cas d'hypertension ou d'autres problèmes cardiaques. La fibrillation atriale peut être dangereuse car si elle n'est pas traitée, elle peut entraîner des conditions plus graves, comme un accident vasculaire cérébral. Une analyse plus poussée sur la fibrillation atriale a été présentée dans la section 1.2.1. Durant un flutter auriculaire, les battements cardiaques sont plus rythmiques et constants que les battements cardiaques dans la fibrillation atriale. Pourtant, le flutter auriculaire peut aller et venir en rafales soudaines et

peut être mortel. Ce type d'arythmie survient le plus souvent chez les personnes atteintes d'une maladie cardiaque. Il apparaît également souvent dans les premières semaines après une chirurgie cardiaque. La tachycardie ventriculaire est une arythmie qui commence dans les ventricules du cœur. La plupart des tachycardies ventriculaires surviennent chez des personnes qui ont eu une maladie cardiaque ou des problèmes cardiaques, comme une maladie coronarienne ou une crise cardiaque. Elle est souvent causée par une impulsion électrique se déplaçant autour d'une partie cicatricielle du muscle cardiaque. Cela peut provoquer la contraction des ventricules plus de 200 fois par minute. Si elle n'est pas traitée, elle peut augmenter le risque de développer des arythmies ventriculaires plus graves, telles que la fibrillation ventriculaire. La fibrillation ventriculaire peut avoir des signes tels que des battements cardiaques soudains, rapides, irréguliers et chaotiques dans le ventricule. Ces impulsions électriques erratiques, parfois déclenchées par une crise cardiaque, font trembler les ventricules du cœur ce qu'on appelle une fibrillation. Lors d'une telle arythmie, les ventricules ne peuvent pas pomper le sang dans le corps et la fréquence cardiaque diminue rapidement. Cela fait chuter la tension artérielle et diminue l'apport sanguin au corps et aux organes. La fibrillation ventriculaire est la première cause d'arrêt cardiaque soudain. Les battements cardiaques prématurés peuvent donner l'impression que le cœur a sauté un battement. En réalité, le rythme cardiaque normal a été interrompu par un battement trop tôt et un battement supplémentaire entre deux battements de cœur normaux est alors ressenti [10].

### **Insuffisance cardiaque**

L'insuffisance cardiaque se caractérise par l'incapacité du cœur à pomper une quantité suffisante de sang dans le corps. Sans un flux sanguin suffisant, toutes les principales fonctions du corps sont perturbées. L'insuffisance cardiaque est une affection ou un ensemble de symptômes qui affaiblissent votre cœur. Chez certaines personnes souffrant d'insuffisance cardiaque, le cœur a du mal à pomper suffisamment de sang pour soutenir d'autres organes du corps. D'autres personnes peuvent avoir un durcissement et un raidissement du muscle cardiaque lui-même, ce qui bloque ou réduit le flux sanguin vers le cœur. Il peut s'agir d'une affection aiguë (à court terme) ou chronique (à long terme). Dans l'insuffisance cardiaque aiguë, les symptômes apparaissent soudainement mais disparaissent assez rapidement. Cette condition survient souvent après une crise cardiaque. Cela peut également être le résultat d'un problème avec les valves cardiaques qui contrôlent le flux sanguin dans le cœur. Cependant, dans l'insuffisance cardiaque chronique, les symptômes sont continus et ne s'améliorent pas avec le temps. La grande majorité des cas d'insuffisance cardiaque sont chroniques. L'insuffisance cardiaque est une maladie grave qui nécessite un traitement. Un traitement précoce augmente les chances de guérison à long terme avec moins de complications. Les symptômes de l'insuffisance cardiaque peuvent inclure une fatigue excessive, la prise de poids soudaine, une perte d'appétit, une toux persistante, un pouls irrégulier, des palpitations cardiaques, un gonflement abdominal, un essoufflement, un gonflement des jambes et des chevilles et des veines cervicales saillantes.

L'insuffisance cardiaque est le plus souvent liée à une autre maladie. La cause d'insuffisance cardiaque la plus courante est la maladie coronarienne, un trouble qui provoque un rétrécissement des artères qui fournissent le sang et l'oxygène au cœur. Il existe d'autres conditions qui peuvent augmenter le risque de développer une insuffisance cardiaque comme la cardiomyopathie, un trouble du muscle cardiaque qui affaiblit le cœur, une malformation cardiaque congénitale, une crise cardiaque, une maladie des valves cardiaques, certains types d'arythmies ou de rythmes cardiaques irréguliers, l'hypertension artérielle, l'emphysème qui est une maladie pulmonaire, le diabète, une thyroïde hyperactive ou insuffisante, le virus de l'immunodéficience humaine (VIH), le syndrome d'immunodéficience acquise (SIDA), les formes graves d'anémie, certains traitements contre le cancer, comme la chimiothérapie ou l'abus de drogues ou d'alcool.

L'insuffisance cardiaque peut survenir dans le côté gauche ou droit du cœur. Il est également possible que les deux côtés du cœur échouent en même temps. L'insuffisance cardiaque est également classée comme diastolique ou systolique. L'insuffisance cardiaque gauche est le type d'insuffisance cardiaque le plus courant. Le ventricule cardiaque gauche est situé en bas à gauche du

cœur. Cette zone pompe le sang riche en oxygène vers le reste du corps. L'insuffisance cardiaque gauche se produit lorsque le ventricule gauche ne pompe pas efficacement. Cela empêche le corps d'obtenir suffisamment de sang riche en oxygène. Le sang retourne à la place dans les poumons, ce qui provoque un essoufflement et une accumulation de liquide. Le ventricule cardiaque droit est chargé de pomper le sang vers les poumons pour recueillir l'oxygène. L'insuffisance cardiaque droite se produit lorsque le côté droit du cœur ne peut pas effectuer son travail efficacement. Elle est généralement déclenchée par une insuffisance cardiaque gauche. L'accumulation de sang dans les poumons causée par une insuffisance cardiaque du côté gauche rend le ventricule droit plus difficile à travailler. Cela peut stresser le côté droit du cœur et le faire échouer. Une insuffisance cardiaque du côté droit peut également survenir à la suite d'autres conditions, telles qu'une maladie pulmonaire. L'insuffisance cardiaque droite est marquée par un gonflement des membres inférieurs. Cette enflure est causée par un écoulement fluide dans les jambes, les pieds et l'abdomen. L'insuffisance cardiaque diastolique survient lorsque le muscle cardiaque devient plus rigide que la normale. La raideur, qui est généralement due à une maladie cardiaque, signifie que le cœur ne se remplit pas facilement de sang. Ceci est connu sous le nom de dysfonction diastolique et conduit à un manque de circulation sanguine vers le reste des organes du corps. L'insuffisance cardiaque diastolique est plus fréquente chez les femmes que chez les hommes. L'insuffisance cardiaque systolique, quant à elle, survient lorsque le muscle cardiaque perd sa capacité à se contracter. Les contractions du cœur sont nécessaires pour pomper le sang riche en oxygène vers le corps. Ce problème est connu sous le nom de dysfonction systolique, et il se développe généralement lorsque le cœur est faible et hypertrophié. L'insuffisance cardiaque systolique est plus fréquente chez les hommes que chez les femmes. Une insuffisance cardiaque diastolique et systolique peut survenir du côté gauche ou droit du cœur [11].

### 1.3 Analyse épidémiologique

Les maladies cardio-vasculaires sont la principale cause de décès et d'invalidité [12; 13] dans le monde. Dix-huit millions de décès par an sont attribués aux maladies cardio-vasculaires dans le monde [14]. L'incidence des événements cardio-vasculaires majeurs a été signalée comme étant la plus élevée dans les pays à faible revenu, malgré le fait que ces pays présentent des facteurs de risque plus faibles [15]. Les données sur la mortalité sont couramment utilisées dans la surveillance des maladies. Le taux de mortalité d'une maladie peut fournir une indication de son influence au sein d'une population, en particulier pour les maladies chroniques telles que les maladies cardio-vasculaires. De plus, la mortalité est l'un des résultats les plus fiables. La déclaration des décès fait souvent partie intégrante du système de statistiques de l'état civil d'un pays. Dans la plupart des pays, les autorités sanitaires exigent que chaque décès dans la population soit signalé, y compris un enregistrement de la cause du décès. Ces enregistrements peuvent ensuite être compilés afin de produire des mesures de mortalité pour la population, y compris le nombre absolu de décès et les taux qui s'ajustent à la taille et à la distribution de la population. Bien que les systèmes d'enregistrement des décès soient universels, leur qualité et leur exhaustivité peuvent varier. Dans la plupart des pays développés, ce système est relativement complet et les taux de mortalité calculés à partir des données sont assez précis. Cependant, dans de nombreux pays moins développés, les systèmes de dossiers de santé peuvent être incomplets. De nombreux décès peuvent ne pas être déclarés, en particulier s'ils surviennent en dehors du système de santé. Bien que la présentation du nombre total de décès d'une maladie limite notre capacité de comparaison entre les populations, elle nous permet de comparer les maladies d'une même population. Entre 2007 et 2017, 9,57 millions de décès ont été enregistrés dans le monde, dont 4,11 en raison de maladies cardio-vasculaires selon les données de l'organisation mondiale de la santé [2]. Dans la FIGURE 1.12, on remarque que les décès par MCV en 2000 sont concentrés dans les états de l'Asie, et plus légèrement en Afrique, puis, sur la FIGURE 1.13, on relève qu'en 2017, les décès par MCV ont baissés en Asie alors que dans le reste du monde, ils n'ont pas beaucoup évolué.

Pour les maladies associées au vieillissement telles que les maladies cardio-vasculaires, le taux

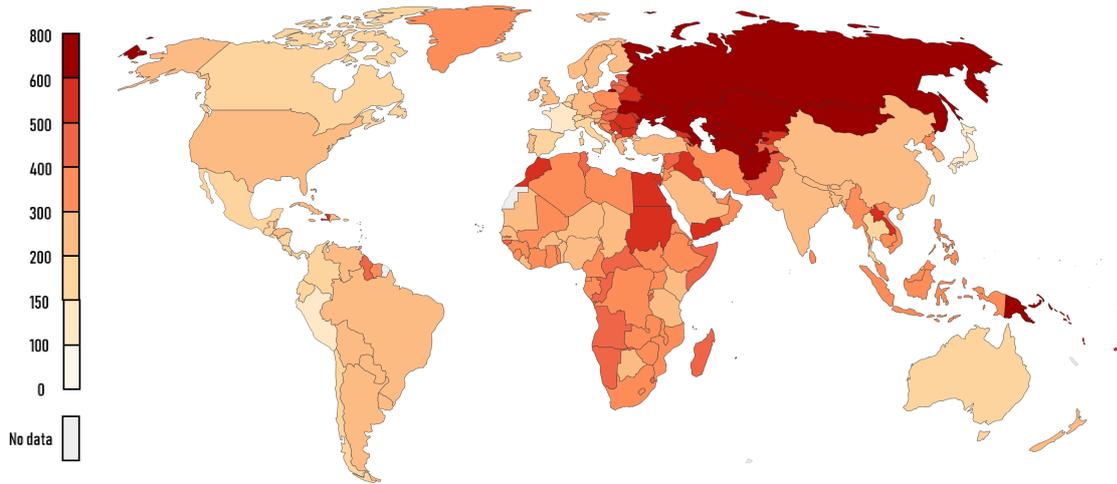


FIGURE 1.12 – Répartition des décès par maladies cardio-vasculaires standardisée selon l'âge par 100000 habitants en 2000. Source : "www.ouronlinedata.org"

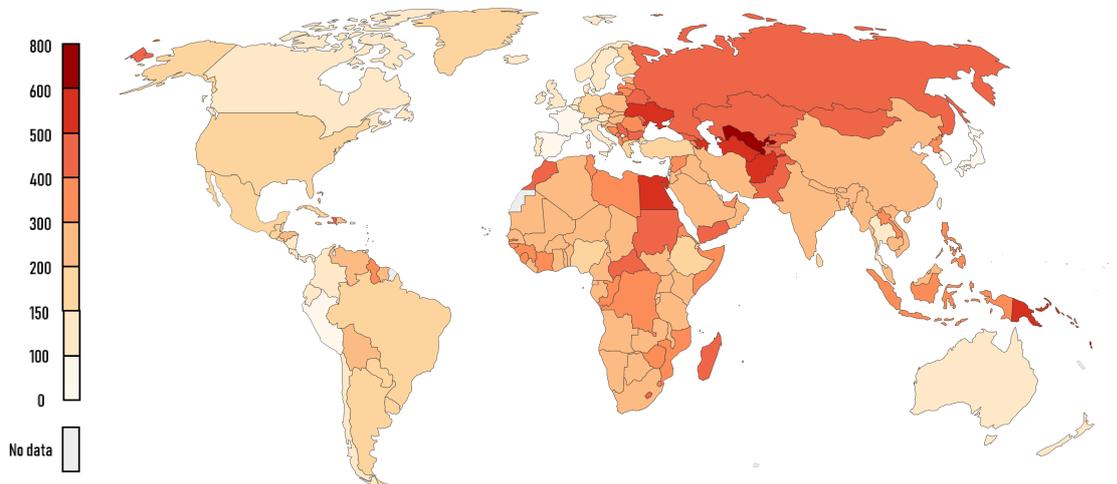


FIGURE 1.13 – Répartition des décès par maladies cardio-vasculaires standardisée selon l'âge par 100000 habitants en 2017. Source : "www.ouronlinedata.org"

de décès pour 100 000 habitants est influencé par la répartition par âge de la population. Une population comptant une plus grande répartition de personnes âgées devrait connaître un plus grand nombre de décès par MCV. Afin de faire des comparaisons entre les populations en calculant les taux bruts pour ajuster la taille de la population, il est souvent nécessaire de calculer des taux normalisés selon l'âge qui s'ajustent également à la distribution de la population [16]. Les taux de mortalité normalisés selon l'âge ajustent les différences dans la répartition par âge des populations en appliquant les taux de mortalité spécifiques à chaque population observés à une population standard [16]. Une comparaison des taux de mortalité standardisés pour l'âge de différents pays ne sera pas affectée par les différences de répartition d'âge de leurs populations. Cela peut être utile pour comparer les pays d'Europe, car l'espérance de vie peut varier considérablement [17]. Les taux de mortalité normalisés selon l'âge calculés pour chaque pays peuvent être considérés comme le taux que les pays auraient s'ils avaient la même répartition de la population que la population standard. La population standard recommandée utilisée pour le calcul des taux de mortalité standardisés pour l'âge en Europe est le "European Standard Population" ou "ESP" [17; 18] (Voir FIGURE 1.14). La FIGURE 1.14 montre le taux de mortalité standardisé selon l'âge des maladies cardio-vasculaires dans le monde de 1990 à 2017 pour 100000 habitants. On note une

décroissance générale du taux de mortalité au cours de la période de 1990 à 2017, dont la médiane reste autour de 300 par 100000.

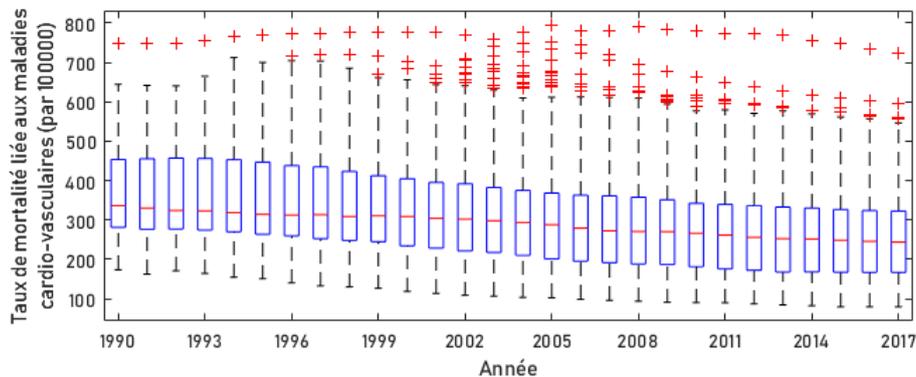


FIGURE 1.14 – Taux de mortalité standardisé selon l'âge des maladies cardio-vasculaires dans le monde de 1990 à 2017 pour 100000 habitants. Source de données : [3]

L'utilisation de mesures de mortalité absolue, comme le nombre de décès ou le nombre total d'années potentielles de vie perdues (APVP), peut induire en erreur car elle ne tient pas compte de la taille de la population. Les comparaisons entre pays et dans le temps devraient nous montrer le taux brut qui prend en compte les différences et les changements de populations. Les taux bruts sont calculés en divisant le nombre total de décès (ou APVP) par la taille de la population, généralement exprimé par 100 000 individus au sein de la population [16].

### 1.3.1 Facteurs de risque

Il existe une relation linéaire entre les niveaux de pression artérielle et le risque d'accident vasculaire cérébral ou d'infarctus du myocarde [19]. Des enquêtes internationales montrent que le taux d'hypertension artérielle systolique ( $\geq 140$  mmHg) a considérablement augmenté entre 1990 et 2015 avec des effets d'entraînement sur les années de vie ajustées sur l'incapacité et les décès attribuables à l'hypertension [20]. L'étude INTERHEART a estimé que 22% des infarctus du myocarde en Europe sont liés à l'hypertension, ce qui double presque le risque par rapport aux personnes sans antécédents d'hypertension [21]. Le traitement pour abaisser la tension artérielle offre une protection significative contre les événements cardio-vasculaires, avec des avantages supplémentaires de traitement plus intensif à l'égard des patients à haut risque (maladie vasculaire concomitante, maladie rénale ou diabète) [22]. En 2015, la prévalence médiane standardisée selon l'âge de l'hypertension artérielle (définie comme la pression artérielle systolique  $\geq 140$  mmHg ou la pression artérielle diastolique  $\geq 90$  mmHg) dans les pays membres de l'ESC était de 24.8% (intervalle inter-quartile (IIQ) [19.8;28.5]%).

Le cholestérol, en particulier le cholestérol LDL, est un déterminant majeur du risque de MCV qui augmente linéairement à mesure que les concentrations sanguines augmentent [23; 24]. Il s'agit d'une cible majeure des programmes de réduction des risques dans lesquels la thérapie par statines chez les personnes sans antécédents de MCV (prévention primaire) peut produire une réduction de 15% du risque de décès vasculaire pour chaque réduction de 1 mmol/L du cholestérol LDL [25; 26]. La prévalence d'un taux de cholestérol total élevé dépasse 50% dans les pays à revenu élevé, dont l'Europe, contre moins de 30% en Afrique et en Asie du Sud-Est [27]. Les données pour 2009 ont montré que la concentration médiane de cholestérol sanguin moyenne dans tous les pays membres de l'ESC était de 5.1 mmol/L chez les femmes et les hommes, allant de 4.4 mmol/L chez les femmes et les hommes du Kirghizistan à 5.6 mmol/L chez les femmes et les hommes d'Islande. La prévalence médiane standardisée d'hypercholestérolémie ( $> 6.2$  mmol/L) en 2008 dans les pays membres était de 15.6% et 14.3% chez les femmes et les hommes, allant de  $< 10\%$  en Azerbaïdjan, Bosnie-Herzégovine, République de Géorgie, Kirghizistan, République de

Moldova et la Turquie à > 20% en Belgique, au Danemark, en Finlande, en France, en Allemagne, en Islande, en Irlande, au Luxembourg, en Norvège et au Royaume-Uni.

L'OMS signale que 422 millions de personnes vivent actuellement avec le diabète dans le monde (principalement de type 2), dont plus de 60 millions vivent dans la région européenne. La prévalence du diabète chez les hommes et les femmes a augmenté de façon spectaculaire au cours des 30 dernières années, sans différence substantielle dans le monde entre les pays à revenu élevé, intermédiaire ou faible [27]. Cela est largement attribué au nombre croissant de personnes en sur-poids et obèses, qui est à son tour motivé par un excès de calories alimentaires et l'inactivité physique. Le diabète double le risque de décès par rapport aux personnes sans diabète. Au moins la moitié de ces décès sont le résultat d'une maladie cardio-vasculaire, généralement une cardiopathie ischémique ou un accident vasculaire cérébral [27]. La prévalence médiane du diabète selon les estimations de 2017 chez les femmes et les hommes âgés de 20 à 79 ans était en moyenne de 6.8% dans les pays membres de l'ESC, allant de  $\leq 4\%$  en Estonie, en Irlande et en Lituanie à  $\geq 10\%$  en Albanie, Bosnie-Herzégovine, Égypte, Liban, Libye, Macédoine du Nord, Monténégro, Serbie et Turquie.

La prévalence du sur-poids et de l'obésité, définie par un indice de masse corporelle (IMC) de  $\geq 25\text{kg}/\text{m}^2$  et  $\geq 30\text{kg}/\text{m}^2$ , respectivement, augmente dans les pays développés et en cours de développement. Davantage de personnes sont maintenant obèses que le contraire globalement dans le monde et dans la plupart des régions du monde [28]. L'épidémie d'obésité est largement due à la libéralisation du commerce mondial, à la croissance économique et à l'urbanisation rapide affectant le mode de vie et l'apport alimentaire, avec une tendance vers une consommation plus importante de graisses animales et de sucre ajouté [29]. En 2016, la méta-analyse de la "Global BMI Mortality Collaboration" (GBMC) a indiqué que pour un IMC  $\geq 25.0\text{kg}/\text{m}^2$ , le rapport de risque pour chaque augmentation de  $5\text{kg}/\text{m}^2$  de l'IMC était de 1.39 parmi les personnes vivant dans la région européenne. La collaboration a conclu que les conséquences d'une augmentation de l'IMC pour toutes causes de la mortalité étaient graves et cohérentes dans toutes les régions du monde [30]. L'IMC médian normalisé selon l'âge dans les pays membres de l'ESC en 2016 était de  $26.6\text{kg}/\text{m}^2$  et était similaire pour les femmes  $26.2\text{kg}/\text{m}^2$  et les hommes  $27.0\text{kg}/\text{m}^2$ . L'IMC moyen chez les femmes variait de  $23.7\text{kg}/\text{m}^2$  en Suisse à  $31.4\text{kg}/\text{m}^2$  en Égypte et chez les hommes de  $24.7\text{kg}/\text{m}^2$  en Algérie à  $28.2\text{kg}/\text{m}^2$  en Hongrie. Les données pour 2016 ont montré que, dans les pays membres de l'ESC, environ une femme et un homme sur cinq étaient obèses ( $\geq 30\text{kg}/\text{m}^2$ ), comme en témoignent les taux de prévalence de 22.8% et 22.3%, respectivement. L'obésité chez les femmes était particulièrement courante en Égypte et en Libye, où elle touchait plus d'une femme sur trois, tandis que l'obésité chez les hommes était courante à Malte, en Hongrie, en République tchèque, au Royaume-Uni et au Liban, où elle touchait plus d'un homme sur quatre.

### 1.3.2 Mauvaises habitudes liées

Le tabac a été décrit comme «le plus grand risque sanitaire évitable dans l'Union Européenne» par la Direction générale de la santé et des consommateurs (DG SANTÉ). Le tabac est lié à de nombreuses formes de cancer et de maladies cardio-vasculaires. De plus, c'est la principale cause de décès de près de 6 millions de personnes par an. Par conséquent, l'UE a encouragé des mesures politiques liées à l'usage du tabac et à la commercialisation des dérivés du tabac au cours des 15 dernières années. Au cours de cette période, la prévalence du tabac a constamment diminué en Europe [31]. La prévalence du tabagisme en 2014 était de 28.6% en France alors qu'en 2017, les prévalences de tabagisme concernant les hommes et les femmes âgés de plus de 15 ans sont exposées dans le graphe de la FIGURE 1.15.

La consommation d'alcool est définie comme la quantité enregistrée (en litres) d'alcool pur consommée par adulte (plus de 15 ans) au cours d'une année civile et est un indicateur de santé de base européen. L'indicateur ne prend en compte que la consommation enregistrée à partir de la production, import, export et vente des données, souvent via la fiscalité. Les limites supérieures recommandées pour la consommation d'alcool varient selon les pays, reflétant peut-être l'incertitude quant aux seuils de risque précis. Aux États-Unis, la consommation excessive d'alcool reste la

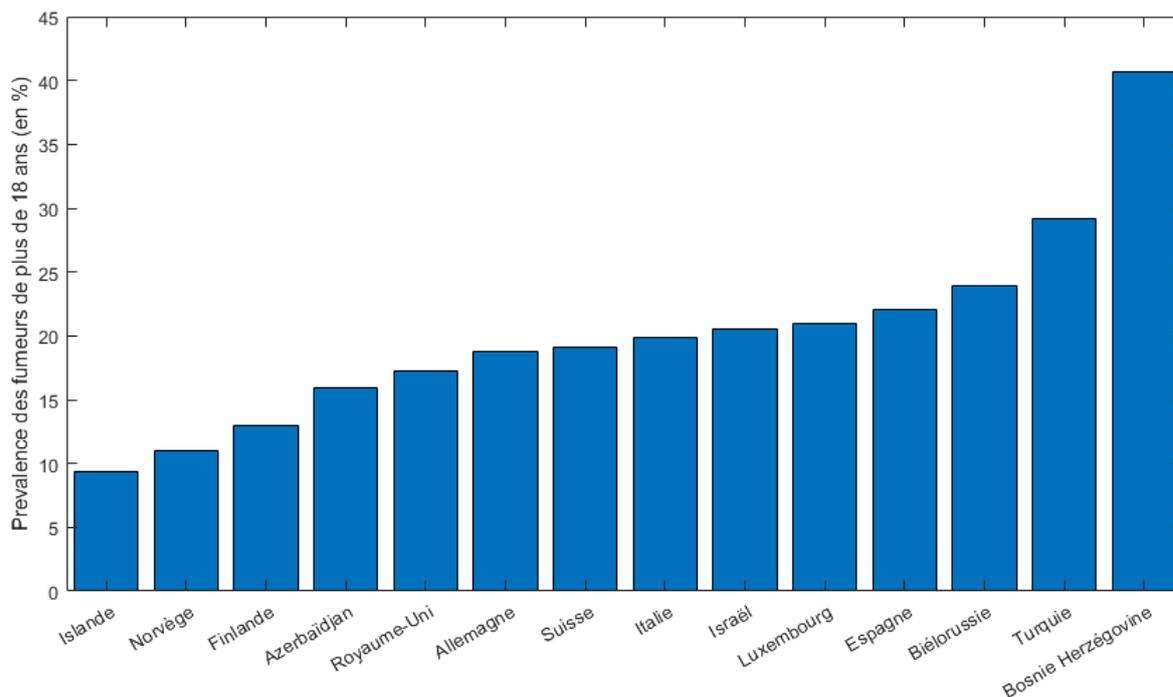


FIGURE 1.15 – Pourcentage de fumeurs quotidiens réguliers dans la population de 15 ans et plus en 2017. Source : Site officiel de l’Organisation Mondiale de la Santé (OMS).

principale cause de décès prématuré, où elle est responsable de 1 décès sur 10 chez les adultes en âge de travailler [32]. Dans l’UE, la consommation nocive d’alcool est la troisième cause de décès prématuré après le tabac et l’hypertension avec l’alcool. On estime que la dépendance est responsable de plus de 60% de la mortalité attribuable à l’alcool [33]. Les données pour 2016 montrent que 66.6% des personnes âgées de plus de 15 ans vivant dans les pays membres de l’ESC avaient consommé de l’alcool au cours des 12 mois précédents, avec une consommation médiane de 10.2 L/habitant/an. Il y avait de grandes différences entre les pays, avec une consommation allant de moins de 4 L/habitant/an, en Algérie, Azerbaïdjan, Égypte, Liban, Libye, Maroc, Syrie, Tunisie et Turquie, à plus de 12 L/habitant/an en Belgique, France, Allemagne, Lettonie, Portugal, Slovénie, Bulgarie, République de Moldova et Roumanie. Des facteurs culturels et religieux sont susceptibles de contribuer à la très faible consommation d’alcool dans de nombreux pays du Moyen-Orient et d’Afrique du Nord.

La faible consommation de fruits et légumes est un facteur de risque reconnu de maladies non transmissibles, comme le cancer et les maladies coronariennes [34; 35]. Dans une revue systématique et une méta-analyse dose-réponse de 95 études prospectives, des réductions progressives du risque de MCV et de la mortalité toutes causes confondues ont été observées jusqu’à un apport de 800 g/jour de fruits et légumes combinés, alors que pour le cancer total, aucune d’autres réductions du risque ont été observées au-dessus de 600g/jour [36]. Comme toutes les études examinant les relations entre la nutrition et la maladie, une interprétation prudente est nécessaire, en partie en raison de la confusion par un mode de vie sain qui rend les contributions nutritionnelles à la réduction du risque difficiles à déterminer et en partie en raison du manque variable de précision dans la mesure diététique. Les dernières estimations disponibles pour la consommation de légumes et de fruits ont été obtenues en 2014 pour les personnes de plus de 15 ans et se limitaient à 22 pays à revenu élevé et à un pays à revenu intermédiaire (Turquie). Dans ces 23 pays membres du CES, 52.2% des personnes consommaient au moins une portion de légumes par jour et 55% consommaient au moins une portion de fruits par jour. La consommation de légumes était la plus élevée en Israël, où 81% de la population consommait au moins une portion par jour, contre moins de 35% des populations des Pays-Bas et de l’Allemagne. La consommation de fruits était la plus élevée en Italie, où plus de 70% de la population consommait au moins une portion

par jour, contre seulement 20.2% en Lettonie.

L'activité physique insuffisante est définie comme la proportion de la population atteignant moins de 150 minutes d'activité physique d'intensité modérée par semaine ou moins de 75 minutes d'activité physique d'intensité vigoureuse par semaine. L'inactivité augmente le risque de plusieurs maladies non transmissibles telles que la cardiopathie ischémique, le diabète de type 2, les cancers du sein et du côlon, et représente près de 10% de tous les décès dans le monde [37]. La promotion de l'exercice pendant les loisirs a constamment démontré qu'elle favorisait la santé cardio-vasculaire [38]. En 2016, la prévalence médiane normalisée selon l'âge de l'activité physique insuffisante auto-déclarée était de 29.32% chez les adultes âgés de plus de 18 ans en France. Les prévalences des pays d'Europe les plus touchés par l'activité physique insuffisante sont affichées dans la FIGURE 1.16. La prévalence était généralement plus élevée chez les femmes que chez les hommes, avec plus de 43% des femmes à Chypre, en Allemagne, en Italie, à Malte et au Portugal déclarant une activité physique insuffisante par rapport à plus de 35% des hommes dans les mêmes pays.

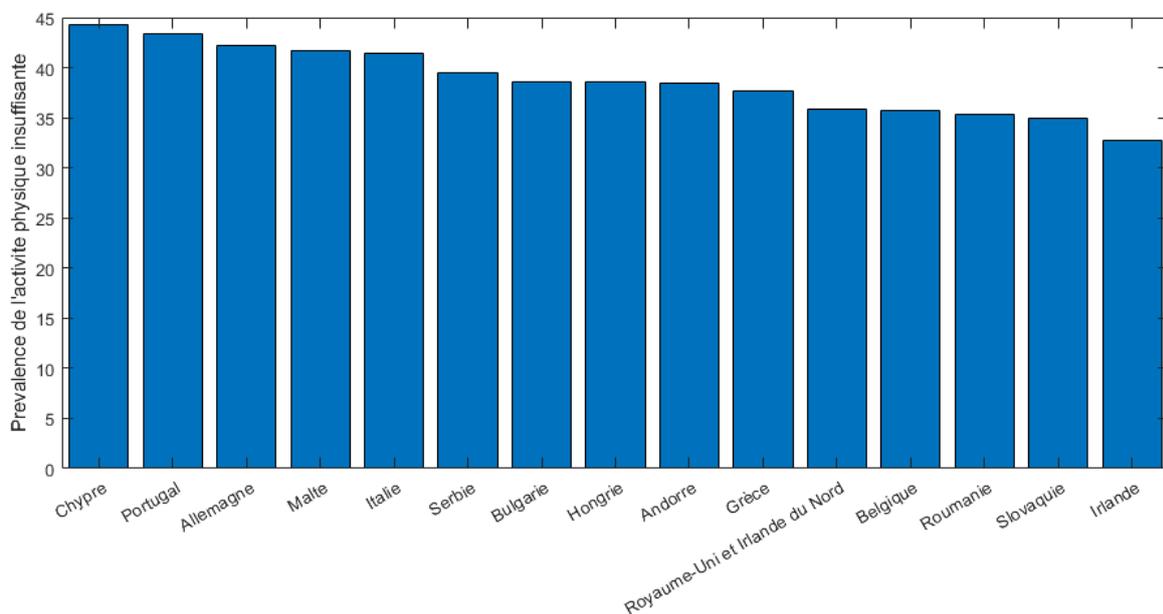


FIGURE 1.16 – Prévalence de l'activité physique insuffisante chez les adultes en Europe en 2016. Source : Site officiel de l'Organisation Mondiale de la Santé (OMS).

### 1.3.3 Analyse de morbidité

Les mesures de la morbidité jouent un rôle important dans la description de l'épidémiologie d'une maladie, car elles représentent le nombre de personnes qui en souffrent au sein d'une population [16]. Les mesures de prévalence décrivent le nombre d'individus qui souffrent actuellement d'une maladie particulière dans une population donnée. Dans le cas d'événements cardio-vasculaires aigus, tels que l'infarctus du myocarde ou l'accident vasculaire cérébral, ils indiquent le nombre de personnes vivant dans la population, qui ont déjà souffert d'un tel événement. L'incidence définit le nombre de nouveaux cas dans une population au cours d'une période donnée, fournissant une mesure de l'occurrence de la maladie. Même pour les conditions avec un taux de mortalité relativement élevé, comme les accidents vasculaires cérébraux et les infarctus aigus du myocarde, la morbidité est une mesure importante à côté de la mortalité, car l'invalidité causée par de tels événements peut être élevée chez ceux qui y survivent et il est difficile de traiter les individus indépendamment de leur niveau de récupération. Cependant, par rapport à la collecte de données sur la mortalité, qui implique d'agrèger le nombre de décès enregistrés, la collecte de données sur la morbidité est plus difficile. Malgré la prolifération des systèmes d'enregistrement

électronique au sein des services de santé, peu de pays utilisent ces systèmes pour signaler des enregistrements précis de morbidité. Il faut donc s'appuyer sur d'autres sources pour les statistiques de la morbidité. Les données de l'étude "Global Burden of Disease", qui modélise les estimations de la morbidité pour les pays en utilisant les données des enquêtes sur la santé, les cohortes prospectives, les données administratives du système de santé et les registres, sont utilisés dans cette section [39; 40]. Ces estimations sont inappropriées à la comparaison de morbidité entre les pays et dans le temps. Il est à noter que les statistiques présentées dans cette section concernent tous les pays membres de l'ESC dont les données ont été rendues disponibles [41; 42].

En 2017, 108.7 millions de personnes vivent avec des maladies cardio-vasculaires dont 8 millions en France et 19.9 millions de nouveaux cas ont été enregistrés dont 1.47 en France. Les contributions nationales des nouveaux cas étaient en partie déterminées par la taille de la population, l'Allemagne et la Fédération de Russie contribuant chacune à environ 2.5 millions de nouveaux cas et l'Islande et Malte à moins de 10 000. Les graphes dans la FIGURE 1.17 et la FIGURE 1.18 montrent respectivement l'incidence et la prévalence standardisée selon l'âge des maladies cardio-vasculaires en 2017, pour plusieurs pays membres de l'ESC.

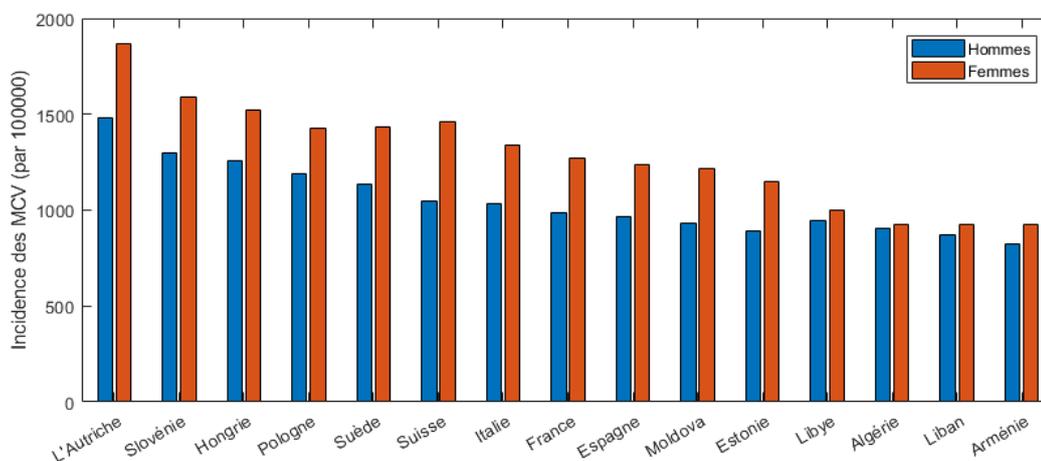


FIGURE 1.17 – Incidence des maladies cardio-vasculaire en 2017 [2].

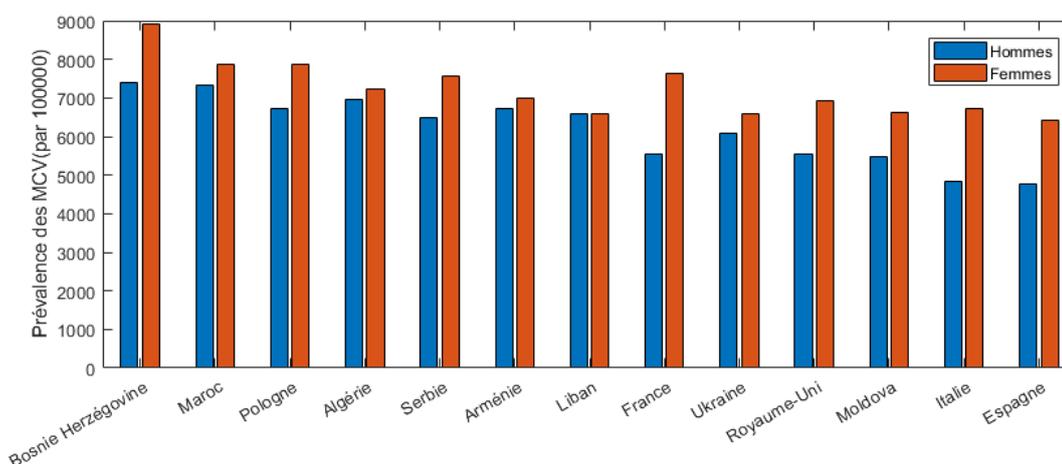


FIGURE 1.18 – Prévalence des maladies cardio-vasculaire en 2017 [2].

### La cardiopathie ischémique

La cardiopathie ischémique est un groupe de troubles résultant d'un apport insuffisant d'oxygène au muscle cardiaque en raison du développement et des complications de l'athérosclérose

dans une ou plusieurs artères coronaires. L'obstruction des artères coronaires peut dans certains cas être plus ou moins complète ou brutale. De plus, des lésions myocardiques de gravité variable ainsi qu'une ischémie entraînant une nécrose myocardique surviennent généralement suite à un manque d'oxygénation. La FIGURE 1.19 montre l'évolution de l'incidence standardisée selon l'âge en France de la cardiopathie ischémique de 1990 à 2017. On remarque qu'en France, les hommes sont généralement plus touchés par la cardiopathie ischémique que les femmes et que la prévalence décroît légèrement avec le temps avant de se stabiliser vers 2005.

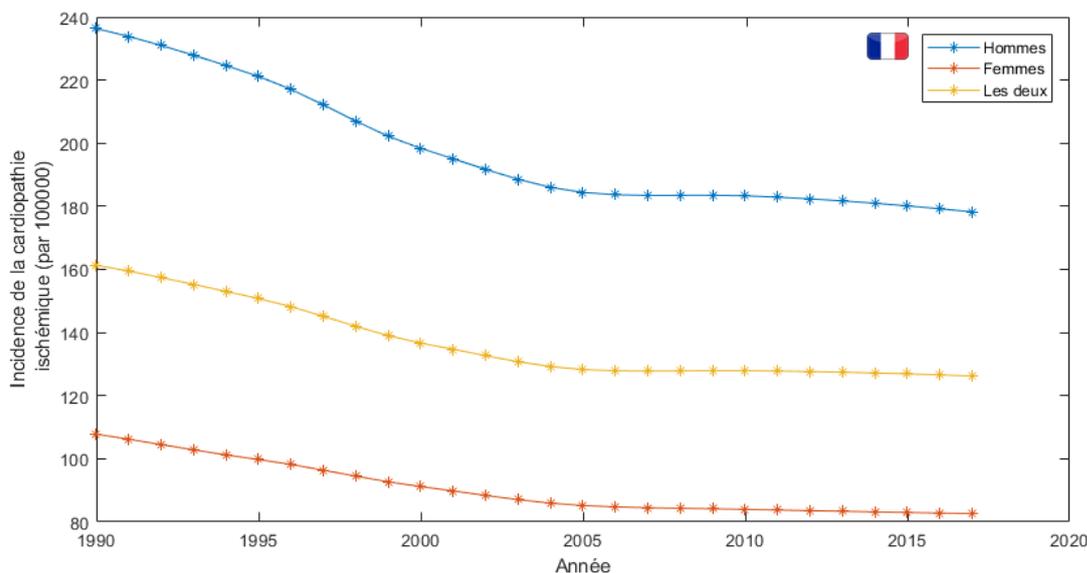


FIGURE 1.19 – Incidence standardisée selon l'âge de la cardiopathie ischémique en France de 1990 à 2017 [2].

La FIGURE 1.20 montre l'évolution de l'incidence standardisée selon l'âge au Liban de la cardiopathie ischémique de 1990 à 2017. On remarque qu'au Liban aussi, les hommes sont plus touchés par la cardiopathie ischémique que les femmes mais aussi que la prévalence décroît légèrement avec le temps en moyenne et que, contrairement à la prévalence pour les femmes qui est pratiquement stable, celle des hommes décroît de façon plus prononcée à partir de 2010.

La prévalence standardisée selon l'âge par 100000 habitants de la cardiopathie ischémique en 2017 était de 1366 pour les femmes et de 2350 pour les hommes dans les pays développés alors qu'elle était de 2067 pour les femmes et de 2920 pour les hommes dans les pays sous-développés. Sur l'ensemble des pays membres de l'ESC, elle était de 1678 chez les femmes et de 2603 chez les hommes.

### L'accident vasculaire cérébral

Un accident vasculaire cérébral (AVC) survient lorsque l'apport sanguin à une partie du cerveau est interrompu ou réduit, empêchant le tissu cérébral d'obtenir de l'oxygène et des nutriments. Les cellules cérébrales commencent à mourir en quelques minutes. Un AVC est une urgence médicale et un traitement rapide est crucial. Une action précoce peut réduire les lésions cérébrales et d'autres complications. La bonne nouvelle est que beaucoup moins de personnes meurent d'un AVC maintenant que par le passé. Des traitements efficaces peuvent également aider à prévenir l'invalidité d'un AVC. En 2017, près de 110 milles cas d'AVC ont été recensés en France, 9.15 milles cas au Liban et 2.31 millions dans l'ensemble des pays membres de l'ESC ce qui correspond à une moyenne de 42.8 milles cas par pays. La FIGURE 1.21 montre la répartition selon le sexe des incidences de l'AVC en 2017. Les secteurs de la FIGURE 1.21 montrent qu'il y a plus de femmes ayant connu un AVC en 2017 que d'hommes en général. Au Liban, bien plus de

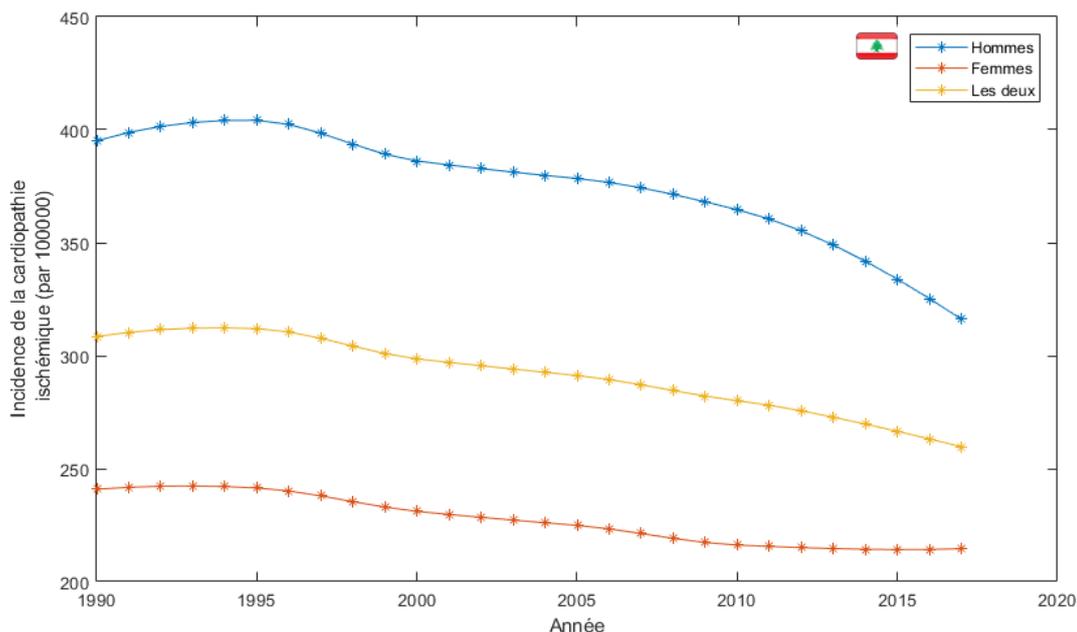


FIGURE 1.20 – Incidence standardisée selon l’âge de la cardiopathie ischémique au Liban de 1990 à 2017 [2].

femmes sont atteintes par l’AVC que les hommes et en France, il y a presque autant d’hommes que de femmes ayant vécu un AVC en 2017.

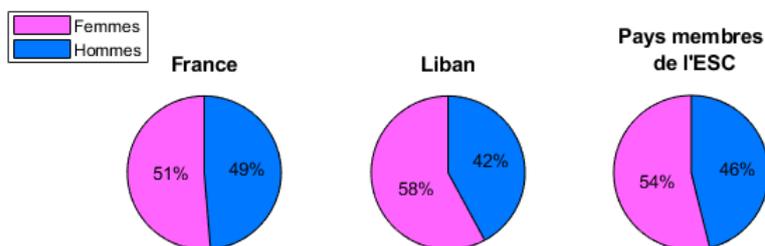


FIGURE 1.21 – Incidence de l’accident vasculaire cérébral en 2017 [2].

La FIGURE 1.22 montre l’évolution de l’incidence standardisée selon l’âge en France de l’accident vasculaire cérébral de 1990 à 2017. On remarque que de moins en moins de personnes connaissent un AVC de 1990 à 2017, et plus précisément, cette décroissance du taux d’incidence se stabilise chez les femmes à partir de 2000, alors que chez les hommes, elle se stabilise en 2005 pour reprendre plus légèrement cette fois en 2010. On remarque aussi que, malgré le même nombre de nouveaux cas d’AVC en 2017, les hommes ont un taux d’incidence plus élevé que celui des femmes ce qui s’explique par la répartition inégale des sexes au sein de la population.

La FIGURE 1.23 montre l’évolution de l’incidence standardisée selon l’âge au Liban de l’accident vasculaire cérébral de 1990 à 2017. On remarque une décroissante générale de l’incidence de l’AVC qui s’arrête en 2005 pour reprendre une croissance jusqu’en 2017. Le taux passe de 167.5 en 1990 jusqu’à 138.4 en 2005, puis remonte à 152 en 2017. On remarque que les femmes sont plus touchées que les hommes par l’AVC au Liban.

La prévalence standardisée selon l’âge par 100000 habitants de l’accident vasculaire cérébral en 2017 était de 1007 pour les femmes et de 1147 pour les hommes dans les pays développés alors qu’elle était de 1517 pour les femmes et de 1539 pour les hommes dans les pays sous-développés. Sur l’ensemble des pays membres de l’ESC, elle était de 1234 chez les femmes et de 1321 chez les hommes.

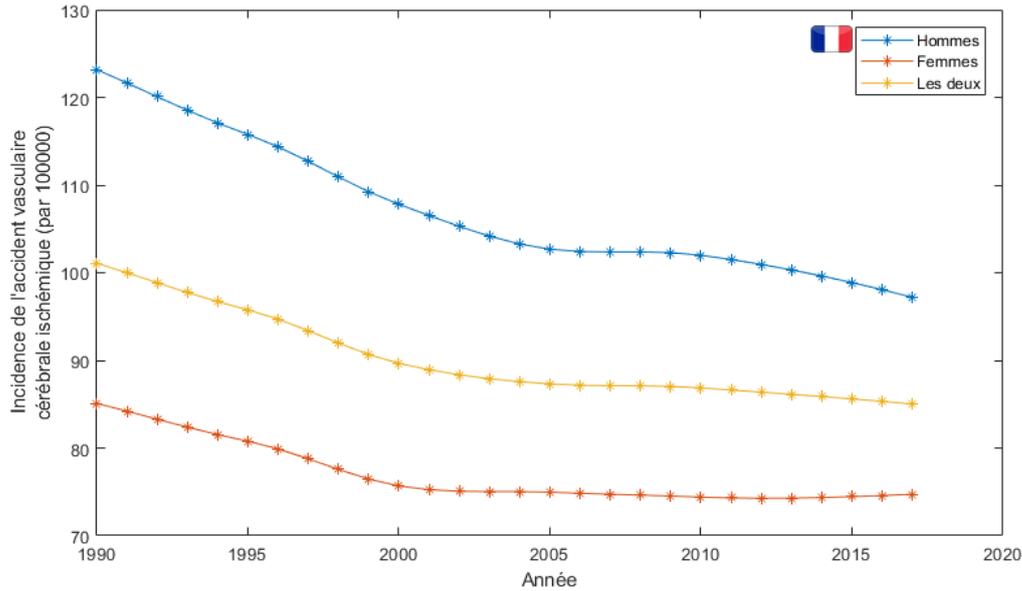


FIGURE 1.22 – Incidence standardisée selon l’âge de l’accident vasculaire cérébral en France de 1990 à 2017 [2].

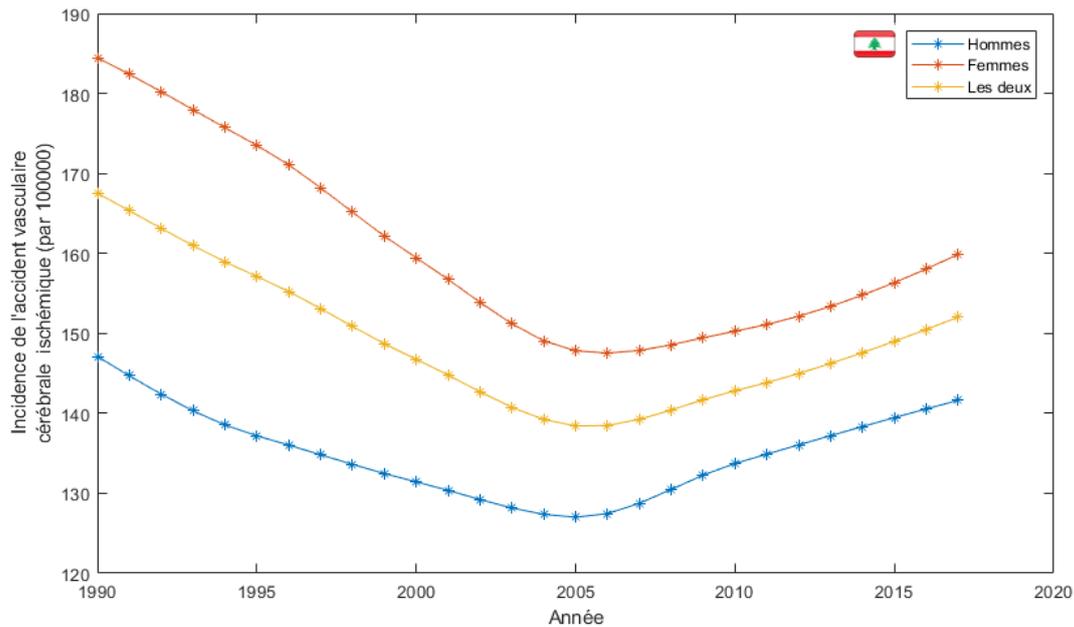


FIGURE 1.23 – Incidence standardisée selon l’âge de l’accident vasculaire cérébral au Liban de 1990 à 2017 [2].

### La maladie vasculaire périphérique

La maladie vasculaire périphérique (PAD) est une affection courante où une accumulation de dépôts gras dans les artères restreint l’apport sanguin aux muscles des jambes. Elle est également connue sous le nom de maladie artérielle périphérique (MAP). Elle est généralement causée par une accumulation de dépôts gras dans les parois des artères des jambes. Les dépôts gras sont constitués de cholestérol et d’autres déchets. L’accumulation de dépôts gras sur les parois des artères rend les artères plus étroites et restreint le flux sanguin vers les jambes. En 2017, près de 150.19 mille cas de maladies vasculaires périphériques ont été recensés en France, 8.11 mille cas au Liban et 2.161 millions dans l’ensemble des pays membres de l’ESC ce qui cor-

respond à une moyenne d'environ 40 mille cas par pays. La FIGURE 1.24 montre la répartition selon le sexe des incidences des maladies vasculaires périphériques en 2017. Les secteurs de la FIGURE 1.24 montrent qu'en 2017, on a enregistré plus de cas de maladies vasculaires périphériques chez les hommes que chez les femmes en France, contrairement au Liban. Sur l'ensemble des pays membre de l'ESC, la répartition des nouveaux cas enregistrés en 2017 selon les sexes est presque équitable.

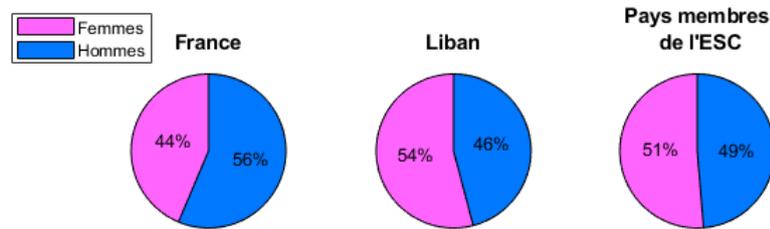


FIGURE 1.24 – Incidence des maladies vasculaires périphériques en 2017 [2].

La FIGURE 1.25 montre l'évolution de l'incidence standardisée selon l'âge en France des maladies vasculaires périphériques de 1990 à 2017. On remarque une décroissance monotone de la courbe d'incidence sur toute la période qui varie de 151.9 en 1990 jusqu'à 128.4 en 2017, ainsi qu'une large supériorité de l'incidence des maladies vasculaires périphériques chez les hommes sur les femmes.

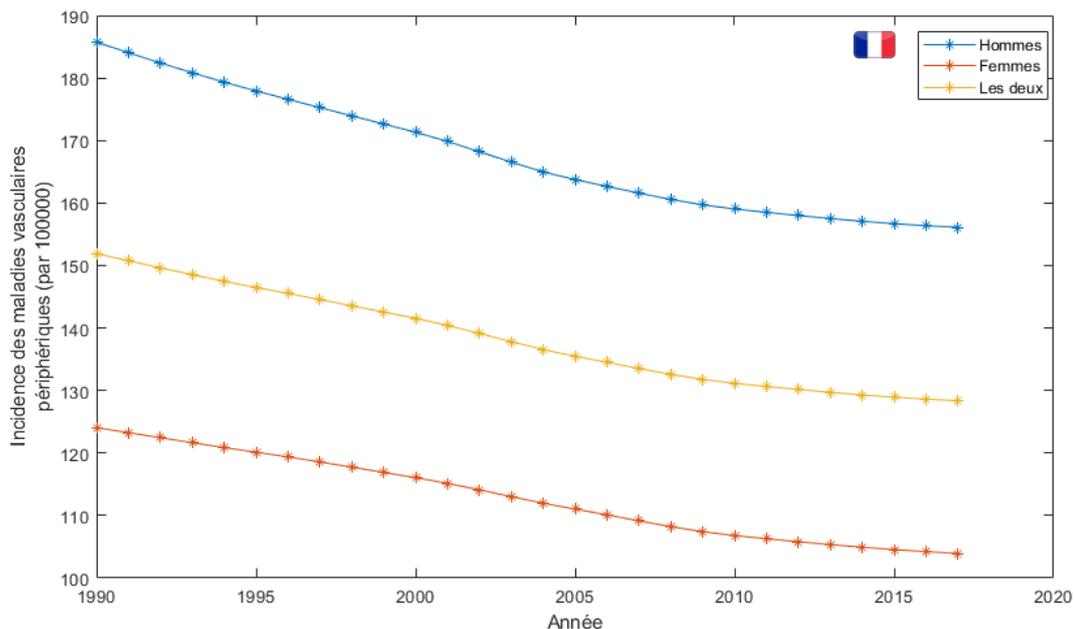


FIGURE 1.25 – Incidence standardisée selon l'âge des maladies vasculaires périphériques en France de 1990 à 2017 [2].

La FIGURE 1.26 montre l'évolution de l'incidence standardisée selon l'âge au Liban des maladies vasculaires périphériques de 1990 à 2017. Les courbes d'incidence au Liban présentent des caractéristiques variées. D'abord, les femmes ont un taux d'incidence plus élevé que celui des hommes. Ensuite, la courbe des femmes décroît fortement de 1990 à 2000 puis se stabilise sur une croissance légère jusqu'à 2017. La courbe des hommes décroît aussi fortement jusqu'en 2005 où elle recroît fortement jusqu'en 2010 et finit sur une croissance faible jusqu'à 2017. La courbe générale passe de 140.1 en 1990, par 132.3 en 2000, 131.2 en 2005 et enfin par 134.2 en 2017.

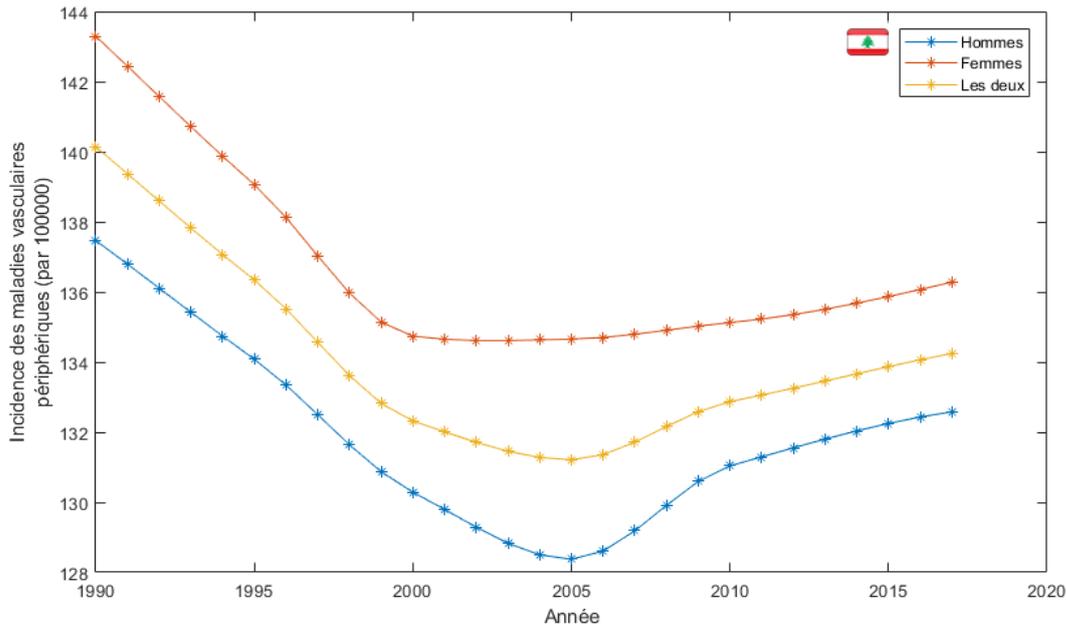


FIGURE 1.26 – Incidence standardisée selon l'âge des maladies vasculaires périphériques au Liban de 1990 à 2017 [2].

La prévalence standardisée selon l'âge par 100000 habitants des maladies vasculaires périphériques en 2017 était de 1395 pour les femmes et de 1722 pour les hommes dans les pays développés alors qu'elle était de 1435 pour les femmes et de 1364 pour les hommes dans les pays sous-développés. Sur l'ensemble des pays membres de l'ESC, elle était de 1413 chez les femmes et de 1563 chez les hommes.

## 1.4 Présentation de la problématique

### 1.4.1 E-santé

Au cours des années 90, alors qu'Internet explose au grand public, un certain nombre de termes électroniques ont fait leur apparition et ont proliféré. Parmi ces termes, on trouve l'e-mail qui a apporté de nouvelles possibilités de communication permettant de contacter rapidement et de partager. On trouve aussi l'e-commerce qui représente le commerce électronique proposant de nouvelles façons de mener les transactions commerciales et financières via Internet. L'introduction de l'e-santé représentait la promesse des technologies de l'information et de la communication pour améliorer la santé et le système de santé [43]. Le terme e-santé est devenu aussitôt un terme indispensable. Comme pour la plupart des néologismes, la signification précise de l'e-santé variait avec le contexte dans lequel le terme a été utilisé. Néanmoins, il a été assez bien compris et est maintenant largement utilisé par de nombreux établissements universitaires, organismes professionnels et organisations. Dans [44], l'auteur reconnu l'impossibilité de trouver une définition formelle universellement acceptable et universellement applicable. Il est encore estimé qu'une compréhension plus claire du terme pourrait être atteinte en examinant la gamme des significations proposées. Deux articles ont traité la question de savoir comment l'e-santé peut être définie [45; 46].

Dans le cadre de cette thèse, nous abordons l'e-santé dans le but de développer un système de surveillance médicale pour le diagnostic et le pronostic placé à l'intérieur du cadre de vie d'un sujet à surveiller. Ce dernier peut se résumer à son lieu d'habitation ou s'étendre sur l'ensemble des lieux qui régissent la vie quotidienne (supermarché, restaurant, parc, etc.). Le projet dont découle cette thèse s'adresse au domaine médical et plus particulièrement à la surveillance des patients

ayant des antécédents cardiaques dans leur cadre de vie. Le projet vise à améliorer la capacité de surveillance des patients en dehors du cadre des infrastructures médicales sans imposer de conditions contraignantes sur le confort de l'utilisateur comme le port d'équipements lourds ou encombrants. Ce projet est parfaitement adapté au développement des systèmes de surveillance continue basés sur les bracelets connectés qui servent d'appareils de mesure médicale. Les travaux de recherche associés à ce projet sont axés vers le développement d'algorithmes capables d'analyser le niveau de risque ou les types de dangers potentiels qui évoluent chez la personne concernée. Le support (hardware) qui s'occupe de lancer ces algorithmes ne sont pas inclus dans les travaux de recherche si bien que plusieurs types de systèmes embarqués peuvent utiliser et appliquer l'essence des algorithmes développés au sein de ce projet.

### 1.4.2 Problématique

Ce projet de thèse se focalise essentiellement sur la surveillance médicale de la santé cardiaque des patients à l'extérieur du cadre des centres chargés de services médicaux. Un ou plusieurs appareils portables et connectés sont alors chargés de réaliser des mesures physiologiques sur le sujet en état de surveillance. Les données recueillies par ces appareils sont alors traitées dans le but d'évaluer le risque potentiel des complications cardiaques et d'en prévenir le corps chargé de l'intervention médicale. Le but est donc de prédire et anticiper des épisodes, des complications ou des troubles cardiaques en se basant sur la lecture des signaux physiologiques qui seront mesurées de manière non-invasive et tout le long de la journée. On entend par mesure non-invasive tout procédé de mesure qui ne demande ni électrodes ni équipements fixes. Ainsi, les signaux physiologiques utilisés comme données d'entrée pour les algorithmes développés dans le cadre de cette thèse sont sélectionnés en se basant sur les technologies des équipements et procédés de mesure pour chaque signal physiologique. Le système de surveillance a pour objectif de détecter l'arrivée d'un malaise ou d'une évolution d'une complication cardiaque et d'en prévenir les unités médicales adéquates pour une intervention prématurée.

Dans [47], l'auteur aborde le développement de dispositifs de surveillance portables avancés qui est un élément essentiel dans la prochaine étape de diagnostic et traitement des maladies cardio-vasculaires. Ces défis seront relevés par des avancées dans des domaines de la nanotechnologie, les sciences des matériaux et les technologies de l'information, et leur application redéfinira l'approche du clinicien en matière d'examen, d'évaluation et de traitement des patients atteints de maladies cardio-vasculaires.

Dans [48], l'auteur décrit l'évolution d'un capteur annulaire qui représente probablement le projet le plus connu dans le domaine des appareils portables. Le résultat de plusieurs années de travail est un capteur d'oxymétrie de pouls qui permet de surveiller en continu la fréquence cardiaque et la saturation en oxygène de manière totalement discrète. L'appareil a la forme d'un anneau et peut donc être porté pendant de longues périodes sans gêne pour le sujet. Le capteur en anneau est équipé d'un émetteur-récepteur de faible puissance qui réalise une communication bidirectionnelle avec une station de base, permettant ainsi de reconfigurer le capteur si nécessaire et de télécharger des données à tout moment. Les signaux physiologiques sont choisis en se basant sur la disponibilité de ce genre de technologie et leur transfert potentiel vers un unique appareil portable connecté.

L'objectif de cette thèse est de développer des algorithmes intelligents de prédiction et de diagnostic de maladies cardio-vasculaires, dont la fibrillation atriale et l'insuffisance cardiaque, grâce à des signaux physiologiques de basses fréquences relevés sur les personnes surveillées dans leur cadre de vie. Un grand nombre d'études se base sur l'analyse de l'électrocardiogramme pour la prédiction et l'anticipation des malaises cardiaques [49–52]. Dans nos études, l'électrocardiographie est une ressource supposée indisponible vu le niveau de confort imposé sur les techniques de mesure des signaux physiologiques.

### 1.4.3 Medical Information Mart for Intensive Care III

Les travaux réalisés dans cette thèse nécessitent des données médicales de patients, telles que leurs états de santé, leurs diagnostics médicaux et leurs signaux physiologiques. La "Medical Information Mart for Intensive Care III" (MIMIC III) est une base de données médicales contenant des informations sur des patients qui ont séjourné dans des unités de soins intensifs (USI) [53]. Il s'agit d'un outil en ligne librement accessible utilisé pour les études de recherche sur les soins de santé. La base de données MIMIC III a fourni toutes les données nécessaires aux bons développement et validation des contributions de cette thèse. Cette base de données est divisée en deux sections. La première est dédiée aux données cliniques telles que la liste des admissions, les diagnostics, les tests de laboratoire, les séjours en USI ou les procédures médicales. Cette section est principalement constituée de tables connectées via des clés primaires qui représentent les identifiants de différentes structures telles que les patients, les admissions, les procédures, etc. Les patients sont anonymisés à des fins de confidentialité et toutes les dates et heures associées sont décalées d'une constante choisie au hasard pour chaque patient. La deuxième section est composée de données sous forme de séries temporelles qui contiennent essentiellement des enregistrements de signaux physiologiques au cours du temps. Nous pouvons trouver dans cette section des signaux haute résolution échantillonnés à 125 Hz comme l'électrocardiogramme et ceux basse résolution échantillonnés à un point par minute comme le rythme cardiaque ou la fréquence cardiaque et la pression artérielle. Les deux sections ne sont pas liées entre elles ce qui signifie que les patients ne peuvent pas être associés à leur données temporelles. Cependant, une base de données "match" est en cours de construction où une partie de la base de données cliniques est liée à la base de données à séries temporelles. Notre étude ne portera que sur cette partie où les patients sont appariés à leurs signaux enregistrés. Dans cette thèse, nous nous intéressons aux signaux basse résolution car ils ne nécessitent pas de capteurs invasifs et peuvent être mesurés à l'aide d'un dispositif portable rendant la procédure d'acquisition réalisable en dehors du centre médical. Les deux sections sont liées via une clé de référence qui associe les données cliniques des patients avec leurs données tracées.

## 1.5 Organisation du manuscrit et contributions

Ce premier chapitre a pour objectif d'introduire les maladies cardio-vasculaires, d'évaluer les enjeux liés à la prédiction des malaises cardiaques, de familiariser le lecteur avec les termes cardiologiques, d'exposer la distribution de ces maladies dans l'espace et dans le temps, d'en détailler les risques, de situer le cadre des travaux de recherche liés à cette thèse et de présenter brièvement les différentes contributions apportées.

Le deuxième chapitre présente la théorie des fonctions de croyance, qui sera utilisée et étendue dans la majorité des études dans le contexte de la prédiction médicale à partir des signaux physiologiques. Cette théorie, nommée également la théorie de Dempster-Shafer, est une extension de l'approche probabiliste qui permet de quantifier l'incertitude et qui sera particulièrement utile lors de l'absence de certains signaux physiologiques.

Le troisième chapitre présente un ensemble de contributions apportées sur le sujet de la prédiction de la fibrillation atriale à l'aide de la théorie des fonctions de croyance. C'est une complication cardiaque très répandue et sa prédiction prématurée permet la prédiction d'autres maladies cardiaques. Dans ce chapitre nous proposons des approches originales basées sur la Théorie des Fonctions de Croyance ou TFC pour la classification des signaux en fonction de l'existence ou non de la FA. L'utilisation de la TFC sur les caractéristiques retenues nous permet une modélisation adéquate de l'information disponible, en tenant compte de la forte variabilité de la disponibilité des signaux, et la précarité de certaines données disponibles. Les travaux menés dans ce chapitre permettent l'orientation des travaux de recherche vers la prédiction multi-pathologiques cardiaques. Les travaux de recherches liés à la prédiction de cette pathologie ont été à la source de trois publications scientifiques dont un papier journal en deuxième révision, une conférence internationale et une nationale [54–56].

Le quatrième chapitre propose une extension des travaux de recherche liés à la prédiction de la fibrillation atriale. Une nouvelle modélisation de la fonction de masse liée à chaque caractéristique est présentée. Cette approche permet de mieux modéliser l'incertitude associée à chaque caractéristique. De plus, la méthode proposée considère certaines propriétés des données liées à leur récurrence, leur disponibilité, et leur pouvoir discriminatif. Dans ce chapitre, nous proposons aussi une méthode associative pour la combinaison des fonctions de masse. Cette méthode permet de remplacer la méthode de combinaison non associative de Yager [57]. En effet, l'utilisation d'une fonction de masse non associative peut dégrader les performances du fait que le résultat de la combinaison dépend dans ce cas de l'ordre de combinaison des sources choisi. Finalement, une dernière contribution de ce chapitre est l'utilisation d'une option de rejet de classification quand les données ne sont pas assez fiables pour faire une bonne classification. Ces travaux ont fait l'objet de deux publications dans deux conférences internationales [58; 59] et d'une revue en cours de rédaction.

Le cinquième chapitre introduit le concept de la classification multi-label. Ce type de classification sera utilisé pour le diagnostic multi-pathologique en considérant les traits caractéristiques de chaque pathologie comme des étiquettes ou des labels. Une approche multi-label originale qu'on a validé grâce aux données extraites de la base médicale MIMIC III a été proposée. L'approche développée se situe dans la continuité des travaux précédents basés sur la théorie des fonctions de croyance. Ces travaux ont fait l'objet d'une publication dans une conférence internationale [60].

Le sixième chapitre expose une contribution d'ordre théorique apportée dans le domaine de la classification multi-label. La méthode vise à améliorer les performances de la prédiction multi-label en favorisant surtout la mise en oeuvre de solutions pour des problèmes de classification dont la prédisposition et la structure des données sont similaires à celle établies par le sujet de la prédiction multi-pathologique cardiaque. Une méthode originale appelée CoLEx pour "Correlation based Labelsets by Exchange strategy", est présentée. Cette méthode est inspirée par la méthode RAKEL qui utilise des sous ensembles de k-labelsets. Le méthode CoLEx optimise le choix des k-labelsets en considérant les corrélations existantes entre les labels. Pour cela, nous avons utilisé un indicateur de corrélation pour choisir la meilleure combinaison de k-labelsets. Des expériences sur des ensembles de données du monde réel montrent l'efficacité de notre méthode par rapport à la méthode RAKEL et à d'autres méthodes de pointe de la littérature. Ces travaux ont fait l'objet d'une publication soumise dans un journal international [61].

Enfin, la conclusion traite et rassemble l'ensemble des contributions réalisées par les travaux de recherche dans le cadre de cette thèse et contextualise la chronologie des apports scientifiques. Il présentera aussi les perspectives de la classification multi-label appliquée au domaine de la prédiction multi-pathologie pour un ensemble donné de maladies cardiaques.



# Chapitre 2

## Théorie des fonctions de croyance

### Sommaire

---

<b>2.1 Introduction</b> . . . . .	<b>36</b>
<b>2.2 Utilité de la théorie et définitions</b> . . . . .	<b>36</b>
2.2.1 Approche probabiliste et ses limitations . . . . .	36
2.2.2 Approche évidentielle et fonctions de masse . . . . .	37
<b>2.3 Fonctions indicatives</b> . . . . .	<b>39</b>
2.3.1 Fonction de croyance . . . . .	39
2.3.2 Fonction de plausibilité . . . . .	40
2.3.3 Fonction de communalité . . . . .	41
2.3.4 Valeurs des fonctions pour les exemples . . . . .	41
<b>2.4 Méthodes de combinaison de sources</b> . . . . .	<b>42</b>
2.4.1 Méthode conjonctive . . . . .	43
2.4.2 Méthode disjonctive . . . . .	44
2.4.3 Méthode de Dempster . . . . .	44
2.4.4 Méthode de Yager . . . . .	45
<b>2.5 Affaiblissement</b> . . . . .	<b>45</b>
<b>2.6 Prise de décision</b> . . . . .	<b>47</b>
<b>2.7 Conclusion</b> . . . . .	<b>48</b>

---

## 2.1 Introduction

La théorie des fonctions de croyance est un outil mathématique qui permet de modéliser, tout comme la théorie des probabilités, un état de connaissance pour un sujet précis [62]. Cette modélisation passe par l'évaluation des croyances, grandeur analogue à la probabilité propre à la théorie des fonctions de croyance, et sa mise en valeur. Le principal avantage de cette théorie est la possibilité de modéliser une connaissance même imprécise. La capacité à modéliser des informations imprécises permet de mieux évaluer les probabilités lorsque plusieurs sources d'informations mettent en commun leur données. Dans ce chapitre, on abordera en premier lieu la différence entre les approches probabilistes et les approches évidentielles, i.e. issues de la théorie des fonctions de croyance. Des exemples seront présentés pour illustrer l'avantage de cette théorie. Ensuite, les fonctions indicatives, qui constituent la base de la représentation mathématique de la connaissance acquis du problème, seront présentées avec leur différentes significations, les liens entre elles et leurs valeurs pour les mêmes exemples donnés. Par la suite, on présentera les méthodes de combinaison de sources d'informations utilisées lorsque plusieurs entités informelles mettent en commun leur contribution. Ceci permet la fusion de données issues de plusieurs sources d'information comme des capteurs par exemple. Puis, un processus nommé affaiblissement sera présenté dans la section suivante. Cette étape permet de pondérer précisément et selon les fiabilités des sources d'information la contribution de chaque source améliorant ainsi la combinaison. Finalement, dans le but de prendre une décision, une transformation de la représentation dictée par la théorie des fonctions de croyance en une approche probabiliste est nécessaire. Cette méthode de transformation sera détaillée et appliquée sur les exemples présentés en début de chapitre.

## 2.2 Utilité de la théorie et définitions

Cette section présente la théorie des fonctions de croyance en la comparant à la théorie des probabilités. Elle en démontre les avantages et discute des nouvelles fonctionnalités qu'offre cette théorie face à celle des probabilités en détaillant les limitations franchies par l'utilisation de celle-ci. Ainsi, le contexte de la présentation de la théorie requiert une introduction aux différents termes et notations scientifiques utilisés. On considère un ensemble de  $n$  réponses potentielles  $e_i$ ,  $1 \leq i \leq n$ , éligibles à un certain problème posé, noté

$$\Omega = \{e_1, e_2, e_3 \dots e_n\}.$$

L'ensemble  $\Omega$  est supposé exhaustif et ses éléments représentant des hypothèses sont considérés exclusifs. L'ensemble de tous les sous-ensembles de  $\Omega$  est noté  $2^\Omega$  et contient tous les singletons de  $\Omega$  mais aussi toutes les disjonctions conductibles à partir des éléments de  $\Omega$ ,

$$2^\Omega = \{\emptyset, \{e_1\}, \{e_2\}, \dots, \{e_n\}, \{e_1, e_2\}, \dots, \Omega\}.$$

On notera par la suite  $P(e_i)$  la probabilité associée à  $e_i$  et on considère l'hypothèse du monde fermé qui stipule qu'au moins l'un des éléments  $e_i$  est valide, c'est-à-dire que  $\Omega$  est exhaustif sans perte de généralité, car dans le cas de l'hypothèse du monde ouvert, il est possible de considérer un élément supplémentaire qui a la propriété d'être valide lorsque tous les éléments de base sont invalides, ce qui nous ramène au cas du monde fermé. Ainsi, en se basant sur l'exhaustivité de  $\Omega$  et l'exclusivité de ses éléments, on a

$$\sum_{1 \leq i \leq n} P(e_i) = 1.$$

### 2.2.1 Approche probabiliste et ses limitations

Pour mettre en évidence les limitations de l'approche probabiliste, plusieurs exemples seront présentés où le but est de modéliser les informations acquises et de déduire les probabilités des

différents résultats potentiels.

**Exemple 2.2.1.** On considère un jeu qui consiste à lancer une pièce de monnaie et selon le résultat, un dé ou un tir à l'arc est lancé. Si la pièce tombe sur pile, le joueur devra lancer un dé parfaitement équilibré et remportera un petit lot s'il obtient un six. Si la pièce tombe sur face, il devra faire du tir à l'arc. Si le joueur met dans le mil, il remporte le gros lot, sinon il remporte le petit lot. N'ayant aucune idée de la difficulté du tir à l'arc, de la dextérité du joueur ni de l'usure du matériel de jeu, la modélisation probabiliste de l'issue du tir à l'arc devient compliquée. Ainsi la FIGURE 2.1 montre l'arbre

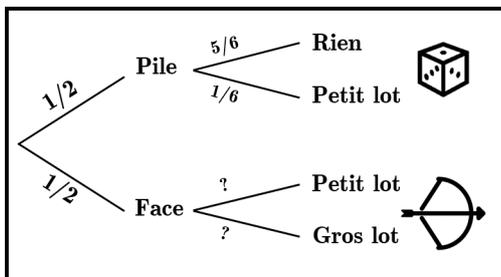


FIGURE 2.1 – Arbre de probabilité du jeu de l'exemple 2.2.1.

de probabilité des différentes issues possibles de ce jeu. On en déduit que l'univers des événements peut être modélisé par un ensemble  $\Omega$  à 3 éléments tel que  $\Omega = \{R, P, G\}$ . R représente l'évènement de ne rien remporter, P représente celui de remporter le petit lot et G celui de remporter le gros lot. La probabilité de jouer au dé et d'obtenir un chiffre inférieur à 6 est de  $5/12$ , et celle d'obtenir un 6 est de  $1/12$ . La probabilité de jouer à l'arc est de  $1/2$ , par contre, aucune information n'est disponible sur la probabilité de mettre la flèche dans le mil. L'approche probabiliste ne nous permet donc pas de modéliser l'ensemble des informations connues. En revanche on sait que  $P(R) = \frac{5}{12}$ ,  $P(P) = \frac{1}{12} + \frac{1-p}{2}$  et  $P(G) = \frac{p}{2}$  avec  $p$  la probabilité de mettre dans le mil. C'est exactement ce manque d'information qui sera modélisé numériquement grâce à un modèle mathématique sur lequel est fondée la théorie des fonctions de croyance sans introduire des variables supplémentaires.

**Exemple 2.2.2.** L'expérience aléatoire de cet exemple consiste à déterminer le phénotype d'un individu concernant un gène en particulier qui ne prend que deux allèles possibles, un récessif et l'autre dominant. L'un des parents de cet individu a un phénotype récessif alors que l'autre en a un dominant. On rappelle qu'un phénotype est récessif lorsque les deux allèles du gène sont récessifs et qu'il est dominant lorsque l'un des deux allèles est dominant. Dans ce cas, l'un des allèles du gène de l'individu qui est celui en provenance du parent ayant le phénotype récessif est de type récessif. L'autre allèle détermine directement le phénotype de l'individu. Le parent ayant le phénotype dominant peut soit avoir un allèle dominant et un autre récessif soit avoir les deux allèles dominants. Le manque d'information dans cet exemple provient du gène du parent qui a le phénotype dominant. On est certain que l'un des deux allèles est dominant mais l'autre peut être récessif comme dominant. Le principe du raisonnement insuffisant nous pousse à estimer la probabilité des deux événements à 50%. Cependant, pour éviter toute supposition précipitée, on considère que la probabilité que le parent 2 ayant le phénotype dominant, possède un allèle récessif est  $p$  avec  $0 \leq p \leq 1$ . De ce fait, on peut calculer la probabilité des événements D et R qui indiquent obtenir une descendance avec phénotype respectivement dominant et récessif. On a alors  $P(D) = 1 - \frac{p}{2}$  et  $P(R) = \frac{p}{2}$ . Le principe du raisonnement insuffisant (PRI) suggère de prendre  $p = 0.5$  et conduit à la distribution de probabilité  $P(D) = 0.75$  et  $P(R) = 0.25$ .

## 2.2.2 Approche évidentielle et fonctions de masse

Dans cette section, on présentera l'approche évidentielle, c'est-à-dire l'approche qui consiste à utiliser la théorie des fonctions de croyance, en l'appliquant aux problèmes posés dans les exemples de la section précédente. Pour cela, il est nécessaire d'introduire la fonction de masse qui est une fonction de base de la théorie des fonctions de croyance permettant d'allouer une

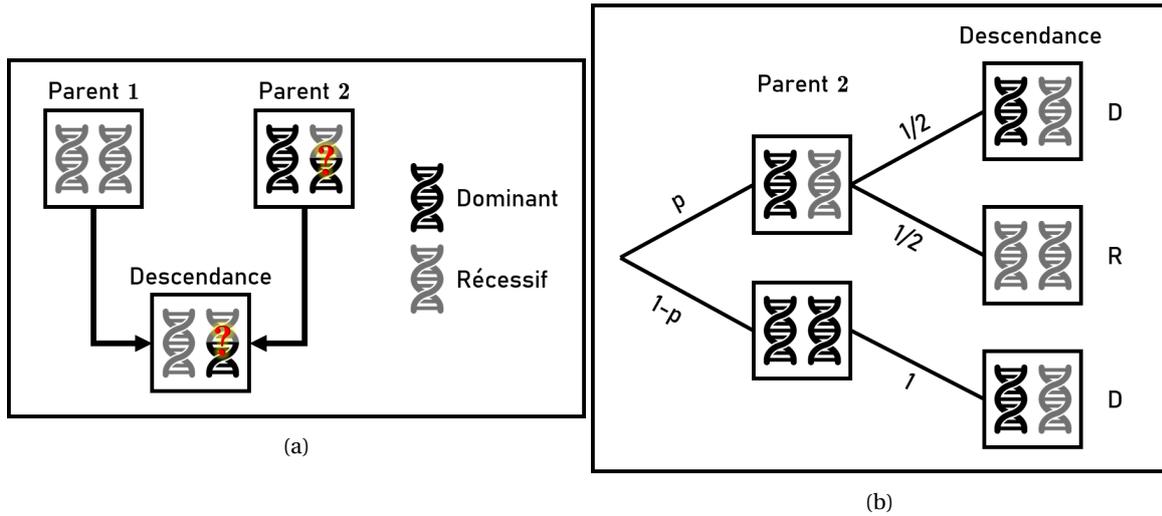


FIGURE 2.2 – Organigramme de l’héritage génétique (a) et arbre de probabilité (b) de l’exemple 2.2.2.

valeur numérique à un ensemble d’évènements. Ainsi, la modélisation évidentielle se démarque de la modélisation probabiliste par la fonction de masse qui attribue une valeur, pas seulement à chaque singleton d’évènement ou d’hypothèse mais à chaque sous-ensemble d’évènements ou d’hypothèses aussi. La fonction de masse est donc formellement définie comme une fonction de  $2^\Omega$  à  $[0;1]$  qui associe à chaque sous-ensemble de  $\Omega$  une valeur numérique réelle représentant la part de croyance attribuée strictement à cet ensemble. On note la fonction de masse  $m(\cdot)$  :

$$m: 2^\Omega \rightarrow [0;1]$$

$$A \mapsto m(A),$$

telle que  $\sum_{A \subseteq \Omega} m(A) = 1$ . Les éléments  $A$  de  $2^\Omega$  tels que  $m(A) \neq 0$  sont appelés éléments focaux de  $m$ . La fonction de masse est souvent liée à une source d’information pouvant être la mesure d’un capteur ou l’avis d’un expert qui définit les valeurs de la fonction pour chaque sous-ensemble  $A \subseteq \Omega$ . On notera que dans l’hypothèse du monde fermé ou *Closed World*, une contrainte est ajoutée  $m(\emptyset) = 0$ .

**Exemple 2.2.3.** Dans le cas du jeu à pile ou face, le lancé du dé et le tir à l’arc, la modélisation en utilisant les fonction de masses permet de prendre en considération l’incertitude liée à l’ignorance des chances de réussite du tir à l’arc. En effet, la masse attribuée au singleton  $\{R\}$  étant la croyance allouée uniquement à R sera la probabilité minimal de R or  $P(R) = \frac{5}{12}$ , donc  $m(\{R\}) = \frac{5}{12}$ . De même, la masse attribuée à  $\{P\}$  sera la probabilité minimale de P, donc  $m(\{P\}) = \frac{1}{12}$ . Finalement, pour les 0.5 restants de masse, ils ne seront pas attribué à un singleton mais à l’ensemble  $\{P,G\}$  puisque cette masse représente la probabilité 0.5 d’obtenir "Face" qui conduit à une part de probabilité répartie entre P et G. La modélisation de l’information peut donc être résumée par les masses des ensembles focaux suivants :  $m(\{R\}) = \frac{5}{12}$ ,  $m(\{P\}) = \frac{1}{12}$  et  $m(\{P,G\}) = \frac{1}{2}$ .

**Exemple 2.2.4.** Dans l’exemple de l’héritage génétique, l’attribution de masse aux singletons étant assimilable à la probabilité minimale pouvant être prise par l’évènement ou l’hypothèse en singleton, il en résulte que les masses de D et de R peuvent être déduites :  $m(\{R\}) = 0$  et  $m(\{D\}) = \frac{1}{2}$ . La masse restante est alors attribuée à l’ensemble des deux évènements,  $m(\{R,D\}) = \frac{1}{2}$ .

La modélisation en utilisant les fonctions de masse permet donc de prendre en compte le manque d’information sans prendre des considérations hâtives. Lorsqu’une partie de la probabilité est répartie sur plusieurs évènements avec des proportions inconnues, elle est attribuée sous forme de masse à l’ensemble de ces évènements. Ainsi, l’équivalence des modélisations entre approches probabiliste et évidentielle peut être établie grâce à des variables supplémentaires qui font office de lacunes d’information. En considérant l’ensemble  $\Omega$  présenté en début de section

2.2, on peut démontrer que :

$$P(e_i) = \sum_{A \subseteq \Omega} \alpha_{e_i}^A m(A) \quad (2.1)$$

avec  $0 \leq \alpha_{e_i}^A \leq 1 \forall e_i \in \Omega, A \subseteq \Omega, \sum_{e_i \in A} \alpha_{e_i}^A = 1$  et  $\alpha_{e_i}^A = 0 \forall e_i \notin A$ . Les variables  $\alpha_{e_i}^A$  représentent la part d'ignorance que l'on a du problème, elles indiquent aussi précisément les proportions de masses des ensembles non-singletons sur l'ensemble des éléments qui les composent. Cette équivalence est ce qui se rapproche le plus de la compréhension formelle de la théorie des fonctions de croyance et plus précisément, l'attribution des masses par la fonction de masse aux différents ensembles d'évènements ou d'hypothèses.

## 2.3 Fonctions indicatives

En plus de la fonction de masse, d'autres fonctions toutes aussi indicatives peuvent être construites à partir de cette dernière. Ces fonctions permettent d'obtenir plusieurs points de vue du même problème et d'étendre le champ compréhensif de l'information en introduisant de nouveaux concepts qualitatifs soutenus par des valeurs numériques.

### 2.3.1 Fonction de croyance

La fonction de croyance indique la croyance réservée à un ou plusieurs évènements, contrairement à la fonction de masse qui exprime la croyance réservée strictement à un ensemble d'évènements. Elle correspond donc à la probabilité minimale que l'un des évènements ou hypothèses de l'ensemble soit valide. On notera cette fonction  $Bel(\cdot)$  et on a donc :

$$Bel(A) = \min_{\alpha} \sum_{e_i \in A} P(e_i), \quad (2.2)$$

et on peut alors déduire en utilisant l'équation 2.1 :

$$Bel(A) = \min_{\alpha} \sum_{e_i \in A} \sum_{A' \subseteq \Omega} \alpha_{e_i}^{A'} m(A') = \min_{\alpha} \sum_{A' \subseteq \Omega} \left( \sum_{e_i \in A} \alpha_{e_i}^{A'} \right) m(A').$$

On peut alors diviser les sous-ensemble  $A' \subseteq \Omega$  en trois parties distinctes où la première rassemble les sous-ensembles  $A'$  inclus dans  $A$ , la deuxième les sous-ensembles  $A'$  entièrement exclus de  $A$  et enfin les sous-ensembles  $A'$  ayant au moins un élément dans  $A$  et un autre dans  $\bar{A}$ . Ceci nous conduit à la forme suivante :

$$\begin{aligned} Bel(A) &= \min_{\alpha} \sum_{A' \subseteq A} \left( \sum_{e_i \in A} \alpha_{e_i}^{A'} \right) m(A') \\ &+ \min_{\alpha} \sum_{A' \subseteq \bar{A}} \left( \sum_{e_i \in A} \alpha_{e_i}^{A'} \right) m(A') \\ &+ \min_{\alpha} \sum_{\substack{A' \cap A \neq \emptyset \\ A' \not\subseteq A}} \left( \sum_{e_i \in A} \alpha_{e_i}^{A'} \right) m(A'), \end{aligned}$$

or on a  $\alpha_{e_i}^{A'} = 0 \forall e_i \notin A'$ , ce qui nous ramène à :

$$\begin{aligned} Bel(A) &= \min_{\alpha} \sum_{A' \subseteq A} \left( \sum_{e_i \in A'} \alpha_{e_i}^{A'} \right) m(A') \\ &+ \min_{\alpha} \sum_{\substack{A' \cap A \neq \emptyset \\ A' \not\subseteq A}} \left( \sum_{e_i \in A' \cap A} \alpha_{e_i}^{A'} \right) m(A'), \end{aligned}$$

et puisque  $\sum_{e_i \in A'} \alpha_{e_i}^{A'} = 1 \forall A' \neq \emptyset$  et  $\alpha_{e_i}^\emptyset = 0 \forall e_i \in \Omega$ , on obtient :

$$\text{Bel}(A) = \sum_{\substack{A' \subseteq A \\ A' \neq \emptyset}} m(A') + \min_{\alpha} \sum_{\substack{A' \cap A \neq \emptyset \\ A' \not\subseteq A}} \left( \sum_{e_i \in A' \cap A} \alpha_{e_i}^{A'} \right) m(A').$$

Puisque les masses et les coefficients  $\alpha$  sont tous positifs, la minimisation de la formule ci-dessus est atteinte lorsque  $\alpha_{e_i}^{A'} = 0 \forall e_i \in A' \cap A$  et donc on finit par obtenir la formule de définition de la croyance :

$$\text{Bel}(A) = \sum_{\substack{A' \subseteq A \\ A' \neq \emptyset}} m(A'). \quad (2.3)$$

La fonction de masse peut aussi être obtenue à partir de la fonction de croyance en considérant la formule suivante :

$$\sum_{A' \subseteq A} (-1)^{|A-A'|} \text{Bel}(A'),$$

avec  $A - A' = A \cap \overline{A'}$  l'ensemble formé des éléments de  $A$  n'appartenant pas à  $A'$ . En remplaçant la fonction de croyance par sa formule en fonction de la répartition des masses, on obtient :

$$\begin{aligned} \sum_{A' \subseteq A} (-1)^{|A-A'|} \sum_{\substack{B \subseteq A' \\ B \neq \emptyset}} m(B) &= \sum_{\substack{A' \subseteq A \\ B \subseteq A' \\ B \neq \emptyset}} (-1)^{|A-A'|} m(B) \\ &= \sum_{\substack{B \subseteq A \\ A' = A - C \\ C \subseteq A - B}} (-1)^{|A-A'|} m(B). \end{aligned}$$

En effet, dans la somme ci-dessus, on considère tous les sous-ensembles  $A'$  de  $A$ , puis tous les sous-ensembles  $B$  de  $A'$ , ce qui est équivalent à considérer tous les sous-ensembles  $B$  de  $A$  puis tous les sous-ensembles  $A'$  de  $A$  englobant  $B$ . Pour cela, on considère le complémentaire de  $A'$  dans  $A$ , qu'on appellera  $C$  tel que  $C \cap B = \emptyset$ . Il suffit alors de considérer tous les sous-ensembles  $C$  de  $A - B$  et les apparier à  $A'$  en appliquant  $A' = A - C$ . On notera aussi que  $A - A' = A - (A - C) = A \cap C = C$ , ce qui nous donne :

$$\sum_{\substack{B \subseteq A \\ C \subseteq A - B}} (-1)^{|C|} m(B) = \sum_{B \subseteq A} \left( \sum_{C \subseteq A - B} (-1)^{|C|} \right) m(B).$$

Pour évaluer la distribution des cardinaux des sous-ensembles de  $A - B$ , on s'intéresse au nombre des sous-ensembles de taille  $i$  en particulier. On sait que le nombre de combinaison de  $i$  éléments dans un ensemble de  $|A - B|$  éléments est de  $C_{|A-B|}^i$ , ce qui prouve que pour  $A - B \neq \emptyset$  :

$$\sum_{C \subseteq A - B} (-1)^{|C|} = \sum_{0 \leq i \leq |A-B|} C_{|A-B|}^i (-1)^i = \sum_{0 \leq i \leq |A-B|} C_{|A-B|}^i (-1)^i (1)^{|A-B|-i} = (1-1)^{|A-B|} = 0,$$

et pour  $A - B = \emptyset$ , c'est-à-dire  $A = B$ , on a  $\sum_{C \subseteq A - B} (-1)^{|C|} = (-1)^0 = 1$ . D'après ce raisonnement, on obtient :

$$m(A) = \sum_{A' \subseteq A} (-1)^{|A-A'|} \text{Bel}(A'). \quad (2.4)$$

### 2.3.2 Fonction de plausibilité

La fonction de plausibilité indique la part de croyance que l'on pourrait assigner à un ensemble d'événements ou d'hypothèses compte tenu des éléments la discréditant. C'est donc la probabilité maximale que l'un des événements ou hypothèses de l'ensemble soit valide. Cette fonction est notée  $Pl(\cdot)$  et vaut :

$$Pl(A) = \max_{\alpha} \sum_{e_i \in A} P(e_i). \quad (2.5)$$

Grâce à un raisonnement similaire à celui présenté dans la section 2.3.1, on peut démontrer que :

$$Pl(A) = \sum_{\substack{A' \subseteq A \\ A' \neq \emptyset}} m(A') + \max_{\alpha} \sum_{\substack{A' \cap A \neq \emptyset \\ A' \not\subseteq A}} \left( \sum_{e_i \in A' \cap A} \alpha_{e_i}^{A'} \right) m(A'),$$

or on a  $\sum_{e_i \in A'} \alpha_{e_i}^{A'} = 1 \forall A' \neq \emptyset$ , donc la valeur maximale que peut prendre  $\sum_{e_i \in A' \cap A} \alpha_{e_i}^{A'}$  est 1, ce qui nous amène à :

$$Pl(A) = \sum_{\substack{A' \subseteq A \\ A' \neq \emptyset}} m(A') + \sum_{\substack{A' \cap A \neq \emptyset \\ A' \not\subseteq A}} m(A').$$

La formule de définition de la fonction de plausibilité est donc :

$$Pl(A) = \sum_{\substack{A' \cap A \neq \emptyset \\ A' \neq \emptyset}} m(A'). \quad (2.6)$$

Sachant que  $\sum_{A' \subseteq \Omega} m(A') = 1$ , on a

$$\sum_{\substack{A' \cap A \neq \emptyset \\ A' \neq \emptyset}} m(A') + \sum_{\substack{A' \subseteq \bar{A} \\ A' \neq \emptyset}} m(A') + m(\emptyset) = 1,$$

ce qui prouve que

$$Pl(A) + Bel(\bar{A}) = 1 - m(\emptyset).$$

Étant dans l'hypothèse du monde fermé ( $m(\emptyset) = 0$ ), on peut alors lier les fonctions de plausibilité et de croyance grâce à la formule suivante :

$$Pl(A) = 1 - Bel(\bar{A}). \quad (2.7)$$

### 2.3.3 Fonction de communalité

La fonction de communalité représente la part de croyance commune qui peut être allouée à chaque élément d'un ensemble d'évènements ou d'hypothèses. C'est donc la part commune de probabilité que s'échange les différents évènements ou hypothèses d'un même ensemble. La communalité d'un sous-ensemble représente donc la somme de toutes les masses d'ensemble contenant ce sous-ensemble. On notera cette fonction  $q(\cdot)$  et elle est définie par :

$$q(A) = \sum_{A' \supseteq A} m(A'). \quad (2.8)$$

On peut aussi, par un raisonnement analogue à celui présenté en section 2.3.1, retrouver la masse d'un ensemble d'évènements ou d'hypothèses à partir de la fonction de communalité :

$$m(A) = \sum_{A' \supseteq A} (-1)^{|A' - A|} q(A'). \quad (2.9)$$

### 2.3.4 Valeurs des fonctions pour les exemples

Dans cette section seront présentées les valeurs des fonctions indicatives définies ci-dessus pour chacun des deux exemples 2.2.1 et 2.2.2.

**Exemple 2.3.1.** Dans le cas du jeu à pile ou face présenté dans l'exemple 2.2.1, les éléments focaux

sont présentés comme suit :

$$m(\{R\}) = \frac{5}{12}$$

$$m(\{P\}) = \frac{1}{12}$$

$$m(\{P,G\}) = \frac{1}{2}$$

Le TABLEAU 2.1 présente les valeurs des différentes fonctions indicatives pour l'exemple 2.2.1.

TABLEAU 2.1 – Tableau récapitulatif des valeurs des fonctions indicatives pour l'exemple 2.2.1.

	$\emptyset$	$\{R\}$	$\{P\}$	$\{G\}$	$\{R,P\}$	$\{P,G\}$	$\{R,G\}$	$\{R,P,G\}$
$m(\cdot)$	0	5/12	1/12	0	0	1/2	0	0
$Bel(\cdot)$	0	5/12	1/12	0	1/2	7/12	5/12	1
$Pl(\cdot)$	1	5/12	7/12	1/2	1	7/12	11/12	1
$q(\cdot)$	1	5/12	7/12	1/2	0	1/2	0	0

**Exemple 2.3.2.** Dans le cas de l'héritage génétique présenté dans l'exemple 2.2.2, les éléments focaux sont présentés comme suit :

$$m(\{D\}) = \frac{1}{2}$$

$$m(\{R,D\}) = \frac{1}{2}$$

Le TABLEAU 2.2 présente les valeurs des différentes fonctions indicatives pour l'exemple 2.2.2.

TABLEAU 2.2 – Tableau récapitulatif des valeurs des fonctions indicatives pour l'exemple 2.2.2.

	$\emptyset$	$\{R\}$	$\{D\}$	$\{R,D\}$
$m(\cdot)$	0	0	1/2	1/2
$Bel(\cdot)$	0	0	1/2	1
$Pl(\cdot)$	1	1/2	1	1
$q(\cdot)$	1	1/2	1	1/2

## 2.4 Méthodes de combinaison de sources

Dans le contexte de la théorie des fonctions de croyance, on modélise une information ou une source d'information par une répartition de masses. Lorsque plusieurs sources d'information au sujet d'un même problème sont disponibles, on obtient plusieurs fonctions de masse qui désignent chacune une partie de la connaissance appréhendée du problème. Une fusion des données est alors nécessaire pour obtenir une fonction de masse unique prenant en considération l'ensemble des autres sources d'information. Les méthodes utilisées dans le cadre de la théorie des fonctions de croyance pour combiner des sources d'information sont appelées les méthodes de combinaison [63] et dans cette section, quatre de ces méthodes seront présentées :

- Méthode de combinaison conjonctive
- Méthode de combinaison disjonctive
- Méthode de Dempster
- Méthode de Yager

Les méthodes de combinaison sont généralement définies par une manipulation mathématique des fonctions de masse issues des sources d'information disponibles. On considère pour simplifier

deux fonctions de masse  $m_1(\cdot)$  et  $m_2(\cdot)$  que l'on souhaite combiner en utilisant les différentes méthodes de combinaison listées ci-dessus. La définition des méthodes de combinaison de masses est basée sur le produit des valeurs des masses entre elles. Ainsi, une nouvelle fonction de masse est habituellement formée en attribuant le produit de chaque couple de masses d'ensemble d'évènements ou d'hypothèses à un autre ensemble. Ceci revient à procéder à la formulation suivante :

$$m_{1*2}(A) = \sum_{B, B' \subseteq \Omega} \beta_{B, B'}^A m_1(B) m_2(B'), \quad (2.10)$$

avec  $\beta_{B, B'}^A \in \{0, 1\}$  et  $\sum_{A \subseteq \Omega} \beta_{B, B'}^A = 1$ . Si  $\beta_{B, B'}^A = 1$ , alors la masse issue du produit  $m_1(B) m_2(B')$  est attribuée à la masse de A dans la fonction de masse  $m_{1*2}(\cdot)$  représentant la fonction de masse obtenue après combinaison des fonctions de masse  $m_1(\cdot)$  et  $m_2(\cdot)$ . Seule la méthode de Dempster fera exception à cette règle.

**Exemple 2.4.1.** *On souhaite évaluer la qualité productive d'un champs de blé en se servant de sa production saisonnière. Pour cela, le produit dérivé est analysé par des machines sur une vaste sélection d'échantillons dont l'une se base sur la couleur du produit par traitement d'image et l'autre sur ses dimensions grâce à une technologie d'évaluation métrique au laser. Ces machines émettent pour chaque unité de produit une décision qui est soit "produit défectueux" que l'on notera D et "produit sain" que l'on notera S. Les machines bien que performantes se trouvent parfois dans l'incapacité de classifier l'échantillon (manque de luminosité pour la première machine, présence de débris volumineux pour la deuxième) ce qui conduit à éviter une classification. Les machines évaluent alors la proportion de produits sains et valables avec un taux d'erreur dû aux échantillons non-classifiés. Les machines donnent les résultats suivant :*

- Machine 1 : 65%  $\pm$  5%
- Machine 2 : 60%  $\pm$  20%

*Dans le but d'émettre une proportion unique prenant en compte les contributions des deux machines, chacune d'elle sera considérée comme une source d'information et la modélisation évidentielle des données offertes par les machines donne lieu aux distributions de masse du TABLEAU 2.3. La combinaison de ces deux fonctions de masses donnent différents résultats selon la méthode de*

TABLEAU 2.3 – Modélisation évidentielle de l'exemple.

	Machine 1	Machine 2
$m(\{D\})$	0.6	0.4
$m(\{S\})$	0.3	0.2
$m(\{D, S\})$	0.1	0.4

*combinaison utilisée comme le montre le TABLEAU 2.4.*

### 2.4.1 Méthode conjonctive

La méthode conjonctive est une méthode de combinaison qui suppose que les sources d'information sont toutes fiables. En effet, la part de croyance issue du produit  $m_1(B) m_2(B')$  représente une portion de probabilité pour laquelle la première source d'information stipule qu'elle devrait

TABLEAU 2.4 – Fonctions de masse issues de différentes méthodes de combinaison.

Méthode	Conjonctive	Disjonctive	Dempster	Yager
$m(\{D\})$	0.52	0.24	0.6842	0.52
$m(\{S\})$	0.2	0.06	0.2632	0.2
$m(\{D, S\})$	0.04	0.7	0.0526	0.28
$m(\emptyset)$	0.24	0	0	0

être distribuée sur les éléments de l'ensemble B et la deuxième source d'information sur les éléments de l'ensemble B'. Comme les deux sources sont supposées fiables, il en résulte que cette part de croyance devra être distribuée sur les éléments en commun de B et de B' et sera donc attribuée à la masse de  $B \cap B'$ . Ainsi,

$$\begin{aligned}\beta_{B,B'}^A &= 1 \text{ si } A = B \cap B', \\ \beta_{B,B'}^A &= 0 \text{ si } A \neq B \cap B' .\end{aligned}$$

On obtient donc la formule explicite de la méthode conjonctive de combinaison des sources d'information modélisées par les fonctions de masse  $m_1(\cdot)$  et  $m_2(\cdot)$  donnant lieu à la nouvelle fonction de masse  $m_{1 \odot 2}(\cdot)$  :

$$m_{1 \odot 2}(A) = \sum_{B \cap B' = A} m_1(B) m_2(B') . \quad (2.11)$$

### 2.4.2 Méthode disjonctive

Contrairement à la méthode conjonctive, la méthode disjonctive est la méthode de combinaison qui suppose que l'une des sources d'information est fiable. En effet, la part de croyance issue du produit  $m_1(B)m_2(B')$  représente une portion de probabilité pour laquelle la première source d'information stipule qu'elle devrait être distribuée sur les éléments de l'ensemble B et la deuxième source d'information sur les éléments de l'ensemble B'. Comme l'une des deux sources est supposée fiable, il en résulte que cette part de croyance devra être distribuée sur les éléments de B et de B' et sera donc attribuée à la masse de  $B \cup B'$ . Ainsi,

$$\begin{aligned}\beta_{B,B'}^A &= 1 \text{ si } A = B \cup B', \\ \beta_{B,B'}^A &= 0 \text{ si } A \neq B \cup B' .\end{aligned}$$

On obtient donc la formule explicite de la méthode disjonctive de combinaison des sources d'information modélisées par les fonctions de masse  $m_1(\cdot)$  et  $m_2(\cdot)$  donnant lieu à la nouvelle fonction de masse  $m_{1 \cup 2}(\cdot)$  :

$$m_{1 \cup 2}(A) = \sum_{B \cup B' = A} m_1(B) m_2(B') . \quad (2.12)$$

### 2.4.3 Méthode de Dempster

La méthode de combinaison de Dempster [64] est similaire à la méthode de combinaison conjonctive à l'exception du rejet des conflits. Le conflit est un terme utilisé dans le cadre de la théorie des fonctions de croyance pour représenter la part de croyance issue du désaccord de deux sources d'information. Il s'agit donc des produits des masses attribuées par la méthode conjonctive à  $\emptyset$ . La méthode de combinaison de Dempster, indiquée ici par le symbole  $\oplus$ , devra donc assurer que  $m_{1 \oplus 2}(\emptyset) = 0$ . La part de croyance normalement allouée à  $\emptyset$  par la méthode conjonctive disparaît donc, ce qui exige une normalisation des autres masses afin de conserver les rapports entre les masses des différents ensembles d'événements ou d'hypothèses. Ainsi, on a :

$$m_{1 \oplus 2}(A) = \frac{m_{1 \odot 2}(A)}{1 - m_{1 \odot 2}(\emptyset)} \quad \forall A \neq \emptyset .$$

La formule explicite de la méthode de combinaison de Dempster se trouve ainsi comme suit :

$$m_{1 \oplus 2}(A) = \begin{cases} \frac{1}{1-K} \sum_{B \cap B' = A} m_1(B) m_2(B') & \text{si } A \neq \emptyset, \\ 0 & \text{si } A = \emptyset, \end{cases} \quad (2.13)$$

avec  $K = \sum_{B \cap B' = \emptyset} m_1(B) m_2(B')$ .

#### 2.4.4 Méthode de Yager

La méthode de combinaison de Yager, bien que moins connue que celle de Dempster, vise aussi à éliminer le conflit. Cependant, au lieu de mettre la masse du conflit à zéro et de normaliser le restant des masses en les multipliant par le facteur adéquat, la méthode de Yager transfère cette masse à celle de l'ignorance absolue, c'est-à-dire à la masse de  $\Omega$ . Ainsi la part de croyance sur laquelle les deux sources se prononcent en créant une impossibilité ou un conflit est considérée comme un manque d'information. La méthode suppose par ailleurs, comme pour la méthode de Dempster, que les sources sont toutes fiables. Ainsi, le conflit est créé lors du produit des masses de deux ensembles qui n'ont aucun élément en commun. Pour les autres cas, la méthode procède tout comme la méthode conjonctive. La distribution des produits de masses revient donc à la formulation suivante :

$$\begin{aligned} \beta_{B,B'}^A &= 1 \text{ si } A = B \cap B' \text{ et } A \neq \Omega, \\ \beta_{B,B'}^A &= 0 \text{ si } A \neq B \cap B' \text{ et } A \neq \Omega, \\ \beta_{B,B'}^\Omega &= 1 \text{ si } B \cap B' = \Omega \text{ ou } B \cap B' = \emptyset, \\ \beta_{B,B'}^\Omega &= 0 \text{ si } B \cap B' \neq \Omega \text{ et } B \cap B' \neq \emptyset. \end{aligned} \quad (2.14)$$

Ainsi, voici comment on obtient la masse combinée d'un ensemble  $A$  à partir de deux fonctions de masses  $m_1(\cdot)$  et  $m_2(\cdot)$  issues de deux sources d'information supposées indépendantes :

$$m_{1 \otimes 2}(A) = \begin{cases} \sum_{B \cap B' = A} m_1(B) m_2(B') & \text{si } A \neq \emptyset \text{ et } A \neq \Omega, \\ 0 & \text{si } A = \emptyset, \\ \sum_{B \cap B' \in \{\emptyset, \Omega\}} m_1(B) m_2(B') & \text{si } A = \Omega. \end{cases} \quad (2.15)$$

## 2.5 Affaiblissement

L'affaiblissement est une étape importante lorsque la fiabilité d'une source d'information est connue [65]. En effet, cette étape permet de pondérer les effets des sources d'information pendant le processus de combinaison. Elle agit en modifiant la fonction de masse des sources d'information selon la fiabilité de chacune. Elle affaiblit les masses des ensembles en attribuant le reste à la masse de l'ensemble univers  $\Omega$ . Puisque la masse de  $\Omega$  représente l'ignorance absolue, ceci a pour effet d'indiquer plus d'ignorance de la part de la source d'information. Un coefficient d'affaiblissement  $\gamma \in [0; 1]$  est utilisé pour modifier la fonction de masse  $m(\cdot)$  en une nouvelle fonction de masse affaiblie  $m'(\cdot)$ . L'affaiblissement procède alors comme suit :

$$\begin{aligned} m'(A) &= (1 - \gamma) m(A) \quad \forall A \neq \Omega, \\ m'(\Omega) &= (1 - \gamma) m(\Omega) + \gamma. \end{aligned} \quad (2.16)$$

On notera que pour  $\gamma = 0$ , aucune modification est apportée à la masse, on suppose alors que la source d'information est totalement fiable. A l'opposé, pour  $\gamma = 1$ , la totalité de la croyance est déplacée vers  $\Omega$  ce qui renvoie à une fonction de masse vide. La fonction de masse vide, qui correspond à une fonction de masse  $m(\cdot)$  tel que  $m(\Omega) = 1$ , est l'élément neutre de toutes les méthodes de combinaison. Ainsi, combiner une fonction de masse avec une autre vide ne changera en rien la distribution des croyances au sein de cette fonction de masse. Elle ne participe donc plus à la fonction de masse issue du processus de combinaison. Un exemple bien connu présenté par Zadeh [66] pour critiquer les limites de la méthode de combinaison de Dempster suggère un cas où le conflit est assez élevé ce qui conduit à une combinaison contre-intuitive.

**Exemple 2.5.1.** On considère un univers de trois événements ou hypothèses  $\Omega = \{e_1, e_2, e_3\}$ . Deux sources d'informations sont disponibles représentées par deux fonctions de masses  $m_1(\cdot)$  et  $m_2(\cdot)$ .

Les éléments focaux de ces fonctions de masse sont :

$$\begin{aligned} m_1(e_1) &= 0.99, \\ m_2(e_3) &= 0.99, \\ m_1(e_2) &= 0.01, \\ m_2(e_2) &= 0.01. \end{aligned}$$

Ainsi, la combinaison selon Dempster conduit à la fonction de masse dont le seul élément focal est :

$$m_{1\oplus 2}(e_2) = 1.$$

Ce résultat est totalement contre-intuitif étant donné que les deux sources d'information témoignent de l'improbabilité de  $e_2$ . Le vrai coupable derrière ce malentendu est le haut niveau de conflit. La fonction de masse  $m_1(\cdot)$  associe à  $e_1$  une masse  $m_1(e_1) = 0.99$  indiquant que l'évènement est très probable alors que la fonction de masse  $m_2(\cdot)$  associe au même élément une masse nulle,  $m_2(e_2) = 0$ , indiquant que l'évènement est impossible. Puisque la méthode de combinaison de Dempster est similaire à la méthode conjonctive, elle suppose que les sources d'information sont toutes fiables, alors l'évènement sera considéré impossible et la masse du conflit sera élevée. La masse du conflit de deux sources d'information est la masse de l'ensemble vide  $\emptyset$  issue de la fonction combinée par la méthode conjonctive. Dans cet exemple, on a la masse de conflit

$$m_{1\otimes 2}(\emptyset) = 0.9999,$$

ce qui explique l'ampleur de l'irrationalisme de la méthode de combinaison de Dempster. Dans ce cas, l'affaiblissement peut être utile pour discréditer les sources d'information. Vu le niveau de conflit, les sources ne peuvent pas être toutes fiables. Le coefficient d'affaiblissement dépend évidemment de la fiabilité de la source d'information mais pour faire simple, on peut les affaiblir avec un coefficient  $\gamma = 0.5$ . Les nouveaux éléments focaux après affaiblissement sont :

$$\begin{aligned} m'_1(e_1) &= 0.495, \\ m'_1(e_2) &= 0.005, \\ m'_1(\Omega) &= 0.5, \\ m'_2(e_3) &= 0.495, \\ m'_2(e_2) &= 0.005, \\ m'_2(\Omega) &= 0.5. \end{aligned}$$

On remarque tout de suite que le taux de conflit a baissé,

$$m'_{1\otimes 2}(\emptyset) = 0.249975,$$

et la fonction de masse obtenue par combinaison selon la règle de Dempster a pour éléments focaux :

$$\begin{aligned} m'_{1\oplus 2}(e_1) &= 0.329989, \\ m'_{1\oplus 2}(e_3) &= 0.329989, \\ m'_{1\oplus 2}(e_2) &= 0.006700, \\ m'_{1\oplus 2}(\Omega) &= 0.333322. \end{aligned}$$

On remarque que les éléments  $e_1$  et  $e_3$ , présentés chacun comme très probables par l'une des sources d'information ont eu une part de croyance de l'ordre de  $1/3$ , alors que l'élément  $e_2$  présenté comme très improbable par chacune des sources d'information a eu une part de croyance négligeable ce qui relève l'incompréhension.

## 2.6 Prise de décision

La modélisation des connaissances selon la théorie des fonctions de croyance est très utile et permet d'exprimer une lacune d'information plus efficacement que par l'approche probabiliste. Cependant, l'approche probabiliste est plus intuitive quant à la prise de décision. L'approche bayésienne stipulant que la meilleure prise de décision revient à décider de l'évènement le plus probable. Pour obtenir une distribution de probabilité à partir de la représentation évidentielle de l'information qui se traduit par une distribution de masses, une règle de transformation est nécessaire [57; 67–69]. La méthode pignistique de transformation de la fonction de masse en une distribution probabiliste repose sur le principe du raisonnement insuffisant. Ceci signifie qu'en cas de croyance attribuée à un ensemble, on choisit de distribuer cette croyance équitablement entre les éléments de cet ensemble pour en déduire la probabilité. En reprenant l'équivalence entre approche probabiliste et approche évidentielle détaillée dans la section 2.2.2, ceci revient à considérer l'égalité de tous les coefficients  $\alpha$ . Ainsi, la méthode pignistique se résume à la condition suivante :

$$\alpha_{e_i}^A = \alpha_{e_j}^A \quad \forall e_i, e_j \in A \quad \forall A \subseteq \Omega, \quad (2.17)$$

et comme on a toujours  $\sum_{e_i \in A} \alpha_{e_i}^A = 1 \quad \forall A \subseteq \Omega$ , on en déduit que

$$\alpha_{e_i}^A = \frac{1}{|A|} \quad \forall e_i \in A \quad \forall A \subseteq \Omega.$$

En appliquant le modèle d'équivalence entre approche probabiliste et approche évidentielle, on obtient la formule de la distribution de probabilité  $P(\cdot)$  obtenue à partir de la fonction de masse  $m(\cdot)$  comme suit :

$$P(e_i) = \sum_{\substack{A \subseteq \Omega \\ e_i \in A}} \frac{m(A)}{|A|} \quad \forall e_i \in A. \quad (2.18)$$

**Exemple 2.6.1.** Dans le cas de l'exemple 2.2.1, les informations connues sur le problème en question sont modélisées comme suit :

$$\begin{aligned} \Omega &= \{R, P, G\}, \\ m(\{R\}) &= 5/12, \\ m(\{P\}) &= 1/12, \\ m(\{P, G\}) &= 1/2. \end{aligned}$$

La méthode de transformation pignistique suggère de diviser la masse de  $\{P, G\}$  en deux parties égales et de les attribuer aux probabilités des évènements  $P$  et  $G$ . On obtient alors la distribution de probabilité suivante :

$$\begin{aligned} P(R) &= 5/12, \\ P(P) &= 1/3, \\ P(G) &= 1/4. \end{aligned}$$

**Exemple 2.6.2.** Dans le cas de l'exemple 2.2.2, les informations connues sur le problème en question sont modélisées comme suit :

$$\begin{aligned} \Omega &= \{R, D\}, \\ m(\{D\}) &= 1/2, \\ m(\{R, D\}) &= 1/2. \end{aligned}$$

En appliquant la méthode pignistique de transformation, on obtient la distribution de probabilité

*suivante :*

$$P(R) = 1/4,$$

$$P(D) = 3/4.$$

Une fois que les probabilités des états sont estimées, la prise de décision consiste à choisir l'état ayant la plus grande probabilité,

$$e^* = \operatorname{argmax}_{e_i \in \Omega} P(e_i), \quad (2.19)$$

tout en ayant un classement des états dans l'ordre décroissant de leurs probabilités.

## 2.7 Conclusion

Dans ce chapitre, la théorie des fonctions de croyance est présentée en commençant par aborder ses avantages sur l'approche probabiliste avec des exemples illustratifs. Puis des fonctions indicatives propres à la théorie sont introduites et appliquées aux exemples. Ensuite, la combinaison des sources d'informations a été présentée en abordant plusieurs méthodes classiques ainsi que l'affaiblissement qui consiste à pondérer les sources selon leur fiabilité. Enfin, la prise de décision est rendu possible grâce à la méthode de transformation pignistique qui ramène à l'approche probabiliste.

## Chapitre 3

# Classification évidentielle pour la prédiction de la fibrillation atriale

### Sommaire

---

<b>3.1 Introduction</b> . . . . .	<b>50</b>
<b>3.2 Base de données médicale MIMIC III</b> . . . . .	<b>50</b>
<b>3.3 Sélection des sujets et des signaux</b> . . . . .	<b>52</b>
<b>3.4 Présentation de l'approche</b> . . . . .	<b>55</b>
<b>3.5 Extraction des caractéristiques</b> . . . . .	<b>55</b>
<b>3.6 Modélisation des fonctions de masse</b> . . . . .	<b>58</b>
3.6.1 Introduction à la classification . . . . .	58
3.6.2 Division de la base de données . . . . .	60
3.6.3 Attribution des masses . . . . .	61
3.6.4 Sélection des caractéristiques . . . . .	62
3.6.5 Combinaison et prise de décision . . . . .	63
<b>3.7 Résultats et discussions</b> . . . . .	<b>66</b>
3.7.1 Illustration des performances . . . . .	66
3.7.2 Impact des paramètres . . . . .	68
3.7.3 Comparaison aux méthodes de la littérature . . . . .	70
<b>3.8 Conclusion</b> . . . . .	<b>71</b>

---

### 3.1 Introduction

Ce chapitre présente une approche originale de classification permettant de prédire l'épisode de Fibrillation Atriale ou FA chez une personne à partir de ses signaux physiologiques. Le problème de prédiction est défini comme un problème de classification, à effectuer en amont de l'occurrence de la FA. Ainsi, le classifieur prend en entrée les signaux physiologiques du patient et prédit si le patient surveillé subira, dans un futur proche, un épisode de FA. Nous rappelons que cette thèse se situe dans le cadre d'un projet de recherche visant à développer un système de suivi médical à domicile, n'imposant aucun changement d'habitude à la personne surveillée. Ainsi, la contrainte posée sur notre approche est le recours à des signaux ne nécessitant que des mesures non-invasives. Il est à noter que le terme «non-invasif» couvre ici «sans électrode» et «sans aucune action recommandée». Par conséquent, des signaux comme l'électrocardiogramme ou la variabilité de la fréquence cardiaque sont considérés comme indisponibles dans notre étude contrairement à certains travaux présentés dans [70–74]. En outre de la complexité du corps humain, et l'absence de signaux très informatifs, les signaux physiologiques peuvent être inaccessibles par moment et contenant des imperfections liées principalement à leurs modes d'acquisition. Ainsi, l'objectif scientifique est la conception d'une approche de classification sous contrainte de données imprécises, hétérogènes, manquantes par moment et indirectement corrélées au phénomène suivi. Cela implique la construction d'un classifieur dont l'entrée est de taille variable et qui fusionne une grande quantité de données hétérogènes moyennement informatives pour en tirer une information efficace sur l'état surveillé. L'objectif applicatif est de pouvoir atteindre des performances de prédiction acceptables en utilisant des dispositifs sans électrode, et plus généralement, tout dispositif multi-fonctionnel, permettant aux personnes d'être prévenues et traitées prématurément lorsqu'un risque est signalé. Ce chapitre propose une première approche, qui extrait les caractéristiques des signaux d'entrée en temps réel et les classe pour indiquer si le sujet subira un épisode de FA ou non. Le classifieur est construit à l'aide de la Théorie des Fonctions de Croyance ou TFC [75–79], en considérant chaque caractéristique comme une source d'information. Une approche novatrice de modélisation de l'information est ainsi proposée, en définissant une fonction de masse originale représentant précisément les connaissances et les lacunes des données. En considérant un modèle par caractéristique, nous surmontons la forte variabilité de la disponibilité des signaux, et le rendons applicable même avec peu de données disponibles. Afin de construire les modèles, la base de données médicale MIMIC III est utilisée en sélectionnant les signaux de patients atteints de FA et les signaux d'autres patients sans FA. Les modèles disponibles sont ensuite combinés conduisant à un modèle plus informatif, qui nous donne une décision en guise de sortie du classifieur. Dans la suite, nous commençons par présenter la base de données MIMIC III et la sélection des sujets et des signaux d'entrée. Nous décrivons ensuite l'approche de classification, comprenant l'extraction de caractéristiques et le développement du classifieur évidentiel. Nous présentons enfin les résultats et la conclusion.

### 3.2 Base de données médicale MIMIC III

La base de données "Medical Information Mart for Intensive Care III" ou MIMIC III est une immense base de données librement accessible comprenant des données sur la santé associées à plus de quarante mille patients anonymisés qui ont séjourné dans des unités de soins intensifs (USI) d'un hôpital entre 2001 et 2012 [80; 81]. La base de données comprend des informations telles que les données démographiques, les mesures des signaux vitaux effectuées au chevet du patient (1 point de données par heure), les résultats des tests de laboratoire, les procédures, les médicaments, les notes du soignant, les rapports d'imagerie et la mortalité (à l'intérieur et à l'extérieur de l'hôpital). La base de données est hébergée sur la plateforme PHISIONET [82]. Elle est divisée en deux grandes parties qui sont les données cliniques et les données temporelles. La première regroupe les données cliniques des patients et contient notamment des informations sur les admissions, les sorties, les diagnostics, les transferts au sein de l'hôpital, le personnel médi-

cal, les procédures médicales, les analyses de laboratoire, les notes des infirmiers, les séjours à l'unité des soins intensifs, etc. Les données cliniques de MIMIC III forment un ensemble de 40 tables, faisant un total de 534 colonnes et 728 556 685 lignes. La deuxième contient les données cartographiques enregistrées pour chaque patient. Ces données comprennent deux types de séries temporelles, l'une échantillonnée à haute fréquence (60 Hz) qui inclut l'électrocardiogramme et la pression artérielle continue, et l'autre qui enregistre en général un point par minute et qui comprend notamment le rythme cardiaque, la pression artérielle, la saturation pulsée en oxygène, etc. La FIGURE 3.1 montre un extrait d'une table de la base de données clinique qui liste les patients ayant fréquenté l'hôpital. Seules quelques colonnes de la table y sont exposées. On peut voir notamment l'identifiant du sujet "SUBJECT\_ID", le genre "GENDER", la date de naissance "DOB" et la date de décès à l'hôpital "DOD\_HOSP". On remarque que les informations temporelles sont toutes décalées d'une durée aléatoire et propre à chaque patient pour des raisons de confidentialité médicale.

ROW_ID	SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN	EXPIRE_FLAG
234	249	F	13/03/2075 00:00				0
235	250	F	27/12/2164 00:00	22/11/2188 00:00	22/11/2188 00:00		1
236	251	M	15/03/2090 00:00				0
237	252	M	06/03/2078 00:00				0
238	253	F	26/11/2089 00:00				0
239	255	M	05/08/2109 00:00				0
240	256	M	31/07/2086 00:00				0
241	257	F	03/04/2031 00:00	08/07/2121 00:00	08/07/2121 00:00	08/07/2121 00:00	1
242	258	F	19/09/2124 00:00				0
243	260	F	23/03/2105 00:00				0
244	261	M	04/08/2025 00:00	29/06/2102 00:00	29/06/2102 00:00	29/06/2102 00:00	1
245	262	M	05/01/2090 00:00				0
246	263	M	18/06/2104 00:00	13/06/2168 00:00	13/06/2168 00:00		1
247	264	F	30/11/2162 00:00				0
248	265	M	01/08/2093 00:00				0
249	266	F	17/12/2090 00:00				0
250	267	F	05/09/2131 00:00				0
251	268	F	21/02/2132 00:00	18/02/2198 00:00	18/02/2198 00:00		1
252	269	M	30/09/2130 00:00				0

FIGURE 3.1 – Extrait de la table "PATIENTS" de la base de données clinique.

Dans la FIGURE 3.2, quelques colonnes d'un extrait de la table "ADMISSIONS" sont affichées. Elle montre notamment les admissions du patient dont le numéro identifiant est 256 et témoigne des instants d'admission "ADMITTIME", de décharge de l'hôpital "DISCHTIME", de décès "DEATHTIME" le cas échéant et d'autres informations. Par exemple, le patient identifié par le numéro 256, présenté comme un sujet mâle né le 31/07/2086 par la table de la FIGURE 3.1, a enregistré 4 admissions le 26/07/2163 (77 ans) pour une durée de 8 jours, le 21/07/2166 (80 ans) pour une durée de 7 jours, le 15/06/2170 (84 ans) pour une durée de 13 jours et le 16/08/2170 (84 ans) pour une durée de 7 jours.

Les informations sur les sujets sont donc extraites en utilisant les tables de la partie clinique de la base de données médicale MIMIC III. Pour les signaux physiologiques, on se réfère à la partie des données cartographiques de MIMIC III. La FIGURE 3.3 montre un exemple de plusieurs signaux physiologiques enregistrés à des fréquences faibles pour une même admission à l'USI.

ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE
323	249	158975	27/04/2156 15:33	14/05/2156 15:30		EMERGENCY
324	250	124271	12/11/2188 09:22	22/11/2188 12:00	22/11/2188 12:00	EMERGENCY
325	251	117937	27/07/2110 06:46	29/07/2110 15:23		EMERGENCY
326	252	190159	31/03/2133 04:24	23/04/2133 15:00		EMERGENCY
327	252	193470	15/08/2133 04:23	19/08/2133 17:30		EMERGENCY
328	253	176189	21/01/2174 20:58	26/01/2174 16:15		EMERGENCY
329	255	112013	12/02/2187 10:30	15/02/2187 10:30		ELECTIVE
330	256	155415	26/07/2163 10:30	02/08/2163 14:14		ELECTIVE
331	256	153771	21/07/2166 23:38	28/07/2166 12:45		EMERGENCY
332	256	188869	15/06/2170 03:06	27/06/2170 16:17		EMERGENCY
333	256	108811	16/08/2170 13:55	22/08/2170 18:00		EMERGENCY
334	257	179006	20/12/2112 20:45	23/12/2112 17:00		EMERGENCY
335	258	189406	19/09/2124 03:59	22/09/2124 15:52		NEWBORN
336	260	190363	23/03/2105 10:23	30/03/2105 11:00		NEWBORN
337	261	118523	27/12/2101 01:50	12/01/2102 13:39		EMERGENCY
338	262	106019	25/09/2153 18:01	28/09/2153 18:48		ELECTIVE
339	263	120845	09/01/2160 19:30	21/01/2160 16:15		EMERGENCY
340	264	136714	30/11/2162 02:19	02/12/2162 13:47		NEWBORN
341	265	101608	20/02/2149 07:15	25/02/2149 15:50		ELECTIVE
342	266	186251	10/07/2168 08:01	14/07/2168 13:03		EMERGENCY
343	267	163714	04/06/2156 05:08	15/06/2156 13:54		EMERGENCY
344	268	110404	11/02/2198 13:40	18/02/2198 03:55	18/02/2198 03:55	EMERGENCY

FIGURE 3.2 – Extrait de la table "ADMISSIONS" de la base de données clinique.

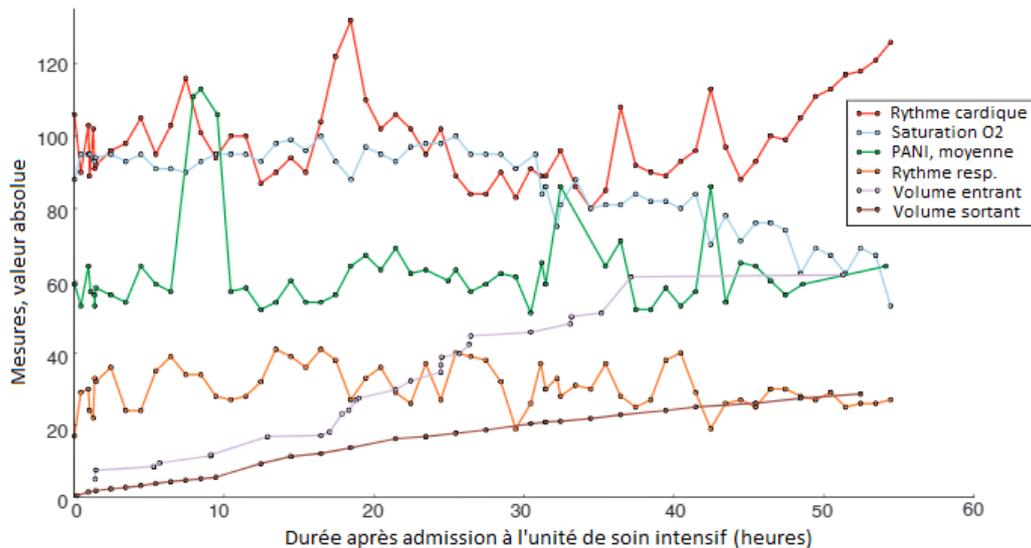


FIGURE 3.3 – Extrait de signaux physiologiques de la base de données cartographique.

### 3.3 Sélection des sujets et des signaux

Pour étudier l'effet de la FA sur les signaux physiologiques avant la phase d'épisode, il est judicieux d'analyser les enregistrements physiologiques des patients diagnostiqués avec une FA. Pour cela, on sélectionne de la base de données médicale les patients en se basant sur la table des diagnostics appelée "DIAGNOSES\_ICD" qui contient essentiellement les différents éléments diagnostiqués pour chaque admission à l'hôpital. La table "DIAGNOSES\_ICD" possède 5 champs ou colonnes :

- ROW\_ID : Entier unique pour chaque ligne ou enregistrement de la table ;
- SUBJECT\_ID : Entier identifiant le sujet ou le patient concerné par le diagnosti ;
- HADM\_ID : Entier identifiant l'admission à l'hôpital concernée par le diagnostic. Cet entier doit être conforme au champ de "SUBJECT\_ID" selon la table "ADMISSIONS" ;

- SEQ\_NUM : Entier indiquant l'ordre dans lequel les diagnostics se rapportent au patient. Les diagnostics sont classés par priorité et l'ordre a un impact sur le remboursement du traitement;
- ICD9\_CODE : Chaîne de caractère représentant un code du classement international des maladies "ICD9" et indiquant l'objet du diagnostic.

Le "ICD9" pour "International Classification of Diseases 9" est un registre de classification des maladies qui attribue à chaque pathologie un code unique. Il est généralement constitué d'un nombre à 3 chiffres suivi d'un point puis d'un autre nombre à deux chiffres au plus. Une autre version consiste à omettre le point et ainsi le code sera composé d'un nombre à 5 chiffres. Par exemple, le syndrome de vol sous-clavier est classifié selon le ICD9 par le code 435.2 ou 43520. La FA a le code 427.31 ou 42731. Pour sélectionner les admissions à l'hôpital dont la FA fait partie des maladies diagnostiquées au patient, on sélectionne les lignes ayant le champ "ICD9\_CODE" égal à 42731 de la table "DIAGNOSES\_ICD" et on garde les champs "HADM\_ID". Ainsi, on obtient la liste des admissions pour les patients atteints de FA. En parallèle, nous sélectionnons au hasard des patients sans cardiopathies et avec peu de pathologies diagnostiquées, pour les sujets témoins. Cela évite d'avoir des similitudes entre les états ou les modèles des signaux physiologiques avec ceux des patients atteints de FA. Dans le groupe de sujets témoins, seuls les patients atteints de troubles mentaux, qui ont reçu un diagnostic de ICD9 entre 290 et 319, sont admis. Nous nous retrouvons alors avec 3605 séjours ICU pour les sujets témoins, 799 d'entre eux ayant des enregistrements associés dans la base de données cartographique. Chaque enregistrement est un ensemble de signaux physiologiques mesurés et possède sa propre durée qui représente la durée du séjour du patient en réanimation. On note qu'une durée du séjour élevée permet de faire des prédictions plus tôt car elle donne l'opportunité d'explorer le comportement du métabolisme bien avant l'épisode de FA. Sur la FIGURE 3.4, les histogrammes des différentes durées de séjour disponibles dans la base de données médicale sont présentés à la fois pour le groupe de patients FA et pour le groupe témoins. La durée du séjour moyenne pour tous les sujets (avec et sans FA) est de 41h avec un écart-type de 37h.

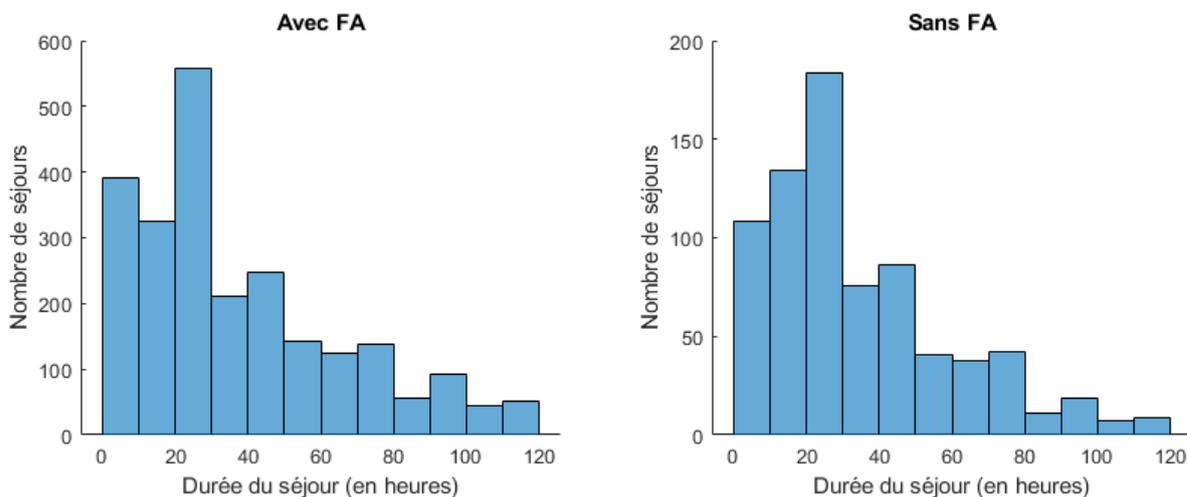


FIGURE 3.4 – Histogrammes des durées de séjour pour les patients avec et sans FA.

Pour détecter la présence d'un épisode de FA et le situer dans le temps, il est nécessaire de se servir d'un modèle de détection qui se base sur les signaux physiologiques disponibles. La condition qui permet d'affirmer la survenue d'un épisode de FA durant le séjour à l'unité de soin intensif est une fenêtre de 15 minutes pendant laquelle le rythme cardiaque reste au-dessus de la barre des 110 battements par minute [83]. L'indisponibilité du signal de rythme cardiaque ou l'inexistence de ce critère témoignant de l'absence d'un épisode de FA durant tout le séjour à l'USI rend l'enregistrement inutile pour notre étude. De ce fait, il ne reste que 701 enregistrements confirmés avec un épisode de FA bien situé dans le temps. Il est à noter que les signaux sont coupés à partir de

l'instant de l'épisode vu que l'étude concerne la prédiction prématurée de l'épisode. Notre intérêt réside alors dans l'allure des signaux physiologiques depuis l'admission à l'USI jusqu'à la survenue de l'épisode.

Les types de signaux considérés doivent également être liés à la FA, plus généralement, ils doivent être liés à la circulation sanguine car elle est assurée par le cœur. De plus, les valeurs prises par le signal doivent être continues car une application d'analyse statistique est prévue à ce signal. Par conséquent, tous les types de signaux avec des valeurs discrètes telles que les signaux d'état sont exclus de notre étude. En tenant compte de tous ces facteurs, une sélection de 10 types de signaux semblent la plus adéquate :

- Fréquence cardiaque (FC)
- Nombre de battements rythmés (NBR)
- Nombre de battements normaux (NBN)
- Moyenne de la tension artérielle non invasive (NBP Mean)
- Tension artérielle systolique non invasive (NBP Sys)
- Tension artérielle diastolique non invasive (NBP Dias)
- Pression veineuse centrale (PVC)
- Moyenne de la tension artérielle (ABP Mean)
- Tension artérielle systolique (ABP Sys)
- Tension artérielle diastolique (ABP Dias)

La FIGURE 3.5 montre la disponibilité pour chaque type de signal. Le nombre moyen de types de signaux disponibles par sujet est de 4.06 pour les sujets atteints de FA et de 4.32 pour les sujets témoins, 102 sujets ayant tous leurs signaux disponibles. On rappelle que, pour les sujets du groupe FA, seules les parties des signaux précédant le début de l'épisode de FA sont conservées pour être traitées ultérieurement, c'est-à-dire pour effectuer la prédiction et non le diagnostic de la FA.

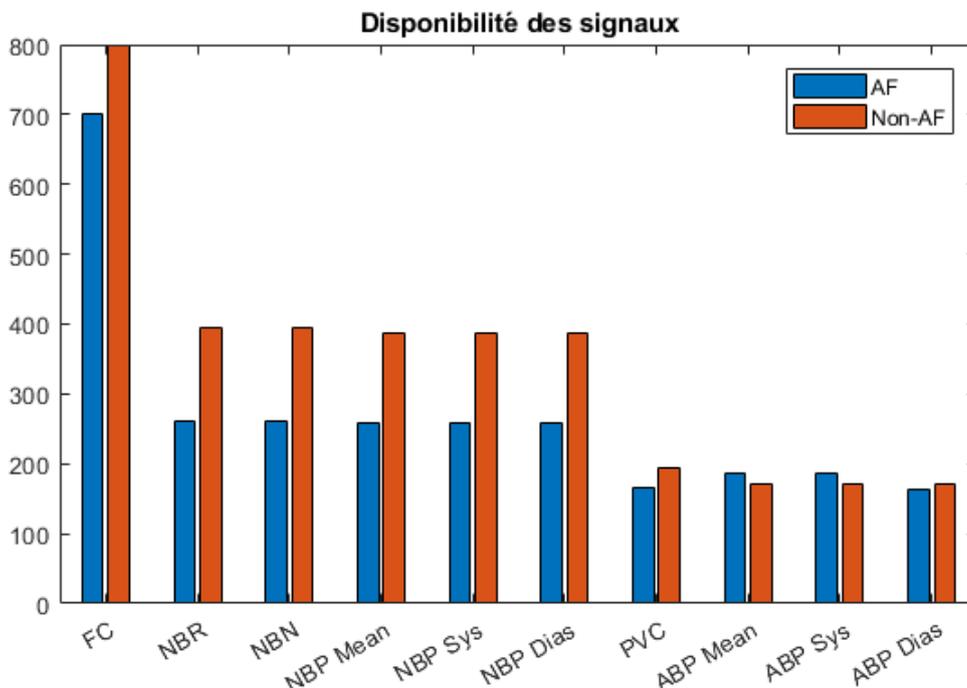


FIGURE 3.5 – Statistiques de disponibilité des signaux physiologiques.

Dans le but de formaliser les données obtenues, certaines variables seront introduites et elles représentent la totalité des informations disponibles pour la tâche de prédiction de la FA. Les 1500

séjours sélectionnés à l'USI seront indexés par  $i$  tel que  $1 \leq i \leq 1500$ . Les signaux physiologiques temporels échantillonnés tous à 1 point par minute sont représentés par des suites numériques  $s_{i,j}(\cdot)$  tel que  $i$  désigne le séjour à l'USI et  $j$  désigne le type de signal ( $1 \leq j \leq 10$ ). Chaque séjour a une durée propre  $T_i$  différente d'un séjour à l'autre, mais identique pour tous les signaux prélevés sur ce même séjour. Ainsi,  $s_{i,j}(t)$  représente la valeur du signal  $j$  prélevée du séjour  $i$  à la minute  $t$  depuis l'admission à l'USI tel que  $1 \leq t \leq T_i$ .

### 3.4 Présentation de l'approche

La méthode présentée dans ce chapitre consiste à analyser des signaux physiologiques d'un sujet dans le but d'en forger une prédiction sur l'avenir sanitaire de celui-ci. Cette méthode est composée de plusieurs étapes. La première consiste à extraire des caractéristiques des signaux pour adapter les informations à l'entrée du classifieur. Cette étape permet au classifieur de traiter des caractéristiques précalculées des signaux au lieu des signaux eux-mêmes et sera détaillée dans la section 3.5. La deuxième étape consiste à se servir de la théorie des fonctions de croyance dans l'optique de construire un classifieur ou prédicteur. Dans ce cadre, chaque caractéristique précédemment extraite sera utilisée, en se référant à une base d'apprentissage, comme une source d'information représentée par une distribution de masses. La troisième étape consiste à combiner les différentes distributions de masses obtenues des caractéristiques pour en déduire une nouvelle plus informative et rassemblant les connaissances apportées par toutes les caractéristiques calculées à partir des signaux physiologiques. La dernière étape consiste à prendre une décision basée sur la distribution de masse obtenue qui est assurée par une transformation de l'approche évidentielle en une approche probabiliste. La méthode globale est illustrée dans la FIGURE 3.6.

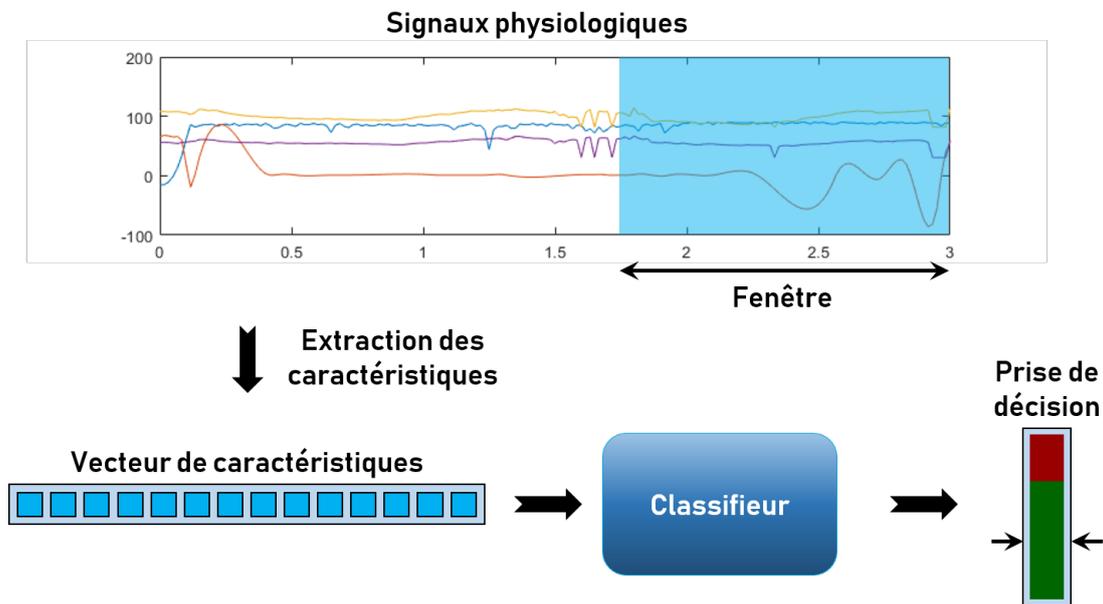


FIGURE 3.6 – Processus de prédiction de la FA.

### 3.5 Extraction des caractéristiques

L'extraction de caractéristiques est une étape-clé des problèmes de reconnaissance de formes. Elle est notamment une étape préalable à la classification, permettant de décrire l'objet à classer grâce à un ensemble d'attributs discriminants [84]. Ce problème réside dans une large variété de domaines d'applications allant de la reconnaissance vocale dans le domaine de la robotique au traitement du signal et de l'image dans le domaine médical. Les caractéristiques peuvent être binaires, catégoriques ou continues, et elles sont de deux types [85] :

- Des attributs ayant un sens physique, tels que des symptômes ou des caractéristiques morphologiques d'objets dans les images, établis généralement grâce à une expertise humaine ;
- Des attributs génériques, sans interprétation physique ou biologique, établis grâce à une construction automatique. Différentes techniques d'extraction automatique de caractéristiques existent. Nous en citons :
  - Les méthodes statistiques, extrayant des attributs statistiques des données brutes [86] ;
  - Les méthodes fréquentielles, telles que la transformée de Fourier et l'analyse par Ondelettes [87] ;
  - Les méthodes linéaires et non-linéaires de réduction de dimensionnalité, telles que l'analyse par composantes principales [88] ou la mise à l'échelle multidimensionnelle [89], où les données brutes sont représentées dans un espace à dimensions réduites ;
  - Les méthodes d'extraction de caractéristiques locales, appliquées sur des données séquentielles, spatiales ou structurées, telles que les méthodes convolutionnelles d'apprentissage profond [90].

Dans cette thèse, nous nous intéressons au domaine médical, et plus particulièrement à la représentation de signaux physiologiques pour le pronostic et le diagnostic médical. Pour mieux analyser les signaux physiologiques d'un sujet, il est nécessaire d'extraire des caractéristiques de ces signaux. Il en résulte un vecteur de caractéristiques dont chaque composante représente un critère numérique bien défini décrivant l'état du sujet observé. Dans ce chapitre, nous nous intéressons aux indicateurs statistiques, choisis pour leur forte dépendance aux différents schémas des signaux physiologiques de l'être humain. Ainsi, les sursauts, les augmentations ou diminutions prolongées, les fortes variations ou la stabilité des signaux physiologiques sont facilement détectés et quantifiés par ces caractéristiques.

La surveillance d'un sujet au cours d'un séjour  $i$  produit 10 séries temporelles ou signaux représentés par  $s_{i,j}(t)$ , avec  $1 \leq j \leq 10$  représentant l'indice du signal et  $t$  l'instant de la mesure. A chaque instant  $t$ , une fenêtre de taille  $N$  est considérée à la fin de chaque signal couvrant les données  $s_{i,j}(t - N + 1), \dots, s_{i,j}(t)$  et un certain nombre  $F$  de caractéristiques statistiques  $y$  est extrait. Soit  $\mathbf{x}_i(t)$  le vecteur de caractéristiques extrait de l'ensemble des fenêtres de taille  $N$  prises à la fin des signaux  $j$  à l'instant  $t$ . Ainsi, la dimension de  $\mathbf{x}_i(t)$  est donnée par  $10 \times F$  et la composante  $F(j - 1) + k$  du vecteur  $\mathbf{x}_i(t)$ , notée  $x_{i,F(j-1)+k}(t)$ , représente la caractéristique de type  $k$  ( $1 \leq k \leq F$ ) du signal  $j$  ( $1 \leq j \leq 10$ ). Les caractéristiques extraites sont les suivantes :

- Moyenne

$$x_{i,F(j-1)+1}(t) = \frac{1}{N} \sum_{r=1}^N s_{i,j}(t - N + r),$$

- Écart-type

$$x_{i,F(j-1)+2}(t) = \sqrt{\frac{1}{N} \sum_{r=1}^N (s_{i,j}(t - N + r) - x_{i,F(j-1)+1}(t))^2},$$

- Skewness

$$x_{i,F(j-1)+3}(t) = \frac{1}{N} \sum_{r=1}^N \left( \frac{s_{i,j}(t - N + r) - x_{i,F(j-1)+1}(t)}{x_{i,F(j-1)+2}(t)} \right)^3,$$

- Kurtosis

$$x_{i,F(j-1)+4}(t) = \frac{1}{N} \sum_{r=1}^N \left( \frac{s_{i,j}(t - N + r) - x_{i,F(j-1)+1}(t)}{x_{i,F(j-1)+2}(t)} \right)^4,$$

- Valeur absolue de la différence

$$x_{i,F(j-1)+5}(t) = \frac{1}{N} \sum_{r=2}^N \left| s'_{i,j}(t - N + r) \right|,$$

avec  $|\cdot|$  la valeur absolue de  $\cdot$  et  $s'_{i,j}(t) = s_{i,j}(t) - s_{i,j}(t - 1)$ ,

— Écart-type de la différence

$$x_{i,F(j-1)+6}(t) = \sqrt{\frac{1}{N} \sum_{r=2}^N \left( s'_{i,j}(t-N+r) - \frac{s_{i,j}(t) - s_{i,j}(t-N+1)}{N} \right)^2},$$

— Skewness de la différence

$$x_{i,F(j-1)+7}(t) = \frac{1}{N} \sum_{r=2}^N \left( \frac{Ns'_{i,j}(t-N+r) - s_{i,j}(t) + s_{i,j}(t-N+1)}{Nx_{i,F(j-1)+6}(t)} \right)^3,$$

— Kurtosis de la différence

$$x_{i,F(j-1)+8}(t) = \frac{1}{N} \sum_{r=2}^N \left( \frac{Ns'_{i,j}(t-N+r) - s_{i,j}(t) + s_{i,j}(t-N+1)}{Nx_{i,F(j-1)+6}(t)} \right)^4,$$

— Moyenne de la valeur minimale sur une fenêtre glissante

$$x_{i,F(j-1)+9}(t) = \frac{1}{N - \lfloor N/5 \rfloor} \sum_{r=1}^{N - \lfloor N/5 \rfloor} \min_{0 \leq \epsilon \leq \lfloor N/5 \rfloor} (s_{i,j}(t-N+r+\epsilon)),$$

avec  $\lfloor \cdot \rfloor$  la valeur arrondie de  $\cdot$ ,

— Moyenne de la valeur maximale sur une fenêtre glissante

$$x_{i,F(j-1)+10}(t) = \frac{1}{N - \lfloor N/5 \rfloor} \sum_{r=1}^{N - \lfloor N/5 \rfloor} \max_{0 \leq \epsilon \leq \lfloor N/5 \rfloor} (s_{i,j}(t-N+r+\epsilon)),$$

— Changement totale

$$x_{i,F(j-1)+11}(t) = s_{i,j}(t) - s_{i,j}(t-N+1),$$

— Maximum

$$x_{i,F(j-1)+12}(t) = \max_{1 \leq r \leq N} s_{i,j}(t-N+r),$$

— Augmentation maximale

$$x_{i,F(j-1)+13}(t) = \max_{2 \leq r \leq N} s'_{i,j}(t-N+r),$$

— Minimum

$$x_{i,F(j-1)+14}(t) = \min_{1 \leq r \leq N} s_{i,j}(t-N+r),$$

— Diminution maximale

$$x_{i,F(j-1)+15}(t) = - \min_{2 \leq r \leq N} s'_{i,j}(t-N+r),$$

— Intervalle

$$x_{i,F(j-1)+16}(t) = x_{i,F(j-1)+12}(t) - x_{i,F(j-1)+14}(t),$$

— Entropie de l'échantillon

Soit  $\mathbf{v}_{i,j,t,N}^{(w,u)}$  le vecteur formé des valeurs du signal  $s_{i,j}(\cdot)$  entre  $t-N+w$  et  $t-N+w+u-1$  et

$$g_{i,j,t,N}(r, u) = \sum_{1 \leq w_1, w_2 \leq N-u+1} \left[ \text{dist} \left( \mathbf{v}_{i,j,t,N}^{(w_1,u)}, \mathbf{v}_{i,j,t,N}^{(w_2,u)} \right) \leq r \right],$$

avec  $r \in \mathbb{R}^+$ ,  $2 \leq u \leq N-1$ ,  $\text{dist}(\cdot, \cdot)$  est la fonction de distance de Chebyshev et  $\llbracket \cdot \rrbracket$  prend la

valeur 1 si  $\cdot$  est vrai et 0 sinon, ainsi l'entropie de l'échantillon est définie comme

$$SE_{i,j,t,N}(r, u) = \log \left( \frac{g_{i,j,t,N}(r, u-1)}{g_{i,j,t,N}(r, u)} \right).$$

Dans notre cas, les paramètres  $r$  et  $u$  sont assignés comme suit :

$$r = \frac{x_{i,F(j-1)+16}(t)}{10},$$

$$u = \lfloor N/8.2 \rfloor.$$

La valeur de  $u$  est choisie de sorte à obtenir 10 fenêtres de  $s_{i,j}(\cdot)$  comprises entre  $t - N + 1$  et  $t$  avec un chevauchement de 20%. On a alors :

$$x_{i,F(j-1)+17}(t) = SE_{i,j,t,N} \left( \frac{x_{i,F(j-1)+16}(t)}{10}, \lfloor N/8.2 \rfloor \right),$$

— Analyse de la fluctuation détendue pour les petites valeurs

Soit  $X_{i,j,t,N}(r) = \sum_{l=1}^r s_{i,j}(t - N + l) - x_{i,F(j-1)+1}(t)$  avec  $1 \leq r \leq N$  et

$$F_{i,j,t,N}(u) = \sqrt{\frac{1}{N} \sum_{r=1}^N \left( X_{i,j,t,N}(r) - Y_{i,j,t,N}^u(r) \right)^2},$$

avec  $Y_{i,j,t,N}^u(r)$  représente la concaténations de plusieurs fonctions linéaires adaptées chacune par la méthode des moindres carrés sur un intervalle de taille  $u$  de  $X_{i,j,t,N}(\cdot)$  sans chevauchement. Ensuite une dernière fonction linéaire est adaptée entre  $\log(u)$  et  $\log(F_{i,j,t,N}(u))$  lorsque  $u$  prend varie dans  $U$  tel que

$$\log(F_{i,j,t,N}(u)) = \hat{a}_{i,j,t,N}(U) \log(u) + \hat{b}_{i,j,t,N}(U).$$

La pente  $\hat{a}_{i,j,t,N}(U)$  dépend de l'ensemble  $U$ . On a alors :

$$x_{i,F(j-1)+18}(t) = \hat{a}_{i,j,t,N}(\{3, \dots, 10\}),$$

— Analyse de la fluctuation détendue pour les grandes valeurs

$$x_{i,F(j-1)+19}(t) = \hat{a}_{i,j,t,N}(\{10, \dots, 30\}).$$

Dans notre cas, on a alors  $F = 19$ , ce qui renvoie à un vecteur de caractéristiques de dimension 190. Notons qu'il est crucial de représenter les données brutes d'une manière fiable, sans perdre des informations pertinentes. Ainsi, il est toujours préférable d'inclure un grand nombre de caractéristiques, au prix de noyer les informations pertinentes au sein d'informations potentiellement non pertinentes. Par ailleurs, dans le domaine médical, beaucoup de données peuvent manquer à un instant donné. L'avantage de notre classifieur, présenté dans la section suivante, réside dans sa capacité à surmonter les problèmes du grand nombre de caractéristiques et de données manquantes grâce à sa modélisation évidentielle de l'information, basée sur la théorie des fonctions de croyance.

## 3.6 Modélisation des fonctions de masse

### 3.6.1 Introduction à la classification

La classification est un domaine de l'intelligence artificielle où le but est d'attribuer un label à une observation. Ce label est choisi parmi un certain nombre de classes ou clusters possibles. On distingue la classification non-supervisée et la classification supervisée. La classifica-

tion non-supervisée ou clustering consiste à regrouper des instances en se basant sur un critère de dissimilarité entre elles. En général, les instances sont représentées par des vecteurs de caractéristiques et le critère utilisé est une fonction de distance entre ces vecteurs. L'avantage de la classification non-supervisée est qu'elle n'exige pas une base d'apprentissage avec des instances préalablement labellisées. Une des méthodes les plus connues de la classification non-supervisée est la K-moyennes ou K-means [91]. En ayant un certain nombre d'instances disponibles, cette méthode consiste à construire K clusters ou groupes d'instances, en appliquant un algorithme itératif. En premier lieu, une initialisation de l'algorithme est effectuée en choisissant aléatoirement un nombre K d'instances parmi les instances disponibles, et en les définissant comme centres initiaux des K clusters à construire. L'algorithme alterne ensuite entre deux étapes. La première étape attribue à chaque instance le cluster ayant le centre le plus proche à l'instance. La deuxième étape consiste à redéfinir les centres en calculant les moyennes des instances ou centre de gravité de chaque cluster. L'algorithme s'arrête lorsque les centres n'évoluent plus. Cela conduit à une partition des instances, avec des clusters bien définis. Une autre approche répandue de clustering est la méthode hiérarchique, qui consiste à regrouper itérativement les instances, en construisant progressivement un arbre appelé dendrogramme [92]. A la racine de cet arbre, se retrouvent les instances élémentaires définissant des clusters unitaires. A chaque itération, les distances inter-clusters sont calculées et la paire de clusters donnant la distance optimale est regroupée en un même cluster. L'algorithme s'arrête quand toutes les instances sont regroupées dans un même cluster. Ceci conduit à une hiérarchisation des instances sous forme d'un dendrogramme. Ceci suppose de savoir calculer, à chaque étape ou regroupement, la distance entre deux clusters. Enfin, le dendrogramme est coupé à un certain niveau optimal où la distance inter- et intra-clusters est optimale selon un critère à définir.

La classification supervisée quant à elle, consiste à attribuer une classe bien définie à une observation en se basant sur un ensemble d'instances déjà labellisées, c'est-à-dire dont les classes sont connues. Elle nécessite ainsi une base d'apprentissage constituée d'instances labellisées, de préférence équilibrées entre les classes, et surtout représentatives de la population. Une des méthodes classiques de classification supervisée est la méthode des K plus proches voisins ou KNN [93]. Elle consiste à attribuer à chaque observation à classer la classe dominante dans le voisinage de cette observation formé des K plus proches instances de la base d'apprentissage en se basant sur une fonction de distance bien définie. On distingue aussi la méthode des machines à vecteurs de support ou "Support Vector Machine" (SVM) qui est une méthode de base conçue pour la classification binaire. Cette méthode vise à entraîner un classifieur en minimisant un critère de marginalité reposant sur les vecteurs dits "de support". L'algorithme est donc apparié à un problème d'optimisation et la répétition de cet algorithme en confrontant différentes classes ou ensemble de classes élargit le domaine d'application de cet algorithme aux problèmes de classification multi-classes [94]. Une autre méthode de classification très performante mais gourmande en données d'apprentissage est la méthode de l'apprentissage profond ou "Deep Learning" basée sur les réseaux de neurones convolutifs ou "Convolutional Neural Networks" (CNN) [95–100]. Cette méthode a la particularité de traiter des données brutes en dénichant de ses propres soins les caractéristiques les plus importantes des données en utilisant des architectures variées de fonctions non-linéaires. Elle a permis un développement considérable de la vision par ordinateur, de la reconnaissance de la parole, des formes et du langage naturel. Toutefois, ces méthodes se voient leur performance dégradées en cas de données manquantes, de sources d'information imprécises et de traitement de signal (pour le CNN).

Dans notre étude, il s'agit d'un cas de classification supervisée puisque l'on dispose d'une base d'apprentissage avec des sujets labellisés. Le processus de prédiction en temps réel de l'épisode de la FA commence par la mesure des signaux physiologiques du sujet. A chaque instant  $t$  et pour chaque signal mesuré, la fenêtre de longueur N minutes se terminant à  $t$  est envoyée à la tâche d'extraction de caractéristiques de la Section 3.5 qui donne un vecteur de caractéristiques  $\mathbf{x}(t)$ . Ce vecteur de caractéristiques est ensuite utilisé comme entrée au classifieur qui génère un niveau de risque d'occurrence d'un épisode de la FA. Deux classes sont considérées, la classe "Sujet témoin"

non prévu de subir un épisode de FA, labellisée 0, et la classe "Sujet FA" prévu de subir un épisode de FA, labellisée 1. Le classifieur proposé est défini dans le cadre de la théorie des fonctions de croyance en utilisant une base de données d'apprentissage. L'avantage de cette théorie réside dans sa capacité à modéliser le problème de manière décente même avec des données manquantes ou peu fiables, sans nécessité d'une base d'apprentissage totalement équilibrée. La construction de la base d'apprentissage et la conception du classifieur sont présentées ci-dessous.

### 3.6.2 Division de la base de données

Afin d'entraîner le classifieur, une base d'apprentissage est construite en sélectionnant au hasard une portion  $\delta\%$  ( $0 \leq \delta \leq 100$ ) de l'ensemble des 701 séjours à l'USI des sujets ayant eu la FA et une portion pareille des 799 séjours de ceux ne l'ayant pas eue. Ce choix est indiqué par les variables  $I_{\mathcal{A}}$  et  $I_{\mathcal{T}}$  où  $I_{\mathcal{A}}$  regroupe l'ensemble des indices des séjours à l'USI utilisé pour l'apprentissage et  $I_{\mathcal{T}}$  représente le groupe restant utilisé pour tester les performances de l'algorithme de prédiction. On a donc  $I_{\mathcal{A}} \cup I_{\mathcal{T}} = \{1, 2, \dots, 1500\}$  et  $|I_{\mathcal{A}}| = \lceil 701 \times \delta \rceil + \lceil 799 \times \delta \rceil$ ,  $\lceil \cdot \rceil$  représentant l'arrondi de  $\cdot$ . Pour les séjours choisis dans la base d'apprentissage, les dernières fenêtres de longueur  $N$  de leurs signaux physiologiques disponibles sont considérées pour calculer leurs vecteurs de caractéristiques, conduisant à un ensemble de vecteurs de caractéristiques avec des labels bien connus. Le choix de la dernière fenêtre permet d'avoir un morceau des signaux dont le comportement décrit le plus fidèlement possible son label. En effet, pour les sujets qui connaissent un épisode de la FA, les fenêtres se terminant au début de l'épisode donnent la représentation la plus proche des perturbations liées à l'épisode de la FA tout en restant dans le domaine de la prédiction c'est-à-dire avant l'occurrence de l'épisode. Tandis que pour les sujets témoins, la fenêtre se termine à la sortie de l'unité de soins intensifs assurant une stabilité de l'état de santé du sujet. Afin de tester la méthode plus tard, nous définissons  $\delta = 0.8$ , c'est-à-dire que nous choisissons au hasard 20% de l'ensemble de données pour le processus de test et les autres 80% sont réservés pour entraîner le classifieur. On définit la base d'apprentissage comme suit :

$$\mathcal{A} = \{(\mathbf{x}_i(T_i), y_i) \mid i \in I_{\mathcal{A}} \text{ et } y_i \in \{0, 1\}\},$$

et la base de test :

$$\mathcal{T} = \{(\mathbf{x}_i(t), y_i) \mid N \leq t \leq T_i, i \in I_{\mathcal{T}} \text{ et } y_i \in \{0, 1\}\}.$$

La base de données  $\mathcal{A}$  est utilisée dans le domaine de la théorie des fonctions de croyance pour construire le classifieur. Dans la phase de test, les sujets des indices dans  $I_{\mathcal{T}}$  sont considérés avec une fenêtre glissante. A chaque instant  $t$ , le vecteur de caractéristiques  $\mathbf{x}_i(t)$ ,  $i \in I_{\mathcal{T}}$ , calculé à  $t$  est utilisé comme entrée du classifieur, comme illustré dans la FIGURE 3.7.

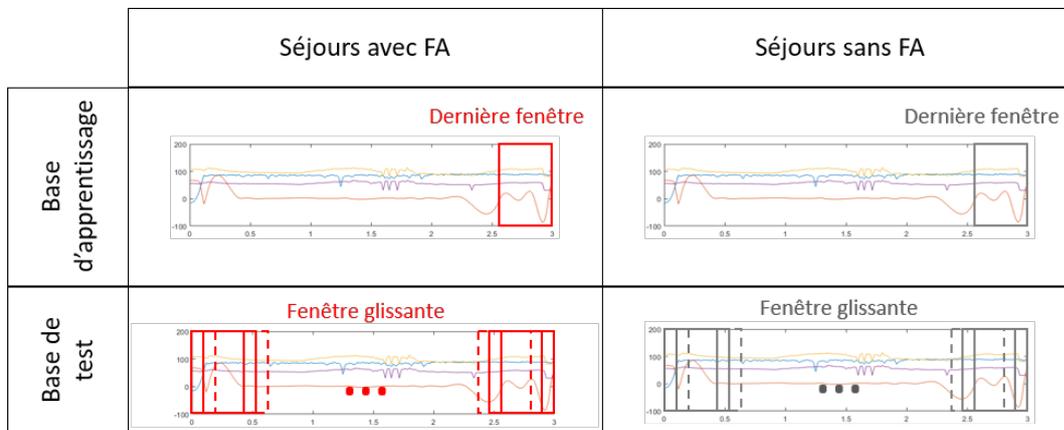


FIGURE 3.7 – Répartition des fenêtres selon les bases d'apprentissage et de test.

### 3.6.3 Attribution des masses

La méthode de classification proposée est une méthode évidentielle qui consiste à utiliser chaque caractéristique comme une source d'information indépendante dans le cadre de la théorie des fonctions de croyance. Une fonction de masse est donc élaborée par caractéristique en se basant sur la base d'apprentissage. Cette source apporte des informations concernant l'occurrence de la fibrillation atriale ou FA. Elle est ainsi définie sur l'ensemble  $\Omega = \{0, 1\}$ , où 0 désigne la classe témoin, de non-prédiction de la FA, et 1 désigne la classe FA, de prédiction d'une occurrence de la FA. Dans le cadre de la théorie des fonctions de croyance, les masses seront donc attribuées aux 4 sous-ensembles possibles de  $\Omega$  qui sont  $\{0\}$ ,  $\{1\}$ ,  $\emptyset$  et  $\Omega = \{0, 1\}$ , où la masse de  $\{0\}$  représente la part d'évidence réservée uniquement à la non-prédiction de la FA, la masse de  $\{1\}$  représente la part d'évidence réservée uniquement à la prédiction de la FA, la masse de  $\emptyset$  représente le conflit et donc la masse sera nulle lors de l'attribution des masses et enfin la masse de  $\Omega$  représente la part d'ignorance apportée par la caractéristique. Afin de pouvoir modéliser les fonctions de masse, nous proposons d'abord de modéliser les données de la base d'apprentissage  $\mathcal{A}$  pour chaque caractéristique  $u$  et pour chaque classe labellisée 0 ou 1 par une distribution de probabilité. En conduisant un test de Student sur les caractéristiques des vecteurs, on remarque qu'en moyenne 90.5% des caractéristiques ne rejettent pas l'hypothèse nulle ce qui confirme une distribution normale des données [101]. On adapte alors une distribution normale pour chaque caractéristique et pour chaque classe. Cette étape a pour but de concrétiser les données pour une mesure de similarité plus simple avec un vecteur de caractéristiques à classifier. Cette adaptation est assurée par la méthode de la maximisation de la fonction de vraisemblance et elle consiste à estimer les paramètres d'une distribution normale, à savoir, l'espérance et l'écart-type. On obtient alors les paramètres suivants  $\forall u \in \{1, 2, \dots, 190\}$ ,

$$\begin{aligned}\hat{\mu}_u^{(1)} &= \frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} x_{i,u}(T_i) y_i}{\sum_{i \in \mathcal{I}_{\mathcal{A}}} y_i}, \\ \hat{\mu}_u^{(0)} &= \frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} x_{i,u}(T_i) (1 - y_i)}{\sum_{i \in \mathcal{I}_{\mathcal{A}}} 1 - y_i}, \\ \hat{\sigma}_u^{(1)} &= \sqrt{\frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} (x_{i,u}(T_i) - \hat{\mu}_u^{(1)})^2 y_i}{\sum_{i \in \mathcal{I}_{\mathcal{A}}} y_i}}, \\ \hat{\sigma}_u^{(0)} &= \sqrt{\frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} (x_{i,u}(T_i) - \hat{\mu}_u^{(0)})^2 (1 - y_i)}{\sum_{i \in \mathcal{I}_{\mathcal{A}}} 1 - y_i}},\end{aligned}$$

avec  $x_{i,u}(T_i)$  la coordonnée  $u$  du vecteur de caractéristiques  $\mathbf{x}_i(T_i)$ . Ainsi, pour chaque caractéristique  $u$  ( $1 \leq u \leq 190$ ), on a deux distributions normales estimées à partir des valeurs des deux classes dans la base d'apprentissage  $\mathcal{A}$ . L'attribution des masses pour chaque caractéristique se fera en se basant sur ces deux distributions et sur la valeur de cette caractéristique dans le vecteur à classifier à chaque instant  $t$ .

Durant la phase de test, une fenêtre glissante de taille  $N$  est extraite du signal à chaque instant  $t$ , ce qui donne lieu à plusieurs fenêtres à analyser par séjour. De chacune de ces fenêtres, est extrait un vecteur de caractéristiques  $\mathbf{x}_i(t)$ . Il est donc question d'élaborer une fonction de masse  $m_u^{i,t}(\cdot)$  pour chaque caractéristique  $u$  répartissant les masses entre les événements  $\{0\}$ ,  $\{1\}$  et  $\Omega$  à l'instant  $t$ , grâce à la valeur de la caractéristique  $x_{i,u}(t)$  et aux distributions normales préalablement définies sur  $\mathcal{A}$ . La masse de l'ignorance est la première calculée en se basant sur le chevauchement entre les courbes des fonctions de densité de probabilité. Cette surface s'exprime comme suit :

$$\mathcal{S}_u = \int_{-\infty}^{+\infty} \min(\phi(x, \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}), \phi(x, \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})) dx,$$

avec  $\phi(\cdot, \mu, \sigma)$  la fonction de densité de probabilité de la loi normale d'espérance  $\mu$  et d'écart-type

$\sigma$ . On a alors  $\forall u \in \{1, 2, \dots, 190\}$  :

$$m_u^{i,t}(\Omega) = (\mathcal{S}_u)^\alpha, \quad (3.1)$$

avec  $\alpha$  un paramètre à adapter selon les données. Nous expliquons cette attribution par le fait que plus les densités des deux classes concernant une caractéristique  $u$  se chevauchent, plus les classes vues par  $u$  se ressemblent, et ainsi plus la capacité informative de la caractéristique  $u$  doit être faible et donc moins elle doit se prononcer. Nous proposons de traduire cela par une masse de l'ignorance définie comme une fonction croissante de la surface de chevauchement. Si  $\alpha < 1$ , la masse de l'ignorance se trouve augmentée ce qui est adéquat pour un nombre important de fonctions de masse. En effet, dans le processus de combinaison, les sources d'information mettent en commun leur données pour construire une fonction de masse plus informative. Une faible valeur de masse d'ignorance s'apparente alors à des sources d'information présentant leurs données avec un certain niveau élevé de confiance. Cet abus de confiance peut conduire lors de la combinaison à une augmentation du niveau de conflits et une perte du noyau informatif des sources réellement dignes de confiance. C'est pourquoi ce terme correcteur  $\alpha$  et surtout dans le cas d'un nombre important de sources permet de rendre les déclarations de celles-ci plus modestes, de limiter les conflits et ainsi d'obtenir une fonction de masse relayant plus fidèlement la significations des données statistiques. Le reste de masses est distribué proportionnellement aux valeurs prises par les fonctions de densité de probabilité au point  $x_{i,u}(t)$  sur les singletons :

$$m_u^{i,t}(\{1\}) = \frac{\phi(x_{i,u}(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})}{\phi(x_{i,u}(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}) + \phi(x_{i,u}(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})} (1 - m_u^{i,t}(\Omega)),$$

$$m_u^{i,t}(\{0\}) = \frac{\phi(x_{i,u}(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})}{\phi(x_{i,u}(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}) + \phi(x_{i,u}(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})} (1 - m_u^{i,t}(\Omega)).$$

La FIGURE 3.8 illustre l'attribution des masses pour une caractéristique  $u$  dont la valeur est de 2.7. Les valeurs prises par les fonctions de densité de probabilité de chaque groupe 1 et 0 sont respectivement 0.09405 et 0.1972. L'aire de la surface de chevauchement entre les courbes illustrées en gris est de 0.4525. En réglant  $\alpha = 1$ , le calcul de la distribution de masses s'effectue alors en commençant par attribuer l'aire de chevauchement élevée à la puissance  $\alpha$ , à l'ignorance ou à  $\Omega$ . Puis le reste de la masse qui est de 0.5475 est distribuée selon la proportion des valeurs prises par les fonctions de densité de probabilité de chaque groupe. Ainsi la masse attribuée à  $\{1\}$  est de  $0.5475 \times \frac{0.09405}{0.09405+0.1972} = 0.1768$  et la masse attribuée à  $\{0\}$  est de  $0.5475 \times \frac{0.1972}{0.09405+0.1972} = 0.3707$ .

### 3.6.4 Sélection des caractéristiques

Dans le but d'améliorer la modélisation des données statistiques lors de l'attribution des masses, la valeur de l'ignorance peut être modifiée selon la disponibilité et la quantité d'information traduite par la base d'apprentissage. Certaines sources d'information présentent une masse d'ignorance très importante. Dans le cas de notre étude, la majorité des sources d'information ne sont utiles que sur de rares plage de valeurs des caractéristiques. Ainsi, une masse d'ignorance élevée permet d'évanouir leur participation lors du processus de combinaison des sources d'information et dans la construction de la fonction de masse finale. Il n'en demeure pas moins vrai que leur nombre élevé finit quand même par perturber les informations issues des sources fiables dans la combinaison. Pour palier à ce mal, un terme  $\beta$  ( $0 \leq \beta \leq 1$ ) est introduit. Les sources ayant une masse d'ignorance supérieure à ce seuil  $\beta$  seront exclues de la combinaison et n'auront donc aucun effet sur la masse finale. Mathématiquement, ceci est équivalent à modifier cette masse d'ignorance et la remettre à 1 ce qui donne l'équation suivante :

$$m_u^{i,t}(\Omega) = \begin{cases} (\mathcal{S}_u)^\alpha & \text{si } (\mathcal{S}_u)^\alpha < \beta, \\ 1 & \text{si } (\mathcal{S}_u)^\alpha \geq \beta, \end{cases}$$

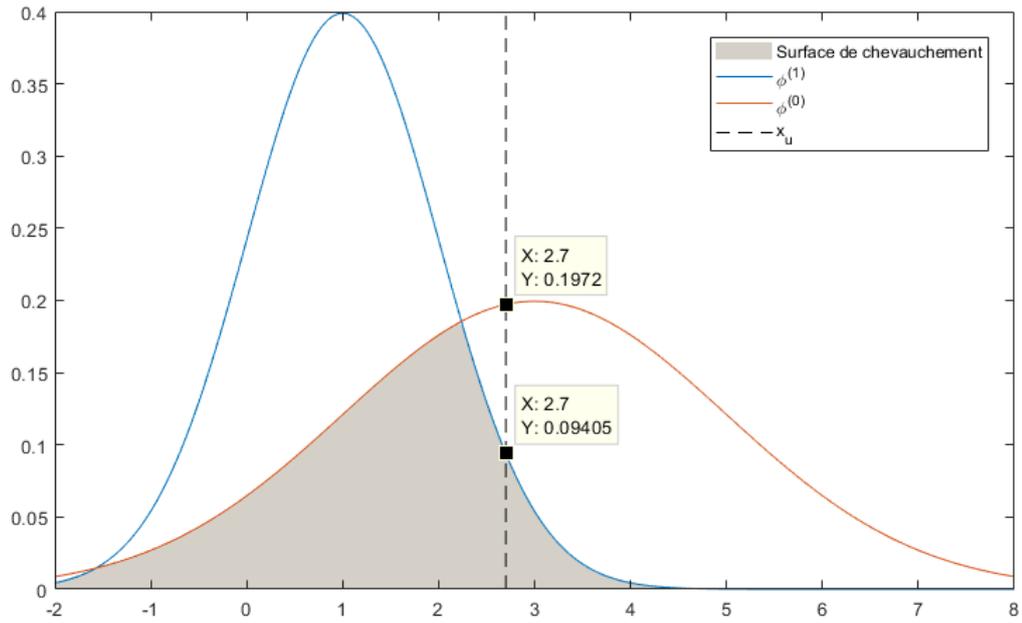


FIGURE 3.8 – Exemple de distribution des masses.

$$m_u^{i,t}(\{r\}) = \frac{\phi(x_{i,u}(t), \hat{\mu}_u^{(r)}, \hat{\sigma}_u^{(r)})}{\sum_{s \in \{0,1\}} \phi(x_{i,u}(t), \hat{\mu}_u^{(s)}, \hat{\sigma}_u^{(s)})} (1 - m_u^{i,t}(\Omega)) \quad \forall r \in \{0, 1\}.$$

En effet, une fonction de masse qui attribue l'unité entière à la masse de l'ignorance totale n'apporte aucune information et se trouve donc l'élément neutre de l'opération de combinaison. On note que la sélection des caractéristiques peut être ignorée en considérant  $\beta = 1$ .

### 3.6.5 Combinaison et prise de décision

Dans le chapitre précédent, plusieurs méthodes de combinaison ont été introduites. Le choix de la méthode de combinaison est une étape cruciale pour le bon fonctionnement du classifieur. Elle définit comment les informations seront mises en commun et influe sur la certitude prononcée par la fonction de masse finale. Dans le cas de notre étude, étant donné que deux événements (0 et 1) seulement sont possibles, la distribution des masses concerne les ensembles d'évènements suivants :  $\emptyset$ ,  $\{0\}$ ,  $\{1\}$  et  $\Omega = \{0, 1\}$ . De plus, on tient à éliminer l'impossibilité pour éviter les conflits souvent trop élevés dans l'analyse de certains types de données et surtout médicaux. Ainsi, la distribution de masse est assimilée à un triplé de valeurs réelles comprises entre 0 et 1 et dont la somme vaut l'unité. Un moyen illustratif de représenter ces distributions de masses est affiché à la FIGURE 3.9 dans laquelle la règle ayant pour longueur l'unité est divisée en trois parties de longueurs correspondant aux masses des trois ensembles d'évènements  $\{0\}$ ,  $\{1\}$  et  $\Omega$ . La masse attribuée à  $\{1\}$  est représentée en rouge, celle attribuée à  $\{0\}$  en vert et la masse attribuée à l'ignorance ou à  $\Omega$  en gris.

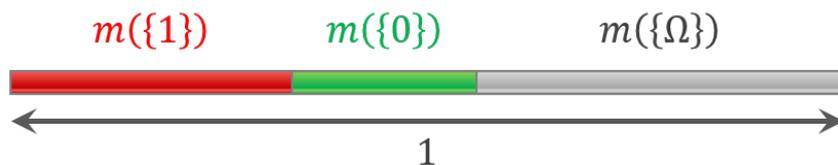


FIGURE 3.9 – Illustration d'une distribution de masse.

La méthode de combinaison retenue pour notre étude est la méthode de combinaison de YAGER [57]. Elle est adaptée à la combinaison d'un nombre important de sources d'information car elle empêche l'évanouissement de la masse de l'ignorance au cours des combinaisons enchaînées. En plus, elle n'accorde aucune masse au conflit vu quelle est transférée à l'ignorance. Il est à noter qu'un certain nombre de signaux peut être manquant sur un séjour donné, conduisant ainsi à un manque d'un certain nombre de caractéristiques. Soit  $I_{\mathcal{C},t}$  l'ensemble des caractéristiques disponibles à un instant donné  $t$ . Les fonctions de masses issues des caractéristiques indexées dans  $I_{\mathcal{C},t}$  sont combinées, pour produire une fonction de masse finale, notée  $m^{i,t}(\cdot)$ . Selon YAGER, la combinaison se fait de façon itérative. On note  $\oplus$  l'opérateur de combinaison de YAGER et  $m_{\oplus_1^s}^{i,t}(\cdot)$  la fonction de masse issue de la combinaison des  $s$  premières fonctions de masse d'indices  $u_1$  à  $u_s$  dans  $I_{\mathcal{C},t}$ . Ainsi, on a :

$$m_{\oplus_1^s}^{i,t}(\cdot) = m_{\oplus_1^{s-1}}^{i,t}(\cdot) \oplus m_{u_s}^{i,t}(\cdot), \quad \forall s \in \{2, \dots, |I_{\mathcal{C},t}|\}.$$

A chaque itération, la règle de combinaison de YAGER permet de combiner deux fonctions de masses différentes qui seront représentées par deux règles différentes que l'on va placer de manière perpendiculaire l'une à l'autre dans le but de mieux illustrer la méthode de combinaison. Le calcul des masses de la source d'information résultante de la combinaison est essentiellement basé sur les produits des masses de chacune des deux sources. Pour mieux illustrer ces produits dans notre cas, on représente dans la FIGURE 3.10 les 9 couples de masses possibles par des rectangles dont les côtés représentent les masses des sous-ensembles concernés et la surface représente le produit de ces masses. Dans cette figure, les fonctions de masse à combiner sont notées  $m_1(\cdot)$  et  $m_2(\cdot)$  par volonté de simplification et la surface  $\mathcal{M}_{\Omega_1} = m_1(\Omega).m_2(\{1\})$  par exemple. Les neuf rectangles sont colorés en rouge s'ils sont attribués à la masse finale de  $\{1\}$ , en vert s'ils sont attribués à la masse finale de  $\{0\}$  et en gris s'ils sont attribués à la masse finale de  $\Omega$  selon la règle de combinaison de YAGER. Les surfaces de conflit  $\mathcal{M}_{01}$  et  $\mathcal{M}_{10}$  sont attribuées à la masse de l'ignorance dans la source d'information résultante. La FIGURE 3.11 illustre la procédure de combinaison de YAGER d'une manière différente, toujours basée sur les produits des masses des paires de sous-ensembles de  $\Omega$ . Ainsi, la combinaison de YAGER se résume aux équations suivantes :

$$\begin{aligned} m_{\oplus_1^s}^{i,t}(\{0\}) &= m_{\oplus_1^{s-1}}^{i,t}(\{0\})m_{u_s}^{i,t}(\{0\}) + m_{\oplus_1^{s-1}}^{i,t}(\Omega)m_{u_s}^{i,t}(\{0\}) + m_{\oplus_1^{s-1}}^{i,t}(\{0\})m_{u_s}^{i,t}(\Omega), \\ m_{\oplus_1^s}^{i,t}(\{1\}) &= m_{\oplus_1^{s-1}}^{i,t}(\{1\})m_{u_s}^{i,t}(\{1\}) + m_{\oplus_1^{s-1}}^{i,t}(\Omega)m_{u_s}^{i,t}(\{1\}) + m_{\oplus_1^{s-1}}^{i,t}(\{1\})m_{u_s}^{i,t}(\Omega), \\ m_{\oplus_1^s}^{i,t}(\Omega) &= m_{\oplus_1^{s-1}}^{i,t}(\Omega)m_{u_s}^{i,t}(\Omega) + m_{\oplus_1^{s-1}}^{i,t}(\{0\})m_{u_s}^{i,t}(\{1\}) + m_{\oplus_1^{s-1}}^{i,t}(\{1\})m_{u_s}^{i,t}(\{0\}). \end{aligned}$$

La masse finale issue de la combinaison de toutes les sources s'exprime par :

$$m^{i,t}(\cdot) = m_{\oplus_1^{|I_{\mathcal{C},t}|}}^{i,t}(\cdot).$$

La prise de décision est basée sur la transformation de la fonction de masse finale en distribution de probabilité. Ceci est fait en utilisant la transformation pignistique, définie au chapitre précédent, qui consiste à répartir la masse d'un ensemble équitablement sur ses éléments. Dans notre cas, la transformation pignistique se résume aux équations suivantes :

$$\begin{aligned} P^{i,t}(0) &= m^{i,t}(\{0\}) + \frac{1}{2} m^{i,t}(\Omega), \\ P^{i,t}(1) &= m^{i,t}(\{1\}) + \frac{1}{2} m^{i,t}(\Omega), \end{aligned}$$

avec  $P^{i,t}(\cdot)$  l'estimation de la probabilité de la classe  $\cdot$  à l'instant  $t$  pour le séjour  $i$ . Le sujet est considéré comme potentiellement en danger si :

$$P^{i,t}(1) \geq \gamma,$$

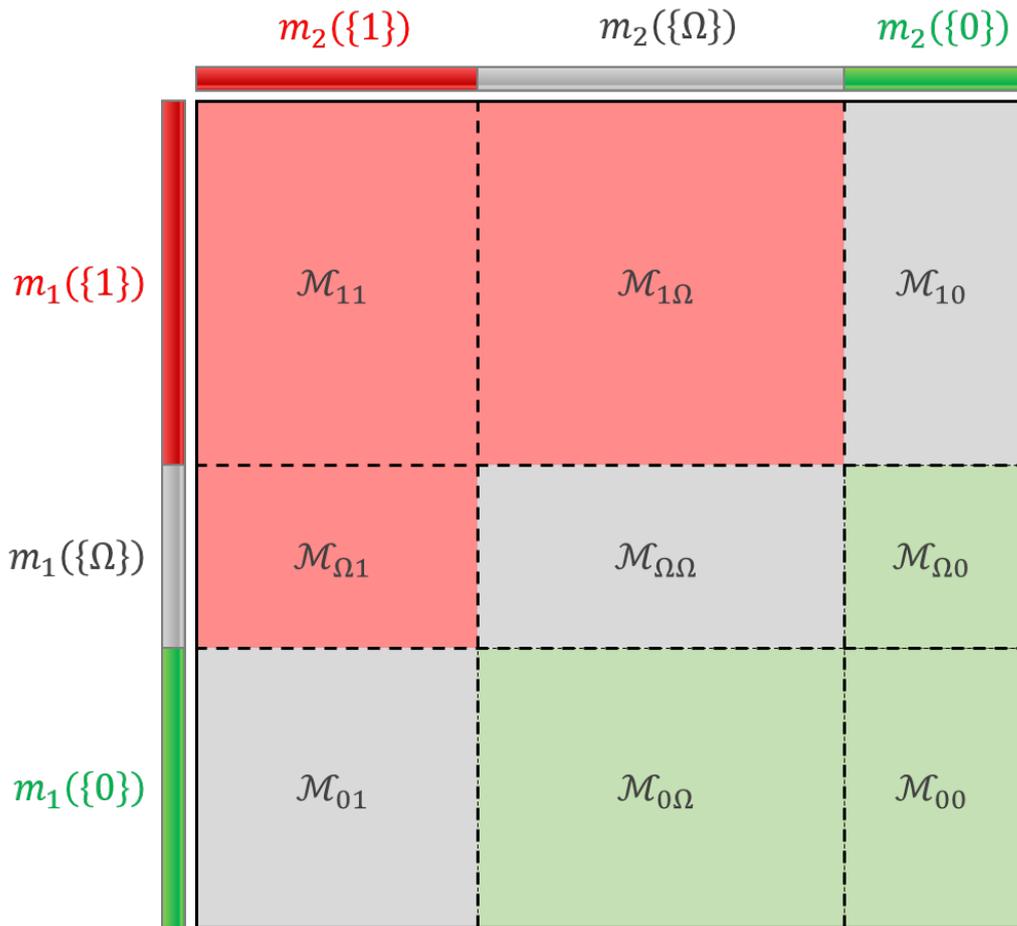


FIGURE 3.10 – Illustration de la combinaison de YAGER.

avec  $0 \leq \gamma \leq 1$  un seuil prédéfini. Lors de l'analyse en temps réel, la détection du danger peut se faire n'importe quand. Alors, pour notre étude, un séjour en USI prélevé du groupe avec FA de la base de test est considéré avoir été prédit correctement si à un certain moment durant son séjour, il est classé en danger. Ainsi, la décision relative à un séjour est :

$$\hat{y}_i = \left[ \sum_{N \leq t \leq T_i} \left[ P^{i,t}(1) \geq \gamma \right] \geq 1 \right],$$

avec  $[\cdot]$  prenant la valeur 1 si  $\cdot$  est vrai et 0 sinon.

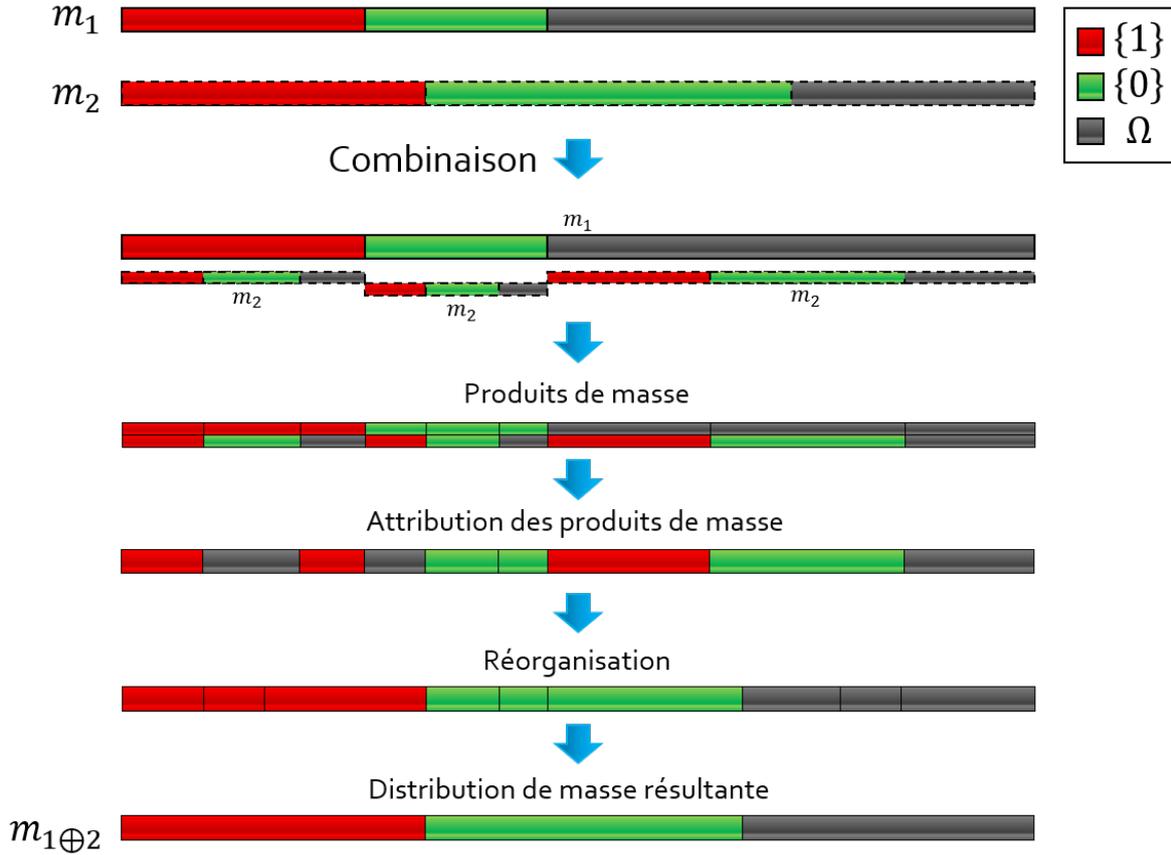


FIGURE 3.11 – Illustration de la méthode de combinaison de YAGER.

### 3.7 Résultats et discussions

#### 3.7.1 Illustration des performances

Plusieurs indicateurs peuvent témoigner des performances de la méthode de prédiction. Dans le domaine médical, certains termes sont particulièrement intéressants comme la sensibilité qui représente le taux de sujets atteints prédits correctement, la spécificité qui représente le taux de sujets témoins prédits correctement et la précision qui représente le taux de sujets prédits correctement parmi tous les sujets de test. Dans notre cas, il est aussi intéressant d'examiner la portée de prédiction qui représente la durée qui sépare l'alerte donnée par la méthode de prédiction et l'occurrence de l'épisode de FA. Ces indicateurs sont calculés à partir d'un ensemble de séjours  $I$ , comme suit :

— Sensibilité

$$Se(I) = \frac{\sum_{i \in I} y_i \hat{y}_i}{\sum_{i \in I} y_i}$$

— Spécificité

$$Sp(I) = \frac{\sum_{i \in I} (1 - y_i)(1 - \hat{y}_i)}{\sum_{i \in I} 1 - y_i}$$

— Précision

$$Pr(I) = \frac{\sum_{i \in I} \mathbb{I}[y_i = \hat{y}_i]}{|I|}$$

— Portée de prédiction

$$Po(I) = \frac{\sum_{i \in I} y_i \hat{y}_i (T_i - \min \{t \mid P^{i,t}(1) \geq \gamma\})}{\sum_{i \in I} y_i \hat{y}_i}$$

Plusieurs paramètres sont réglables avant l'exécution de l'algorithme. Ces paramètres sont tous numériques et sont listés ci-après :

- Longueur de la fenêtre glissante  $N$   
Il s'agit de la plage de durée sur laquelle se base l'algorithme pour construire un vecteur de caractéristiques et le classifier. Évidemment, les vecteurs de la base d'apprentissage sont construits en se basant sur des fenêtres de la même durée.
- Seuil  $\gamma$   
C'est le niveau de risque que le vecteur de caractéristiques doit atteindre pour que l'algorithme décide de déclencher une alerte.
- Facteur d'accentuation de l'incertitude  $\alpha$   
Ce facteur permet d'accentuer l'incertitude et de limiter le conflit dans le cas de nombreuses sources d'informations néfastes.
- Seuil de sélectivité des caractéristiques  $\beta$   
Dans la phase de la sélection des caractéristiques, celles ayant un niveau d'incertitude supérieur à  $\beta$  sont éliminées.

La méthode présentée dans ce chapitre est évaluée sur les bases d'apprentissage et de test. Dans un premier temps, nous fixons les paramètres  $N$ ,  $\alpha$  et  $\beta$  à  $N = 60$  minutes,  $\alpha = 1$  et  $\beta = 1$  et nous optimisons la valeur de  $\gamma$  sur la base d'apprentissage. Pour cela, nous faisons varier  $\gamma$  de 0 à 1 avec un pas de 0.01 et nous appliquons la méthode de la validation croisée à 10 blocs, c'est-à-dire la base d'apprentissage  $\mathcal{A}$  est divisée en 10 blocs de même taille et chaque bloc est classé en utilisant le modèle fourni par l'ensemble de tous les autres blocs de  $\mathcal{A}$ , puis la précision est calculée. Enfin, la précision moyenne pour chaque valeur de  $\gamma$  sur les 10 blocs de la base d'apprentissage est calculée et la valeur de  $\gamma$  qui donne le maximum de précision moyenne est sélectionnée. Cela conduit à  $\gamma^* = 0.91$  où la précision maximale prend la valeur de 71.77%. Le TABLEAU 3.1 illustre les valeurs de précision  $Pr$ , sensibilité  $Se$ , spécificité  $Sp$  et portée de prédiction  $Po$  sur les bases d'apprentissage (moyennes sur les 10 blocs) et de test pour un balayage des valeurs de  $\gamma$  autour de  $\gamma^*$ .

TABLEAU 3.1 – Performances de la méthode pour des valeurs autour  $\gamma^* = 0.91$  sur les bases de test et d'apprentissage.

$\gamma$	Apprentissage				Test			
	Pr	Se	Sp	Po	Pr	Se	Sp	Po
0.85	69.85%	54.30%	83.01%	148 min	70.75%	61.86%	78.26%	164 min
0.86	69.44%	51.65%	84.50%	148 min	71.46%	60.31%	80.87%	164 min
0.87	69.44%	50.11%	85.81%	151 min	71.46%	58.25%	82.61%	163 min
0.88	68.93%	47.47%	87.12%	150 min	72.64%	57.22%	85.65%	156 min
0.89	70.05%	47.02%	89.55%	143 min	73.11%	55.67%	87.83%	158 min
0.90	71.16%	47.02%	91.59%	136 min	73.58%	54.64%	89.57%	147 min
<b>0.91</b>	<b>71.77%</b>	<b>46.14%</b>	<b>93.47%</b>	<b>134 min</b>	75.00%	54.12%	92.61%	133 min
0.92	71.66%	45.27%	94.02%	135 min	75.47%	53.09%	94.35%	134 min
0.93	71.66%	44.16%	94.95%	134 min	75.24%	52.06%	94.78%	136 min
0.94	71.36%	42.17%	96.07%	138 min	<b>75.71%</b>	<b>51.55%</b>	<b>96.09%</b>	<b>122 min</b>
0.95	71.26%	40.62%	97.19%	135 min	74.76%	48.97%	96.52%	122 min
0.96	70.25%	37.74%	97.75%	130 min	73.82%	46.91%	96.52%	124 min
0.97	67.61%	29.80%	99.63%	125 min	69.58%	34.02%	99.57%	115 min

Les résultats du tableau montre que les performances varient légèrement autour de la valeur optimale de  $\gamma$ , ce qui montre une stabilité de la méthode. Par ailleurs, les résultats sur la base de test sont cohérents avec ceux de la base d'apprentissage, ce qui montre l'absence du surapprentissage du modèle, évité grâce à la validation croisée. L'utilisation de  $\gamma^*$  pour le test aboutit à une précision de 75%. La précision maximale obtenue sur la base de test est de 75.71% qui ne diffère que de 0.71% de la valeur de la précision pour  $\gamma = \gamma^*$ . Ces résultats préliminaires peuvent être améliorés en faisant varier  $\alpha$  et  $\beta$ . Le TABLEAU 3.1 montre que, à mesure que  $\gamma$  augmente, la sensibilité diminue tandis que la spécificité augmente ce qui reflète le fait qu'il devient plus difficile

pour un sujet d'être classé dans le groupe positif avec un seuil plus élevé. Selon le problème de classification,  $\gamma$  peut être ajusté pour privilégier l'une des trois mesures de performance. Pour les sujets du groupe AF prédits dans le groupe 1 (vrai positif), l'alerte est lancée en moyenne, 2 heures et 13 minutes avant l'épisode de la FA ce qui est relativement suffisant pour une intervention prématurée.

### 3.7.2 Impact des paramètres

Dans cette section, les paramètres  $\alpha$  et  $N$  représentant respectivement le facteur d'accentuation de l'incertitude et la longueur de la fenêtre de laquelle nous extrayons les caractéristiques sont optimisés sur l'ensemble des données d'apprentissage  $\mathcal{A}$  avant la phase de test alors que le paramètre  $\beta$  est fixé à  $\beta = 1$ . L'objectif de cette tâche d'optimisation est de trouver la combinaison  $(\alpha, N)$  qui maximise les performances sur la base d'apprentissage pour ensuite l'utiliser sur la base de test. Afin de réaliser cette tâche, nous nous basons sur les fonctions d'efficacité du récepteur dites courbes ROC. Ainsi, nous considérons cinq valeurs de  $N$  en minutes et trois valeurs de  $\alpha$ ,  $N \in \{15, 20, 30, 60, 120\}$  et  $\alpha \in \{1, 0.1, 0.01\}$ . Ensuite, pour chaque couple de valeurs  $(\alpha, N)$ , nous variations  $\gamma$  entre 0 et 1, par pas de 0.01, nous appliquons la méthode de validation croisée à 10 blocs sur la base d'apprentissage et calculons les performances. Enfin, nous traçons la courbe ROC, représentant  $Se$  en fonction de  $1 - Sp$  sur les différentes valeurs de  $\gamma$  pour le même couple  $(\alpha, N)$ , et nous calculons la surface sous la courbe ou AUC pour *area under the curve* qu'il faut maximiser. Le TABLEAU 3.2 montre les différentes valeurs d'AUC pour toutes les combinaisons de  $\alpha$  et  $N$ . Les résultats montrent que la combinaison  $(\alpha, N)$  donnant l'AUC maximale de 0.8585 est donnée par  $(\alpha^* = 0.1, N^* = 60 \text{ min})$ . En fixant  $N = N^*$  et en faisant varier  $\gamma$  de 0 à 1 avec un pas de 0.01, nous obtenons les courbes de la FIGURE 3.12, représentant la précision en fonction de  $\gamma$  sur la base d'apprentissage pour  $\alpha = 1$ ,  $\alpha = 0.1$  et  $\alpha = 0.01$ .

TABLEAU 3.2 – AUC pour différentes combinaisons de  $\alpha$  et  $N$ .

N	$\alpha$		
	1	0.1	0.01
15	0.7346	0.8497	0.8338
20	0.7108	0.8386	0.8479
30	0.8019	0.8315	0.8098
60	0.7362	<b>0.8585</b>	0.7323
120	0.7548	0.8474	0.8358

La FIGURE 3.12 montre qu'il est possible d'atteindre des valeurs de précision élevées en diminuant  $\alpha$  mais la bande des précisions élevées devient plus étroite, ce qui rend la valeur de la précision instable lorsque  $\gamma$  varie autour de sa valeur optimale. Afin de tester cette stabilité, les valeurs de précision, sensibilité et spécificité sont calculées sur la base de test  $\mathcal{T}$  pour  $\gamma = \gamma^*$  qui donne la précision maximale sur la base d'apprentissage avec  $\alpha = \alpha^* = 0.1$ , et pour son voisinage  $\gamma^* \pm 0.15$ . Le même calcul est fait pour  $\alpha = 1$  et  $\alpha = 0.01$ , en fixant  $N = N^* = 60 \text{ min}$ . Pour chaque valeur de  $\alpha$ , la valeur de  $\gamma$  est balayée autour de sa valeur optimale, celle maximisant la précision sur la base d'apprentissage pour le couple  $(\alpha, N)$  considéré. Le TABLEAU 3.3 montre les valeurs de précision, sensibilité et spécificité obtenues. On remarque que l'écart-type de la précision sur le voisinage du seuil optimal  $\gamma^*$  augmente lorsque  $\alpha$  diminue indiquant une forte variation de la précision autour de la valeur optimale de  $\gamma$  et donc une perte de stabilité c'est-à-dire que la précision diminue considérablement lorsque  $\gamma$  varie autour de sa valeur optimale. L'utilisateur peut favoriser la stabilité de la précision ou sa valeur en réglant  $\alpha$ .

Le dernier paramètre à optimiser est  $\beta$ , qui est lié à l'étape de sélection des caractéristiques dans la méthode proposée. Jusqu'à présent  $\beta = 1$ , cela ne correspond à aucune sélection de caractéristique. Cependant, lorsque  $\beta$  varie de 1 à 0, de plus en plus de caractéristiques sont ignorées comme expliqué dans la section 3.6.4. En fonction de la valeur de  $\alpha$ ,  $\beta$  varie dans différents inter-

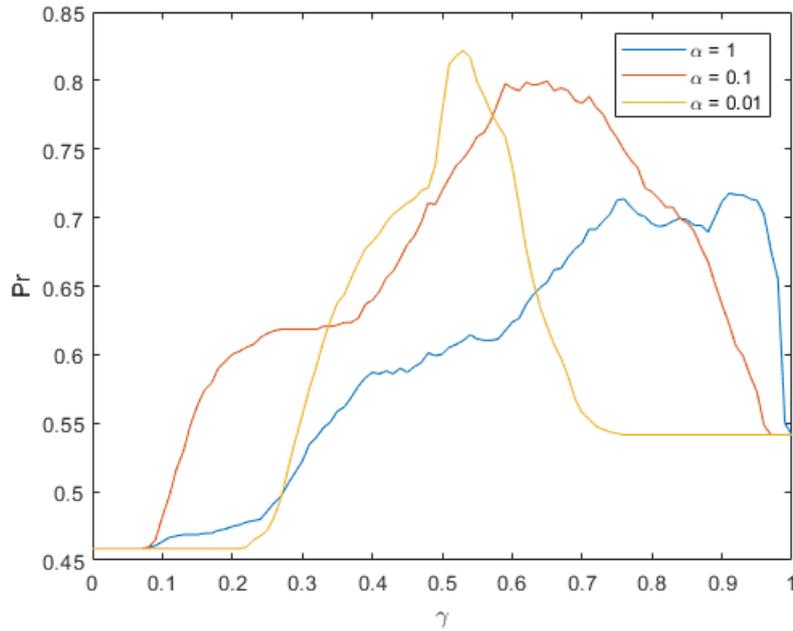


FIGURE 3.12 – Précision moyenne sur les 10 blocs de la base d'apprentissage en fonction de  $\gamma$  pour  $\beta = 1$  and  $N = 60$  min.

TABEAU 3.3 – Performance de la méthode proposée sur la base de test pour  $N=60$  min.

$\alpha$	$\gamma$	Pr	Se	Sp	$\sigma(\text{Pr})$
1	0.76	70.75%	72.68%	69.13%	0.0179
	0.81	71.70%	66.49%	76.09%	
	0.86	71.46%	60.31%	80.87%	
	<b>0.91</b>	75.00%	54.12%	92.61%	
	0.96	73.82%	46.91%	96.52%	
0.1	0.50	71.93%	97.94%	50.00%	0.0304
	0.55	78.30%	92.78%	66.09%	
	0.60	79.01%	86.08%	73.04%	
	<b>0.65</b>	80.19%	75.26%	84.35%	
	0.70	77.59%	60.82%	91.74%	
	0.75	75.94%	52.58%	95.65%	
	0.80	73.35%	45.36%	96.96%	
0.01	0.38	65.33%	100.00%	36.09%	0.0878
	0.43	69.34%	100.00%	43.48%	
	0.48	70.99%	99.48%	46.96%	
	<b>0.53</b>	83.96%	71.65%	94.35%	
	0.58	79.25%	56.70%	98.26%	
	0.63	63.92%	21.13%	100.00%	
	0.68	58.96%	10.31%	100.00%	

valles avec des pas appropriés. La précision, la sensibilité et la spécificité sur la base d'apprentissage sont indiquées dans le TABLEAU 3.4 avec la valeur optimale  $\gamma^*$  pour chaque combinaison de  $\alpha$  et  $\beta$ ,  $N$  étant fixé à 60 minutes.

D'après le TABLEAU 3.4, la variation de  $\beta$  permet d'améliorer la précision dans certains cas même si cette amélioration reste mineure. Cette stabilité de la précision montre que la sélection des caractéristiques a peu d'effet sur les performances de classification et ceci revient en partie à la bonne modélisation de l'ignorance qui empêche les caractéristiques inutiles, voir néfastes,

TABLEAU 3.4 – Performances de la méthode proposée pour différentes combinaisons de  $\alpha$  et  $\beta$  sur la base d'apprentissage pour  $N=60$  min.

$\alpha$	$\beta$	$\gamma^*$	Pr	Se	Sp
1	1.0000	0.91	71.77%	46.14%	93.47%
	0.8500	0.77	70.96%	68.86%	72.76%
	0.7000	0.79	72.48%	69.30%	75.18%
	0.5500	0.83	73.29%	69.53%	76.48%
	0.4000	0.86	77.63%	64.03%	89.18%
0.1	1.0000	0.65	79.96%	70.64%	87.85%
	0.9750	0.60	79.56%	80.14%	79.07%
	0.9500	0.61	81.07%	78.35%	83.38%
	0.9250	0.59	81.88%	70.87%	91.23%
	0.9000	0.56	80.87%	68.65%	91.23%
0.01	1.0000	0.53	81.99%	66.67%	94.96%
	0.9975	0.52	81.99%	74.17%	88.61%
	0.9950	0.53	81.58%	65.78%	94.96%
	0.9925	0.51	83.30%	72.62%	92.35%
	0.9900	0.51	82.29%	68.43%	94.04%

de participer activement à la construction de la décision finale. On choisit donc la valeur optimale de  $\beta^* = 1$  de sorte à ignorer la phase de sélection des caractéristiques. En utilisant les valeurs optimales de tous les paramètres sur la base de test, qui sont  $\alpha^* = 0.1$ ,  $\beta^* = 1$ ,  $N^* = 60$  min et  $\gamma^* = 0.65$ , nous obtenons une valeur de précision de 80.19%, une valeur de sensibilité de 75.26% et une valeur de spécificité de 84.35%. Le délai de prédiction moyen, qui est le délai entre l'alerte déclenchée par la méthode proposée et le début de l'épisode, est de 130.79 minutes ce qui signifie que la méthode proposée donne, en moyenne, une alerte 2 heures et 10 minutes avant le début de l'épisode.

### 3.7.3 Comparaison aux méthodes de la littérature

Dans cette section, la méthode proposée est d'abord comparée à d'autres méthodes de classification classique, à savoir la méthode des K plus proches voisins (KNN) [102], la méthode des K plus proches voisins pondérés (WKNN) [103] et la méthode de la machine à vecteurs de support (SVM) [104]. Afin de comparer leurs performances, un processus de sélection de caractéristiques couplé à un processus de sélection d'instances est réalisé de manière à garantir une disponibilité totale de toutes les caractéristiques de la base d'apprentissage. Par conséquent, des méthodes telles que SVM, qui nécessitent un ensemble de données entièrement disponible, peuvent accomplir la tâche de classification. Nous sélectionnons alors des caractéristiques avec un taux de disponibilité supérieur à un certain seuil. Dans notre cas, ce seuil est fixé à 95% ce qui nous donne un ensemble de 18 caractéristiques pour une longueur de fenêtre de 60 minutes. Ensuite, la base de données est nettoyée en supprimant toutes les instances ayant au moins une fonctionnalité manquante. Le TABLEAU 3.5 montre les résultats des méthodes classiques pour une longueur de fenêtre de 60 minutes. Le tableau montre également les performances moyennes avec leur écart-type (moyenne  $\pm$  écart-type) de la méthode proposée sur 15 différentes partitions aléatoires de la base de données (apprentissage/test) pour  $\alpha = \alpha^* = 0.1$  et  $\beta = \beta^* = 1$ . Les résultats du TABLEAU 3.5 montrent que les données sont insuffisantes pour construire un classifieur robuste puisque toutes les méthodes classiques conduisent à des classifieurs de faible qualité. La valeur élevée de sensibilité et la faible valeur de spécificité pour KNN et WKNN montrent que le classifieur produit presque toujours au moins une prédiction positive au cours du temps pour chaque sujet. Les méthodes qui nécessitent une disponibilité totale dans la base de données de formation sont donc inefficaces dans notre cas. La méthode proposée atteint la précision maximale du fait de sa capacité à s'adapter au manque d'informations.

TABLEAU 3.5 – Comparaison aux méthodes classiques de classification.

Méthode	Précision	Sensibilité	Spécificité
KNN	0.596 ± 0.03	0.995 ± 0.005	0.261 ± 0.032
WKNN	0.601 ± 0.027	0.997 ± 0.004	0.271 ± 0.027
SVM	0.605 ± 0.027	0.582 ± 0.454	0.623 ± 0.38
Méthode proposée	<b>0.793 ± 0.023</b>	0.742 ± 0.062	0.837 ± 0.038

La méthode proposée est ensuite comparée à d'autres méthodes récentes de prédiction de la FA à l'aide de mesures plus invasives, comme l'électrocardiogramme (ECG). Dans [105], la variabilité du contenu spectral de l'onde P est quantifiée en calculant à partir de l'estimation spectrale auto-régressive de l'onde P, l'évolution temporelle de plusieurs métriques estimant l'énergie absolue et les rapports de puissance haute fréquence à basse fréquence sur différentes bandes entre 20 et 200 Hz. Dans [72] et [71], l'algorithme génétique est utilisé pour optimiser l'ensemble de fonctionnalités extrait de la variabilité de la fréquence cardiaque, puis la machine à vecteurs de support est utilisée pour la classification. Le TABLEAU 3.6 montre la précision de ces méthodes, telles que données dans la littérature, par rapport à la nôtre. Il est clair que la précision de la méthode proposée reste dans le même ordre de grandeur que celles utilisant la mesure plus invasive ce qui valide l'utilité de cette méthode.

TABLEAU 3.6 – Comparaison à des méthodes de prédiction basées sur l'ECG dans la littérature.

Méthode	Invasivité	Précision de la prédiction
Analyse haute fréquence de l'énergie et de la durée des ondes P [105]	ECG	88.07%
Analyse des courtes séquences de variabilité de fréquence cardiaque [72]	ECG	80.20%
Analyse de variabilité de fréquence cardiaque et algorithme génétique non-dominé de trie III [71]	ECG	87.70%
Méthode proposée	Signaux physiologiques mesurables sans électrode	79.34%

### 3.8 Conclusion

Le but de ce chapitre est de proposer une nouvelle méthode pour prédire l'épisode de fibrillation atriale qui est une maladie cardiaque de type arythmie pouvant entraîner la mort. La tâche de prédiction est basée sur des techniques d'apprentissage automatique appliquées sur des données médicales cartographiées de sujets dont l'état est bien connu grâce à des fiches de diagnostic disponibles dans la base de données médicale. Le prédicteur prend en entrée les signaux physiologiques, mesurés par des capteurs non-invasifs, extrait des caractéristiques de ces signaux et les transmet à la méthode de classification. Cette dernière est basée sur la théorie des fonctions de croyance, qui est une extension de l'approche probabiliste avec la capacité de prendre en compte le manque de connaissances. Pour chaque caractéristiques, deux ensembles de valeurs sont collectés à partir de la base de données, l'un du groupe AF et l'autre du groupe non-AF. La distribution de chaque ensemble de valeurs est estimée pour pouvoir classer une nouvelle valeur à laquelle on associe une masse pour chaque état, c'est-à-dire positive ou négative, et une masse pour l'ignorance ou l'incertitude apportée par la caractéristique. Les fonctions de masse sont ensuite combinées à l'aide de la règle de combinaison YAGER et une étape de transformation nous ramène à l'approche probabiliste. La probabilité est comparée à un seuil prédéfini et une décision est prise

sur la base de cette comparaison. La précision du classificateur atteint, après optimisation de la longueur de la fenêtre et d'autres paramètres, une précision de 80.19% avec un délai de prédiction d'environ 2 heures et 10 minutes.

## Chapitre 4

# Classification évidentielle avec rejet pour la prédiction de la fibrillation atriale

### Sommaire

---

<b>4.1 Introduction</b> . . . . .	<b>74</b>
<b>4.2 Préambule et notations</b> . . . . .	<b>74</b>
<b>4.3 Modélisation avancée des fonctions de masse</b> . . . . .	<b>75</b>
<b>4.4 Méthode de combinaison associative</b> . . . . .	<b>78</b>
4.4.1 Présentation de la méthode proposée . . . . .	78
4.4.2 Transformation et décision . . . . .	79
<b>4.5 Option de rejet</b> . . . . .	<b>81</b>
<b>4.6 Résultats et discussions</b> . . . . .	<b>82</b>
4.6.1 Méthode d'attribution des masses . . . . .	82
4.6.2 Méthode associative de combinaison des masses . . . . .	83
4.6.3 Option de rejet . . . . .	84
4.6.4 Validation de l'approche globale . . . . .	85
<b>4.7 Conclusion</b> . . . . .	<b>86</b>

---

## 4.1 Introduction

Le chapitre précédent a introduit une approche de classification basée sur la théorie des fonctions de croyance. Celle-ci consiste en une extraction de caractéristiques des signaux d'entrée, puis en une conception d'un classifieur fondée sur i) une représentation gaussienne des données par caractéristique, ii) une attribution de masses à l'ignorance et aux états possibles également par caractéristique, iii) une fusion des données de toutes les fonctions de masse des caractéristiques disponibles grâce à la règle de Yager, et enfin iv) une transformation des masses en probabilités et une prise de décision. Ce chapitre présente une extension de l'approche de classification proposée dans le chapitre précédent à tous ses niveaux. En effet, il apporte différentes améliorations à chacune des étapes de la conception du classifieur pour une meilleure adaptation aux données, une meilleure interprétation des évidences, une transformation probabiliste plus ajustée et une prise de décision introduisant le rejet de décision quand celle-ci n'est pas suffisamment fiable. D'abord, la représentation gaussienne des caractéristiques profite de la disponibilité de leurs données dans la base d'apprentissage. Par ailleurs, la méthode d'attribution des fonctions de masse prend en considération plusieurs facteurs, tels que le manque de données pour chaque caractéristique de façon individuelle, la relativité des valeurs obtenues par les fonctions de densité de probabilité et le rejet des valeurs très différentes de celles présentes plus fréquemment dans la base d'apprentissage. Il s'en suit une proposition d'une méthode de combinaison plus flexible que la règle de Yager, tenant compte de l'associativité. La méthode de transformation proposée par la suite permet d'éviter l'ajustement du seuil de décision en l'estimant directement à partir de la base d'apprentissage. Finalement, afin d'utiliser efficacement la théorie des fonctions de croyance, une option de rejet est ajoutée au système de classification, permettant de mieux traduire l'ignorance finale dans la décision. Ces contributions sont évaluées à la fin du chapitre grâce à la base de données MIMIC III introduite dans le chapitre précédent.

## 4.2 Préambule et notations

Cette section introduit les notations mathématiques nécessaires pour les modélisations proposées dans ce chapitre. On dispose d'un sujet  $i$  dont une série de signaux physiologiques est mesurée en temps réel, noté  $t$ . En tronquant la dernière fenêtre d'enregistrement, dont la taille  $N$  est prédéfinie et se terminant à  $t$ , on extrait des caractéristiques qui forment un vecteur de caractéristiques  $\mathbf{x}_i(t)$  à chaque instant  $t$ . Pour une raison de simplification, on retirera l'indice  $i$  quand il s'agit d'un sujet à évaluer en temps réel et son vecteur de caractéristiques sera noté  $\mathbf{x}(t)$  dans ce chapitre. La dimension de ce vecteur correspond au nombre total de caractéristiques, noté  $d$ ,  $\mathbf{x}(t) \in \mathbb{R}^d$ . Les composantes du vecteur  $\mathbf{x}(t)$  sont notées  $x_u(t)$  avec  $1 \leq u \leq d$ . On a donc  $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_d(t))$ . Il est à noter que certaines caractéristiques peuvent être indisponibles à l'instant  $t$ , soit à cause de l'indisponibilité du signal lui-même dont la caractéristique est extraite, soit à cause des contraintes mathématiques imposées par la caractéristique et insatisfaites par le signal. Il est alors nécessaire d'indexer les valeurs disponibles du vecteur de caractéristiques dans un ensemble d'indices  $I_{\mathcal{C},t}$ . Dans ce chapitre, on introduit le vecteur de disponibilité  $\mathbf{a}(t) \in \{0, 1\}^d$ . Les composantes de ce vecteur sont notées  $a_u(t)$  avec  $1 \leq u \leq d$ . On pose  $a_u(t) = 1$  si la caractéristique  $x_u(t)$  est disponible, c'est-à-dire si  $u \in I_{\mathcal{C},t}$ , et  $a_u(t) = 0$  sinon. On dispose pour l'entraînement du classifieur d'une base d'apprentissage

$$\mathcal{A} = \{(\mathbf{x}_i, y_i) \mid 1 \leq i \leq m\}$$

avec  $m$  le nombre d'instances d'apprentissage,  $\mathbf{x}_i \in \mathbb{R}^d$  et  $y_i \in \{0, 1\} \quad \forall 1 \leq i \leq m$ . Le vecteur  $\mathbf{x}_i$  est le vecteur de caractéristiques extrait de la dernière fenêtre de taille  $N$  disponible avant la décharge du sujet de l'USI pour les sujets non-atteints de FA et avant l'instant de l'épisode de FA pour les sujets ayant subi ce malaise. La variable  $y_i$  prend la valeur 1 si le sujet  $i$  a subi un épisode de FA durant son séjour à l'USI et prend la valeur 0 sinon. Dans ce chapitre, on prend en compte la disponibilité des caractéristiques dans la base d'apprentissage, ce qui n'était pas le cas dans le

chapitre précédent où les caractéristiques étaient supposées toutes disponibles pour l'apprentissage. Des vecteurs de disponibilités  $\mathbf{a}_i \in \{0, 1\}^d$  sont alors introduits pour  $1 \leq i \leq m$ , avec  $a_{i,u} = 1$  si la caractéristique  $u$  de la donnée d'apprentissage  $i$  est disponible et 0 sinon.

### 4.3 Modélisation avancée des fonctions de masse

Afin de modéliser les données d'apprentissage, deux distributions gaussiennes sont adaptées aux valeurs disponibles dans la base d'apprentissage pour chaque caractéristique  $u$ ,  $1 \leq u \leq d$ . La première  $\mathcal{N}(\hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})$  est réservée au groupe positif, atteint de FA, alors que la deuxième  $\mathcal{N}(\hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})$  est réservée au groupe négatif, dit groupe de contrôle. On a alors,

$$\begin{aligned}\hat{\mu}_u^{(1)} &= \frac{1}{\sum_{1 \leq i \leq m} a_{i,u} y_i} \sum_{1 \leq i \leq m} x_{i,u} y_i a_{i,u}, \\ \hat{\mu}_u^{(0)} &= \frac{1}{\sum_{1 \leq i \leq m} a_{i,u} (1 - y_i)} \sum_{1 \leq i \leq m} x_{i,u} (1 - y_i) a_{i,u}, \\ \hat{\sigma}_u^{(1)} &= \sqrt{\frac{1}{\sum_{1 \leq i \leq m} a_{i,u} y_i} \sum_{1 \leq i \leq m} (x_{i,u} - \hat{\mu}_u^{(1)})^2 y_i a_{i,u}}, \\ \hat{\sigma}_u^{(0)} &= \sqrt{\frac{1}{\sum_{1 \leq i \leq m} a_{i,u} (1 - y_i)} \sum_{1 \leq i \leq m} (x_{i,u} - \hat{\mu}_u^{(0)})^2 (1 - y_i) a_{i,u}}.\end{aligned}$$

Dans la suite,  $\phi(\cdot, \mu, \sigma)$  désignera la fonction de densité de probabilités de la loi gaussienne de moyenne  $\mu$  et d'écart-type  $\sigma$ , appliquée à l'argument  $\cdot$ .

L'objectif suivant est de modéliser les fonctions de masses  $m_u^t(\cdot)$  proposant une répartition des masses entre l'ignorance  $\Omega$ , l'évènement  $\{1\}$  et l'évènement  $\{0\}$  pour chaque caractéristique  $u$  en se basant sur sa valeur  $x_u(t)$ , telle que :

$$m_u^t(\{1\}) + m_u^t(\{0\}) + m_u^t(\Omega) = 1.$$

La modélisation du chapitre précédent se basait sur la surface de chevauchement entre les courbes des densités de probabilités des groupes pour attribuer la masse à l'ignorance. Cette modélisation ne profitait ni de la valeur ponctuelle de la caractéristique, ni de la disponibilité de ses valeurs. Ce chapitre propose une nouvelle modélisation prenant en compte diverses notions pour la représentation de l'ignorance comme suit :

- La disponibilité des caractéristiques dans la base d'apprentissage. Moins la caractéristique est disponible pour la construction des distributions, moins elle doit se prononcer, plus son ignorance doit être élevée;
- La différence entre les probabilités de la valeur ponctuelle de la caractéristique  $x_u(t)$  selon les distributions des groupes 0 et 1. Plus les probabilités sont proches, moins on est sûr de l'appartenance de la mesure à l'une ou l'autre des groupes, moins la caractéristique doit se prononcer ponctuellement, plus son ignorance doit être élevée;
- L'éloignement de la valeur ponctuelle de la caractéristique  $x_u(t)$  de celles de la base d'apprentissage. Plus la caractéristique ponctuelle à  $t$  diverge de la base d'apprentissage, moins elle doit se prononcer ponctuellement, plus son ignorance doit être élevée.

Ces facteurs peuvent ainsi rendre une caractéristique inefficace dans le processus de classification. Cela doit se traduire par une masse élevée attribuée à l'ignorance pour réduire l'impact de cette caractéristique sur la décision finale.

Afin de construire l'expression mathématique de la masse de l'ignorance, le raisonnement suivant est proposé. D'abord, pour certaines caractéristiques, la disponibilité est tellement faible que les valeurs restantes sont insuffisantes pour représenter ou discriminer les deux groupes. Le

nombre de valeurs disponibles pour l'adaptation d'une fonction de densité de probabilité est donc un atout essentiel à la bonne estimation de cette dernière. Il faut donc à la fois que les nombres des valeurs disponibles pour les deux groupes soient assez élevés pour réduire l'ignorance. On note  $n_u^{(1)}$  le nombre de valeurs disponibles pour la caractéristique  $u$  dans la base d'apprentissage pour le groupe positif et  $n_u^{(0)}$  celui pour le groupe négatif, alors :

$$n_u^{(1)} = \sum_{1 \leq i \leq m} a_{i,u} y_i,$$

$$n_u^{(0)} = \sum_{1 \leq i \leq m} a_{i,u} (1 - y_i).$$

La masse de l'ignorance doit être une fonction décroissante de ces deux nombres. En plus, cette décroissance doit être plus importante lorsque ces termes sont relativement faibles, ce qui implique que la variation de la masse d'ignorance par rapport à ces deux termes doit être proportionnelle à la valeur de la masse d'ignorance elle-même. Cette disposition d'équation différentielle du premier ordre conduit à une relation exponentielle et le terme suivant sera donc utilisé dans le calcul de la masse de  $\Omega$ ,

$$e^{-\frac{n_u^{(1)} n_u^{(0)}}{\nu^2}},$$

avec  $\nu$  un terme de régularisation à optimiser. Ainsi,  $m_u^t(\Omega)$  sera une fonction croissante de  $e^{-\frac{n_u^{(1)} n_u^{(0)}}{\nu^2}}$ .

Un autre facteur important pour la modélisation de  $m_u^t(\Omega)$  est le rapport entre les valeurs ponctuelles prises par les fonctions de densité de probabilité à la mesure  $x_u(t)$ ,  $\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})$  et  $\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})$ . Plus ces valeurs sont espacées, plus l'appartenance de la mesure à l'un des groupes est sûre et moins il y a d'ambiguïté. La notion d'espacement ici évoquée, réfère à un espacement géométrique et non arithmétique car la différence arithmétique entre des valeurs issues de fonctions de densité de probabilité ne représente pas effectivement la notion de dominance d'une distribution de probabilité sur une autre en un point précis. Cette dominance est mieux représentée par la métrique géométrique qui en se basant sur le quotient de la division entre les deux valeurs, indique un facteur multiplicatif et sans unité entre ces valeurs. On a donc recours à l'écart-type géométrique pour représenter le niveau d'ignorance qu'une source d'information peut induire à l'issue des valeurs des deux fonctions de densité de probabilité. La moyenne géométrique et l'écart-type géométrique de  $\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})$  et de  $\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})$  sont définies comme suit :

$$\mu_{g_u}^t = \sqrt{\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}) \cdot \phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})},$$

$$\sigma_{g_u}^t = \exp \sqrt{\frac{(\log \phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}) - \log \mu_{g_u}^t)^2 + (\log \phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)}) - \log \mu_{g_u}^t)^2}{2}}.$$

L'écart-type géométrique de ces deux valeurs peut être simplifié comme suit :

$$\sigma_{g_u}^t = \exp \frac{1}{2} \left| \log \frac{\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})}{\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})} \right|.$$

La quantité  $\sigma_{g_u}^t$  est supérieure à 1 et plus elle est grande, plus les probabilités sont différentes l'une de l'autre, plus la masse de  $\Omega$  doit être faible. Afin de limiter la valeur de la masse de  $\Omega$  entre 0 et 1, ce terme est inversé puis, pour contrôler l'impact de ce facteur, élevé à une puissance  $\varphi$ . Le terme suivant est ainsi proposé dans le calcul de la masse de  $\Omega$  :

$$e^{-\frac{\varphi}{2} \left| \log \frac{\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})}{\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})} \right|}.$$

La masse  $m_u^t(\Omega)$  sera une fonction croissante de ce facteur.

Enfin, lorsque la valeur de la caractéristique dans le vecteur à classifier  $x_u(t)$  est très différente de toutes celles présentes dans la base d'apprentissage, l'ignorance doit être élevée car il est préférable que cette caractéristique ne participe pas activement à la décision finale. Afin de représenter correctement cette notion, on introduit la distribution de mélange gaussien bimodale dont la fonction de densité de probabilité est

$$f_u(x) = \frac{n_u^{(1)}}{m} \phi(x, \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}) + \frac{n_u^{(0)}}{m} \phi(x, \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)}).$$

Plus la valeur de la densité de probabilité de cette distribution appliquée à la mesure  $x_u(t)$  est petite, plus la valeur  $x_u(t)$  est loin de la base d'apprentissage, plus l'ignorance doit être grande. Pour assurer que la valeur du terme à utiliser dans la masse est comprise entre 0 et 1, la fonction est divisée par son maximum. En outre, pour contrôler l'effet de ce facteur, on élève le rapport à la puissance  $\rho$ , un paramètre à réguler. L'ignorance dépend donc aussi du terme :

$$\left( \frac{f_u(x_u(t))}{\max_{x \in \mathbb{R}} (f_u(x))} \right)^\rho,$$

la masse  $m_u^t(\Omega)$  devant être une fonction décroissante de ce facteur.

Malgré le fait que les différents facteurs doivent agir ensemble, il est intuitif d'affirmer que lorsque l'un des facteurs annonce une ignorance totale, l'ignorance doit être élevée indépendamment des valeurs des autres facteurs. Pour cette raison, les termes seront ajustés dans un produit de manière à obtenir un élément absorbant lorsqu'un facteur déclare une ignorance totale. La masse de l'ignorance sera alors modélisée comme suit :

$$m_u^t(\Omega) = 1 - \left( 1 - e^{-\frac{n_u^{(1)} n_u^{(0)}}{v^2}} \right) \left( 1 - e^{-\frac{\rho}{2} \left| \log \frac{\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})}{\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})} \right|} \right) \left( \frac{f_u(x_u(t))}{\max_{x \in \mathbb{R}} (f_u(x))} \right)^\rho. \quad (4.1)$$

Même si ce modèle prend en compte trois facteurs simultanément, il est possible d'annuler l'effet de chaque facteur en procédant aux paramétrages suivants :

$$v \rightarrow 0,$$

$$\rho \rightarrow +\infty,$$

$$\rho \rightarrow 0.$$

Cette manipulation nous permet d'identifier l'impact de chaque facteur sur les performances du classifieur.

Le reste des masses est calculé comme expliqué au chapitre précédent. En effet, lorsque la fonction de densité de probabilité pour une certaine caractéristique est élevée, elle indique une concentration importante des valeurs de cette caractéristique dans la base d'apprentissage du groupe qu'elle représente. On a donc choisi d'attribuer les masses aux singletons  $\{0\}$  et  $\{1\}$  pour chaque caractéristique proportionnellement aux valeurs prises par les fonctions de densité de probabilité pour la valeur de cette caractéristique du vecteur à classifier. Ainsi, on doit assurer que :

$$\frac{m_u^t(\{1\})}{m_u^t(\{0\})} = \frac{\phi(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)})}{\phi(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)})},$$

avec  $m_u^t(\{1\}) + m_u^t(\{0\}) + m_u^t(\Omega) = 1$ . Cela conduit à :

$$m_u^t(\{1\}) = \frac{\phi\left(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}\right)}{\phi\left(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}\right) + \phi\left(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)}\right)} (1 - m_u^t(\Omega)),$$

$$m_u^t(\{0\}) = \frac{\phi\left(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)}\right)}{\phi\left(x_u(t), \hat{\mu}_u^{(1)}, \hat{\sigma}_u^{(1)}\right) + \phi\left(x_u(t), \hat{\mu}_u^{(0)}, \hat{\sigma}_u^{(0)}\right)} (1 - m_u^t(\Omega)).$$

## 4.4 Méthode de combinaison associative

### 4.4.1 Présentation de la méthode proposée

La méthode de combinaison utilisée dans le chapitre précédent est la méthode de combinaison de Yager. Selon cette règle, la combinaison d'un grand nombre de sources d'information est réalisée, de manière itérative, de sorte à ce que la masse du conflit soit transférée à l'ignorance à chaque itération. Les sources d'information ont donc été combinées les unes après les autres. L'opérateur de combinaison n'étant pas associatif, ceci est susceptible de pénaliser les performances de la combinaison en rendant la masse finale dépendante de l'ordre de combinaison des sources choisi. L'objectif de la combinaison étant de fusionner les données de toutes les sources simultanément, il est préférable que la construction de la décision finale soit indépendante d'un quelconque choix de permutation des sources d'information. Pour palier à ce problème de dépendance à l'ordre de combinaison des sources, cette section propose une nouvelle méthode qui combine simultanément plusieurs sources d'information, tout en associant le conflit à l'ignorance. Afin de visualiser cette méthode, on considère tous les produits des masses issues des différentes fonctions de masse à combiner. Chaque distribution de masse  $u$  dans le cadre de notre étude est représentée par trois nombres réels  $m_u^t(\{1\})$ ,  $m_u^t(\{0\})$  et  $m_u^t(\Omega)$ , dont la somme vaut l'unité,  $1 \leq u \leq d$ . Un produit  $\ell$  des masses est de la forme :

$$\prod_{u=1}^d m_u^t(A_{\ell,u}) = m_1^t(A_{\ell,1}) \cdot m_2^t(A_{\ell,2}) \cdot m_3^t(A_{\ell,3}) \dots m_{d-1}^t(A_{\ell,d-1}) \cdot m_d^t(A_{\ell,d}),$$

avec  $A_{\ell,u} \in \{\{1\}, \{0\}, \Omega = \{0, 1\}\}$ ,  $\forall 1 \leq u \leq d$ . Ce produit est caractérisé par la séquence  $\mathbf{A}_\ell = (A_{\ell,1}, A_{\ell,2}, A_{\ell,3}, \dots, A_{\ell,d-1}, A_{\ell,d})$ . On dénombre  $3^d$  séquences différentes ( $\mathbf{A}_\ell$ ) qui sont liées chacune à un produit de masse différent dont la somme vaut l'unité :

$$\sum_{\mathbf{A}_\ell \in \{\{0\}, \{1\}, \Omega\}^d} \prod_{u=1}^d m_u^t(A_{\ell,u}) = 1.$$

Une procédure pour combiner les sources d'informations et obtenir une fonction de masse unique serait de distribuer tous ces produits de masses sur les masses finales de  $\{0\}$ ,  $\{1\}$  et  $\Omega$  selon la méthode de combinaison choisie. Par exemple, l'application itérative de la méthode de combinaison de Yager revient à attribuer chaque produit de masse  $\ell$  caractérisé par la séquence  $\mathbf{A}_\ell$  à la masse de  $\{1\}$  si les deux derniers singletons ( $\{0\}$  ou  $\{1\}$ ) dans  $\mathbf{A}_\ell$  sont  $\{1\}$  ou si un seul singleton existe dans  $\mathbf{A}_\ell$  et qu'il est  $\{1\}$ , à la masse de  $\{0\}$  si les deux derniers singletons dans  $\mathbf{A}_\ell$  sont  $\{0\}$  ou si un seul singleton existe dans  $\mathbf{A}_\ell$  et qu'il est  $\{0\}$  et finalement à  $\Omega$  si les deux derniers singletons dans  $\mathbf{A}_\ell$  sont différents ou s'il n'existe aucun singleton dans  $\mathbf{A}_\ell$ . Cette référence aux deux derniers éléments de cardinal 1 dans la séquence  $\mathbf{A}_\ell$  démontre bien que la méthode de combinaison de Yager n'est pas associative et que la fonction de masse finale peut être différente si les fonctions de masse sont combinées en suivant un ordre différent.

Pour introduire la méthode proposée, il est nécessaire de définir la fonction  $\kappa(\cdot)$  qui associe à  $\{0\}$  la valeur  $\kappa(\{0\}) = -1$ , à  $\{1\}$  la valeur  $\kappa(\{1\}) = 1$  et à  $\Omega$  la valeur  $\kappa(\Omega) = 0$ . Ainsi, la méthode proposée attribue un produit  $\ell$  de masse issue de la séquence  $\mathbf{A}_\ell$  à la masse finale de  $\{1\}$  si

$\sum_{u=1}^d \kappa(A_{\ell,u}) > c$ , à la masse finale de  $\{0\}$  si  $\sum_{u=1}^d \kappa(A_{\ell,u}) < -c$  et finalement à la masse finale de  $\Omega$  si  $|\sum_{u=1}^d \kappa(A_{\ell,u})| \leq c$  avec  $c$  un paramètre à optimiser qui indique la fiabilité générale des sources d'information. Évidemment, si les éléments dans  $A_{\ell}$  sont permutés, ceci n'aura pas d'effet sur la somme  $\sum_{u=1}^d \kappa(A_{\ell,u})$  et ne changera donc pas la fonction de masse finale.

Afin de reformuler cette méthode, on s'appuie sur une approche polynomiale, équivalente à utiliser le produit de convolution discret pour effectuer la combinaison. Chaque distribution de masse dans le cadre de notre étude est représentée par trois nombres réels dont la somme vaut l'unité. La distribution de masse est alors représentée dans cette section par un polynôme. Ainsi, pour une fonction de masse  $m_u^t(\cdot)$ , le polynôme la représentant est noté  $Q_u^t(z)$  et suit l'équation :

$$Q_u^t(z) = m_u^t(\{0\}) + m_u^t(\Omega)z + m_u^t(\{1\})z^2.$$

La combinaison est réalisée en calculant le produit de tous les polynômes représentatifs des sources d'information à combiner :

$$Q^t(z) = \prod_{u=1}^d Q_u^t(z).$$

On considère par la suite la somme des coefficients des monômes de  $Q^t(z)$  dont la puissance est strictement supérieure à  $d$ , égale à  $d$ , puis strictement inférieure à  $d$  qu'on attribue à la masse de  $\{1\}$ ,  $\{0\}$  et  $\Omega$  respectivement. Soit  $c_j^t$  le  $j^{\text{ème}}$  coefficient du polynôme  $Q^t(z)$ , multipliant  $z^j$ , avec  $0 \leq j \leq 2d$ . Les équations ci-dessous présentent un moyen mathématique de calculer les masses finales :

$$m^t(\{1\}) = \sum_{j=d+1}^{2d} c_j^t,$$

$$m^t(\Omega) = c_d^t,$$

$$m^t(\{0\}) = \sum_{j=0}^{d-1} c_j^t.$$

Comme le produit polynomial est un opérateur associatif, la méthode de combinaison l'est aussi. Si certaines caractéristiques sont manquantes à un instant  $t$ , elles seront omises de la combinaison et la méthode proposée sera très facilement adaptée comme suit :

$$Q^t(z) = \prod_{s=1}^{d'} Q_{u_s}^t(z),$$

avec  $d' = \sum_1^d a_u(t)$  le nombre des caractéristiques disponibles et  $u_s$ ,  $1 \leq s \leq d'$ , leurs indices pris dans  $\{1, \dots, d\}$  de sorte que  $a_{u_s}(t) = 1$ . Enfin,

$$m^t(\{1\}) = \sum_{j=d'+1}^{2d'} c_j^t,$$

$$m^t(\Omega) = c_{d'}^t,$$

$$m^t(\{0\}) = \sum_{j=0}^{d'-1} c_j^t.$$

#### 4.4.2 Transformation et décision

La transformation, dans le cadre de la théorie des fonctions de croyance, est la manipulation qui consiste à transformer une approche évidentielle, modélisée par des masses, en une approche probabiliste. La méthode de transformation pignistique repose sur le partage des masses des ensembles de cardinalité supérieure ou égale à deux équitablement sur les éléments le constituant. Dans le cas de notre étude, l'ensemble des évènements est défini par  $\Omega = \{0, 1\}$ . Ainsi, ayant une

fonction de masse  $m^t(\cdot)$ , on obtient la distribution de probabilité suivante :

$$P^t(0) = m^t(\{0\}) + \frac{1}{2}m^t(\Omega),$$

$$P^t(1) = m^t(\{1\}) + \frac{1}{2}m^t(\Omega).$$

Ce raisonnement est fondé sur le principe du raisonnement insuffisant (PRI) qui implique qu'en cas d'ignorance sur les distributions probabilistes, on suppose la probabilité distribuée équitablement sur les évènements possibles. Dans le cadre de la théorie des fonctions de croyance, la masse de  $\Omega$  représente une part de probabilité dont on ignore l'appartenance, on sait seulement qu'elle est distribuée sur les éléments de  $\Omega$ . Puisque dans notre étude, en cas d'ignorance totale sur le diagnostic d'un patient (comme par exemple avant toute analyse médicale et observations quelconques de symptômes), on peut toujours modéliser la distribution de probabilité selon les connaissances statistiques sur la maladie à diagnostiquer, on peut alors trancher la masse de  $\Omega$  d'une manière plus fidèle à la réalité en se penchant sur la distribution dans l'échantillon d'apprentissage. En effet, un épisode de FA ou plus généralement, un malaise quelconque lié à une maladie cardio-vasculaire peut avoir des répercussions de durées différentes sur les signaux physiologiques ce qui les rend plus ou moins rares. Il est donc judicieux de traiter la masse d'ignorance selon cette probabilité "a priori". La FIGURE 4.1 présente la méthode de transformation proposée

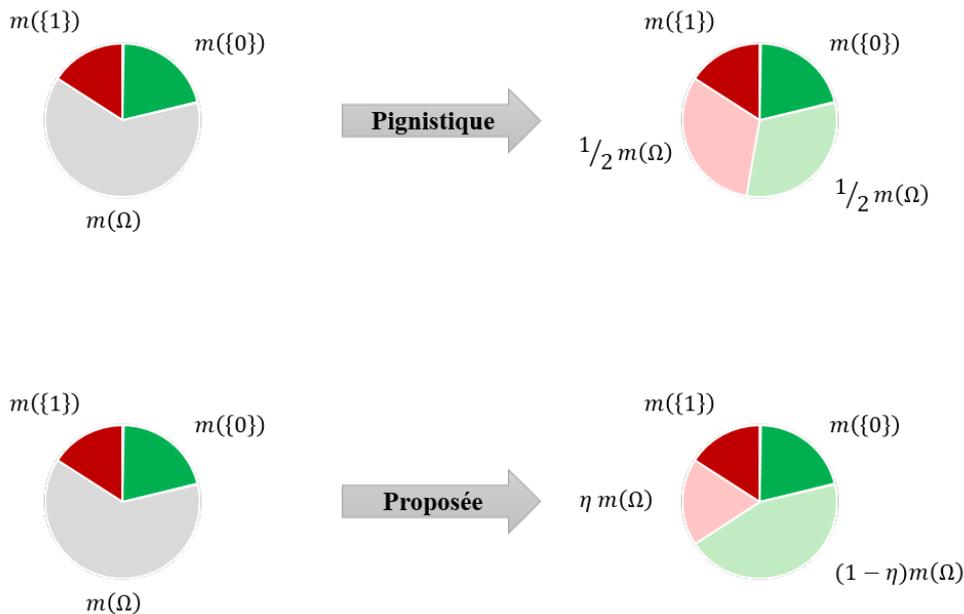


FIGURE 4.1 – Illustration de la méthode de transformation proposée.

en soulignant la différence avec la méthode de transformation pignistique. La variable  $\eta$  est ajustée selon les connaissances a priori de la maladie. Dans notre cas, on se sert de la répartition des éléments dans la base d'apprentissage sur les deux évènements 0 et 1 :

$$\eta = \frac{1}{m} \sum_{i=1}^m y_i,$$

et on a alors :

$$P^t(0) = m^t(\{0\}) + (1 - \eta)m^t(\Omega),$$

$$P^t(1) = m^t(\{1\}) + \eta m^t(\Omega).$$

Finalement, la décision instantanée est obtenue en comparant la probabilité  $P^t(1)$  au seuil  $\gamma \in [0, 1]$  :

$$h(\mathbf{x}(t)) = \llbracket P^t(1) \geq \gamma \rrbracket,$$

avec  $\llbracket \cdot \rrbracket = 1$  si  $\cdot$  est vraie et 0 sinon. Dans le but d'obtenir le seuil  $\gamma$  optimal qui sert à la décision, on valide sa valeur sur la base d'apprentissage. Le modèle est ainsi testé sur la base d'apprentissage en faisant varier  $\gamma$  jusqu'à obtenir celle qui maximise la précision (minimise l'erreur) :

$$\gamma^* = \operatorname{argmin}_{\gamma \in [0,1]} \frac{1}{m} \sum_{i=1}^m |h(\mathbf{x}_i) - y_i|.$$

Il est à noter que dans la base d'apprentissage à ce stade, les instances sont intentionnellement équilibrées sur les deux groupes FA et non-FA. Cela conduit à  $\eta = 0.5$  et donc à une transformation équivalente à la transformation pignistique dans ce chapitre. Cette méthode de transformation se révèle plus utile dans le cas d'introduction des données statistiques de la maladie ou dans une approche multi-pathologique comme dans le chapitre suivant.

## 4.5 Option de rejet

L'approche proposée ci-dessus rend une décision précise appartenant à l'ensemble  $\{0, 1\}$ . Afin d'ajouter une option de rejet, on introduit une sortie supplémentaire traduisant la capacité de l'approche à ne prendre aucune décision si l'instance est ambiguë. En effet, partant du vecteur de caractéristiques  $\mathbf{x}(t)$  d'un sujet de la base de test, le classifieur avec rejet aura trois sorties possibles,  $h_r(\mathbf{x}(t)) = 1$  lorsque le sujet présente un risque considéré d'épisode AF,  $h_r(\mathbf{x}(t)) = 0$  lorsque le sujet a un très faible risque et  $h_r(\mathbf{x}(t)) = -1$  pour "état inconnu" lorsque la décision semble difficile à prendre,  $h_r(\cdot)$  étant la fonction de décision instantanée avec rejet. Ceci permet au système d'éviter de prendre des décisions hâtives et peut se traduire par la proposition du système de plus d'analyses médicales ou l'obtention des mesures de signaux physiologiques supplémentaires plus invasives. Pour aboutir à cet objectif, on définit le risque instantané comme étant la probabilité de l'état 1, obtenue grâce à la transformation pignistique,

$$R(t) = P^t(1) = m^t(\{1\}) + \frac{1}{2} m^t(\Omega).$$

Si une décision binaire est nécessaire (sans rejet), un sujet est considéré comme positif à la FA lorsque sa courbe de risque franchit un certain seuil  $\gamma$ . Cependant, dans les applications médicales et lorsque l'enjeu est assez important, il est recommandé d'éviter de prendre une décision si le résultat n'est pas sûr. Pour ce faire, on construit un intervalle autour de ce risque, conduisant à deux courbes de risque pour chaque sujet au cours du temps. L'une aura une allure optimiste et l'autre pessimiste. Le système décide de rejeter une classification si les deux points de vue optimiste et pessimiste ne s'accordent pas à sortir le même résultat de classification.

Afin de représenter les courbes de risque optimiste et pessimiste, on se base sur la masse de l'ignorance  $m^t(\Omega)$ , étant la partie de la probabilité dont on ignore l'appartenance. Soit  $\epsilon \in [0, 1]$  un paramètre d'optimisme. Alors, les risques optimiste et pessimiste  $\bar{R}(t)$  et  $\underline{R}(t)$  sont définis comme suit :

$$\begin{aligned} \bar{R}(t) &= m^t(\{1\}) + \frac{1+\epsilon}{2} m^t(\Omega) = R(t) + \frac{\epsilon}{2} m^t(\Omega), \\ \underline{R}(t) &= m^t(\{1\}) + \frac{1-\epsilon}{2} m^t(\Omega) = R(t) - \frac{\epsilon}{2} m^t(\Omega). \end{aligned}$$

Grâce au paramètre  $\epsilon$ , un intervalle de risque ayant  $\underline{R}(t)$  et  $\bar{R}(t)$  comme bornes inférieure et supérieure est défini. Cet intervalle est centré sur la probabilité pignistique  $P^t(1)$  de 1 avec une largeur de  $\epsilon \cdot m(\Omega)$ . La fonction de décision  $h_r(\cdot)$  est alors donnée par :

$$h_r(\mathbf{x}(t)) = 1 \text{ si } \bar{R}(t) \geq \underline{R}(t) \geq \gamma,$$

$$h_r(\mathbf{x}(t)) = 0 \text{ si } \underline{R}(t) \leq \bar{R}(t) \leq \gamma,$$

$$h_r(\mathbf{x}(t)) = -1 \text{ si } \underline{R}(t) < \gamma < \bar{R}(t).$$

Si  $\epsilon = 0$ , les frontières de l'intervalle de risque se confondent et ceci revient à une seule courbe égale à la probabilité pignistique  $P^t(1)$ , les deux points de vue sont alors confondus et l'on aura  $h_r(\cdot) = h(\cdot)$ . Si  $\epsilon = 1$ , l'intervalle est le plus grand possible avec une largeur de  $m(\Omega)$ . Dans ce cas, c'est la masse entière de  $\Omega$  qui est attribuée différemment sur les probabilités des évènements 0 et 1 selon les points de vue optimiste et pessimiste. Ce paramètre représente donc l'intensité de la divergence entre les points de vue et plus sa valeur est élevée, plus le système sera amené à rejeter des instances. Enfin, la prédiction est annoncée pour le patient suivant les valeurs de  $h_r(\mathbf{x}_t)$  pour  $t$  variable. Si  $\exists t \mid h_r(\mathbf{x}(t)) = 1$ , le patient est classé dans le groupe FA ( $\hat{y} = 1$ ), si  $\forall t \mid h_r(\mathbf{x}(t)) = 0$ , le patient est classé dans le groupe non-FA ( $\hat{y} = 0$ ) et enfin si  $\forall t \mid h_r(\mathbf{x}(t)) \leq 0$  et  $\exists t \mid h_r(\mathbf{x}_t) = -1$  le patient n'est pas classé ( $\hat{y} = -1$ ).

## 4.6 Résultats et discussions

Cette section illustre les performances de l'approche proposée en l'appliquant à la base de données médicale MIMIC III. Pour cela, chaque étape de l'approche est évaluée séparément, en la couplant à l'approche proposée au chapitre précédent. D'abord, la méthode d'attribution de masses est considérée et sa pertinence est évaluée en la comparant à celle exploitant la surface de chevauchement des distributions des caractéristiques. Ensuite, la méthode de combinaison associative est comparée à la méthode de Yager. Puis, l'option de rejet est appliquée à la suite de l'approche précédente. L'approche complète est enfin illustrée avec son option de rejet.

### 4.6.1 Méthode d'attribution des masses

Dans ce paragraphe, nous évaluons la modélisation de masse proposée dans ce chapitre. Pour cela, nous l'utilisons d'abord pour la représentation de l'information par caractéristique dans l'approche du chapitre 3. Ainsi, nous appliquons la méthode de Yager pour la combinaison et la transformation des masses en probabilités grâce à la transformation pignistique. Afin de régler les paramètres de la méthode, nous varions  $\nu$ ,  $\phi$  et  $\rho$  et pour chaque triplet de valeurs, le seuil optimal  $\gamma^*$  est obtenu par la méthode de validation croisée à 10 blocs, appliquée sur la base d'apprentissage. Dans un premier temps, l'effet des facteurs que contrôlent les paramètres  $\phi$  et  $\rho$  est atténué en fixant leur valeurs à  $\phi = +\infty$  et  $\rho = 0$  et seul  $\nu$  est varié. Le TABLEAU 4.1 montre la moyenne des métriques d'évaluation de cette configuration sur les 10 blocs de la base d'apprentissage et on remarque que la précision maximale de 70.65% est obtenue pour  $\nu = 100$ . On obtient, pour cette valeur, sur la base de test, une précision de 72.64%, une sensibilité de 50.51%, une spécificité de 91.30% et une portée de décision de 130 minutes.

TABLEAU 4.1 – Précision moyenne sur les 10 blocs sur la base d'apprentissage en fonction de  $\nu$  pour  $\phi = +\infty$  et  $\rho = 0$ .

$\nu$	$\gamma^*$	Pr
1	0.96	68.23%
20	0.95	68.73%
60	0.95	70.55%
100	0.88	<b>70.65%</b>
250	0.69	68.73%
500	0.58	75.10%

La valeur optimale de  $\nu$ ,  $\nu^* = 100$ , sera donc utilisée pour la suite. On fait ensuite varier  $\phi$  et  $\rho$  ensemble car leurs effets sont intriqués, dépendant de  $\mathbf{x}$ . Les résultats sur la base d'apprentissage par validation croisée sont affichés dans le TABLEAU 4.2. La précision maximale  $\text{Pr} = 82.29\%$  sur

l'apprentissage est atteinte pour  $\phi^* = 0.01$  et  $\rho^* = 0.5$ . Les performances pour ces valeurs optimales sur la base de test sont consignées dans le TABLEAU 4.3. Celles-ci sont meilleures que celles de la méthode précédente du chapitre 3, avec une précision de 80.19%. On remarque également que la valeur de la précision 85.14% n'est pas issue d'un déséquilibre entre la sensibilité et la spécificité ce qui confirme la démarche théorique de la méthode. En effet, ces deux valeurs sont très proches avec 83.5% pour la sensibilité et 86% pour la spécificité. La prédiction a néanmoins perdu 10 min de sa portée car on passe à 2 heures seulement.

TABLEAU 4.2 – Précision moyenne sur les 10 blocs sur la base d'apprentissage en fonction de  $\phi$  et  $\rho$  pour  $v = v^* = 100$ .

		$\phi$						
		0.00	0.01	0.05	0.10	0.50	1.00	2.00
$\rho$	0.00	54.15%	78.03%	76.83%	74.10%	69.54%	69.03%	69.33%
	0.01	54.15%	78.24%	77.23%	74.81%	70.04%	69.54%	69.23%
	0.05	54.15%	79.45%	77.64%	75.71%	70.05%	69.94%	69.54%
	0.10	54.15%	80.06%	78.35%	76.52%	70.86%	70.25%	70.05%
	0.50	54.15%	82.29%	81.58%	80.77%	75.11%	74.20%	71.36%
	1.00	54.15%	81.27%	81.48%	80.36%	74.19%	72.78%	70.95%
	2.00	54.15%	79.66%	79.86%	79.86%	74.09%	72.47%	72.06%

TABLEAU 4.3 – Performances sur la base de test pour les valeurs optimales des paramètres obtenues à partir de la base d'apprentissage.

$v$	$\rho$	$\phi$	$\gamma^*$	Pr	Se	Sp	Po
100	0.5	0.01	0.54	85.14%	83.50%	86.00%	120 min

#### 4.6.2 Méthode associative de combinaison des masses

Afin d'étudier l'impact de la méthode de combinaison proposée et la comparer à la méthode de combinaison de Yager, on reprend les performances optimales du chapitre précédent obtenues avec la méthode de combinaison de Yager avec optimisation des paramètres par la méthode de validation croisée à 10 blocs. On rappelle le paramétrage optimal obtenu :

$$N^* = 60, \alpha^* = 0.1, \beta^* = 1,$$

qui a donné les résultats suivants :

$$\text{Pr} = 80.19\%, \text{Se} = 75.26\%, \text{Sp} = 84.35\%, \text{ et Po} = 130.79 \text{ minutes.}$$

Pour ce même paramétrage, on applique la méthode de combinaison proposée au lieu de celle de Yager, à la suite de la modélisation de masse du chapitre 3, et on obtient pour la valeur  $\gamma^* = 0.59$  les performances suivantes :

$$\text{Pr} = 81.13\%, \text{Se} = 80.92\%, \text{Sp} = 81.30\%, \text{ et Po} = 121.67 \text{ minutes.}$$

On remarque alors qu'on obtient le même ordre de performance que la méthode de combinaison de Yager. L'apport demeure néanmoins dans le fond théorique de la méthode de combinaison qui a obtenu une immunité à la dépendance au classement des sources lors de la combinaison due au caractère associatif de l'opérateur de combinaison.

### 4.6.3 Option de rejet

Afin d'exprimer l'impact de l'option de rejet sur les performances de la prédiction des épisodes de la FA, on part, encore une fois, de la méthode du chapitre précédant avec le paramétrage élu optimal par la méthode de validation croisée à 10 blocs et on fait varier le paramètre d'optimisme  $\epsilon$  allant de 0 (pas de rejet) à 1 (maximum de rejet). Il est nécessaire pour cette comparaison d'introduire un indicateur de performance supplémentaire qui exprime le taux de patients pour lesquels la méthode a décidé de se prononcer. Cet indicateur est noté Tc pour Taux de classification et est défini comme suit :

$$Tc(I) = \frac{\sum_{i \in I} \llbracket \hat{y}_i > -1 \rrbracket}{|I|}.$$

Le TABLEAU 4.4 démontre que plus la valeur de  $\epsilon$  est élevée, meilleure est la précision et moins de patients sont classés. Ceci traduit le fait qu'au fur et à mesure que  $\epsilon$  augmente, le classifieur ne se prononce que sur les patients pour lesquels l'issue de la classification est de plus en plus certaine.

TABLEAU 4.4 – Impact de l'option pour plusieurs valeurs de  $\epsilon$  sur la base de test.

$\epsilon$	Pr	Tc	Se	Sp	Po
0.00	80.19%	100.00%	75.26%	84.35%	131 min
0.20	85.67%	82.31%	83.87%	87.11%	135 min
0.40	87.77%	65.57%	97.24%	77.44%	134 min
0.60	94.44%	50.94%	98.31%	89.80%	133 min
0.80	94.39%	46.23%	99.08%	88.51%	127 min
1.00	95.56%	42.45%	98.88%	92.31%	112 min

L'utilisateur peut alors favoriser l'une des deux métriques de performance au détriment de l'autre en ajustant la valeur de  $\epsilon$ . En effet, une performance est considérée meilleure lorsque la précision ou le taux de classification sont plus élevés. Il s'agit donc d'un problème d'optimisation bi-objectif et il est commun dans ce cas, de déterminer l'optimum de Pareto. Le principe d'optimum de Pareto réside dans la division de l'ensemble des solutions en deux parties. La première comporte des solutions qui sont dominées par une ou plusieurs autres, c-à-d qu'il existe au moins une autre solution qui suggère de meilleures performances selon les deux fonctions objectives à la fois. La deuxième comporte les solutions restantes non-dominées et ce sont ces solutions qui forment l'optimum de Pareto. En général, l'optimum de Pareto est utilisé lors d'une optimisation à plusieurs fonctions objectives qu'il est question de minimiser. On définit alors, pour notre application les deux fonctions objectives Pc et Tr représentant respectivement la perte de classification et le taux de rejet des patients (à minimiser) qui sont définis comme suit :

$$Pc(I) = 1 - Pr(I),$$

$$Tr(I) = 1 - Tc(I).$$

Afin de déterminer les valeurs de  $\epsilon$  qui conduisent à des performances non-dominées, on fait varier  $\epsilon$  de 0 à 1 avec un pas de 0.01 et pour chaque valeur, la méthode est appliquée sur la base d'apprentissage, et on place un point dans un repère dont les axes représentent les fonctions objectives à minimiser. Les valeurs de ces fonctions objectives sont issues de la moyenne des résultats obtenues sur chacun des 10 blocs de la validation croisée. Pour chaque valeur de  $\epsilon$ , on optimise la valeur de  $\gamma$  en sélectionnant celle qui maximise la précision car toute autre valeur de  $\gamma$  conduit au même taux de rejet (même Tr) mais à une perte de classification supérieure (plus grand Pc) ce qui correspond donc à une solution dominée. La Liste des figures 4.2 montre l'ensemble des solutions obtenues et les points en rouge représentent les solutions non-dominées. L'ensemble de ces valeurs forment ce qu'on appelle le front de Pareto qui est tracé sur la Liste des figures 4.3.

Il revient alors à l'utilisateur de choisir la valeur de  $\epsilon$  qui correspond à ses besoins. Il peut alors

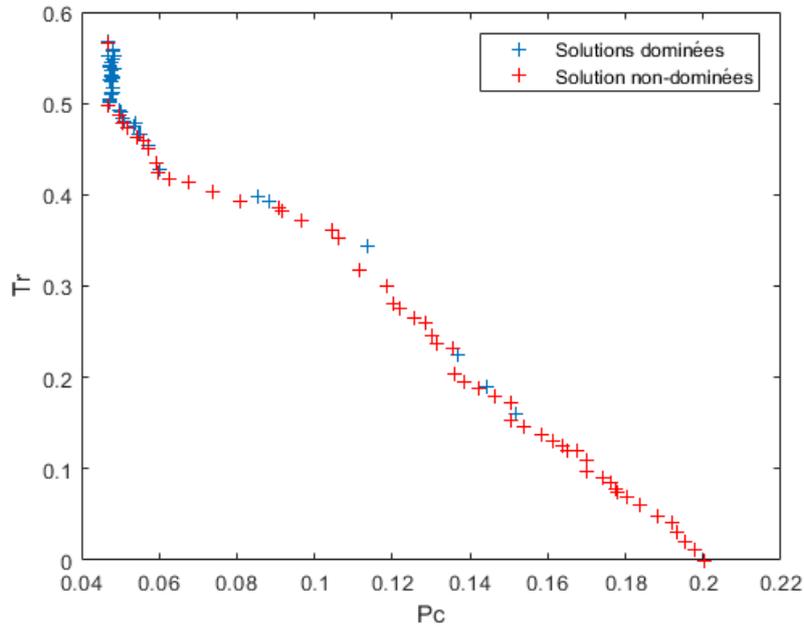


FIGURE 4.2 – Solutions obtenues sur la base d'apprentissage en faisant varier  $\epsilon$ .

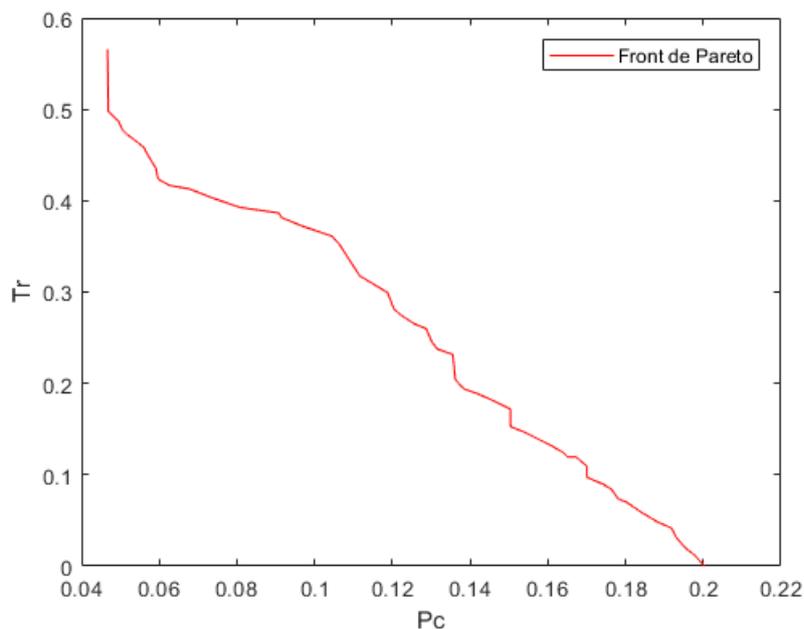


FIGURE 4.3 – Front de Pareto pour  $\epsilon$  variable de 0 à 1.

choisir de favoriser une fonction objective sur l'autre. Par ailleurs, il est possible aussi d'exprimer d'autres conditions indépendantes des fonctions objectives. En effet, un professionnel de santé peut, par exemple, choisir de sélectionner la valeur de  $\epsilon$  qui maximise la sensibilité et minimise le taux de rejet. Il s'agira alors d'un autre problème d'optimisation multi-objectif.

#### 4.6.4 Validation de l'approche globale

Afin de valider l'ensemble des contributions présentées dans ce chapitre, elles seront utilisées simultanément avec les valeurs optimales des paramètres obtenues dans chaque section. L'option de rejet ne procure pas de valeur optimale pour  $\epsilon$  car il existe deux fonctions objectives donc deux

optimums extrêmes. On évalue alors les performances pour plusieurs valeurs de  $\epsilon$  sur la base de test avec  $N = 60$ ,  $v = 100$ ,  $\phi = 0.01$ ,  $\rho = 0.5$  et  $\beta = 1$ . Les résultats sont affichés dans le TABLEAU 4.5 où on remarque que la précision peut augmenter jusqu'à 95.8% au détriment du taux de classification qui atteint 61.8%. A propos de la portée de classification, elle varie légèrement de 2 heures à 1 heure et 53 minutes.

TABLEAU 4.5 – Évaluation des contributions du chapitre sur la base de test

$\epsilon$	$\gamma^*$	Pr	Se	Sp	Tc	Po
0.00	0.54	85.14%	83.51%	86.52%	100.00%	120 min
0.02	0.53	86.87%	89.62%	84.51%	93.40%	120 min
0.04	0.51	86.55%	97.73%	74.70%	80.66%	124 min
0.06	0.52	90.03%	97.55%	82.28%	75.71%	119 min
0.08	0.53	92.95%	97.28%	88.74%	70.28%	116 min
0.10	0.53	95.80%	98.55%	92.74%	61.79%	113 min

On compare ensuite les performances obtenues à celles des approches de la littérature utilisant l'ECG, comme dans le chapitre précédent. On peut voir que pour  $\epsilon = 0.06$ , la précision dépasse celles des méthodes présentées, avec un taux de rejet d'environ 25%, comme le montre le TABLEAU 4.6.

TABLEAU 4.6 – Comparaison à des méthodes de prédiction basées sur l'ECG dans la littérature.

Méthode	Invasivité	Précision de la prédiction
Analyse haute fréquence de l'énergie et de la durée des ondes P	ECG	88.07%
Analyse des courtes séquences de variabilité de fréquence cardiaque	ECG	80.20%
Analyse de variabilité de fréquence cardiaque et algorithme génétique non-dominé de trie III	ECG	87.70%
Méthode proposée avec les contributions de ce chapitre.	Signaux physiologiques mesurables sans électrode	90.03%

## 4.7 Conclusion

Dans ce chapitre, plusieurs contributions ont été apportées au niveau de l'approche présentée dans le chapitre précédant. Il s'agit toujours de mesurer des signaux physiologiques d'un sujet et d'en extraire des caractéristiques dans le but de prédire un éventuel épisode de fibrillation atriale (FA). L'idée était de considérer chaque caractéristique comme une source d'information dans le cadre de la théorie des fonctions de croyance. Ainsi, chaque caractéristique est modélisée par une fonction de masse. Ces fonctions sont ensuite combinées puis transformées pour obtenir une probabilité de subir un épisode de FA. Cette probabilité est seuillée pour obtenir une décision. La première contribution réside dans la modélisation de la fonction de croyance liée à chaque caractéristique. En effet, afin de mieux modéliser l'incertitude associée à chaque caractéristique, plusieurs facteurs sont pris en compte comme la disponibilité, la discriminabilité et la récurrence des données. La deuxième contribution est au niveau de la combinaison. Une méthode de combinaison a été proposée ayant la particularité d'être associative contrairement à la méthode de combinaison de Yager précédemment utilisée. Finalement, une dernière contribution est l'ajout d'une option de rejet qui donne la possibilité au système de refuser de classer une instance quand

les données n'apportent pas assez de certitude sur l'issue de la classification. Ces méthodes ont été testées et comparées à celle du chapitre précédent grâce à différentes simulations pour en vérifier la qualité des contributions. Les résultats ont permis de valider l'approche, qui dépasse en performances celles utilisant des signaux invasifs, tels que l'ECG. Le chapitre suivant traitera le problème de détection de plusieurs maladies en même temps, grâce aux mêmes signaux physiologiques.



# Chapitre 5

## Classification multi-label évidentielle pour le suivi médical

### Sommaire

---

<b>5.1 Introduction</b>	<b>90</b>
<b>5.2 Définitions et formalisation mathématique</b>	<b>90</b>
5.2.1 Classification multi-label	90
5.2.2 Formalisation mathématique	90
5.2.3 Caractéristiques des données multi-label	92
5.2.4 Réglage du seuil de classification	93
<b>5.3 Métriques d'évaluation</b>	<b>93</b>
5.3.1 Les métriques basées sur les instances	93
5.3.2 Les métriques basées sur les labels	95
<b>5.4 Méthodes de classification multi-label</b>	<b>96</b>
5.4.1 Méthodes de transformation	97
5.4.2 Méthodes d'adaptation	102
<b>5.5 Classification multi-label de maladies cardio-vasculaires</b>	<b>105</b>
5.5.1 Construction des données multi-label	105
5.5.2 Fonctions de masse et règles d'association	108
<b>5.6 Résultats et discussion</b>	<b>110</b>
<b>5.7 Conclusion</b>	<b>113</b>

---

## 5.1 Introduction

Les chapitres 3 et 4 ont introduit deux approches de classification traitant d'une maladie unique, qui est la fibrillation atriale (FA). Ce chapitre s'intéresse au diagnostic médical de plusieurs maladies cardio-vasculaires (MCV), pouvant arriver au même moment. En effet, les mêmes signaux physiologiques en lien avec la FA sont capables de nous fournir des indications sur d'autres complications cardiaques et donc induire un diagnostic médical plus complet, portant sur plusieurs MCV. Notamment, ce chapitre porte sur le diagnostic en temps réel de 55 MCV, identifiées dans la base de données médicale MIMIC III. Ces maladies étant représentées par des descriptifs communs, on propose d'associer des étiquettes ou labels à chaque patient pour définir sa ou ses maladies. Le problème de classification est ainsi défini comme un problème de classification multi-label visant à déterminer les étiquettes correspondant à chaque patient à partir de ses signaux physiologiques. La suite de ce chapitre commence par introduire la classification multi-label et en établir une formalisation mathématique et un état de l'art. Ensuite, la modélisation du problème de diagnostic médical en classification multi-label est détaillée. Une approche multi-label basée sur la théorie des fonctions de croyance est enfin introduite et validée grâce aux données extraites de la base médicale MIMIC III.

## 5.2 Définitions et formalisation mathématique

### 5.2.1 Classification multi-label

La classification standard mono-label consiste à attribuer une seule classe ou label à une observation représentée le plus souvent par un vecteur de caractéristiques réelles. La classe ou label est un élément catégorique souvent issu d'un ensemble prédéfini de classes appelé aussi espace de classes. Pour ce type de classification, la base d'apprentissage est composée d'un ensemble d'instances étiquetées avec des étiquettes ou labels uniques. On parle de classification binaire si l'espace des classes est formé de deux classes et de classification multi-classe lorsque l'espace des classes est formé de plusieurs classes.

La classification multi-label fait référence à un type plus général de classification où une même instance peut être associée à plusieurs classes simultanément. La résolution du problème de classification revient alors à attribuer à une observation, un ensemble de classes ou labels en se basant sur ses caractéristiques. Un classifieur multi-label est ainsi une fonction qui associe à un vecteur de caractéristiques réelles un ensemble d'étiquettes ou labels inclus dans un espace de classes ou labels. Dans la FIGURE 5.1, on présente un exemple de données multi-label où plusieurs images sont présentées avec les labels correspondants. Dans le domaine de la classification d'images, il est commun qu'une image puisse être associée à plusieurs labels à la fois. D'autres exemples du monde réel impliquant des données de type multi-label sont la catégorisation du texte, où un article d'actualité peut couvrir plusieurs sujets simultanément comme le sport, les jeux olympiques et la natation, la classification d'un son pouvant contenir plusieurs étiquettes comme le piano, la musique classique et Mozart, l'annotation d'une vidéo qui peut être liée à plusieurs aspects comme un anniversaire, une fête et une célébration. De nombreux travaux de recherche existent sur la catégorisation de texte [106–113], la bio-informatique [114–122], la fouille du WEB (WEB mining) [123–125], la fouille de règles [125–133] et la recommandation des marques [134–137]. Dans ce qui suit, nous allons définir le cadre formel de la classification en précisant les notations mathématiques et les définitions de base de la classification multi-label.

### 5.2.2 Formalisation mathématique

On note  $\mathcal{X} = \mathbb{R}^d$  l'espace des caractéristiques et  $\mathbf{x} \in \mathcal{X}$  un vecteur de caractéristiques quelconque, avec  $d$  sa dimension. Les composantes du vecteur de caractéristiques  $\mathbf{x}$  sont notées  $x_k$ , avec  $1 \leq k \leq d$ . D'autre part, on note  $\mathcal{Y} = \{\ell_1, \ell_2, \dots, \ell_q\}$  l'espace des labels et  $\ell_j \in \mathcal{Y}$  un label ou une

	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Mer</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Ciel</div> </div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px; margin-top: 5px;">Soleil</div>
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Montagne</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Ciel</div> </div>
	<div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px; margin-left: 20px;">Neige</div>
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Montagne</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Mer</div> </div>
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Forêt</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Nuage</div> </div>
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Forêt</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Montagne</div> </div>
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Neige</div> <div style="border: 1px solid blue; padding: 2px 10px; background-color: #4a7ebb; color: white; border-radius: 5px;">Montagne</div> </div>

FIGURE 5.1 – Exemple de données multi-label.

étiquette, avec  $1 \leq j \leq q$  et  $q = |\mathcal{Y}|$  un entier strictement positif indiquant le nombre total de labels potentiels. Un ensemble quelconque de labels associé au vecteur  $\mathbf{x}$  sera noté généralement  $Y \subseteq \mathcal{Y}$  et  $\hat{Y}$  désignera une prédiction de cet ensemble par un classifieur, représentée par la fonction  $h(\cdot)$ . Ainsi,  $\hat{Y} = h(\mathbf{x})$  où  $h(\cdot)$  est une fonction de  $\mathcal{X}$  dans  $\mathcal{P}(\mathcal{Y})$ , avec  $\mathcal{P}(\mathcal{Y})$  l'ensemble des sous-ensembles de  $\mathcal{Y}$ .

$$h: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$$

$$\mathbf{x} \mapsto \hat{Y}.$$

Dans le contexte de la classification multi-label, il est très commun de travailler avec une fonction score qui attribue un nombre réel faisant office de niveau de confiance à chaque label de  $\mathcal{Y}$ . Cette fonction sera notée  $f$  et est définie comme suit :

$$f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$$

$$(\mathbf{x}, \ell_j) \mapsto f(\mathbf{x}, \ell_j).$$

On note respectivement  $\mathcal{A}$  et  $\mathcal{T}$  les bases d'apprentissage et de test. Chaque élément des bases de données multi-label est un couple formé d'un vecteur de caractéristiques  $\mathbf{x}_i$  et de l'ensemble

de labels qui lui correspond  $Y_i$ . Ainsi les bases d'apprentissage et de test sont représentées comme suit :

$$\mathcal{A} = \{(\mathbf{x}_i, Y_i) \mid i \in I_{\mathcal{A}}\},$$

$$\mathcal{T} = \{(\mathbf{x}_i, Y_i) \mid i \in I_{\mathcal{T}}\},$$

avec  $I_{\mathcal{A}}$  l'ensemble des indices réservés à l'apprentissage et  $I_{\mathcal{T}}$  l'ensemble des indices réservés au test. Le nombre d'éléments disponibles pour l'apprentissage est  $|I_{\mathcal{A}}| = m$  et pour le test  $|I_{\mathcal{T}}| = n$ . De plus, on définit la variable binaire  $\delta_j(Y) = \llbracket \ell_j \in Y \rrbracket$  qui indique si un certain label  $\ell_j$  appartient à l'ensemble  $Y$ , avec la fonction  $\llbracket \cdot \rrbracket$  indiquant la valeur 1 si la proposition  $\cdot$  est vraie et 0 sinon. On a alors :

$$\delta_j(Y) = \begin{cases} 1 & \text{si } \ell_j \in Y, \\ 0 & \text{si } \ell_j \notin Y. \end{cases}$$

Le TABLEAU 5.1 résume l'ensemble des notations présentées dans cette section.

TABLEAU 5.1 – Récapitulatif des notations.

Notation	Signification	Remarque
$\mathcal{X}$	Espace des caractéristiques	$\mathcal{X} = \mathbb{R}^d$
$\mathcal{Y}$	Espace des labels	$ \mathcal{Y}  = q$
$\ell_j$	Un label dans $\mathcal{Y}$	$1 \leq j \leq q$
$\mathbf{x}$	Vecteur de caractéristiques quelconque	$\mathbf{x} \in \mathcal{X}$
$\mathcal{P}(\cdot)$	Ensemble des sous-ensembles de $\cdot$	
$h(\cdot)$	Fonction de prédiction d'un classifieur	$h(\mathbf{x}) \subseteq \mathcal{Y}$
$f(\cdot, \cdot)$	Fonction score	$\exists t \in \mathbb{R} : h(\mathbf{x}) = \{y \in \mathcal{Y} \mid f(\mathbf{x}, y) \geq t\}$
$\mathcal{A}$	Base d'apprentissage	$ \mathcal{A}  = m$
$\mathcal{T}$	Base de test	$ \mathcal{T}  = n$
$I_{\mathcal{A}}$	Ensemble des indices des instances de $\mathcal{A}$	$(\mathbf{x}_i, Y_i) \in \mathcal{A} \Leftrightarrow i \in I_{\mathcal{A}}$
$I_{\mathcal{T}}$	Ensemble des indices des instances de $\mathcal{T}$	$(\mathbf{x}_i, Y_i) \in \mathcal{T} \Leftrightarrow i \in I_{\mathcal{T}}$
$\hat{Y}_i$	Prédiction pour l'observation $\mathbf{x}_i$	$h(\mathbf{x}_i) = \hat{Y}_i$
$x_{i,k}$	Composante $k$ du vecteur $\mathbf{x}_i$	$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$
$\llbracket \cdot \rrbracket$	Prend la valeur 1 si $\cdot$ est vrai et 0 sinon	
$\delta_j(\cdot)$	Prend la valeur 1 si $\ell_j \in \cdot$ et 0 sinon	

### 5.2.3 Caractéristiques des données multi-label

Pour caractériser les propriétés de tout ensemble de données multi-label, plusieurs indicateurs multi-label peuvent être utilisés. Le moyen le plus simple pour mesurer le degré du caractère multi-label d'une base de données multi-label  $\mathcal{D} = \{(\mathbf{x}_i, Y_i) \mid i \in I_{\mathcal{D}}\}$  est la cardinalité des labels qui représente le nombre de labels moyen pour chaque instance et qui est notée  $\text{LCard}(\cdot)$  :

$$\text{LCard}(\mathcal{D}) = \frac{1}{|I_{\mathcal{D}}|} \sum_{i \in I_{\mathcal{D}}} |Y_i|.$$

La densité des labels  $\text{LDens}(\cdot)$  normalise la cardinalité des labels en divisant par le nombre de labels total  $|\mathcal{Y}| = q$  :

$$\text{LDens}(\mathcal{D}) = \frac{1}{|I_{\mathcal{D}}|} \sum_{i \in I_{\mathcal{D}}} \frac{|Y_i|}{q}.$$

Un autre indicateur assez populaire pour la mesure du degré du caractère multi-label d'une base de données est la diversité des labels  $\text{LDiv}(\cdot)$  qui représente le nombre d'ensembles de labels distincts au sein de cette base de données :

$$\text{LDiv}(\mathcal{D}) = |\{Y \subseteq \mathcal{Y} \mid \exists \mathbf{x} \in \mathcal{X} : (\mathbf{x}, Y) \in \mathcal{D}\}|.$$

De même, la diversité peut être normalisée pour obtenir la diversité proportionnelle  $PLDiv(\cdot)$  :

$$PLDiv(\mathcal{D}) = \frac{LDiv(\mathcal{D})}{|\mathcal{D}|}.$$

### 5.2.4 Réglage du seuil de classification

Comme mentionné plus tôt, il est assez commun que des méthodes de classification multi-label produisent une fonction score  $f(\cdot, \cdot)$  avant de prédire un ensemble de labels. Dans ce cas, la prédiction est assurée en comparant la fonction score à un certain seuil  $\tau$  pour déterminer l'appartenance de chaque étiquette à l'ensemble de labels prédits. Ainsi, l'ensemble de labels prédit est  $h(\mathbf{x}) = \{y \in \mathcal{Y} \mid f(\mathbf{x}, y) \geq \tau\}$ . Il existe plusieurs stratégies pour le choix du seuil  $\tau$ . Certaines se basent sur une valeur constante et d'autres sur une valeur dépendante du vecteur  $\mathbf{x}$ . Pour le premier cas, cette constante peut par exemple être mise à 0.5 lorsque la fonction score trouve ses valeurs dans  $[0, 1]$  comme pour une fonction de probabilité. Lorsque toutes les instances de test sont disponibles, il est possible de fixer ce seuil constant en minimisant la différence d'un indicateur multi-label précis, notamment la cardinalité des étiquettes, entre la base d'apprentissage et la base de test :

$$\tau^* = \underset{\tau \in \mathbb{R}}{\operatorname{argmin}} \left| \text{LCard}(\mathcal{L}) - \text{LCard}(\{\mathbf{x}_i, \hat{Y}_i \mid i \in \mathcal{I}_{\mathcal{T}}\}) \right|.$$

Pour le second cas, il est judicieux de compter sur la base d'apprentissage pour choisir le seuil. Une approche assez commune pour faire ce choix est le modèle linéaire où le seuil est supposé être une fonction linéaire du vecteur  $\mathbf{f}(\mathbf{x}) = (f(\mathbf{x}, \ell_1), f(\mathbf{x}, \ell_2) \dots f(\mathbf{x}, \ell_q))$  :

$$\tau(\mathbf{x}) = \langle \mathbf{f}(\mathbf{x}), \mathbf{w}^T \rangle + b,$$

où  $\mathbf{w}$  est un vecteur de coefficients réels de dimension  $q$  et  $b$  une constante. Ces paramètres sont régis par un problème d'optimisation qui vise à minimiser la différence entre le seuil  $\tau(\mathbf{x}_i)$  et le seuil optimal  $s(\mathbf{x}_i)$  pour l'instance  $(\mathbf{x}_i, Y_i)$  sur la base d'apprentissage ( $i \in \mathcal{I}_{\mathcal{A}}$ ) :

$$\min_{\{\mathbf{w}, b\}} \sum_{i \in \mathcal{I}_{\mathcal{A}}} |\langle \mathbf{f}(\mathbf{x}_i), \mathbf{w}^T \rangle + b - s(\mathbf{x}_i)|$$

avec  $s(\mathbf{x}_i) = \arg \min_{t \in \mathbb{R}} |Y_i \Delta h(\mathbf{x}_i)|$  et  $\Delta$  l'opérateur de la différence symétrique sur l'ensemble  $\mathcal{Y}$ .

## 5.3 Métriques d'évaluation

Dans l'apprentissage supervisé traditionnel, les performances générales du système d'apprentissage sont évaluées avec des mesures conventionnelles telles que la précision, la mesure F, l'aire sous la courbe ROC (AUC), etc. Cependant, l'évaluation des performances dans l'apprentissage multi-label est beaucoup plus compliquée que celle du mono-label, car chaque exemple peut être associé à plusieurs labels simultanément. Par conséquent, un certain nombre de métriques d'évaluation spécifiques à l'apprentissage multi-label est proposé. Ces métriques peuvent être classées en deux groupes, à savoir les métriques basées sur des instances [138–140], et les métriques basées sur les labels [141].

Les mesures basées sur des instances fonctionnent en évaluant les performances du système d'apprentissage sur chaque instance de test séparément, puis en renvoyant la valeur moyenne sur l'ensemble de test. D'autre part, les mesures basées sur des labels fonctionnent en évaluant les performances du système d'apprentissage sur chaque label séparément, puis en renvoyant la valeur moyenne sur tous les labels de l'espace des labels  $\mathcal{Y}$ .

### 5.3.1 Les métriques basées sur les instances

Plusieurs métriques basées sur les instances peuvent être définies :

### Subset Accuracy

Le "Subset Accuracy" évalue la fraction d'exemples correctement classés, c'est-à-dire lorsque l'ensemble des labels prédit est identique celui de la base de test  $\mathcal{S}$ .

$$\text{SubAcc}(h) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \mathbb{1}[\hat{Y}_i = Y_i].$$

Intuitivement, la précision du sous-ensemble peut être considérée comme un équivalent multi-label de la métrique de précision traditionnelle, et a tendance à être trop stricte, en particulier lorsque la taille de l'espace des labels  $\mathcal{Y}$  est grande.

### Hamming Loss

Le "Hamming Loss" (perte de Hamming) évalue la fraction des labels d'instances mal classées, c'est-à-dire qu'un label non-pertinent est prédit ou qu'un label pertinent ne l'est pas.

$$\text{HLoss}(h) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{S}}} |\hat{Y}_i \Delta Y_i|,$$

avec  $\Delta$  représentant la différence symétrique entre deux ensembles.

### Precision

Le "Precision" (Précision) est le taux de labels pertinents parmi les prédits.

$$\text{Precision}(h) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \frac{|Y_i \cap \hat{Y}_i|}{|\hat{Y}_i|}.$$

### Recall

Le "Recall" (Rappel) est le taux de labels prédits parmi les pertinents.

$$\text{Recall}(h) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \frac{|Y_i \cap \hat{Y}_i|}{|Y_i|}.$$

### $F^\beta$

La métrique "F" (F-mesure) est la moyenne harmonique de la précision et du rappel. Le " $F^\beta$ " est la moyenne harmonique pondérée par  $\beta$  de ces deux métriques ce qui est équivalent à calculer la moyenne harmonique simple de  $\beta \cdot \text{Precision}(h)$  et de  $\frac{1}{\beta} \cdot \text{Recall}(h)$ .

$$F^\beta(h) = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}.$$

### Accuracy

Le "Accuracy" donne le rapport entre les labels pertinents prédits et les labels pertinents ou prédits. Il existe une application bijective entre cette métrique d'évaluation et celle du "F" qui est la moyenne harmonique de "Precision" et de "Recall" :  $\text{Accuracy}(h) = \frac{F(h)}{2 - F(h)}$ .

$$\text{Accuracy}(h) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \frac{|Y_i \cap \hat{Y}_i|}{|Y_i \cup \hat{Y}_i|}.$$

### One-Error

Le "One-Error" évalue la fraction d'exemples dont le label le mieux classé n'est pas dans l'ensemble des labels pertinents :

$$\text{One-Error}(f) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{D}}} \left[ \arg \max_{y \in \mathcal{Y}} f(\mathbf{x}, y) \notin Y_i \right].$$

### Coverage

Le "Coverage" (couverture) évalue le nombre de pas nécessaires, en moyenne, pour descendre dans la liste des labels classés de manière à couvrir tous les labels pertinents de l'exemple.

$$\text{Cov}(f) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{D}}} \left( \max_{y \in Y_i} \sum_{y' \in \mathcal{Y}} |\{y' \in \mathcal{Y} \mid f(\mathbf{x}, y') \geq f(\mathbf{x}, y)\}| \right) + 1.$$

### Ranking Loss

Le "Ranking Loss" (perte de classement) évalue la fraction de paires de labels inversés, c'est-à-dire qu'un label non pertinent est classé plus haut qu'un autre pertinent.

$$\text{Rloss}(f) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{D}}} \frac{1}{|Y_i| |\bar{Y}_i|} \left| \{(y, y') \in Y_i \times \bar{Y}_i \mid f(\mathbf{x}_i, y) \leq f(\mathbf{x}_i, y')\} \right|,$$

avec  $\bar{Y}_i$  l'ensemble complémentaire de  $Y_i$  dans  $\mathcal{Y}$ .

### Average precision

Le "Average precision" (précision moyenne) évalue la fraction moyenne de labels pertinents classés plus haut qu'un label particulier  $y \in Y_i$ .

$$\text{AvgPrec}(f) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{D}}} \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\{y' \in Y_i \setminus \{y\} \mid f(\mathbf{x}_i, y) \leq f(\mathbf{x}_i, y')\}|}{\sum_{y' \in \mathcal{Y}} \mathbb{1}\{f(\mathbf{x}_i, y) \leq f(\mathbf{x}_i, y')\}}$$

Pour le "One-Error", le "Coverage" et le "Ranking Loss", plus la valeur de la métrique est petite, meilleures sont les performances du système, avec une valeur optimale de  $\frac{1}{n} \sum_{i \in \mathcal{I}_{\mathcal{D}}} |Y_i| - 1$  pour le "Coverage" et de 0 pour le "One-Error" et le "Ranking Loss". Pour les autres métriques multi-label basées sur des exemples, plus la valeur de la métrique est élevée, meilleure est la performance du système, avec une valeur optimale de 1.

### 5.3.2 Les métriques basées sur les labels

Pour un certain label  $\ell_j \in \mathcal{Y}$ , on peut définir 4 quantités caractérisant la performance de classification binaire de ce label par  $h(\cdot)$  :

$$\text{VP}_j = \sum_{i \in \mathcal{I}_{\mathcal{D}}} \mathbb{1}\{\ell_j \in Y_i \cap \hat{Y}_i\},$$

$$\text{FP}_j = \sum_{i \in \mathcal{I}_{\mathcal{D}}} \mathbb{1}\{\ell_j \in \bar{Y}_i \cap \hat{Y}_i\},$$

$$\text{VN}_j = \sum_{i \in \mathcal{I}_{\mathcal{D}}} \mathbb{1}\{\ell_j \in \bar{Y}_i \cap \bar{\hat{Y}}_i\},$$

$$\text{FN}_j = \sum_{i \in \mathcal{I}_{\mathcal{D}}} \mathbb{1}\{\ell_j \in Y_i \cap \bar{\hat{Y}}_i\}.$$

$VP_j$ ,  $FP_j$ ,  $VN_j$  et  $FN_j$  représentent respectivement les vrais positifs, faux positifs, vrais négatifs et faux négatifs du label  $\ell_j$ . On a donc  $VP_j + FP_j + VN_j + FN_j = n$ . En se basant sur ces 4 quantités, la plupart des métriques de classification binaire peuvent être dérivées. Soit  $M(VP_j, FP_j, VN_j, FN_j)$  une métrique de classification binaire dépendante de  $VP_j$ ,  $FP_j$ ,  $VN_j$  et  $FN_j$ . Une métrique de classification multi-label peut être obtenue à partir de  $M$  en utilisant l'un de ces deux modes :

— Macro-moyenne :

$$M_{\text{macro}}(h) = \frac{1}{q} \sum_{j=1}^q M(VP_j, FP_j, VN_j, FN_j),$$

— Micro-moyenne :

$$M_{\text{micro}}(h) = M\left(\sum_{j=1}^q VP_j, \sum_{j=1}^q FP_j, \sum_{j=1}^q VN_j, \sum_{j=1}^q FN_j\right).$$

Voici quelques métriques de classification binaire que peut représenter  $M$  :

— Accuracy

$$\text{Accuracy}(VP_j, FP_j, VN_j, FN_j) = \frac{VP_j + VN_j}{n},$$

— Precision :

$$\text{Precision}(VP_j, FP_j, VN_j, FN_j) = \frac{VP_j}{VP_j + FP_j},$$

— Recall :

$$\text{Recall}(VP_j, FP_j, VN_j, FN_j) = \frac{VP_j}{VP_j + FN_j},$$

— Specificity :

$$\text{Specificity}(VP_j, FP_j, VN_j, FN_j) = \frac{VN_j}{VN_j + FP_j},$$

—  $F^\beta$  :

$$F^\beta(VP_j, FP_j, VN_j, FN_j) = \frac{(1 + \beta^2) VP_j}{(1 + \beta^2) VP_j + \beta^2 FN_j + FP_j}.$$

## 5.4 Méthodes de classification multi-label

Le développement d'algorithmes est toujours au cœur des recherches sur l'apprentissage automatique dont la classification multi-label est un cas particulier. Au cours de la dernière décennie, une quantité importante d'algorithmes a été proposée pour apprendre à partir de données multi-label. Dans un souci de cohérence notationnelle et de rigueur mathématique, chaque algorithme est présenté sous des notations communes introduite dans le TABLEAU 5.1. Dans ce chapitre, une simple catégorisation des algorithmes d'apprentissage multi-label est adoptée.

### Méthodes de transformation

Cette catégorie d'algorithmes aborde le problème d'apprentissage multi-label en le transformant en d'autres problème d'apprentissage bien établis. Les algorithmes représentatifs comprennent le "Binary Relevance" [142] et les chaînes de classifieur [143] qui transforment la tâche d'apprentissage multi-label en plusieurs tâches de classification binaire, le "Calibrated Label Ranking" [144] qui transforme la tâche d'apprentissage multi-label en tâche de classement des labels, la méthode Label Powerset (LP) [141] qui transforme la tâche d'apprentissage multi-label en tâches de classification multi-classes et la méthode RAKEL (Random  $k$ -labelset) qui est une méthode d'ensemble, c-à-d qu'elle se base sur des sous-ensembles de l'espace des labels pour générer plusieurs classifieurs avant de mettre en commun leur contributions et d'obtenir un classifieur unique plus performant.

## Méthodes d'adaptation

Cette catégorie d'algorithmes s'attaque au problème d'apprentissage multi-label en adaptant les techniques d'apprentissage populaires pour traiter directement les données multi-label. Les algorithmes représentatifs incluent le "multi-label Decision Tree" (ML-DT) [145] adaptant les techniques d'arbre de décision, le "Rank-SVM" [146] adaptant les séparateurs à vaste marge et le "Collective multi-label Classifier" (CML) [138] adaptant les techniques de la théorie de l'information.

### 5.4.1 Méthodes de transformation

#### Binary Relevance (BR)

Le concept de cette méthode est de transformer un problème de classification multi-label en  $q$  problèmes de classification binaire en considérant chaque étiquette  $\ell_j$  indépendamment,  $1 \leq j \leq q$ . Pour cela, la méthode construit en premier lieu une base d'apprentissage pour chaque étiquette  $\ell_j$  que l'on notera  $\mathcal{A}_j$  telle que

$$\mathcal{A}_j = \{(\mathbf{x}_i, \delta_j(Y_i)) \mid i \in I_{\mathcal{A}}\}.$$

Ensuite, un algorithme de classification binaire est utilisé pour induire un classifieur  $g_j(\cdot)$  pour chaque étiquette  $\ell_j$  :

$$g_j: \mathcal{X} \rightarrow \{0, 1\}$$

$$\mathbf{x} \mapsto g_j(\mathbf{x}).$$

Ainsi, pour une instance représentée par le vecteur  $\mathbf{x}$ , l'ensemble de labels prédit  $h(\mathbf{x})$  est

$$h(\mathbf{x}) = \{\ell_j \in \mathcal{Y} \mid g_j(\mathbf{x}) = 1, 1 \leq j \leq q\}.$$

La FIGURE 5.2 illustre le procédé de cette méthode.

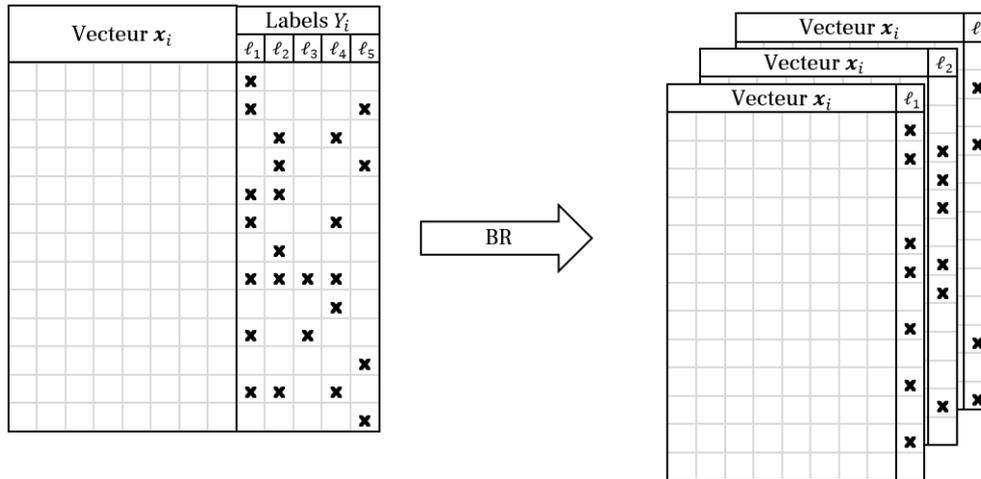


FIGURE 5.2 – Illustration de la méthode "Binary Relevance".

Pour éviter de sortir une décision vide ( $h(\mathbf{x}) = \emptyset$ ), on peut avoir recours à la fonction score  $f_j(\cdot)$  des classifieurs binaires correspondant à chaque label  $\ell_j$ . On rappelle que la décision est construite en seuillant la fonction score à une valeur  $\tau$  :

$$h(\mathbf{x}) = \{\ell_j \mid f_j(\mathbf{x}) \geq \tau\}.$$

Une règle simple pour éviter une décision vide est la règle du "T-criterion" qui consiste à prédire l'étiquette qui affiche le score maximal parmi les autres étiquettes même si cette valeur de score

est inférieure au seuil choisi :

$$h(\mathbf{x}) = \{\ell_j \mid f_j(\mathbf{x}) \geq \tau\} \cup \left\{ \ell_{j^*} \mid j^* = \underset{1 \leq j \leq q}{\operatorname{arg\,max}} f_j(\mathbf{x}) \right\}.$$

La méthode BR ignore complètement les corrélations potentielles entre les labels, et le classifieur binaire pour chaque label  $g_j(\cdot)$  peut souffrir du problème de déséquilibre des classes surtout lorsque  $q$  est grand et que la densité des labels  $\text{LDens}(\mathcal{D})$  est faible. Comme le montre l'ALGORITHME 1, la méthode BR a une complexité de calcul de  $\mathcal{O}(q\mathcal{F}_g(n, d))$  pour l'apprentissage et  $\mathcal{O}(q\mathcal{F}'_g(d))$  pour le test. Ici,  $\mathcal{F}_g(n, d)$  et  $\mathcal{F}'_g(d)$  représentent les complexités de calcul du classifieur binaire pour l'apprentissage et pour le test respectivement.

---

**Algorithm 1:** Méthode BR
 

---

```

Y ← ∅
Pour j ← 1 à q
    Entraîner  $g_j(\cdot)$  par  $\mathcal{A}_j$ 
    Si  $g_j(\mathbf{x}) = 1$ 
        | Y ← Y ∪ { $\ell_j$ }
    FinSi
FinPour
Retourner Y
    
```

---

### Chaînes de classifieurs

L'idée de base de cet algorithme est de transformer le problème de classification multi-label en une chaîne de problèmes de classification binaire où les classifieurs binaires sont construits en se basant sur les précédents dans une séquence bien définie [143; 147]. Soit  $\pi$  une bijection de  $\{1, 2, \dots, q\}$  dans  $\{1, 2, \dots, q\}$  qui sera utilisée pour définir une liste ordonnée de labels tel que  $\ell_{\pi(1)} < \ell_{\pi(2)} < \dots < \ell_{\pi(q)}$ . On définit les bases d'apprentissage  $\mathcal{A}_j^\pi$ ,  $1 \leq j \leq q$ , pour chaque maillon de la chaîne comme suit :

$$\mathcal{A}_j^\pi = \left\{ \left( \mathbf{x}_{i,j}^\pi, \llbracket \ell_{\pi(j)} \in Y_i \rrbracket \right) \mid i \in I_{\mathcal{A}} \right\},$$

avec  $\mathbf{x}_{i,j}^\pi = [\mathbf{x}_i, \llbracket \ell_{\pi(1)} \in Y_i \rrbracket, \llbracket \ell_{\pi(2)} \in Y_i \rrbracket, \dots, \llbracket \ell_{\pi(j-1)} \in Y_i \rrbracket]$  concaténant le vecteur  $\mathbf{x}_i$  et les éléments binaires qui le suivent dans  $[\cdot]$ . Par la suite,  $q$  classifieurs binaires  $g_j$  sont induits par apprentissage sur les bases  $\mathcal{A}_j^\pi$  tel que

$$g_j : \mathcal{X} \times \{0, 1\}^{j-1} \rightarrow \{0, 1\}$$

$$\mathbf{x} \mapsto g_j(\mathbf{x}).$$

Soit  $\mathbf{x}$  un vecteur de caractéristiques représentant une instance multi-label à classifier. La classification est assurée par applications successives des classifieurs  $g_j$  pour  $j$  allant de 1 à  $q$  :

$$\lambda_j = g_j([\mathbf{x}, \lambda_1, \lambda_2, \dots, \lambda_{j-1}]).$$

On obtient finalement le classifieur multi-label :

$$h(\mathbf{x}) = \{\ell_j \mid \lambda_j = 1, 1 \leq j \leq q\}.$$

La FIGURE 5.3 illustre le procédé de cette méthode.

Par rapport à la méthode BR, la méthode des chaînes de classifieurs a l'avantage d'exploiter les corrélations des étiquettes mais elle perd l'avantage de l'implémentation parallèle en raison de sa propriété de chaînage. Pendant la phase d'apprentissage, les chaînes de classifieurs augmentent la dimension des instances avec des caractéristiques supplémentaires issues des étiquettes qui leur correspondent. Au lieu de garder des caractéristiques supplémentaires à valeur binaire, une autre possibilité est de les affecter aux sorties probabilistes du classifieur binaire lorsque son modèle

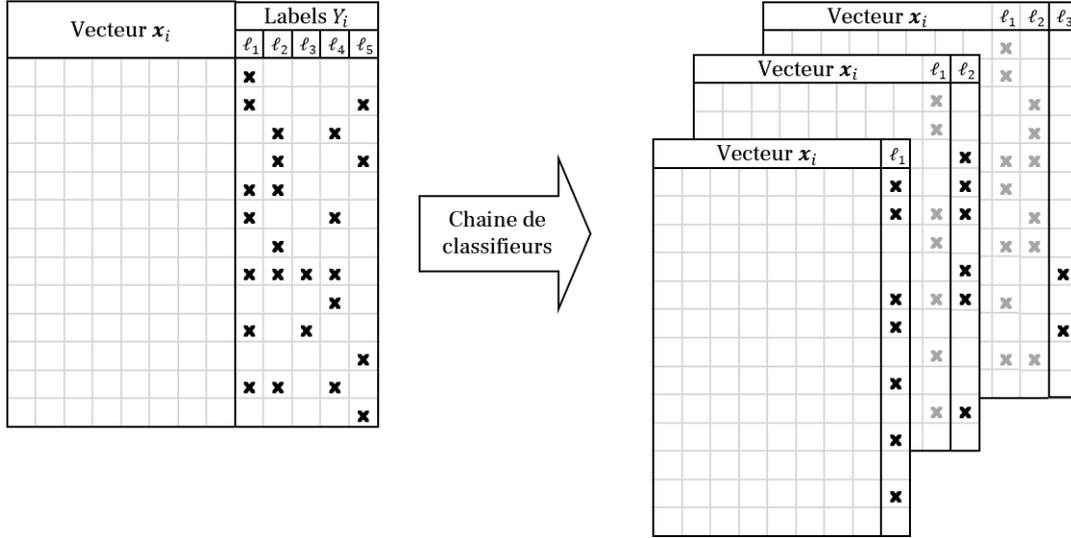


FIGURE 5.3 – Illustration de la méthode des chaînes de classifieurs.

est capable de produire une probabilité a posteriori [148]. Comme le montre l'algorithme 2, les chaînes de classifieurs ont une complexité de calcul dans le pire des cas, de  $\mathcal{O}(q\mathcal{F}_g(m, d + q))$  pour l'apprentissage et de  $\mathcal{O}(q\mathcal{F}'_g(d + q))$  pour le test.

---

**Algorithm 2:** Méthode des chaînes de classifieurs
 

---

```

Y ← ∅
Pour j ← 1 à q
    Entraîner gj(·) par  $\mathcal{A}_j^\pi$ 
    Si  $\lambda_j = 1$ 
        | Y ← Y ∪ {lj}
    FinSi
FinPour
Retourner Y
    
```

---

**Calibrated Label Ranking (CLR)**

Cette méthode vise à transformer le problème de classification multi-label en un problème de classement des étiquettes en comptant sur la comparaison des labels par paires. Pour  $q$  labels, on construit un total de  $\frac{q(q-1)}{2}$  classifieurs chacun basé sur une paire précise de labels. D'abord, une base d'apprentissage est construite pour chaque paire  $(\ell_j, \ell_{j'})$ ,  $1 \leq j < j' \leq q$  :

$$\mathcal{A}_{j,j'} = \{(\mathbf{x}_i, \delta_j(Y_i)) \mid i \in I_{\mathcal{A}}, \delta_j(Y_i) \neq \delta_{j'}(Y_i)\}.$$

Seules les instances ayant uniquement un label de la paire considérée sont retenues pour la base d'apprentissage propre à cette paire. Ensuite, un classifieur  $g_{j,j'}(\cdot)$  est induit pour chaque paire de labels  $(\ell_j, \ell_{j'})$  par un algorithme de classification binaire. En plus, le classifieur  $g_j(\cdot)$  de la méthode "Binary Relevance" est aussi induit pour chaque label dans le but de comparer les votes d'un label à un certain seuil. Pour une instance de vecteur de caractéristiques  $\mathbf{x}$  à classer, on obtient l'accumulation des votes de chaque label durant sa confrontation avec toutes les autres en utilisant les classifieurs  $g_{j,j'}(\cdot)$  en plus de la contribution du classifieur individuel  $g_j(\cdot)$  :

$$f(\mathbf{x}, \ell_j) = g_j(\mathbf{x}) + \sum_{j'=1}^{j-1} 1 - g_{j',j}(\mathbf{x}) + \sum_{j'=j+1}^q g_{j,j'}(\mathbf{x}).$$

Ainsi le classifieur multi-label est obtenu en comparant, pour chaque label  $\ell_j$ ,  $f(\mathbf{x}, \ell_j)$  au nombre de labels prédits impertinents par la méthode BR :

$$h(\mathbf{x}) = \left\{ \ell_j \in \mathcal{Y} \mid f(\mathbf{x}, \ell_j) + \sum_{j=1}^q g_j(\mathbf{x}) > q \right\}$$

Comparé aux algorithmes précédemment introduits [142; 143], qui construisent des classifieurs binaires d'une manière un contre tous, la méthode CLR construit des classifieurs binaires d'une manière un contre un et a donc l'avantage d'atténuer l'influence négative de la question du déséquilibre des classes. D'autre part, le nombre de classifieurs binaires construits par la méthode CLR passe d'une échelle linéaire à une échelle quadratique en fonction du nombre d'étiquettes. Les améliorations de la méthode CLR se focalisent principalement sur la réduction du nombre quadratique de classifieurs à interroger en phase de test par élagage exact [149] ou élagage approximatif [150; 151]. En exploitant le caractère de l'algorithme d'apprentissage binaire sous-jacent  $g$ , comme la double représentation pour Perceptron [152], le nombre quadratique de classifieurs peut être induit plus efficacement en phase d'apprentissage [153]. Comme le montre l'algorithme 3, la méthode CLR a une complexité de calcul de  $O(q^2 \mathcal{F}_g(m, d))$  pour l'apprentissage et  $\mathcal{O}(q^2 \mathcal{F}'_g(d))$  pour les tests.

---

**Algorithm 3:** Méthode CLR
 

---

```

scores ← Tableau( $q$ )
Pour  $j \leftarrow 1$  à  $q - 1$ 
    | scores( $j$ ) ← 0
FinPour
Pour  $j \leftarrow 1$  à  $q - 1$ 
    | Pour  $j' \leftarrow j$  à  $q$ 
        | | Entraîner  $g_{j,j'}(\cdot)$  par  $\mathcal{A}_{j,j'}$ 
        | | Si  $g_{j,j'}(\mathbf{x}) = 1$ 
        | | | scores( $j$ ) ← scores( $j$ ) + 1
        | | FinSi
        | | Si  $g_{j,j'}(\mathbf{x}) = 0$ 
        | | | scores( $j'$ ) ← scores( $j'$ ) + 1
        | | FinSi
    | FinPour
FinPour
seuil ←  $q$ 
Pour  $j \leftarrow 1$  à  $q$ 
    | Entraîner  $g_j(\cdot)$  par  $\mathcal{A}_j$ 
    | Si  $g_j(\mathbf{x}) = 1$ 
    | | scores( $j$ ) ← scores( $j$ ) + 1
    | | seuil ← seuil - 1
    | FinSi
FinPour
 $Y \leftarrow \emptyset$ 
Pour  $j \leftarrow 1$  à  $q$ 
    | Si scores( $j$ ) > seuil
    | |  $Y \leftarrow Y \cup \{\ell_j\}$ 
    | FinSi
FinPour
Retourner  $Y$ 
    
```

---

**Label Powerset (LP)**

Cet algorithme a pour objectif de transformer le problème de classification multi-label en un problème de classification multi-classes. Pour cela, chaque combinaison différente de labels est

affiliée à une classe d'un ensemble de classes  $C = \{c_1, c_2 \dots c_{2^q}\}$  qui lui est propre par le relais d'une fonction bijective  $\Gamma$  :

$$\begin{aligned} \Gamma: \mathcal{P}(\mathcal{Y}) &\rightarrow C \\ Y &\mapsto \Gamma(Y) \end{aligned}$$

tel que  $Y \neq Y' \Leftrightarrow \Gamma(Y) \neq \Gamma(Y')$  et  $\forall Y \subseteq \mathcal{Y}, \exists c \in C \mid \Gamma(Y) = c$ . Une seule base d'apprentissage est construite :

$$\mathcal{A}' = \{(\mathbf{x}_i, \Gamma(Y_i)) \mid i \in I_{\mathcal{A}}\}.$$

Un classifieur  $g(\cdot)$  multi-classe est induit sur cette base d'apprentissage et la prédiction du classifieur multi-label  $h(\cdot)$  est obtenue en appliquant la fonction inverse de  $\Gamma$ ,  $\Gamma^{-1}$  sur la sortie du classifieur multi-classe  $g(\cdot)$ . Ainsi, pour un vecteur de caractéristique  $\mathbf{x}$  :

$$h(\mathbf{x}) = \Gamma^{-1}(g(\mathbf{x})).$$

Un moyen d'améliorer la méthode LP consiste à élaguer les étiquettes dans  $\mathcal{A}$  ayant une fréquence inférieure à un seuil de comptage prédéfini [154]. Comme le montre l'algorithme 4, la méthode LP a une complexité de calcul de  $\mathcal{O}(\mathcal{F}_g(m, d, 2^q))$  pour l'apprentissage et  $\mathcal{O}(\mathcal{F}'_g(d, 2^q))$  pour les tests.

---

**Algorithm 4:** Méthode LP
 

---

```

dictionnaire  $\leftarrow \emptyset$ 
id  $\leftarrow 0$ 
Pour  $i \leftarrow 1$  à  $m$ 
    Si  $Y_i \notin$  dictionnaire
        id  $\leftarrow$  id + 1
        dictionnaire  $\leftarrow$  dictionnaire  $\cup \{(Y_i, id)\}$ 
    FinSi
FinPour
 $\mathcal{A}' \leftarrow \emptyset$ 
Pour  $i \leftarrow 1$  à  $n$ 
    Pour  $j \leftarrow 1$  à id
        Si dictionnaire(j)(1) =  $Y_i$ 
            idTrouvé  $\leftarrow$  dictionnaire(j)(2)
        FinSi
    FinPour
     $\mathcal{A}' \leftarrow \mathcal{A}' \cup \{(\mathbf{x}_i, idTrouvé)\}$ 
FinPour
Entraîner  $g(\cdot)$  par  $\mathcal{A}'$ 
 $Y \leftarrow$  dictionnaire( $g(\mathbf{x})$ )(1)
Retourner  $Y$ 
    
```

---

**Random  $k$ -labelset (RAkEL)**

Cette méthode fait partie d'une catégorie de méthodes appelées les méthodes d'ensemble. Elle prend en considération plusieurs ensembles de labels sur lesquels elle applique un classifieur LP. Ces ensembles sont au nombre de  $N$  et chacun est composé d'une combinaison de  $k$  labels d'où le nom " $k$ -labelset". On commence donc par choisir au hasard  $N$  ensembles de labels  $\mathcal{Y}_p \subseteq \mathcal{Y}$ ,  $1 \leq p \leq N$  différents de taille  $k$ . On construit ensuite les bases d'apprentissage :

$$\mathcal{A}'_p = \{(\mathbf{x}_i, \Gamma_p(Y_i \cap \mathcal{Y}_p)) \mid i \in I_{\mathcal{A}}\},$$

avec  $\Gamma_p$  une fonction bijective :

$$\begin{aligned} \Gamma_p: \mathcal{P}(\mathcal{Y}_p) &\rightarrow C \\ Y &\mapsto \Gamma_p(Y) \end{aligned}$$

et  $|C| = 2^k$ . Pour un vecteur de caractéristique  $\mathbf{x}$ , un ensemble de  $N$  classifieurs LP est appliqué que l'on notera dans cette section  $g_p(\cdot)$ ,  $1 \leq p \leq N$  :

$$g_p: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y}_p) \\ \mathbf{x} \mapsto g_p(\mathbf{x}).$$

Pour finaliser la classification, chaque label est traité séparément en comparant les votes obtenus par les  $N$  classifieurs avec le nombre de ceux d'entre eux qui le prennent en considération. Le classifieur multi-label  $h(\cdot)$  de la méthode RAKEL est donc construit sur la base des  $N$  classifieurs multi-label  $g_p(\cdot)$  :

$$h(\mathbf{x}) = \left\{ \ell_j \in \mathcal{Y} \mid \frac{\sum_{p=1}^N \mathbb{1}[\ell_j \in g_p(\mathbf{x})]}{\sum_{p=1}^N \mathbb{1}[\ell_j \in \mathcal{Y}_p]} \geq \tau \right\},$$

avec  $\tau$  le seuil à optimiser (Voir section 5.2.4). Pour un ensemble créé par  $N$   $k$ -labelsets, le nombre maximum de votes pour chaque étiquette est de  $\frac{Nk}{q}$  en moyenne. Une règle empirique pour la méthode RAKEL est de prendre  $k = 3$  et  $N = 2q$  [141; 155]. Bien que la méthode RAKEL intègre le concept d'ensemble comme son caractère inhérent pour palier aux inconvénients majeurs de la méthode LP, l'apprentissage d'ensemble pourrait être utilisé comme une stratégie de méta-niveau pour faciliter l'apprentissage multi-label en englobant les composantes multi-label homogènes [143; 156], ou hétérogènes [157; 158]. Comme le montre l'algorithme 5, la méthode RAKEL a une complexité de calcul de  $\mathcal{O}(N\mathcal{F}_g(m, d, 2^k))$  pour l'apprentissage et  $\mathcal{O}(N\mathcal{F}'_g(d, 2^k))$  pour le test.

## 5.4.2 Méthodes d'adaptation

### Multi-label Decision Tree (ML-DT)

L'idée de base de cette famille de méthodes est d'adopter des techniques d'arbre de décision pour traiter des données multi-label, où un critère de gain d'information basé sur l'entropie multi-label est utilisé pour construire l'arbre de décision de manière récursive. Le gain d'information obtenue par la partition d'une base  $\mathcal{D}$  le long de la caractéristique  $k$  à la valeur de séparation  $\omega$  est :

$$\text{GI}(\mathcal{D}, k, \omega) = E(\mathcal{D}) - \frac{|\mathcal{D}^+|}{|\mathcal{D}|} E(\mathcal{D}^+) - \frac{|\mathcal{D}^-|}{|\mathcal{D}|} E(\mathcal{D}^-),$$

avec  $E(\cdot)$  la quantité d'information présente dans  $\cdot$ ,

$$\mathcal{D}^+ = \{(\mathbf{x}_i, Y_i) \in \mathcal{D} \mid x_{i,k} > \omega\},$$

$$\mathcal{D}^- = \{(\mathbf{x}_i, Y_i) \in \mathcal{D} \mid x_{i,k} \leq \omega\}.$$

On définit l'entropie d'une base  $\mathcal{D}$  :

$$E(\mathcal{D}) = \sum_{j=1}^q -p_j(\mathcal{D}) \log(p_j(\mathcal{D})) - (1 - p_j(\mathcal{D})) \log(1 - p_j(\mathcal{D})),$$

avec  $p_j(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} \mathbb{1}[\ell_j \in Y_i]$ . Un arbre est alors construit en débutant par  $\mathcal{D} = \mathcal{A}$  puis en séparant  $\mathcal{D}$  suivant la dimension et à la valeur de séparation qui induisent un gain d'information maximal :

$$(k^*, \omega^*) = \max_{k, \omega} \text{GI}(\mathcal{D}, k, \omega).$$

Ensuite la même séparation est effectuée sur les bases de données  $\mathcal{D}^+$  et  $\mathcal{D}^-$  itérativement jusqu'à ce que l'une des bases obtenue après séparation rencontre un certain critère d'arrêt. Ensuite, pour un vecteur de caractéristiques  $\mathbf{x}$  à classifier, l'arbre est balayée pour atteindre la feuille qui correspond aux valeurs des coordonnées de  $\mathbf{x}$ , puis pour chaque label, la prédiction est décidée suivant

**Algorithm 5:** Méthode RAKEL

---

```

nombreVotes ← Tableau(q)
votes ← Tableau(q)
Pour  $p \leftarrow 1$  à  $N$ 
    Pour  $j \leftarrow 1$  à  $q$ 
        Si  $\ell_j \in L_p$ 
            nombreVotes( $j$ ) ← nombreVotes( $j$ ) + 1
        FinSi
    FinPour
    dictionnaire ←  $\emptyset$ 
    id ← 0
    Pour  $i \leftarrow 1$  à  $m$ 
        Si  $Y_i \notin \text{dictionnaire}$ 
            id ← id + 1
            dictionnaire ← dictionnaire  $\cup \{(Y_i \cap L_p, id)\}$ 
        FinSi
    FinPour
     $\mathcal{A}'_p \leftarrow \emptyset$ 
    Pour  $i \leftarrow 1$  à  $m$ 
        Pour  $j \leftarrow 1$  à id
            Si dictionnaire( $j$ )(1) =  $Y_i \cap L_p$ 
                idTrouvé ← dictionnaire( $j$ )(2)
            FinSi
        FinPour
         $\mathcal{A}'_p \leftarrow \mathcal{A}'_p \cup \{(\mathbf{x}_i, \text{idTrouvé})\}$ 
    FinPour
    Entraîner  $g_p(\cdot)$  par  $\mathcal{A}'_p$ 
     $Y_p \leftarrow \text{dictionnaire}(g(\mathbf{x}))(1)$ 
    Pour  $j \leftarrow 1$  à  $q$ 
        Si  $\ell_j \in Y_p$ 
            votes( $j$ ) ← votes( $j$ ) + 1
        FinSi
    FinPour
FinPour
 $Y \leftarrow \emptyset$ 
Pour  $j \leftarrow 1$  à  $q$ 
    Si votes( $j$ ) / nombreVotes( $j$ )  $\geq \tau$ 
         $Y \leftarrow Y \cup \{\ell_j\}$ 
    FinSi
FinPour
Retourner  $Y$ 

```

---

la taux de présence du label dans la feuille :

$$h(\mathbf{x}) = \left\{ \ell_j \in \mathcal{Y} \mid \sum_{i \in \mathcal{D}} \delta_j(Y_i) \geq \frac{|\mathcal{D}|}{2} \right\},$$

avec  $\mathcal{D}$  la feuille de l'arbre dont les intervalles de chaque label correspondent aux valeurs des coordonnées de  $\mathbf{x}$ . Un avantage important de la méthode ML-DT réside dans sa grande efficacité à induire le modèle d'arbre de décision à partir de données multi-label. Les améliorations possibles sur cette méthode comprennent notamment l'utilisation d'une stratégie d'élagage [145] ou de méthodes d'apprentissage d'ensemble [159; 160]. La méthode ML-DT a une complexité de calcul de  $\mathcal{O}(mdq)$  pour l'apprentissage et  $\mathcal{O}(mq)$  pour le test.

### Rank-SVM

L'idée de base de cette méthode est d'adapter la stratégie de la marge maximale pour traiter les données multi-label, où un ensemble de classificateurs linéaires est recherché en minimisant une fonction de perte empirique. La généralisation de la méthode pour générer des classificateurs non-linéaires est réalisée en utilisant l'astuce du noyau. On obtient la marge maximale d'un classificateur linéaire  $\mathcal{W} = \{(\mathbf{w}_j, b_j) \mid 1 \leq j \leq q\}$  avec  $\mathbf{w}_j$  le vecteur de poids de dimension  $d$  et  $b_j$  le biais du label  $j$ , sur une instance  $(\mathbf{x}_i, Y_i)$  de la base d'apprentissage ( $i \in I_{\mathcal{A}}$ ) en traitant le problème d'optimisation suivant :

$$\min_{(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i} \frac{\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'}}{\|\mathbf{w}_j - \mathbf{w}_{j'}\|},$$

avec  $\langle \cdot, \cdot \rangle$  le produit scalaire et  $\|\cdot\|$  la norme du vecteur  $\cdot$ . Géométriquement parlant, pour chaque paire de labels  $(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i$ , leur frontière de discrimination correspond à l'hyperplan  $\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x} \rangle + b_j - b_{j'} = 0$ . Par conséquent, l'équation ci-dessus considère la distance euclidienne entre  $\mathbf{x}_i$  et les hyperplans de chaque paire d'étiquettes dont l'une est pertinente et l'autre non-pertinente, puis renvoie le minimum comme marge sur  $(\mathbf{x}_i, Y_i)$ . Par conséquent, la marge du système d'apprentissage sur la base d'apprentissage  $\mathcal{A}$  suit naturellement :

$$\min_{(\mathbf{x}_i, Y_i) \in \mathcal{A}} \min_{(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i} \frac{\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'}}{\|\mathbf{w}_j - \mathbf{w}_{j'}\|}.$$

En ajustant les valeurs paramétriques du classificateur proportionnellement, on peut assurer pour les éléments de la marge (instance d'apprentissage et pair de label) que  $\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_{i^*} \rangle + b_j - b_{j'} = 1$  et on a alors automatiquement  $\forall i \in I_{\mathcal{A}}, (\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i, \langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'} \geq 1$ . Par la suite, le problème de la maximisation de la marge peut être exprimé comme suit :

$$\max_{\mathcal{W}} \min_{(\mathbf{x}_i, Y_i) \in \mathcal{A}} \min_{(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i} \frac{1}{\|\mathbf{w}_j - \mathbf{w}_{j'}\|^2}.$$

Supposons que nous ayons suffisamment d'exemples d'apprentissage de sorte que pour chaque paire d'étiquettes  $(\ell_j, \ell_{j'})$  ( $j \neq j'$ ), il existe  $(\mathbf{x}_i, Y_i) \in \mathcal{A}$  satisfaisant  $(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i$ . Ainsi, l'objectif du problème de maximisation ci-dessus devient équivalent à

$$\max_{\mathcal{W}} \min_{1 \leq j < j' \leq q} \frac{1}{\|\mathbf{w}_j - \mathbf{w}_{j'}\|^2},$$

et le problème d'optimisation peut être réécrit comme :

$$\min_{\mathcal{W}} \max_{1 \leq j < j' \leq q} \|\mathbf{w}_j - \mathbf{w}_{j'}\|^2,$$

sous condition que :

$$\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'} \geq 1 \quad \forall i \in I_{\mathcal{A}}, (\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i.$$

Pour surmonter la difficulté apportée par l'opérateur max, la méthode "Rank-SVM" choisit de simplifier l'équation ci-dessus en approximant l'opérateur max avec l'opérateur somme :

$$\min_{\mathcal{W}} \sum_{j=1}^q \|\mathbf{w}_j\|^2$$

sous condition que

$$\langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'} \geq 1 \quad \forall i \in I_{\mathcal{A}}, (\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i.$$

Pour s'adapter aux scénarios du monde réel où les contraintes de l'équation ci-dessus ne peuvent pas être entièrement satisfaites, des variables d'écart peuvent être incorporées dans celle-ci :

$$\min_{\mathcal{W}, \Omega} \sum_{j=1}^q \|\mathbf{w}_j\|^2 + \alpha \sum_{i \in I_{\mathcal{A}}} \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i} \zeta_{i,j,j'}$$

sous condition que :

$$\begin{aligned} \langle \mathbf{w}_j - \mathbf{w}_{j'}, \mathbf{x}_i \rangle + b_j - b_{j'} &\geq 1 - \zeta_{i,j,j'} \\ \zeta_{i,j,j'} &\geq 0 \\ \forall i \in I_{\mathcal{A}}, (\ell_j, \ell_{j'}) &\in Y_i \times \bar{Y}_i, \end{aligned}$$

avec  $\Omega = \{\zeta_{i,j,j'}, i \in I_{\mathcal{A}}, (\ell_j, \ell_{j'}) \in Y_i \times \bar{Y}_i\}$ . La méthode Rank-SVM bénéficie des noyaux pour gérer les problèmes de classification non linéaires, et d'autres variantes peuvent être obtenues. Premièrement, la stratégie de seuillage peut être réalisée avec des techniques autres que la procédure de style empilement [161]. Deuxièmement, pour éviter le problème de la sélection du noyau, plusieurs techniques d'apprentissage de noyau peuvent être employées pour apprendre à partir de données multi-label [162; 163].

## 5.5 Classification multi-label de maladies cardio-vasculaires

### 5.5.1 Construction des données multi-label

La base de données médicale MIMIC III contient des admissions à l'hôpital et les pathologies diagnostiquées relatives à chacune de ces admissions. Ces dernières sont identifiées grâce à la classification internationale des maladies, 9<sup>ème</sup> révision (ICD9) qui attribue à chaque pathologie un code numérique spécifique. Les maladies cardio-vasculaires sont incluses dans les maladies du système circulatoire dont les codes ICD9 varient de 39000 à 45900. Puisque le but est de concevoir un système capable de prédire plusieurs maladies cardio-vasculaires simultanément, on focalise notre étude sur un ensemble de 55 maladies choisies dans l'intervalle des codes ICD9 des maladies du système circulatoire. Une approche systématique pour labéliser chaque patient est de choisir les maladies qui lui sont diagnostiquées comme labels. Cependant, cette méthode conduit à une base de données ayant une très faible cardinalité de labels car le nombre de maladies diagnostiquées à chaque patient n'est généralement pas très élevé ce qui minimise son caractère multi-label. De plus, l'avantage principal de la classification multi-label est de pouvoir se servir des corrélations entre les labels. Il est donc judicieux de concevoir des labels portant une charge sémantique plus significative. Nous proposons ainsi l'élaboration d'un ensemble différent de descriptifs propre à chacune des 55 maladies sélectionnées pour notre étude, qui indexent essentiellement leur phase, leur sévérité, leur localisation dans le corps humain et d'autres indicateurs plus spécifiques. La cardinalité de l'intersection de ces ensembles issus de maladies différentes peut alors témoigner du degré de similitudes médicales entre ces maladies. Ce processus a conduit à la mise

en place d'un total de 26 labels. Pour chacune des 55 maladies, on affecte alors un sous-ensemble des 26 labels selon les descriptifs de chacune d'elles. Les 26 labels sont détaillés et numérotés dans le TABLEAU 5.2. Le tableau de la FIGURE 5.4 liste les 55 maladies cardio-vasculaires concernées par notre étude sur la première colonne au moyen de leur code ICD9 et les 26 labels sont sur les autres colonnes. Une valeur de 1 dans le tableau signifie que le label de la colonne est attribué à la maladie de la ligne et une valeur de 0 signifie le contraire. Sur ce tableau, nous avons une moyenne de 4.16 labels par maladie.

TABLEAU 5.2 – Numérotation des labels.

	<b>Description</b>
1	Hypertension
2	Cardiopathie
3	Maladie rénale chronique
4	Hypertension secondaire
5	Virulent
6	Étape 1 à 4 / au-dessus 4
7	Insuffisance cardiaque
8	Rénovasculaire
9	Aigu
10	Infarctus du myocarde
11	Antérieur
12	Postérieur
13	Latéral
14	Inférieur
15	Épisode initial / suivant
16	Dysrythmies cardiaques
17	Paroxystique
18	Ventriculaire / supraventriculaire
19	Fibrillation
20	Battements prématurés
21	Congestif
22	Systolique
23	Diastolique
24	Chronique
25	Bénin
26	Atrial

Une fois les maladies et leurs labels définis, la seconde étape du processus consiste à sélectionner dans la base de données MIMIC III tous les patients pour lesquels au moins l'une des 55 maladies sélectionnées a été diagnostiquée. On labélise ensuite chaque patient par l'union des ensembles des labels issus de chaque maladie qui lui a été diagnostiquée. Cela conduit à 46520 patients labélisés dont l'objectif sera de considérer les signaux physiologiques et d'en extraire des caractéristiques qui représenteront le signal. Le problème de classification des maladies diagnostiquées est alors reformulé comme un problème de classification multi-label. Comme mentionné plus tôt, l'avantage de l'approche multi-label est de tirer profit de la corrélation entre les maladies en orientant la prédiction vers les traits des maladies plutôt que les maladies elles-mêmes.

Dans le chapitre 3, la méthode présentée consiste à choisir des signaux physiologiques pour chaque patient disponible dans la base de données et à lui associer une valeur binaire indiquant si le patient est atteint de fibrillation atriale ou non. Dans le cadre de la classification multi-label, chaque patient lui est diagnostiqué plusieurs pathologies dont on retiendra ceux faisant partie du groupe des 55 pathologies présélectionnées. Chaque pathologie étant liée à un sous-ensemble de labels, on associe à chaque patient l'union des labels apportés par toutes les pathologies qui lui

# ICD9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
40100	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
40110	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40200	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40201	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40210	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40211	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40300	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40301	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40310	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40311	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40400	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40401	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40402	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40403	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40410	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40411	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40412	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40413	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40501	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40509	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40511	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
40519	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
41001	0	1	0	0	0	0	0	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
41002	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41011	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
41012	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41021	0	1	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
41022	0	1	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
41031	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
41032	0	1	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
41041	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
41042	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
41051	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
41052	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41200	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42700	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
42710	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
42731	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
42732	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
42741	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
42742	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
42761	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
42800	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
42820	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
42821	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
42822	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
42823	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
42830	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
42831	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
42832	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
42833	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
42840	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
42841	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
42842	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
42843	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0

FIGURE 5.4 – Tableau regroupant les labels accumulés par chaque maladie.

ont été diagnostiquées et faisant l’objet de notre étude.

Afin de construire les bases de données d’apprentissage et de test, les signaux physiologiques des 46520 patients sont considérés et des caractéristiques en sont extraites, similairement au chapitre 3, en faisant un balayage sur toute la longueur des séjours par fenêtre glissante de longueur  $N$  sans chevauchement entre les fenêtres. Ensuite, les fenêtres patients de durée  $N$  sont représentées par leurs vecteurs de caractéristiques  $\mathbf{x}_i$  regroupant les caractéristiques extraites de tous les types de signaux physiologiques disponibles. Chacun de ces vecteurs est associé à un ensemble de labels  $Y_i \subseteq \mathcal{Y}$  selon les pathologies qui ont été diagnostiquées. La base de données globale se

présente alors sous la forme suivante :

$$\mathcal{H} = \{(\mathbf{x}_i, Y_i) \mid 1 \leq i \leq N_{\mathcal{H}}\},$$

avec  $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^d$ ,  $Y_i \subseteq \mathcal{Y} = \{\ell_1, \ell_2, \dots, \ell_{26}\}$  et  $N_{\mathcal{H}}$  le nombre total de fenêtres obtenues et donc d'instance. Pour les besoins de l'apprentissage du classifieur et de sa validation, la base de données  $\mathcal{H}$  sera divisée en deux parties, l'une pour l'apprentissage  $\mathcal{A}$  et l'autre pour le test  $\mathcal{T}$  :

$$\mathcal{A} = \{(\mathbf{x}_i, Y_i) \in \mathcal{H} \mid i \in I_{\mathcal{A}}\},$$

$$\mathcal{T} = \{(\mathbf{x}_i, Y_i) \in \mathcal{H} \mid i \in I_{\mathcal{T}}\},$$

avec  $I_{\mathcal{A}} \cup I_{\mathcal{T}} = \{1, 2, \dots, N_{\mathcal{H}}\}$  et  $\frac{|I_{\mathcal{A}}|}{N_{\mathcal{H}}} = \frac{2}{3}$ .

## 5.5.2 Fonctions de masse et règles d'association

### Attribution des fonctions de masses selon les caractéristiques

Pour réaliser la tâche de classification multi-label, une classification binaire (BR) est utilisée pour traiter chaque label indépendamment des autres. Cela conduit à  $q = 26$  classifieurs binaires, chacun estimant la pertinence d'un label  $\ell_j$ ,  $1 \leq j \leq q$ , à partir de l'ensemble  $\mathcal{Y}$ . Un classifieur binaire prend le vecteur de caractéristiques d'un patient  $i$  et donne des informations sur la pertinence de l'étiquette  $\ell_j$ . Pour ce faire, chaque caractéristique est traitée comme une source d'information indépendante dans le domaine de la théorie des fonctions de croyance, et l'information de chaque source est modélisée par une fonction de masse. Soit  $\mathbf{x} = \{x_1, x_2, \dots, x_d\}$  le vecteur de caractéristiques d'une instance à prédire. Pour chaque caractéristique  $u \in 1, \dots, d$  et pour chaque label  $j$ , deux distributions normales sont ajustées pour les groupes d'instances pertinents et non pertinents sur les données disponibles dans la base de données d'apprentissage  $\mathcal{A}$  en maximisant la fonction de vraisemblance. Les moyennes et les écarts-types des distributions du groupe pertinent  $\hat{\mu}_{j,u}^{(1)}$  et  $\hat{\sigma}_{j,u}^{(1)}$ , et du groupe non pertinent  $\hat{\mu}_{j,u}^{(0)}$  et  $\hat{\sigma}_{j,u}^{(0)}$  sont donnés comme suit,

$$\begin{aligned} \hat{\mu}_{j,u}^{(1)} &= \frac{\sum_{i \in I_{\mathcal{A}}} x_{i,u} \llbracket \ell_j \in Y_i \rrbracket}{\sum_{i \in I_{\mathcal{A}}} \llbracket \ell_j \in Y_i \rrbracket}, \\ \hat{\mu}_{j,u}^{(0)} &= \frac{\sum_{i \in I_{\mathcal{A}}} x_{i,u} \llbracket \ell_j \notin Y_i \rrbracket}{\sum_{i \in I_{\mathcal{A}}} \llbracket \ell_j \notin Y_i \rrbracket}, \\ \hat{\sigma}_{j,u}^{(1)} &= \sqrt{\frac{\sum_{i \in I_{\mathcal{A}}} (x_{i,u} - \hat{\mu}_{j,u}^{(1)})^2 \llbracket \ell_j \in Y_i \rrbracket}{\sum_{i \in I_{\mathcal{A}}} \llbracket \ell_j \in Y_i \rrbracket}}, \\ \hat{\sigma}_{j,u}^{(0)} &= \sqrt{\frac{\sum_{i \in I_{\mathcal{A}}} (x_{i,u} - \hat{\mu}_{j,u}^{(0)})^2 \llbracket \ell_j \notin Y_i \rrbracket}{\sum_{i \in I_{\mathcal{A}}} \llbracket \ell_j \notin Y_i \rrbracket}}. \end{aligned}$$

Soient  $E_j$  et  $\bar{E}_j$  les événements indiquant si le label  $\ell_j$  est pertinent ou non respectivement, et soit  $\Omega = \{E_j, \bar{E}_j\}$  l'ensemble des deux événements. L'objectif est d'attribuer des masses aux ensembles d'événements  $\{E_j\}$ ,  $\{\bar{E}_j\}$  et  $\Omega$  de manière à quantifier l'information. D'abord, la masse attribuée à l'ensemble  $\Omega$  notée  $m_{j,u}^{\mathbf{x}}(\Omega)$  est calculée grâce aux équations (3.1) ou (4.1) selon les représentations proposées aux Chapitres 3 ou 4 respectivement. Puisque la fonction de densité de probabilité est plus élevée là où les données sont les plus susceptibles d'être, les masses attribuées à  $\{E_j\}$  et  $\{\bar{E}_j\}$  sont proportionnellement distribuées aux valeurs des fonctions de densité de probabilité par

rapport à la valeur de cette caractéristique pour l'instance  $\mathbf{x}$ . On a alors :

$$m_{j,u}^{\mathbf{x}}(\{E_j\}) = \frac{\Phi(x_u, \hat{\mu}_{j,u}^{(1)}, \hat{\sigma}_{j,u}^{(1)})}{\sum_{s \in \{0,1\}} \Phi(x_u, \hat{\mu}_{j,u}^{(s)}, \hat{\sigma}_{j,u}^{(s)})} (1 - m_{j,u}^{\mathbf{x}}(\Omega)),$$

$$m_{j,u}^{\mathbf{x}}(\{\bar{E}_j\}) = \frac{\Phi(x_u, \hat{\mu}_{j,u}^{(0)}, \hat{\sigma}_{j,u}^{(0)})}{\sum_{s \in \{0,1\}} \Phi(x_u, \hat{\mu}_{j,u}^{(s)}, \hat{\sigma}_{j,u}^{(s)})} (1 - m_{j,u}^{\mathbf{x}}(\Omega)),$$

avec  $\phi(\cdot, \mu, \sigma)$  la fonction de densité de probabilité d'une distribution normale d'espérance  $\mu$  et d'écart-type  $\sigma$  et appliquée à  $\cdot$ . Après avoir calculé toutes les masses pour toutes les entités, le processus de combinaison a lieu en combinant séquentiellement les masses à l'aide de la méthode de combinaison de YAGER [57]. La masse finale issue de la combinaison de toutes les sources est notée  $m_j^{\mathbf{x}}(\cdot)$ .

### Utilisation des règles d'association

Les règles d'association [164–171] sont des méthodes qui visent à trouver des relations intéressantes entre les éléments d'une base de données de transactions. Dans notre cas, nous cherchons à trouver des labels qui peuvent conduire à la pertinence des autres. La méthode proposée utilise les règles extraites comme une nouvelle source d'information et les combine avec la fonction de masse  $m_j^{\mathbf{x}}(\cdot)$  finale obtenue à partir des caractéristiques. Ainsi, la classification prendra en compte la dépendance entre les labels pour améliorer les résultats. Les règles d'association sont généralement définies par une implication d'un ensemble d'items à un autre ( $A \Rightarrow B$ ). Dans notre cas, on s'intéresse aux relations entre des labels individuels ( $E_j \Rightarrow E_{j'}$ ). La première étape de l'extraction de règles d'association est de trouver des ensembles d'éléments fréquents. Cela se fait en calculant le support du côté gauche de la relation. En général, le support est défini pour tout ensemble d'éléments  $Y \subseteq \mathcal{Y}$  sur la base d'apprentissage régi par l'ensemble d'indices  $I_{\mathcal{A}}$  comme étant :

$$\text{Supp}(\mathcal{A}, Y) = \frac{|\{(\mathbf{x}_i, Y_i) \in \mathcal{A} \mid Y \subseteq Y_i\}|}{|\mathcal{A}|}$$

où  $|\cdot|$  désigne le cardinal de l'ensemble  $\cdot$ . Seules les relations avec un support supérieur à une valeur seuil prédéfinie  $\text{Supp}_{\min}$  sont conservées. Dans notre approche, nous nous intéressons aux relations où l'occurrence d'un ou plusieurs labels implique l'occurrence d'un autre label  $\ell_j$ . Soit  $Y_{\Rightarrow E_j}$  l'ensemble des labels dont le support avec  $\ell_j$  est supérieur à  $\text{Supp}_{\min}$  :

$$Y_{\Rightarrow E_j} = \{y \in \mathcal{Y} \mid \text{Supp}(\mathcal{A}, \{\ell_j, y\}) \geq \text{Supp}_{\min}\}.$$

Pour évaluer le poids d'une association d'un label  $\ell_{j'}$  à un label  $\ell_j$ , nous utilisons la confiance de cette règle qui peut être calculée comme suit,

$$\text{Conf}(\mathcal{A}, E_{j'} \Rightarrow E_j) = \frac{\text{Supp}(\mathcal{A}, \{\ell_j, \ell_{j'}\})}{\text{Supp}(\mathcal{A}, \{\ell_{j'}\})}.$$

La valeur de la confiance rend la relation plus utile lorsqu'elle s'approche des bornes de son intervalle (0 et 1). Lorsque la valeur est proche de 1, la règle  $E_{j'} \Rightarrow E_j$  est considérée comme une source d'information, et si elle est proche de 0, la règle  $E_{j'} \Rightarrow \bar{E}_j$  est utilisée à la place. La règle est considérée comme plus informative si le côté gauche  $E_{j'}$  est plus susceptible d'être pertinent. La valeur de la masse finale pour l'étiquette  $\ell_{j'}$  est utilisée pour confirmer cette validité. Cependant, si la valeur est proche de 0.5, la règle devient inutile. Afin d'exprimer cette idée par une source d'information dans le domaine de la théorie des fonctions de croyance, nous proposons une fonction de masse  $m_{(\ell_{j'} \Rightarrow \ell_j)}^{\mathbf{x}}(\cdot)$  à partir de chaque label  $\ell_{j'}$  de  $Y_{\Rightarrow E_j}$  propageant la masse obtenue de  $\ell_{j'}$  en utilisant sa

confiance comme suit :

$$\begin{aligned} m'_{(\ell_{j'} \Rightarrow \ell_j)}(\{E_j\}) &= m_{j'}^{\mathbf{x}}(\{E_{j'}\}) (\text{Conf}(\mathcal{A}, E_{j'} \Rightarrow E_j)), \\ m'_{(\ell_{j'} \Rightarrow \ell_j)}(\{\bar{E}_j\}) &= m_{j'}^{\mathbf{x}}(\{E_{j'}\}) (1 - \text{Conf}(\mathcal{A}, E_{j'} \Rightarrow E_j)), \\ m'_{(\ell_{j'} \Rightarrow \ell_j)}(\Omega) &= 1 - m_{j'}^{\mathbf{x}}(\{E_{j'}\}). \end{aligned}$$

Pour les labels n'ayant pas de relations directes avec  $\ell_j$ ,  $\ell_{j'} \notin Y_{\Rightarrow E_j}$ , on choisit la fonction de masse telle que la totalité de la masse est attribuée à l'ignorance. Cette fonction de masse est un élément neutre de la combinaison de YAGER et ne perturbera pas donc la combinaison. On a donc pour  $\ell_{j'} \notin Y_{\Rightarrow E_j}$  :

$$\begin{aligned} m'_{(\ell_{j'} \Rightarrow \ell_j)}(\{E_j\}) &= m'_{(\ell_{j'} \Rightarrow \ell_j)}(\{E_{j'}\}) = 0, \\ m'_{(\ell_{j'}, \ell_j)}(\Omega) &= 1. \end{aligned}$$

Ensuite, nous combinons toutes les distributions de masses construites grâce à la règle d'association pour tout label  $\ell_j$  en utilisant la méthode de combinaison de YAGER d'opérateur  $\oplus$ ,

$$m_j^{\mathbf{x}}(\cdot) = \bigoplus_{j' \in Y_{\Rightarrow E_j}} m'_{(\ell_{j'} \Rightarrow \ell_j)}(\cdot) = \bigoplus_{j' \in \{1, \dots, q\}, j' \neq j} m'_{(\ell_{j'} \Rightarrow \ell_j)}(\cdot),$$

avec  $m_j^{\mathbf{x}}(\cdot)$  la masse finale issue de la combinaison de toutes les sources. On obtient deux distributions de masses pour chaque label  $\ell_j$ , la première  $m_j^{\mathbf{x}}(\cdot)$  provient de l'étude des caractéristiques et la seconde  $m_j^{\mathbf{x}}(\cdot)$  de la corrélation des labels. Ces deux sources d'information sont fusionnées avec la même méthode de combinaison de YAGER :

$$\mathbf{m}_j^{\mathbf{x}}(\cdot) = m_j^{\mathbf{x}}(\cdot) \oplus m_j^{\mathbf{x}}(\cdot).$$

La distribution de masse finale est transformée en probabilité en utilisant la méthode de transformation du Chapitre 4 :

$$\begin{aligned} \mathbb{P}^{\mathbf{x}}(E_j) &= \mathbf{m}_j^{\mathbf{x}}(\{E_j\}) + \mathbf{m}_j^{\mathbf{x}}(\Omega) \frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} \llbracket \ell_j \in Y_i \rrbracket}{|\mathcal{A}|}, \\ \mathbb{P}^{\mathbf{x}}(\bar{E}_j) &= \mathbf{m}_j^{\mathbf{x}}(\{0\}) + \mathbf{m}_j^{\mathbf{x}}(\Omega) \left( 1 - \frac{\sum_{i \in \mathcal{I}_{\mathcal{A}}} \llbracket \ell_j \in Y_i \rrbracket}{|\mathcal{A}|} \right). \end{aligned}$$

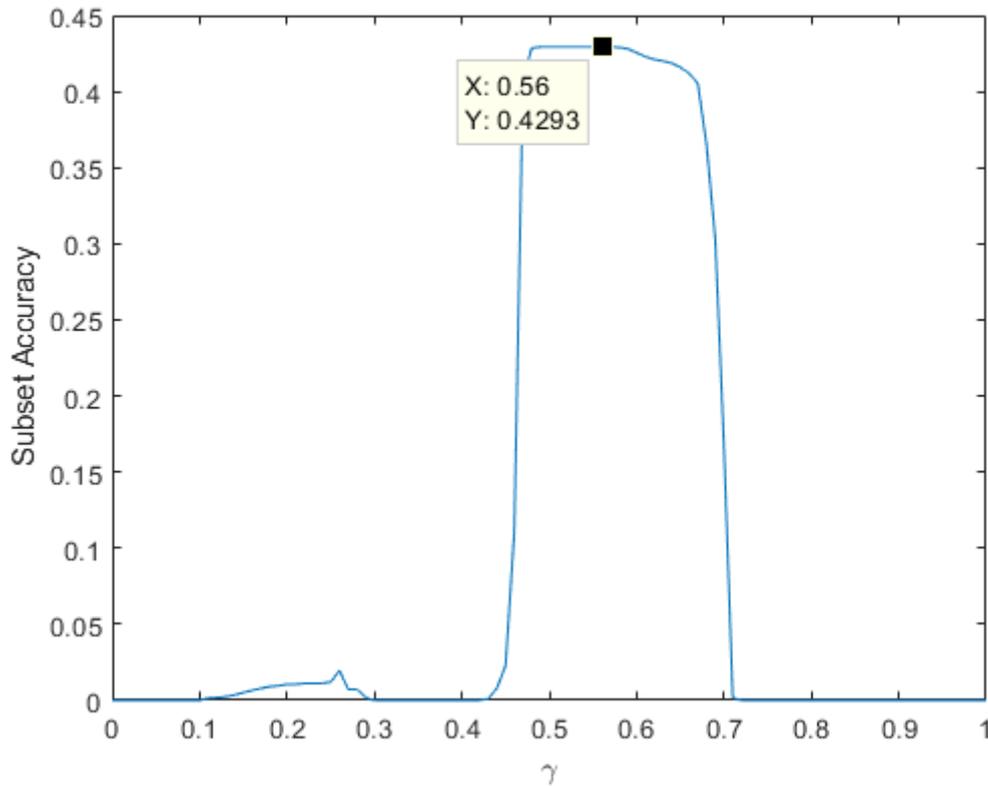
Finalement la prédiction d'un label  $\ell_j$  est annoncée pertinente si  $\mathbb{P}^{\mathbf{x}}(E_j) \geq \mathbb{P}^{\mathbf{x}}(\bar{E}_j)$ , ce qui permet d'obtenir l'ensemble de labels prédit par le classifieur pour l'instance  $\mathbf{x}$  :

$$\hat{Y} = \left\{ \ell_j \in \mathcal{Y} \mid \mathbb{P}^{\mathbf{x}}(E_{\ell_j}) \geq 0.5 \right\}.$$

## 5.6 Résultats et discussion

Cette section évalue les performances de la méthode proposée en utilisant des signaux extraits de MIMIC III. Les paramètres des simulations réalisées dans cette étude :

- Longueur de la fenêtre N  
Cela représente la taille des fenêtres prises à partir des différents signaux. C'est le temps nécessaire pour collecter les données de tout nouveau patient pour donner un diagnostic de son état. Dans cette simulation, cette valeur est fixée à  $N = 12$  heures.
- Division de la base de données  
La fraction d'instances choisie pour la phase d'apprentissage est fixée à  $\frac{2}{3}$ .
- Seuil de support  $\text{Supp}_{\min}$   
Cette valeur sera ajustée afin d'obtenir la meilleure amélioration apportée par les règles d'association.


 FIGURE 5.5 – Valeurs moyennes du "Subset Accuracy" pour  $0 \leq \text{Supp}_{\min} \leq 0.1$  en fonction de  $\gamma$ 

Quelques statistiques de la base de données multi-label obtenue avec la configuration détaillée ci-dessus sont présentées dans le TABLEAU 5.3. La méthode proposée est appliquée à cette base de données pour chaque instance de test et pour chaque label.

Statistique	Valeur
Cardinalité des labels	4.1614
Densité des labels	0.1601
Diversité des labels	145
Diversité proportionnelle	0.0256

TABLEAU 5.3 – Statistiques de la base de données multi-label issue de MIMIC III.

Dans un premier temps, les performances de la méthode proposée sont évaluées pour plusieurs valeurs de  $\text{Supp}_{\min}$  allant de 0 (toutes les règles d'associations sont utilisées) à 0.1 (aucune des règles d'association n'est utilisée) et pour  $\gamma$  allant de 0 à 1. On remarque que certaines métriques de performances sont influencées par la valeur de  $\text{Supp}_{\min}$  et d'autre par celle de  $\gamma$ . On choisit alors d'optimiser la valeur de  $\gamma$  suivant celle du "Subset Accuracy" et la valeur de  $\text{Supp}_{\min}$  suivant celle du "Average Precision". La FIGURE 5.5 montre les valeurs moyennes du "Subset Accuracy" sur toutes les valeurs de  $\text{Supp}_{\min}$  en fonction de  $\gamma$  et on remarque que la valeur maximale est obtenue pour  $\gamma^* = 0.56$  qui sera donc considérée comme la valeur optimale de  $\gamma$ . Le même procédé est appliqué en suivant les valeurs du "Average Precision" pour  $\gamma = \gamma^* = 0.56$  en fonction de  $\text{Supp}_{\min}$  et on remarque que la valeur maximale est obtenue pour  $\text{Supp}_{\min}^* = 0.019$  qui sera donc considérée comme la valeur optimale de  $\text{Supp}_{\min}$ . En utilisant les valeurs optimales sur la base de test, on obtient les résultats du TABLEAU 5.4.

Les TABLEAUX 5.5 et 5.6 montrent la moyenne de quelques métriques d'évaluation multi-label qui sont le "Hamming Loss", le "One-Error", le "Coverage", le "Ranking Loss", le "Subset Accuracy", le "Recall" et le "Average Precision", avec leur écart type pour 15 simulations de divisions

TABLEAU 5.4 – Métriques de performances pour les valeurs optimales de  $\gamma$  et  $\text{Supp}_{\min}$

$\gamma$	$\text{Supp}_{\min}$	Hloss	One Error	Cov	Rloss	Subsetacc	Recall	Avgprec
0.56	0.019	0.1074	0.2537	5.6512	0.0721	0.4416	0.5918	0.8041

différentes apprentissage/test de la base de données (moyenne  $\pm$  écart-type) en comparant la méthode proposée avec plusieurs méthodes classiques de transformations multi-label couplées avec plusieurs méthodes classiques de classification mono-label que sont :

- Naïve bayésien (NB)
- Méthode des k plus proches voisins (KNN)
- Machine à vecteurs de support (SVM)
- Arbre de décision (DT)
- Analyse discriminante linéaire (LDA)

Une amélioration significative est notée en utilisant la méthode proposée. Sur les 7 métriques d'évaluation, la méthode proposée gagne le premier rang sur toutes les autres méthodes pour 5 d'entre elles. Le rang moyen des métriques d'évaluation de la méthode proposée est de 3.86. Ceci s'explique par la capacité de la méthode proposée à traiter les données manquantes. En effet pour les méthodes classiques où une disponibilité totale des instances est nécessaire, il a fallu supprimer des instances et des caractéristiques afin d'obtenir une base d'apprentissage sans données manquantes et traitable par les méthodes classiques. La base de données médicale ayant beaucoup de données manquantes les performances s'en trouvent remarquablement dégradées. On remarque dans le tableau que le choix du classifieur n'agit souvent pas sur les performances contrairement au choix de la méthode de transformation ce qui indique la présence de classifieurs factices. Ceci s'explique par le manque d'informations utiles de la base de données qui est due, en partie, à la suppression d'une partie des données indispensable au fonctionnement des classifieurs classiques.

TABLEAU 5.5 – Comparaison des méthodes classiques de classification multi-label avec la méthode proposée.

Méthode de transformation	Classifieur	Hloss ↓	One Error ↓	Coverage ↓	Rloss ↓
BR	NB	0.325 $\pm$ 0.005(17)	1.000 $\pm$ 0.000(12)	23.779 $\pm$ 0.494(17)	0.904 $\pm$ 0.037(22)
	KNN	0.325 $\pm$ 0.005(18)	1.000 $\pm$ 0.000(13)	23.779 $\pm$ 0.494(18)	0.904 $\pm$ 0.037(23)
	SVM	0.325 $\pm$ 0.005(19)	1.000 $\pm$ 0.000(14)	23.779 $\pm$ 0.494(19)	0.904 $\pm$ 0.037(24)
	DT	0.325 $\pm$ 0.005(20)	1.000 $\pm$ 0.000(15)	23.779 $\pm$ 0.494(20)	0.904 $\pm$ 0.037(25)
	LDA	0.325 $\pm$ 0.005(21)	1.000 $\pm$ 0.000(16)	23.779 $\pm$ 0.494(21)	0.904 $\pm$ 0.037(26)
CC	NB	0.803 $\pm$ 0.001(22)	1.000 $\pm$ 0.000(17)	22.834 $\pm$ 0.033(12)	0.004 $\pm$ 0.001(1)
	KNN	0.803 $\pm$ 0.001(23)	1.000 $\pm$ 0.000(18)	22.834 $\pm$ 0.033(13)	0.004 $\pm$ 0.001(2)
	SVM	0.803 $\pm$ 0.001(24)	1.000 $\pm$ 0.000(19)	22.834 $\pm$ 0.033(14)	0.004 $\pm$ 0.001(3)
	DT	0.803 $\pm$ 0.001(25)	1.000 $\pm$ 0.000(20)	22.834 $\pm$ 0.033(15)	0.004 $\pm$ 0.001(4)
	LDA	0.803 $\pm$ 0.001(26)	1.000 $\pm$ 0.000(21)	22.834 $\pm$ 0.033(16)	0.004 $\pm$ 0.001(5)
CLR	NB	0.259 $\pm$ 0.010(12)	1.000 $\pm$ 0.000(22)	23.974 $\pm$ 0.111(22)	0.777 $\pm$ 0.033(17)
	KNN	0.259 $\pm$ 0.010(13)	1.000 $\pm$ 0.000(23)	23.974 $\pm$ 0.111(23)	0.777 $\pm$ 0.033(18)
	SVM	0.259 $\pm$ 0.010(14)	1.000 $\pm$ 0.000(24)	23.974 $\pm$ 0.111(24)	0.777 $\pm$ 0.033(19)
	DT	0.259 $\pm$ 0.010(15)	1.000 $\pm$ 0.000(25)	23.974 $\pm$ 0.111(25)	0.777 $\pm$ 0.033(20)
	LDA	0.259 $\pm$ 0.010(16)	1.000 $\pm$ 0.000(26)	23.974 $\pm$ 0.111(26)	0.777 $\pm$ 0.033(21)
LP	NB	0.126 $\pm$ 0.001(2)	0.269 $\pm$ 0.009(2)	20.655 $\pm$ 0.043(2)	0.025 $\pm$ 0.001(6)
	KNN	0.126 $\pm$ 0.001(3)	0.269 $\pm$ 0.009(3)	20.655 $\pm$ 0.043(3)	0.025 $\pm$ 0.001(7)
	SVM	0.126 $\pm$ 0.001(4)	0.269 $\pm$ 0.009(4)	20.655 $\pm$ 0.043(4)	0.025 $\pm$ 0.001(8)
	DT	0.126 $\pm$ 0.001(5)	0.269 $\pm$ 0.009(5)	20.655 $\pm$ 0.043(5)	0.025 $\pm$ 0.001(9)
	LDA	0.126 $\pm$ 0.001(6)	0.269 $\pm$ 0.009(6)	20.655 $\pm$ 0.043(6)	0.025 $\pm$ 0.001(10)
RAKEL	NB	0.154 $\pm$ 0.008(8)	0.356 $\pm$ 0.213(9)	21.630 $\pm$ 1.358(9)	0.025 $\pm$ 0.016(11)
	KNN	0.154 $\pm$ 0.013(9)	0.276 $\pm$ 0.011(7)	21.566 $\pm$ 1.391(7)	0.029 $\pm$ 0.015(13)
	SVM	0.158 $\pm$ 0.007(11)	0.466 $\pm$ 0.325(11)	22.280 $\pm$ 1.439(11)	0.038 $\pm$ 0.047(15)
	DT	0.152 $\pm$ 0.011(7)	0.346 $\pm$ 0.164(8)	21.570 $\pm$ 1.310(8)	0.027 $\pm$ 0.018(12)
	LDA	0.157 $\pm$ 0.012(10)	0.389 $\pm$ 0.229(10)	21.924 $\pm$ 1.551(10)	0.029 $\pm$ 0.017(14)
Méthode proposée		0.108 $\pm$ 0.002(1)	0.260 $\pm$ 0.009(1)	5.695 $\pm$ 0.102(1)	0.071 $\pm$ 0.003(16)

TABLEAU 5.6 – Comparaison des méthodes classiques de classification multi-label avec la méthode proposée (suite).

Méthode de transformation	Classifieur	Subset Accuracy ↑	Recall ↑	Average Precision ↑	Average Rank ↓
BR	NB	0.000 ± 0.000(26)	0.370 ± 0.022(21)	0.114 ± 0.003(26)	20.14
	KNN	0.000 ± 0.000(19)	0.370 ± 0.022(2)	0.114 ± 0.003(18)	15.86
	SVM	0.000 ± 0.000(18)	0.370 ± 0.022(5)	0.114 ± 0.003(17)	16.57
	DT	0.000 ± 0.000(17)	0.370 ± 0.022(1)	0.114 ± 0.003(16)	16.29
	LDA	0.000 ± 0.000(20)	0.370 ± 0.022(4)	0.114 ± 0.003(20)	18.29
CC	NB	0.000 ± 0.000(16)	0.996 ± 0.001(3)	0.156 ± 0.001(19)	12.86
	KNN	0.000 ± 0.000(25)	0.996 ± 0.001(15)	0.156 ± 0.001(25)	17.29
	SVM	0.000 ± 0.000(24)	0.996 ± 0.001(14)	0.156 ± 0.001(24)	17.43
	DT	0.000 ± 0.000(23)	0.996 ± 0.001(13)	0.156 ± 0.001(23)	17.57
	LDA	0.000 ± 0.000(22)	0.996 ± 0.001(12)	0.156 ± 0.001(22)	17.71
CLR	NB	0.000 ± 0.000(21)	0.455 ± 0.030(11)	0.133 ± 0.005(21)	18.00
	KNN	0.000 ± 0.000(15)	0.455 ± 0.030(20)	0.133 ± 0.005(10)	17.43
	SVM	0.000 ± 0.000(14)	0.455 ± 0.030(19)	0.133 ± 0.005(9)	17.57
	DT	0.000 ± 0.000(13)	0.455 ± 0.030(18)	0.133 ± 0.005(8)	17.71
	LDA	0.000 ± 0.000(12)	0.455 ± 0.030(17)	0.133 ± 0.005(7)	17.86
LP	NB	0.015 ± 0.002(11)	0.390 ± 0.005(16)	0.509 ± 0.004(6)	6.43
	KNN	0.015 ± 0.002(10)	0.390 ± 0.005(26)	0.509 ± 0.004(15)	9.57
	SVM	0.015 ± 0.002(9)	0.390 ± 0.005(25)	0.509 ± 0.004(14)	9.71
	DT	0.015 ± 0.002(8)	0.390 ± 0.005(24)	0.509 ± 0.004(13)	9.86
	LDA	0.015 ± 0.002(7)	0.390 ± 0.005(23)	0.509 ± 0.004(12)	10.00
RAKEL	NB	0.000 ± 0.000(6)	0.066 ± 0.096(22)	0.408 ± 0.111(11)	10.86
	KNN	0.002 ± 0.006(5)	0.066 ± 0.143(10)	0.463 ± 0.085(5)	8.00
	SVM	0.000 ± 0.002(4)	0.021 ± 0.080(9)	0.353 ± 0.129(4)	9.29
	DT	0.001 ± 0.004(3)	0.091 ± 0.125(8)	0.393 ± 0.102(3)	7.00
	LDA	0.001 ± 0.004(2)	0.056 ± 0.115(7)	0.402 ± 0.111(2)	7.86
Méthode proposée		0.435 ± 0.008(1)	0.585 ± 0.008(6)	0.801 ± 0.006(1)	3.86

## 5.7 Conclusion

Ce chapitre a abordé le diagnostic médical multi-pathologique pour les maladies cardio-vasculaires. Les mêmes signaux physiologiques du chapitre précédent ont été repris afin de nous fournir des informations sur d'autres éventuelles complications cardiaques. Un ensemble de 55 maladies cardio-vasculaires, identifiées dans la base de données médicale MIMIC III, ont été choisies pour l'étude de leur diagnostic en temps réel. Afin de mieux prendre en considération les similarités opérationnelles entre maladies, on a proposé d'associer des étiquettes ou labels à chaque patient selon les maladies qui lui ont été diagnostiquées. Ces maladies étant représentées par des descriptifs communs, on a associé à chacune un ensemble de labels. On a commencé par introduire la classification multi-label, puis on a établi une formalisation mathématique et un état de l'art. On a ensuite détaillé la modélisation du problème de classification multi-label. Enfin, on a introduit une approche multi-label basée sur la théorie des fonctions de croyance qu'on a validé grâce aux données extraites de la base médicale MIMIC III.



## Chapitre 6

# Classification multi-label par ensemble pour le suivi médical

### Sommaire

---

<b>6.1 Introduction</b> . . . . .	<b>116</b>
<b>6.2 Approche proposée : COLEX</b> . . . . .	<b>116</b>
6.2.1 La méthode RAKEL . . . . .	117
6.2.2 Le coefficient de corrélation binaire $\phi$ . . . . .	117
6.2.3 Facteur de corrélation globale . . . . .	118
6.2.4 Stratégie d'échange : Méthode COLEX . . . . .	119
<b>6.3 Résultats et discussions</b> . . . . .	<b>122</b>
6.3.1 Application de la méthode CoLEx à la base de données médicale . . . . .	122
6.3.2 Illustration des performances de la méthode proposée sur des bases de données réelles multi-label . . . . .	123
6.3.3 Efficacité de la stratégie d'échange . . . . .	126
6.3.4 Impact des paramètres $k$ et $n$ . . . . .	127
6.3.5 Comparaison avec d'autres méthodes . . . . .	128
<b>6.4 Conclusion</b> . . . . .	<b>132</b>

---

## 6.1 Introduction

La classification multi-label fait partie de l'apprentissage automatique où la même instance peut appartenir à plusieurs classes simultanément. Dans ce cas, les classes ne sont pas exclusives. L'une des méthodes de transformation les plus populaires est la méthode "Binary Relevance" (BR) [172], où chaque label génère un problème de classification binaire. En conséquence, un problème de classification multi-label est donc transformé en différentes tâches de classification binaire dont le nombre est égal au nombre de labels. Cependant, un inconvénient majeur de cette méthode est qu'elle gère les labels séparément et qu'elle ignore donc les dépendances entre eux. Cela peut entraîner certaines conséquences, telles que la dégradation des performances et la perte d'informations importantes. Si la méthode du "Label Powerset" (LP) prend en compte les dépendances existantes entre les labels, son utilisation sur des bases de données caractérisées par un grand nombre de labels conduit à certains inconvénients. En effet, à mesure que le nombre de labels augmente, le nombre de sous-ensembles de labels augmente de façon exponentielle. Afin de surmonter ce problème, la méthode "Random  $k$ -labelsets" (RAkEL) est introduite [155]. Cette méthode considère un certain nombre de sous-ensembles aléatoires de labels de taille  $k$  et utilise la méthode LP sur chacun de ces sous-ensembles. Avec un nombre approprié de  $k$ -labelsets ou ensemble de labels de taille  $k$ , cette méthode a tendance à gérer les corrélations entre les labels. Dans la littérature, plusieurs études sont présentées sur les méthodes d'ensemble, certaines d'entre elles utilisent la corrélation entre les labels, d'autres sont basées sur la sélection de caractéristiques [156].

Le chapitre 5 a présenté une approche de classification multi-label, profitant à la fois de la classification binaire et de la dépendance des labels entre eux. Elle transforme le problème multi-label en des problèmes de classifications binaires, une par label, qu'elle corrige grâce à la règle d'association proposée pour profiter de la dépendance entre les labels. Afin de profiter aussi des dépendances éventuelles entre les caractéristiques et d'éviter de les supposer indépendantes, il est préférable de construire des classifieurs utilisant plusieurs caractéristiques à la fois, donc exigeant l'apprentissage d'un ou de plusieurs classifieurs multi-classe. L'approche proposée dans les chapitres précédents utilisant la théorie des fonctions de croyance n'est pas adaptée aux problèmes de classification multi-classe. Il est donc nécessaire de passer à une méthode de classification classique. Néanmoins, une contribution peut être apportée au niveau de la transformation du problème de classification multi-label en un ou plusieurs problèmes de classification multi-classe. Ce chapitre propose une méthode originale appelée CoLEx pour "Correlation based Labelsets by Exchange strategy", inspirée par la méthode RAkEL. CoLEx optimise le choix des  $k$ -labelsets en considérant les corrélations existantes entre les labels. Pour y parvenir, un indicateur de corrélation est calculé, puis une procédure d'optimisation est appliquée pour choisir la meilleure combinaison de  $k$ -labelsets. Le choix est en réalité basé sur une stratégie d'échange qui vise à maximiser les corrélations entre les labels d'un même  $k$ -labelset en échangeant des labels entre les différents  $k$ -labelsets. Ce processus est répété jusqu'à atteindre un état dans lequel tout échange diminue le facteur de corrélation global. Un classifieur est ensuite entraîné pour chaque  $k$ -labelset à l'aide de la méthode LP. Pour la classification d'une nouvelle instance, tous les classifieurs entraînés sont utilisés et un système de vote est ensuite appliqué pour donner une décision sur chacun des  $q$  labels. Des expériences sur des ensembles de données du monde réel montrent l'efficacité de notre méthode par rapport à la méthode RAkEL et à d'autres méthodes de pointe de la littérature.

## 6.2 Approche proposée : COLEX

L'apprentissage des relations entre les labels est un problème difficile en classification multi-label. En effet, les classes se chevauchent et sont corrélées, comme l'association d'un label a une instance peut fournir des informations sur l'appartenance de cette instance à d'autres labels. En

l'absence de relations entre les labels, l'aspect multi-label devient sans intérêt et les données peuvent être traitées par une méthode mono-label sans perte de généralité. Différentes approches ont été proposées dans la littérature pour modéliser la corrélation entre les labels [156; 173; 174]. Les relations entre les labels peuvent être considérées sous forme de relations binaires entre chaque paire d'labels, ou bien des relations d'ordre supérieur entre chaque label et les labels restants. Les relations d'ordre supérieur sont plus complexes à représenter que les relations binaires, qui peuvent être mesurées, e.g., par la probabilité conditionnelle d'une label étant donnée l'existence d'une autre.

### 6.2.1 La méthode RAKEL

On rappelle que la méthode Label Powerset (LP) est une méthode de transformation qui considère chaque sous-ensemble de labels comme une classe unique dans un problème mono-label équivalent. Ainsi, s'il y a  $q$  labels dans l'espace des labels  $\mathcal{Y}$  du problème multi-label, on peut obtenir un maximum de  $2^q$  classes dans le problème correspondant en mono-label. Pour une nouvelle instance à classifier, toute méthode de classification mono-label peut être utilisée pour prédire sa classe. La sortie du classifieur mono-label est alors remplacée par le sous-ensemble équivalent de labels pour obtenir la sortie du classifieur multi-label. La méthode LP offre un moyen simple d'utiliser la méthode mono-label pour résoudre les problèmes de classification multi-label. Cependant, elle présente quelques défauts, en particulier lorsque l'espace des labels contient un nombre important de labels. Dans ce cas, le nombre d'instances associées à chaque classe diminue et rend la méthode mono-label incapable de généraliser correctement pour les nouvelles instances.

La méthode "Random  $k$ -labelsets" (RAKEL) a été proposée avec l'objectif principal de trouver un compromis entre les méthodes "Binary Relevance" et "Label Powerset" pour exploiter les avantages des deux méthodes. Cette méthode applique la stratégie LP sur un plusieurs sous-ensembles de labels de taille  $k$ , appelés  $k$ -labelsets. Ces sous-ensembles sont choisis aléatoirement parmi  $C_q^k = \binom{q}{k}$  différents  $k$ -labelsets possibles au total. Le nombre  $n$  de  $k$ -labelsets sélectionnés par la méthode RAKEL est un paramètre allant de 2 à  $C_q^k - 1$  spécifié par l'utilisateur en fonction de l'application à traiter. Ainsi, en considérant des sous-ensembles de labels au lieu de considérer des labels individuellement, RAKEL prend en compte les dépendances entre les labels surmontant les défauts de la méthode BR. En limitant le nombre de  $k$ -labelsets considérés, RAKEL surmonte le coût élevé de la méthode "Label Powerset" en terme de puissance de calcul. Ceci étant, on peut conclure intuitivement que la sélection aléatoire des  $k$ -labelsets est l'une des difficultés majeures de la méthode.

Dans la suite, nous présentons une méthode originale qui consiste à entraîner, comme la méthode RAKEL, plusieurs classifieurs mono-label en utilisant chacun un  $k$ -labelset différent, mais au lieu de les sélectionner aléatoirement, ils seront choisis en se basant sur la corrélation entre les labels au sein d'un même  $k$ -labelset et sur la corrélation globale entre tous les  $k$ -labelsets. Pour cela, il est nécessaire d'introduire le coefficient de corrélation binaire  $\phi$ .

### 6.2.2 Le coefficient de corrélation binaire $\phi$

Afin de mieux introduire le coefficient de corrélation binaire, on représente les labels de chaque instance  $\mathbf{x}_i$  de la base de données multi-label par un vecteur binaire  $\mathbf{v}_i$  contenant une composante  $\lambda_{i,j}$  pour chaque label du domaine des labels  $\mathcal{Y}$  qui prend la valeur  $\lambda_{i,j} = 1$  si cette instance est labélisée par  $\ell_j$  et 0 sinon. Dans le cadre de variables binaires, la mesure de la corrélation entre les variables est généralement effectuée en utilisant le coefficient  $\phi$ . C'est l'équivalent du coefficient de Pearson appliqué à la référence des variables binaires. Soit  $\mathcal{D}$  une base de données multi-label étiquetée telle que

$$\mathcal{D} = \{(\mathbf{x}_i, Y_i) \mid \mathbf{x}_i \in \mathcal{X}, Y_i \subseteq \mathcal{Y}, 1 \leq i \leq M\}$$

avec  $\mathcal{X} = \mathbb{R}^d$  et  $\mathcal{Y} = \{\ell_1, \ell_2 \dots \ell_q\}$ . On définit les quantités  $c_{(j,j')}$ ,  $c_{(-j,j')}$ ,  $c_{(j,-j')}$  et  $c_{(-j,-j')}$  comme suit

$$\begin{aligned} c_{(j,j')} &= \frac{1}{M} \sum_{i=1}^M \lambda_{i,j} \lambda_{i,j'}, \\ c_{(-j,j')} &= \frac{1}{M} \sum_{i=1}^M (1 - \lambda_{i,j}) \lambda_{i,j'}, \\ c_{(j,-j')} &= \frac{1}{M} \sum_{i=1}^M \lambda_{i,j} (1 - \lambda_{i,j'}), \\ c_{(-j,-j')} &= \frac{1}{M} \sum_{i=1}^M (1 - \lambda_{i,j}) (1 - \lambda_{i,j'}), \quad \forall j, j' \in \{1, 2 \dots q\}. \end{aligned}$$

En utilisant ces notations, le coefficient de corrélation binaire  $\phi$  entre les labels  $\ell_j$  et  $\ell_{j'}$  peut être calculé par :

$$\phi_{j,j'} = \frac{c_{(j,j')}c_{(-j,-j')} - c_{(-j,j')}c_{(j,-j')}}{\sqrt{(c_{(j,j')} + c_{(-j,j')})(c_{(j,-j')} + c_{(-j,-j')})(c_{(-j,j')} + c_{(j,-j')})(c_{(-j,-j')} + c_{(j,j')}}}.$$

Sous l'hypothèse de variables indépendantes, ce coefficient est égal à zéro. En fait, plus il est proche de zéro, moins les labels sont corrélés. Contrairement au coefficient de Pearson, le coefficient  $\phi$  n'est pas limité entre -1 et 1. Dans notre travail, nous utilisons ce coefficient comme indicateur de corrélation pour chaque paire de labels possible dans l'espace  $\mathcal{Y}$ , comme il sera montré ci-dessous.

### 6.2.3 Facteur de corrélation globale

La méthode proposée dans ce chapitre consiste à sélectionner un certain nombre de  $k$ -labelsets ayant une forte corrélation entre leurs labels. En effet, avec des corrélations élevées, ces labels sont plus susceptibles de tomber ensemble pour les nouvelles instances. En les prenant ensemble, nous évitons d'avoir des prédictions contradictoires par différents prédicteurs. Afin de mesurer la corrélation des labels au sein d'un  $k$ -labelset particulier, nous calculons d'abord la corrélation entre toute les paires de labels dans celui-ci en utilisant le coefficient  $\phi$  présenté dans la section précédente.

Ce coefficient est égal à zéro lorsqu'il n'y a pas de corrélation entre les deux labels considérées et a une valeur absolue élevée lorsqu'il existe une forte corrélation entre eux. Une matrice de corrélation symétrique  $\Phi$  est construite à partir de la valeur absolue de ce coefficient de corrélation :

$$\begin{aligned} \Phi(\ell_j, \ell_{j'}) &= |\phi_{j,j'}| \quad \forall j, j' \in \{1, 2 \dots q\}, j \neq j' \\ \Phi(\ell_j, \ell_j) &= 0 \quad \forall j \in \{1, 2 \dots q\}, \\ \Phi &= \begin{pmatrix} 0 & |\phi_{1,2}| & \dots & |\phi_{1,q}| \\ |\phi_{2,1}| & 0 & \dots & |\phi_{2,q}| \\ \vdots & & \ddots & \\ |\phi_{q,1}| & |\phi_{q,2}| & \dots & 0 \end{pmatrix}, \end{aligned}$$

avec  $q$  le nombre de labels dans  $\mathcal{Y}$ . Cette matrice nécessite le calcul du coefficient  $\phi$  pour  $\frac{q(q-1)}{2}$  paires de labels. Pour un labelset  $S$  quelconque de taille  $k$ , le facteur de corrélation qui lui est propre est défini par :

$$C(S) = \frac{1}{k(k-1)} \sum_{\substack{(\ell_j, \ell_{j'}) \in S^2 \\ \ell_j \neq \ell_{j'}}} \Phi(\ell_j, \ell_{j'}).$$

Afin de mieux introduire la méthode, on définit la matrice  $\mathbf{S}$  formée de  $n$  lignes chacune représen-

tant un  $k$ -labelset  $S_u$  ( $1 \leq u \leq n$ ). L'élément de la  $u^{\text{ème}}$  et de la  $v^{\text{ème}}$  colonne de cette matrice est notée  $s_{u,v}$ .  $s_{u,v}$  est alors un label appartenant à  $S_u$ . Ainsi, le problème peut être reformulé comme un problème d'optimisation où le paramètre à maximiser est la corrélation totale de l'ensemble des  $n$   $k$ -labelsets sélectionnés :

$$\zeta(\mathbf{S}) = \frac{1}{n} \sum_{S_i \in \mathbf{S}} C(S_i),$$

où les variables sont les ensembles  $S_i$  ( $1 \leq i \leq n$ ). Le but du problème d'optimisation proposé est d'obtenir un ensemble de  $n$   $k$ -labelsets  $\mathbf{S}^*$  avec la valeur la plus élevée de la corrélation entre labels :

$$\max_{\mathbf{S}} \zeta(\mathbf{S})$$

Une solution simpliste à ce problème reste de considérer tous les  $k$ -labelsets possibles dont le nombre totale est égal à  $\binom{q}{k}$ , et sélectionner ensuite les ensembles ayant les corrélations les plus élevées. Au lieu de faire ces calculs exhaustifs, nous proposons, dans ce qui suit, une stratégie d'échange pour résoudre le problème d'optimisation.

#### 6.2.4 Stratégie d'échange : Méthode COLEX

La méthode CoLex ("Correlation based Labelsets by Exchange strategy") est proposée dans le but d'optimiser la sélection des  $n$  différents  $k$ -labelsets sur lesquels la méthode LP sera appliquée individuellement. Comme son nom l'indique, cette méthode est basée sur des échanges de labels entre les  $k$ -labelsets et l'objectif de ces échanges est d'aboutir à une nouvelle sélection des  $n$   $k$ -labelsets ayant le facteur de corrélation  $\zeta$ , présenté dans la section 6.2.3, le plus élevée. Concrètement, la méthode consiste à tester chaque échange de labels possible, c-à-d chaque paire de labels issues de différents labelsets et de déterminer la variation du facteur de corrélation globale  $\Delta\zeta$  qui s'en suit. Si ce dernier est strictement positif, l'échange est appliqué et l'algorithme est lancé à nouveau, sinon, on passe au test de l'échange suivant. Il s'agit d'un moyen rapide de trouver une solution pseudo-optimale car l'optimisation absolue est coûteuse en puissance de calcul. De plus, compte tenu du fait que la méthode est basée sur des échanges de labels uniquement, le nombre d'occurrence totale de chaque label dans l'ensemble des  $n$  des  $k$ -labelsets reste invariant au cours de l'optimisation ce qui empêche d'aboutir à une sélection déséquilibrée entre labels. Cette matrice est initialisée à  $\mathbf{S}^{(0)}$  en premier lieu en suivant l'algorithme 6. Elle est construite ligne par ligne en partant de la première colonne et en affectant les labels de sorte que  $s_{(u,v)}^{(0)} = \ell_r \quad \forall 1 \leq u \leq n$  et  $1 \leq v \leq k$  avec  $r = ((u-1).k + v) \% q + 1$  et  $\%$  étant l'opérateur modulo. En initialisant cette matrice comme indiquée, l'équilibre entre les labels est garanti et maintenu vu qu'elle contient des occurrences de labels dont l'écart est au plus 1. On obtient une matrice équilibré. Par exemple, pour  $n = 5$ ,  $k = 4$  et  $q = 7$ , l'algorithme 6 donne

$$\mathbf{S}^{(0)} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 & \ell_4 \\ \ell_5 & \ell_6 & \ell_7 & \ell_1 \\ \ell_2 & \ell_3 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_1 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix},$$

ce qui signifie que  $S_1^{(0)} = \{\ell_1, \ell_2, \ell_3, \ell_4\}$ ,  $S_2^{(0)} = \{\ell_1, \ell_5, \ell_6, \ell_7\}$ ,  $S_3^{(0)} = \{\ell_2, \ell_3, \ell_4, \ell_5\}$ ,  $S_4^{(0)} = \{\ell_1, \ell_2, \ell_6, \ell_7\}$ ,  $S_5^{(0)} = \{\ell_3, \ell_4, \ell_5, \ell_6\}$ . On remarque que tous les labels sont répétés 3 fois sauf  $\ell_7$  qui est répété 2 fois ce qui garantie qu'aucun label n'est favorisé par rapport au autre dans la sélection des  $n$   $k$ -labelsets.

Une fois la matrice  $\mathbf{S}^{(0)}$  initialisée, chaque couple de labels issue de deux  $k$ -labelsets différents est évalué pour savoir si le fait d'échanger la position de ses labels dans la matrice  $\mathbf{S}^{(0)}$  provoque une augmentation du facteur de corrélation globale  $\zeta$ . On trouve le premier couple  $(s_{(u,v)}, s_{(u',v)})$  tel que  $1 \leq u < u' \leq n$  et  $1 \leq v, v' \leq k$  et qui assure les conditions suivantes :

$$s_{(u,v)} \notin S_{u'},$$

**Algorithm 6:** Matrice d'initialisation **S**
**FONCTION** Matrice d'initialisation( $n, k, q$ )

**S**  $\leftarrow$  Matrice( $n \times k$ )

 $r \leftarrow 1$ 
**Pour**  $u \in 1..n$ 

     **Pour**  $v \in 1..k$ 

          $s_{(u,v)} \leftarrow \ell_r$ 

          $r \leftarrow r + 1$ 

         **Si**  $r = q+1$ 

              $r \leftarrow 1$ 

         **FinSi**

     **FinPour**
**FinPour**
**Retourner S**

$$s_{(u',v')} \notin S_u,$$

$$\Delta\zeta_{(u,v) \Leftrightarrow (u',v')}(\mathbf{S}^{(0)}) > 0,$$

avec  $\Delta\zeta_{(u,v) \Leftrightarrow (u',v')}(\mathbf{S}^{(0)})$  la variation du facteur de corrélation globale issue de l'échange des labels  $s_{(u,v)}^{(0)}$  et  $s_{(u',v')}^{(0)}$  des  $k$ -labelsets respectifs  $S_u^{(0)}$  et  $S_{u'}^{(0)}$ . Les positions des éléments  $s_{(u,v)}^{(0)}$  et  $s_{(u',v')}^{(0)}$  sont alors échangées dans la matrice  $\mathbf{S}^{(0)}$  et on obtient une nouvelle matrice  $\mathbf{S}^{(1)}$  pour l'itération suivante. En reprenant l'exemple plus haut, on obtient les matrices suivantes chacune issue d'un échange particulier :

$$\begin{pmatrix} \ell_5 & \ell_2 & \ell_3 & \ell_4 \\ \ell_1 & \ell_6 & \ell_7 & \ell_1 \\ \ell_2 & \ell_3 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_1 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix}, \begin{pmatrix} \ell_1 & \ell_5 & \ell_3 & \ell_4 \\ \ell_2 & \ell_6 & \ell_7 & \ell_1 \\ \ell_2 & \ell_3 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_1 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix}, \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 & \ell_1 \\ \ell_5 & \ell_6 & \ell_7 & \ell_4 \\ \ell_2 & \ell_3 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_1 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix} \text{ et } \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 & \ell_4 \\ \ell_5 & \ell_6 & \ell_7 & \ell_1 \\ \ell_2 & \ell_1 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_3 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix}.$$

La première matrice est issue d'un échange entre les labels  $\ell_1$  et  $\ell_5$  des lignes respectives 1 et 2, mais  $\ell_1$  existe déjà sur la ligne 2 donc cet échange est inutile car il conduit à une configuration illogique. La deuxième est issue d'un échange entre les labels  $\ell_2$  et  $\ell_5$  des lignes respectives 1 et 2, par contre, cet échange aboutit à la variation de la corrélation globale  $\Delta\zeta_{(1,2) \Leftrightarrow (2,1)}(\mathbf{S}^{(0)}) = -0.3$  ce qui ne satisfait pas les conditions d'application de l'échange. Le troisième échange introduit un doublet de  $\ell_1$  sur la première ligne; cet échange est donc ignoré. Enfin, pour la dernière matrice, on a  $\Delta\zeta_{(3,2) \Leftrightarrow (4,3)}(\mathbf{S}^{(0)}) = +0.2$ ; cet échange est pris en compte et on aura alors :

$$\mathbf{S}^{(1)} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 & \ell_4 \\ \ell_5 & \ell_6 & \ell_7 & \ell_1 \\ \ell_2 & \ell_1 & \ell_4 & \ell_5 \\ \ell_6 & \ell_7 & \ell_3 & \ell_2 \\ \ell_3 & \ell_4 & \ell_5 & \ell_6 \end{pmatrix}.$$

Cette opération est ainsi répétée jusqu'à ce qu'on aboutisse à une matrice  $\mathbf{S}^{(t^*)}$  pour laquelle tout échange de labels conduit à une diminution du facteur de corrélation globale  $\zeta$ , c-à-d

$$\Delta\zeta_{(u,v) \Leftrightarrow (u',v')}(\mathbf{S}^{(t^*)}) \leq 0 \quad \forall 1 \leq u < u' \leq n, 1 \leq v, v' \leq k.$$

Cette procédure est représentée par l'algorithme 7 qui prend en entrée une matrice **S** initiale et qui donne en sortie la matrice optimisée issue d'une succession d'échanges sur la matrice initiale.

Il est à noter que l'échange du label  $s_{(u,v)}^{(t)}$  du labelset  $S_u^{(t)}$  avec le label  $s_{(u',v')}^{(t)}$  du labelset  $S_{u'}^{(t)}$ ,

**Algorithm 7:** Stratégie d'échange

---

```

FONCTION Optimisation matrice(S)
    unstable ← true
    Tant que unstable
        unstable ← false
        Pour  $u \in 1..n - 1$ 
            Pour  $u' \in u + 1..n$ 
                Pour  $v \in 1..k$ 
                    Pour  $v' \in 1..k$ 
                        Si  $s_{(u,v)} \notin S_{u'} \text{ Et } s_{(u',v')} \notin S_u \text{ Et } \Delta\zeta_{(u,v) \Leftrightarrow (u',v')}(\mathbf{S}) > 0$ 
                            aide ←  $s_{(u,v)}$ 
                             $s_{(u,v)} \leftarrow s_{(u',v')}$ 
                             $s_{(u',v')} \leftarrow \textit{aide}$ 
                            aide ←  $s_{(u,v)}$ 
                            unstable ← true
                        FinSi
                    FinPour
                FinPour
            FinPour
        FinPour
    FinTantque
    Retourner S
    
```

---

provoque une variation  $\Delta\zeta_{(u,v) \Leftrightarrow (u',v')}(\mathbf{S}^{(t)})$  qui exige pour sa détermination peu d'effort de calcul. En effet, il est possible de démontrer que :

$$\Delta\zeta_{(u,v) \Leftrightarrow (u',v')}^{(t+1)} \propto \sum_{\substack{r=1 \\ r \neq v}}^k \Phi(s_{(u,r)}^{(t)}, s_{(u',v')}^{(t)}) - \sum_{r=1}^k \Phi(s_{(u,r)}^{(t)}, s_{(u,v)}^{(t)}) +$$

$$\sum_{\substack{r=1 \\ r \neq v'}}^k \Phi(s_{(u',r)}^{(t)}, s_{(u,v)}^{(t)}) - \sum_{r=1}^k \Phi(s_{(u',r)}^{(t)}, s_{(u',v')}^{(t)})$$

Les lignes de la matrice  $\mathbf{S}^{(t^*)}$  représentent alors les  $n$   $k$ -labelsets qui seront chacun utilisés par la méthode LP pour l'apprentissage de  $n$  différents classifieurs  $h_u(\cdot)$  ( $1 \leq u \leq n$ ). Ces classifieurs  $h_u(\cdot)$  sortent pour un vecteur de caractéristiques  $\mathbf{x}$  à classifier, un ensemble de labels chacun  $\hat{Y}^{(u)} = h_u(\mathbf{x})$  tel que  $\hat{Y}^{(u)} \subseteq S_u^{(t^*)}$ . Comme la méthode RAKEL, afin de fusionner les sorties de tous les classifieurs  $h_u(\cdot)$  et obtenir une prédiction générale pour  $\mathbf{x}$ , un système de vote est mis en place. En effet, pour chaque label  $\ell_j \in \mathcal{Y}$ , une valeur  $\mathbb{V}(\ell_j)$  ( $1 \leq j \leq q$ ) est calculée en fonction des sorties des classifieurs  $h_i(\cdot)$  :

$$\mathbb{V}(\ell_j) = \frac{\sum_{u=1}^n \mathbb{1}[\ell_j \in \hat{Y}^{(u)}]}{\sum_{u=1}^n \mathbb{1}[\ell_j \in S_u^{(t^*)}]}$$

L'ensemble de labels prédit pour le vecteur de caractéristiques  $\mathbf{x}$  est alors définit :

$$\hat{Y} = \{\ell_j \in \mathcal{Y} \mid \mathbb{V}(\ell_j) \geq \mathbb{V}_{th}\},$$

avec  $\mathbb{V}_{th}$  un seuil que l'on peut optimiser.

## 6.3 Résultats et discussions

### 6.3.1 Application de la méthode CoLEx à la base de données médicale

Dans le chapitre précédant, une base de données multi-label a été construite en se basant sur les données médicales de MIMIC III. Cette base regroupe un ensemble d'admissions à l'unité de soin intensif d'un hôpital pour lesquels un diagnostic incluant une pathologie cardiaque a été enregistré. En préservant la même notation, cette base de donnée sera notée  $\mathcal{H}$ . Vu que cette base de données a été configurée pour convenir à l'apprentissage du modèle évidentiel, son fort niveau d'indisponibilité de l'ordre de 47% ne présentait pas un inconvénient insurmontable. Cependant, dans une optique d'apprentissage basée sur la méthode SVM, ce n'est plus le cas. En effet, toutes les instances de la base de données doivent avoir toutes leur caractéristiques disponibles. En éliminant les instances ayant au moins une caractéristique indisponible, la base de donnée se vide littéralement. Alors, pour palier à ce problème, nous avons opté pour une adaptation de la base de donnée qui se fait en deux étapes. La première consiste à sélectionner des caractéristiques selon leur niveaux de disponibilité respectives et par la suite, la deuxième comprend la suppression de toutes les instances n'ayant pas toutes les caractéristiques disponibles. La sélection des caractéristiques est basée sur un seuil prédéfinie  $\xi$  où les seules caractéristiques retenues sont celles dont la disponibilité est supérieure ou égale à  $\xi$ . Soit  $\mathcal{H}$  la base d'origine définie comme suit :

$$\mathcal{H} = \{(\mathbf{x}_i, Y_i) \mid 1 \leq i \leq n\},$$

avec  $\mathbf{x}_i \in \mathcal{X}$ ,  $Y_i \subseteq \mathcal{Y} = \{\ell_1, \ell_2, \dots, \ell_q\} \quad \forall 1 \leq i \leq M$ . La dimension des vecteurs de caractéristiques est  $d$ , la taille de l'espace des labels est  $q$  et le nombre d'instances dans la base de données est  $M$ . On définit  $\mathbf{A}$ , la matrice de taille  $n \times d$ , indiquant les disponibilités des caractéristiques tel que  $A_{ij} \in \{0, 1\}$  prenant la valeur 1 lorsque  $x_i^{(j)}$  la  $j^{\text{ème}}$  composante de  $\mathbf{x}_i$  est disponible et 0 sinon. La disponibilité de la caractéristique  $j$  ( $1 \leq j \leq d$ ) est définie par :

$$\pi_j = \frac{1}{M} \sum_{i=1}^M A_{ij}.$$

On s'intéresse à la proportion des caractéristiques maintenues après élimination de celles qui présentent une disponibilité inférieure à un seuil  $\xi$  qui sera définie :

$$\omega_f(\xi) = \frac{1}{d} |\Gamma(\xi)|,$$

avec  $\Gamma(\xi) = \{j \mid \pi_j \geq \xi\}$ .

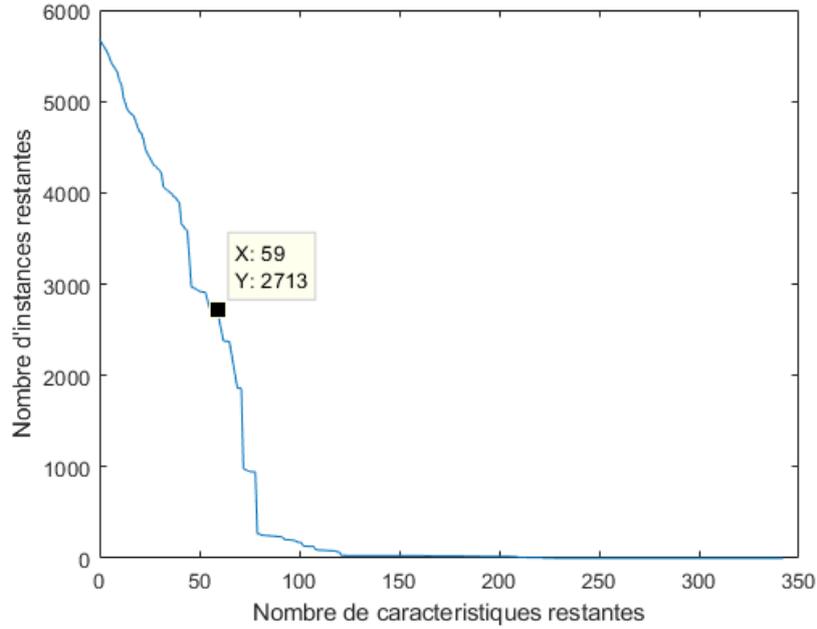
Après suppression des instances de la base  $\mathcal{H}$  n'ayant pas toutes les caractéristiques restantes disponibles, le rapport entre le nombre d'instances restants et celui d'origine représente la proportion des instances maintenues donnée par :

$$\omega_i(\xi) = \frac{1}{M} \left| \left\{ (\mathbf{x}_i, Y_i) \in \mathcal{H} \mid \prod_{j \in \Gamma(\xi)} A_{ij} = 1 \right\} \right|.$$

Notre objectif est de maximiser les proportions d'instances et de caractéristiques maintenues en faisant varier  $\xi$ . Ce problème d'optimisation qui concerne la maximisation de deux variables simultanément est intéressant à analyser sur un graphe dont les axes représentent les deux variables de sorties à maximiser.

Dans la FIGURE 6.1, les abscisses représentent le nombre de caractéristiques retenues et les ordonnées, le nombre d'instances retenues. Afin de maximiser les deux quantités simultanément, on s'intéresse au produit des deux nombres qu'on maximise en faisant varier  $\xi$ . Le point optimal trouvé se présente comme suit :

$$\text{— } \xi = 0.69$$


 FIGURE 6.1 – Graphe des proportions d’instances et de caractéristiques restantes en variant  $\xi$ .

- $\omega_f(0.69) = 0.173$
- $\omega_i(0.69) = 0.478$
- $d.\omega_f(0.69) = 59$
- $M.\omega_i(0.69) = 2713$

Pour cette valeur de  $\xi$ , la nouvelle base de données est établie :

$$\mathcal{H}^* = \left\{ (\tilde{\mathbf{x}}_i, Y_i) \mid \prod_{j \in \Gamma(\xi)} A_{ij} = 1 \right\}$$

avec  $\tilde{\mathbf{x}}_i$  le vecteur issu de la réduction de la dimension du vecteur  $\mathbf{x}_i$  en conservant uniquement les composantes indexées par  $\Gamma(\xi)$ .

Les méthodes RAKEL et CoLEX sont appliquées sur la base de données  $\mathcal{H}^*$ . La simulation est lancée pour plusieurs combinaisons de valeurs possibles pour  $k$  et  $n$  en veillant à ce que  $nk \geq q$ . Le TABLEAU 6.1 résume les performances de la méthode CoLex en la comparant à la méthode RAKEL pour plusieurs configurations de paramètres.

Dans le TABLEAU 6.1, on note une amélioration quasi-globale sur toutes les métriques et suivant toutes les configurations de paramètres. La métrique présentant le plus de dégradations pour notre méthode est le Recall où, sur les 18 configurations de paramètres simulées, 5 présentent une dégradation du Recall avec une moyenne de  $-6.46\%$  par rapport à la méthode RAKEL alors que pour les 13 autres, on note une amélioration de  $+15.42\%$  en moyenne.

### 6.3.2 Illustration des performances de la méthode proposée sur des bases de données réelles multi-label

Afin de valider l’efficacité de l’approche CoLEX proposée, nous considérons différentes bases de données multi-label bien connues que sont Yeast, Genbase, Scene, Medical, Enron et Emotions [175–177]. La première base de données *Yeast* contient des informations sur un ensemble de cellules de levure, le but étant de déterminer le site de localisation de chaque cellule. Dans la deuxième, appelée *Genbase*, il s’agit de déterminer la famille protéique de chaque protéine. Le troisième, *Scene*, fait référence à des scènes pour lesquelles plusieurs étiquettes peuvent être

TABLEAU 6.1 – Tableau comparatif des améliorations apportées par la méthode CoLex sur RAKEL.

$k$	$n$	Hloss ↓	One-Error ↓	Coverage ↓	SubAcc ↑	Recall ↑	P. moy. ↑	
2	13	0.128 ± 0.014	0.376 ± 0.096	18.162 ± 4.043	0.104 ± 0.177	0.484 ± 0.158	0.520 ± 0.124	
		0.118 ± 0.003	0.286 ± 0.014	14.402 ± 0.297	0.392 ± 0.013	0.551 ± 0.013	0.641 ± 0.010	
		-8.2%	-23.9%	-20.7%	+278.8%	+13.9%	+23.4%	
	15	0.130 ± 0.010	0.365 ± 0.090	20.180 ± 3.844	0.077 ± 0.160	0.452 ± 0.118	0.502 ± 0.092	
		0.118 ± 0.003	0.286 ± 0.014	14.402 ± 0.297	0.392 ± 0.013	0.551 ± 0.013	0.641 ± 0.010	
		-9.7%	-21.7%	-28.6%	+407.5%	+21.8%	+27.6%	
	20	0.122 ± 0.010	0.353 ± 0.092	16.417 ± 3.243	0.154 ± 0.195	0.542 ± 0.133	0.577 ± 0.099	
		0.118 ± 0.003	0.286 ± 0.014	14.401 ± 0.297	0.392 ± 0.013	0.551 ± 0.013	0.641 ± 0.010	
		-3.8%	-19.1%	-12.3%	+154.4%	+1.6%	+11.0%	
	30	0.124 ± 0.010	0.366 ± 0.097	17.694 ± 4.082	0.077 ± 0.157	0.548 ± 0.118	0.568 ± 0.099	
		0.118 ± 0.003	0.286 ± 0.014	14.402 ± 0.297	0.392 ± 0.013	0.551 ± 0.013	0.641 ± 0.010	
		-5.0%	-21.8%	-18.6%	+406.5%	+0.5%	+12.8%	
40	0.117 ± 0.005	0.345 ± 0.093	15.212 ± 2.457	0.104 ± 0.176	0.615 ± 0.086	0.630 ± 0.055		
	0.118 ± 0.003	0.286 ± 0.014	14.401 ± 0.297	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010		
	+0.3%	-17.1%	-5.3%	+276.9%	-10.5%	+1.7%		
3	9	0.131 ± 0.016	0.324 ± 0.076	18.624 ± 4.006	0.133 ± 0.189	0.423 ± 0.174	0.508 ± 0.134	
		0.118 ± 0.003	0.286 ± 0.014	14.411 ± 0.295	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-9.9%	-11.8%	-22.6%	+195.4%	+30.2%	+26.0%	
	15	0.126 ± 0.011	0.349 ± 0.091	18.772 ± 4.286	0.081 ± 0.164	0.516 ± 0.125	0.547 ± 0.105	
		0.118 ± 0.003	0.286 ± 0.013	14.406 ± 0.295	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-6.5%	-18.1%	-23.3%	+384.0%	+6.8%	+17.1%	
	20	0.120 ± 0.009	0.324 ± 0.081	16.554 ± 3.687	0.053 ± 0.137	0.585 ± 0.115	0.614 ± 0.090	
		0.118 ± 0.003	0.286 ± 0.014	14.404 ± 0.296	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-2.3%	-11.8%	-13.0%	+636.3%	-6.0%	+4.4%	
	25	0.118 ± 0.006	0.375 ± 0.098	15.447 ± 2.172	0.029 ± 0.101	0.626 ± 0.090	0.607 ± 0.062	
		0.118 ± 0.003	0.286 ± 0.014	14.403 ± 0.298	0.391 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-0.3%	-23.7%	-6.8%	+1271.3%	-12.1%	+5.5%	
4	7	0.129 ± 0.013	0.316 ± 0.070	18.833 ± 4.431	0.184 ± 0.201	0.430 ± 0.137	0.525 ± 0.125	
		0.118 ± 0.003	0.286 ± 0.014	14.417 ± 0.296	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-8.7%	-9.6%	-23.5%	+112.6%	+27.9%	+22.0%	
	10	0.129 ± 0.014	0.329 ± 0.071	18.612 ± 4.227	0.157 ± 0.199	0.449 ± 0.138	0.533 ± 0.126	
		0.118 ± 0.003	0.286 ± 0.014	14.412 ± 0.295	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-8.9%	-13.0%	-22.6%	+150.0%	+22.4%	+20.3%	
	20	0.120 ± 0.007	0.301 ± 0.056	15.997 ± 3.186	0.160 ± 0.197	0.567 ± 0.100	0.628 ± 0.061	
		0.118 ± 0.004	0.286 ± 0.014	14.415 ± 0.295	0.392 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-1.9%	-4.9%	-9.9%	+145.0%	-3.1%	+2.0%	
	5	6	0.130 ± 0.012	0.331 ± 0.076	19.132 ± 4.087	0.155 ± 0.196	0.426 ± 0.143	0.514 ± 0.116
			0.118 ± 0.004	0.287 ± 0.014	14.417 ± 0.295	0.391 ± 0.013	0.550 ± 0.013	0.641 ± 0.010
			-9.3%	-13.5%	-24.6%	+153.1%	+29.1%	+24.6%
10		0.126 ± 0.011	0.316 ± 0.069	17.591 ± 3.594	0.156 ± 0.196	0.497 ± 0.110	0.572 ± 0.097	
		0.118 ± 0.004	0.287 ± 0.014	14.418 ± 0.294	0.391 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-6.3%	-9.4%	-18.0%	+151.2%	+10.6%	+12.0%	
15		0.119 ± 0.004	0.300 ± 0.049	14.868 ± 1.773	0.339 ± 0.138	0.553 ± 0.036	0.645 ± 0.052	
		0.118 ± 0.004	0.286 ± 0.014	14.414 ± 0.293	0.391 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-0.7%	-4.5%	-3.1%	+15.4%	-0.6%	-0.7%	
10		3	0.126 ± 0.013	0.294 ± 0.019	17.144 ± 3.655	0.238 ± 0.200	0.453 ± 0.139	0.562 ± 0.116
			0.118 ± 0.004	0.287 ± 0.014	14.421 ± 0.294	0.391 ± 0.013	0.550 ± 0.013	0.640 ± 0.010
			-6.5%	-2.5%	-15.9%	+64.8%	+21.4%	+14.0%
	5	0.122 ± 0.008	0.287 ± 0.014	16.187 ± 3.008	0.291 ± 0.172	0.502 ± 0.084	0.602 ± 0.068	
		0.118 ± 0.004	0.287 ± 0.014	14.421 ± 0.294	0.391 ± 0.013	0.550 ± 0.013	0.640 ± 0.010	
		-3.1%	0.0%	-10.9%	+34.4%	+9.4%	+6.4%	
	10	0.120 ± 0.006	0.290 ± 0.015	15.224 ± 2.173	0.339 ± 0.138	0.525 ± 0.066	0.627 ± 0.052	
		0.118 ± 0.004	0.287 ± 0.014	14.411 ± 0.294	0.391 ± 0.013	0.550 ± 0.013	0.641 ± 0.010	
		-1.9%	-1.2%	-5.3%	+15.4%	+4.8%	+2.1%	

associées, comme par exemple, une même photographie peut contenir une montagne, la mer et la plaine. Pour la base de données *Medical*, il s'agit d'associer des labels à un texte libre de rapports médicaux. La base de données *Enron* vise à labeliser des email et enfin la base de données *Emotions* associe à chaque musique les émotions qu'elle provoque. Le TABLEAU 6.2 regroupe des caractéristiques sur les bases de données considérées.

Dans ce qui suit, la méthode COLEX proposée est d'abord illustrée à l'aide des deux premières bases de données. Elle est également comparée à la méthode RAKEL, pour évaluer ses performances. Pour les deux méthodes, l'algorithme SVM est utilisé pour la construction des  $n$  classifieurs  $h_u(\cdot)$ . L'efficacité de la stratégie d'échange est illustrée dans la section 6.3.3. Ensuite, les mesures de performance de la section 6.3.4 sont évaluées et comparées pour différentes valeurs des paramètres de réglage  $n$  et  $k$  sur toutes les bases de données. Enfin, la méthode

proposée est comparée à d'autres méthodes de l'état de l'art.

TABLEAU 6.2 – Caractéristiques des bases de données multi-label du monde réel (Voir Section 5.2.3).

	Nombre d'instances	Nombre de caractéristiques	Nombre de labels	LCard	LDens	LDiv	PLDiv
Yeast	2417	103	14	4.24	0.30	198	0.082
Genbase	662	1185	27	1.25	0.05	32	0.048
Scene	2407	294	6	1.07	0.18	15	0.006
Medical	978	1449	45	1.25	0.03	94	0.096
Enron	1702	1001	53	3.38	0.06	753	0.442
Emotions	593	72	6	1.87	0.31	27	0.046

Dans ce qui suit, nous présentons les performances de notre méthode et ses avantages par rapport à la méthode RAKEL standard et la méthode exacte pour maximiser la corrélation des labelsets. Nous donnons les résultats obtenus pour les deux ensembles de données à des valeurs  $k$  et  $n$  différentes. Dans le TABLEAU 6.3, plusieurs métriques de performances sont affichées pour les deux méthodes RAKEL et celle proposée pour la base de données Genbase avec  $n = 9$  labelsets et avec  $k = 3$  labels par labelset. Nous pouvons constater une amélioration significative de l'efficacité globale de la tâche de classification dans notre méthode. Cela montre que notre méthode tire parti des corrélations entre labels et réussit à créer un meilleur groupe de labelsets. Dans le TABLEAU 6.4, nous avons utilisé  $n = 18$  labelsets au lieu de 9 et nous remarquons toujours de meilleurs résultats en utilisant notre méthode. Au contraire, la méthode RAKEL donne de meilleures performances avec ce nombre plus élevé de labelsets. Cela peut s'expliquer par le fait que RAKEL a plus de chance d'attraper les corrélations entre labels aléatoirement lorsqu'il y a plus de labelsets.

TABLEAU 6.3 – Comparaison entre RAKEL et COLEX pour la base de données Genbase  $k = 3$  et  $n = 9$ .

Métriques	RAKEL	COLEX
↓ Hloss	0.0252 ± 0.0057	<b>0.0112 ± 0.0014</b>
↓ One-Error	0.4154 ± 0.1271	<b>0.1608 ± 0.0228</b>
↓ Coverage	3.8613 ± 1.0213	<b>1.8556 ± 0.2827</b>
↓ Rloss	0.0000 ± 0.0000	0.0000 ± 0.0000
↑ Subset Accuracy	0.4422 ± 0.1470	<b>0.8080 ± 0.0214</b>
↑ Recall	0.4769 ± 0.1460	<b>0.8142 ± 0.0225</b>
↑ Average Precision	0.6534 ± 0.1055	<b>0.8684 ± 0.0165</b>

TABLEAU 6.4 – Comparaison entre RAKEL et COLEX pour la base de données Genbase  $k = 3$  et  $n = 18$ .

Métriques	RAKEL	COLEX
↓ Hloss	0.0168 ± 0.0047	<b>0.0121 ± 0.0024</b>
↓ One-Error	0.2670 ± 0.1325	<b>0.1799 ± 0.0261</b>
↓ Coverage	2.6402 ± 0.5702	<b>2.0352 ± 0.4188</b>
↓ Rloss	0.0000 ± 0.0000	0.0000 ± 0.0000
↑ Subset Accuracy	0.6422 ± 0.1269	<b>0.7943 ± 0.0293</b>
↑ Recall	0.6727 ± 0.1291	<b>0.8035 ± 0.0275</b>
↑ Average Precision	0.7777 ± 0.1046	<b>0.8545 ± 0.0227</b>

Dans le TABLEAU 6.5, nous appliquons notre méthode sur la base de données Yeast en choisissant  $k = 2$ ,  $n = 7$  et seule le One-Error est mieux évaluée pour la méthode RAKEL. Dans un deuxième essai, nous avons testé notre méthode pour  $k = 5$ ,  $n = 3$  et nous notons à nouveau une amélioration significative de toutes les métriques d'évaluation. Nous notons que la performance de notre méthode est peu affectée par la variation de  $k$ . La méthode proposée remédie donc aux problèmes

liées à la dépendance au nombre de modèles  $n$  car elle a tendance à gérer les corrélations même avec un nombre réduit de modèle.

TABLEAU 6.5 – Comparaison entre RAKEL et COLEX pour la base de données Yeast  $k = 2$  et  $n = 7$ .

Métriques	RAKEL	COLEX
↓ Hloss	0.2208 ± 0.0211	<b>0.1910 ± 0.0050</b>
↓ One-Error	0.2780 ± 0.0282	<b>0.2432 ± 0.0130</b>
↓ Coverage	8.2969 ± 1.9637	<b>6.1076 ± 0.0957</b>
↓ Rloss	<b>0.0458 ± 0.0093</b>	0.0488 ± 0.0031
↑ Subset Accuracy	0.0398 ± 0.0535	<b>0.2063 ± 0.0118</b>
↑ Recall	0.4257 ± 0.1019	<b>0.5696 ± 0.0093</b>
↑ Average Precision	0.6825 ± 0.0548	<b>0.7533 ± 0.0075</b>

TABLEAU 6.6 – Comparaison entre RAKEL et COLEX pour la base de données Yeast  $k = 5$  et  $n = 3$ .

Métriques	RAKEL	COLEX
↓ Hloss	0.2261 ± 0.0243	<b>0.1927 ± 0.0049</b>
↓ One-Error	0.3280 ± 0.0371	<b>0.3088 ± 0.0279</b>
↓ Coverage	8.2366 ± 2.0022	<b>6.2319 ± 0.0651</b>
↓ Rloss	0.0649 ± 0.0101	<b>0.0610 ± 0.0049</b>
↑ Subset Accuracy	0.0738 ± 0.0691	<b>0.2320 ± 0.0120</b>
↑ Recall	0.4793 ± 0.1013	<b>0.6143 ± 0.0201</b>
↑ Average Precision	0.6728 ± 0.0598	<b>0.7415 ± 0.0085</b>

### 6.3.3 Efficacité de la stratégie d'échange

Afin de montrer l'efficacité de la stratégie d'échange dans la recherche d'ensembles de labels corrélés, nous exécutons la méthode exacte (force brute) pour trouver le groupe d'ensembles de labels maximisant la corrélation totale  $\zeta$ . Nous pouvons montrer que le nombre de groupes de labels possibles est

$$\frac{(nk)!}{n!(k!)^n \left( \left\lfloor \frac{nk}{q} \right\rfloor ! \right)^{q - \text{mod}(nk, q)} \left( \left\lceil \frac{nk}{q} \right\rceil ! \right)^{\text{mod}(nk, q)}}$$

où ! est le factoriel de  $\cdot$  et  $\text{mod}(\cdot, \cdot)$  est la fonction modulo. Dans le TABLEAU 6.7, nous montrons le nombre de possibilités pour la méthode exacte, le nombre d'échanges pour la méthode proposée, le temps de calcul et la valeur  $\zeta$  pour l'ensemble de données Yeast à différentes valeurs de  $n$  et  $k$ .

TABLEAU 6.7 – Comparaison de la méthode exacte à la méthode proposée avec la stratégie d'échange.

	Méthode exacte		Stratégie d'échange	
	k=2	k=3	k=2	k=3
	n=7	n=5	n=7	n=5
Possibilités / Échanges	135135	700700	2	5
Temps (s)	200	2819	0.036	0.020
$\zeta$	4.0483	5.6385	4.0483	5.6385

On note que la stratégie proposée a convergé dans les deux cas vers sa solution optimale avec 2 cycles (2 échanges pour  $k = 2$   $n = 7$  et 5 échanges pour  $k = 3$   $n = 5$ ) et la durée de l'optimisation est considérablement réduite.

### 6.3.4 Impact des paramètres $k$ et $n$

Dans la FIGURE 6.2, nous montrons 6 graphiques différents représentant chacun une métrique d'évaluation pour la base de données Genbase et pour  $k = 3$  par rapport au nombre de labelsets  $n$ . Ces graphiques montrent une amélioration significative des performances globales de classification lors de l'utilisation de notre méthode au lieu de RAKEL. Nous remarquons sur cette figure que la méthode RAKEL donne de meilleures performances lorsque  $n$  augmente, expliquant un besoin d'un grand nombre de labelsets pour atteindre les mêmes performances que notre méthode. Un autre fait est que notre méthode est insensible aux variations de  $n$ . Par conséquent, elle conserve les mêmes performances pour tous les nombres de labelsets, ce qui nous permet d'utiliser le nombre minimum de labelsets pour réduire le coût en temps de calcul sans sacrifier la valeur des performances. On peut également noter que même avec deux fois plus de labelsets, notre méthode donne quand même de meilleures performances que la méthode RAKEL.

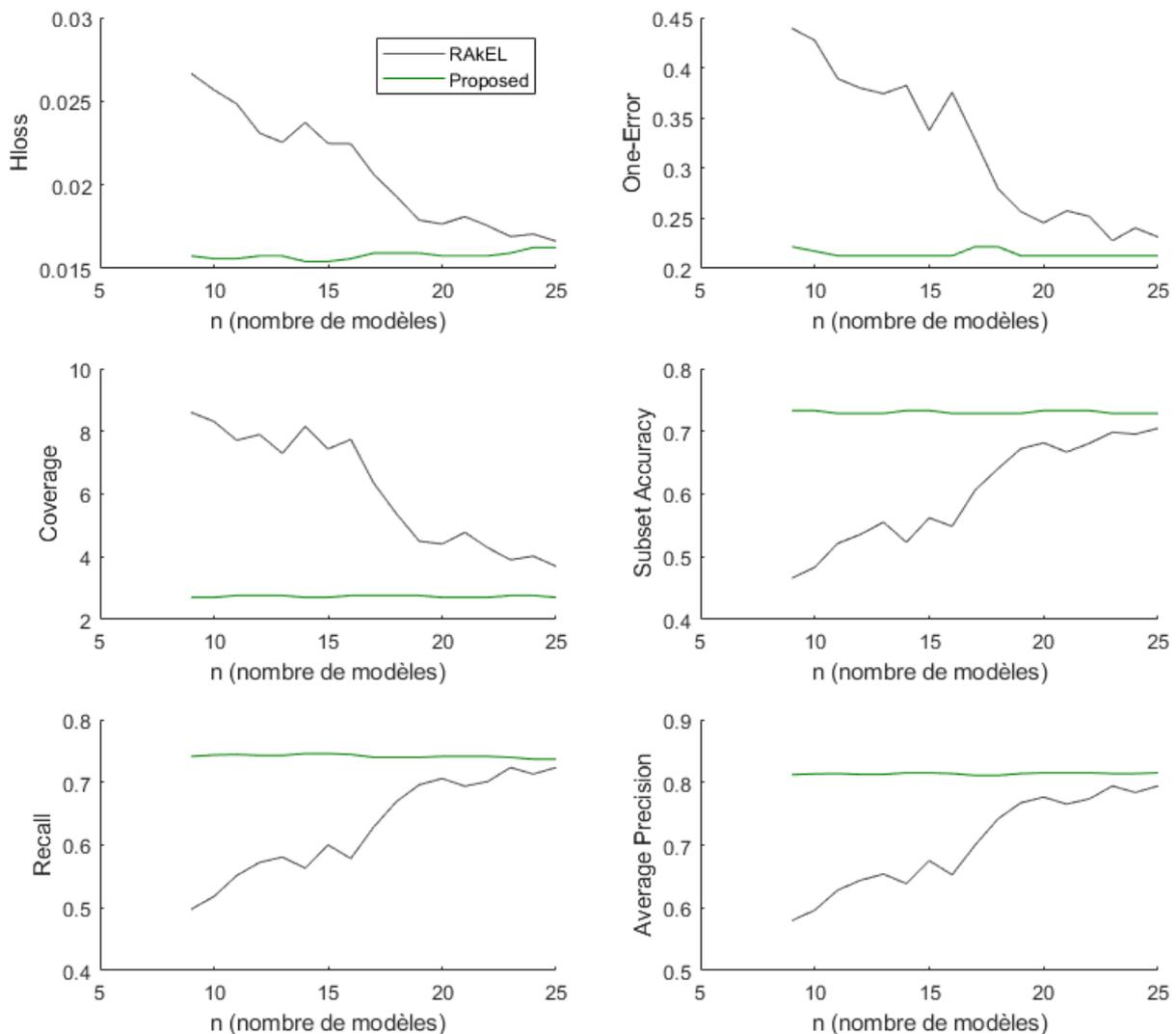


FIGURE 6.2 – Comparaison entre RAKEL et COLEX pour différentes valeurs de  $n$  avec  $k = 3$  pour la base de données Genbase.

Dans la FIGURE 6.3, nous faisons le même processus que précédemment mais pour  $k = 10$ , ce qui diminue l'effet de notre méthode puisque nous n'avons que 14 labels dans cette base de

données et les labels corrélés sont plus susceptibles d’être capturés dans les mêmes labelsets sans l’intervention de la méthode proposée. Cependant, nous pouvons remarquer sur cette figure que notre méthode fonctionne toujours mieux que RAKEL jusqu’à  $n = 8$  après quoi les performances des deux méthodes convergent presque au même niveau.

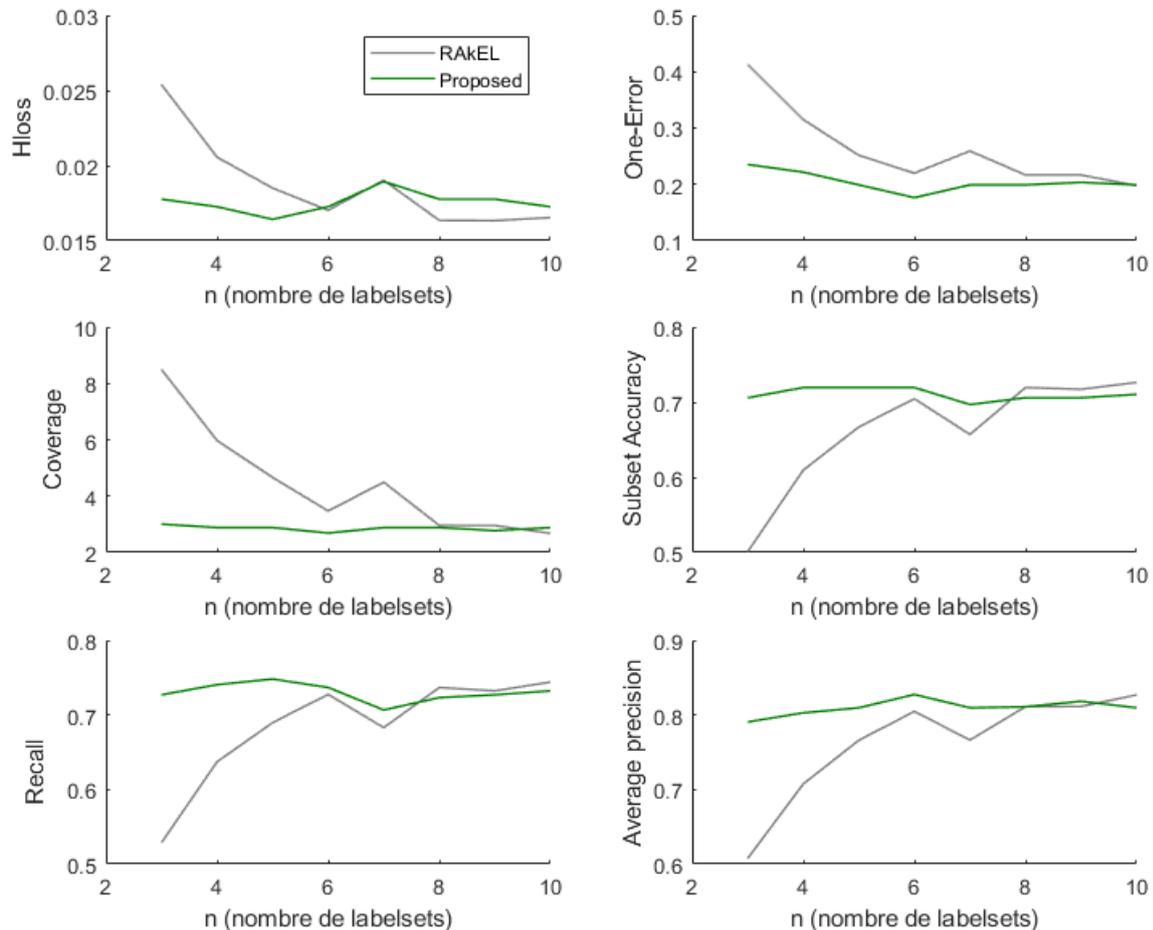


FIGURE 6.3 – Comparaison entre RAKEL et COLEX pour différentes valeurs de  $n$  avec  $k = 10$  pour la base de données Genbase.

Nous faisons une étude similaire sur la base de données *Yeast*, qui a plus de labels (27) que la base de données Genbase (14), ce qui permet de mieux mettre en valeur la contribution de notre méthode. La FIGURE 6.4 montre les performances de notre méthode pour  $k = 2$  en faisant varier  $n$ . Nous pouvons remarquer que RAKEL a besoin de plus de 15 labelsets pour atteindre les mêmes performances de notre méthode à  $n = 7$  labelsets. Dans un autre essai, nous avons fixé  $n$  et fait varier  $k$  afin d’étudier l’impact de  $k$  sur la classification. Les résultats de cette exécution sont présentés dans la FIGURE 6.5 où nous pouvons noter que les deux méthodes convergent vers le même niveau de performance après  $k = 5$ .

### 6.3.5 Comparaison avec d’autres méthodes

Dans cette section, la méthode CoLex est comparée à plusieurs méthodes de classification multi-label. Le TABLEAU 6.8 fait référence à toutes les méthodes utilisées pour la comparaison avec leurs abréviations correspondantes. EME est un ensemble évolutif multi-label qui utilise un algorithme génétique pour optimiser le choix des labelsets [178]. EML est un ensemble de classifieurs multi-label utilisant différentes méthodes pour combiner les sorties de chaque classifieur

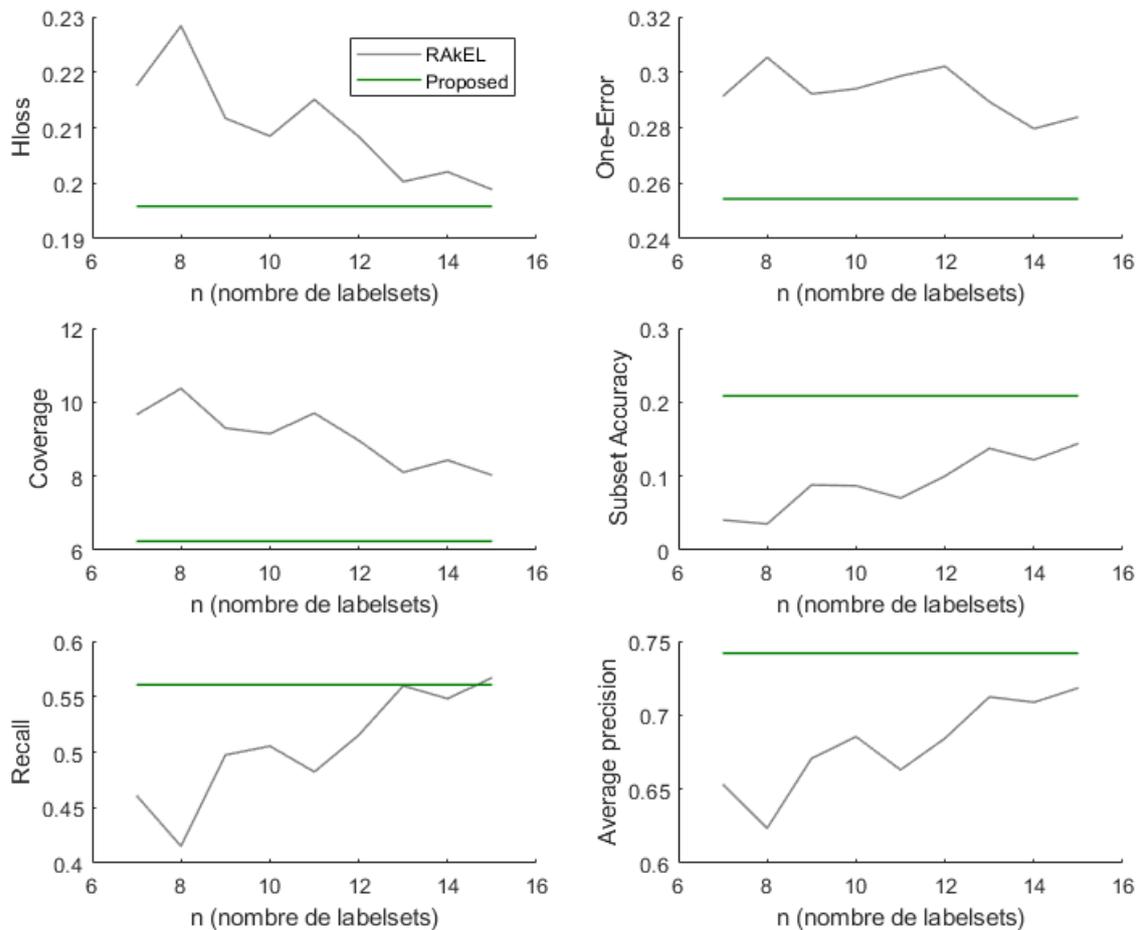


FIGURE 6.4 – Comparaison entre RAKEL et COLEX pour différentes valeurs de  $n$  avec  $k = 2$  pour la base de données Yeast.

[179]. ECC génère une prédiction en sommant les sorties de différents ordres de chaînes aléatoires et effectue l'apprentissage classique des classifieurs en chaîne, c'est-à-dire que chaque label est entraîné seul avec les prédictions des labels précédents utilisés comme caractéristiques supplémentaires [147]. CLR considère la classification binaire par paires où chaque paire de labels est confrontée aux autres par un classifieur approprié [144]. IBLR considère les labels des instances voisines comme des caractéristiques d'échantillons de test et réduit ainsi l'apprentissage basé sur les instances à la régression logistique [180]. RF-PCT applique une forêt aléatoire sur les arbres de regroupement prédictifs [181]. HOMER travaille sur une hiérarchie de classificateurs multi-label [182]. TSA est une architecture en deux étapes utilisant deux couches d'apprentissage différentes telles que le *BinaryRelevance* et CLR [151]. De plus, une méthode d'ensemble d'arbres multi-label pour la classification multi-label appelée ML-Forest [183] est utilisée pour comparer les méthodes de l'état de l'art présentées dans cet article. Les comparaisons sont effectuées à l'aide de plusieurs mesures d'évaluation sur différentes bases de données. Des statistiques sur des bases de données utilisées sont données dans le TABLEAU 6.2.

Dans le TABLEAU 6.9, CoLex est d'abord comparé à EME en utilisant les bases de données Emotions, Yeast, Genebase et Medical sur cinq mesures d'évaluation, à savoir le *HammingLoss* (Hloss), le *Subset Accuracy* (SubAcc), la macro-précision (MaP), le macro-Recall (MaR) et la macro-spécificité (MaS). Les résultats de ce tableau montrent de meilleures performances pour CoLex. Le tableau donne également, dans sa dernière colonne, le classement moyen de CoLex et EME sur toutes les métriques d'évaluation. Notez qu'ici, le rang ne peut être que 1 ou 2 et le clas-

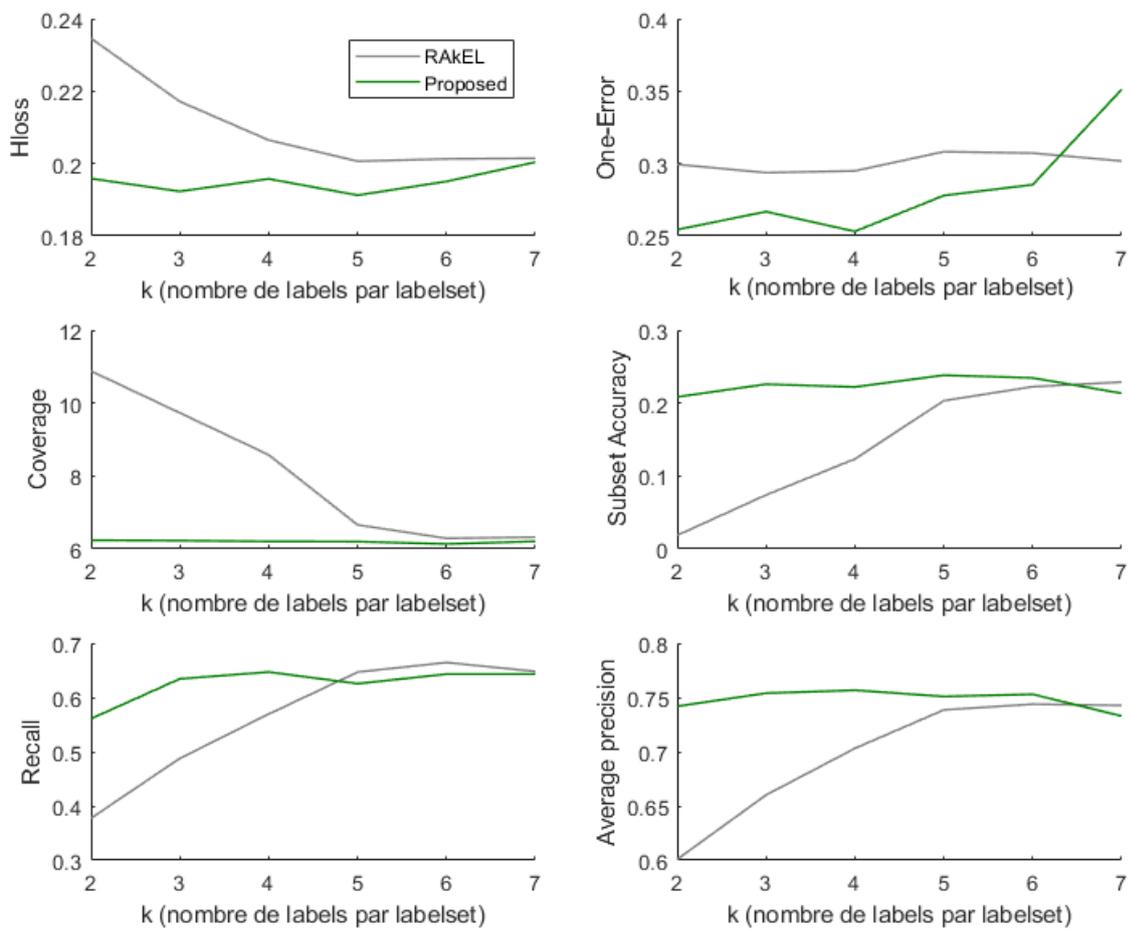

 FIGURE 6.5 – Comparaison entre RAKEL et COLEX pour différentes valeurs de k avec  $n = 7$  pour la base de données Yeast.

TABLEAU 6.8 – Méthodes multi-label utilisées dans la partie expérimentale.

Méthode	Abréviation
Evolutionary multi-label Ensemble	EME
Ensemble of multi-label classifiers with mean rule	EML <sub>A</sub>
Ensemble of multi-label classifiers with threshold selection	EML <sub>T</sub>
Ensemble of multi-label classifiers with static weighting	EML <sub>S</sub>
Ensemble of multi-label classifiers with Shepard's rule	EML <sub>P</sub>
Ensemble of multi-label classifiers with Dudani's rule	EML <sub>D</sub>
Ensemble of Classifier Chains	ECC
Calibrated Label Ranking	CLR
Instance Based Logistic Regression	IBLR
Random Forest of Predictive Clustering Trees	RF-PCT
Hierarchy Of multi-label classifiers	HOMER
Two Stage Architecture	TSA
multi-label Forest	ML-F
Correlation based labelsets optimized by exchange strategy (proposed method)	CoLEx

TABLEAU 6.9 – Comparaison des performances entre les méthodes EME et CoLEx.

		HLoss ↓	SubAcc ↑	MaP ↑	MaR ↑	MaS ↑	ClassM
Emotions	EME	0.22	0.248	<b>0.657</b>	0.592	<b>0.858</b>	1.6
	CoLEx	<b>0.21</b>	<b>0.304</b>	0.646	<b>0.704</b>	0.822	<b>1.4</b>
Yeast	EME	0.21	0.137	<b>0.51</b>	0.361	<b>0.803</b>	1.6
	CoLEx	<b>0.2</b>	<b>0.212</b>	-	<b>0.363</b>	0.793	<b>1.4</b>
Genbase	EME	0.001	0.966	0.929	0.934	1	2
	CoLEx	<b>0.00064</b>	<b>0.984</b>	<b>0.997</b>	<b>0.938</b>	1	<b>1</b>
Medical	EME	<b>0.01</b>	0.649	0.651	<b>0.65</b>	0.995	1.6
	CoLEx	0.011	<b>0.654</b>	<b>0.817</b>	0.425	<b>0.996</b>	<b>1.4</b>

sement moyen est alors compris entre 1 et 2. Le classement moyen (ClassM) sur toutes les bases de données est égal à 1.3 pour CoLEx, alors qu’il est de 1.65 pour EME, ce qui valide l’efficacité de CoLEx.

Dans le TABLEAU 6.10 ( suite dans le TABLEAU 6.11), les autres méthodes du TABLEAU 6.8 sont comparées à CoLEx, sur les bases de données Yeast, Emotions, Medical, Scene et Enron, en utilisant plusieurs métriques d’évaluation. Nous pouvons remarquer que, à l’exception du *HammingLoss*, toutes les métriques d’évaluation sont en faveur de CoLEx. Le tableau montre également le classement moyen de chaque méthode sur les différentes métriques. Ici, avec 10 méthodes comparées, le rang varie entre de 1 à 10. CoLEx prend le premier rang moyen sur toutes les autres méthodes pour l’ensemble de données utilisé. Ceci permet de montrer que notre méthode peut fonctionner dans différentes circonstances reliées au nombre initial d’étiquettes et les corrélations entre elles.

TABLEAU 6.10 – Comparaison entre CoLEx et d’autres méthodes de l’état de l’art (1).

		Hloss ↓	One-error ↓	Coverage ↓	Rloss ↓	Prém ↑	ClassM
Yeast	EML <sub>A</sub>	<b>0.193(1)</b>	0.227(5)	6.241(5)	0.162(3)	0.768(3)	3.4
	EML <sub>T</sub>	0.199(4)	0.227(5)	6.241(5)	0.162(3)	0.768(3)	4.0
	EML <sub>S</sub>	0.198(3)	0.226(4)	6.236(4)	0.162(3)	0.768(3)	3.4
	EML <sub>P</sub>	0.197(2)	0.224(2)	6.173(2)	0.161(2)	0.769(2)	2.0
	EML <sub>D</sub>	0.198(3)	0.225(3)	6.19(3)	0.162(3)	0.768(3)	3.0
	ECC	0.216(10)	0.259(12)	7.086(12)	0.214(9)	0.735(8)	10.2
	CLR	0.211(8)	0.239(8)	6.632(10)	0.181(7)	0.75(7)	8.0
	IBLR	0.198(3)	0.237(7)	6.331(7)	0.17(4)	0.759(4)	5.0
	RF-PCT	0.207(7)	0.244(10)	6.244(6)	0.171(5)	0.754(5)	6.6
	HOMER	0.212(9)	0.25(11)	6.96(11)	0.204(8)	0.734(9)	9.6
	TSA	0.206(6)	0.236(6)	6.54(9)	0.176(6)	0.752(6)	6.6
	ML-F	0.199(4)	0.241(9)	6.489(8)	0.176(6)	0.754(5)	6.4
	CoLEx	0.2(5)	<b>0.133(1)</b>	<b>2.552(1)</b>	<b>0.069(1)</b>	<b>0.867(1)</b>	<b>1.8</b>
	Emotions	EML <sub>A</sub>	<b>0.185(1)</b>	0.249(4)	1.68(2)	0.143(2)	0.818(2)
EML <sub>T</sub>		0.199(5)	0.249(4)	1.68(2)	0.143(2)	0.818(2)	3.0
EML <sub>S</sub>		0.2(6)	0.248(3)	1.68(2)	0.143(2)	0.818(2)	3.0
EML <sub>P</sub>		0.199(5)	0.251(5)	1.68(2)	0.144(3)	0.817(3)	3.6
EML <sub>D</sub>		0.198(4)	0.225(2)	6.19(10)	0.162(5)	0.768(10)	6.2
ECC		0.204(8)	0.275(7)	1.9(5)	0.181(9)	0.795(6)	7.0
CLR		0.205(9)	0.271(6)	1.73(3)	0.154(4)	0.807(4)	5.2
IBLR		0.201(7)	0.279(8)	1.77(4)	0.164(6)	0.798(5)	6.0
RF-PCT		0.197(3)	0.297(9)	1.921(6)	0.171(7)	0.794(7)	6.4
HOMER		0.211(10)	0.352(11)	2.663(9)	0.311(11)	0.709(11)	10.4
TSA		0.263(12)	0.302(10)	1.951(7)	0.176(8)	0.783(8)	9.0
ML-F		0.256(11)	0.302(10)	2.055(8)	0.203(10)	0.769(9)	9.6
CoLEx		0.196(2)	<b>0.157(1)</b>	<b>0.83(1)</b>	<b>0.078(1)</b>	<b>0.888(1)</b>	<b>1.2</b>

TABLEAU 6.11 – Comparaison entre CoLEx et d'autres méthodes de l'état de l'art (2).

		<b>Hloss</b> ↓	<b>One-error</b> ↓	<b>Coverage</b> ↓	<b>Rloss</b> ↓	<b>Prém</b> ↑	<b>ClassM</b>
<b>Medical</b>	EML <sub>A</sub>	<b>0.01(1)</b>	0.151(4)	1.54(4)	0.023(2)	0.887(3)	2.8
	EML <sub>T</sub>	<b>0.01(1)</b>	0.151(4)	1.54(4)	0.023(2)	0.887(3)	2.8
	EML <sub>S</sub>	0.011(2)	0.15(2)	1.53(3)	0.023(2)	0.888(2)	2.2
	EML <sub>P</sub>	0.012(3)	0.151(4)	1.52(2)	0.023(2)	0.888(2)	2.6
	EML <sub>D</sub>	0.2(7)	0.254(10)	1.694(5)	0.147(10)	0.815(9)	8.2
	ECC	0.012(3)	0.171(5)	2.92(9)	0.048(6)	0.862(4)	5.4
	CLR	0.014(5)	0.208(8)	1.77(6)	0.027(3)	0.849(6)	5.6
	IBLR	0.026(6)	0.414(11)	4.58(11)	0.084(8)	0.686(11)	9.4
	RF-PCT	0.013(4)	0.202(7)	2.146(7)	0.034(4)	0.839(8)	6.0
	HOMER	0.014(5)	0.247(9)	5.992(12)	0.103(9)	0.758(10)	9.0
	TSA	0.013(4)	0.172(6)	2.375(8)	0.037(5)	0.857(5)	5.6
	ML-F	0.012(3)	0.1504(3)	3.2(10)	0.051(7)	0.844(7)	6.0
CoLEx	<b>0.01(1)</b>	<b>0.073(1)</b>	<b>0.202(1)</b>	<b>0.002(1)</b>	<b>0.983(1)</b>	<b>1.0</b>	
<b>Scene</b>	EML <sub>A</sub>	<b>0.08(1)</b>	0.206(3)	0.401(2)	0.064(2)	0.881(2)	2.0
	EML <sub>T</sub>	0.088(2)	0.206(3)	0.401(2)	0.064(2)	0.881(2)	2.2
	EML <sub>S</sub>	0.088(2)	0.206(3)	0.401(2)	0.064(2)	0.881(2)	2.2
	EML <sub>P</sub>	0.088(2)	0.205(2)	0.403(3)	0.064(2)	0.881(2)	2.2
	EML <sub>D</sub>	0.089(3)	0.208(4)	0.418(4)	0.067(3)	0.878(3)	3.4
	ECC	0.098(6)	0.253(8)	0.56(9)	0.094(8)	0.845(8)	7.8
	CLR	0.113(8)	0.255(9)	0.501(7)	0.083(6)	0.85(7)	7.4
	IBLR	0.089(3)	0.235(5)	0.491(6)	0.081(5)	0.86(4)	4.6
	RF-PCT	0.097(5)	0.248(7)	0.528(8)	0.086(7)	0.853(6)	6.6
	HOMER	0.118(9)	0.329(10)	1.12(11)	0.2(10)	0.762(10)	10.0
	TSA	0.104(7)	0.237(6)	0.49(5)	0.078(4)	0.859(5)	5.4
	ML-F	0.097(5)	0.248(7)	0.681(10)	0.115(9)	0.835(9)	8.0
CoLEx	0.091(4)	<b>0.183(1)</b>	<b>0.242(1)</b>	<b>0.039(1)</b>	<b>0.958(1)</b>	<b>1.6</b>	
<b>Enron</b>	EML <sub>A</sub>	<b>0.048(1)</b>	0.24(3)	12.04(3)	0.079(2)	0.686(2)	2.0
	EML <sub>T</sub>	0.052(4)	0.24(3)	12.04(3)	0.079(2)	0.686(2)	2.8
	EML <sub>S</sub>	0.051(3)	0.239(2)	12.03(2)	0.079(2)	0.686(2)	2.2
	EML <sub>P</sub>	0.054(5)	0.25(5)	12.29(4)	0.081(3)	0.677(3)	4.0
	EML <sub>D</sub>	0.056(6)	0.247(4)	12.58(5)	0.084(4)	0.675(4)	4.6
	ECC	0.061(7)	0.329(7)	20.34(8)	0.15(7)	0.614(6)	7.0
	CLR	0.065(9)	0.323(6)	12.68(6)	0.089(5)	0.63(5)	6.2
	IBLR	0.064(8)	0.469(8)	16.05(7)	0.12(6)	0.564(7)	7.2
	CoLEx	0.05(2)	<b>0.111(1)</b>	<b>2.121(1)</b>	<b>0.011(1)</b>	<b>0.928(1)</b>	<b>1.2</b>

## 6.4 Conclusion

Dans ce chapitre, une méthode de transformation, appelée CoLEx pour "Correlation based Labelsets by Exchange strategy", et se basant sur l'optimisation du choix des  $k$ -labelsets en considérant les corrélations existantes entre les labels, a été proposée. Un indicateur de corrélation a été définie, ainsi qu'une procédure d'optimisation qui permet de choisir la meilleure combinaison des  $n$   $k$ -labelsets. La méthode utilise la stratégie d'échange en maximisant les corrélations entre les labels d'un même  $k$ -labelset. Cette stratégie fonctionne en échangeant des labels entre les différents  $k$ -labelsets. Ensuite, on a entraîné un classifieur pour chaque  $k$ -labelset en se servant de la méthode LP. On a utilisé tous ces classifieurs entraînés, en plus d'un système de vote pour donner une décision sur chacun des  $q$  labels. Des simulations sur la base de données médicale du chapitre précédant ainsi que sur des bases de données multi-label du monde réel ont montré l'efficacité de notre méthode par rapport à la méthode RAKEL et à d'autres méthodes de la littérature.

# Conclusion générale et perspectives

Cette thèse s'est intéressée au problème de classification mono- et multi-label dans un contexte médical lié au suivi des maladies cardio-vasculaires (MCV). L'objectif de cette thèse était de proposer des approches intelligentes qui anticipent l'évolution des MCV chez des patients dans leur lieu de vie quotidien sans imposer des contraintes gênantes. Ces approches se basent sur des mesures de signaux physiologiques en temps réel pour les analyser et en sortir une décision. Dans cette thèse, nous avons opté pour l'utilisation de deux formalismes différents pour répondre à ce problème. Le premier adopte la notion de classification mono-label qui vise la prédiction d'épisode de fibrillation atriale (FA) basée sur la théorie de fonctions de croyance, et le second adopte la notion multi-label pour pouvoir effectuer un diagnostic multi-pathologique. Le défi scientifique à relever dans cette thèse était la proposition d'approches de classification permettant de répondre aux contraintes de données manquantes par moment, hétérogènes, imprécises et indirectement liées au phénomène observé.

Le manuscrit commence par une introduction aux MCV démontrant l'ampleur de leur impact sur la santé et mettant en évidence les enjeux des travaux de recherche de la thèse tout en introduisant le matériel utilisé dans ces travaux qui demeure principalement la base de données médicales MIMIC III. Ensuite, Le manuscrit introduit la théorie des fonctions de croyance, étant l'outil technique de base de la majorité des travaux développés. Quatre contributions sont présentées par la suite. La première porte sur une approche de classification visant à prédire un épisode de la FA. Cette méthode est basée sur la théorie des fonctions de croyance, qui est une théorie permettant de modéliser l'information d'une manière différente de la théorie de probabilités en tenant en compte des spécificités de l'application relatives au manque d'information. La mise en œuvre de la méthode développée a donné, après optimisation des paramètres et sur un ensemble de patients test et un groupe de contrôle, une précision de 79.3%, une sensibilité de 74.2%, une spécificité de 83.7% ainsi qu'une portée de prédiction d'environ 2 heures et 10 minutes avant l'occurrence de l'épisode. On a apporté par la suite des développements améliorant les performances de cette méthode en proposant une nouvelle modélisation de l'information et une option de rejet supplémentaire. Le manuscrit introduit aussi une approche originale pour la fusion des données se traduisant par une méthode de combinaison plus flexible que la règle de YAGER. On a montré qu'on peut atteindre avec ces améliorations des précisions allant jusqu'à 95.8% pour un taux de rejet d'environ 38.21%. La méthode est évaluée sur la base de données MIMIC III.

Par ailleurs, on a introduit la notion de classification multi-label avant de l'appliquer au diagnostic multi-pathologique. Les signaux physiologiques de la base de données médicale MIMIC III ont été utilisés pour faire le diagnostic des éventuelles complications cardiaques relatives à 55 maladies cardio-vasculaires. On a aussi introduit une approche multi-label basée sur la théorie des fonctions de croyance qu'on a validée grâce aux données extraites de la base médicale MIMIC III. La méthode a été comparée à des méthodes multi-label classiques basées sur une transformation de plusieurs méthodes de classification mono-label. Il s'est avéré que la méthode proposée est meilleure que toutes les autres méthodes en utilisant plusieurs critères d'évaluation. Ceci montre la capacité de la méthode proposée à traiter les données manquantes. Enfin, on a proposé une méthode originale appelée COLEX pour la classification multi-label puis on l'a appliquée à notre problème de diagnostic. Cette méthode s'inscrit dans le cadre des méthodes de transformation, comme elle transforme le problème de classification multi-label en plusieurs sous-problèmes mono-label. Elle est une généralisation de la méthode RAKEL qui se base sur l'optimisation du

choix des  $k$ -labelsets en considérant les corrélations existantes entre les labels. En effet, la méthode permet un réglage automatique de la répartition des  $k$ -labelsets sur les  $n$  modèles de la méthode RAKEL. Les résultats obtenus sur des données du monde réel ont montrées l'efficacité de notre méthode par rapport à la méthode RAKEL classique et à d'autres méthodes de la littérature.

Des perspectives sont envisageables pour la suite des travaux de recherche sur ce sujet :

- D'abord, les approches proposées extraient des caractéristiques statistiques simples des signaux physiologiques. Dans des travaux futurs, il est envisageable d'extraire des caractéristiques en utilisant d'autres approches linéaires ou non linéaires, telles que celles basées sur l'analyse par composantes principales.
- Les caractéristiques extraites de la base d'apprentissage sont modélisées individuellement par des distributions de probabilité normales. Il serait intéressant de proposer des distributions de probabilité plus adéquates, adaptées à chaque caractéristique. On peut par exemple se servir de l'estimation à noyau des fonctions de densité ou de concevoir une sélection automatique parmi plusieurs distributions de probabilité standards bien connues pour chaque caractéristique indépendamment. Il est intéressant aussi d'envisager une modélisation multi-dimensionnelle par signal, en considérant toutes les caractéristiques d'un même signal comme un vecteur. Cela permettra de profiter de la dépendance entre les caractéristiques et préservera la robustesse des méthodes aux signaux manquants.
- Une autre perspective consiste à adapter l'approche de la théorie des fonctions de croyance à la classification multi-classe permettant ainsi de résoudre des problèmes de classification multi-label en appliquant une méthode de transformation en plusieurs problèmes de classification multi-classe. Cette extension doit prendre en considération la conception d'une modélisation adéquate pour chaque sous-ensemble de classes.
- D'autre part, la théorie des fonctions de croyance permet d'attribuer des poids à des sous-ensembles de l'ensemble des labels possibles. Cela laisse présager que cette théorie pourrait être convenable à la conception d'une méthode de classification multi-label par adaptation, agissant directement sur les données multi-label.
- Il est aussi possible d'intégrer dans le modèle de nouvelles sources d'information qui s'appuient sur des données statistiques de la maladie (prévalence, incidence, symptômes connus, etc.) ou catégoriques propres au patient (séjours précédents, fiche de soins, antécédents médicaux, etc.).
- De nos jours, grâce à la domotique et l'affluence des objets connectés dans la vie quotidienne, il est possible de collecter de grands volumes de données liées d'une manière directe ou indirecte à l'état de santé d'un sujet. En présence de ces données, il est intéressant d'utiliser l'apprentissage profond (deep learning) à plusieurs niveaux comme au niveau de l'extraction de caractéristiques, le remplissage des données manquantes ou la classification elle-même. En fonction du volume de données, il est aussi possible que plusieurs de ces tâches soient accomplies simultanément par un réseau de neurones utilisant l'apprentissage profond. Il serait alors intéressant d'étudier la transformation des signaux en images cohérentes pour tirer profit de ces réseaux.
- Les approches proposées permettent le suivi d'un patient quelconque grâce à un classifieur pré-entraîné. Il est judicieux de calibrer le système aux informations du patient, en exploitant le transfert d'apprentissage.
- L'apprentissage proposé utilise la base de données MIMIC III, formée de signaux collectés en unités de soins intensifs. Cela nous a permis de valider nos approches grâce à des signaux réels. Toutefois, ces signaux ne sont pas représentatifs du cadre applicatif auquel est dédié le système, à savoir la surveillance à domicile. Une collaboration avec une équipe de cardiologues est envisagée pour l'accès à des données plus réalistes.
- Enfin, les développements menés à ce jour se situent au niveau méthodologique. D'un point de vue technologique, d'autres perspectives visent à développer un système de surveillance

à domicile respectant les contraintes pratiques d'acceptabilité, éthiques et économiques. Il est nécessaire que le produit soit accepté et validé par la communauté visée.



# Bibliographie

- [1] Ameli.fr, “Qu’est-ce que le risque cardiovasculaire?.” [vii, 4](#)
- [2] A. Timmis, N. Townsend, C. P. Gale, A. Torbica, M. Lettino, S. E. Petersen, E. A. Mossialos, A. P. Maggioni, D. Kazakiewicz, H. T. May, *et al.*, “European society of cardiology : cardiovascular disease statistics 2019,” *European Heart Journal*, vol. 41, no. 1, pp. 12–85, 2020. [vii](#), [viii](#), [ix](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [19](#), [25](#), [26](#), [27](#), [28](#), [29](#), [30](#)
- [3] G. B. of Disease Collaborative Network, “Global burden of disease study 2016 (gbd 2016) results,” 2017. [vii, 21](#)
- [4] A. Freeman, S. Tyrovolas, A. Koyanagi, S. Chatterji, M. Leonardi, J. L. Ayuso-Mateos, B. Tobiasz-Adamczyk, S. Koskinen, C. Rummel-Kluge, and J. M. Haro, “The role of socio-economic status in depression : results from the courage (aging survey in europe),” *BMC public health*, vol. 16, no. 1, p. 1098, 2016. [4](#)
- [5] A. Da Costa, C. Roméyer-Bouchard, L. Bisch, L. Khris, and K. Isaz, “Fibrillation atriale : enjeux épidémiologiques, définition, nosologie, médico-économie,” in *Annales de cardiologie et d'angiologie*, vol. 58, pp. S3–S5, Elsevier, 2009. [5](#)
- [6] D. Kotecha and P. Kirchhof, “What’s new in the 2016 esc guidelines on atrial fibrillation?,” *European heart journal*, vol. 37, no. 38, pp. 2851–2852, 2016. [6](#)
- [7] Z. Goldberger and R. Lampert, “Implantable Cardioverter-Defibrillators Expanding Indications and Technologies,” *JAMA*, vol. 295, pp. 809–818, 02 2006. [15](#)
- [8] D. Sullivan, *Acute Myocardial Infarction*, 2018. <https://www.healthline.com/health/acute-myocardial-infarction>. [15](#)
- [9] D. Sullivan, *Stable Angina*, 2018. <https://www.healthline.com/health/stable-angina>. [16](#)
- [10] D. Sullivan, *What Is Arrhythmia?*, 2017. <https://www.healthline.com/health/arrhythmia>. [18](#)
- [11] D. Sullivan, *Heart Failure*, 2020. <https://www.healthline.com/health/heart-failure>. [19](#)
- [12] I. Metrics, “Evaluation : the global burden of disease : generating evidence, guiding policy,” *WA : IHME Seattle*, 2013. [19](#)
- [13] T. Vos, A. D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, S. Abdalla, V. Aboyans, *et al.*, “Years lived with disability (ylds) for 1160 sequelae of 289 diseases and injuries 1990–2010 : a systematic analysis for the global burden of disease study 2010,” *The lancet*, vol. 380, no. 9859, pp. 2163–2196, 2012. [19](#)
- [14] C. J. Murray, T. Vos, R. Lozano, M. Naghavi, A. D. Flaxman, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, S. Abdalla, *et al.*, “Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990–2010 : a systematic analysis for the global burden of disease study 2010,” *The lancet*, vol. 380, no. 9859, pp. 2197–2223, 2012. [19](#)

- [15] S. Yusuf, S. Rangarajan, K. Teo, S. Islam, W. Li, L. Liu, J. Bo, Q. Lou, F. Lu, T. Liu, *et al.*, “Cardiovascular risk and events in 17 low-, middle-, and high-income countries,” *New England Journal of Medicine*, vol. 371, no. 9, pp. 818–827, 2014. [19](#)
- [16] D. Coggon, D. Barker, and G. Rose, *Epidemiology for the Uninitiated*. John Wiley & Sons, 2009. [20](#), [21](#), [24](#)
- [17] S. Tadayon, K. Wickramasinghe, and N. Townsend, “Examining trends in cardiovascular disease mortality across europe : how does the introduction of a new european standard population affect the description of the relative burden of cardiovascular disease?,” *Population health metrics*, vol. 17, no. 1, p. 6, 2019. [20](#)
- [18] M. Pace, G. Lanzieri, M. Glickman, E. Grande, T. Zupanic, B. Wojtyniak, M. Gissler, E. Cayotte, and L. Agafitei, “Revision of the european standard population,” *Report of the Eurostat’s task force. (Publications Office of the European Union, 2013)*, 2013. [20](#)
- [19] S. Lewington, R. Clarke, N. Qizilbash, R. Peto, and R. Collins, “Age-specific relevance of usual blood pressure to vascular mortality : a meta-analysis of individual data for one million adults in 61 prospective studies. *lancet*. 2002 ; 360 : 1903–1913,” *Hellenic Journal of Cardiology) HJC*, vol. 365, 2003. [21](#)
- [20] M. H. Forouzanfar, P. Liu, G. A. Roth, M. Ng, S. Biryukov, L. Marczak, L. Alexander, K. Estep, K. H. Abate, T. F. Akinyemiju, *et al.*, “Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm hg, 1990-2015,” *Jama*, vol. 317, no. 2, pp. 165–182, 2017. [21](#)
- [21] S. Yusuf, S. Hawken, S. Ôunpuu, T. Dans, A. Avezum, F. Lanas, M. McQueen, A. Budaj, P. Pais, J. Varigos, *et al.*, “Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the interheart study) : case-control study,” *The lancet*, vol. 364, no. 9438, pp. 937–952, 2004. [21](#)
- [22] X. Xie, E. Atkins, J. Lv, A. Bennett, B. Neal, T. Ninomiya, M. Woodward, S. MacMahon, F. Turnbull, G. S. Hillis, *et al.*, “Effects of intensive blood pressure lowering on cardiovascular and renal outcomes : updated systematic review and meta-analysis,” *The Lancet*, vol. 387, no. 10017, pp. 435–443, 2016. [21](#)
- [23] J. Stamler, D. Wentworth, and J. D. Neaton, “Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? : findings in 356 222 primary screenees of the multiple risk factor intervention trial (mrfit),” *Jama*, vol. 256, no. 20, pp. 2823–2828, 1986. [21](#)
- [24] B. A. Ference, W. Yoo, I. Alesh, N. Mahajan, K. K. Mirowska, A. Mewada, J. Kahn, L. Afonso, K. A. Williams, and J. M. Flack, “Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease : a mendelian randomization analysis,” *Journal of the American College of Cardiology*, vol. 60, no. 25, pp. 2631–2639, 2012. [21](#)
- [25] C. T. Trialists, “(cct) collaborators, mihaylova b, emberson j,” *Blackwell L, Keech A, Simes J, Barnes EH, Voysey M, Gray A, Collins R, Baigent C. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease : meta-analysis of individual data from*, vol. 27, pp. 581–590, 2012. [21](#)
- [26] M. G. Silverman, B. A. Ference, K. Im, S. D. Wiviott, R. P. Giugliano, S. M. Grundy, E. Braunwald, and M. S. Sabatine, “Association between lowering ldl-c and cardiovascular risk reduction among different therapeutic interventions : a systematic review and meta-analysis,” *Jama*, vol. 316, no. 12, pp. 1289–1297, 2016. [21](#)

- [27] W. H. Organization *et al.*, “Global health observatory (gho) data. 2016,” *Child mortality and causes of death*. WHO, Geneva, 2016. 21, 22
- [28] N. R. F. Collaboration *et al.*, “Worldwide trends in blood pressure from 1975 to 2015 : a pooled analysis of 1479 population-based measurement studies with 19· 1 million participants,” *Lancet (London, England)*, vol. 389, no. 10064, p. 37, 2017. 22
- [29] V. S. Malik, W. C. Willett, and F. B. Hu, “Global obesity : trends, risk factors and policy implications,” *Nature Reviews Endocrinology*, vol. 9, no. 1, p. 13, 2013. 22
- [30] E. Di Angelantonio, S. N. Bhupathiraju, D. Wormser, P. Gao, S. Kaptoge, A. B. de Gonzalez, B. J. Cairns, R. Huxley, C. L. Jackson, G. Joshy, *et al.*, “Body-mass index and all-cause mortality : individual-participant-data meta-analysis of 239 prospective studies in four continents,” *The Lancet*, vol. 388, no. 10046, pp. 776–786, 2016. 22
- [31] S. J. Hoffman and C. Tan, “Overview of systematic reviews on the health-related effects of government tobacco control policies,” *BMC public health*, vol. 15, no. 1, p. 744, 2015. 22
- [32] M. Stahre, J. Roeber, D. Kanny, R. D. Brewer, and X. Zhang, “Peer reviewed : Contribution of excessive alcohol consumption to deaths and years of potential life lost in the united states,” *Preventing chronic disease*, vol. 11, 2014. 23
- [33] P. Anderson and B. Baumberg, “alcohol in europe,” *London : Institute of alcohol studies*, vol. 2, pp. 73–75, 2006. 23
- [34] V. Miller, A. Mente, M. Dehghan, S. Rangarajan, X. Zhang, S. Swaminathan, G. Dagenais, R. Gupta, V. Mohan, S. Lear, *et al.*, “Fruit, vegetable, and legume intake, and cardiovascular disease and deaths in 18 countries (pure) : a prospective cohort study,” *The Lancet*, vol. 390, no. 10107, pp. 2037–2049, 2017. 23
- [35] D. E. Threapleton, D. C. Greenwood, C. E. Evans, C. L. Cleghorn, C. Nykjaer, C. Woodhead, J. E. Cade, C. P. Gale, and V. J. Burley, “Dietary fibre intake and risk of cardiovascular disease : systematic review and meta-analysis,” *Bmj*, vol. 347, p. f6879, 2013. 23
- [36] D. Aune, E. Giovannucci, P. Boffetta, L. T. Fadnes, N. Keum, T. Norat, D. C. Greenwood, E. Riboli, L. J. Vatten, and S. Tonstad, “Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies,” *International journal of epidemiology*, vol. 46, no. 3, pp. 1029–1056, 2017. 23
- [37] I.-M. Lee, E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, P. T. Katzmarzyk, L. P. A. S. W. Group, *et al.*, “Effect of physical inactivity on major non-communicable diseases worldwide : an analysis of burden of disease and life expectancy,” *The lancet*, vol. 380, no. 9838, pp. 219–229, 2012. 24
- [38] M. Hamer, G. O’Donovan, and M. Murphy, “Physical inactivity and the economic and health burdens due to cardiovascular disease : exercise as medicine,” in *Exercise for Cardiovascular Disease Prevention and Treatment*, pp. 3–18, Springer, 2017. 24
- [39] P. Jeemon, G. Mini, K. Thankappan, and P. Sylaja, “Gbd 2017 mortality collaborators. global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017 : a systematic analysis for the global burden of disease study 2017,” 2018. 25
- [40] C. J. Murray, R. M. Barber, K. J. Foreman, A. A. Ozgoren, F. Abd-Allah, S. F. Abera, V. Aboyans, J. P. Abraham, I. Abubakar, L. J. Abu-Raddad, *et al.*, “Global, regional, and national disability-adjusted life years (dalys) for 306 diseases and injuries and healthy life expectancy (hale) for 188 countries, 1990–2013 : quantifying the epidemiological transition,” *The Lancet*, vol. 386, no. 10009, pp. 2145–2191, 2015. 25

- [41] D. Dicker, G. Nguyen, A. D. Lopez, and C. J. Murray, “Another blow to credibility in published data sources—author’s reply,” *The Lancet*, vol. 394, no. 10192, pp. 27–28, 2019. [25](#)
- [42] M. Rigby, S. Deshpande, and M. Blair, “Another blow to credibility in published data sources,” *The Lancet*, vol. 394, no. 10192, pp. 26–27, 2019. [25](#)
- [43] R. C. Alvarez, “The promise of e-health—a canadian perspective,” *Ehealth international*, vol. 1, no. 1, p. 4, 2002. [30](#)
- [44] H. Oh, C. Rizo, M. Enkin, and A. Jadad, “What is ehealth (3) : a systematic review of published definitions,” *Journal of medical Internet research*, vol. 7, no. 1, p. e1, 2005. [30](#)
- [45] V. Della Mea, “What is e-health (2) : The death of telemedicine?,” *Journal of medical Internet research*, vol. 3, no. 2, p. e22, 2001. [30](#)
- [46] A. R. J. Gunther Eysenbach, “Evidence-based patient choice and consumer health informatics in the internet age,” *J Med Internet Res*, vol. 3, no. 2, p. e19, 2001. [30](#)
- [47] P. F. Binkley, “Predicting the potential of wearable technology,” *IEEE engineering in medicine and biology magazine*, vol. 22, no. 3, pp. 23–27, 2003. [31](#)
- [48] P. Bonato, “Wearable sensors/systems and their impact on biomedical engineering,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 3, pp. 18–20, 2003. [31](#)
- [49] Y. ElSaadany, A. J. A. Majumder, and D. R. Ucci, “A wireless early prediction system of cardiac arrest through iot,” in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 690–695, IEEE, 2017. [31](#)
- [50] X. Liu, Y. Zheng, M. W. Phyu, B. Zhao, M. Je, and X. Yuan, “Multiple functional ecg signal is processing for wearable applications of long-term cardiac monitoring,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 380–389, 2010.
- [51] M. Hadjem, O. Salem, and F. Naït-Abdesselam, “An ecg monitoring system for prediction of cardiac anomalies using wban,” in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 441–446, IEEE, 2014.
- [52] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng, “A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 734–740, 2010. [31](#)
- [53] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, “Mimic-iii, a freely accessible critical care database,” *Scientific data*, vol. 3, p. 160035, 2016. [32](#)
- [54] M. Mohamed, M.-C. Farah, and A. Fahed, “Prédiction de la fibrillation atriale à l’aide des signaux physiologiques,” in *Jetsan 2019 : Journées d’Etude sur la TéléSanté*, pp. 1–6, 2019. [32](#)
- [55] M. Mohamed, M.-C. Farah, and A. Fahed, “Evidence-based prediction of atrial fibrillation using physiological signals,” in *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pp. 1–4, IEEE, 2019.
- [56] M. Mohamed, M.-C. Farah, and A. Fahed, “Prediction of atrial fibrillation episode from physiological signals using belief functions theory,” pp. 1–12, 2020. [32](#)
- [57] R. R. Yager, “Decision making under dempster-shafer uncertainties,” *International Journal of General System*, vol. 20, no. 3, pp. 233–245, 1992. [33](#), [47](#), [64](#), [109](#)

- [58] M. Mohamed, M.-C. Farah, and A. Fahed, "Atrial fibrillation predictor with reject option using belief functions theory," in *2020 IEEE 15th International Symposium on Medical Measurements and Applications*, pp. 1–6, 2020. 33
- [59] M. Mohamed, M.-C. Farah, and A. Fahed, "Evidential prediction of atrial fibrillation with rejection," in *2020 Smart Health International Conference*, pp. 1–2, 2020. 33
- [60] M. Mohamed, M.-C. Farah, and A. Fahed, "Multi-label classification and evidential approach in diseases diagnoses using physiological signals," in *2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering*, pp. 1–5, 2020. 33
- [61] M. Mohamed, M.-C. Farah, and A. Fahed, "Prediction of atrial fibrillation episode from physiological signals using belief functions theory," pp. 1–13, 2020. 33
- [62] G. Shafer, *A mathematical theory of evidence*, vol. 42. Princeton university press, 1976. 36
- [63] K. Sentz, S. Ferson, *et al.*, *Combination of evidence in Dempster-Shafer theory*, vol. 4015. Sandia National Laboratories Albuquerque, 2002. 42
- [64] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic works of the Dempster-Shafer theory of belief functions*, pp. 57–72, Springer, 2008. 44
- [65] D. Mercier, B. Quost, and T. Dencœur, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Information fusion*, vol. 9, no. 2, pp. 246–258, 2008. 45
- [66] L. A. Zadeh, "Review of a mathematical theory of evidence," *AI magazine*, vol. 5, no. 3, pp. 81–81, 1984. 45
- [67] J.-Y. Jaffray, "Linear utility theory for belief functions," *Operations Research Letters*, vol. 8, no. 2, pp. 107–112, 1989. 47
- [68] P. Smets and R. Kennes, "The transferable belief model," *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [69] T. M. Strat, "Decision analysis using belief functions," *International Journal of Approximate Reasoning*, vol. 4, no. 5-6, pp. 391–417, 1990. 47
- [70] E. Ebrahimzadeh, M. Kalantari, M. Joulani, R. S. Shahraki, F. Fayaz, and F. Ahmadi, "Prediction of paroxysmal atrial fibrillation : A machine learning based approach using combined feature vector and mixture of expert classification on hrv signal," *Computer methods and programs in biomedicine*, vol. 165, pp. 53–67, 2018. 50
- [71] K. Boon, M. Khalil-Hani, and M. Malarvili, "Paroxysmal atrial fibrillation prediction based on hrv analysis and non-dominated sorting genetic algorithm iii," *Computer methods and programs in biomedicine*, vol. 153, pp. 171–184, 2018. 71
- [72] K. Boon, M. Khalil-Hani, M. Malarvili, and C. Sia, "Paroxysmal atrial fibrillation prediction method with shorter hrv sequences," *Computer methods and programs in biomedicine*, vol. 134, pp. 187–196, 2016. 71
- [73] P. Vizza, A. Curcio, G. Tradigo, C. Indolfi, and P. Veltri, "A framework for the atrial fibrillation prediction in electrophysiological studies," *Computer methods and programs in biomedicine*, vol. 120, no. 2, pp. 65–76, 2015.
- [74] M. Mohebbi and H. Ghassemian, "Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal," *Computer methods and programs in biomedicine*, vol. 105, no. 1, pp. 40–49, 2012. 50

- [75] T. Denoeux, “Distributed combination of belief functions,” *Information Fusion*, 2020. 50
- [76] P. P. Shenoy, “An expectation operator for belief functions in the dempster–shafer theory,” *International Journal of General Systems*, vol. 49, no. 1, pp. 112–141, 2020.
- [77] H. Zhang and Y. Deng, “Weighted belief function of sensor data fusion in engine fault diagnosis,” *Soft computing*, vol. 24, no. 3, pp. 2329–2339, 2020.
- [78] F. Xiao, “Generalized belief function in complex evidence theory,” *Journal of Intelligent & Fuzzy Systems*, no. Preprint, pp. 1–9, 2020.
- [79] F. Xiao, “A new divergence measure for belief functions in d–s evidence theory for multisensor data fusion,” *Information Sciences*, vol. 514, pp. 462–483, 2020. 50
- [80] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, “Mimic-iii, a freely accessible critical care database,” *Scientific data*, vol. 3, p. 160035, 2016. 50
- [81] A. E. Pollard, Tom J abd Johnson, “The MIMIC-III Clinical Database, doi=10.13026/C2XW26, year=2016, howpublished= <http://dx.doi.org/10.13026/c2xw26>.” 50
- [82] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet : components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. 50
- [83] A. Moskowitz, K. P. Chen, A. Z. Cooper, A. Chahin, M. M. Ghassemi, and L. A. Celi, “Management of atrial fibrillation with rapid ventricular response in the intensive care unit : a secondary analysis of electronic health record data,” *Shock (Augusta, Ga.)*, vol. 48, no. 4, p. 436, 2017. 53
- [84] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction : foundations and applications*, vol. 207. Springer, 2008. 55
- [85] A. Zubiolo, *Extraction de caractéristiques et apprentissage statistique pour l’imagerie biomédicale cellulaire et tissulaire*. PhD thesis, Nice, 2015. 55
- [86] A. Taoum, F. Mourad-Chehade, and H. Amoud, “Evidence-based model for real-time surveillance of ards,” *Biomedical Signal Processing and Control*, vol. 50, pp. 83–91, 2019. 56
- [87] R. G. Lyons, *Understanding digital signal processing, 3/E*. Pearson Education India, 2004. 56
- [88] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley interdisciplinary reviews : computational statistics*, vol. 2, no. 4, pp. 433–459, 2010. 56
- [89] J. B. Kruskal, *Multidimensional scaling*. No. 11, Sage, 1978. 56
- [90] A. Romero, C. Gatta, and G. Camps-Valls, “Unsupervised deep feature extraction for remote sensing image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1349–1362, 2015. 56
- [91] S. Lloyd, “Least squares quantization in pcm,” *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982. 59
- [92] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu, “Hierarchical clustering : Objective functions and algorithms,” *Journal of the ACM (JACM)*, vol. 66, no. 4, pp. 1–42, 2019. 59

- [93] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, “A methodology for applying k-nearest neighbor to time series forecasting,” *Artificial Intelligence Review*, vol. 52, no. 3, 2019. 59
- [94] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998. 59
- [95] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016. 59
- [96] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [97] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, “Deep learning for health informatics,” *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.
- [98] M. A. Nielsen, *Neural networks and deep learning*, vol. 2018. Determination press San Francisco, CA, 2015.
- [99] L. Deng and D. Yu, “Deep learning : methods and applications,” *Foundations and trends in signal processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [100] J. Schmidhuber, “Deep learning in neural networks : An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015. 59
- [101] J. D. Gibbons and S. Chakraborti, “Comparisons of the mann-whitney, student’s t, and alternate t tests for means of normal distributions,” *The Journal of Experimental Education*, vol. 59, no. 3, pp. 258–267, 1991. 61
- [102] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009. 70
- [103] S. Dudani, “The distance-weighted k-nearest neighbor rule,” *IEEE trans. on systems, man and cybernetics*, vol. 8, no. 4, pp. 311–313, 1978. 70
- [104] W. S. Noble, “What is a support vector machine?,” *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006. 70
- [105] R. Alcaraz, A. Martínez, and J. J. Rieta, “Role of the p-wave high frequency energy and duration as noninvasive cardiovascular predictors of paroxysmal atrial fibrillation,” *Computer methods and programs in biomedicine*, vol. 119, no. 2, pp. 110–119, 2015. 71
- [106] B. Al-Salemi, S. A. M. Noah, and M. J. Ab Aziz, “Rfboost : an improved multi-label boosting algorithm and its application to text categorisation,” *Knowledge-Based Systems*, vol. 103, pp. 104–117, 2016. 90
- [107] S. Gao, W. Wu, C.-H. Lee, and T.-S. Chua, “A mfom learning approach to robust multiclass multi-label text categorization,” in *Proceedings of the twenty-first international conference on Machine learning*, p. 42, 2004.
- [108] G. Chen, D. Ye, Z. Xing, J. Chen, and E. Cambria, “Ensemble application of convolutional and recurrent neural networks for multi-label text categorization,” in *2017 international joint conference on neural networks (IJCNN)*, pp. 2377–2383, IEEE, 2017.
- [109] A. F. De Souza, F. Pedroni, E. Oliveira, P. M. Ciarelli, W. F. Henrique, L. Veronese, and C. Baudue, “Automated multi-label text categorization with vg-ram weightless neural networks,” *Neurocomputing*, vol. 72, no. 10-12, pp. 2209–2217, 2009.

- [110] J.-Y. Jiang, S.-C. Tsai, and S.-J. Lee, “Fsknn : multi-label text categorization based on fuzzy similarity and k nearest neighbors,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 2813–2821, 2012.
- [111] H. Elghazel, A. Aussem, O. Gharroudi, and W. Saadaoui, “Ensemble multi-label text categorization based on rotation forest and latent semantic indexing,” *Expert Systems with Applications*, vol. 57, pp. 1–11, 2016.
- [112] P. M. Ciarelli, E. Oliveira, C. Badue, and A. F. De Souza, “Multi-label text categorization using a probabilistic neural network,” *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 1, no. 133-144, p. 51, 2009.
- [113] A. Fujino, H. Isozaki, and J. Suzuki, “Multi-label text categorization with model combination based on f1-score maximization,” in *Proceedings of the Third International Joint Conference on Natural Language Processing : Volume-II*, 2008. 90
- [114] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, “Hierarchical multi-label prediction of gene function,” *Bioinformatics*, vol. 22, no. 7, pp. 830–836, 2006. 90
- [115] L. Schietgat, C. Vens, J. Struyf, H. Blockeel, D. Kocev, and S. Džeroski, “Predicting gene function using hierarchical multi-label decision tree ensembles,” *BMC bioinformatics*, vol. 11, no. 1, p. 2, 2010.
- [116] Y.-X. Li, S. Ji, S. Kumar, J. Ye, and Z.-H. Zhou, “Drosophila gene expression pattern annotation through multi-instance multi-label learning,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 1, pp. 98–112, 2011.
- [117] S. Wan, M.-W. Mak, and S.-Y. Kung, “mgoasvm : Multi-label protein subcellular localization based on gene ontology and support vector machines,” *BMC bioinformatics*, vol. 13, no. 1, p. 290, 2012.
- [118] N. Cesa-Bianchi, M. Re, and G. Valentini, “Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference,” *Machine Learning*, vol. 88, no. 1-2, pp. 209–241, 2012.
- [119] B. Jin, B. Muller, C. Zhai, and X. Lu, “Multi-label literature classification based on the gene ontology graph,” *BMC bioinformatics*, vol. 9, no. 1, p. 525, 2008.
- [120] D. Stojanova, M. Ceci, D. Malerba, and S. Dzeroski, “Using ppi network autocorrelation in hierarchical multi-label classification trees for gene function prediction,” *BMC bioinformatics*, vol. 14, no. 1, p. 285, 2013.
- [121] G.-Z. Li, M. You, L. Ge, J. Y. Yang, and M. Q. Yang, “Feature selection for semi-supervised multi-label learning with application to gene function analysis,” in *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, pp. 354–357, 2010.
- [122] S. Feng, P. Fu, and W. Zheng, “A hierarchical multi-label classification algorithm for gene function prediction,” *Algorithms*, vol. 10, no. 4, p. 138, 2017. 90
- [123] L. Tang, S. Rajan, and V. K. Narayanan, “Large scale multi-label classification via metabeler,” in *Proceedings of the 18th international conference on World wide web*, pp. 211–220, 2009. 90
- [124] J. Pestian, C. Brew, P. Matykiewicz, D. J. Hovermale, N. Johnson, K. B. Cohen, and W. Duch, “A shared task involving multi-label classification of clinical free text,” in *Biological, translational, and clinical language processing*, pp. 97–104, 2007.

- [125] C.-L. Chen, F. S. Tseng, and T. Liang, "An integration of wordnet and fuzzy association rule mining for multi-label document clustering," *Data & Knowledge Engineering*, vol. 69, no. 11, pp. 1208–1226, 2010. [90](#)
- [126] J. Arunadevi and V. Rajamani, "An evolutionary multi label classification using associative rule mining for spatial preferences," *IJCA Special Issue on Artificial Intelligence Techniques- Novel Approaches and Practical Applications*, vol. 3, pp. 28–37, 2011.
- [127] K. Patel, N. Kapadia, and M. Parikh, "Discover multi-label classification using association rule mining," *International journal of Advance Engineering and Research Development*, vol. 1, no. 1, 2014.
- [128] C. Prathibhamol, G. Amala, and M. Kapadia, "Anomaly detection based multi label classification using association rule mining (admlcar)," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2703–2707, IEEE, 2016.
- [129] A. Tazaree, A.-M. Eftekhari-Moghadam, and S. Sajjadi-Ghaem-Maghani, "A semantic image classifier based on hierarchical fuzzy association rule mining," *Multimedia tools and applications*, vol. 69, no. 3, pp. 921–949, 2014.
- [130] R. Patel, J. Vala, and K. Patel, "Classification on multi-label dataset using rule mining technique," *IJRET : International Journal of Research in Engineering and Technology*, vol. 3, no. 06, 2014.
- [131] K. Song and K. Lee, "Predictability-based collective class association rule mining," *Expert Systems with Applications*, vol. 79, pp. 1–7, 2017.
- [132] A. Taher, "Rule mining and prediction using the flek machine—a new machine learning engine," in *Canadian Conference on Artificial Intelligence*, pp. 388–393, Springer, 2018.
- [133] B. V. Balaji and V. V. Rao, "Improved classification based association rule mining," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 5, pp. 2211–2221, 2013. [90](#)
- [134] W. Shi, X. Liu, and Q. Yu, "Correlation-aware multi-label active learning for web service tag recommendation," in *2017 IEEE International Conference on Web Services (ICWS)*, pp. 229–236, IEEE, 2017. [90](#)
- [135] C. Shen, J. Jiao, Y. Yang, and B. Wang, "Multi-instance multi-label learning for automatic tag recommendation," in *2009 IEEE International Conference on Systems, Man and Cybernetics*, pp. 4910–4914, IEEE, 2009.
- [136] L. Li, Y. Yao, F. Xu, and J. Lu, "Matar : Keywords enhanced multi-label learning for tag recommendation," in *Asia-Pacific Web Conference*, pp. 268–279, Springer, 2015.
- [137] Y. Yamada and T. Nakatoh, "Tag recommendation for open government data by multi-label classification and particular noun phrase extraction," in *10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2018*, pp. 83–91, SciTePress, 2018. [90](#)
- [138] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 195–200, 2005. [93](#), [97](#)
- [139] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," in *Pacific-Asia conference on knowledge discovery and data mining*, pp. 22–30, Springer, 2004.

- [140] R. E. Schapire and Y. Singer, “Boostexter : A boosting-based system for text categorization,” *Machine learning*, vol. 39, no. 2-3, pp. 135–168, 2000. [93](#)
- [141] G. Tsoumakas and I. Vlahavas, “Random k-labelsets : An ensemble method for multilabel classification,” in *European conference on machine learning*, pp. 406–417, Springer, 2007. [93](#), [96](#), [102](#)
- [142] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, “Learning multi-label scene classification,” *Pattern recognition*, vol. 37, no. 9, pp. 1757–1771, 2004. [96](#), [100](#)
- [143] J. Read, B. Pfahringer, G. Holmes, and E. Frank, “Classifier chains for multi-label classification,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 254–269, Springer, 2009. [96](#), [98](#), [100](#), [102](#)
- [144] J. Fürnkranz, E. Hüllermeier, E. L. Mencía, and K. Brinker, “Multilabel classification via calibrated label ranking,” *Machine learning*, vol. 73, no. 2, pp. 133–153, 2008. [96](#), [129](#)
- [145] A. Clare and R. D. King, “Knowledge discovery in multi-label phenotype data,” in *European conference on principles of data mining and knowledge discovery*, pp. 42–53, Springer, 2001. [97](#), [104](#)
- [146] A. Elisseeff and J. Weston, “A kernel method for multi-labelled classification,” in *Advances in neural information processing systems*, pp. 681–687, 2002. [97](#)
- [147] J. Read, B. Pfahringer, G. Holmes, and E. Frank, “Classifier chains for multi-label classification,” *Machine learning*, vol. 85, no. 3, p. 333, 2011. [98](#), [129](#)
- [148] J. C. Zaragoza, E. Sucar, E. Morales, C. Bielza, and P. Larranaga, “Bayesian chain classifiers for multidimensional classification,” in *Twenty-second international joint conference on artificial intelligence*, Citeseer, 2011. [99](#)
- [149] E. L. Mencía, S.-H. Park, and J. Fürnkranz, “Efficient voting prediction for pairwise multilabel classification,” *Neurocomputing*, vol. 73, no. 7-9, pp. 1164–1176, 2010. [100](#)
- [150] G. Madjarov, D. Gjorgjevikj, and T. Delev, “Efficient two stage voting architecture for pairwise multi-label classification,” in *Australasian Joint Conference on Artificial Intelligence*, pp. 164–173, Springer, 2010. [100](#)
- [151] G. Madjarov, D. Gjorgjevikj, and S. Džeroski, “Two stage architecture for multi-label learning,” *Pattern Recognition*, vol. 45, no. 3, pp. 1019–1034, 2012. [100](#), [129](#)
- [152] E. L. Mencía and J. Furnkranz, “Pairwise learning of multilabel classifications with perceptrons,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2899–2906, IEEE, 2008. [100](#)
- [153] E. L. Mencia and J. Fürnkranz, “Efficient pairwise multilabel classification for large-scale problems in the legal domain,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 50–65, Springer, 2008. [100](#)
- [154] J. Read, B. Pfahringer, and G. Holmes, “Multi-label classification using ensembles of pruned sets,” in *2008 eighth IEEE international conference on data mining*, pp. 995–1000, IEEE, 2008. [101](#)
- [155] G. Tsoumakas, I. Katakis, and I. Vlahavas, “Random k-labelsets for multilabel classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079–1089, 2010. [102](#), [116](#)

- [156] C. Shi, X. Kong, S. Y. Philip, and B. Wang, “Multi-label ensemble learning,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 223–239, Springer, 2011. [102](#), [116](#), [117](#)
- [157] C. Sanden and J. Z. Zhang, “Enhancing multi-label music genre classification through ensemble techniques,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 705–714, 2011. [102](#)
- [158] L. Tenenboim-Chekina, L. Rokach, and B. Shapira, “Identification of label dependencies for multi-label classification,” in *Working Notes of the Second International Workshop on Learning from Multi-Label Data*, pp. 53–60, 2010. [102](#)
- [159] D. Kocev, C. Vens, J. Struyf, and S. Džeroski, “Ensembles of multi-objective decision trees,” in *European conference on machine learning*, pp. 624–631, Springer, 2007. [104](#)
- [160] X. Zhang, Q. Yuan, S. Zhao, W. Fan, W. Zheng, and Z. Wang, “Multi-label classification without the multi-label cost,” in *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 778–789, SIAM, 2010. [104](#)
- [161] A. Jiang, C. Wang, and Y. Zhu, “Calibrated rank-svm for multi-label image categorization,” in *2008 IEEE international joint conference on neural networks (IEEE world Congress on Computational Intelligence)*, pp. 1450–1455, IEEE, 2008. [105](#)
- [162] S. Ji, L. Sun, R. Jin, and J. Ye, “Multi-label multiple kernel learning,” *Advances in neural information processing systems*, vol. 21, pp. 777–784, 2008. [105](#)
- [163] S. Bucak, R. Jin, and A. Jain, “Multi-label multiple kernel learning by stochastic approximation : Application to visual object recognition,” *Advances in Neural Information Processing Systems*, vol. 23, pp. 325–333, 2010. [105](#)
- [164] R. Srikant and R. Agrawal, “Mining generalized association rules,” 1995. [109](#)
- [165] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. I. Verkamo, *et al.*, “Fast discovery of association rules,” *Advances in knowledge discovery and data mining*, vol. 12, no. 1, pp. 307–328, 1996.
- [166] H. Toivonen *et al.*, “Sampling large databases for association rules,” in *Vldb*, vol. 96, pp. 134–145, 1996.
- [167] R. Agrawal and J. C. Shafer, “Parallel mining of association rules,” *IEEE Transactions on knowledge and Data Engineering*, vol. 8, no. 6, pp. 962–969, 1996.
- [168] B. Ozden, S. Ramaswamy, and A. Silberschatz, “Cyclic association rules,” in *Proceedings 14th International Conference on Data Engineering*, pp. 412–421, IEEE, 1998.
- [169] P. B. Bafna and J. R. Saini, “Identification of significant challenges faced by the tourism and hospitality industries using association rules,” in *Data Management, Analytics and Innovation*, pp. 121–129, Springer, 2020.
- [170] E. V. Altay and B. Alatas, “Intelligent optimization algorithms for the problem of mining numerical association rules,” *Physica A : Statistical Mechanics and its Applications*, vol. 540, p. 123142, 2020.
- [171] G. J. Krishna and V. Ravi, “Mining top high utility association rules using binary differential evolution,” *Engineering Applications of Artificial Intelligence*, vol. 96, p. 103935, 2020. [109](#)
- [172] M.-L. Zhang, Y.-K. Li, X.-Y. Liu, and X. Geng, “Binary relevance for multi-label learning : an overview,” *Frontiers of Computer Science*, vol. 12, no. 2, pp. 191–202, 2018. [116](#)

- [173] J. Read, *Scalable multi-label classification*. PhD thesis, University of Waikato, 2010. [117](#)
- [174] Z. Younes, F. Abdallah, T. Denoeux, and H. Snoussi, “A dependent multilabel classification method derived from the k-nearest neighbor rule,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, pp. 1–14, 2011. [117](#)
- [175] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, “Keel data-mining software tool : data set repository, integration of algorithms and experimental analysis framework,” *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, 2011. [123](#)
- [176] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas, “Multi-label classification of music into emotions,” in *ISMIR*, vol. 8, pp. 325–330, 2008.
- [177] G. Tsoumakas, I. Katakis, and I. Vlahavas, “Mining multi-label data,” in *Data mining and knowledge discovery handbook*, pp. 667–685, Springer, 2009. [123](#)
- [178] J. M. Moyano, E. L. Gibaja, K. J. Cios, and S. Ventura, “An evolutionary approach to build ensembles of multi-label classifiers,” *Information Fusion*, vol. 50, pp. 168–180, 2019. [128](#)
- [179] M. A. Tahir, J. Kittler, and A. Bouridane, “Multilabel classification using heterogeneous ensemble of multi-label classifiers,” *Pattern Recognition Letters*, vol. 33, no. 5, pp. 513–523, 2012. [129](#)
- [180] W. Cheng and E. Hüllermeier, “Combining instance-based learning and logistic regression for multilabel classification,” *Machine Learning*, vol. 76, no. 2-3, pp. 211–225, 2009. [129](#)
- [181] D. Kocev, C. Vens, J. Struyf, and S. Džeroski, “Tree ensembles for predicting structured outputs,” *Pattern Recognition*, vol. 46, no. 3, pp. 817–833, 2013. [129](#)
- [182] G. Tsoumakas, I. Katakis, and I. Vlahavas, “Effective and efficient multilabel classification in domains with large number of labels,” in *Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD’08)*, vol. 21, pp. 53–59, sn, 2008. [129](#)
- [183] Q. Wu, M. Tan, H. Song, J. Chen, and M. K. Ng, “MI-forest : A multi-label tree ensemble method for multi-label classification,” *IEEE transactions on knowledge and data engineering*, vol. 28, no. 10, pp. 2665–2680, 2016. [129](#)

# Mohamed MROUEH DIT INJIBAR

Doctorat : Optimisation et Sûreté des Systèmes

Année 2021

## Classification évidentielle mono- et multi-label : application à la détection de maladies cardio-vasculaires

Cette thèse porte sur la détection de maladies cardio-vasculaires grâce à la surveillance de signaux physiologiques. L'objectif est de développer des approches de classification mono- et multi-label, basées sur la théorie des fonctions de croyance, pour prédire ou diagnostiquer une complication liée à une ou plusieurs maladies cardio-vasculaires. D'abord, une approche proposant une extraction de paramètres et une modélisation de l'information dans un cadre évidentiel est développée pour prédire la fibrillation atriale, une arythmie cardiaque. Une extension de cette approche utilise une option de rejet de classification et une modélisation alternative de l'information. La thèse élargit par la suite le domaine d'application pour couvrir plusieurs maladies cardio-vasculaires à la fois. Le problème est alors défini comme une classification multi-label où les labels représentent des traits caractéristiques des maladies. Une approche de classification multi-label est développée dans le domaine évidentiel qui se sert des corrélations entre les maladies pour accroître la précision du diagnostic. Enfin, une approche théorique de classification multi-label qui profite de la corrélation entre les labels a été proposée. Cette méthode d'ensemble permet de classifier en multi-label de manière efficace. Les approches proposées sont validées grâce à une base de données médicale publique, MIMIC III, hébergée sur Physionet.

**Mots clés :** théorie de Dempster-Shafer – traitement du signal – apprentissage supervisé (intelligence artificielle) – appareil cardiovasculaire, maladies – pronostic (médecine).

## Evidential Mono- and Multi-label Classification: Application to Cardiovascular Diseases Detection

This thesis focuses on the detection of cardiovascular diseases through the monitoring of physiological signals. The objective is to develop mono- and multi-label classification approaches, based on the theory of belief functions, to predict or diagnose a complication linked to one or more cardiovascular diseases. First, an approach providing parameter extraction and information modeling in an evidential framework is developed to predict atrial fibrillation, a cardiac arrhythmia. An extension of this approach uses a reject classification option and alternative information modeling. The thesis then broadens the field of application to cover several cardiovascular diseases at the same time. The problem is thus defined as a multi-label classification where the labels represent features of the diseases. A multi-label classification approach is developed in the evidential domain which makes use of correlations between diseases to increase diagnostic accuracy. Finally, a theoretical approach of multi-label classification, which takes advantage of the correlation between labels, has been proposed. This ensemble method allows for efficient multi-label classification. The proposed approaches are validated using a public medical database, MIMIC III, hosted on Physionet.

**Keywords:** Dempster-Shafer theory – signal processing – supervised learning (machine learning) – cardiovascular system, diseases – prognosis.

Thèse réalisée en partenariat entre :

