

Chapitre III

INTERPRETATION

&&

DISCUSSION

DES RÉSULTATS

Présentation

Ce chapitre dresse la synoptique de la démarche préconisée, la présentation des différentes interfaces de l'application réalisée et le modèle HMM proposé dans notre application, ainsi que la discussion des expérimentations effectuées et l'interprétation des résultats obtenus.

III.1. Introduction

Dans ce chapitre nous allons implémenter le modèle des HMMs déjà présenté dans le chapitre deux et ensuite nous discutons les résultats obtenus de la classification des arythmies cardiaques.

III.2. Synoptique de la démarche préconisée

Cette démarche est caractérisée par **03** phases implantées :

- ✓ Préparation et codification des données.
- ✓ Apprentissage.
- ✓ Test et évaluation.

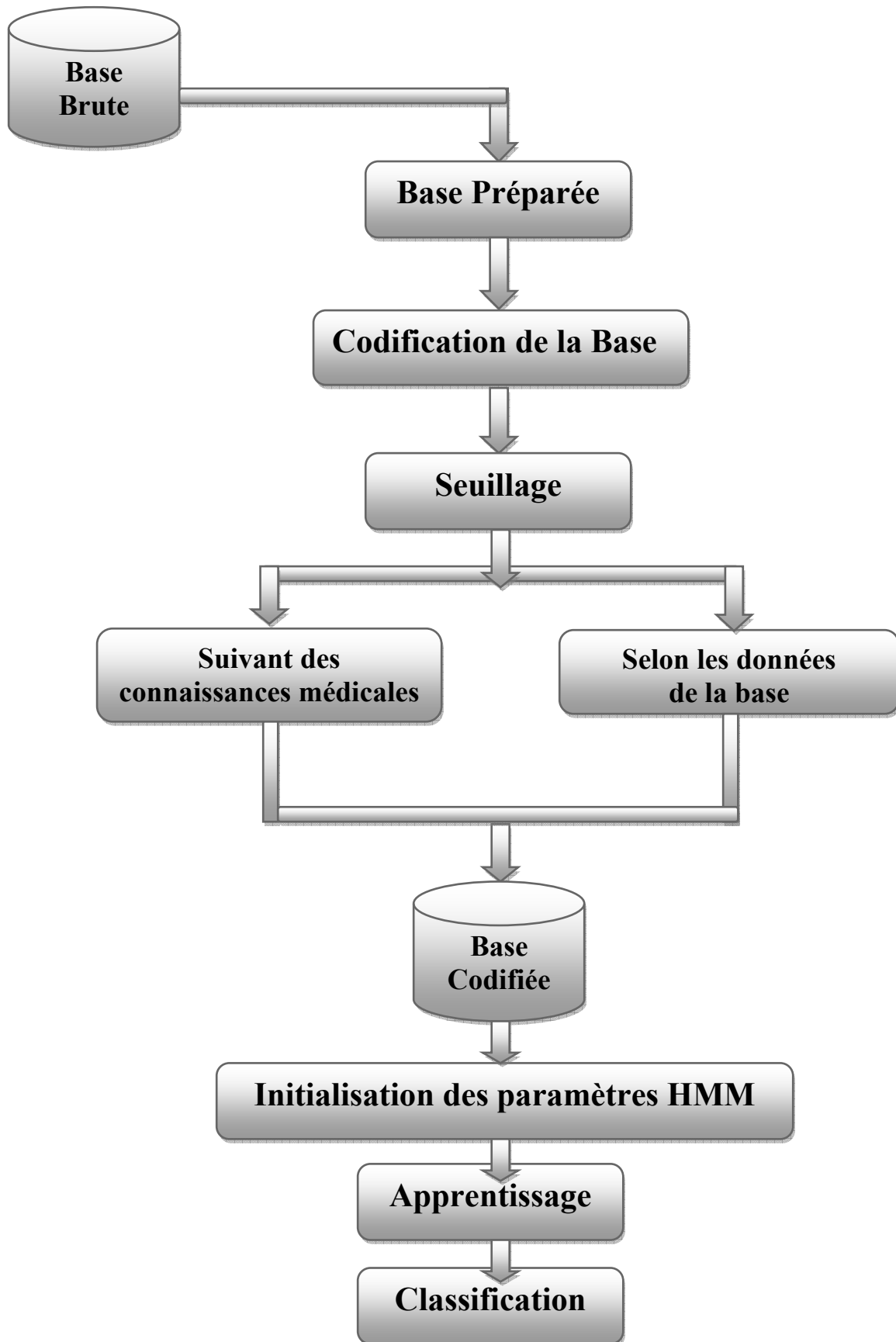


Figure III.1 : La synoptique de notre démarche.

III.3. Préparation de la base de données

III.3.1. Présentation de la base de données

Depuis 1975, les laboratoires de l'hôpital de Beth Israël à Boston et MIT ont réalisé une base de données MIT/BIH, qui a été commencée à être distribuée en 1980. Cette base de données contient 48 enregistrements extraits d'une demi-heure des enregistrements ambulatoires à deux voies d'ECG, obtenus à partir de 47 sujets étudiés par le laboratoire d'arythmie de BIH entre 1975 et 1979. Vingt-trois enregistrements ont été choisis au hasard d'un ensemble de 4000 enregistrements ambulatoires de 24 heures d'ECG rassemblées d'une population mélangée des patients hospitalisés (60 %) et des patients non hospitalisés (40 %) à l'hôpital de Beth Israël à Boston, les 25 enregistrements restants ont été choisis parmi les mêmes enregistrements mais qui en considération des arythmies rarement observés qui ont une signification cliniques. Les enregistrements ont été échantillonné à une fréquence $f_e = 360 \text{ Hz}$ avec une résolution de 11 bits sur une gamme de 10mV. Deux cardiologues ou plus ont indépendamment annoté chaque enregistrement, environ 110.000 annotations ont été inclus avec la base de données.

Enregistrement	Nombre Des Battements 'N'	Nombre Des Battements 'V'	Nombre Des Battements 'A'	Nombre Des Battements 'J'
200	100	100	0	0
234	100	0	0	50
114	0	0	10	0
232	0	0	100	0

Afin de construire une base de données, nous avons utilisé la base données MIT-BIH comme base de données brute, et la base de données LTSI caractérisée par l'algorithme tompkins (algorithme pour la détection des différentes ondes du signal ECG).

Sachant que notre base de données est annotée, donc on a conçu une matrice qui contient plusieurs battements depuis des enregistrements bien choisi afin d'avoir le maximum d'exemple pour chaque classe.

III. 3.2. Caractérisation du signal ECG

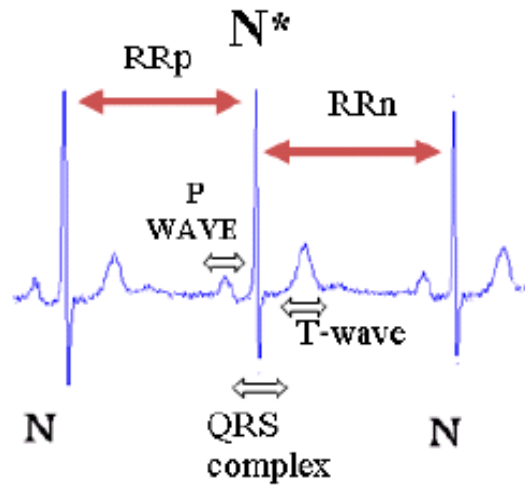


Figure III.2 : Différentes ondes du signal électrocardiogramme (ECG).

Paramètres	Définition
RRo	depuis le pique R précédent et le pique R présent
RRs	depuis le pique R présent et le pique R suivant.
Rapport	représente le rapport entre le RRo et le RRs
QRS	représente la largeur du complexe QRS

Figure III.3 : signification des paramètres contenus dans la base de données.

Chaque battement cardiaque est caractérisé par des descripteurs pertinents qui sont indispensables lors de la conception et l'implémentation de tout modèle de reconnaissance d'une anomalie cardiaque. La figure III.3 résume les quatre paramètres utilisés dans notre application et qui assurent une bonne distinction entre les différentes classes ciblées (N, V, A, J).

NB: 0 : c'est le cas Normal (**N**).

1 : c'est le cas extrasystole Ventriculaire (**V**).

2 : c'est le cas extrasystole Auriculaire (**A**).

3 : c'est le cas Jonctionnelle (**J**).

III.3.3.Choix des paramètres

La caractérisation du battement cardiaque par des descripteurs pertinents est indispensable lors de la conception et l'implémentation de tout modèle de reconnaissance d'une anomalie cardiaque. De nombreuses approches citées dans la littérature ont porté sur la difficulté que représentent la mesure et le choix des paramètres pertinents du signal ECG et leur classification.

Donc, le bon choix des paramètres du vecteur d'entrée du classifieur est très important. Pour cela nous procédons à une analyse géométrique des données (bidimensionnelle et tridimensionnelle) pour voir le degré d'appartenance de chaque paramètre par rapport aux différentes classes ciblées (**N, V, A, J**).

III.3.3.1. Exploration de la base de données : Nous avons effectué deux expérimentations, en premier lieu une représentation bidimensionnelle des paramètres en fonction de classes et en second lieu une représentation tridimensionnelle de ces mêmes paramètres afin de cibler les paramètres qui assurent une bonne distinction entre les différentes arythmies.

III.3.3.1.1 : Représentation bidimensionnelle : En premier lieu nous avons testé séparément les paramètres du vecteur d'entrée du classifieur.

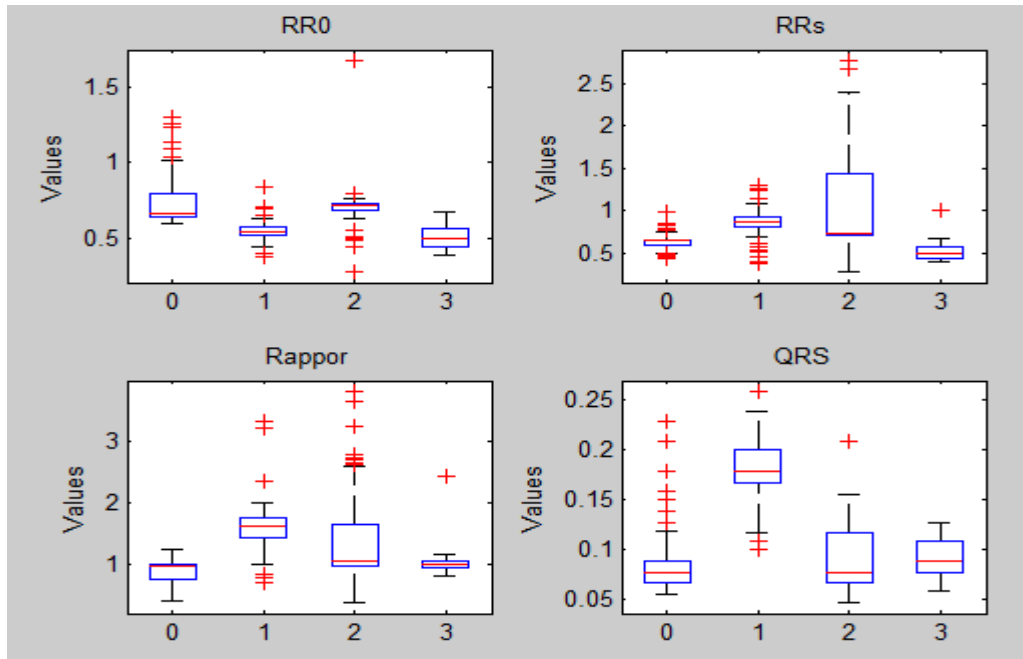


Figure III.4 : Distribution des paramètres en fonction de classes.

- + Concernant le paramètre **RR0**, nous voyons clairement que la distribution varie de façon similaire (figure III.4) pour les quatre classes (**N, V, A, J**), mais on remarque aussi qu'il y a une distinction entre la classe **V** et **A**.
- + Avec le paramètre **RRs**, on peut distinguer la classe **V** de la classe **J** (figure III.4), par contre si on prend les quatre classes (**N, V, A, J**), la distinction devient impossible et donc son effet est négligeable pour l'entrée du classifieur.
- + Le rapport, c'est la distribution des données RR_s/RR_0 .
- + Avec le paramètre **QRS**, nous différencions entre la classe **N** et **V** (figure III.4), ce qui est très logique du point de vue physiologique (dans le cas d'une extrasystole ventriculaire (**V**), ou l'onde **QRS** est large).

III.3.3.1.2. Représentation tridimensionnelle : Dans cette 2^{ème} expérimentation, nous avons testé l'effet de deux paramètres ensemble sur les 4 classes ciblées.

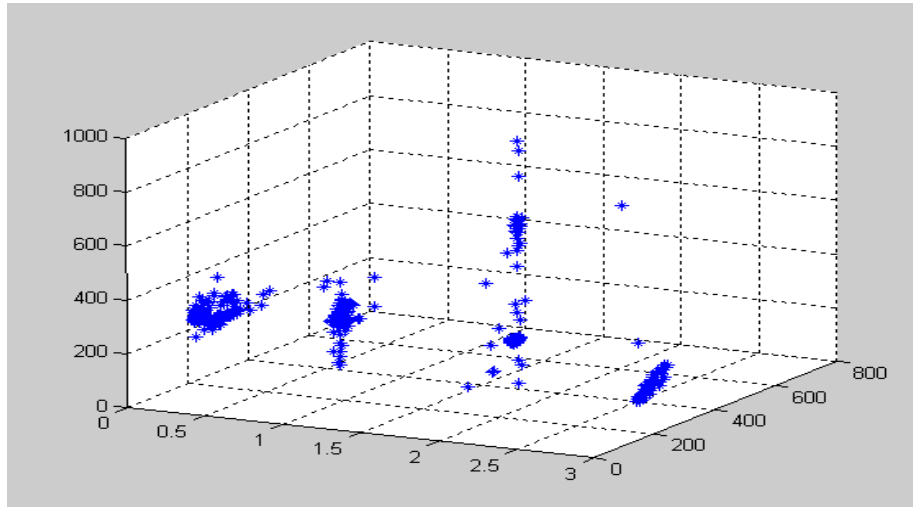


Figure III.5 : Distribution des données RR_O et RR_S en fonction de la classe.

Dans cette figure (figure III.5), on a regroupé le paramètre RR_O avec le paramètre RR_S pour chaque classe et on voit que chaque nuage de point est situé dans une région distincte de l'espace.

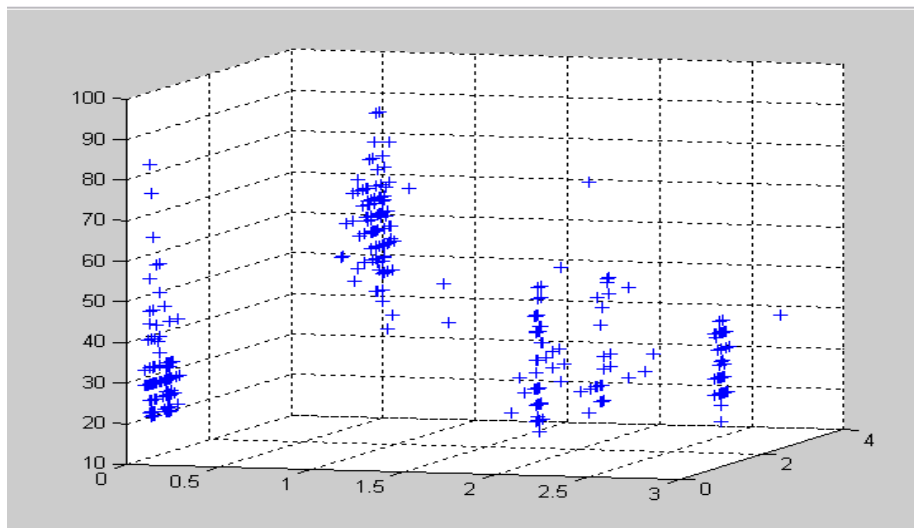


Figure III.6 : Distribution des données RR_0 et Rapport en fonction de la classe.

Dans cette figure (figure III.6), on voit une parfaite distribution des deux paramètres RR_0 et PP pour chaque classe du moment que chaque nuage de point est distingué par rapport aux autres.

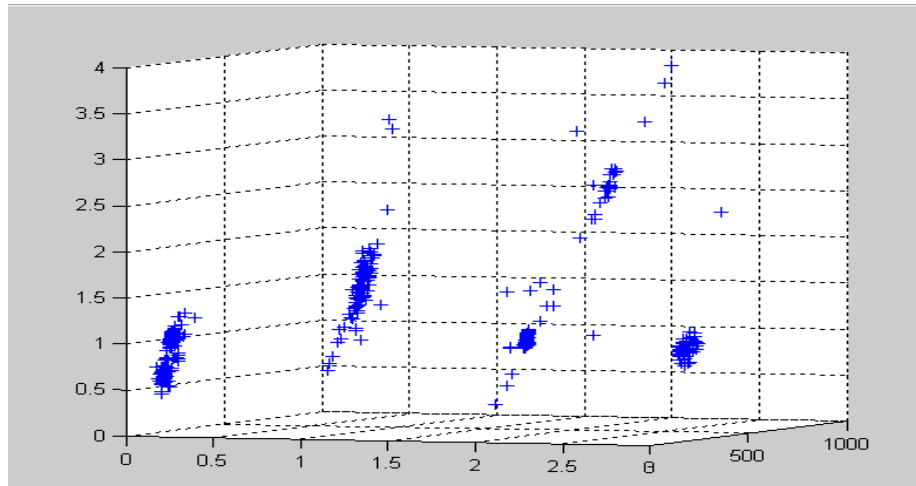


Figure III.7: Distribution des données RR, et Rapport en fonction de la classe.

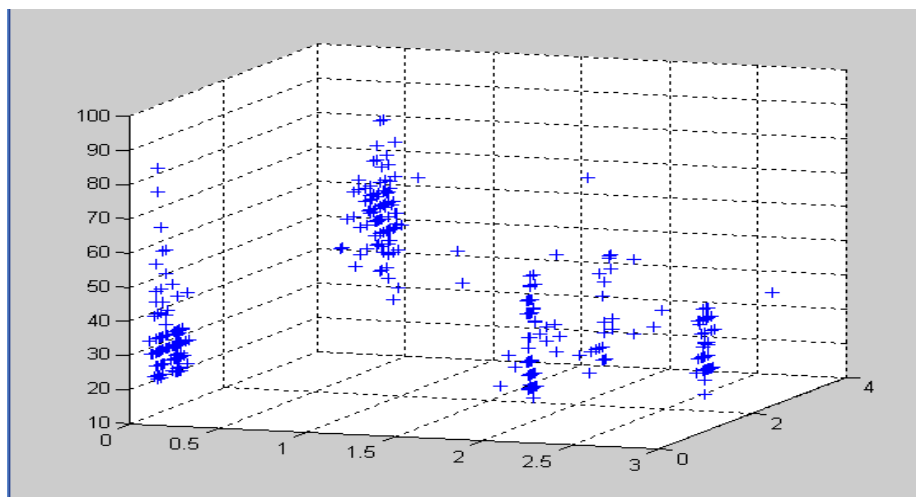


Figure III.8 : Distribution des données Rapport et QRS en fonction de la classe.

L'utilisation d'un seul paramètre au niveau des différentes classes qui a été évoqué dans la première expérimentation a donné un résultat peu satisfaisant, puisqu'il y a eu une confusion entre les classes alors qu'en associant deux paramètres dans la deuxième expérimentation nous avons pu constater une meilleure visibilité des données dans chaque classe, ceux qui nous amène à déduire que plus on augmente le nombre de paramètres plus on aura une distinction entre les différentes arythmies.

III.4. Application HMM

III.4.1. Problématique

Il est nécessaire pour notre application de choisir un modèle pour les séquences à reconnaître. Cette étape n'est pas facile ni évidente, vu le vocabulaire de cette méthode qui définit : les états, les transitions et les observations.

III.4.2. La démarche de la classification par la méthode HMM

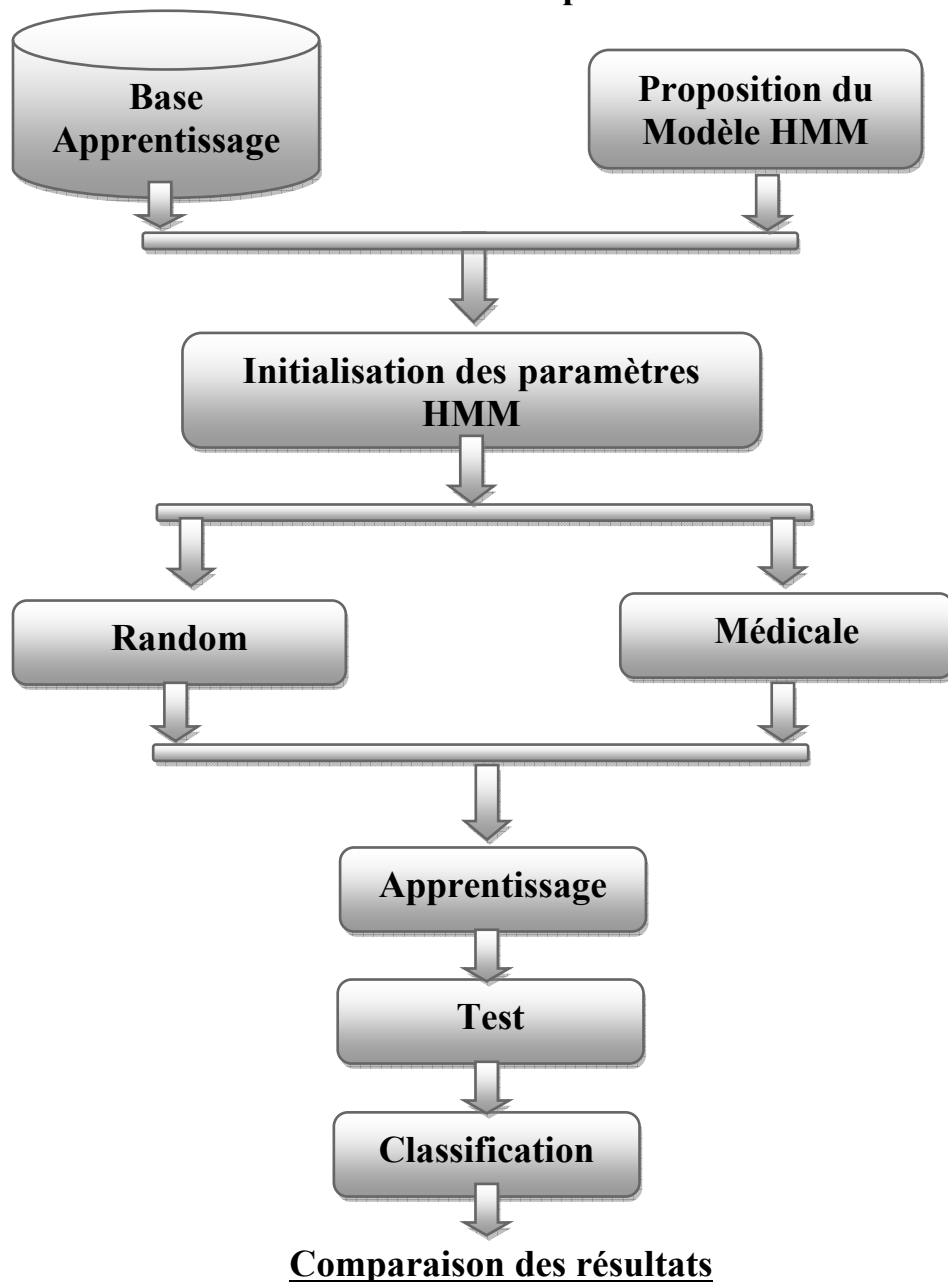


Figure III.9. La démarche de la classification par la méthode HMM.

III.4.3. Adaptation du modèle

III.4.3.1. Les états

Qui donnent à un instant donné, la description du système. Ils correspondent aux classes de notre modèle (classe Normale, Ventriculaire, Auriculaire ou Jonctionnelle). Ces quatre états modélisent l'état caché du système.

III.4.3.2. Les transitions

Qui sont les changements d'état. Ceux sont les probabilités de passer d'une classe vers une autre.

Nous avons proposé deux modèles, le premier correspond aux trois premières expérimentations qui seront présentées par la suite et le deuxième à la quatrième et la dernière expérimentation.

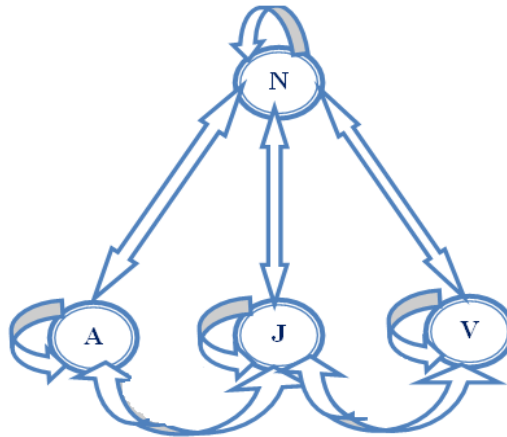


Figure III.10. Modèle1 HMM

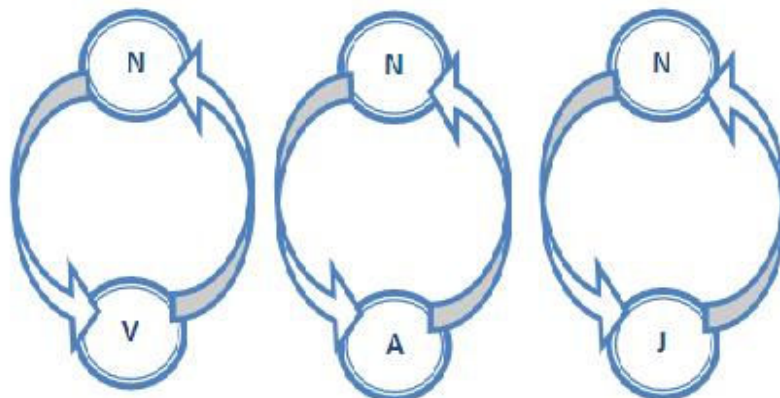


Figure III.11. Modèle2 HMM

III.4.3.3. Les observations

Cette méthode de reconnaissance de séquence repose sur les données acquises à partir de la base de données MIT. Celles-ci sont des données continues de bas niveau (en fait ce sont des données échantillonnées pour être traitée numériquement mais considérées comme continues parce qu'à valeurs réelles). Les observations correspondent donc aux différents paramètres du signal ECG (dans notre cas ils correspondent aux quatre paramètres cités précédemment **RR₀, RRs, Rapport, QRS**).

Ce type de données ne convient pas au modèle HMM. Il faut donc mettre en place un système de conversion des données, qui permet d'accorder une signification aux valeurs des données à tout instant. Cela passe par une méthode de discrétisation que nous appellerons transformation numérique-symbolique. La conversion d'une variable numérique en variable symbolique, c'est à dire définie sur un ensemble discret d'hypothèses, nécessite l'utilisation d'un modèle permettant de définir une classe de valeurs numériques par hypothèse. Nous présentons quelques techniques qui permettent d'effectuer cela:

III.4.3.3.1. Seuillage

Cette approche consiste à définir des seuils sur l'ensemble continu des valeurs numériques possibles, définissant des sous ensembles stricts sur lesquels une hypothèse particulière sera vraie. On définit ainsi un certain nombre (nombre de seuils + 1) de sous-ensembles **X_i** auxquels on peut associer une fonction caractéristique **μ_i** telle que :

$$\text{Pour tout point } x, \mu_i(x) = \begin{cases} 1 & \text{si } x \in X_i \\ 0 & \text{si } x \notin X_i \end{cases}$$

La

12: Modèle conversion seuillage. représente conversion

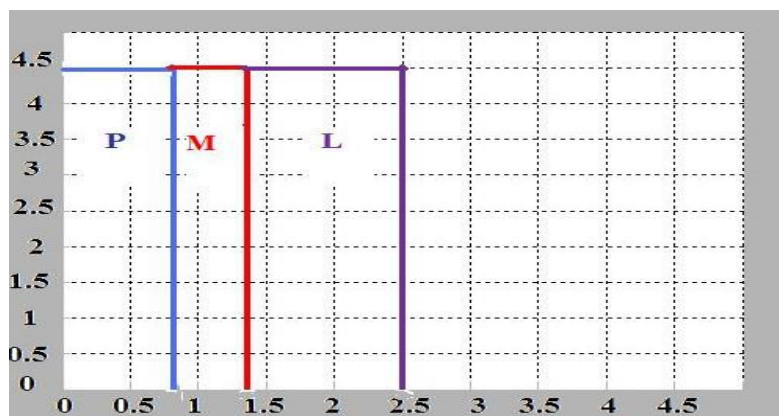


Figure III. de du QRS par la règle de du paramètre

QRS en une variable symbolique à 3 états.

Figure III. 12: Modèle de conversion du QRS par seuillage.

Les seuils sont définis à partir de connaissances expertes sur le système : soit avec l'analyse des données de la base (**minimum**, **medium**, **maximum**), soit par l'acquisition des connaissances médicales. Ce qui permet de faire une correspondance entre la valeur des données et l'état du système.

III.4.3.3.2. Sous-ensembles flous

Dans le cas précédent, la fonction caractéristique μ_i est une fonction binaire, spécifiant l'appartenance booléenne de chaque point à X_i . La théorie des sous-ensembles flous traite l'appartenance graduelle. Un sous-ensemble flou est défini par sa fonction d'appartenance μ_i de l'espace de définition des données S dans $[0,1]$. Pour tout x de S , $\mu_i(x)$ est la valeur dans $[0,1]$ représentant le degré d'appartenance de x au sous-ensemble flou X_i .

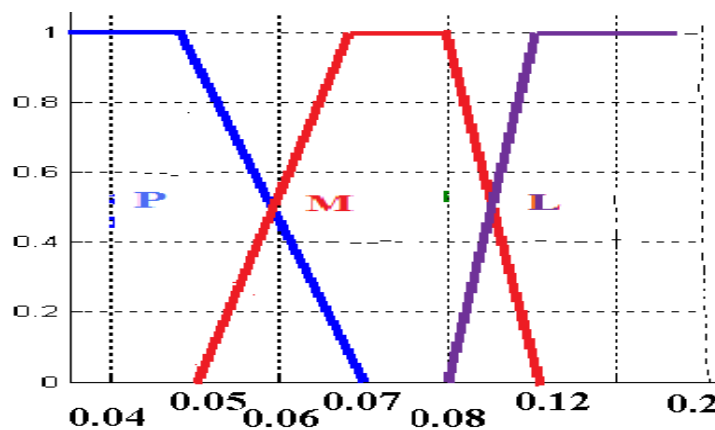


Figure III.13: Modèle de conversion du QRS par sous-ensembles flous.

Là encore, la définition des sous-ensembles flous est effectuée par expertise et elle est totalement dépendante de l'application.

III.5. Présentation du système

III.5.1. Le choix du langage de programmation

Nous avons réalisé notre application dans deux environnements de développement puissants : Java et Matlab.

Apparu en 1991, **Java un langage orienté objet**, un langage portable, c'est à dire qu'il s'adapte à une foule de plates-formes différentes, c'est sa qualité la plus importante. Une grande variété et versions de ce langage existent, nous avons choisi le **NetBeans**.

NetBeans est un environnement de développement intégré (EDI) pour Java. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, éditeur graphique d'interfaces et de pages web). C'est un langage de programmation visuel permettant la conception d'objets graphiques directement à l'écran, désignés par une bibliographie des composants. NetBeans ressemble à un atelier où l'on dispose d'une boîte à outils et d'un ensemble d'objets qui permet la construction des interfaces graphiques. Ce langage est multilingue; la version avec laquelle on travaille, NetBeans 6.8, est disponible en 23 langues (la plupart étant réalisées par la communauté d'utilisateurs). En plus NetBeans est open source, c'est un produit gratuit, sans aucune restriction quant à son usage (commercial et non-commercial, son code source est disponible pour la réutilisation sous la *Common Development and Distribution License CDDL*).

MATLAB est une abréviation de *Matrix LABORatory*. Écrit à l'origine, en Fortran, par *C. Mole*. La version actuelle, écrite en C par the MathWorks Inc.

MATLAB est un environnement puissant, complet, destiné au calcul scientifique. Il apporte aux ingénieurs, chercheurs et à tout scientifique un système interactif intégrant le calcul numérique et la visualisation. C'est un environnement performant, ouvert et programmable qui permet de remarquables gains de productivité et de créativité. Il permet de traiter les données sans aucune limitation de taille et de réaliser des calculs numériques et symboliques de façon fiable et rapide. Nous avons intégré une bibliothèque qui s'appelle "jmatlink", c'est une bibliothèque faramineuse qui permet de basculer entre Matlab et Java.

III.5.2. Description de l'interface

La réalisation technique commence par la présentation de la fenêtre principale, il s'agit de ce que voit l'utilisateur « un Menu » qui comporte quatre items horizontaux, représentant les différents traitements applicables, commençant par le chargement et la codification de la base de données ensuite l'apprentissage et la classification et enfin l'aide concernant l'application.

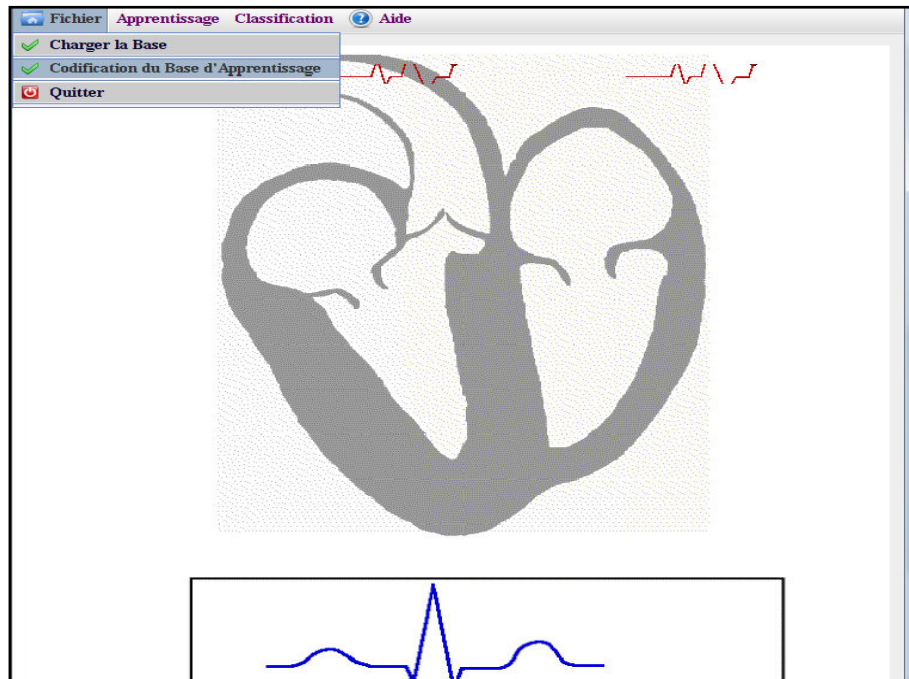


Figure III.14. Représentation de la fenêtre principale.

➤ Le chargement de la base de données

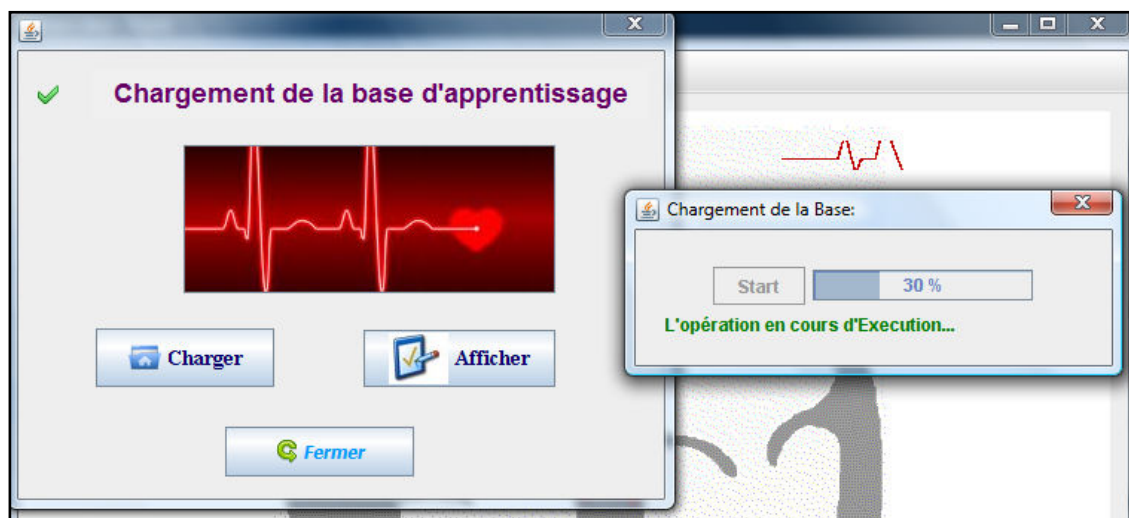


Figure III.15. Représentation du chargement de la base.

➤ Apprentissage

The screenshot displays a software window titled 'Chargement de la Base:'. A progress bar shows '100%' completion with the text 'L'opération en cours d'Execution...'. The main interface is divided into several sections:

- Modèle 1**: A dropdown menu for selecting the model.
- Observations:**: A plot showing an ECG signal with peaks labeled P, Q, R, S, and T. Below it are radio buttons for **RR0**, **RRs**, **QRS**, and **Rapport**.
- Affichage des Matrices de Transition et Emission**: Two tables are shown.

TRANSITION :				
	N	A	J	V
N	0.9	0	0.1	0
A	0	0.9	0.1	0
J	0.1	0	0.9	0
V	0	0.1	0	0.9

EMISSION :					
	P	M	L	R	I
N	0.9	0	0.1	0	0
A	0	0.9	0.1	0	0
J	0.1	0	0.9	0	0
V	0	0.1	0	0.9	0
- Initialisation**: A button to start the process.
- Selectionner l'une des Méthodes**: Radio buttons for **RANDOM**, **Double apprentissage**, and **A BASE MEDICAL**.
- Charger la Base d'Apprentissage**: A button to load the training data.
- Train**: A button to start training.
- Tester**: A button to test the model.
- Fermer**: A button to close the window.

Figure III.18. Représentation de la phase d'apprentissage.

Avant de lancer l'apprentissage, il faut tout d'abord sélectionner l'un des quatre modèles et choisir la méthode de codification (Random, Double apprentissage, à base médicale) pour pouvoir initialiser les paramètres d'HMM (la matrice de transition et la matrice d'émission qui seront affichées en cliquant sur le bouton OK). Après l'apprentissage des modèles nous pouvons les tester par la suite.

➤ Classification

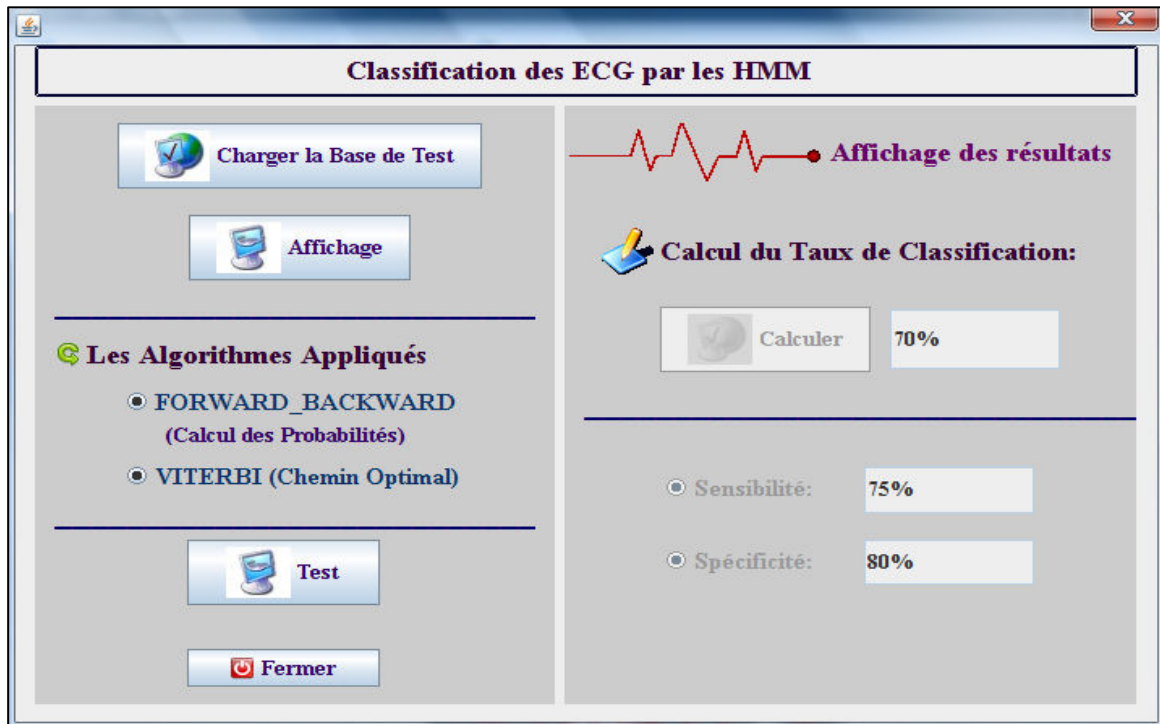


Figure III.19. Représentation de la classification.

III.6. Résultats et interprétations

III.6.1. 1^{ère} expérimentation

Notre premier essai consiste à créer un seul HMM pour toutes les classes. Cette expérience nous a confirmé que nous ne pouvons pas suivre ce principe, vu que le résultat obtenu nous a permis de donner un taux de classification global de 70% (pour les quatre classes) ce qui reste insuffisant par rapport à notre étude qui nécessite un taux de classification pour chacune des classes. De ce fait, nous avons décidé d'utiliser la méthode de reconnaissance par HMM présentée dans la littérature. Ce qui nous a conduits à lancer la deuxième expérimentation.

Performance	R	D
Taux %	70	70

Figure III.20. Tableau de critères de performance de l'expérimentation 1.

III.6.2. 2^{ème} expérimentation

Dans cette expérimentation nous créons un HMM pour chaque classe pour discriminer les différentes arythmies. Les battements anormaux que nous distinguons prioritairement des battements normaux sont les **ESV**, **ESH**, **ESA**.

Pour cela, nous employons quatre HMMs, les quatre modèles emploient quatre états et cinq observations :

- Le premier HMM est notre modèle de reconnaissance de battements normaux.
- Le second un modèle conçu pour la reconnaissance de battements d'extrasystoles ventriculaires.
- Le troisième est créé pour la reconnaissance de battements d'extrasystoles auriculaires.
- Le quatrième est fait pour la reconnaissance de battements d'extrasystoles jonctionnelles.

Ceci nous permet de calculer la vraisemblance d'une séquence d'observations par rapport à chaque HMM. Lorsqu'une séquence est analysée par le modèle HMM, nous calculons sa vraisemblance d'avoir été générée par ce dernier. Par conséquent étant donné une séquence d'observations, nous sommes capables de déterminer le modèle le plus probable qui a pu générer la séquence, au sens du maximum de vraisemblance.

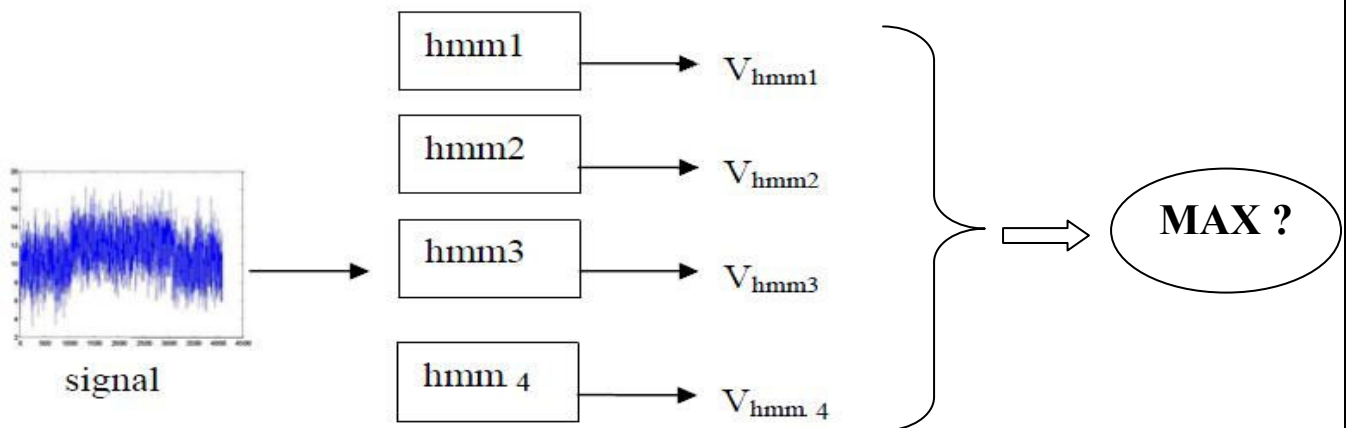


Figure III.21. Principe de la classification de signal par l'emploi de modèles HMM. Le signal est analysé par 4 modèles en termes de vraisemblances. On décide alors de la classe du signal à partir de la vraisemblance la plus élevée parmi les 4 modèles.

CHAPITRE III : IMPLEMENTATION && DISCUSSION DES RESULTATS

Les résultats de cette classification sont satisfaisants. En effet, les vraisemblances en sortie, calculées par nos modèles HMM contient des informations comparables, exploitables et facilement interprétables.

Classe	N		V		A		J	
	R	D	R	D	R	D	R	D
Taux %	50	50	63,33	100	60,25	60,25	60	60
Sensibilité %	50	50	63,33	100	60,25	60,25	60	60
Spécificité %	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure III.22. Tableau de critères de performance de l'expérimentation 2.

Pour la classe V elle est bien reconnue, ce qui est tout à fait normal puisqu'on a utilisé la largeur du QRS, considéré comme un paramètre discriminant, il distingue bien la classe V par rapport aux autres (QRS large → symbole L → Battement provienne d'une extrasystole ventriculaire).

Ce qui n'est pas le cas pour la classe N qui a été reconnue à 50%, n'empêche pas que c'est un bon résultat mais il n'est pas suffisant vue la richesse de notre base par la classe N, cela s'explique par le choix d'échantillons d'apprentissage et par le rapprochement des pourcentages de N, A et J, causés par les symboles P, M existants dans les trois classes citées précédemment, on a tendance à avoir une confusion des classes.

Le pouvoir discriminatoire du modèle HMM pour la reconnaissance des observations d'une séquence est bien justifié dans cette expérimentation.

III.6.3. 3^{ème} expérimentation

L'idée dernière cette expérimentation est de mesurer le degré d'importance de chaque symbole par rapport aux classes ciblées. Cependant, nous observons que chacun des symboles doit être pondéré par un HMM, donc on pourrait effectuer cette pondération au niveau du calcul de vraisemblance.

CHAPITRE III : IMPLEMENTATION && DISCUSSION DES RESULTATS

Rigoureusement, une telle expérience consiste à créer cinq HMMs correspondant aux symboles utilisés. Les cinq modèles emploient quatre états et cinq observations :

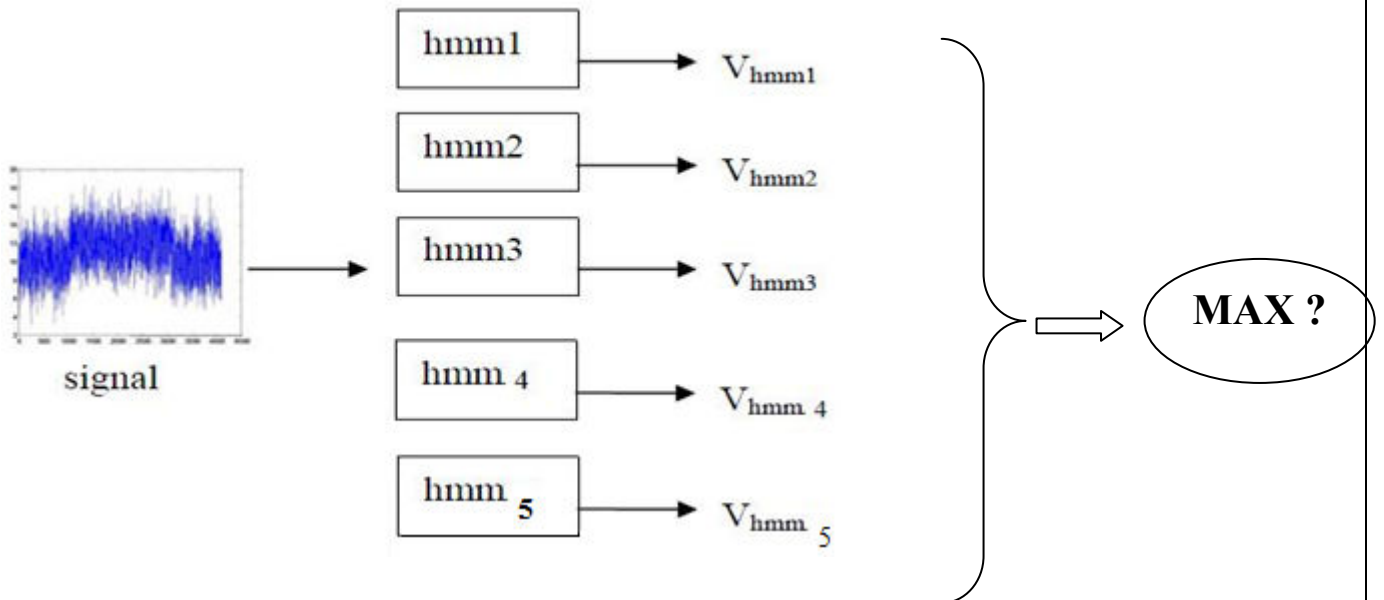


Figure III.23. Principe de la pondération des symboles définis pour chaque état par l'emploi de modèles HMM.

✓ *HMM qui mesure le degré d'importance du symbole P*

Classe	N		V		A		J	
Codification	R	D	R	D	R	D	R	D
Taux %	100	100	0	0	86,67	86,67	40	40
Sensibilité %	100	100	0	0	86,66	86,66	40	40
Spécificité %	NaN	NaN	0	0	NaN	NaN	NaN	NaN

Figure III.24. Tableau de critères de performance de l'expérimentation 3.1.

Le classifieur a bien reconnu le symbole P par rapport à la classe A et moyennement par rapport à la classe J et bien et mieux reconnu ce symbole pour la classe normale. Alors le P est un paramètre discriminant pour ces trois classes (c'est un paramètre). Or qu'il n'a pas reconnu du tout le cas V.

✓ *HMM qui fait la pondération du symbole M*

Classe	N		V		A		J	
	R	D	R	D	R	D	R	D
Codification								
Taux %	0	0	0	0	42,85	42,85	40	40
Sensibilité %	0	0	0	0	42,85	42,85	40	40
Spécificité %	0	0	0	0	NaN	NaN	NaN	NaN

Figure III.25. Tableau de critères de performance de l'expérimentation 3.2.

Pour le N, il n'y avait pas des échantillons de test. Et pour le V un nombre d'échantillons minimal(8).

✓ *HMM qui fait la pondération du symbole L*

Classe	N		V		A		J	
	R	D	R	D	R	D	R	D
Codification								
Taux %	0	0	91,67	100	0	0	0	0
Sensibilité %	0	0	91,66	100	0	0	0	0
Spécificité %	0	0	NaN	NaN	0	0	0	0

Figure III.26. Tableau de critères de performance de l'expérimentation 3.3.

Pour la classe N et la classe J, il n'y avait pas des échantillons de test. Et pour le A un nombre d'échantillons minimal(3).

✓ *HMM qui fait la pondération du symbole R*

Classe	N		V		A		J	
	R	D	R	D	R	D	R	D
Codification								
Taux %	70	70	0	0	45	45	40	40
Sensibilité %	65	65	0	0	41,34	41,34	41	41
Spécificité %	50	50	0	0	50,60	50,60	45,33	45,33

Figure III.27. Tableau de critères de performance de l'expérimentation 3.4.

Pour la classe V, il n'y avait pas des échantillons de test. Et pour la classe J un nombre d'échantillons minimal(8).

✓ *HMM qui fait la pondération du symbole I*

Classe	N		V		A		J	
	R	D	R	D	R	D	R	D
Codification								
Taux %	0	0	95	95	0	0	0	0
Sensibilité %	0	0	95	95	0	0	0	0
Spécificité %	0	0	NaN	NaN	0	0	0	0

Figure III.28. Tableau de critères de performance de l'expérimentation 3.5.

Pour la classe N et la classe J, il n'y avait pas des échantillons de test. Et pour le A un nombre d'échantillons minimal(10).

Ce symbole est considéré comme un paramètre bien discriminant pour la classe V.

III.5.4. 4^{ème} expérimentation

Dans la quatrième expérimentation, nous avons lancé trois HMMs, chacun correspond aux deux classes. Chaque modèle emploie donc deux états (suivant les deux classes étudiées) et cinq observations (figure III.11).

CHAPITRE III : IMPLEMENTATION && DISCUSSION DES RESULTATS

- Le premier HMM est un modèle de reconnaissance de battements normaux et des extrasystoles ventriculaires.
- Le second, un modèle conçu pour la reconnaissance de battements normaux et des extrasystoles auriculaires.
- Le troisième est créé pour la reconnaissance de battements normaux et des extrasystoles jonctionnelles.

Classe	N&V		N&A		N&J	
	R	D	R	D	R	D
Taux %	71.43	57.14	54.88	44.30	50.48	52.38
Sensibilité %	100	66.67	82.29	45.45	76.92	80
Spécificité %	60	50	15.79	42.86	24.53	27.27

Figure III.29. Tableau de critères de performance de l'expérimentation 4.

Les résultats de cette expérimentation ont validé ceux du 2^{ème} et 3^{ème} test puisque nous avons obtenu un taux de **54.88%** pour la classification entre la classe **N** et **A** et **50.48%** entre **N** et **J**, ce qui signifie que le classifieur n'a pas pu différencier entre ces trois classes d'une manière exacte, or qu'il a bien distingué entre la classe **N** et **V**, vu que le taux de classification de cette dernière est de **71%**.

Résumé :

Nous avons détaillé les traitements à appliquer dans toutes les étapes depuis la proposition du modèle jusqu'à la classification. Nous nous sommes focalisés ensuite sur l'interprétation des résultats et de l'état du système au cours de l'apprentissage pour montrer l'apport de cette méthode stochastique qui modélise les états et les observations. Celle-ci a été traitée avec des données réelles, provenant de la base de données **MIT/BIH** du laboratoire de l'hôpital de Beth Israël à Boston et du Mach Institut Technology.