

الجمهورية الجزائرية الديمقراطية الشعبية

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

وزارة التعليم العالي و البحث العلمي

**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

جامعة أبي بكر بلقايد - تلمسان

Université Aboubakr Belkaïd – Tlemcen –



**THESE**

Présentée pour l'obtention du **grade de DOCTORAT 3<sup>ème</sup> Cycle**

**En** : Electronique Biomédicale

**Spécialité** : Electronique Biomédicale

**Par** : BOUCHIKHI Sarra

**Sujet**

**Classification des Données Médicales  
par Les Modèles de Markov Cachés  
et La Logique Floue**

Soutenue publiquement, le 17/05/2016 , devant le jury composé de :

Mr. ABDERRAHIM M. Amine  
Pr. CHIKH M. Amine  
Pr. AMINE Abdelmalek  
Pr. BERRICHI Zakaria

MCA  
Professeur  
Professeur  
Professeur

Univ. Tlemcen  
Univ. Tlemcen  
Univ. Saida  
Univ. SBA

Président  
Directeur de thèse  
Examineur 1  
Examineur 2

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITÉ ABOU BEKR BELKAID  
FACULTÉ DE TECHNOLOGIE  
DÉPARTEMENT GÉNIE BIOMÉDICAL

## THÈSE

pour obtenir le grade de

**DOCTEUR EN ÉLECTRONIQUE BIOMÉDICALE**

présenté et soutenu publiquement

par

**Melle BOUCHIKHI Sarra**

le 17 Mai 2016

Titre:

# Classification des Données Médicales par Les Modèles de Markov Cachés et La Logique Floue

## Jury

Président du jury. Mr. ABDERRAHIM M. Amine,	MCA Université de Tlemcen
Examineurs. Mr. EL BERRICHI Zakaria,	Professeur Université de Sidi Belabes
Examineurs. Mr. AMINE Abdelmalek,	Professeur Université de Saida
Directeur de thèse. Mr. CHIKH M. Amine,	Professeur Université de Tlemcen



*À mes parents*

# Remerciements

*Je tiens à remercier avant tous, le bon Dieu qui m'a donné la force et la patience tout au long de la préparation de cette thèse. Je tiens à remercier en premier lieu Mr CHIKH Med Amine, mon directeur de thèse, pour son encadrement pour les remarques constructives qu'il m'a fournis ainsi que pour ses précieux conseils durant toute la période de mon travail. En dehors de son apport scientifique, je n'oublierai pas aussi de le remercier pour ses qualités humaines et son soutien qui m'ont permis de mener à bien ce travail.*

*J'adresse ma plus vive gratitude aux membres de mon jury. Mr ABDERRAHIM Mohammed El Amine, MCA à l'université de Tlemcen, Mr EL BERRICHI Zakaria Professeur à l'université de Sidi BelAbbes et Mr ABDELMALEK Amine, Professeur à l'université de Saida ; qui ont eu la gentillesse d'avoir accepté la lourde tâche de rapporteur auprès du jury, je les remercie tout particulièrement pour l'attention et le temps qu'ils y ont consacré pour lire ce manuscrit.*

*Mes pensées vont également vers l'ensemble des personnes avec qui j'ai partagé mes années de recherches au sein du laboratoire de génie biomédical. C'était très agréable de travailler en leur compagnie.*

*Je remercie les membres de l'équipe CREDOM que j'ai pu côtoyer durant la période de ma thèse et qui ont su rendre mon travail agréable, par leur simple présence et l'ambiance qu'ils ont su créer. Je pense à Amine , Sara, Amina, Amina et Nesma pour leur bonne humeur et leur soutien.*

*Je termine en remerciement tout particulièrement Houaria, nouira, souad et wafaa pour nos nombreux cafés et le plaisir partagé à discuter de la recherche ensemble.*

*Sarra BOUCHIKHI*

# Résumé

Dans cette thèse nous présentons nos travaux qui s'articule autour du domaine d'aide au diagnostic médical plus particulièrement en traitement de séquences, réalisés sur deux types de données (signal ECG et image cytologique) avec des tâches différentes (classification, segmentation). Nos contributions concernent d'une part la proposition d'un classifieur supervisé dédié à la reconnaissance des extrasystoles ventriculaires par les modèles de Markov cachés où nous avons intégré une tâche de symbolisation des différents attributs utilisés afin d'augmenter l'interprétabilité de notre système pour justifier la prise de décision. Aussi, un classifieur neuro-floue (ANFIS) est implémenter pour pouvoir comparer l'aspect probabiliste dans les HMM et l'aspect interprétable de la logique floue. Les HMMs sont des modèles probabilistes qui fournis une information de probabilité sur la classe reconnues tandis que le classifieur neuro-floue (ANFIS) nous donne une transparence dans le système de reconnaissance par les règles générés. D'autre part, l'apprentissage par ces mêmes approches est appliqué à la classification pixellaire des cellules cytologique dans un but de segmentation pour une classification future des cellules sanguines.

## Mots clés

Classification, Interprétabilité, HMM, ANFIS, ECG, Images cytologiques, Segmentation.

# Abstract

In this thesis we present our works that are focused on the field of medical diagnosis more particularly in the treatment of sequences, performed on two types of data (ECG signal and cytological images) on different tasks (classification and segmentation). Our contributions concern the proposal of a supervised classifier for the recognition of the ventricular premature beats by hidden Markov models, where we have integrated a symbolization stage for the different used attributes to increase transparency of our system to justify making of the decision . Also, a neuro-fuzzy classifier (ANFIS) is implemented to compare probabilistic side in the HMM and interpretable aspect of fuzzy logic. HMMs are probabilistic models that provided information on the likelihood of recognized class while the neuro-fuzzy classifier (ANFIS) gives transparency in the recognition system by the generation of the different rules. Also, the learning by the same approach is applied to the classification of pixels of the cytological images for segmentation goals for future classification of blood cells.

## Keywords

Classification, Interpretability, HMM, ANFIS, ECG, Cytological images, Segmentation.

# Table des matières

Remerciements . . . . .	i
Résumé . . . . .	ii
Abstract . . . . .	iii
Table des matières . . . . .	iv
Table des figures . . . . .	vii
Liste des tableaux . . . . .	ix
Glossaire . . . . .	x
<b>Introduction Générale</b>	<b>1</b>
1 Contexte . . . . .	2
2 Problématique . . . . .	3
3 Organisation de la thèse . . . . .	3
<b>1 Classification et Représentation des Données Médicales</b>	<b>5</b>
1 Introduction . . . . .	6
2 Représentation des données . . . . .	6
2.1 Collecte et prétraitement des données . . . . .	6
2.2 Prétraitement des données . . . . .	7
3 Sélection des variables . . . . .	8
4 Les différents types de données rencontrés . . . . .	8
4.1 Description classique d'une variable . . . . .	9
5 Discrétisation des données . . . . .	10
6 Intelligence artificielle . . . . .	10
6.1 Classification automatique . . . . .	10
6.2 Classification et apprentissage . . . . .	10
6.3 Le problème de la classification . . . . .	11
6.4 Clustering et apprentissage non-supervisé . . . . .	12
6.5 L'apprentissage supervisé . . . . .	12
7 La classification des données séquentielles . . . . .	13
8 Aspects probabilistes d'une classification automatique . . . . .	13
8.1 Modélisation probabiliste en classification . . . . .	14



9	Interprétabilité . . . . .	14
10	Conclusion . . . . .	15
<b>2</b>	<b>La théorie des Modèles de Markov cachés</b>	<b>16</b>
1	Introduction . . . . .	17
2	Définitions . . . . .	17
2.1	Les chaines de Markov discrètes . . . . .	18
2.2	Les modèles de Markov cachés discrets . . . . .	21
2.3	Évaluation de la probabilité d'une séquence d'observations . . . . .	23
2.4	Décodage/ segmentation de séquences d'observations . . . . .	27
2.5	Maximisation de la vraisemblance . . . . .	29
3	Application des HMMs dans le domaine biomédical . . . . .	32
3.1	Classification des arythmies cardiaque par les HMM dans la littérature . . . . .	32
4	Conclusion . . . . .	33
<b>3</b>	<b>Classification Floue des données médicales</b>	<b>34</b>
1	Introduction . . . . .	35
2	Principe de la logique floue . . . . .	35
2.1	Notion d'appartenance partielle . . . . .	35
2.2	Fonctions d'appartenance . . . . .	36
2.3	Les variables floues . . . . .	38
2.4	Les règles d'inférence . . . . .	40
3	Système D'inférence Floue . . . . .	41
3.1	Principe de fonctionnement du SIF . . . . .	42
3.2	Type des SIF . . . . .	44
4	Vers les systèmes Neuro-Flous . . . . .	45
4.1	Hybridation Neuro-floue . . . . .	45
4.2	Cas du modèle ANFIS . . . . .	46
5	Conclusion . . . . .	48
<b>4</b>	<b>Résultats et Discussion</b>	<b>49</b>
<b>I</b>	<b>Considération 1 : Détection des Extrasystoles ventriculaires</b>	<b>50</b>
1	Proposition 1 : Reconnaissance des Extrasystoles ventriculaire par les HMMs . . . . .	51
1.1	Introduction . . . . .	51
1.2	LE SIGNAL ELECTROCARDIOGRAMME . . . . .	51

1.3	Extrasystole ventriculaire (ESV) . . . . .	55
1.4	Diagnostic à partir du rythme . . . . .	56
1.5	La base de données MIT-BIH . . . . .	57
1.6	Description du système de classification . . . . .	58
1.7	Méthode HMM adaptée . . . . .	58
1.8	Reconnaissance des extrasystoles ventriculaires par HMM . . .	59
2	Proposition 2 : Reconnaissance des Extrasystoles ventriculaire par un modèle Neuro Floue . . . . .	64
2.1	Le choix des fonctions d'appartenances . . . . .	64
2.2	Choix des points modaux . . . . .	65
2.3	Les méthodes interprétables et probabilistes . . . . .	69
<b>II</b>	<b>Considération 2 : Segmentation d'images par HMM</b>	<b>72</b>
3	Introduction . . . . .	73
4	Parcours de Hilbert-Peano . . . . .	73
5	Travaux de l'état de l'art . . . . .	75
6	Expérimentation . . . . .	77
6.1	Base d'apprentissage . . . . .	77
6.2	Résultats . . . . .	78
7	Conclusion . . . . .	79
	<b>Conclusion Générale</b>	<b>82</b>
	<b>Bibliographie</b>	<b>85</b>

# Table des figures

2.1	Représentation graphique de la chaîne de Markov $(\Pi, A)$ . . . . .	20
2.2	Relations de dépendance entre les variables aléatoires d'un HMM. . .	22
2.3	Calcul de la variable Forward. . . . .	25
2.4	Calcul de la variable Backward. . . . .	26
2.5	Représentation de décodage. . . . .	27
2.6	Le retour en arrière. . . . .	29
2.7	Exemple d'HMM à cinq états de gauche à droite. . . . .	32
3.1	Comparaison d'un ensemble classique et d'un ensemble flou . . . . .	35
3.2	Comparaison entre la formalisation en logique booléenne (à gauche) et en logique floue (à droite) . . . . .	37
3.3	Représentation de l'activité journalière d'un individu(logique classique)	38
3.4	Représentation de l'activité journalière d'un individu(logique floue) .	39
3.5	Structure d'un SIF . . . . .	41
3.6	Système d'inférence flou [1] . . . . .	43
3.7	Exemple du SIF de Mandani [2] . . . . .	44
3.8	Exemple du SIF de Sugeno [2] . . . . .	45
3.9	System Neuro-Flou (NFS) . . . . .	46
3.10	Architecture ANFIS . . . . .	47
4.1	Un cycle cardiaque . . . . .	52
4.2	La ligne isoélectrique . . . . .	52
4.3	Segment PR . . . . .	53
4.4	Segment ST . . . . .	53
4.5	Paramètres d'intérêt pour la description d'un battement . . . . .	55
4.6	Extrasystole ventriculaire (ESV). . . . .	56
4.7	Les étapes du système de reconnaissance . . . . .	59
4.9	Courbe de la probabilité de vraisemblance . . . . .	61
4.8	Diagramme de la phase d'apprentissage et de test . . . . .	62
4.10	les fonctions d'appartenances pour les trois entrées (RRp,RRs/RRp,QRS)	66
4.11	L'évolution de l'erreur globale . . . . .	67

---

4.12	les fonctions d'appartenances finaux pour les trois entrées (RRp,RRs/RRp,QRS)	68
4.13	caractéristique d'un battement normal . . . . .	69
4.14	Les règles activé par un battement normal . . . . .	70
4.15	Caractéristique d'un battement ESV . . . . .	70
4.16	Caractéristique d'un battement normal . . . . .	71
4.17	Caractéristique d'un battement ESV bien détecté par approche HMM et non détecté par approche neuro floue . . . . .	71
4.18	Construction du parcours de Peano pour une image de 8x8 . . . . .	74
4.19	Image cytologique médullaire . . . . .	75
4.20	(a) image originale (b) image vérité terrain . . . . .	77
4.21	Transformation 2D en 1D de l'image couleur . . . . .	78
4.22	Modèle HMM pour chaque classe . . . . .	78
4.23	Résultats de segmentation. . . . .	80

# Liste des tableaux

3.1	Les opérateurs logiques . . . . .	40
3.2	La classification des paramètres du SIF [3] . . . . .	42
4.1	Valeurs habituelles des différents paramètres caractérisant un battement cardiaque [4] . . . . .	55
4.2	Les patients choisis pour constituer la base de données . . . . .	57
4.3	Valeurs des intervalles des attributs utilisés . . . . .	60
4.4	Nombre de battements pour chaque enregistrement . . . . .	61
4.5	Performances du classifieur HMM . . . . .	63
4.6	Erreurs quadratiques moyennes pour différentes configurations . . . . .	64
4.7	les points modaux initiaux . . . . .	65
4.8	les points modaux finaux . . . . .	67
4.9	Performances du classifieur Neuro-Flou . . . . .	68
4.10	Taux de reconnaissance des différentes classes par HMM . . . . .	79

# Glossaire

ACI : Analyse en Composantes Indépendantes  
ACP : Analyse en composante principale  
ANFIS : Adaptative Neuro fuzzy inference systeme  
AVC : Accident Vasculaire Cérébrale  
CHMM : continuous hidden markov model  
DHMM : discrete hidden markov model  
ECG : Electrocardiogramme  
EM Expectaion Maximization  
ESV : Extrasystole ventriculaire  
HMM : hidden Markov model  
HTA : hyper tension artérielle  
MAP : maximum a posteriori  
ML : Maximum likelihood  
SIF : systeme d'inférence flou  
SFC : Space Filling Curves

# Introduction Générale

# 1 Contexte

La prise de décision lors d'un diagnostic réalisé par des médecins est une tâche qui nécessite d'une part des données médicales validées par des connaissances des experts en santé et d'autre part l'analyse du nombre important de ces informations qui sont en croissance continue. Cette difficulté confirme le besoin d'automatiser certaines tâches médicales.

Le développement de systèmes d'aide au diagnostic, simulant le raisonnement médical, nécessite la modélisation de ce dernier [5], d'où la collecte de données expertisées par des médecins experts est une phase indispensable afin d'extraire des connaissances qui sont décrites comme un processus interactif de préparation des données brutes (sélection de descripteurs, préparations de base de données, discrétisation des données, filtrage des bruits, etc. . . ), d'extraction de connaissances à l'aide d'algorithmes d'apprentissage, de visualisation et d'interprétation des résultats, lors de l'utilisation par l'expert.

Les erreurs dans le traitement ou la saisie de données peuvent être à l'origine d'importantes erreurs dans la prise de décision ce qui influe sur la fiabilité du système. Une traçabilité concernant des résultats vient pour s'imposer face à ce problème en faisant appel à des méthodes interprétables et probabilistes pour justifier la prise de décision.

Pour exploiter au mieux des données, et en tirer des informations pertinentes, nous utilisons principalement des méthodes statistiques. Cependant la statistique ne se limite pas à l'analyse de données, mais elle s'occupe aussi de leur collecte et de leur interprétation. De plus, la statistique possède deux principales méthodes d'analyse, l'analyse exploratoire des données et l'inférence statistique. Cette dernière est un ensemble de méthodes permettant de tirer des conclusions fiables à partir d'échantillons statistiques (théorie des probabilités) Les approches probabilistes se fondent essentiellement sur l'analyse de données. Elles ont largement été appliquées au diagnostic médical. D'où vient l'intérêt de l'emploi des modèles Markoviens qui constituent un très bon compromis entre richesse d'expression et facilité de mise en œuvre. Logiquement ils se sont avérés comme des modèles les plus performants dans de nombreuses tâches de modélisation et de classification de signaux et de séquences. Les modèles de Markov cachés (Hidden Markov Models : HMM) en générant une probabilité qui maximise une séquence donnée.

Aussi les interprétations fournis par un système d'inférence flou permettent un traitement des données médicales transparent, explicite et proche du raisonnement



humain. Cette qualité fait la puissance des systèmes d'inférence flous dans de nombreux domaines d'application. L'une des motivations pour l'utilisation de systèmes à base de logique floue est due à leurs interprétabilité, ce qui permet de vérifier la plausibilité d'un système, conduisant à une maintenance aisée de ce dernier. Il peut également être utilisé pour acquérir des connaissances à partir d'un problème caractérisé par des données numériques [6].

La modélisation floue, développée par les systèmes basés sur les règles floues linguistiques, nous permet de traiter des systèmes en établissant un modèle linguistique compréhensible par des utilisateurs.

## 2 Problématique

Dans le cadre de cette thèse de doctorat nous faisons des traitement de séquences, réalisés sur deux types de données (signaux ECG et image cytologique) avec des tâches distinctes (classification, segmentation). Nos contributions majeure concernent d'une part la proposition d'un classifieur supervisé pour la reconnaissance des extrasystoles ventriculaires par les modèles de Markov cachés et par un classifieur neuro-floue (ANFIS). Les HMMs sont des modèles probabilistes qui fournissent une information de probabilité sur la classe reconnue tandis que le classifieur neuro-floue (ANFIS) nous donne une transparence dans le système de reconnaissance par les règles générés.

D'autre part, la segmentation pixellaire par les HMMs est appliquée sur des images des cellules cytologiques. La segmentation d'images basée sur les HMMs nécessite tout d'abord la transformation de l'image 2D en une séquence d'observations 1D, pour garder la notion du voisinage de chaque pixel, ensuite une classification pixellaire est réalisée afin d'obtenir une image segmenté ce qui facilite leurs caractérisations et classification future.

## 3 Organisation de la thèse

Les travaux menés dans le cadre de cette thèse et les résultats obtenus sont structurés en cinq chapitres. Le chapitre 2 présente un état de l'art général des travaux qui traitent le problème de la classification automatique et la représentation des données médicales. A ce titre, nous introduisons les différents types de données pouvant être soumis à une approche de classification automatique. Nous présentons également une synthèse des méthodes classiques de classification automatique four-

nies par la littérature.

Le troisième chapitre, est consacré à présenter les concepts théoriques et les outils requis pour la compréhension et l'utilisation des modèles de Markov cachés. Nous exposons en détails les formalismes utilisés et les différentes étapes de l'approche. Nous présenterons également un état de l'art général des travaux qui traite le problème de la classification par les modèles de Markov cachés en domaine biomédical.

Au niveau du quatrième chapitre nous introduisons les termes et concepts essentiels de la logique floue et nous montrons l'importance derrière la mise en œuvre d'un classifieur floue pour fournir des résultats interprétable.

Dans le cinquième chapitre, nous abordons nos contributions relatives à la mise en œuvre, d'une part, de deux classifieurs supervisés dédiés à la reconnaissance des extrasystoles ventriculaire, et d'autre part, la segmentation des images cytologique. Enfin, une conclusion générale synthétise le contenu de cette thèse et présente les perspectives de ce thème de recherche.

# Chapitre 1

## Classification et Représentation des Données Médicales

# 1 Introduction

Avec l'expansion des outils informatiques au cours des dernières décennies, un nombre considérable de données de la vie de tous les jours est mis à disposition des analystes. L'archivage de ces données crée la mémoire de notre société, mais la mémoire ne vaut pas grande chose sans l'intelligence. L'intelligence nous permet de mettre de l'ordre dans notre mémoire en observant des formes, en établissant des règles, en trouvant des nouvelles idées qui méritent d'être essayé et en faisant des prédictions.

La fouille de données (Data Mining en anglais) [7, 8], ou encore analyse intelligente des données, désigne l'ensemble de méthodes destinées à l'exploration et l'analyse des données informatiques, de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information pertinente. La fouille de données permet d'ajouter de l'intelligence aux archives de données pour être capable ensuite de décider.

Avant d'atteindre l'étape de la prise de décision, plusieurs traitements peuvent être réalisés sur les données numérisées, cet ensemble de méthodes de traitement et de classification sont présentées dans ce chapitre.

## 2 Représentation des données

### 2.1 Collecte et prétraitement des données

La première étape est évidemment la collecte des données. Deux cas peuvent se présenter :

1. le modèle doit être conçu à partir d'une base de données préexistante, que l'on ne peut pas enrichir ; là encore, il faut distinguer deux situations :
  - données sont peu nombreuses ; il faut construire des modèles aussi parcimonieux que possible en nombre de paramètres en s'efforçant de tirer la meilleure partie de ces données ;
  - données sont très nombreuses : on peut alors mettre en œuvre des méthodes dites de planification expérimentale ou d'apprentissage actif, afin de ne retenir

que les exemples qui apportent une réelle information au modèle [9].

2. le concepteur du modèle peut spécifier les expériences qui doivent être effectuées pour améliorer le modèle. Dans un tel cas, il est très souhaitable de mettre en œuvre des méthodes de planification expérimentale, surtout si les expériences sont longues ou coûteuses. Les plans d'expérience permettent en effet de limiter le nombre d'expériences, en n'effectuant que celles qui sont réellement utiles pour la conception du modèle.

Une fois les données sont disponibles, il convient de les traiter de manière à rendre la modélisation aussi efficace que possible.

## 2.2 Prétraitement des données

Une fois les données sont disponibles, il faut effectuer un prétraitement qui permette de rendre la modélisation aussi efficace que possible. Ces prétraitements dépendent de la tâche à effectuer et des particularités des données que l'on manipule. Dans tous les cas, le prétraitement minimal consiste à normaliser et à centrer les données, de manière à éviter, par exemple, que certaines variables aient de très grandes valeurs numériques par rapport à d'autres, ce qui rendrait les algorithmes d'apprentissage inefficaces [10]. Le prétraitement le plus simple consiste donc à effectuer le changement de variables suivant, pour les variables  $x$  comme pour la grandeur à modéliser  $y^p$  :

$$u' = \frac{u - \langle u \rangle}{s_u}$$

Où  $\langle u \rangle$  désigne la moyenne de la grandeur  $u$  considérée

$$\langle u \rangle = \frac{1}{N} \sum_{k=1}^N u_k;$$

et  $s_u$  est l'estimateur de l'écart-type de  $u$  :

$$s_u = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (u - \langle u \rangle)^2}$$

### 3 Sélection des variables

la complexité du modèle croît avec le nombre de variables : par exemple, la dimension de Vapnik-Chervonenkis de polynômes de degré  $d$  vaut  $\frac{(n+d)!}{n!d!}$  où  $n$  est le nombre de variables ; elle croît donc très rapidement avec  $n$ . Conserver un contrôle sur le nombre de variables est donc un élément important dans une stratégie de modélisation qui cherche à maîtriser la complexité des modèles.

Il existe de nombreux avantages de la sélection de variables : faciliter la visualisation des données et la compréhension des données, ce qui réduit les exigences en matière de mesure et de stockage, ce qui réduit le temps d'apprentissage, défiant la malédiction de la dimensionnalité pour améliorer les performances de classification.

Les résultats de la sélection de variables sont susceptibles de remettre en cause des idées reçues concernant le phénomène à modéliser, ou, au contraire, de conforter des conjectures ou des intuitions concernant l'influence des variables candidates sur la grandeur à modéliser.

On peut également souhaiter diminuer le nombre de variables en réduisant la dimension de l'espace de représentation de la grandeur que l'on cherche à modéliser. Les principales méthodes utilisées dans ce but sont l'Analyse en Composantes Principales (ACP), l'Analyse en Composantes Indépendantes (ACI, ou ICA pour Independent Component Analysis) ou encore l'Analyse en Composantes Curvilignes (ACC) [10].

### 4 Les différents types de données rencontrés

La classification intervient sur des données qui résultent d'une suite de choix qui vont influencer les résultats de l'analyse. Classiquement, les données sont décrites dans un tableau individus/variables par une valeur unique. On parlera alors de « *tableau de descriptions univaluées ou classiques* ». Dans les applications réelles, où le grand souci est de prendre en compte la variabilité et la richesse d'informations au sein des données, il est courant d'avoir affaire à des données complexes et hétérogènes (ou mixtes). Ce qui se traduit par le fait que chaque case du tableau de descriptions peut contenir non seulement une valeur unique mais également un ensemble de valeurs, un intervalle de valeurs ou une distribution sur un ensemble de valeurs. On dira alors que la classification va porter sur un « *tableau de descriptions symboliques* ».

## 4.1 Description classique d'une variable

Classiquement, une variable  $Y_h$  est définie par une application :

$$Y_h : X \rightarrow O_h$$

$$X_i \in X \rightarrow Y_h(X_i)$$

Où  $X = \{X_1, X_2, \dots, X_n\}$  est l'ensemble des individus. L'ensemble d'arrivée  $O_h$  est appelée domaine d'observation de la variable  $Y_h$ . Un individu est alors décrit sur une variable  $Y_h$  par une valeur unique de  $O_h$ . [11]

On distingue schématiquement deux types de variables : les variables *quantitatives* dites aussi numériques et les variables *qualitatives* dites aussi catégorielles.

### Les variables quantitatives

Une variable quantitative prend des valeurs ordonnées (comparable par la relation d'ordre  $\leq$ ) pour lesquelles des opérations arithmétiques telles que différence et moyenne aient un sens. Une variable quantitative peut être binaire, continue ou discrète. Les variables binaires ne peuvent prendre que deux valeurs, le plus souvent associées à  $\{0,1\}$ ,  $\{\text{absence, présence}\}$  ou  $\{\text{succès, échec}\}$  (*exemple* : le sexe d'un nouveau-né). Les variables continues ou d'échelle sont les variables dont les valeurs forment un sousensemble infini de l'ensemble  $\mathbb{R}$  des réels (*exemple* : la durée d'un intervalle RR, la fréquence d'un signal). Les variables discrètes sont celles dont les valeurs forment un sous ensemble fini ou infini de l'ensemble  $\mathbb{N}$  des entiers naturels (*exemple* : l'âge, le nombre d'enfants) [11].

### Les variables qualitatives

Une variable qualitative (ou aussi catégorielle) est une donnée dont l'ensemble des valeurs est fini. Elle prend des valeurs symboliques qui désignent en fait des catégories ou aussi modalités (*exemple* : le code de la ville, la couleur des cheveux). On ne peut effectuer aucune opération arithmétique sur ces variables. Les variables catégorielles sont parfois ordonnées : on parle alors de variables ordinales (*exemple* : faible, moyen, fort). Les variables ordinales peuvent être rangées dans la famille des variables discrètes et traitées comme telles. Les variables catégorielles non ordonnées sont dites nominales et lorsque ces variables ont pour valeur des textes non codés, écrits en langage naturel, elles sont dites variables textuelles (*exemple* : rapports, nom de film) [11].

## 5 Discrétisation des données

Dans les méthodes purement bayésiennes, les tests donnant des valeurs continues doivent être discrétisés, impliquant ainsi une perte d'information. L'analyse discriminante permet d'éviter cette perte d'information que représente la transformation des variables continues en variables discrètes et évite ainsi de créer des incertitudes supplémentaires. Certains auteurs estiment au contraire, qu'une discrétisation des variables permet de se rapprocher du raisonnement de l'expert mais elle pose le problème de l'interprétation des valeurs frontières et nécessite d'intégrer des graphes de distribution des fréquences [12].

## 6 Intelligence artificielle

### 6.1 Classification automatique

La classification automatique cherche à identifier une typologie des individus en groupes homogènes et bien séparés. Appliquée aux séquences de données temporelles, les individus manipulés sont cette fois des séquences dont on cherche à mettre en relation statistique leurs différents enregistrements. De façon générale, la classification automatique est utile comme préalable à d'autres opérations de fouille de données. En effet, une démarche fréquente du data mining est de trouver des classes d'individus qui ont le même type de comportement, afin d'ensuite pouvoir identifier la classe à laquelle appartient un nouvel individu en exploitant son historique (classement) et évaluer ainsi son comportement futur (prévision).

### 6.2 Classification et apprentissage

Le terme classification est souvent associé à la notion d'apprentissage. En effet on associe souvent l'action de classer à celle d'apprendre à classer. Selon [13] qui ont un point de vue axé apprentissage, le terme classification peut avoir deux significations distinctes. La classification peut être l'action de regrouper en différentes catégories des objets ayant certains points communs ou faisant partie d'un même concept, on parle alors de problème de clustering et apprentissage non supervisé. Cependant la classification peut aussi correspondre à l'action d'affecter des objets à des catégories bien définies, on parle dans ce cas de problème d'affectation et d'apprentissage supervisé. L'apprentissage non supervisé consiste donc à découvrir des catégories d'objets (ou clusters) alors que l'apprentissage supervisé consiste à mettre à jour une fonction



d'affectation d'objets à des catégories et à l'utiliser en vue de d'affecter de nouveaux objets. Il convient de noter que la notion d'apprentissage n'est pas nécessairement présente dans les méthodes de classification.

### 6.3 Le problème de la classification

Le problème de la classification étant d'organiser un ensemble d'objets en classes homogènes, il nous faut donc définir ce que sont une partition, une classe (on utilisera également la notion de groupe) et ce que sont les éléments que l'on cherche à classer.

#### Le problème de la généralisation

L'objectif de la classification est de fournir une procédure ayant un bon pouvoir prédictif c'est-à-dire garantissant des prédictions fiables sur les nouveaux exemples qui seront soumis au système. La qualité prédictive d'un modèle peut être évaluée par le risque réel ou l'espérance du risque, qui mesure la probabilité de mauvaise classification d'une hypothèse  $h$ .

#### Évaluation d'une hypothèse de classification

Pour avoir une estimation non optimiste de l'erreur de classification, il faut recourir à une base d'exemples qui n'ont pas servi pour l'apprentissage : il s'agit de la base de test.

La base de test contient elle aussi des exemples étiquetés qui permettent de comparer les prédictions d'une hypothèse  $h$  avec la valeur réelle de la classe. Cette base de test est généralement obtenue en réservant une partie des exemples supervisés initiaux et en ne les utilisant pas pour la phase d'apprentissage [14].

#### Partition et classe

La question se pose alors de savoir ce qu'est une classe et quel type de structure va être utilisé pour modéliser l'espace des objets. D'un point de vue mathématique, une classe est un sous-ensemble de l'ensemble  $E$  des objets à classer. Il existe deux approches différentes pour décrire une classe : l'approche générative et l'approche discriminative. La première décrit une classe par les propriétés caractéristiques des objets qui la composent alors que la seconde décrit une classe par sa frontière avec ses voisines. La structure la plus couramment utilisée est celle de la partition qui

se définit de la façon suivante :

**Définition :** L'ensemble  $P = C_1, \dots, C_k$  est une partition de l'ensemble  $E$  en  $k$  classes si et seulement si :

1.  $C_i \neq \emptyset$  pour  $i = 1, \dots, k$ ,
2.  $\bigcup_{i=1}^k C_i = E$ ,
3.  $C_i \cap C_l = \emptyset$  pour tout  $i \neq l$

## 6.4 Clustering et apprentissage non-supervisé

La classification non supervisée désigne un corpus de méthodes ayant pour objectif de dresser ou de retrouver une typologie existante caractérisant un ensemble de  $n$  observations, à partir de  $p$  caractéristiques mesurées sur chacune des observations. Par typologie, on entend que les observations, bien que collectées lors d'une même expérience, ne sont pas toutes issues de la même population homogène, mais plutôt de  $K$  populations. Deux exemples peuvent être considérés :

- l'ensemble des clients d'une banque est une collection de  $n$  observations, chacune de ces observations étant caractérisée par la nature des  $p$  transactions bancaires qu'elle réalise. Il existe certainement différents  $K$  profils types de clients. L'objectif est alors d'une part de retrouver ces profils types à partir de l'information sur les transactions bancaires, et d'autre part de déterminer, pour chaque observation, à quel profil type elle correspond.
- une cohorte de patients représente un ensemble de  $n$  observations, chacune décrite par  $p$  mesures physiologiques. Bien qu'ayant tous la même pathologie, ces patients n'ont pas tous le même historique médical. On souhaite donc dans un premier temps établir une typologie de ces patients en  $K$  groupes selon leurs caractéristiques physiologiques. Dans un deuxième temps, on étudiera si la réponse au traitement diffère pour des patients issus de groupes différents.

## 6.5 L'apprentissage supervisé

L'apprentissage supervisé, utilise des exemples étiquetés ou classés. Ces étiquettes ou ces classes peuvent être vues comme fournies par un expert ou un superviseur, d'où le nom d'apprentissage supervisé. Le but de l'apprentissage est alors de produire une fonction de classification, appelée hypothèse, permettant de déterminer la classe d'un exemple.

L'apprentissage supervisé a donc pour but de déterminer une représentation en intention d'un concept (l'hypothèse) à partir d'un sous ensemble de son extension (les exemples). Il réalise un saut inductif en passant des exemples particuliers à une fonction de classification générale.

Autrement dit, étant donné un ensemble d'exemples sous la forme  $\{(x_i, y_i)\}$  avec  $x_i$  des objets d'un domaine  $D$  et  $y_i$  les classes associées, le but de l'apprentissage supervisé est de déterminer une hypothèse  $h$  tel que  $y = h(x)$  pour tout  $x$  de  $D$ . Les  $x_i$  sont appelés les observations.

Notons que l'on utilise classiquement le même terme d'exemples pour désigner d'une part les exemples d'apprentissage qui servent à fabriquer une hypothèse, et d'autre part les objets sur lesquels cette hypothèse sera appliquée pour déterminer leur classe. En cas d'ambiguïté, nous appellerons ces objets des nouveaux exemples à classer, pour les différencier des exemples d'apprentissage.

## 7 La classification des données séquentielles

Plusieurs techniques de classification automatique des données séquentielles ont été développées ces dernières années. Elles ont été appliquées dans différents domaines comme l'analyse de séquences biologiques [15], l'étude de la mobilité des objets dans les vidéos [16], ou d'autres domaines pour lesquels l'individu joue un rôle central, à savoir la classification des patients sur la base de leurs groupes sanguins [15], la modélisation du comportement d'utilisateurs sur le Web [17], etc. Parmi cette variété de méthodes de classification, les plus répandues sont les approches markoviennes.

## 8 Aspects probabilistes d'une classification automatique

L'aspect probabiliste d'une classification se rapporte à l'évaluation d'une similarité ou d'une association entre 2 entités comme un objet et un attribut, ou une classe et un paramètre de fonction de distribution. Le calcul est donc souvent indépendant de la représentation des données. En effet, on trouve des coefficients d'association probabiliste aussi bien avec des modèles de vecteurs de caractéristiques qu'avec des descriptions d'attributs symboliques.

Il est presque impossible de présenter un algorithme unique ou générique concernant

la description probabiliste d'extraction de classes puisque chaque méthode pose ses propres hypothèses de fonctions de distribution et de dépendances conditionnelles avec son lot d'hypothèses de simplification pour éviter l'explosion exponentielle de recherche des paramètres.

## 8.1 Modélisation probabiliste en classification

Les premières approches qui ont été proposées en classification étaient algorithmiques, heuristiques ou géométriques et reposaient essentiellement sur la dissimilarité entre les objets à classer.

L'approche statistique, plus récente, se base sur des modèles probabilistes qui formalisent l'idée de classe. Cette approche permet en outre d'interpréter de façon statistique la classification obtenue.

Les modèles probabilistes prennent en compte des fonctions de distribution paramétriques (fonction gaussienne, fonctions multinomiales...) exigeant de trouver les paramètres optimum (moyenne, écart-type, poids...) pour maximiser la vraisemblance. Par exemple une classe sera représentée par une fonction de probabilité  $f$ , le but sera d'estimer des paramètres concernant les classes grâce à la vraisemblance (probabilité d'avoir la totalité d'un système de variables) en estimant les paramètres grâce aux données observées.

## 9 Interprétabilité

Le concept de l'interprétabilité apparaît dans la littérature sous différents noms : transparence, intelligibilité, compréhensibilité. L'interprétabilité d'un système qui est la facilité d'analyse de son comportement, dépend à la fois du nombre de ses variables et la complexité du modèle, ainsi qu'aux mécanismes d'apprentissage.

L'interprétabilité représente la force motrice la plus importante derrière la mise en œuvre des systèmes de classification automatique pour les problèmes d'application médicale.

L'expert devrait être capable de comprendre le classifieur et d'évaluer ses résultats, elle permet de vérifier la plausibilité d'un système, conduisant à une maintenance aisée de ce dernier. Il peut également être utilisé pour acquérir des connaissances à partir d'un problème caractérisé par des exemples numériques [6]. Une amélioration de l'interprétabilité peut augmenter les performances de généralisation lorsque l'ensemble des données est petit.

## 10 Conclusion

L'automatisation de la prise de décision exige un traitement et une analyse des données collectées en allant vers une classification via différentes méthodes. Ce chapitre a été consacré à une présentation détaillée des techniques de traitement des données médicales d'une part et d'autre par des outils d'intelligence artificielle dans le but d'aide au diagnostic.

## Chapitre 2

# La théorie des Modèles de Markov cachés

# 1 Introduction

Les modèles de Markov cachés (HMMs : Hidden Markov Models) constituent des outils statistiques qui fondent sur une théorie solide. Ils se caractérisent par une grande efficacité et souplesse dans les algorithmes qu'ils offrent pour la modélisation des problèmes traités.

Ces modèles sont récemment largement utilisés en domaine de reconnaissance profitant de leurs grand succès dans la reconnaissance de la parole où les HMMs sont initialement et fréquemment employés.

Grâce à leur relative simplicité, robustesse et élasticité, les HMMs ont été utilisés dans plusieurs domaines. Parmi ces domaines : la modélisation des langages, la modélisation de la météorologie ou de climats, la cardiologie, l'évaluation des scénarios lors de la simulation du déplacement d'un robot, la théorie de codage, la modélisation des finances, le contrôle biologique, la bioinformatique, . . .etc. Le but de ce chapitre est de présenter les concepts théoriques et les outils requis pour la compréhension et l'utilisation des HMMs. Il est organisé comme suit :

La première partie rappelle les différentes définitions nécessaires pour la compréhension des concepts liés aux HMMs. La deuxième partie présente l'application de cet outil en domaine médical.

## 2 Définitions

Les modèles de Markov cachés sont une famille d'outils mathématiques probabilistes parfaitement adaptés à la modélisation de séquences temporelles. Il existe plusieurs types de modèles de Markov cachés afin de répondre à des problèmes spécifiques. Dans le cadre de cette études et plus particulièrement de ce chapitre, nous nous intéresserons principalement aux modèles de Markov cachés discrets du premier ordre, que nous abrègerons par la suite HMM. Pour pouvoir présenter les HMM, il est nécessaire de commencer par présenter les modèles de Markov et les propriétés qui leur sont associées.

## 2.1 Les chaines de Markov discrètes

En calcul de probabilités, on définit une variable aléatoire (*v.a.*) réelle comme une fonction mesurable  $X : \Omega \rightarrow \mathbb{R}$ .  $\Omega$  est appelé l'univers. Dans de nombreux cas de figure,  $\Omega$  est l'ensemble des réels  $\mathbb{R}$ , l'ensemble des entiers positifs ou un de leurs sous ensembles.

**Définition 1** un processus stochastique est une famille  $\{X_t\}_{t \in \mathbb{T}}$  de v.a. définies sur  $\Omega$  ( $X_t : \Omega \rightarrow \mathbb{R}$ ).

L'ensemble  $\mathbb{T}$  représente souvent la notion de temps mais il peut également correspondre à la notion de position spatiale de dimension 2 ou à toute autre notion en autant de dimension que nécessaire. Dans le cas où  $\mathbb{T}$  représente la notion de temps et si  $\mathbb{T}$  est discret, on parle de processus stochastique en temps discret, tandis que le processus est dit en temps continu, lorsque  $\mathbb{T}$  est continu.

**Définition 2** Les états d'un processus stochastique défini par les v.a.  $X_t : \Omega \rightarrow \mathbb{R}$  pour tout  $t \in \mathbb{T}$  sont les valeurs prises par ces v.a. lorsque  $t$  varie. On note  $\mathbb{S}$  l'ensemble des « états » du processus.

A.A. Markov fut le premier à étudier et à poser les bases mathématiques permettant l'étude des chaines qui portent son nom. La définition de ces chaines est la suivante :

**Définition 3** Un processus  $\{S_t\}_{t \in \mathbb{T}}$  ( $S_t : \Omega \rightarrow \mathbb{S}$ ) est une chaine de Markov s'il vérifie les trois conditions suivantes :

1.  $\mathbb{T}$  est dénombrable ou fini. Dans ce cas et pour simplifier les notations ultérieurs, il est toujours possible de prendre  $\mathbb{T} \subseteq \mathbb{N}^* = \{1, 2, \dots\}$ . Cette condition signifie que le processus ne change de valeur qu'à des instants déterminés à priori.
2. L'ensemble  $\mathbb{S}$  des états du processus est dénombrables. Dans la suite, supposons également que  $\mathbb{S}$  est fini. Nous pouvons alors définir  $\mathbb{S} = \{s_1, \dots, s_N\}$  cet ensemble.
3. Le processus est associé à une fonction de probabilité  $P$  vérifiant la propriété markovienne : «la probabilité que le processus soit dans un état particulier à un instant  $t$  ne dépend que de l'état dans lequel se trouve le processus au temps  $t - 1$ ». Soit  $Q = (q_t)_{t \in \mathbb{T}}$  une suite d'états du processus ( $q_t \in \mathbb{S}$ ). La propriété de Markov vérifie la relation suivante, pour toute suite d'états  $Q$  et pour tout



instant  $t \in \mathbb{T}$  :

$$P(S_t = q_t / S_{t-1} = q_{t-1}, \dots, S_1 = q_1) = P(S_t = q_t / S_{t-1} = q_{t-1}) \quad (2.1)$$

La probabilité  $P(S_t = q_t / S_{t-1} = q_{t-1})$  correspond à la probabilité de transition de l'état  $q_{t-1}$  à l'instant  $t - 1$  vers l'état  $q_t$  à l'instant  $t$ .

**Définition 4** Une chaîne de Markov est homogène (dans le temps) si et seulement si les probabilités de transition ne dépendent pas du temps  $t$  (les probabilités de transition sont stationnaire), c'est-à-dire que pour tout  $(t, t') \in \mathbb{T}^2$ , on a :

$$P(S_{t+1} = s_j / S_t = s_i) = P(S_{t'+1} = s_j / S_{t'} = s_i) \quad (2.2)$$

On note  $a_{i,j}$  cette probabilité.

Une chaîne de Markov homogène est donc totalement définie par la donnée des états, des probabilités des états initiaux  $\Pi$  et des probabilités de transitions entre états  $A$  avec :

$$\Pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} = (\pi_1, \dots, \pi_N)' \quad \text{et } \pi_i = P(S_1 = s_i) \quad (2.3)$$

$$A = (a_{i,j})_{1 \leq i, j \leq N} \quad \text{et } a_{i,j} = P(S_{t+1} = s_i / S_t = s_j) \quad (2.4)$$

**Définition 5** Vecteurs et matrices stochastiques :

- Un vecteur  $V = (v_1, \dots, v_N)$  de dimension  $N$  (ou, de manière équivalente, son transposé) est stochastique si et seulement si :
  - Pour tout  $i = 1 \dots N$ ,  $0 \leq v_i \leq 1$ .
  - $\sum_{i=1}^N v_i = 1$
- Une matrice  $M = (m_{i,j})_{1 \leq i, j \leq N}$  de dimension  $N \times N$  est dite stochastique si et seulement si :
  - pour tout  $i = 1 \dots N$  et  $j = 1 \dots N$ ,  $0 \leq m_{i,j} \leq 1$ ,
  - pour tout  $i = 1 \dots N$ ,  $\sum_{j=1}^N m_{i,j} = 1$ .

Caractéristiques d'une chaîne de Markov :

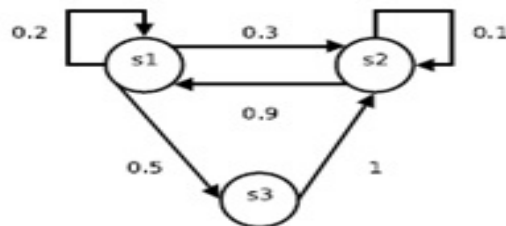
- Une matrice est stochastique si et seulement si les lignes qui la composent sont des vecteurs stochastiques.
- Le système est forcément dans un et un seul état particulier au départ donc  $\Pi$  est un vecteur stochastique.  $A$  est une matrice stochastique car, en partant d'un état  $s_i$  à l'instant  $t$ , le processus transite forcément vers l'un des  $N$  états du système au temps  $t + 1$ .
- A tout couple formé d'un vecteur stochastique  $V$  de dimension  $N$  et d'une matrice stochastique  $M$  de dimension  $N \times N$ , il est possible d'associer une chaîne de Markov caractérisée par le couple  $(V, M)$ .

**Définition 6** Une chaîne de Markov peut être représentée graphiquement. Pour cela, on a associé à la chaîne de Markov  $\{S_t\}_{t \in \mathbb{T}}$  un graphe  $G$  dont l'ensemble des sommets est en bijection avec l'ensemble des états  $S$  et dont l'ensemble des arcs  $U$  (orientés dans le sens des transitions) est défini par

$$(s_i, s_j) \in U \iff a_{i,j} > 0 \tag{2.5}$$

Afin de simplifier les notations, l'ensemble des sommets du graphe  $G$  sera représenté par l'ensemble  $S$ .

La figure 4.1 représente la représentation graphique associée à la chaîne de Markov  $(\Pi, A)$ .



$$A = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.9 & 0.1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \Pi = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.4 \end{pmatrix}$$

FIGURE 2.1 – Représentation graphique de la chaîne de Markov  $(\Pi, A)$ .

Les chaînes de Markov peuvent servir à modéliser de nombreux processus stochastiques.

Voici quelques exemples :

- Modélisation des jours de pluie et de soleil sous certains climats. Pour cela, on suppose qu'une journée ne peut être que pluvieuse ou qu'ensoleillé et on estime avec une distribution de Bernoulli les probabilités en fonction du jour  $t$  : soleil avec une probabilité  $p(t)$  et pluie avec une probabilité  $1 - p(t)$  avec l'hypothèse que le temps au jour  $t$  ne dépend que du temps au jour  $t - 1$  et que, sur une période assez longue, il y a une certaine stabilité statistique du temps (homogénéité du processus).
- Modélisation du nombre d'absents et de présents dans une entreprise. Ce nombre change aléatoirement chaque jour. On a  $S_t = S_{t-1} + \Delta S_t$ . On suppose alors que  $\Delta S_t$  ne dépend que de  $\Delta S_{t-1}$ .
- Modélisation stochastique de la génération de la parole utilisée en reconnaissance de la parole par ordinateur. Par exemple,  $S_t$  correspond à la classification grammaticale (substantif, adjectif, verbe, pronom, ...) du  $t$ -ième mot de la phrase.
- Modélisation et traitement d'images numérique,  $S_t$ , correspondant à l'état d'un pixel (son niveau de gris), est supposé ne dépendre que des quelques pixels précédents.

Cependant, dans certains cas, ces modèles ne permettent pas d'exprimer le comportement du système avec suffisamment de précision. Pour améliorer cette précision, les modèles de Markov cachés ont été développés.

## 2.2 Les modèles de Markov cachés discrets

### Définitions

Un modèle de Markov caché discret correspond à la modélisation de deux processus stochastique : un processus caché parfaitement modélisé par une chaîne de Markov discrète et un processus observé dépendant des états du processus caché.

**Définition 7** soit  $\mathbb{S} = \{s_1, \dots, s_N\}$  l'ensemble des  $N$  états cachés du système. Soit  $\mathbb{S} = (S_1, \dots, S_T)$  un  $T$ -uplet de *v.a.* définies sur  $\mathbb{S}$ . Soit  $\mathbb{V} = \{v_1, \dots, v_M\}$  l'ensemble de  $M$  symboles émissibles par le système. Soit  $\mathbb{V} = \{V_1, \dots, V_T\}$  un  $T$ -uplet de *v.a.* définies sur  $\mathbb{V}$ .

Un modèle de Markov caché discret du premier ordre est alors défini par les probabilités suivantes :

- Les probabilités d'initialisation des états cachés :  $P(S_1 = s_1)$
- Les probabilités de transition entre états cachés :  $P((S_t = s_j)/(S_{t-1} = s_i))$ .
- Les probabilités d'émission des symboles dans chaque état caché :  $P((V_t = v_j)/(S_t = s_i))$ .

Si le modèle est de Markov caché est stationnaire alors les probabilités de transition entre états cachés et les probabilités d'émission des symboles dans chaque état caché sont indépendantes du temps  $t > 1$ . On peut alors définir, pour tout  $t > 1$  quelconque,  $A = (a_{i,j})_{1 \leq i,j \leq N}$  avec  $a_{i,j} = P(S_t = s_j / S_{t-1} = s_i)$ ,  $B = (b_i(j))_{1 \leq i \leq N, 1 \leq j \leq M}$  avec  $b_i(j) = P(V_t = v_j / S_t = s_i)$  et  $\Pi = (\pi_1, \dots, \pi_N)$  avec  $\pi_i = P(S_1 = s_i)$ .

Un modèle de Markov caché stationnaire du premier ordre est défini par le triplet  $(A, B, \Pi)$ . Par la suite, nous utiliserons la notation  $\lambda = (A, B, \Pi)$  et nous emploierons le terme HMM pour des modèles de Markov caché du premier ordre. Les relations de dépendance entre les différentes variables aléatoires d'un HMM sont schématisées par la figure 2.2. Dans cette présentation, les flèches partent de la *v.a.* qui conditionne et se termine au niveau de la variable aléatoire conditionnée. Dans la figure, seuls les transitions au temps  $t - 1$ , et  $t + 1$  sont représentées.

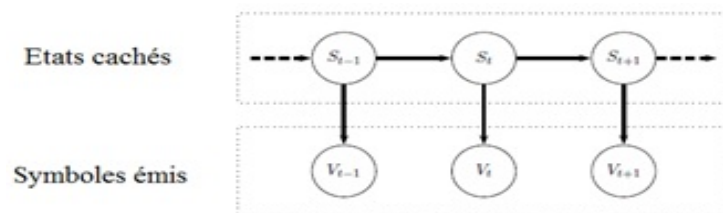


FIGURE 2.2 – Relations de dépendance entre les variables aléatoires d'un HMM.

On note que  $Q = (q_1, \dots, q_T) \in \mathbb{S}^T$  une séquence d'états cachés et  $O = (o_1, \dots, o_T) \in \mathbb{V}^T$  une séquence de symboles observés. La probabilité de réalisation de la séquence d'états cachés  $Q$  et la séquence d'observation  $O$  par rapport au *HMM*  $\lambda$  est alors

$$P(V = O, S = Q / A, B, \Pi) \tag{2.6}$$

Ou plus simplement

$$P(V = O, S = Q/\lambda) \quad (2.7)$$

En utilisant les dépendances de probabilités conditionnelles, on déduit que :

$$P(V = O, S = Q/\lambda) = P(V = O, S = Q/\lambda) P(S = Q/\lambda) \quad (2.8)$$

De plus,

$$P(V = O, S = Q/\lambda) = \prod_{t=1}^T P(V_t = o_t/S_t = q_t, \lambda) \quad (2.9)$$

$$P(S = Q/\lambda) = P(S_1 = q_1/\lambda) = \prod_{t=1}^{T-1} P(S_{t+1} = q_{t+1}/S_t = q_t, \lambda) \quad (2.10)$$

A partir d'un *HMM*  $\lambda$ , d'une séquence d'états cachés  $Q$  et d'une séquence d'observations  $O$ , il est possible de calculer l'adéquation entre le modèle  $\lambda$  et les deux séquences  $Q$  et  $O$ . Pour cela, il suffit de calculer la probabilité que la séquence d'observation  $O$  ait effectivement été engendrée par le modèle  $\lambda$  en suivant la séquence d'états cachés  $Q$ .

Lorsque la séquence d'états cachés n'est pas connue, il est possible d'évaluer la vraisemblance d'une séquence d'observations  $O$  par rapport à un modèle  $\lambda$ . La vraisemblance correspond à la probabilité  $P(V = O/\lambda)$  que la séquence d'observations ait été engendrée par le modèle pour l'ensemble des séquences d'états cachés possibles. On remarque alors que la formule suivante est vérifiée :

$$P(V = O/\lambda) = \sum_{Q \in \mathcal{S}^T} P(V = O, S = Q/\lambda) \quad (2.11)$$

Les *HMM*, pour être utiles, nécessitent la résolution de plusieurs problèmes principaux : le calcul de l'évaluation, le décodage/ segmentation de séquence d'observation et l'apprentissage.

### 2.3 Évaluation de la probabilité d'une séquence d'observations

Il existe plusieurs techniques d'évaluer la vraisemblance d'une séquence d'observation, parmi lesquelles on peut citer l'évaluation directe et l'évaluation par les fonctions Forward backward [18].

La question posée est : Étant donné une séquence d'observations  $O = \{o_1, \dots, o_n\}$  et un modèle  $\lambda = A, B, \pi$  comment peut-on calculer efficacement la probabilité de la séquence  $O$  par rapport au modèle  $\lambda$ , c'est-à-dire  $P(O/\lambda)$ .

### Évaluation directe

L'évaluation directe consiste à déterminer  $P(O/\lambda)$ , en sommant sur tous les chemins d'états possibles  $Q = q_1, \dots, q_N$  des probabilités conjointes de  $O$  et de  $Q$  soit :

$$P(O/\lambda) = \sum_Q P(O, Q/\lambda) = \sum_Q P(O/Q, \lambda)P(Q/\lambda) \quad (2.12)$$

La probabilité d'emprunter le chemin  $Q$  et en utilisant les hypothèses Markoviennes est :

$$P(Q/\lambda) = P(s_1, \dots, s_n|\lambda) = \pi_1 \times a_{12} \times a_{23} \cdots \times a_{n-1n} \quad (2.13)$$

Et la probabilité pour que cette séquence d'états  $Q$  émette les observations de la séquence.

$O$  est :

$$P(O/Q, \lambda) = P(O/S_1, S_2 \dots, S_n, \lambda) = b_1(o_1) \times b_2(o_2) \cdots \times b_n(o_n) \quad (2.14)$$

(Hypothèse d'indépendance des observations entre elles conditionnellement à la séquence d'états cachés  $Q$ ) :

On obtient finalement :

$$P(O/\lambda) = \sum_Q \pi_1 \times a_{12} \times b_1(o_1) \times a_{23} \times b_2(o_2) \times \cdots \times a_{n-1n} \times b_n(o_n) \quad (2.15)$$

Pour un modèle à  $N$  états et pour  $T$  observations, cet algorithme nécessite  $(2T - 1)N^T$  multiplications et  $N^T - 1$  additions. Soit de  $2TN^T$  opérations, c'est-à-dire une complexité en  $o(N^T)$ . Ce qui est pratiquement incalculable pour des valeurs de  $T$  raisonnables. Par exemple, pour  $N = 5$  états et  $T = 100$  observations, il faut, approximativement,  $2.100.5^{100} \approx 10^{72}$  opérations.

### Procédure Forward-Backward

Constitue une autre manière d'évaluation de la probabilité  $P(O/\lambda)$ , le principe considère que l'observation peut se faire en deux étapes suivantes :

La variable Forward : représente l'émission de la séquence d'observations  $O = o_1, \dots, o_t$  et la réalisation de l'état  $q_t$  au temps  $t$ , soit :

$$\alpha_t(i) = P(O, q_t = s_i/\lambda) \tag{2.16}$$

Qui constitue la probabilité de générer  $O = o_1, \dots, o_t$  et de se trouver dans l'état  $q_t$  à l'instant  $t$  (Figure 2.3).

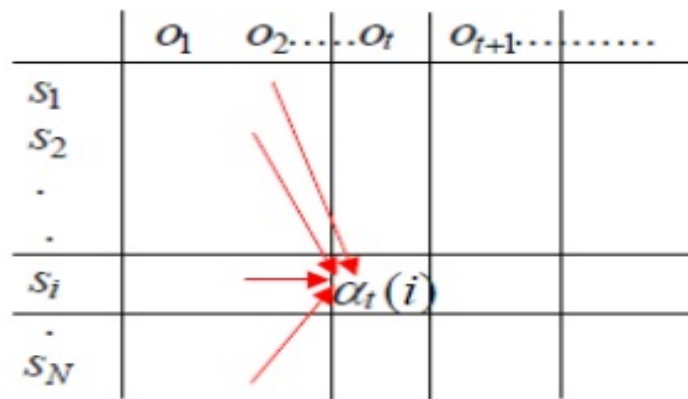


FIGURE 2.3 – Calcul de la variable Forward.

$$\alpha_t(i) \xrightarrow{\text{Forward}} \alpha_{t+1}(j) \tag{2.17}$$

La probabilité  $\alpha_t(i)$  est calculée de manière récursive comme suit (Algorithme 1) :

---

**Algorithm 1** Algorithme de Forward.

---

**Pour**  $i = 1$  à  $N$  **Faire**

$$\alpha_1(i) = \pi_i b_i(o_1)$$

**Fin pour**

**Pour**  $t = 1$  à  $T - 1$  **Faire**

**Pour**  $j = 1$  à  $N$  **Faire**

$$\alpha_{t+1}(j) = (\sum_{i=1}^N \alpha_t(i) a_{ij}) b_j(o_{t+1})$$

**Fin pour**

**Fin pour**

**Sortie :**  $P(V = O/\lambda = \sum_{i=1}^N \alpha_T(i))$

---

La variable Backward : représente l'émission de la séquence d'observations

$O = o_{t+1}, \dots, o_T$  en partant de l'état  $q_t$  au temps  $t$ , soit :

$$\beta_t(i) = P(O, q_t = s_i / \lambda) \quad (2.18)$$

Qui constitue la probabilité de générer  $O = o_{t+1}, \dots, o_T$  en sachant que l'on est dans l'état  $q_t$  à l'instant  $t$

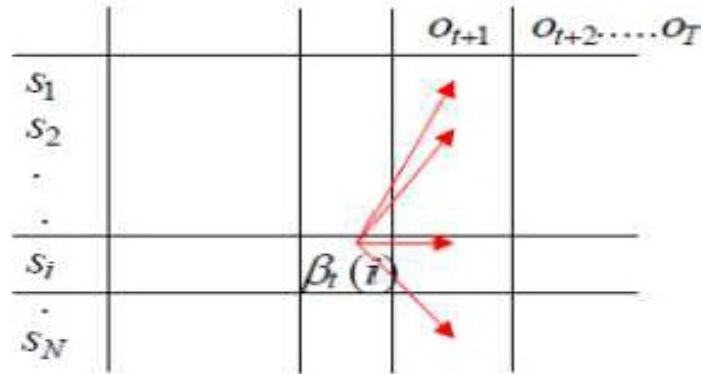


FIGURE 2.4 – Calcul de la variable Backward.

$$\beta_t(i) \xrightarrow{\text{Backward}} \beta_{t+1}(j) \quad (2.19)$$

On déduit  $\beta_t(i)$  et  $\beta_{t+1}$  par l'algorithme suivant (Algorithme 2) :

---

**Algorithm 2** Algorithme de Backward.

---

**Pour**  $i = 1$  à  $N$  **Faire**

$$\beta_T(i) = 1$$

**Fin pour**

**Pour**  $t = T - 1$  à  $1$  **Faire**

**Pour**  $i = 1$  à  $N$  **Faire**

$$\beta_t(i) = \left( \sum_{j=1}^N a_{ij} \beta_{t+1}(j) b_j(o_{t+1}) \right)$$

**Fin pour**

**Fin pour**

**Sortie :**  $P(V = O / \lambda = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i))$

---

L'algorithme *Forward – backward* considère que l'observation peut se faire par l'émission au début de la séquence  $O(1 : t)$  et d'aboutir à  $q_t$  à l'instant  $t$ , puis, émission de la fin de l'observation  $O = (t + 1 : T)$  en partant de l'état  $q_t$  au temps  $t$ , connaissant  $\lambda$ . Le calcul de  $\alpha_t(i)$  se fait avec  $t$  croissant tandis que celui de  $\beta_t(i)$  se fait avec  $t$  décroissant, d'où l'expression *Forward – Backward*.



$P(O/\lambda)$  peut être défini à chaque instant  $t \in [1, T]$  par :

$$P(O/\lambda) = \sum_{i=1}^N \alpha_t(i) \times \beta_t(i) \tag{2.20}$$

Dans le cas où  $(t = 0)$   $(T = 0)$  on obtient :

$$P(O/\lambda) = \sum_{i=1}^N \alpha_T(i) = \sum_{i=1}^N \pi_i \beta_0(i) \tag{2.21}$$

En totalité, le nombre d'opérations dans ces deux algorithmes, se réduit à  $N(N + 1)(T - 1) + N$  multiplications et  $N(N - 1)(T - 1)$  additions soit une complexité en  $o(N^2T)$ . Ainsi, pour l'exemple de l'évaluation direct précédent ( $N = 5, T = 100$ ), nous avons environ 3000 opérations à exécuter, contre 1072 opérations par la méthode directe, ce qui permet donc un gain très important en complexité de calcul.

## 2.4 Décodage/ segmentation de séquences d'observations

Le décodage d'un modèle consiste à déterminer le chemin correspond à l'observation  $O = o_1, \dots, o_t$ , c'est-à-dire de chercher et trouver la meilleure suite d'états  $Q = q_1, \dots, q_t$  qui maximise la vraisemblance de la séquence d'observations  $O = o_1, \dots, o_t$  (Figure 2.5)

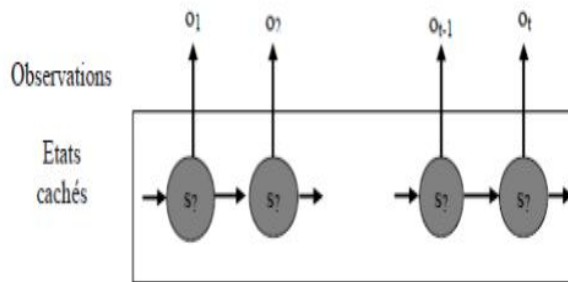


FIGURE 2.5 – Représentation de décodage.

Pour résoudre ce problème on a recourt à l'algorithme de Viterbi [19, 20]. Cette solution est la plus utilisée. Elle est basée sur les techniques de programmation dynamique. Viterbi est un algorithme récursif qui permet de trouver à partir d'une suite d'observations, une solution optimale au problème d'estimation de la suite d'états.

- Probabilité du meilleur début de chemin de taille  $t$  finissant par  $s_i$  est donnée par :

$$\delta_t(i) = \max_I P(o_1, o_2, \dots, o_t, s_t = s_i | \lambda) \quad (2.22)$$

- Règle d'induction

$$\delta_1(i) = \pi_i * b_i(o_1) \quad (2.23)$$

$$\delta_{t+1}(i) = \max_{j=1..n} (\delta_t(j) * a_{j,i}) * b_i(o_{t+1}) \quad (2.24)$$

On sauve aussi, à chaque  $t$ , l'état optimal qui mène à l'état  $i$  au temps  $t+1$

$$\Psi_{t+1}(i) = \operatorname{argmax}_{j=1..n} (\delta_t(j) * a_{j,i}) \quad (2.25)$$

On stocke ainsi les  $n - 1$  meilleurs états successifs intermédiaires

Résultat final : Prendre le chemin qui maximise  $\delta_n(i)$  (Algorithme 3)

---

**Algorithm 3** Algorithme de Viterbi.

---

**Pour**  $i = 1$  à  $N$  **Faire**

$$\alpha_1(i) = \pi_i b_i(o_1)$$

**Fin pour**

**Pour**  $t = 2$  à  $T$  **Faire**

**Pour**  $j = 1$  à  $N$  **Faire**

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} \{\delta_{t-1}(i) a_{ij}\}$$

$$\delta_t(j) = \max_{1 \leq i \leq N} \{\delta_{t-1}(i) a_{ij}\} b_j(o_t) = \delta_{t-1}(\psi_t(j)) a_{\psi_t(j)j} b_j(o_t)$$

**Fin pour**

**Fin pour**

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} \{\delta_T(i)\}$$

$$O(V = O, S = Q^*) = \operatorname{argmax}_{1 \leq i \leq N} \{\delta_T(i)\} = \delta_T(q_T^*)$$

**Pour**  $t = T - 1$  à  $1$  **Faire**

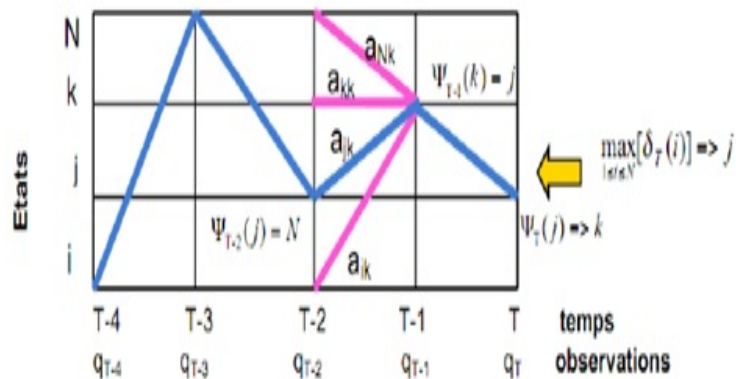
$$q_t^* = \psi_{t+1}(q_{t+1}^*)$$

**Fin pour**

---

Cet algorithme est basé sur la programmation dynamique qui optimise le coût d'une trajectoire entre deux points. Si une trajectoire peut à chaque instant prendre une valeur parmi  $N$ , pour chaque point intermédiaire, chaque sous-trajectoire est aussi optimale. La figure 2.6 ci-dessous présente un schéma illustratif de cet algorithme. Ainsi, le calcul direct d'une trajectoire de longueur  $L$  qui peut prendre  $N$  valeurs aurait une complexité en  $N_L$ , mais le calcul d'une demi-trajectoire est en  $N_{L/2}$ . En itérant  $L$  fois ce processus, on obtient

une approche "diviser pour régner" dont le coût est en  $LN_2$ .



en arriere.jpg

FIGURE 2.6 – Le retour en arrière.

## 2.5 Maximisation de la vraisemblance

Le critère de maximum de vraisemblance consiste à trouver le modèle  $\lambda^*$  maximisant la probabilité  $P(O/\lambda)$  [21]. En général, il n'est pas possible de trouver ce modèle optimal. Néanmoins, pour tenter de résoudre ce problème, il existe principalement une méthode : utiliser l'algorithme Expectation-Maximisation.

### Introduction à l'algorithme Expectation Maximization

L'algorithme Expectation Maximization (*EM*) est une méthode générale d'optimisation en présence d'information incomplète. L'algorithme permet, à partir d'un modèle initial  $m'$ , de trouver  $n$  modèle  $m$  qui augmente la vraisemblance. Dans cette section, nous ne démontrerons pas l'algorithme *EM*, nous nous contenterons juste d'exposer les principes et formules qui nous seront nécessaires par la suite. Particulièrement bien adapter à des probabilités, l'algorithme *EM* repose sur deux hypothèses principales :

- Maximiser  $P(X = x/M = m)$  est équivalent à maximiser  $\ln P(X = x/M = m)$  .
- L'introduction de variables non observées ou cachées  $Y$  définies sur  $\mathcal{Y}$  dans l'expression de la vraisemblance permet d'effectuer les calculs plus facilement.

Dans le cas de variables aléatoires discrètes, on définit  $\Gamma_x(m, m')$  par :

$$\Gamma_x(m, m') = \sum_{y \in \mathcal{Y}} P(Y = y/X = x, M = m') \ln P(X = x, Y = y/M = m)$$

$=E_{y \in \mathbb{Y}} [\ln P(X = x, Y = y / M = m) / X = x, M = m']$  (2.26) Avec  $E_{\mathbb{Y}}[f]$  l'espérance mathématique de  $f$  sur l'ensemble  $\mathbb{Y}$ .

L'algorithme *EM* [22, 23] consiste donc à construire, à partir d'un modèle initial  $m_1$ , une suite de modèles  $(m_t)_{t>0}$  vérifiant

$$\Gamma_x(m_{t+1}, m_t) \geq \Gamma_x(m_t, m_t) \quad (2.27)$$

Une condition suffisante est alors de rechercher le modèle  $m_{t+1}$  qui maximise la fonction  $\Gamma_x(m_{t+1}, m_t)$ . La suite  $(m_t)_{t>0}$  vérifie, pour tout  $t > 1$  et  $m_{t+1} m_t$ , la relation

$$P(X = x, M = m_{t+1}) > P(X = x, M = m_t) \quad (2.28)$$

L'une des plus célèbres applications de l'algorithme *EM* est l'algorithme Baum-Welch permettant l'apprentissage des *HMM* [24].

### L'algorithme Baum-Welch

Dans le cas des *HMM*, on cherche à maximiser  $P(O/\lambda)$  où  $O$  désigne une séquence de  $T$  observations. En appliquant l'algorithme *EM* à la maximisation de cette probabilité [24], on est amené à maximiser  $\Gamma(\lambda, \lambda')$ , avec  $\lambda = (A, B, \Pi)$  le nouveau modèle et  $\lambda'$  le modèle connu (ou actuel) :

$$\Gamma_O(\lambda, \lambda') = \sum_{Q \in \mathbb{S}^T} P(S = Q / V = O, \lambda') \ln P(V = O, S = Q / \lambda) \quad (2.29)$$

En effectuant les différents calculs, on obtient :

$$\pi_i = P(S_1 = s_i / O, \lambda') \quad (2.30)$$

$$a_{i,j} = \frac{\sum_{t=1}^{T-1} P(S_t = s_i, S_{t+1} = s_j / V = O, \lambda')}{\sum_{t=1}^{T-1} P(S_t = s_i / V = O, \lambda')} \quad (2.31)$$

$$b_i(j) = \frac{\sum_{t=1}^T P(S_t = s_i / V = O, \lambda') \delta(o_t = j)}{\sum_{t=1}^T P(S_t = s_i / V = O, \lambda')} \quad (2.32)$$

Les formules de ré-estimation obtenus ci-dessus peuvent s'interpréter de la façon suivante :

$$\pi_i = \text{Probabilit  d'etre dans l'etat } s_i \text{ l'instant } t = 1 \quad (2.33)$$

$$a_{i,j} = \frac{\text{nombre de transition de l'etat } s_i \text{ vers l'etats}_j}{\text{nombre de fois où l'on quitte l'etat } s_i} \quad (2.34)$$

$$b_i(j) = \frac{\text{nombre d'apparitions simultanées de l'etat } s_j \text{ et du symbole } v_k}{\text{nombre d'apparition de l'etat } s_j} \quad (2.35)$$

Les probabilités utilisées pour la ré-estimation des matrices peuvent être obtenues par les algorithmes *Forward* et *Backward*. L'algorithme de Baum-Welch [25] est donné par l'algorithme 4. Sa complexité est  $O(N^2T + NMT)$ .

---

**Algorithm 4** Algorithme Baum-Welch.
 

---

Choisir un modèle initial  $\lambda_0$

$t = 0$

**Répéter**

$t \leftarrow t + 1$

Calculer les variable *Forward* et *Backward* pour le modèle  $\lambda_{t-1}$

Calculer  $\Pi$  de  $\lambda_t$

Calculer  $A$  de  $\lambda_t$

Calculer  $B$  de  $\lambda_t$

**Tant que**  $P(V = O/\lambda_t \Rightarrow P(V = O/\lambda_{t-1}))$  et  $(t > t_{max})$

---

**Remarques :**

- Le choix du modèle initial influe sur les résultats. Par exemple, si certaines valeurs de  $A$  et  $B$  sont égales à 0 au départ, elles le resteront jusqu'à la fin d'apprentissage. Ceci permet en particulier de garder la topologie des modèles notamment pour la structure gauche- droite qui est très utilisée en reconnaissance de l'écriture.
- L'algorithme converge vers des valeurs de paramètres qui assurent un maximum local de  $P(O/l)$ . Il est donc important, si l'on veut être aussi près que possible du maximum global, de bien choisir la structure et les valeurs initiales des paramètres.
- Le nombre d'itérations est fixé empiriquement. L'expérience prouve que, si le point précédent a été correctement traité, la stabilisation des paramètres ne correspond pas à un sur apprentissage : il n'y a donc en général pas besoin de contrôler la convergence par un ensemble de validation.
- Il existe d'autres méthodes d'apprentissage basé sur le principe *EM*, parmi ces méthodes on cite l'apprentissage Viterbi : qui représente une variante modifiée de la procédure de Baum-Welsh dans laquelle on réalise le comptage par l'algorithme de Viterbi plutôt que par les fonctions Backward-Forward.

Dans ce cas, on ne s'intéresse plus à tous les chemins ayant généré une observation, mais seulement au plus probable.

### 3 Application des HMMs dans le domaine biomédical

#### 3.1 Classification des arythmies cardiaque par les HMM dans la littérature

L'emploi des HMM en classification est une méthode générative consistant à entraîner (algorithme de Baum-Welch) dans un premier temps un modèle par classe de battement puis, pour une série d'observations donnée, de déterminer quel est le modèle parmi ceux précédemment établis le plus à même de produire cette séquence d'observations, enfin pour un modèle donnée il est possible de retrouver (algorithme de Viterbi) la séquence d'états cachés la plus à même de produire la séquence d'observations en question.

Le cas général d'un HMM est celui où l'automate des états cachés forme un graphe entièrement connexe. En reconnaissance de battement cardiaque, on contraindra le modèle à un graphe où les états transitent uniquement de gauche à droite avec le temps [26] (Figure 2.7)

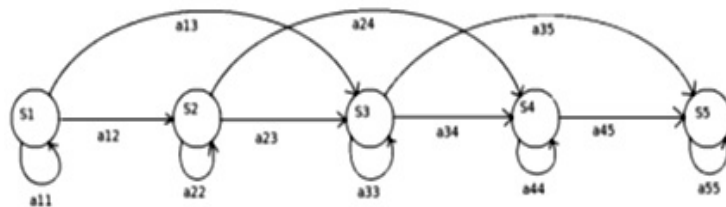


FIGURE 2.7 – Exemple d'HMM à cinq états de gauche à droite.

Les HMM sont largement employés dans le traitement des signaux biomédicaux [27–29] et plus particulièrement dans l'analyse de l'ECG.

A. Rodrigo et al. [26] ont implémenté un modèle markovien incrémental pour l'analyse du signal ECG. Les HMM sont utilisés pour modéliser le signal ECG comme une séquence d'ondes élémentaires connectées. Une fois que ces modèles sont disponibles, un processus d'adaptation est mis en œuvre pour les adapter au signal ECG de chaque individu en particulier, grâce à un apprentissage incrémental des HMM en ligne. Les méthodes abordées sont ba-

sés sur le maximum de vraisemblance (ML) et maximum a posteriori (MAP) comme critères d'optimisation. L'évaluation des performances a été faite pour la détection d'ischémie et ils ont obtenu une sensibilité de 86%. S.Krimi et al. [30] ont proposé une méthode qui consiste en deux étapes : à la première étape ; l'extraction d'un ensemble de caractéristiques, y compris ceux morphologiques, et la deuxième étape est d'utiliser les arbres HMM pour la classification des arythmies cardiaques, les performances ont été mesurées en termes de spécificité et sensibilité où ils ont obtenus 97,46% et 99,22%.

Selon le type de la probabilité de distributions utilisées dans les HMM, elles peuvent être classées comme modèle de Markov caché continu (CHMM) et modèle de Markov caché discret (DHMM). Le DHMM fournit des résultats de reconnaissance plus stables et un apprentissage plus rapide avec la précision de reconnaissance qui n'est pas moins de CHMM [31]. Dans ce même travail, les auteurs ont utilisé un DHMM pour la classification de 4 anomalies cardiaques et ils ont obtenu un taux de reconnaissance de 73%. Thomas et al. [32], 2006, ont utilisé une approche d'apprentissage pour la segmentation du signal ECG consiste à construire un modèle pour le signal, et en utilisant la séquence d'état la plus probable pour une séquence d'observation donnée afin de trouver les transitions d'onde. Pour être en mesure de segmenter un signal ECG quelconques avec une grande précision, ils ont mis en œuvre une approche multi-HMM. Cette méthode consiste à effectuer un regroupement bayésien de la base d'apprentissage [33]. La base d'apprentissage est divisée en  $K$  classes de ECG qui ont des similitudes et  $K$  HMM sont formés pour l'apprentissage et exploités pour fournir de grandes capacités de généralisation et une grande précision en même temps. Cet algorithme de clustering HMM est tout simplement une variation de l'algorithme K-Means, où les clusters sont définies par HMM plutôt que par les centres dans l'espace de données [32].

## 4 Conclusion

Les approches probabilistes permettent d'interpréter de façon statistique la classification obtenue. Les approches markoviennes se fondent sur des bases mathématiques solides, qui sont explicitement abordées dans ce chapitre d'une part, et d'autre part des concepts théoriques ainsi que des outils requis pour la compréhension et l'utilisation des HMMs ont été aussi présentés.

# Chapitre 3

## Classification Floue des données médicales



# 1 Introduction

La logique floue a pour objectif l'étude de la représentation des connaissances imprécises, des raisonnements approchés [34] et elle cherche à modéliser les notions vagues du langage naturel pour pallier l'inadéquation de la théorie des ensembles classiques dans ce domaine [35].

La logique floue fut développée par Lofti A. Zadeh en 1965 à partir de sa théorie des sous-ensembles flous [36]. Les sous-ensembles flous sont une manière mathématique de représenter l'imprécision de la langue naturelle, ils peuvent être considérés comme une généralisation de la théorie des ensembles classiques [36] [37] [38] [39] [40] [35].

Dans ce chapitre nous introduisons les termes et concepts essentiels de la logique floue et nous montrons l'importance derrière la mise en œuvre d'un classifieur floue pour fournir des résultats interprétable.

## 2 Principe de la logique floue

### 2.1 Notion d'appartenance partielle

Dans la théorie des ensembles, un élément appartient ou n'appartient pas à un ensemble.

La théorie des ensembles flous repose sur la notion d'appartenance partielle : chaque élément appartient partiellement ou graduellement aux ensembles flous qui ont été définis. Les contours de chaque ensemble flou 3.1 ne sont pas « nets », mais « flous » ou « graduels ».

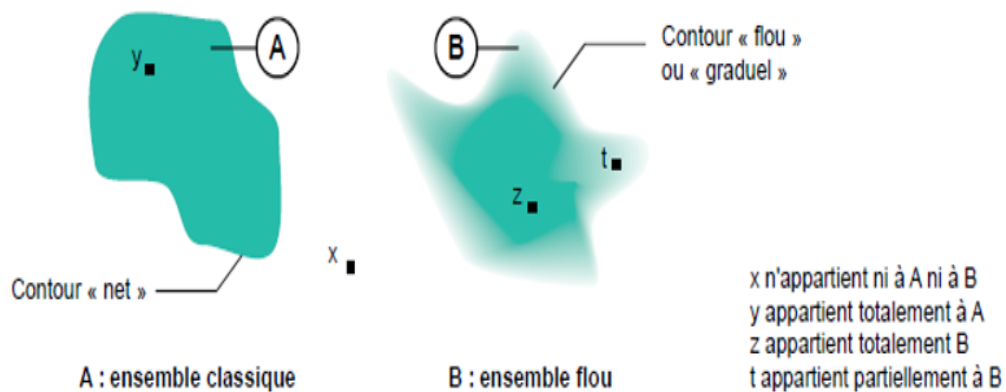


FIGURE 3.1 – Comparaison d'un ensemble classique et d'un ensemble flou

## 2.2 Fonctions d'appartenance

Soit  $X$  un ensemble, un sous-ensemble  $A$  de  $X$  est caractérisé par une fonction d'appartenance  $\mu_A : X \rightarrow [0, 1]$ , cette fonction d'appartenance est l'équivalent de la fonction caractéristique d'un ensemble classique. Le choix de la fonction d'appartenance est fait de façon arbitraire au fil des tests on choisit si c'est la forme sigmoïde, hyperbolique, tangente, exponentielle, gaussienne, triangulaire ou autres, qui convient le mieux à notre problème. Prenons comme exemple l'évaluation de risque d'AVC, il nous faudra redéfinir des fonctions d'appartenance pour chaque sous-ensemble flou de chacune des trois variables :

- **Input1** : Données biochimiques sanguine (glycémie, urémie, ...)  
**Sous ensemble** : { Bonnes – mauvaises - Très mauvaise }
- **Input2** : Comportement à risque (Tabagisme – sédentarité -alcoolisme)  
**Sous ensemble** : { Fréquent, occasionnel, rare }
- **Input3** : Anomalies cardiologiques (Hyper Tension Artérielle (HTA), essoufflement, tachycardies ...)  
**Sous-ensemble** : { Existante – inexistante – épisodique }
- **Input4** : Données divers (âges, sexe, antécédent)  
**Sous-ensemble** : { Elevé – Très élevé - modéré – bas }
- **Output** : Risque de survenue d'attaque cérébrale  
**Sous-ensemble** : { important – modéré – faible }

Dans ce même exemple, l'HTA constitue un risque minime en dessous de 12.7 et élevée au-dessus de cette valeur. On souhaite caractériser le risque d'AVC en répondant par exemple à la question « L'HTA est-elle élevée ? ».

En logique booléenne, la réponse à cette question s'énonce de la manière suivante 4.2 :

- Le risque est élevé à 100% au-dessus de 20.10 et à 0% en dessous.
- L'HTA est normale en dessous de 13.7. On peut donc dire qu'en dessous de ce chiffre, Le risque est bas avec un taux de confiance de 0 %.
- L'HTA est élevée au-dessus de 20.10, le risque est grand avec un taux de

confiance de 100 % au-dessus de ces chiffres.

- Aux stades intermédiaires, on considère que le risque est moyen à 50 % de confiance à 16.9 et à 25 % de confiance à 15.3.

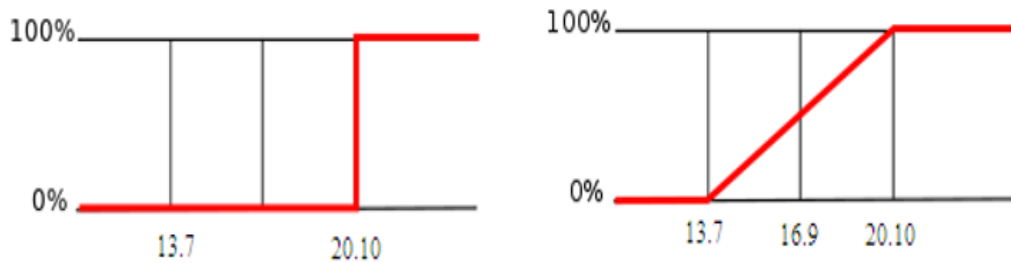


FIGURE 3.2 – Comparaison entre la formalisation en logique booléenne (à gauche) et en logique floue (à droite)

Soit  $X$  un ensemble,  $\mu_A$  est la fonction d'appartenance caractérisant  $A$  lesous-ensemble flou de  $X$  ;

- **La hauteur** : La hauteur de  $A$  notée  $h(A)$ , correspond à la borne supérieure de l'ensemble d'arrivée de sa fonction d'appartenance

$$h(A) = \text{Sup} \{ \mu_A(x) | x \in X \}$$

- **La normalité** :  $A$  est dit normaliser, si et seulement si la hauteur  $h(A)$  est égale à 1. En pratique, il est extrêmement rare de travailler sur des ensembles flous non normalisés.

- **Le support** : Le support de  $A$  est l'ensemble des éléments de  $X$  appartenant au moins un peu à  $A$ . Autrement dit, c'est l'ensemble

$$\text{Sup}(A) = \{ x \in X | \mu_A(x) > 0 \}$$

- **Le noyau** : le noyau de  $A$  c'est l'ensemble des éléments de  $X$  appartenant totalement à  $A$ , autrement dit c'est l'ensemble des noyaux :

$$\text{Noy}(A) = \{ x \in X | \mu_A(x) = 1 \}$$

Par construction  $\text{Noy}(A) \subset \text{Sup}(A)$

- **Une  $\alpha$  coupe** : C'est un sous-ensemble classique des éléments ayant un degré d'appartenance supérieure ou égale à  $\alpha$  :

$$\alpha - \text{coupe}(\alpha) = \{ x \in X | \mu_A(x) > \alpha \}$$

Remarquons que si  $A$  était un ensemble classique, nous aurions simplement  $\text{Sup}(A) = \text{Noy}(A)$  et  $h(A) = 1$  (ou  $h(A) = 0$  si  $A = \emptyset$ ). Donc on retrouve bien les propriétés usuelles des ensembles classiques.

## 2.3 Les variables floues

A l'inverse de la logique classique où les variables binaires sont définies par deux états « vrai » ou « faux », Les variables floues supportent toute une gradation entre ces deux possibilités, l'exemple qui suit permet de mieux comprendre la différence entre les variables binaires et les variables floues : Si l'on désire classer un groupe de personne suivant leur activité physique journalière en définissant la catégorie des sédentaires avec en moyenne 30 mn de marche, la catégorie de actifs avec en moyenne 75 mn de marche et enfin la catégorie des dynamiques avec plus de 75 mn de marche par jour. La logique binaire donne la représentation de la figure 4.3 pour les trois variables « sédentaire », « actif » et « dynamique ».

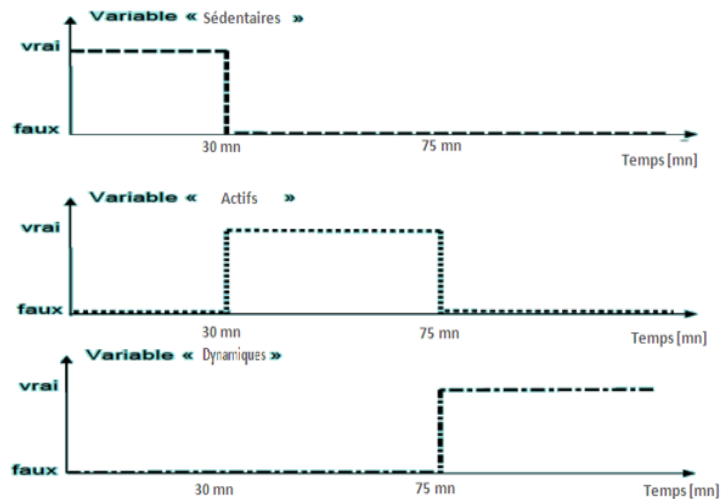


FIGURE 3.3 – Représentation de l'activité journalière d'un individu(logique classique)

De prime abord nous relevons deux failles :

- On représente l'état de la variable à l'aide de son degré de vérité, en associant la valeur 1 (100%) à l'état « vrai » et la valeur nulle (0%) à l'état « faux ».

- On constate que cette façon de faire, est très éloignée de ce qu'on fait . L'être humain lorsqu'il résout ce genre de problème ne fait pas naturellement une distinction franche entre les variables en termes clairs Il utilise des expressions du genre « je suis plutôt actif » pour qualifier une activité quotidienne légèrement inférieure à 75 mn, et « plutôt dynamique » pour une activité légèrement supérieure à cette valeur.

En conclusion, la logique binaire malgré sa simplicité, elle reste assez éloignée de la logique utilisée naturellement par l'être humain.

En logique floue, les variables ne sont plus binaires mais présentent une multitude de valeurs possibles entre le « vrai » et le « Faux » (Figure4.4).

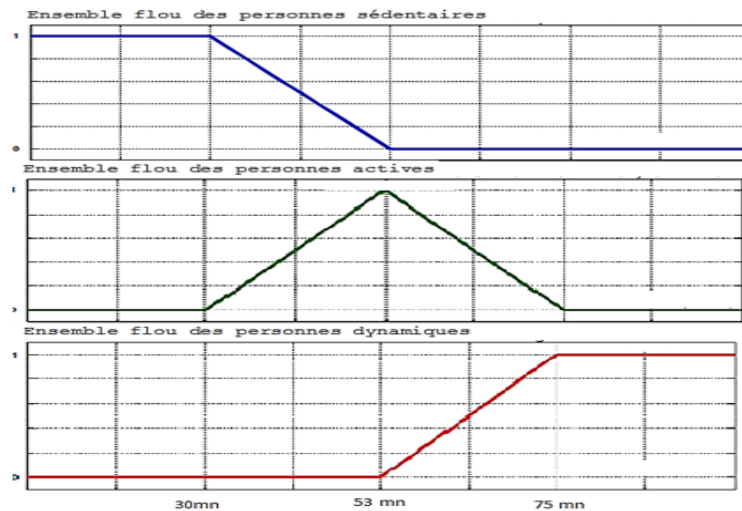


FIGURE 3.4 – Représentation de l'activité journalière d'un individu(logique floue)

### Les variables linguistiques

Les variables linguistiques sont fortement liées au concept de fonction d'appartenance vu précédemment qui permet d'établir le lien entre elles et la logique floue.

Dans un système flou, les variables linguistiques sont représentées par un triplé  $(V, X, T_v)$ , où  $V$  est une variable,  $X$  est la plage des valeurs de la variable et  $T_v$  un ensemble fini ou infini de sous ensembles flous.

### Exemple

$V$  : Comportement à risques (Tabagisme – sédentarité – alcoolisme...).

$X$  :  $[1, 30]$  jours.

$T_v$  : Rare –fréquent- chronique

### Les opérateurs flous

Appliquer les opérateurs de la théorie des ensembles classique sur la logique floue nécessite certaines modifications afin de les adapter aux fonctions d'appartenance propres à celle-ci et dont les valeurs varient entre l'intervalle 0 et 1, comme nous pouvons le voir sur le tableau 3.1

TABLE 3.1 – Les opérateurs logiques

Dénomination	Intersection ET : $\mu_{A \cap B}(x)$	Réunion OU : $\mu_{A \cup B}(x)$	Complément NON : $\mu_{\bar{A}}(x)$
Opérateurs de Zadeh MIN/MAX	$\min(\mu_A(x), \mu_B(x))$	$\max(\mu_A(x), \mu_B(x))$	$1 - \mu_A(x)$
Probabiliste PROD/PROBOR	$\mu_A(x) \times \mu_B(x)$	$\mu_A(x) + \mu_B(x)$ $\mu_A(x) \times \mu_B(x)$	$1 - \mu_A(x)$

Les propriétés de commutativité, distributivité et associativité sont les mêmes pour les opérateurs de la logique classique à deux exception près.

- **Le tiers exclus contredit** : en logique floue  $A \cup \bar{A} \neq X$ , c.à.d  $\mu_{A \cup \bar{A}} \neq 1$ .
- **L’ubiquité** : en logique floue un élément peut appartenir à  $A$  et  $\bar{A}$  au même temps, cela donne  $A \cap \bar{A} \neq \emptyset$ , autrement dit  $\mu_{A \cap \bar{A}} \neq 0$ , notons que ces éléments correspondent à l’ensemble à  $Sup(A) - Noy(A)$ .

## 2.4 Les règles d’inférence

On appelle règles d’inférence, l’ensemble des différentes règles reliant les variables floues d’entrée d’un système aux variables floues de sortie de ce système. Ces règles sont représentées sous la forme :

**Si** condition 1 et/ou condition 2 (et/ou...) **Alors** action sur les sorties.

### Exemple

**Si** la tension artérielle est très élevée ou il y a des antécédents familiaux d’AVC **alors** le risque est majeur

**Si** la tension artérielle est bonne et pas de comportements à risque **alors** le risque est mineur

**Si** la tension artérielle est moyennement élevée et pas d’antécédents familiaux ou tabagisme **alors** le risque est moyen.

Ces règles résument en fait l’expérience de l’expert, le médecin dans notre cas, de ce fait elles ne sont pas immuables.

$$\left\{ \begin{array}{l} Si...alors...ou \\ Si...alors...ou \\ \dots \end{array} \right.$$

Bien que l'opérateur **ou** ne soit pas explicitement mentionné.

### 3 Système D'inférence Floue

Un classifieur est considéré comme « opaque » lorsque sa modélisation n'est pas interprétable et que la façon dont il est parvenu à ce résultat ne peut être expliquée a posteriori [41]. On parle aussi de boîte noire, ce qui est le cas des réseaux de neurones. Les modèles flous, sont très souvent utilisés en intelligence artificielle notamment pour la reconnaissance ou la classification. D'où l'intérêt de leurs applications dans le domaine médicale, notamment lorsque le mécanisme de décision doit être interprétable pour l'utilisateur. Un Système d'Inférence Floue (SIF) a comme but de transformer les données d'entrée en données de sortie à partir de l'évaluation d'un ensemble de règles. Les entrées sont issues du processus de fuzzification et l'ensemble de règles normalement sont définies par le savoir-faire de l'expert. Un SIF (Figure 4.5) est constituée de trois étapes : a) Fuzzification, b) Inférence et c) Défuzzification.

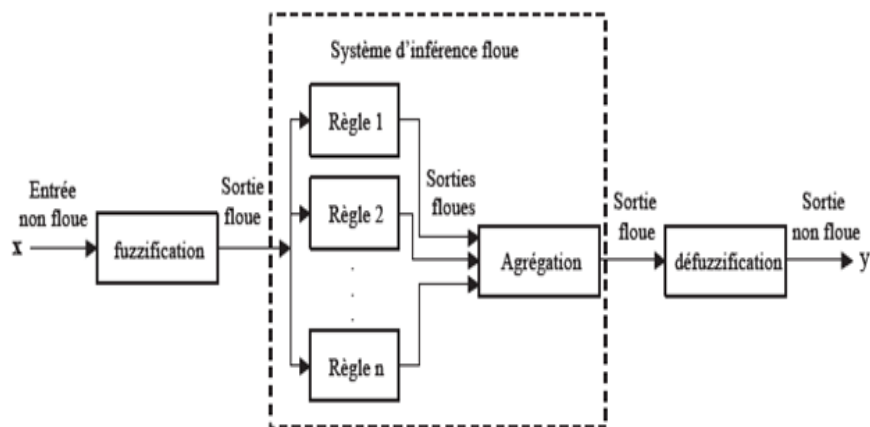


FIGURE 3.5 – Structure d'un SIF

Un système d'inférence floue peut être classifié selon quatre catégories présentées dans le tableau 4.2 suivant [3] :

TABLE 3.2 – La classification des paramètres du SIF [3]

Classe	Paramètre
Logique	Le mécanisme de raisonnement Les opérateurs flous Les types des fonctions d'appartenance La méthode de déffuzzification
Structurelle	Le nombre des fonctions d'appartenance, des règles, des variables linguistiques
Connective	Antécédents et conséquences des regles.
Opérationnel	Les valeurs des fonctions d'appartenances.

- Les paramètres logiques : incluent les opérateurs de la logique floue appliquée pour les méthodes d'implication, d'agrégation et de defuzzification ainsi que le type des opérateurs adaptés pour And, Or . . .
- Les paramètres structurels : incluent le nombre des variables participant à l'inférence, le nombre des fonctions d'appartenance qui définit chaque variable linguistique ainsi que le nombre des règles utilisées.
- Les paramètres connectifs : définis la topologie du système .Ceci incluent la connexion entre les différentes instances linguistiques (les antécédents, conclusions, et les poids des règles).
- Les paramètres opérationnels : Ces paramètres définissent la correspondance entre la représentation linguistique et numérique des variables.

Un système d'inférence flou, aussi appelé : le modèle flou, le système à base de règles floues, ou : le système flou est construit a partir de trois composants conceptuels :

- *Base de règles* : Une base des règles floues est composée de règles qui sont généralement déclenchées en parallèle.
- *Base de données* : définit les fonctions d'appartenance utilisées pour les règles floues.
- *Le mécanisme de raisonnement* : représente la procédure d'inférence appelée aussi la généralisation du *modus-ponens* ou raisonnement approximative.

### 3.1 Principe de fonctionnement du SIF

Généralement, un système d'inférence flou comprend trois étapes [1], décrit dans la figure 4.7



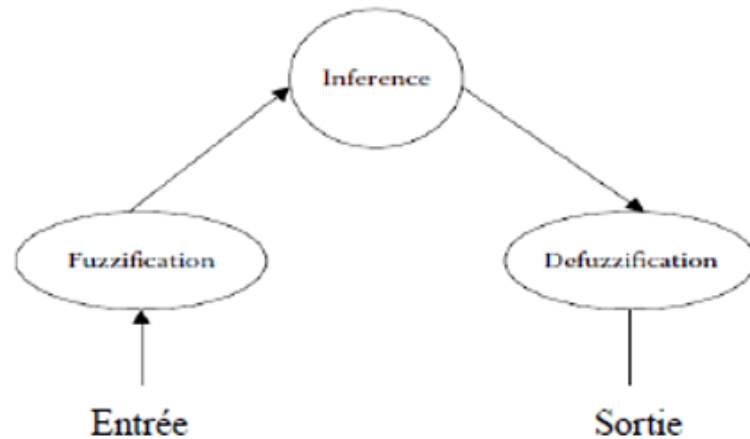


FIGURE 3.6 – Système d'inférence flou [1]

### Fuzzification

L'opération de fuzzification permet de passer du domaine réel au domaine du flou. Elle consiste à déterminer le degré d'appartenance d'une valeur (mesurée par exemple) à un ensemble flou. Par exemple Une donnée provenant d'un capteur pourrait être : tension artérielle = 13.8, après Fuzzification on aurait donc tension = 55% moyenne ,42 % Bonne.

### Le moteur d'inférence

La deuxième étape est le moteur d'inférence, qui est un mécanisme permettant de condenser l'information d'un système au travers d'un ensemble de règles définies pour la représentation d'un problème quelconque. Chaque règle délivre une conclusion partielle qui est ensuite agrégée aux autres règles pour fournir une conclusion (agrégation).

### La Défuzzification

C'est l'étape qui permet de fusionner les différentes commandes générés par le moteur d'inférence pour lui donner qu'une seule commande de sortie et de transformer cette variable linguistique qualitative de sortie en donnée numérique quantitative , ce qui revient donc à trouver la meilleure valeur quantitative en fonction des fonctions d'appartenances des variables linguistiques.

### 3.2 Type des SIF

Il existe principalement deux types de systèmes flous :

#### Système flou de Mandani

Dans ce type de systèmes flous, la prémisse et la conclusion sont floues. Après la réalisation de l'inférence floue, une étape du « défuzzification » est obligatoire pour le passage du symbolique au numérique. Un exemple de ce SIF est illustré dans la figure4.8 [2].

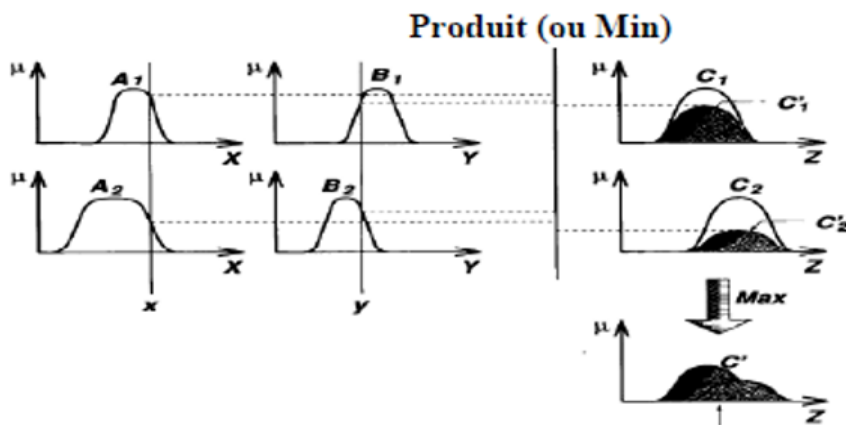


FIGURE 3.7 – Exemple du SIF de Mandani [2]

#### Système flou de Takagi-Sugeno

Ce modèle est proposé par Takagi, Sugeno et Kang [39]. Il est composé d'une base de règles floues de la forme :

Si  $x$  est  $A_i$  et  $y$  est  $B_i$  Alors  $z = f(x, y)$ . Avec  $f$  est une fonction nette.

Dans ce type de SIF, la conclusion correspond a une constante ou une expression polynomiale .Il permet d'obtenir directement la sortie défuzzifiée à partir des règles linguistiques. La figure 4.9 montre un exemple de SIF de Sugeno.

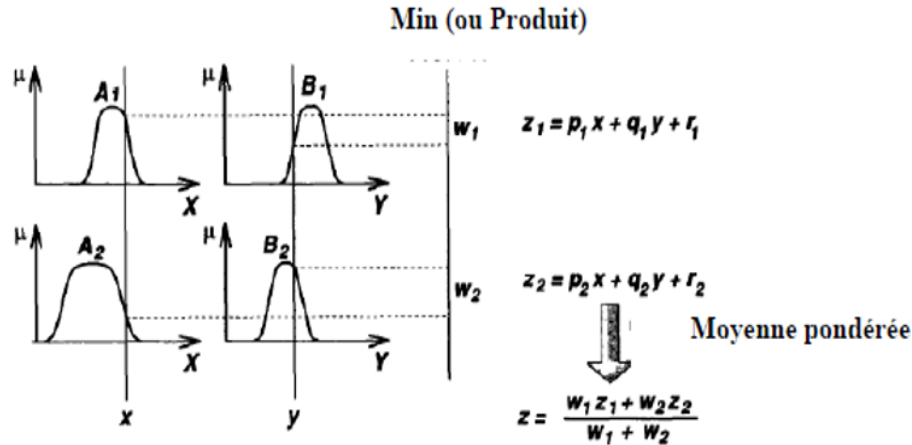


FIGURE 3.8 – Exemple du SIF de Sugeno [2]

Les systèmes flous sont connues par leurs capacités de traitement qui se fait au niveau symbolique ainsi que leurs représentation de l'imprécision et l'incertitude d'un expert humain [42]. Ceci a augmenté le nombre des domaines d'application qui utilisent la logique floue parmi les quelles, on trouve la médecine.

La logique floue et les réseaux de neurones forment aujourd'hui la base de la majorité des systèmes intelligents d'aide au diagnostic et à la décision. Il serait intéressant de fusionner les deux pour exploiter la richesse des deux approches.

## 4 Vers les systèmes Neuro-Flous

La logique floue et les réseaux de neurones artificiels sont des technologies complémentaires. Ces techniques ont chacune leurs forces et leurs faiblesses. La combinaison de ces deux approches dans un système intégré semble plus prometteuse pour le développement des systèmes intelligents [43].

### 4.1 Hybridation Neuro-floue

#### Types d'hybridation

Il existe plusieurs méthodes pour mettre en œuvre la technique neuro-floue de modélisation. En général, l'hybridation neuro-floue est faite de

deux façons [44] :

- Un réseau de neurones équipé de la capacité de manutention du traitement de l'information floue, est appelé un réseau de neurones flou (FNN)
- Un système flou combiné avec les réseaux de neurones pour l'amélioration de certaines de ses caractéristiques, comme la flexibilité, la vitesse et la capacité d'adaptation, est appelé un système neuro flou (NFS)

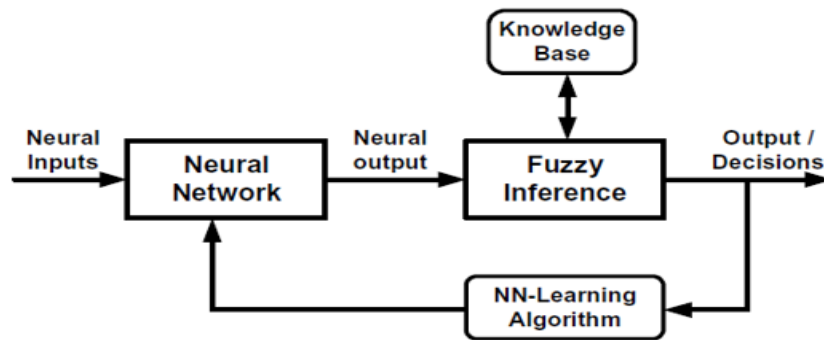


FIGURE 3.9 – System Neuro-Flou (NFS)

Dans le cadre de notre travail nous nous intéresserons au second modèle (Figure 4.10), le bloc de réseau de neurones entraîne le système d'inférence floue pour générer les décisions correspondantes.

## 4.2 Cas du modèle ANFIS

Le modèle ANFIS est un réseau de neurones flou proposé par Jang [45]. Ce réseau est composé de cinq couches représentées sur la Figure 4.11. ci-dessous avec deux entrées X et Y. Pour simplifier la présentation de l'architecture d'ANFIS, deux règles floues « si alors » du modèle de Sugeno de premier ordre sont considérées :

Règle 1 : Si  $(x \text{ est } A_1)$  et  $(y \text{ est } B_1)$  alors  $(f1 = a_1x + b_1y + c_1)$

Règle 2 : Si  $(x \text{ est } A_2)$  et  $(y \text{ est } B_2)$  alors  $(f2 = a_2x + b_2y + c_2)$

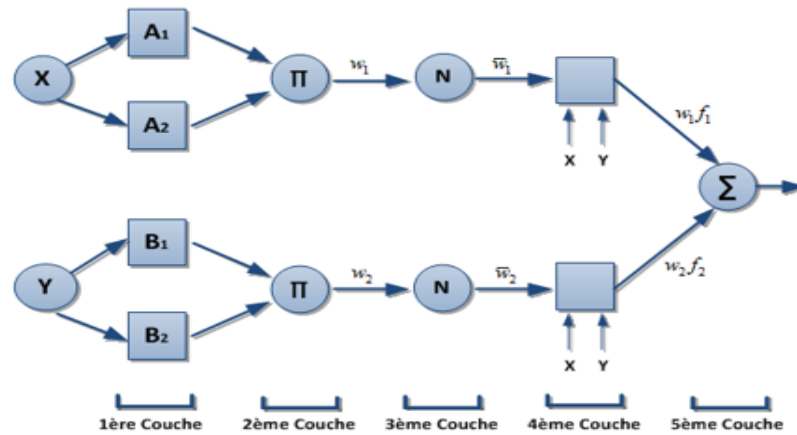


FIGURE 3.10 – Architecture ANFIS

**1<sup>ère</sup> couche ( Fuzzification) :** Cette couche permet de convertir les variables d’entrées dans l’espace flou à l’aide des fonctions d’appartenance.

$$O_i^1 = \mu_{A_i}(X) \quad \text{avec : } i = 1, 2$$

$$O_i^1 = \mu_{B_{i-2}}(Y) \quad \text{avec : } i = 3, 4$$

où  $\mu_{A_i}(X)$  et  $\mu_{B_{i-2}}(Y)$  sont les degrés d’appartenance du *ime* terme de la partition floue des variables d’entrée. Différents types de fonctions d’appartenance peuvent être utilisés : triangle, trapèze ou gaussienne par exemple. Les paramètres définissant ces fonctions sont appelés prémisses.

**2<sup>ème</sup> couche ( Pondération des règles floues) :** Chaque neurone est associé à une règle floue et permet de définir le poids de cette règle à l’aide d’une t-norme. La t-norme la plus employée est le produit :

$$O_i^2 = w_i = \mu_{A_i}(X)\mu_{B_{i-2}}(Y) \quad i = 1, 2$$

**3<sup>ème</sup> couche (Normalisation) :** Cette couche a pour but de normaliser les différents poids.

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1+w_2} \quad i = 1, 2$$

**4<sup>ème</sup> couche (Défuzzification) :** Cette couche donne une sortie définie par une fonction du premier ordre des entrées pondérées (approche de Takagi-Sugeno).

$$O_i^4 = w_i f_i \text{ avec } f = a_i.X + b_i.Y + c_i \quad \text{ou } i=1,2$$

La fonction  $f_i$  peut aussi être constante. Dans ce cas, tous les  $a_i = b_i = 0$ . Les paramètres  $a_i$  ;  $b_i$  et  $c_i$  des fonctions  $f_i$  sont appelés paramètres conséquents.

**5<sup>e</sup>me couche (Calcul de la sortie) :** La sortie  $O_i^5$  est définie par la somme des sorties de la couche précédente :

$$O_i^5 = \sum_{i=1}^2 w_i f_i = \frac{\sum_{i=1}^2 w_i f_i}{w_1 + w_2}$$

### Phase d'Apprentissage

L'ajustement des paramètres de l'ANFIS est réalisé lors de la phase d'apprentissage. Pour cela un ensemble de données associant les séquences d'entrées et de sorties est nécessaire. Pour la réalisation de cette phase, l'algorithme d'apprentissage hybride est utilisé.

L'algorithme d'apprentissage hybride utilisé est une combinaison de deux méthodes celle de descente de gradient et la méthode d'estimation des moindres carrés. La première permet d'ajuster les prémisses en fixant les paramètres conséquents tandis que la seconde LSM (Least square Method) ajuste les paramètres conséquents en fixant les prémisses [46].

## 5 Conclusion

Les outils fournis par la logique floue permettent une modélisation des phénomènes pouvant en un certain sens s'approcher du raisonnement humain. Le fait de transcender le tout ou rien des ordinateurs introduit une souplesse faisant la puissance des outils flous dans de nombreux domaines. Dans ce chapitre, nous avons soulevé l'importance derrière la mise en œuvre d'un classifieur floue pour fournir des résultats interprétable.

# Chapitre 4

## Résultats et Discussion

## Première partie

### Considération 1 : Détection des Extrasystoles ventriculaires



---

# 1 Proposition 1 : Reconnaissance des Extrasystoles ventriculaire par les HMMs

## 1.1 Introduction

L'objectif de cette partie du chapitre est de développer deux classifieurs supervisés pour la détection des ESV. Le premier classifieur se base sur les modèles de Markov cachés tandis que le deuxième est un classifieur neuro-flou. Dans cette partie, nous décrivons brièvement les modèles de classification utilisés et les étapes de prétraitement, puis nous discutons les résultats obtenus de la classification des signaux ECG.

## 1.2 LE SIGNAL ELECTROCARDIOGRAMME

### Principes

Une fibre cardiaque en cours de dépolarisation peut être assimilée à un dipôle de courant. A un instant donné, le front de l'onde d'activation formé par l'ensemble des dipôles élémentaires crée un champ électrique qui est fonction des moments dipolaires.

L'enregistrement de l'évolution temporelle du champ électrique résultant, effectué au moyen d'électrodes cutanées, se nomme l'électrocardiogramme de surface.

### Ondes de l'ECG enregistrées et terminologie

Pour chaque battement cardiaque, l'électrocardiogramme enregistre successivement 5 ondes : P, Q, R, S et T (Figure 4.1). Généralement les ondes Q, R et S sont groupées et nous parlons plutôt du complexe QRS.

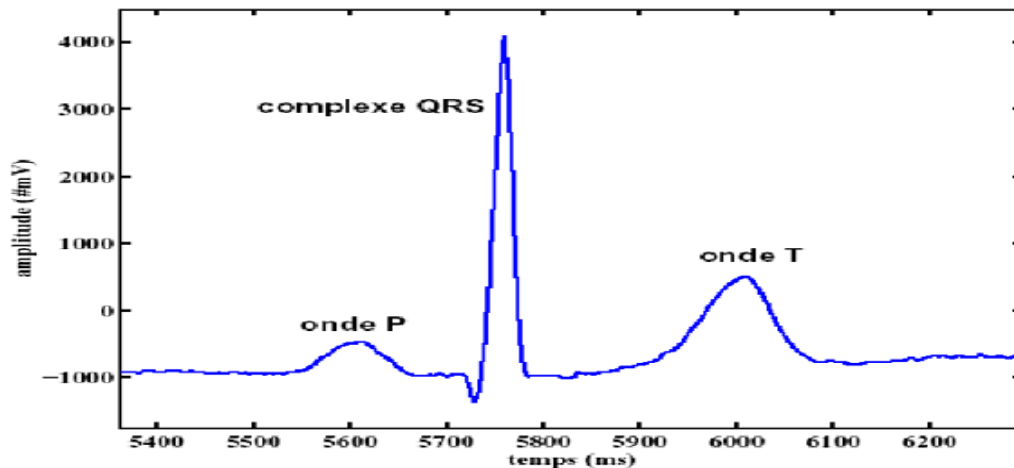


FIGURE 4.1 – Un cycle cardiaque

Sur chaque battement cardiaque nous pouvons identifier :

- La ligne isoélectrique (Figure4.2) : Le premier repère sur un tracé électrocardiographique est la ligne isoélectrique. Elle est la ligne de base correspondante à l'absence de phénomène électrique. Au-dessus de celle-ci, on parle d'onde positive, en dessous, d'onde négative. Une onde peut être aussi diphasique si une partie de celle-ci se situe au-dessus et l'autre partie au-dessous de la ligne isoélectrique. Toutes les ondes se mesurent du début de leur phase initiale, à la ligne isoélectrique.



FIGURE 4.2 – La ligne isoélectrique

- L'onde P4.1 : Elle représente la dépolarisation auriculaire. Sa durée est de l'ordre de 90 ms. C'est une onde positive dont l'amplitude est normalement inférieure ou égale à 0.2 mV.
- Le complexe QRS (Figure4.1) : Il correspond à la dépolarisation ventriculaire précédant l'effet mécanique de contraction. Sa durée normale est comprise entre 85 et 95 ms.

- L'onde T4.1 : Elle correspond à la repolarisation des ventricules. Elle est asymétrique, d'une branche ascendante légèrement oblique et d'une branche descendante plus abrupte. Son amplitude est inférieure à 2 mm. L'onde T normale à une amplitude plus faible que le complexe QRS.
- Le segment PR (Figure4.3) : Il correspond à la pause de 1/10e de seconde entre l'activation auriculaire et l'activation ventriculaire, par le passage de l'influx du nœud auriculo-ventriculaire au faisceau de His. Il se mesure de la fin de l'onde P jusqu'au début du QRS et correspond à 0,03 à 0,04 secondes.

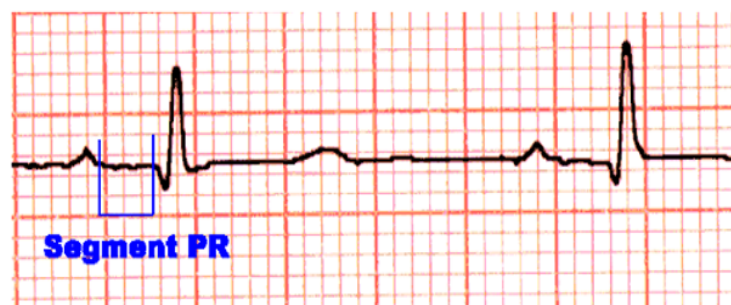


FIGURE 4.3 – Segment PR

- Le segment ST (Figure4.4) : Il correspond à la période d'excitation uniforme des ventricules jusqu'à la phase de récupération des ventricules. On le mesure de la fin de l'onde S ou R jusqu'au début de l'onde T. Il est normalement horizontal ou légèrement oblique +/- isoélectrique. Un sur-décalage ou un sous-décalage de plus de 1 mm par rapport à la ligne isoélectrique est anormal.

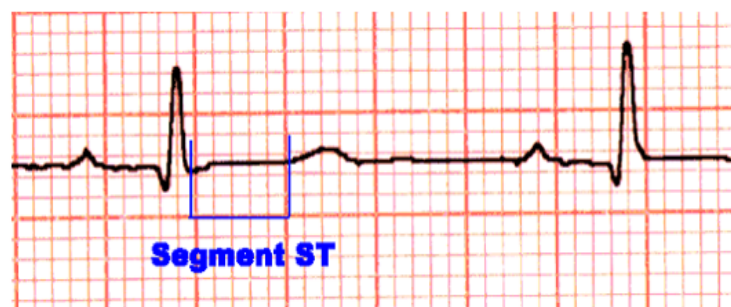


FIGURE 4.4 – Segment ST

## Le rythme cardiaque

L'étude du rythme cardiaque se fait à partir du repérage des ondes R ; ce rythme est caractérisé par deux propriétés : la fréquence des ondes R, exprimée en nombre de battements par minute (bpm), et leur régularité.

---

En l'absence de toute pathologie, le rythme est régulier et sa fréquence est comprise entre 60 et 100 bpm la journée et 40 et 80 bpm la nuit. Hors de ces limites, il peut y avoir trouble du rythme. Cependant, toute irrégularité n'est pas pathologique : en effet, le système nerveux autonome, exerçant un contrôle permanent, peut fortement accélérer le rythme en réponse à un contexte particulier : une période de stress ou d'effort, par exemple. Il est donc essentiel de prendre en considération l'activité du patient avant de poser un diagnostic.

### **Le battement cardiaque standard et ses caractéristiques**

Une analyse pertinente du signal ECG est fondée sur l'analyse de quelques battements cardiaques successifs ; l'étude d'un seul battement ne fournit que peu d'indications pour la pose d'un diagnostic, mais les variations des paramètres caractéristiques de chaque battement au cours de l'enregistrement constituent une source d'information essentielle.

Parmi ces paramètres caractéristiques nous citons :

- Les durées des ondes P, Q, R, S et T, et les amplitudes mesurées par rapport à la ligne de base (Figure 4.5). La ligne de base est la ligne isoélectrique du cœur au repos, qui est prise comme référence pour mesurer l'amplitude des ondes : pendant l'inactivité cardiaque, le potentiel mesuré est donc normalement nul par rapport à cette référence. C'est le cas au niveau :
  - de l'intervalle entre les ondes T et P de deux battements successifs,
  - l'intervalle entre les ondes P et Q d'un même battement,
  - l'intervalle entre les ondes S et T (en l'absence de pathologie)
- Les distances entre ces ondes.

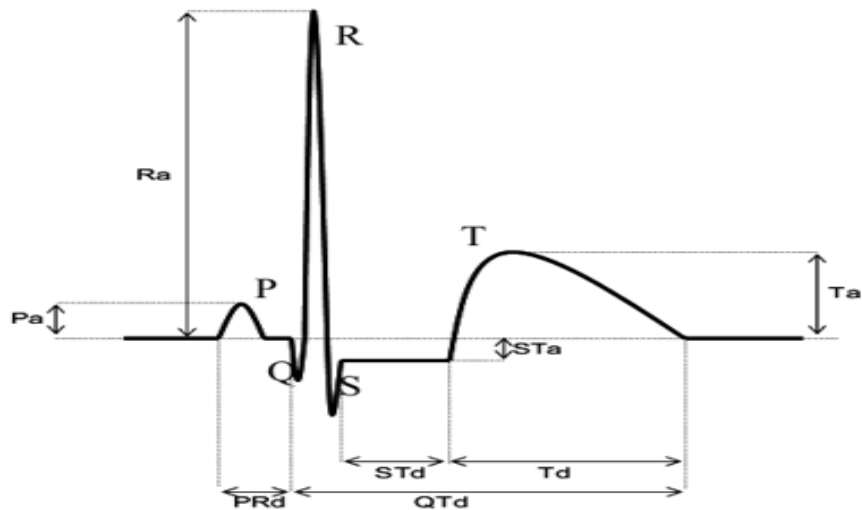


FIGURE 4.5 – Paramètres d'intérêt pour la description d'un battement

Les valeurs des paramètres du signal ECG de la figure 4.5 couramment constatées chez l'adulte en bonne santé sont présentées dans le tableau 3.1.

TABLE 4.1 – Valeurs habituelles des différents paramètres caractérisant un battement cardiaque [4]

	Onde P	Intervalle PQ	Complexe QRS	Intervalle ST	Intervalle QT	Onde T
Durée (s)	0.08-0.1	0.12-0.2	0.06-0.1	0.2	0.36	0.2
Amplitude (mV)	0.2	Isoélectrique	Q<0 R>0 S<0	Isoélectrique	-	T>0

### 1.3 Extrasystole ventriculaire (ESV)

L'extrasystole ventriculaire (ESV) est un battement anormal et qui s'observent sur quasiment tous les enregistrements, principalement en période de récupération après un effort. Bien que leur présence n'indique aucune pathologie particulière, si, de façon récurrente, leur nombre par minute est supérieur à 6, elles peuvent être un signe précurseur d'une tachycardie ventriculaire, qui, constitue une pathologie majeure.

Contrairement aux battements normaux qui ont pour origine la dépolarisation des cellules sinusales, l'ESV naît de la dépolarisation spontanée d'un petit groupe de cellules ventriculaires, appelé alors foyer ectopique ventriculaire. L'impulsion électrique créée n'emprunte pas la voie normale de conduction (faisceau de His), et se propage donc plus lentement dans les ventricules. La contraction ventriculaire ainsi étalée dans le temps perd de son efficacité.

---

Le tracé d'un battement ESV est caractérisé par deux propriétés : l'onde R n'est pas précédée d'une onde P, puisqu'il n'y a pas eu d'activité auriculaire préalable, et la durée du complexe est supérieure à la durée d'un complexe QRS normal (figure 4.7).

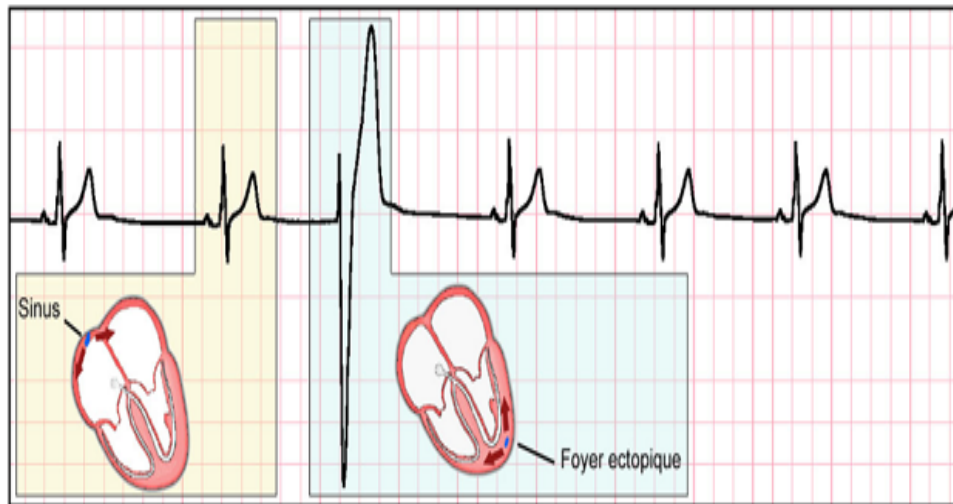


FIGURE 4.6 – Extrasystole ventriculaire (ESV).

Le battement ESV a pour origine la dépolarisation spontanée d'un groupe de cellules des ventricules : il n'est donc pas précédé d'une onde P, et il est plus large qu'un complexe QRS. Les ESV sont pathologiques si leur nombre par minute est supérieur à 6.

## 1.4 Diagnostic à partir du rythme

Deux points essentiels sont à étudier lors d'un diagnostic d'un enregistrement Holter : le rythme cardiaque (fréquence et régularité) et la forme des ondes (distances relatives, amplitudes, durées, ...).

L'analyse du rythme ne nécessite que le repérage des ondes R : elle fut donc l'un des premiers traitements automatiques du signal cardiaque. Cette analyse est fondée sur l'extraction, à partir du signal, des deux paramètres caractéristiques suivants : la fréquence des battements et leur régularité.

Le tracé d'un battement ESV est caractérisé par certaines propriétés dont il faut tenir en compte pour le détecter :

- L’onde R n’est pas précédée d’une onde P, puisqu’il n’y a pas eu d’activité auriculaire préalable.
- La durée du complexe QRS est supérieure à la durée du complexe QRS normal car l’impulsion électrique n’emprunte pas la voie normale de conduction et se propage donc lentement dans les ventricules.
- Le complexe QRS présente une forme atypique.
- Ces battements sont prématurés.

## 1.5 La base de données MIT-BIH

La base de données physiologiques MIT/BIH est formée par différentes bases dont la plus utilisée est la base MIT/BIH d’arythmie [47]. Cette base contient 48 enregistrements de signaux ECG de 47 cas différents. Chaque enregistrement dure environ 30mn, la fréquence d’échantillonnage est à 360 Hz et la résolution est de 200 échantillons par  $mV$ . L’ensemble des enregistrements regroupe plus de 109 000 battements classés en 15 types de battements cardiaques différents. Les différentes pathologies sont divisées selon la base MIT-BIH en classe d’arythmies. Approximativement, la classe la plus abondante est la classe des rythmes normaux (NOR) avec 70% des battements.

Le choix des enregistrements utilisés est lié essentiellement à la présence des pathologies ciblés disponibles mais aussi nous avons essayé de collecté des patients de même sexe et d’âge semblable (les hommes qui ont plus de 55ans) sachant que les caractéristiques du signal ECG dépendent de l’âge et du sexe.

Le tableau 4.2 représente les patients choisis de la base MIT :

TABLE 4.2 – Les patients choisis pour constituer la base de données

Enregistrement	Sexe	Age	Sous médicament anti-arythmique
116	M	68	non
200	M	64	oui
201	M	68	oui
210	M	89	non
223	M	73	non
233	M	57	oui

Notons que la base de données MIT-BIH est formée principalement des enregistrements acquis par la dérivation DII, ce qui limite la généralisation de ce travail vers d’autres types de pathologies qui apparaissent mieux sur d’autres voies que la voie

## 1.6 Description du système de classification

Les modèles de Markov cachés (HMMs) ont été largement appliqués dans la reconnaissance de la parole et les résultats sont prometteurs [48, 49]. Dans le cadre des travaux de cette thèse nous avons envisagé de voir l'apport de ces modèles dans le domaine de diagnostic médical.

Notons que le système de classification comporte 2 blocs : le premier consiste à collecter les mesures avec les paramètres de reconnaissance extraits à partir des signaux ECG de la base MIT-BIH. Ces mesures sont utilisées pour le second bloc comme corpus d'apprentissage et corpus de test servant à la conception du classifieur HMM. Dans la phase d'apprentissage, les deux tiers des données choisis sont utilisées comme corpus d'apprentissage. Les algorithmes de Baum- Welch 4 et forward-Backward1 assurent ces deux tâches.

Dans la phase de reconnaissance, le tiers restant des données est utilisé comme corpus de test. L'évaluation de ce classifieur considère le taux de reconnaissance comme critère de performance qui représente le pourcentage des cardiopathies correctement reconnus. Cette phase fournit aussi une matrice de confusion qui nous donne un rapport détaillé des résultats de classification.

Les modèles automatiques d'analyse et la classification des signaux ECG ont été développés depuis le début des années 1980 et plusieurs approches statistiques ont été adoptées pour améliorer la précision des systèmes de classification. Dans le cadre de notre travail de thèse, nous présentons un classifieur HMM, qui constitue une bonne base mathématique pour résoudre le problème de classification, dans le but d'améliorer la transparence de la décision finale de notre système.

## 1.7 Méthode HMM adaptée

Supposons que nous avons un ensemble de classes  $K$  à classifier (battements normaux et battements ECG pathologiques), chaque classe doit être modélisée par un modèle HMM distinct. Sachant que, pour chaque classe dans notre base de données, nous avons un ensemble d'apprentissage  $L_{train}$  (les battements de l'ECG) de chaque classe où chaque cas de la classe constitue une séquence d'observation, et les observations sont des représentations appropriées des caractéristiques de la classe. Nous réalisons les opérations suivantes pour chaque tâche de classification :



- Pour chaque classe  $k_{in}$  dans notre base de données, nous devons construire un modèle HMM  $\lambda^k$ , nous estimons les paramètres du modèle  $(A, B, \Pi)$  qui permettent d'optimiser la probabilité de l'observation de l'ensemble d'apprentissage pour la  $k^{th}$  classe.
- Pour chaque nouvelle observation à classifier, le traitement de signal représenté sur la figure 4.7 doit être effectuée. Comme le filtrage du bruit, ensuite la détection du complexe QRS [50] et enfin une extraction de caractéristiques est réalisée. Le bloc de pré-traitement est suivi par le calcul des probabilités pour tous les modèles possibles,  $P(O|\lambda^k)$ ,  $1 \leq k \leq K$  et enfin le modèle dont la probabilité d'observation est la plus élevé, est choisi, à savoir

$$k^* = \arg \max_{1 \leq k \leq K} P(O|\lambda^k)$$

L'étape de calcul de la probabilité est effectuée en utilisant l'algorithme de Forward-Backward 1.

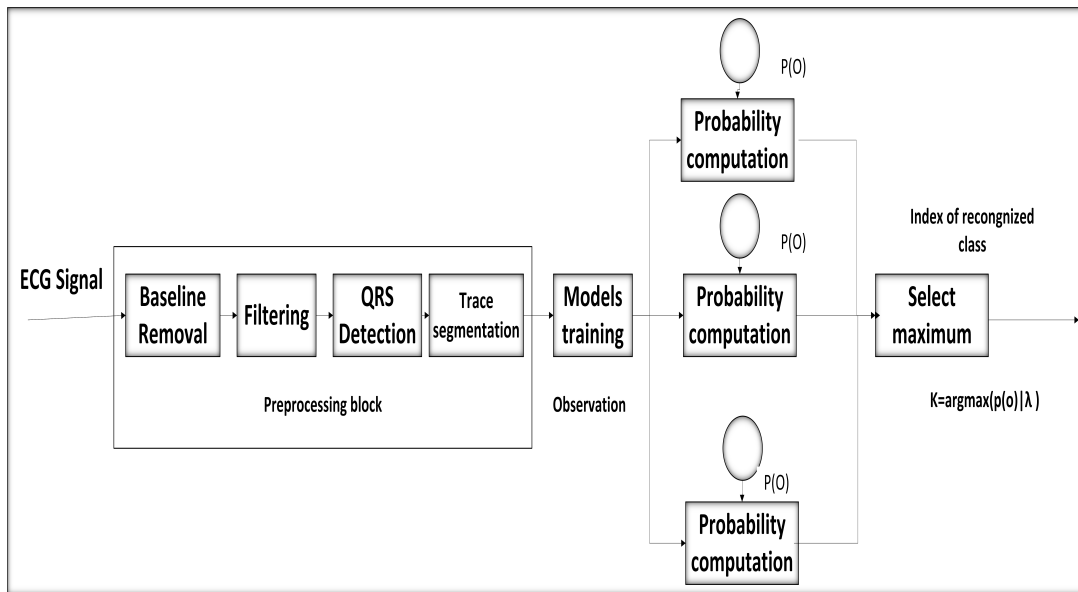


FIGURE 4.7 – Les étapes du système de reconnaissance

## 1.8 Reconnaissance des extrasystoles ventriculaires par HMM

Le problème de la classification de battement cardiaque peut être effectué directement par le HMM [51]. L'idée principale dans les travaux de Koski et al. [51] est de considérer chaque modèle de battement spécifique à chaque type de battement et en utilisant la fonction de densité de probabilité continue pour l'analyse du signal d'ECG.

Dans notre travail, les différents attributs sont discrétisés et transformés en symboles selon des intervalles de temps choisis par les experts dans le but de concevoir le modèle HMM avec les observations interprétables. Les intervalles de différents attributs sont présentés dans le tableau 4.3.

TABLE 4.3 – Valeurs des intervalles des attributs utilisés

Attribut	Symbole	Signification	Intervalle (s)
Intervalle RR	R	Régulier	1-1.6
	IR	Irrégulier	<1 et >1.6
Complexe QRS	N	Normal	<0.1
	L	Large	>0.1

Il est connu que les battements ESV ont deux caractéristiques bien définies 1.3 qui sont suffisantes pour les distinguer des autres classes de battements. La première caractéristique est liée au battement prématurité. Cela signifie que le rythme ESV est caractérisé par un intervalle R-R plus long que le précédent (voir la figure 4.7). La deuxième caractéristique concerne la morphologie du complexe QRS. En fait, un battement de ESV est aussi un type de battement ventriculaire, qui est caractérisé par un complexe QRS plus large que ceux des battements normaux (N).

A partir des informations extraites du type de battement, le modèle de battement est construit, et pour classer la séquence de battement, deux modèles de battement sont nécessaires : un pour les battements normaux et l'autre pour les battements de ESV.

Deux ensembles d'apprentissage sont générés, en fonction de deux modèles, correspondant chacune à l'une des deux classes.

Pour notre système de classification, nous avons utilisé les 4 attributs les plus pertinents :

- RRp : La distance entre le pique R actuel et le pique R précédent .
- RRs : La distance entre le pique R suivant et le pique R suivant.
- RRs/RRp : le rapport entre les distances RR suivant et RR précédent.
- QRS : La durée du complexe QRS.

L'évaluation de notre classifieur HMM est faite sur la base de données MIT-BIH, nous avons choisi les enregistrements dont les battements de type normaux ou ESV.

A partir des données ECG présentés dans le tableau 4.4, on choisit au hasard deux tiers de la base de données pour la phase d'apprentissage pour chacune des deux classes (ESV) et normales. De la base de données d'apprentissage, nous générons un

modèle initial, qui sera construit pour obtenir les deux modèles HMM finaux.

TABLE 4.4 – Nombre de battements pour chaque enregistrement

Enregistrement	Normal	ESV
116	2175	108
200	1726	815
201	1325	176
210	2100	181
223	1986	449
233	1029	813
Total	12041	2545

Le diagramme d'apprentissage et de la classification est présenté dans la figure 4.8. Dans la première étape de la phase d'apprentissage, le type de modèle et le nombre d'états sont identifiés.

Ensuite, les paramètres du modèle ( $A, BetII$ ) obtenus pour chaque classe sont enregistrés pour être utilisé dans la phase de classification.

Quand un modèle est donné, la probabilité d'observer un ensemble de données particulier de ce modèle est obtenu par le calcul la probabilité de vraisemblance. Les probabilités de vraisemblances trouvées lors de la phase d'apprentissage des modèles DHMM sont présentées dans la Figure4.9.

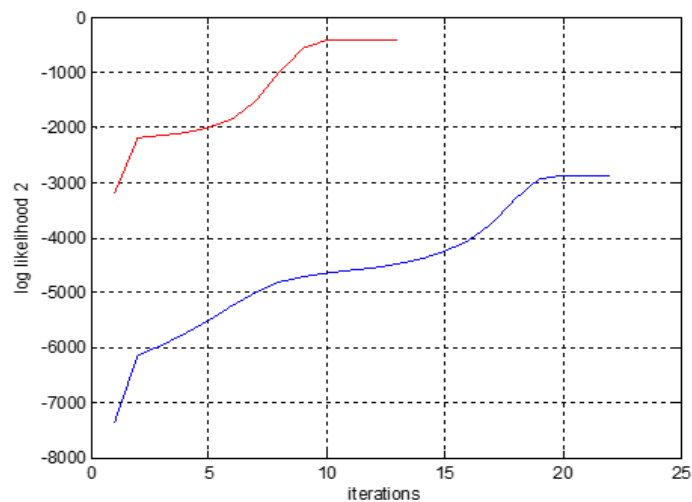


FIGURE 4.9 – Courbe de la probabilité de vraisemblance

D'après le tableau 4.5, nous remarquons que notre modèle de classification a obtenu de bons résultats avec un taux de classification de 96,59%.

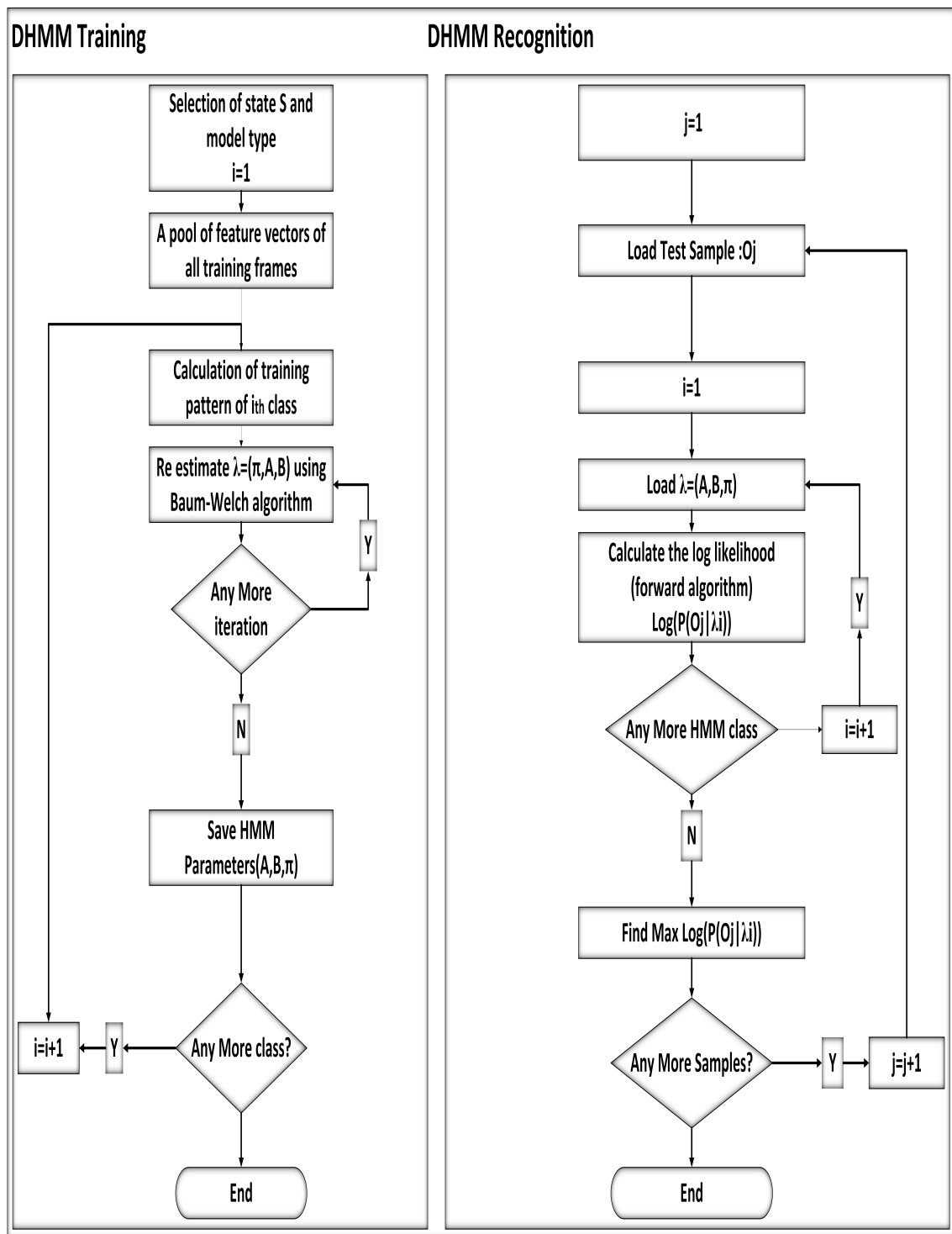


FIGURE 4.8 – Diagramme de la phase d'apprentissage et de test

Le taux de classification des battements de l'enregistrement 210 est égale à 89,25%. Notons que cet enregistrement présente beaucoup de battement succesives.

TABLE 4.5 – Performances du classifieur HMM

Enregistrement	FN	FP	TN	TP	Se(%)	Sp(%)	Tc(%)
116	0	1	757	36	100	99.87	99.87
200	9	9	570	264	96.70	98.45	97.89
201	0	18	519	66	100	96.65	97.01
210	5	88	716	56	91.80	89.05	89.25
223	2	7	665	156	98.73	98.96	98.92
233	5	14	726	270	98.18	98.11	98.13
Moyenne					97.57	96.85	96.59

Dans la littérature un grand nombre de travaux ont été réalisé pour la classification des battements ESV. Ebrahimzadeh et al. [52] ont utilisé un réseau de neuronaux (MLP) avec un nombre différent de couches, et neuf algorithmes d'apprentissage ont été appliqués. Les performances des réseaux en termes de vitesse de convergence et de précision de classifications ont été validées sur sept enregistrements de la base de données MIT-BIH, ils ont obtenu un taux de classification de 95,4%.

N. Asyiqin et al. [53] ont proposé une transformée en ondelette Discrète optimale (DWT) pour classifier les battements ESV et les battements normaux en utilisant un classifieur neuronale. Les auteurs ont remarqué que les ondelettes «haar», «db3» et «sym3» produisent des résultats prometteurs en testant 400 battements.

V. mahesh et al. [54] ont présenté un classifieur des arythmies cardiaque à partir des données de l'ECG, en utilisant le model LMT (Logistic Model Tree); ils ont obtenus une prédiction de 98% en utilisant seulement 1281 battements de la base de données MIT-BIH, par contre le taux de prédiction était de 86,48% en restant quatre enregistrements de la base de données MIT-BIH.

P. ghorbanian et al. [55] ont implémenté un modèle de classification basé sur la transformée en ondelettes continue (CWT), l'analyse en composante principale (PCA) et un réseau de neurones multicouches (MNP) pour classifier six classes de l'ECG en utilisant seulement 600 battements à partir de différents enregistrement de la base de données MIT-BIH, ils ont obtenu un taux de classification de 99,17%.

A partir des travaux réalisés dans la littérature, nous remarquons clairement que les résultats de classificaton obtenus dans les travaux de V. mahesh et al. et P. ghorbanian et al [54] et [55] sont très performants mais en utilisant un nombre réduit de battement. D'autre part, dans notre travail nous avons utilisé six enregistrements complets de la base de données MIT-BIH avec un nombre total de 14 586 batte-

---

ments, nous avons obtenus une meilleure précision de la classification avec un taux de 96.59%.

## 2 Proposition 2 : Reconnaissance des Extrasystoles ventriculaire par un modèle Neuro Floue

Le classifieur neuro-flou que nous avons développé met en application un système d'inférence floue (SIF) de type Takagi-Sugeno d'ordre zero (la fonction de conclusion est une constante) et un réseau de neurones adaptatif. Une description des systèmes d'inférences floues est déjà présentée dans le chapitre 3.

Le modèle neuro-flou implémenté est un modèle hybride qui possède la capacité de modélisation et d'adaptation des réseaux de neurones et la facilité d'interprétation et d'intégration des connaissances des systèmes d'inférences floue.

### 2.1 Le choix des fonctions d'appartenances

La détermination du nombre et du type des fonctions d'appartenances pour chaque variable d'entrée (RRp, RRs /RRp, QRS) est faite d'une manière empirique, et en tenant compte des critères de précision (erreur quadratique moyenne globale) et de complexité (nombre de règles).

Le tableau 4.6 représente les erreurs quadratiques moyennes atteintes et le nombre de règles générées lors de l'apprentissage pour différents choix du nombre des fonctions d'appartenance.

TABLE 4.6 – Erreurs quadratiques moyennes pour différentes configurations

Configuration	Nombre de règles	Erreur atteinte lors de l'apprentissage
$2 \times 2 \times 2$	8	0.210
$2 \times 3 \times 3$	18	0.105
$3 \times 2 \times 3$	18	0.173
$3 \times 3 \times 2$	18	0.201
$3 \times 3 \times 3$	27	0.103

D'après les résultats que nous avons obtenus, nous remarquons que plus le nombre des fonctions d'appartenances est grand, plus l'erreur entre la sortie réelle du classifieur et la classe à prédire est minimale. Cette précision est achevée au prix d'une complexité au niveau des règles générées, ce qui rend la lisibilité de la base de règle

---

difficile avec trop de règles globalement similaires.

Nous avons donc choisis un compromis entre la précision du classifieur et la complexité de la base des règles.

Il est à noter qu'une partition de plus de trois fonctions d'appartenances pour les trois variables d'entrées n'apporte aucune amélioration au niveau des résultats.

Ainsi, le classifieur neuro-flou retenu vérifie la configuration suivante :

Nombre de fonctions d'appartenances pour chaque variable :

- RRp : 2 fonctions d'appartenance (petit, moyen)
- RRs /RRp : 3 fonctions d'appartenance (petit, moyen, grand)
- QRS : 3 fonctions d'appartenance (petit, moyen, grand)

Cette configuration est la meilleure car elle donne une erreur minimale de (0.1052) tout en gardant un nombre raisonnable de règles (18 règles).

En ce qui concerne le type des fonctions d'appartenance, nous avons choisi une fonction « trapézoïde » pour toutes les variables du vecteur d'entrée. Ce type de fonction est préférable aux fonctions gaussiennes car elles permettent une meilleure lisibilité des différentes règles qui contribuent à la décision finale.

## 2.2 Choix des points modaux

L'initialisation des points modaux est faite d'une façon manuelle on se basant sur les connaissances médicales de l'expert.

Les points modaux initiaux sont résumés dans le tableau 4.7 et les fonctions d'appartenances initiales sont illustrées dans la figure 4.10

TABLE 4.7 – les points modaux initiaux

	Petit		Moyen				Grand	
RRp	0.55	1	0.5		0.95			
Rapport	0.6	0.95	0.75	0.9	1.1	1.25	1.1	1.35
QRS	0.045	0.075	0.055	0.075	0.1	0.135	0.1	0.135

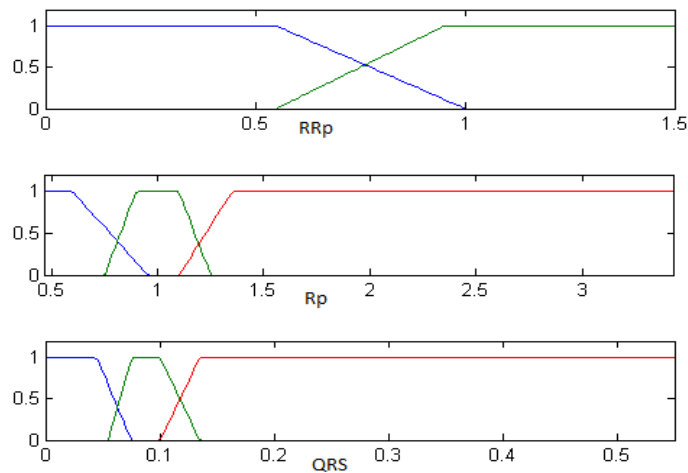


FIGURE 4.10 – les fonctions d'appartenances pour les trois entrées (RRp,RRs/RRp,QRS)

Après avoir choisi la configuration initiale de notre classifieur, nous lançons l'apprentissage par la méthode hybride qui se base sur la combinaison de l'algorithme de rétropropagation de l'erreur pour modifier les fonctions d'appartenances dans la première couche (couche de fuzzification) et la méthode de l'estimation des moindres carrés pour l'optimisation de la constante de sortie (4ème couche).

L'apprentissage est lancé en fixant les critères d'arrêt suivant :

- Le nombre d'itération fixé à 100 a permis un bon apprentissage. En effet, si nous choisissons un nombre trop grand, il va y avoir un sur-apprentissage, par contre s'il est trop petit, l'apprentissage ne sera pas conduit à son terme.
- Le seuil de l'erreur globale est fixé à 0.01.

La valeur de l'erreur globale atteinte après 100 itérations d'apprentissage est de 0.16, l'évolution de cette erreur pendant l'apprentissage en fonction des itérations est illustrée par la figure4.11



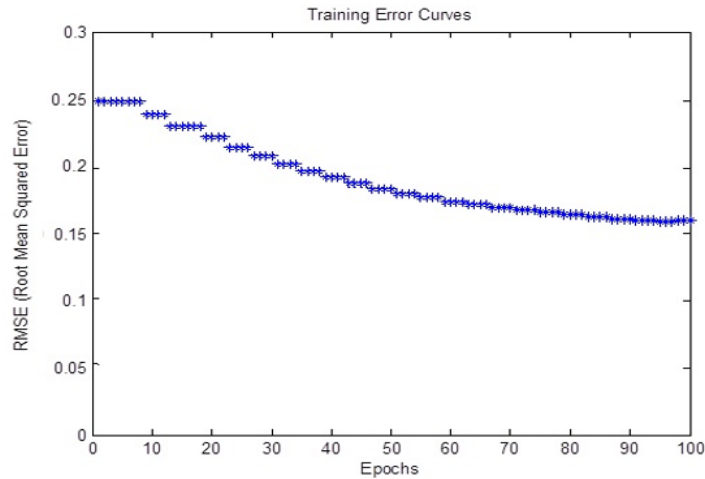


FIGURE 4.11 – L'évolution de l'erreur globale

L'apprentissage du modèle a été réalisé avec une base d'exemple contenant 1400 battements (700 battements de chaque classe N et ESV). Ce choix est établi pour éviter le problème de la spécialisation vers la reconnaissance des battements prédominants dans la base d'exemple (les battements normaux dans notre cas).

L'apprentissage a permis une optimisation du classifieur neuro-flou par la modification de l'emplacement des fonctions d'appartenances des trois variables d'entrée. Les points modaux finaux sont présentés dans le tableau 4.8

TABLE 4.8 – les points modaux finaux

	Petit		Moyen				Grand	
RRp	0.41	0.08	0.52		0.80			
Rapport	0.59	0.97	0.76	0.9	1.09	1.26	1.06	1.32
QRS	0.05	0.088	0.079	0.087	0.116	0.135	0.134	0.149

Nous avons testé notre classifieur neuro-flou sur la totalité des battements présentés dans le tableau 2 et le test a été effectué sur chaque enregistrement afin de calculer la sensibilité et la spécificité et le taux de classification. 4.10

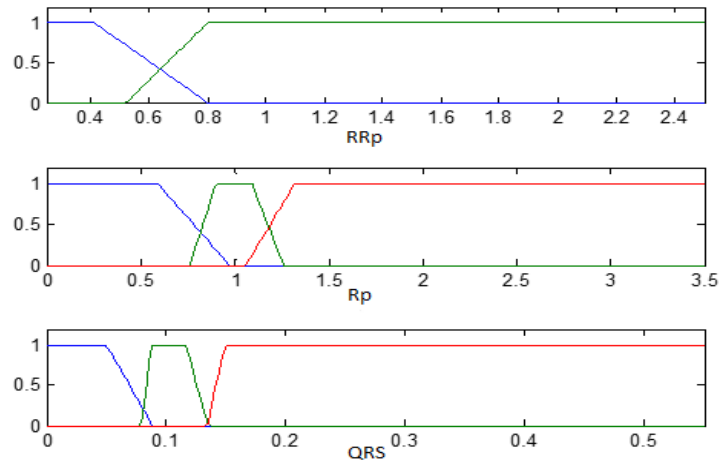


FIGURE 4.12 – les fonctions d'appartenances finaux pour les trois entrées (RRp,RRs/RRp,QRS)

Les résultats obtenus par le classifieur Neuro-Flou sont explicites et interprétables, il permet de générer une base de connaissance (18 règles) pour justifier la classification.

TABLE 4.9 – Performances du classifieur Neuro-Flou

Enregistrement	FN	FP	TN	TP	Se	Sp	Tc
116	9	2	2262	9	98.14	99.60	99.54
200	379	46	1357	379	94.02	78.16	83.35
201	313	55	1296	313	72.22	80.55	79.63
210	699	15	1712	699	91.80	71.01	72.47
223	18	45	1997	18	90.48	99.11	97.47
233	564	28	1655	564	96.60	74.58	80.54
Moyenne					90.54	83.83	85.50

Nous calculons le degré de sollicitation de chaque règle par le rapport du nombre d'exemples qui active une règle (avec degré d'activation supérieur à 50%) sur le nombre total des battements pour chaque classe (N, ESV). Parmi les 18 règles, la règle 4 est la règle la plus sollicité avec 40.07% des battements normaux et la règles 9 est la règle la plus sollicité avec 73,61% des battements ESV. Ces deux règles sont très conformes avec le raisonnement du cardiologue.

Les battements correctement classés par le classifieur Neuro-Flou sont devisés en deux groupes :

- Battements reconnu en activant une seule règle (les prototypes)
- Battements reconnu en activant plusieurs règles (le plus souvent deux),

---

ces battement sont eux même devisés en deux sous groupes :

- Battement reconnu en activant plusieurs règles avec une seule règle dominante (prototype)
- Battements reconnu en activant plusieurs règles avec plus d'une règles dominante ; ce type de battement nécessite une intervention du médecin expert.

### 2.3 Les méthodes interprétables et probabilistes

Dans notre travail nous avons opté pour deux approches de classification ; les modèles de Markov cachés et la logique floue. L'approche probabiliste nous a permis d'extraire des probabilités afin de pouvoir différencier les battements ESV des battements normaux par le calcul de maximum de vraisemblance de chaque séquence présentée.

Le passage par une étape de symbolisation nous a permis d'intégrer dans notre système de classification la notion d'interprétabilité.

Dans la figure 4.13 nous présentons un battement normal dont les caractéristiques sont : *RRp régulier*, *RRs régulier* et un *QRS normal*, ces caractéristiques sont compréhensives par le médecin expert et cette même séquence nous a donné une probabilité de  $-0,1696$  pour le modèle normal et  $-10,2607$  pour le modèle ESV, alors dans ce cas l'utilisateur dispose de l'information probabiliste et de l'information interprétable avec un taux de reconnaissance de 96,59%.

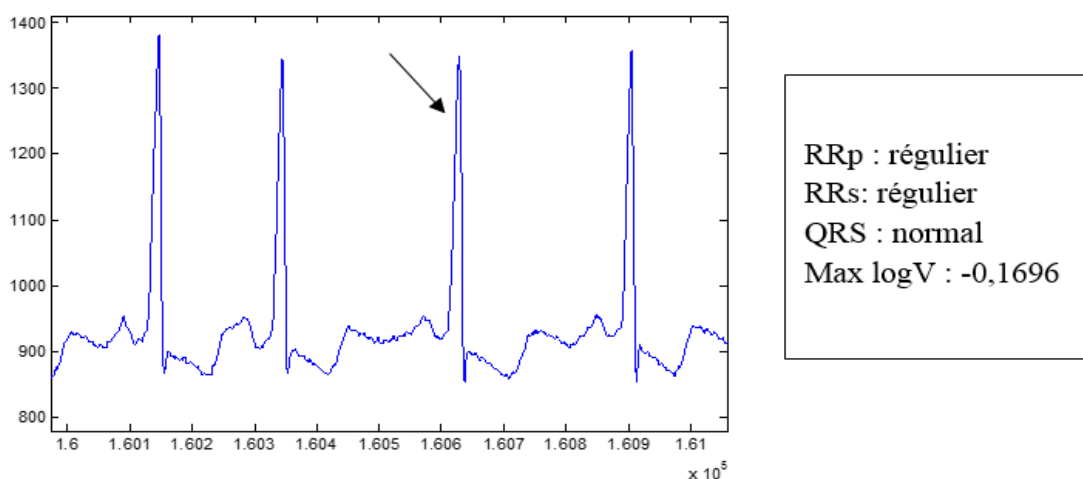


FIGURE 4.13 – caractéristique d'un battement normal

Le même battement illustré dans la figure 4.13 a été détecté en activant les règles 10 et 13 avec la règle 13 dominante (Figure 4.14)

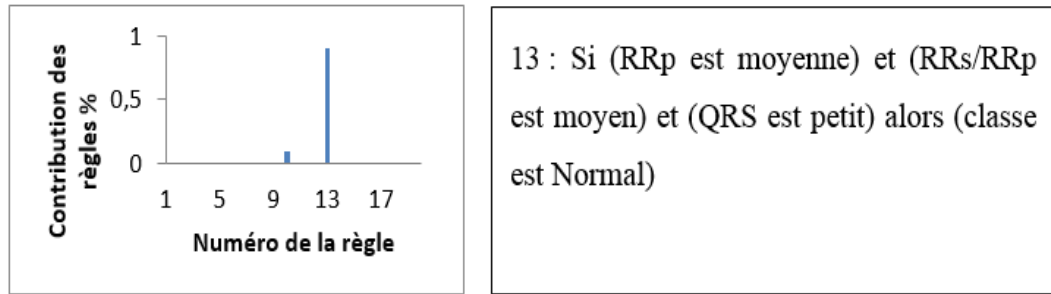


FIGURE 4.14 – Les règles activé par un battement normal

Dans la figure4.15 nous présentons un battement ESV dont les caractéristiques sont : *RRp régulier*, *RRs irrégulier* et un *QRS large*, cette séquence d'observation nous a donné une probabilité de  $-6,678$  pour le modèle normal et  $-0,129$  pour le modèle ESV.

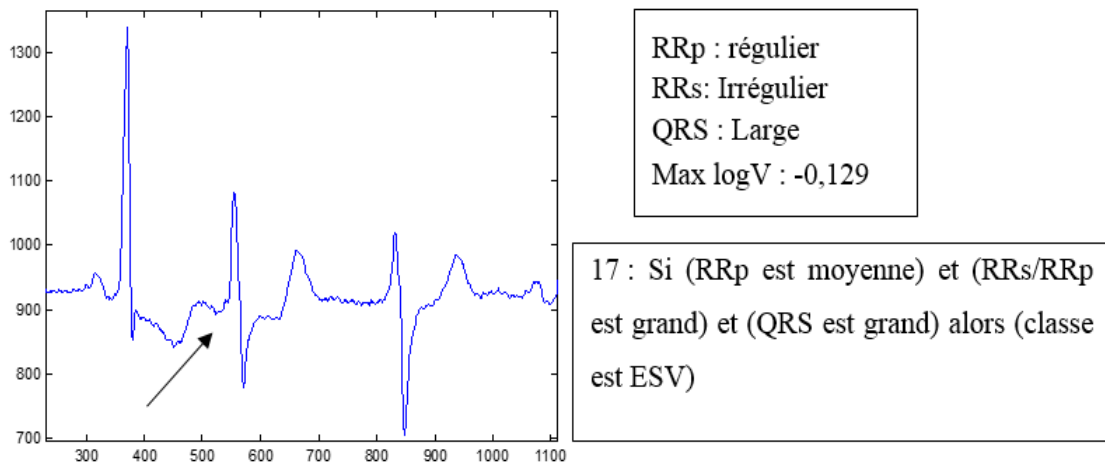


FIGURE 4.15 – Caractéristique d'un battement ESV

La figure 4.16 illustre un battement normal avec un repos compensatoire qui a été bien détecté par le classifieur HMM mais qui a été mal classifié par le classifieur neuro flou (Faux Négatif) en activant la règle 16 et 6 avec la règle 16 dominante. Le classifieur HMM nous a donné une probabilité de  $-4,212$  Pour le modèle normal est  $-5,900$  pour le modèle ESV ; qui sont des probabilités proche mais reste la probabilité du modèle normal la plus élevée.

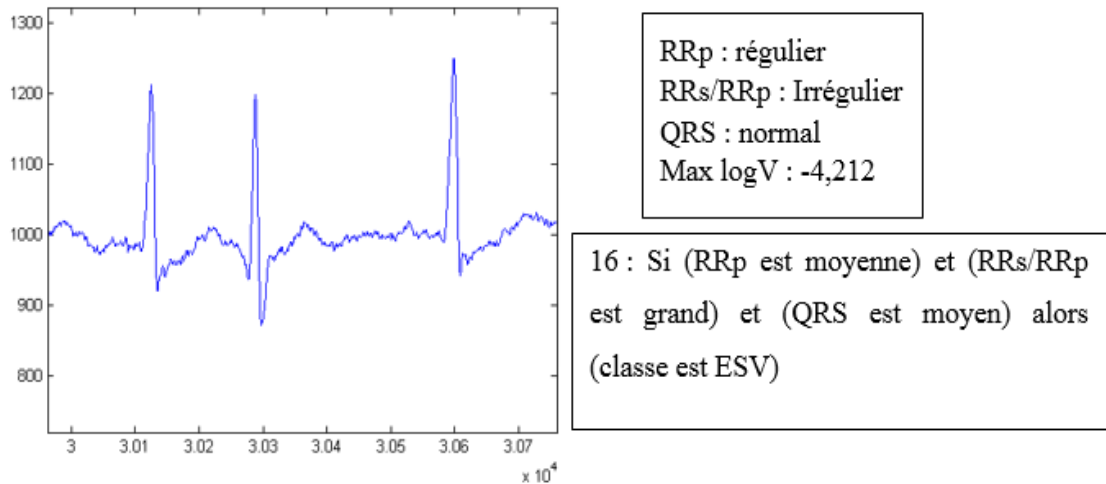


FIGURE 4.16 – Caractéristique d’un battement normal

La figure 4.17 illustre un battement ESV qui a été bien détecté par le classifieur HMM mais qui a été mal classifié par le classifieur neuro flou (Faux positif) en activant la règle 14. Le classifieur HMM nous a donné une probabilité de  $-3,401$  pour le modèle ESV et  $-4,417$  pour le modèle normal; qui sont des probabilités proches mais reste la probabilité du modèle ESV la plus élevée.

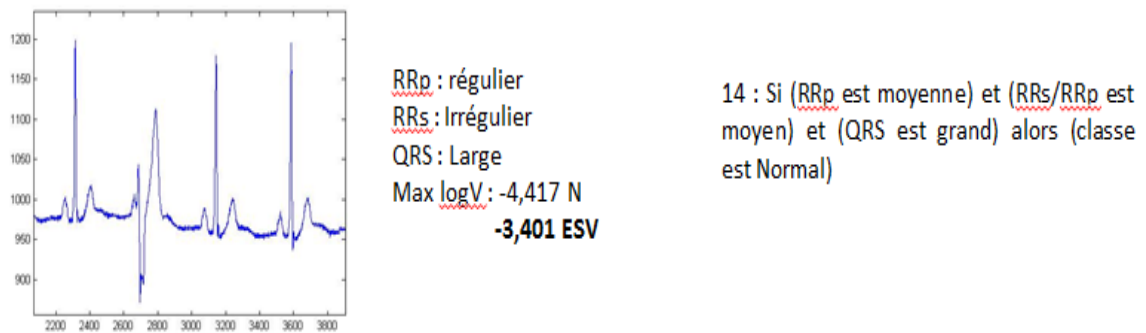


FIGURE 4.17 – Caractéristique d’un battement ESV bien détecté par approche HMM et non détecté par approche neuro floue

L’utilisation d’un classifieur neuro-flou peut nous fournir un système transparent et interprétable en générant une base de connaissance (18 règles) pour justifier la classification mais avec un taux de reconnaissance moins significatif (85,50%).

## Deuxième partie

### Considération 2 : Segmentation d'images par HMM

---

## 3 Introduction

Les méthodes de segmentation ont été l'objet de nombreuses recherches, car elles débouchent sur de multiples applications pratiques. L'imagerie médicale est tout particulièrement concernée. La reconnaissance des cellules, notamment des cellules tumorales tels les plasmocytes et les blastes est indispensable pour l'orientation diagnostique de certaines maladies cancéreuses comme le myélome et la leucémie respectivement. La segmentation des images cytologiques sanguines et médullaires constitue un axe de recherche qui prend de plus en plus d'ampleur, cet engouement est dû essentiellement aux informations cruciales que peut apporter l'analyse de telles images aux experts humains (médecins cytologistes).

Plusieurs raisons nous ont motivées pour adopter l'approche des chaînes de Markov (HMMs) comme méthode de segmentation. L'une d'entre elles est l'attention croissante portée sur le rôle du contexte spatial dans la segmentation des images. De même les modèles statistiques et plus précisément les modèles markoviens montrent leur efficacité pour la segmentation des images en régions.

L'objectif de cette deuxième partie du travail est de présenter de façon détaillée le principe de la segmentation supervisée d'images basée sur le modèle des chaînes de Markov cachées. Les principes fondamentaux nécessaires à la compréhension des nouveaux modèles présentés dans le deuxième chapitre y sont entièrement utilisés.

## 4 Parcours de Hilbert-Peano

La première étape de tous les algorithmes de segmentation d'images basés sur le modèle des HMM consiste donc à transformer l'image en un vecteur. Tous les traitements s'effectuent sur ce vecteur, puis, une fois la segmentation de la chaîne achevée, l'image thématique est finalement reconstruite. Dans la suite, nous utilisons donc une structure 1D pour représenter une image et pour s'adapter à la structure monodimensionnelle d'une chaîne. Quand nous parlerons d'image, il faudra donc comprendre la séquence issue de la transformation de l'image par le parcours de Hilbert-Peano.

Le parcours de Hilbert-Peano appartient à la famille des « Space Filling Curves » (SFCs), qui incluent les parcours en zigzag, le balayage en Z et les parcours standards (ligne par ligne, colonne par colonne) [56]. Les SFCs ont été employées dans une grande variété d'applications de traitement d'image telles que la compression,

le tramage demi-tons (« half-toning »), la reconnaissance de formes et l'analyse de texture. Tous ces parcours peuvent être employés pour convertir un ensemble 2D de pixels en séquence 1D.

Néanmoins, le parcours de Hilbert-Peano fournit une bonne exploitation du voisinage 2D, essentiellement grâce à la contrainte qu'il respecte : « tout pixel de la chaîne possède comme voisins, deux pixels voisins sur l'image ». Cette propriété, ainsi que l'aspect pseudo-aléatoire des changements de directions, implique que le parcours de Hilbert-Peano fonctionne bien (du point de vue statistique) pour de nombreuses familles d'images. La Figure 4.18 présente la construction du parcours de Hilbert-Peano pour une image 8x8.

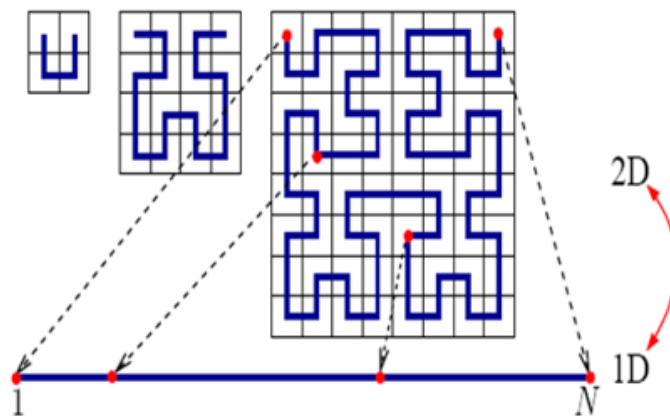


FIGURE 4.18 – Construction du parcours de Peano pour une image de 8x8

La segmentation des images microscopiques de cytologie consiste en l'extraction des différents constituants cellulaires. L'image acquise depuis le microscope englobe les globules rouges (hématies), les globules blancs (leucocyte), les plaquettes et le plasma (fond ou background). Les plaquettes(ou thrombocytes) sont de très petites cellules sans noyau que nous avons jugé utile de les supprimer. L'annulation de cette classe est justifiée par deux points : ce type de cellules ne contribue nullement dans le diagnostic du médecin expert et leurs tailles sont négligeables par rapport aux autres constituants de l'image. Le leucocyte est formé d'un noyau et d'un cytoplasme ; ce qui revient à dire que notre image sera composée de quatre classes : noyau, cytoplasme, globule rouge et fond 4.19.



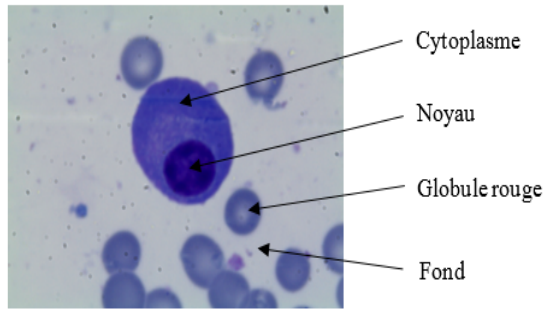


FIGURE 4.19 – Image cytologique médullaire

## 5 Travaux de l'état de l'art

La majorité des travaux dans la littérature ayant trait à ce domaine n'excède pas deux décennies, car cette discipline a beaucoup bénéficié du progrès d'acquisition d'images et des moyens informatiques, ainsi que le foisonnement des techniques de segmentation qui ont permis d'ouvrir une voie de développement très prometteuse. Nous présentons dans cette partie quelques travaux de segmentation et classification des globules blancs dans les images sanguines :

Une méthode non supervisée pour détecter la région du noyau était proposé par [57], en calculant l'histogramme dans le canal vert. Le cytoplasme est obtenu en utilisant une segmentation supervisée dans l'espace CIELab et les régions d'intérêt sont découpées à partir des noyaux. La caractérisation des cellules est réalisée en calculant la matrice de co-occurrence GLCM (Gray Level Co-occurrence Matrix). Pour réduire le nombre de caractéristiques une sélection des attributs avec forward sequential selection (fss) et backward sequential selection (bss) est réalisée avant d'entraîner le classifieur les différentes classes de cellules.

H. Zhang et al. et N.Guo et al. [58] et [59] ont utilisé des images multispectrales. Ces images sont stockées en norme RGB, ils ont ajouté des attributs spatiaux pour tirer une information spectrale pour chaque pixel. Ils ont effectué un seuillage sur l'image convertie au niveau de gris afin d'éliminer le fond (classe plasma). Ils ont utilisé l'algorithme sequential minimal optimization (smo) pour la sélection des attributs issus des bandes spectrales dans le but de réduire le temps d'apprentissage. L'approche de segmentation adaptée est basée sur une classification des pixels par l'algorithme SVM.

T. Markiewicz et al. [60] ont proposé un dénombrement des globules blancs dans

---

des images médullaires, dix différents types de globules blancs sont considérés. En premier lieu ils ont fait appel à des opérations morphologiques de fermetures et d'érosions afin de lisser les contours des cellules, après ils ont réalisé une segmentation couleur par ligne de partage des eaux en identifiant chaque bassin dans une imagerie, cette dernière contiendra une seule cellule. Une étape de caractérisation est nécessaire afin de calculer les attributs comme : taille, forme, texture, granulation, couleur et intensité des images associées aux différents types de cellules. Un classifieur de type SVM a été utilisé pour identifier les différentes classes.

M. Hamghalam [61] ont adopté une segmentation avec un seuillage pour identifier les limites du noyau, et ils ont utilisé l'algorithme des contours actifs pour reconnaître avec précision le cytoplasme. Ils ont fixé dans ce travail cinq types de globules blancs dans le frottis sanguin. Plusieurs caractéristiques (forme, texture, couleur) ont été utilisées pour identifier le leucocyte.

E. Wang et al. [62] considèrent que l'image compte six couleurs de base, une pour le cytoplasme, deux pour les globules rouges et trois pour les plaquettes et leucocytes. À l'aide de l'algorithme fuzzy clustering ils réalisent une distribution des pixels en fonction des couleurs considérés, ensuite ils ont fait une projection pour éliminer les globules rouges et le cytoplasme et ne garde que les leucocytes et les plaquettes. Enfin, une reconnaissance automatique des globules blancs est réalisée en supprimant les plaquettes à cause de leurs petites tailles.

La méthode de segmentation proposée par SH. Rezaatofghi et al. [63] est basée sur l'algorithme Gram-Schmidt orthogonalization, suivie d'une combinaison avec l'algorithme snake pour délimiter les régions noyau et cytoplasme. Plusieurs caractéristiques sont extraites, ils ont utilisé que les plus pertinents après une sélection par SFS (sequential forward selection) à partir des régions segmentées. Enfin une classification des cellules est faite par un classifieur neuronal afin de classer cinq types de globules blancs.

Les travaux de M. Benazzouz et al. [64] se sont basés sur la classification pixelaire des cellules médullaires afin de séparer les différents composants de l'image cellulaire (globule rouge, cytoplasme, noyau, fond).

Le modèle adopté est basé sur les deux stratégies suivantes ; classifier séparément les données issues de différentes sources pour ensuite fusionner les décisions ou bien de combiner ces données en vue de les classifier. Une caractérisation est réalisée afin de

---

fournir des descripteurs pertinents aux constituants cellulaires précédemment segmentés : cytoplasme et noyau. Ensuite un classifieur est construit pour identifier et dénombrer les différents types de cellules ce qui va aider à la reconnaissance de la pathologie ciblée.

Après leurs travaux basés essentiellement sur la texture, LB. Dorini et al. [65] simplifient l'image par régularisation des contours, cette opération est rendue possible grâce à l'opération Self-Dual Multiscale Morphological Toggle (SMMT). Ils segmentent le noyau avec deux méthodes bien connues : la ligne de partage des eaux et la méthode level set. Quant au cytoplasme, ils utilisent une analyse des granules pour le définir.

## 6 Expérimentation

Nous avons effectué un apprentissage sur un ensemble de 10 images cytologiques. Avec l'assistance d'un expert en cytopathologie, autant d'images ont été manuellement segmentées pour créer ce qu'on appelle image vérité terrain (figure 4.20) qui vont servir pour la comparaison avec les travaux de la littérature [64].

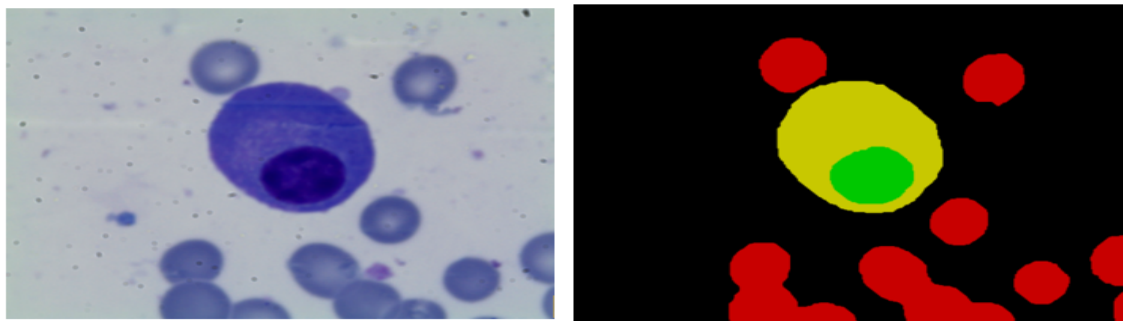


FIGURE 4.20 – (a) image originale (b) image vérité terrain

Dans la figure 4.20, nous avons opté pour des couleurs afin de distinguer les différentes classes (Noyau : Vert, Cytoplasme : Jaune, Globule rouge : Rouge et Background : Noir).

### 6.1 Base d'apprentissage

Nous avons remarqué l'existence de variabilité intra classe, particulièrement dans la classe cytoplasme avec des couleurs différentes, ce qui nous a poussés à penser et

créer une base d'apprentissage composée de 10 images.

La segmentation d'images basée sur les HMM nécessite tout d'abord la transformation de l'image 2D en une séquence d'observations 1D. Nous effectuons cette transformation à l'aide du parcours de Hilbert-Peano de l'image.

Chaque espace de l'image est transformée en vecteur 1D tout en respectant la notion de voisinage de chaque pixel (Figure 4.21).

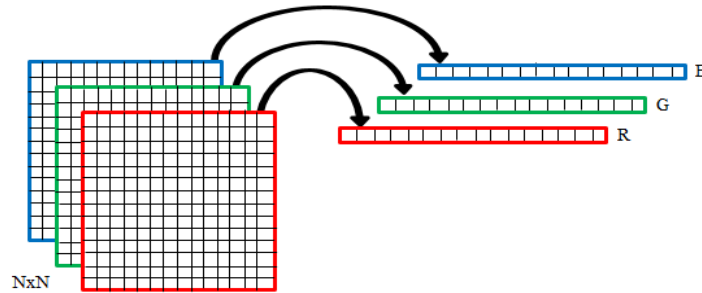


FIGURE 4.21 – Transformation 2D en 1D de l'image couleur

Nous avons préparé notre base d'apprentissage de façon qu'à partir de portions sélectionnées de chaque classe, nous enregistrons leurs coordonnées spatiales (n° ligne, n° colonne) et la classe de chaque pixel dans une base.

Chaque classe est représenté par un modèle HMM ; dont chaque modèle est composé de trois états cachés (R-G-B) et de 255 observations (intensité de couleur) (Figure 4.22).

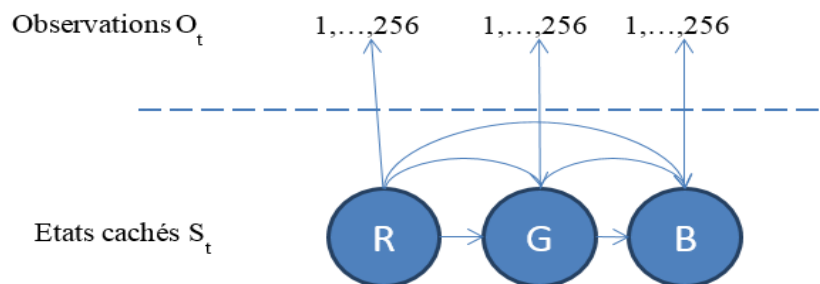


FIGURE 4.22 – Modèle HMM pour chaque classe

## 6.2 Résultats

Nous présentons dans la table 4.10 les résultats de la classification des pixels dans chaque canal distinctement. Nous pouvons rapidement statuer qu'il y a impossibilité de trouver un espace parfait ; c'est-à-dire qui se comporte bien en matière de

---

classification avec l'intégralité des classes.

La première observation est, que les scores plus faibles sont ceux de la classe des globules rouges, cela est justifié par la variabilité de ses pixels constitutifs, créant très souvent une confusion avec le cytoplasme.

TABLE 4.10 – Taux de reconnaissance des différentes classes par HMM

	Taux de reconnaissance du cytoplasme	Taux de reconnaissance du noyau	Taux de reconnaissance des globules rouges	Taux de reconnaissance du fond
Canal rouge	70,39%	23,93%	68,24%	88,52%
Canal vert	91,93 %	55,47%	70,57%	96,50%
Canal bleu	33,25%	56,62%	41,58%	90,25%
RGB	90,92%	93,85%	72,66%	95,79%

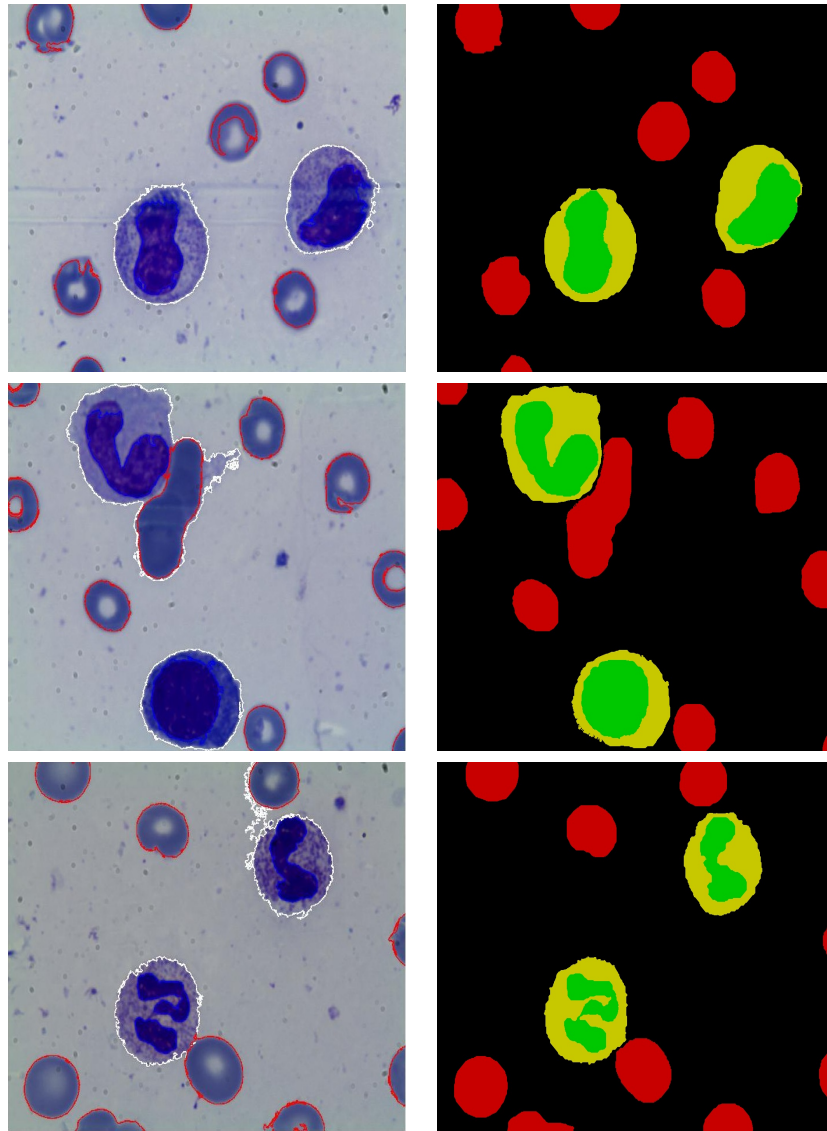
Une partie des résultats de segmentation visuels est présentée dans la figure 4.23. Un post-traitement est nécessaire pour améliorer la qualité de la segmentation obtenue par élimination des artefacts, et aussi en éliminant les petites zones rouges dans la région de cytoplasme.

Les principales erreurs ont été causées ou provoquées par les raisons suivantes. Tout d'abord, une confusion dans la classification se fait principalement entre les classes qui sont étroitement liées (globules rouges et dans certaines régions de cytoplasme), aucune distinction n'est faite entre elles puisque les données sont encore une fois similaire.

Les résultats les plus prometteurs ont été obtenus par la segmentation de la classe noyau, néanmoins la couleur du cytoplasme ressemble souvent à des globules rouges comme le montre le chevauchement entre ces deux classes [64] par conséquent, les résultats de segmentation sont parfois inexacts.

## 7 Conclusion

La première partie de ce chapitre présente la classification en utilisant les HMM sur des ensembles de battements cardiaques. Les modèles HMMs représentent deux types de battement en utilisant la base de données d'arythmie MIT / BIH. Nous avons constaté que le HMM est une méthode très appropriée pour classer les signaux ECG en produisant des valeurs de probabilité, qui donnent plus d'informations comparé à une simple décision binaire (oui / non) comme elle donne de meilleures performances par rapport à un classifieur à base de règles floue qui est un classifieur interprétable



Images segmentées

Images vérité terrain

FIGURE 4.23 – Résultats de segmentation.

---

mais cette transparence diminue ses performance de reconnaissance.

Mais d'un autre côté, dans le cas d'un classifieur HMM, quand un cardiologue reçoit deux probabilités très proches les uns des autres, les modèles ne sont pas discriminatoires donc un examen plus approfondi est nécessaire. Aussi, si les probabilités sont très faibles, nous pouvons distinguer la présence d'un nouveau type de battement qui n'a pas déjà été dans les données de l'apprentissage.

Nous avons traité en deuxième partie de ce chapitre un autre type de données qui sont les images cytologiques, ou nous avons opté pour une classification pixelaire dans le but de segmenter des images bruitées (aucun prétraitement), qui contiennent souvent des scènes complexes, traduites par un chevauchement ou une condensation de cellules. Les résultats obtenus ont pu atteindre une précision de segmentation satisfaisante.

# Conclusion Générale



Dans ce manuscrit nous nous sommes intéressés à deux approches de classification, une approche probabiliste basée sur les modèles de Markov cachés, et une approche floue basée sur les systèmes d'inférences flou. En effet, les HMMs permettent de modéliser chaque cas indépendamment (modèle HMM pour cas sain et d'autres pour les cas pathologiques) ainsi d'extraire une information probabiliste sur les données traités tandis que l'approche flou nous a permis d'avoir des résultats interprétables et modélisable par des règles linguistique.

Pour valider notre démarche, nous avons appliqué nos propositions sur des données séquentielles pour une classification supervisée des signaux ECG afin de détecter les battements extrasystoles ventriculaires et pour une classification pixellaire des images cytologiques afin de segmenter et localiser le noyau, les globules blanches et rouges.

Les modèles de Markov nous ont permis d'extraire des probabilités afin de pouvoir différencier les battements ESV des battements normaux par le calcul de maximum de vraisemblance de chaque séquence présentée. Le passage par une étape de symbolisation nous a permis d'intégrer dans notre système de classification la notion d'interprétabilité.

D'une autre part, un classifieur flou peut être un outil pratique dans le processus de diagnostic. Il se compose de règles linguistiques qui sont faciles à interpréter par l'expert humain. Ceci est très important pour les systèmes de prise de décision, car les experts n'acceptent pas une évaluation sur ordinateur, à moins qu'ils comprennent pourquoi et comment une recommandation a été donnée mais d'une autre part le taux de reconnaissance décroît. Les avantages des classifieurs flous peuvent être principalement vus dans leur interprétabilité linguistique, la manipulation intuitive et la simplicité, qui sont des facteurs importants pour l'acceptation et l'utilisation d'une solution.

L'utilisation d'un classifieur neuro-flou peut nous fournir un système transparent et interprétable en générant une base de connaissance (les règles) pour justifier la classification mais avec un taux de reconnaissance moins significatif.

La deuxième partie dans ce manuscrit a été consacrée à la classification pixellaire par les modèles de Markov cachés dans le but de segmenter des images bruitées (aucun prétraitement), qui contiennent souvent des scènes complexes, traduites par un chevauchement ou une condensation de cellules. La première étape de tous les algorithmes de segmentation d'images basés sur le modèle des HMM consiste à transformer l'image en un vecteur par l'algorithme d'hillbert et Péano. Tous les traitements s'effectuent sur ce vecteur, puis, une fois la segmentation de la chaîne

achevée, l'image thématique est finalement reconstruite.

## **Perspectives**

Dans la continuité directe de notre travail de thèse, deux groupes de travaux sont envisagés, Le premier est lié à l'aspect représentation de la connaissance et le deuxième est lié à la classification où nous pouvons proposer plusieurs pistes de recherche nécessitant plus d'attention.

Concernant la représentation, une piste intéressante est l'exploitation ( la prise en compte d'autres attributs) d'images (contours, texture, . . .) et de signaux, aussi la généralisation de notre modèle de classification sur d'autres pathologies cardiaques peut enrichir le système afin de couvrir la majorité de cardiopathies existantes.

D'autre part une hybridation avec d'autres classifieurs pour fixer les paramètres initiaux des modèles HMMs créés peut améliorer les résultats obtenus. Aussi, une quantification floue des intervalles des différents attributs peuvent apporter une meilleur interprétabilité et transparence au système de classification.

Une autre alternative intéressante pour les travaux futures concernerait, l'investigation dans l'aspect relative à l'apprentissage en ligne dans les systèmes ambulatoires par les modèles Markoviens font partie de nos perspective.

# Bibliographie

- [1] Paul Mather and Brandt Tso, *Classification methods for remotely sensed data*, CRC press, 2009.
- [2] Yue Wu, Biaobiao Zhang, Jiabin Lu, and KL Du, “Fuzzy logic and neuro-fuzzy systems : A systematic introduction,” *International Journal of Artificial Intelligence and Expert Systems*, vol. 2, no. 2, pp. 47–80, 2011.
- [3] Disha Sharma, “Designing and modeling fuzzy control systems,” *International Journal of Computer Applications*, vol. 16, no. 1, pp. 46–53, 2011.
- [4] Andrew R Houghton and David Gray, *Maîtriser l’ECG : de la théorie à la clinique*, Elsevier Health Sciences, 2011.
- [5] J Robert Beck, “Medical decision making : 20 years of advancing the field,” *Medical Decision Making*, vol. 21, no. 1, pp. 73–75, 2001.
- [6] Ke-Lin Du and Madisetti NS Swamy, *Neural networks in a softcomputing framework*, Springer Science & Business Media, 2006.
- [7] Michel Jambu, *Introduction au data mining : analyse intelligente des données*, Eyrolles, 1998.
- [8] Stéphane Tufféry, *Data mining et statistique décisionnelle : l’intelligence dans les bases de données*, Editions Technip, 2005.
- [9] Nicholas Roy and Andrew McCallum, “Toward optimal active learning through monte carlo estimation of error reduction,” *ICML, Williamstown*, pp. 441–448, 2001.
- [10] Gérard Dreyfus, Jean-Marc Martinez, Manuel Samuelides, Mirta B Gordon, Fouad Badran, and Sylvie Thiria, *Apprentissage statistique : Réseaux de neurones-Cartes topologiques-Machines à vecteurs supports*, Editions Eyrolles, 2011.
- [11] Haytham Elghazel, *Classification et Prévission des données hétérogènes : Application aux trajectoires et séjours hospitaliers*, Ph.D. thesis, Lyon 1, 2007.

- 
- [12] Mireille Cléret, Pierre Le Beux, and Franck Le Duff, “Les systèmes d’aide à la décision médicale,” *Les Cahiers du numérique*, vol. 2, no. 2, pp. 125–154, 2001.
- [13] Donald Michie, David J Spiegelhalter, and Charles C Taylor, “Machine learning, neural and statistical classification,” 1994.
- [14] Edmundo Bonilla Huerta, *Logique floue et algorithmes génétiques pour le pré-traitement de données de biopuces et la sélection de gènes*, Ph.D. thesis, Université d’Angers, 2008.
- [15] Igor V. Cadez, Scott Gaffney, and Padhraic Smyth, “A general probabilistic framework for clustering individuals and objects,” in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2000, KDD ’00, pp. 140–149, ACM.
- [16] Dan Buzan, Stan Sclaroff, and George Kollios, “Extraction and clustering of motion trajectories in video,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. IEEE, 2004, vol. 2, pp. 521–524.
- [17] Igor Cadez, David Heckerman, Christopher Meek, Padhraic Smyth, and Steven White, “Visualization of navigation patterns on a web site using model-based clustering,” in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2000, KDD ’00, pp. 280–284, ACM.
- [18] Emmanuel Augustin, *Reconnaissance de mots manuscrits par systèmes hybrides : Réseaux de neurones et modèles de Markov cachés*, Ph.D. thesis, Paris 5, 2001.
- [19] G David Forney Jr, “The viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [20] G. David Forney Jr., “The viterbi algorithm : A personal history,” *CoRR*, vol. abs/cs/0504020, 2005.
- [21] L.R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

- [23] Tood K Moon, “The expectation-maximization algorithm,” *Signal processing magazine, IEEE*, vol. 13, no. 6, pp. 47–60, 1996.
- [24] Jeff A. Bilmes, “A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models,” Tech. Rep., 1997.
- [25] Leonard E. Baum and J. A. Eagon, “An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology,” *BULLETIN OF AMERICAN MATHEMATICAL SOCIETY*, vol. 73, pp. 360–363, 1967.
- [26] Rodrigo V. Andreão, Sandra M. T. Muller, Jérôme Boudy, Bernadette Dorizzi, Teodiano F. Bastos-Filho, and Mário Sarcinelli-Filho, “Incremental hmm training applied to ecg signal analysis,” *Comput. Biol. Med.*, vol. 38, no. 6, pp. 659–667, June 2008.
- [27] Ping Wang, Chu Sing Lim, Sunita Chauhan, Jong Yong A Foo, and Venkataraman Anantharaman, “Phonocardiographic signal analysis method using a modified hidden markov model,” *Annals of biomedical engineering*, vol. 35, no. 3, pp. 367–374, 2007.
- [28] Sheng-Fu Liang, Ching-Fa Chen, Jian-Hong Zeng, and Shing-Tai Pan, “Application of genetic algorithm and fuzzy vector quantization on eeg-based automatic sleep staging by using hidden markov model,” in *Machine Learning and Cybernetics (ICMLC), 2014 International Conference on*. IEEE, 2014, vol. 2, pp. 567–572.
- [29] C Xydeas, P Angelov, S Chiao, and M Reoulas, “Advances in eeg signals classification via dependant hmm models and evolving fuzzy classifiers,” *International Journal on Computers in Biology and Medicine, special issue on Intelligent Technologies for Bio-informatics and Medicine*, vol. 36, no. 10, pp. 1064–1083, 2006.
- [30] Samar Krimi, Kaïs Ouni, and Noureddine Ellouze, “Using hidden markov toolkit for arrhythmia recognition,” in *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*. IEEE, 2012, pp. 1–5.
- [31] Shing-Tai Pan, Tzung-Pei Hong, and Hung-Chin Chen, “Ecg signal analysis by using hidden markov model,” in *Fuzzy Theory and it’s Applications (iFUZZY), 2012 International Conference on*. IEEE, 2012, pp. 288–293.
- [32] Julien Thomas, Cédric Rose, and François Charpillet, “A support system for ecg segmentation based on hidden markov models,” in *En-*

- gineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE. IEEE, 2007, pp. 3228–3231.*
- [33] Cen Li and Gautam Biswas, “A bayesian approach to temporal data clustering using hidden markov models,” in *ICML*, 2000, pp. 543–550.
- [34] Louis Gacôgne, “Eléments de logique floue,” *Hermes*, 1997.
- [35] Jean-Raphaël Tong-Tong, *La logique floue*, Hermes, 1995.
- [36] Lotfi A Zadeh, “Fuzzy sets,” *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
- [37] Arnold Kaufmann, “Introduction à la théorie des sous-ensembles flous à l’usage des ingénieurs (fuzzy sets theory),” 1973.
- [38] Chin-Teng Lin, *Neural fuzzy systems : a neuro-fuzzy synergism to intelligent systems*, Prentice hall PTR, 1996.
- [39] Jyh-Shing Roger Jang, Chuen-Tsai Sun, and Eiji Mizutani, “Neuro-fuzzy and soft computing ; a computational approach to learning and machine intelligence,” 1997.
- [40] Timothy J Ross, *Fuzzy logic with engineering applications*, John Wiley & Sons, 2009.
- [41] MA Chikh, *Analyse du signal ECG par les reseaux de neuronrs et la logique floue : Applicationa la reconnaissance des VPC*, Ph.D. thesis, Doctorate Thesis, Tlemcen University, 2005.
- [42] Lotfi A Zadeh, “Fuzzy logic= computing with words,” *Fuzzy Systems, IEEE Transactions on*, vol. 4, no. 2, pp. 103–111, 1996.
- [43] Anne Magaly de Paula Canuto, *Combining neural networks and fuzzy logic for applications in character recognition.*, Ph.D. thesis, University of Kent at Canterbury, 2001.
- [44] Sushmita Mitra and Yoichi Hayashi, “Neuro-fuzzy rule generation : survey in soft computing framework,” *Neural Networks, IEEE Transactions on*, vol. 11, no. 3, pp. 748–768, 2000.
- [45] Jyh-Shing Roger Jang, “Anfis : adaptive-network-based fuzzy inference system,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, no. 3, pp. 665–685, 1993.
- [46] N Benamrane, A Aribi, L Kraloua, and A Fréville, “Interprétation des images médicales par une approche neuro-flou-génétiques,” in *6ème*

*Conférence Francophone de Modélisation et Simulation. MOSIM06, Rabat, Maroc, 2006.*

- [47] “Mit-bih arrhythmia database,” <https://www.physionet.org/physiobank/database/mitdb/>, Accessed : 8 October 2015.
- [48] Stavros Tsakalidis, Vassilios Digalakis, and Leonardo Neumeyer, “Efficient speech recognition using subvector quantization and discrete-mixture hmms,” in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on.* IEEE, 1999, vol. 2, pp. 569–572.
- [49] Tin Lay Nwe, Say Wei Foo, and Liyanage C De Silva, “Speech emotion recognition using hidden markov models,” *Speech communication*, vol. 41, no. 4, pp. 603–623, 2003.
- [50] Valtino X Afonso, Willis J Tompkins, Truong Q Nguyen, and Shen Luo, “Ecg beat detection using filter banks,” *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 2, pp. 192–202, 1999.
- [51] Antti Koski, “Modelling ecg signals with hidden markov models,” *Artificial intelligence in medicine*, vol. 8, no. 5, pp. 453–471, 1996.
- [52] Ataollah Ebrahimzadeh and Ali Khazaei, “Detection of premature ventricular contractions using mlp neural networks : a comparative study,” *Measurement*, vol. 43, no. 1, pp. 103–112, 2010.
- [53] Nur Asyiqin Amir Hamzah and Rosli Besar, “Comparison of different mother wavelets in pvc detection using pnn,” in *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on.* IEEE, 2010, pp. 496–499.
- [54] Venkatachalam Mahesh, Arumugam Kandaswamy, Chandran Vimal, Balakrishnan Sathish, et al., “Ecg arrhythmia classification based on logistic model tree,” *Journal of Biomedical Science and Engineering*, vol. 2, no. 06, pp. 405, 2009.
- [55] Parham Ghorbanian, Ali Jalali, Ali Ghaffari, and Chandrasekhar Nataraj, “An improved procedure for detection of heart arrhythmias with novel pre-processing techniques,” *Expert systems*, vol. 29, no. 5, pp. 478–491, 2012.
- [56] Stéphane Derrode and Wojciech Pieczynski, “Signal and image segmentation using pairwise markov chains,” *Signal Processing, IEEE Transactions on*, vol. 52, no. 9, pp. 2477–2489, 2004.
- [57] Daniela Mayumi Ushizima Sabino, Luciano da Fontoura Costa, Edgar Gil Rizzatti, and Marco Antonio Zago, “A texture approach to

- leukocyte recognition,” *Real-Time Imaging*, vol. 10, no. 4, pp. 205 – 216, 2004, Imaging in Bioinformatics : Part {III}.
- [58] Hongbo Zhang, Libo Zeng, Hengyu Ke, Hong Zheng, and Qiongshui Wu, *Advances in Natural Computation : First International Conference, ICNC 2005, Changsha, China, August 27-29, 2005, Proceedings, Part II*, chapter A Novel Multispectral Imaging Analysis Method for White Blood Cell Detection, pp. 210–213, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [59] Ningning Guo, Libo Zeng, and Qiongshui Wu, “A method based on multispectral imaging technique for white blood cell segmentation,” *Computers in Biology and Medicine*, vol. 37, no. 1, pp. 70 – 76, 2007.
- [60] Tomasz Markiewicz, Stanisław Osowski, and Bożena Mariańska, *Adaptive and Natural Computing Algorithms : 8th International Conference, ICANNGA 2007, Warsaw, Poland, April 11-14, 2007, Proceedings, Part II*, chapter White Blood Cell Automatic Counting System Based on Support Vector Machine, pp. 318–326, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [61] Mohammad Hamghalam, M. Motameni, and A.E. Kelishomi, “Leukocyte segmentation in giemsa-stained image of peripheral blood smears based on active contour,” in *2009 International Conference on Signal Processing Systems*, May 2009, pp. 103–106.
- [62] En-yong Wang, Zhengpin Gou, Ai-min Miao, Shu-qing Peng, Zhenyang Niu, and Xin-lin Shi, *Fuzzy Information and Engineering Volume 2*, chapter Recognition of Blood Cell Images Based on Color Fuzzy Clustering, pp. 69–75, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [63] Seyed Hamid RezaTofighi and Hamid Soltanian-Zadeh, “Automatic recognition of five types of white blood cells in peripheral blood,” *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333 – 343, 2011.
- [64] Mourtada Benazzouz, Ismahan Baghli, and Med Amine Chikh, “Microscopic image segmentation based on pixel classification and dimensionality reduction,” *International Journal of Imaging Systems and Technology*, vol. 23, no. 1, pp. 22–28, 2013.
- [65] L.B. Dorini, R. Minetto, and N.J. Leite, “Semiautomatic white blood cell segmentation based on multiscale analysis,” *Biomedical and*



*Health Informatics, IEEE Journal of*, vol. 17, no. 1, pp. 250–256,  
Jan 2013.



## Résumé

Dans cette thèse nous présentons nos travaux qui s'articule autour du domaine d'aide au diagnostic médical plus particulièrement en traitement de séquences, réalisés sur deux types de données (signal ECG et image cytologique) avec des tâches différentes (classification, segmentation). Nos contributions concernent d'une part la proposition d'un classifieur supervisé dédié à la reconnaissance des extrasystoles ventriculaires par les modèles de Markov cachés où nous avons intégré une tâche de symbolisation des différents attributs utilisés afin d'augmenter l'interprétabilité de notre système pour justifier la prise de décision. Aussi, un classifieur neuro-floue (ANFIS) est implémenter pour pouvoir comparer l'aspect probabiliste dans les HMM et l'aspect interprétable de la logique floue. Les HMMs sont des modèles probabilistes qui fournis une information de probabilité sur la classe reconnues tandis que le classifieur neuro-floue (ANFIS) nous donne une transparence dans le système de reconnaissance par les règles générés.

D'autre part, l'apprentissage par ces mêmes approches est appliqué à la classification pixellaire des cellules cytologique dans un but de segmentation pour une classification future des cellules sanguines.

**Mots clefs:** Classification, interprétabilité, HMM, ANFIS, ECG, images cytologiques, segmentation

## Abstract

In this thesis we present our works that are focused on the field of medical diagnosis more particularly in the treatment of sequences, performed on two types of data (ECG signal and cytological images) on different tasks (classification and segmentation). Our contributions concern the proposal of a supervised classifier for the recognition of the ventricular premature beats by hidden Markov models, where we have integrated a symbolization stage for the different used attributes to increase transparency of our system to justify making of the decision. Also, a neuro-fuzzy classifier (ANFIS) is implemented to compare probabilistic side in the HMM and interpretable aspect of fuzzy logic. HMMs are probabilistic models that provided information on the likelihood of recognized class while the neuro-fuzzy classifier (ANFIS) gives transparency in the recognition system by the generation of the different rules. Also, the learning by the same approach is applied to the classification of pixels of the cytological images for segmentation goals for future classification of blood cells

**Keywords:** classification, interpretability, HMM, ANFIS, ECG, cytological images, segmentation

## المخلص

في هذه الأطروحة نقدم عملنا الذي يركز على ميدان المساعدة على التشخيص الطبي وخاصة في مجال معالجة المعلومات المتسلسلة، التي أجريت على نوعين من البيانات (تخطيط القلب والصور الخليوية) مع مهام مختلفة (تصنيف وتجزئة). مساهماتنا تتعلق من ناحية باقتراح مصنف مخصص للتعرف على نبضات القلب غير الطبيعية بواسطة نماذج ماركوف المخفية (HMM). حيث أننا أضفنا مهمة ترميز مختلف الخصائص المستعملة لزيادة إمكانية تفسير نظامنا لدعم اتخاذ القرار. أيضاً، يتم تطبيق المصنف العصبي الغامض (ANFIS) لمقارنة الجانب الاحتمالي الذي تمنحه نماذج ماركوف المخفية والجانب التفسيري للمصنفات الغامضة. نماذج ماركوف المخفية هي نماذج احتمالية تقدم معلومات عن أرجحية التعرف على الصنف في حين أن مصنف العصبي غامض يعطي الشفافية في النظام من خلال مختلف القواعد المنشأة. أيضاً التعلم بنفس هذه النماذج طبق من أجل تصنيف النقطة في الصور الخليوية لهدف التجزئة من أجل تصنيف الخلايا الدموية.

**الكلمات المفتاحية:** التصنيف، التفسير، HMM، ANFIS، ECG، الصور الخليوية، التجزئة.