



**HAL**  
open science

# Méthodes bio-inspirées appliquées à l'analyse temps réel des signaux biomédicaux

Mounia Hendel

► **To cite this version:**

Mounia Hendel. Méthodes bio-inspirées appliquées à l'analyse temps réel des signaux biomédicaux. Informatique [cs]. Université Mohamed Boudiaf des Sciences et de la Technologie - Mohamed Boudiaf d'Oran, 2017. Français. NNT: . tel-03042290

**HAL Id: tel-03042290**

**<https://hal.univ-lorraine.fr/tel-03042290v1>**

Submitted on 6 Dec 2020

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

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

# THÈSE

## En vue de l'obtention du Diplôme de Doctorat en Sciences

*Présenté par : Hendel Mounia*

### *Intitulé*

Méthodes bio-inspirées appliquées à l'analyse temps réel des signaux biomédicaux.

*Faculté* : *Mathématiques et Informatiques*  
*Département* : *Informatique*  
*Spécialité* : *Informatique*  
*Option* : *Reconnaissance des formes et intelligence artificielle*

*Devant le Jury Composé de :*

<i>Membres de Jury</i>	<i>Grade</i>	<i>Qualité</i>	<i>Domiciliation</i>
<i>Benyettou Mohammed</i>	<i>Professeur</i>	<i>Président</i>	<i>USTO-MB</i>
<i>Benyettou Abdelkader</i>	<i>Professeur</i>	<i>Encadrant</i>	<i>USTO-MB</i>
<i>Yann Guermmer</i>	<i>Professeur</i>	<i>Co-encadrant</i>	<i>LORIA-Nancy</i>
<i>Chikh Mohammed Amine</i>	<i>Professeur</i>		<i>Université de Tlemcen</i>
<i>Bessaid Abdelhafid</i>	<i>Professeur</i>	<i>Examineurs</i>	<i>Université de Tlemcen</i>
<i>Ahmed Fouatih Zoubir</i>	<i>Professeur</i>		<i>USTO-MB</i>
<i>Lebbah Yahia</i>	<i>Professeur</i>		<i>Université d'Oran</i>

*Année Universitaire : 2016-2017*

# Dédicace

*A mon epoux et mes enfants.*

*A mes parents et mon frère.*

*A ma belle famille.*

*A mes amis.*

# Remerciements

J'adresse mes vifs remerciements à mon encadreur M. le Professeur Abdelkader BENYETTOU. Je lui fais part de ma gratitude pour m'avoir accueillie dans son laboratoire de recherche SIMPA (Signal, Parole, et Image), et pour m'avoir fait profiter de son expérience de recherche. Je le remercie aussi pour sa disponibilité et la solidarité qu'il dispense à ses doctorants.

J'exprime également ma reconnaissance à M. le Professeur Yann GUERMEUR responsable de l'équipe ABC (Apprentissage et Biologie Computationnelle). Sa sollicitude a été permanente. Il a déployé beaucoup d'énergie en ma faveur pour co-diriger cette thèse. Sa rigueur et sa grande qualité scientifique m'ont été très bénéfiques. Je voudrais qu'il trouve ici l'expression de mon immense gratitude.

Je me dois d'associer à ces remerciements les différents membres de l'équipe ABC. Les stages que j'ai effectués à Nancy m'ont beaucoup apporté à la fois sur le plan scientifique et sur le plan humain. J'ai cotoyé au sein de l'équipe ABC, des personnes sympathiques qui m'ont bien accueillie, m'ont soutenue et encouragée durant mes séjours. Mes remerciements vont particulièrement à : Fabien LAUER, Khadija MUSAYEVA, Rémi BONIDAL, Fabienne THOMARAT, Emmanuel DI-DIOT, Hafida CHOUARFIA, Christelle LEVEQUE.

Mes vifs remerciements vont également à M. le Professeur Mohammed BENYETTOU, pour l'honneur qu'il m'a fait en acceptant de présider ce jury.

Je tiens également à remercier M. le Professeur Mohammed Amine CHIKH, d'avoir accepté d'examiner ce travail et de faire partie du jury.

---

Je suis très reconnaissante envers M. le Professeur Zoubir Ahmed FOUATIH, d'être rapporteur de cette thèse et de faire partie du jury.

Ma profonde reconnaissance va encore à M. le Professeur Abdelhafid BESSAID, d'avoir accepté d'examiner et juger ce modeste travail.

Je remercie très sincèrement M. le Professeur Yahia LEBBAH, de m'avoir fait l'honneur de rapporter sur cette thèse.

# Table des matières

<b>Table des figures</b>	<b>7</b>
<b>Liste des tableaux</b>	<b>9</b>
Abreviations . . . . .	11
<b>1 Introduction</b>	<b>14</b>
1.1 Présentation et motivation de l'étude . . . . .	16
1.2 Problématiques et contributions de l'étude . . . . .	18
1.3 Structure de la thèse . . . . .	21
<b>2 Introduction à l'électrophysiologie cérébrale</b>	<b>23</b>
2.1 Le cerveau humain . . . . .	23
2.1.1 Structures anatomiques de l'encéphale et leurs fonctions . . . . .	24
2.1.2 La neurophysiologie du cerveau humain . . . . .	26
2.1.3 Origine de l'électrogenèse cérébrale . . . . .	27
2.2 L'électroencéphalographie . . . . .	30
2.2.1 l'enregistrement électroencéphalographie (EEG) . . . . .	31
2.2.2 Les capteurs . . . . .	32
2.2.3 Nature et fréquences des signaux EEGs . . . . .	34
2.3 Aperçu sur le concept d'interface cerveau-machine (BCI) "machine Brain Computer Interfaces" . . . . .	35
2.3.1 Architecture des systèmes BCIs . . . . .	37

2.3.2	Application des interfaces cerveau machine . . . . .	39
<b>3</b>	<b>Introduction à l'électrophysiologie cardiaque</b>	<b>41</b>
3.1	Anatomie et fonctionnement du cœur humain . . . . .	42
3.2	L'engendrement du signal électrique cardiaque . . . . .	43
3.3	L'électrocardiographie . . . . .	45
3.3.1	L'enregistrement (ECG) . . . . .	46
3.3.2	Analyse de L'ECG . . . . .	49
3.3.3	Arythmies cardiaques étudiées . . . . .	54
<b>4</b>	<b>Machines à Vecteurs de Support Multi Classes Probabilistes</b>	<b>60</b>
4.1	Cadre théorique et notions de base . . . . .	62
4.2	Les M-SVMs basées sur les méthodes de décomposition . . . . .	64
4.3	Les M-SVMs directes . . . . .	68
4.3.1	Le modèle générique . . . . .	69
4.3.2	Caractérisation des quatre principales M-SVMs . . . . .	78
4.4	Post-traitement des sorties des M-SVMs . . . . .	80
4.4.1	Les M-SVMs indirectes . . . . .	81
4.4.2	les M-SVMs directes . . . . .	83
<b>5</b>	<b>Application 1 : Hybridation des SVMs multi-classes et les cartes de kohonen pour l'interprétation de la pensée.</b>	<b>85</b>
5.1	Description de la base de données . . . . .	86
5.2	Etat de l'art . . . . .	89
5.3	Méthode proposée pour l'extraction des paramètres des signaux EEGs	91
5.3.1	Prétraitement . . . . .	92
5.3.2	Estimation des paramètres . . . . .	93
5.4	Structure hybride proposée . . . . .	96
5.4.1	Sélection et regroupement des paramètres en utilisant une SOM . . . . .	97
5.4.2	Classification par les M-SVMs probabilistes . . . . .	99

5.5	Résultats . . . . .	101
5.5.1	Sélection des paramètres . . . . .	101
5.5.2	Évaluation des systèmes de classification implémentés . . . . .	103
5.6	Discussion . . . . .	111
<b>6</b>	<b>Application 2 : Classification des signaux ECGs pour la détection des pathologies cardiovasculaires</b>	<b>114</b>
6.1	Description de la base de données . . . . .	117
6.2	Etat de l'art . . . . .	117
6.3	Méthode d'extraction des paramètres des signaux ECGs proposée .	121
6.3.1	Localisation et extraction des battements . . . . .	122
6.3.2	Estimation des paramètres . . . . .	124
6.4	Classification des signaux ECGs par les M-SVMs probabilistes . . .	127
6.5	Résultats . . . . .	128
6.5.1	Sélection des paramètres . . . . .	128
6.5.2	Évaluation des systèmes de classification implémentés . . . . .	128
6.6	Discussion . . . . .	133
	<b>Conclusion générale</b> . . . . .	<b>137</b>
	<b>Bibliographie</b>	<b>141</b>



# Table des figures

2.1	Les régions anatomiques de l'encéphale. . . . .	25
2.2	Structure simplifiée d'un neurone. . . . .	26
2.3	Système de communication interneurone. . . . .	28
2.4	Neurones pyramidaux. Organisation en palissades [1, 2]. . . . .	29
2.5	Différences entre les neurones pyramidaux [3, 2]. . . . .	29
2.6	Emplacement des électrodes sur le scalpe selon le système international 10-20. . . . .	33
2.7	Les 5 classes de rythme cérébrale. . . . .	36
2.8	Architecture d'une interface cerveau machine. . . . .	38
3.1	Anatomie du coeur humain.[4] . . . . .	42
3.2	Potentiel cellulaire et mouvement ionique. Les cinq phases représentent : la phase 0 la dépolarisation de la cellule, la phase 1 la repolarisation initiale, la phase 2 la repolarisation lente (à phase en plateau), la phase 3 à la repolarisation rapide, et la phase 4 le potentiel de repos. . . . .	44
3.3	Triangle d'Einthoven. . . . .	48
3.4	Dérivations précordiales des membres. . . . .	49
3.5	Processus de propagation du potentiel d'action dans le coeur et tracé ECG généré. L'onde P, représente la contraction des deux oreillettes. Le complexe QRS, représente la contraction des ventricules. L'onde T, représente la repolarisation des ventricules. . . . .	51
3.6	Ondes, intervalles et segments d'un signal ECG . . . . .	52

---

3.7	Extrasystole ventriculaire. . . . .	55
3.8	Extrasystole auriculaire. . . . .	56
3.9	Bloc de branche gauche. . . . .	57
3.10	Bloc de branche droit. . . . .	58
5.1	Diagramme général de la méthode proposée pour un enregistrement EEG. . . . .	87
5.2	Resultat du filtre passe bande et de la segmentation. . . . .	92
5.3	Les coefficients des détails (D2, D3, D4 and D5) et de la cinquième approximation A5 pour un segment donné. . . . .	95
5.4	Structure hybride proposée. . . . .	96
5.5	Résultat du clustering et la sélection des variables en utilisant la carte SOM. . . . .	100
6.1	Diagramme général de la méthode proposée pour l'identification de la classe des battements. . . . .	116
6.2	Localisation des pics R des complexes QRS du signal EEG 100. . . . .	123
6.3	Battement échantillonné sur 360 Hz centré sur le pic R (ECG 100) . . . . .	124
6.4	Les coefficients des 5 niveaux de décomposition et la dernière approximation de la DWT pour un battement normal (ECG 100). . . . .	126

# Liste des tableaux

4.1	les hyperparamètres des quatre M-SVMs . . . . .	80
5.1	Résultats de la classification de la structure hybride proposée (%) (sujet1). . .	104
5.2	Résultats de la classification de la structure hybride proposée (%) (sujet2). . .	104
5.3	Résultats de la classification de la structure hybride proposée (%) (sujet3). . .	105
5.4	Résultats de la classification de la structure hybride proposée (%) (sujet5). . .	106
5.5	Résultats de la classification de la structure hybride proposée (%) (sujet6). . .	107
5.6	Résultats de la classification de la structure hybride proposée (%) (sujet7). . .	107
5.7	Resultats relatifs au : taux de la classe rejet + taux globale de classification. . .	108
5.8	Temps d'apprentissage (en sec.) des M-SVMs directes probabilistes. . . . .	109
5.9	Temps de prédiction (en sec.) des M-SVMs directes probabilistes. . . . .	110
5.10	Les résultats des travaux précédants (%). . . . .	112
6.1	Patients et nombre de battements utilisés dans cette étude. . . . .	118
6.2	Les gammes de fréquences des sous-bandes (D3, D4, D5, et A5) relatives à un signal ECG échantillonné à 360 hz. . . . .	125
6.3	Matrice de confusion 1-c-1. . . . .	129
6.4	Matrice de confusion 1-c-tous. . . . .	130
6.5	Matrice de confusion WW. . . . .	130
6.6	Matrice de confusion CS. . . . .	130
6.7	Matrice de confusion LLW. . . . .	131
6.8	Matrice de confusion M-SVM <sup>2</sup> . . . . .	131

6.9	Resultat de classification des M-SVMs (%).	131
6.10	Temps d'entrainement et de prédiction (en sec.) des M-SVMs directes probabi- listes.	133
6.11	Les résultats des travaux précédents (%)	135

# Abbreviations

ALG :	Logistic Regression Algorithms	Algorithmes de Regression Logistique.
ANN :	Artificial Neural Network.	Réseaux de Neurones Artificiels.
AP :	Action Potential.	Potentiel d'Action.
APB :	Atrial Premature Contraction	Extrasystole Auriculaire.
AR :	AutoRegressive.	AutoRégressif.
ASR :	ASymmetry Ratio.	Ratio d'Asymétrie.
BCI :	Brain Computer Interface.	Interface Cerveau-Machine.
BQ :	Quadratic Bayesian Classifier	Classifieur Bayésien Quadratique.
ComT :	Counting Tasks.	Tâche de Comptage.
DWT :	Discrete Wavelet Transform.	Transformée par Ondelettes Discrètes.
ECG :	ElectroCardioGram.	Signal ElectroCardioGraphique (ElectroCardioGramme).
ECOC :	Error-Correcting Output Coding.	Codes Correcteurs d'Erreurs.
EROS :	Event-Related Optical Signal.	Signal Optique Lié à l'Evenement.
EEG :	ElectroEncephaloGram.	Signal ElectroEncephaloGraphique (ElectroEncephaloGramme).
ELM :	Extreme Learning Machine.	Machine à Apprentissage Extrême.
EMD :	Empirical Mode Decomposition.	Décomposition Modale Empirique.
ENN :	Elman Neural Network.	Réseaux de Neurones d'Elman.

DFT :	Discrete Fourier Transform.	Transformation de Fourier Discrète.
FFT :	Fast Fourier Transform.	Transformation de Fourier Rapide.
FIR :	Finite Impulse Response Filter.	Réponse Impulsionnelle Finie.
FN :	False Negative.	Faux Négatifs.
fMRI :	Imaging Functional Magnetic Resonance.	Imagerie par Résonance Magnétique Fonctionnelle.
GA :	Genetic Algorithm.	Algorithme Génétique.
CS :	Crammer et Singer.	Crammer et Singer.
IA :	Artificial Intelligence.	Intelligence Artificielle.
ICA :	Independent Component Analysis.	Analyse en Composantes Indépendantes.
IRM :	Magnetic Resonance Imaging.	Imagerie par Résonance Magnétique.
kNN :	k-Nearest Neighbors.	k plus Proche Voisins.
LBBB :	Left Bundle Branch Bloc.	Bloc de Branche Gauche.
LetT :	Letter Task.	Tâche Lettre.
LDA :	Linear Discriminant Analysis.	Analyse Discriminante Linéaire.
LLW :	Lee, Lin et Wahba.	Lee, Lin et Wahba.
LS-SVM :	Lost Square Support Vector Machine.	Machine à Vecteurs de Supports à Moins Carré.
MEG :	Magneto EncephaloGraphy.	Magneto EncéphaloGraphie.
MI :	Motor Imaging.	Imagerie Motrice.
MLP :	Multi Layer Perceptro.	Perceptron Multi Couche
M-SVM :	Multi-class Support Vector Machine.	Machine à Vecteurs de Supports Multiclasses.
M-SVM <sup>2</sup> :	Quadratic M-SVM.	M-SVM à Coût Quadratique.
MSR :	Mean of Square Root.	Moyenne de la Racine Carrée
MulT :	Multiplication Task.	Tâche de Multiplication.
NIRS :	Near-Infrared SpectroScopy.	Spéctroscopie Proche InfraRouge.
EOG :	ElectroOculoGraphy.	ElectroOculoGraphie.

PB :	Paced Beat.	Pacemaker.
PCA :	Principal Component Analysis.	Analyse en Composantes Principales.
PET :	Positron Emission Tomography.	Tomoscintigraphie par Emission de Positons.
PL :	Linear Prediction.	Prédiction Linéaire.
PS :	Power Spectrum.	Spectre de Puissance.
PSP :	PostSynaptic Potentials.	Potentiels PostSynaptique.
PVC :	Premature Ventricular Contraction.	Extrasystole Ventriculaire.
QDA :	Quadratic Discriminant Analysis.	Analyse Discriminante Quadratic.
RBBB :	Right Bundle Branch Block.	Bloc de Branche Droit.
RelT :	Baseline Task.	Tâche de Relaxation.
RKHS :	Reproducing kernel Hilbert Space.	Espace de Hilbert à Noyau Reproduisant.
RN :	Neural Network.	Réseau de Neurone.
RotT :	Rotation Task.	Tâche Rotation.
SOM :	Self Organizing Map of Kohonen.	Carte Auto-Organisatrice de Kohonen.
SD :	Standard Derivation.	Dérivation Standard.
ST :	Stockwell Transform.	Transformée Stockwell.
STFT :	Short-Time Fourier Transform.	Transformée de Fourier à Court Terme.
SVM :	Support Vector Machine.	Machine à Vecteurs de Supports.
WL :	Wilks Lambda.	Wilks Lambda.
WW :	Weston et Watkins.	Weston et Watkins.
1-c-1 :	One-Again-One.	Un-Contre-Un.
1-c-tous :	One-Again-All.	Un-Contre-Tous.

# Chapitre 1

## Introduction

Les organismes vivants sont constitués d'un ensemble de systèmes. Le corps humain, par exemple, comprend le système musculaire, l'appareil cardiovasculaire, l'appareil respiratoire, le système nerveux, etc. Les processus physiologiques sont des mécanismes qui permettent à l'ensemble de ces systèmes de collaborer entre eux pour permettre à un organisme de survivre, grandir, et se développer. La plupart des processus physiologiques se manifestent sous forme de signaux. Ces signaux peuvent être de plusieurs types : biochimiques sous la forme d'hormones et neuro-médiateurs, électriques sous la forme du potentiel ou de courant, et physiques sous la forme de pression ou de température. Dans cette étude, nous nous concentrons sur l'étude des signaux ElectroCardioGraphiques (ECGs) et ElectroEncephaloGraphiques (EEGs) en utilisant les méthodes bio-inspirées, dans le cadre de deux applications très répandues. La première application, concerne l'exploitation des signaux ECGs pour la détection des anomalies cardiovasculaires, tandis que la seconde application, concerne la discrimination entre différentes tâches mentales dans les Interfaces Cerveau-Machine (BCIs) par le biais des signaux EEGs.

Depuis quelques années, les méthodes bio-inspirées ont déjà contribué à des applications dans des domaines variés de l'Intelligence Artificielle (IA) tels que : la médecine, la robotique, l'imagerie, etc.. L'analyse automatique des signaux bio-



médicaux par le biais des méthodes bio-inspirées, doit permettre d'accéder à des résultats quantitatifs objectifs et précis, qui contribueront à réduire la subjectivité des médecins et biologistes. En plus, l'analyse des signaux biomédicaux doit donner accès à l'extraction d'observations ignorées et/ou cachées aux cliniciens et qui possèdent une signification importante pour la prise de décision.

Par "Méthodes bio-inspirées", nous entendons tous systèmes ou architectures qui s'inspirent des inventions et astuces mises en place dans la nature par les êtres vivants, et qui peuvent être utilisés pour exploiter des données et créer des modèles explicatifs, prédictifs ou de classification. Il s'agit donc, entre autres, des techniques inspirées de la structure cellulaire du système nerveux, à savoir, les réseaux de neurones. Il s'agit aussi d'approches composées de multiples éléments (systèmes multi-agents) telle une fourmière (algorithmes de fourmis). Il existe également les algorithmes immuno-inspirés ou les algorithmes évolutionnistes, dont le but est de mimer le processus de sélection naturelle. Ces algorithmes s'inspirent du fait que l'évolution des êtres vivants, tend à générer des organismes plus adéquats à leur environnement.

Par ailleurs, les méthodes bio-inspirées incluent particulièrement les techniques dotées de capacité d'apprentissage et qui évoluent dans le temps : les réseaux de neurones, les Machines à Vecteurs de Supports (SVMs). En fait, l'approche traditionnelle de l'IA utilisée était basée principalement sur la programmation, toute l'intelligence dans un système provient de son concepteur. Cette approche n'est pas nécessairement la meilleure, du fait que son principe étant d'envisager l'ensemble des cas possibles pendant la programmation. Cependant, la réponse d'un système pré-programmé face à une situation nouvelle est incertaine. Il a donc fallu trouver des solutions alternatives à l'approche mathématique pour permettre aux machines d'être plus autonomes. Par conséquent, l'apprentissage automatique (artificiel) est apparu comme réponse aux faiblesses de l'approche traditionnelle de l'IA. Ainsi, l'apprentissage artificiel peut être défini comme la capacité d'un système à améliorer ses performances via un processus d'apprentissage, et par la suite, remplir

des tâches difficiles ou impossibles à remplir par des moyens algorithmiques plus classiques.

## 1.1 Présentation et motivation de l'étude

Le coeur et le cerveau représentent à la fois les organes les plus importants et les plus critiques du corps humain.

Le cerveau est un organe indispensable car il représente le centre de contrôle de tous les autres organes de l'organisme. Il est responsable de la perception, la cognition, l'attention, l'émotion, la mémoire et l'action. Quand une personne pense, lit, ou regarde la télévision, différentes parties du cerveau sont stimulées et des signaux électriques sont générés. Ces signaux électriques peuvent être explorés à l'aide de plusieurs techniques invasives et non invasives, on peut citer : l'ElectroEncephaloGramme (EEG), Imagerie par Resonance Magnétique fonctionnelle (fMRI), Magneto EncéphaloGraphie (MEG), Spéctroscopie Proche InfraRouge (NIRS), Tomoscintigraphie par Emission de Positons (PET), Signal Optique Lié à l'Evenement (EROS) [5].

Parmi ces techniques, l'EEG est la plus utilisée en raison de sa capacité à enregistrer les signaux cérébraux de façon non invasive, et à son faible coût. En effet, L'EEG est exploité par des spécialistes pour suivre en temps réel la chronologie des opérations mentales, et d'étudier la dynamique des phénomènes cérébraux, par exemple dans des cas des atteintes mentales comme l'épilepsie. L'EEG trouve aussi application dans les BCIs, il fournit un moyen efficace pour aider les personnes touchées par un handicap moteur sévère à communiquer avec le monde extérieur juste à travers les signaux émis par le cerveau (sans avoir à parler ou à effectuer des activités musculaires). Un système BCI mesure en temps réel (directement) l'activité cérébrale associée à l'intention de l'utilisateur, puis la traduit en un signal de commande correspondant [6, 5].

Le coeur est aussi un organe essentiel pour la simple raison que le corps ne peut pas fonctionner sans sa présence. Son rôle est de faire circuler le sang dans

l'organisme en agissant comme une pompe par des contractions rythmiques. Le sang véhicule, les nutriments, l'oxygène, et bien d'autres molécules dans tout l'organisme. Ce qui permettra d'alimenter les cellules et assurer le bon fonctionnement du cerveau, des muscles, ou de tout autre organe. Ainsi, même le coeur est vascularisé. Le coeur peut être affecté par plusieurs pathologies, qui constituent la première causes de mortalité dans le monde. Plusieurs techniques ont été développées pour la détection précoce des anomalies cardiaques, comme : l'ElectroCardioGramme (ECG), la radiographie thoracique, l'échocardiographie, la tomographie ("scanner"), l'Imagerie par Résonance Magnétique (IRM), le cathétérisme cardiaque, les examens du sang, et dans quelques cas rares, la biopsie du muscle cardiaque.

Parmi l'ensemble de ces techniques, l'ECG est l'examen le plus utilisé. Le diagnostic des maladies cardiovasculaires à partir de l'ECG, inclut deux types de procédures : la procédure invasive et la procédure non-invasive. Dans la première on parle d'enregistrements endocavitaires. Ces types d'enregistrements nécessitent une intervention chirurgicale, et consistent à placer des électrodes directement à l'intérieur du coeur. Cette procédure d'enregistrement peut s'avérer, très efficace et précise pour la détection des lésions du tissu cardiaque, mais elle est cependant très contraignante pour les patients. La procédure non-invasive, consiste à recueillir l'activité cardiaque au moyen d'électrodes posées sur la peau (au dessus des poignets et des chevilles, sur le thorax ou dans le dos dans certains cas). L'ECG non-invasif est l'examen le plus courant, car il est rapide à mettre en place, peu coûteux et surtout non traumatique donc très peu contraignant pour le patient (Note : dans ce qui suit "ECG" fera référence à l'ECG non-invasif). L'examen ECG permet d'avoir un aperçu sur les différentes caractéristiques du coeur. De ce fait on peut déterminer s'il est affecté ou pas par une pathologie. Néanmoins, dans certains cas, les signes électrocardiographiques de la maladie ne sont pas constants à chaque examen, mais se manifestent à certains intervalles irréguliers au cours de la journée. Par conséquent, pour un diagnostic correct, l'étude d'un signal ECG

doit être effectuée sur plusieurs heures. Ceci nécessite un traitement automatique du signal ECG en temps réel, afin d’assister les cardiologues dans le traitement des enregistrements de longue durée.

La mise en oeuvre d’un système BCI typique, et l’implémentation d’un système intelligent (temps réel) pour la détection des anomalies cardiovasculaires, nécessitent un bon système de traitement de signal. Les performances de ce dernier sont principalement liées à la technique utilisée pour sélectionner les meilleures caractéristiques des signaux, mais aussi à la méthode de classification mise en place pour interpréter les composants obtenus. C’est dans ce contexte général que s’insère l’essentiel du travail présenté dans cette thèse.

## 1.2 Problématiques et contributions de l’étude

En général, les caractéristiques sélectionnées pour une application donnée, doivent être très significatives afin d’avoir une bonne discrimination entre les différentes catégories. Dans ce contexte, de nombreuses approches intéressantes ont été proposées pour l’extraction des paramètres à partir des signaux EEGs et des signaux ECGs. Parmi elles, la Transformée par Ondelettes Discrètes (DWT) qui a été utilisée avec succès dans l’extraction de caractéristiques des signaux EEGs [7, 8, 9] et des signaux ECGs [10], cette méthode fournit une approche mathématique rigoureuse qui lui donne la capacité de représenter les signaux non stationnaires dans le temps et dans le domaine fréquentiel. Par conséquent, la DWT a été utilisée dans cette étude. Cependant, Les signaux EEGs sont obtenus à partir de plusieurs canaux, et les vecteurs caractéristiques sont construits par concaténation des descriptions qui ont été obtenues à partir des multiples canaux utilisés. Ce qui se traduit par un grand vecteur de caractéristiques, tandis que le nombre d’échantillons de l’ensemble de la formation est limité (résultant le fléau de la dimension : curse of dimensionality). Ainsi, une méthode de sélection de variables est nécessaire. En effet, la sélection des paramètres pertinents permet de réduire le volume d’informations qui contribuera à réduire la complexité des algorithmes

de classification et également améliorer leur performance de généralisation.

La première contribution apportée dans cette étude consiste à proposer une nouvelle alternative de sélection de variables en utilisant une carte auto organisatrice de Kohonen (SOM). Ce choix repose principalement sur le fait que l'algorithme SOM a la particularité de regrouper des ensembles de grandes dimensions d'une manière non supervisée. Le principe de l'algorithme SOM est inspiré de la biologie plus précisément de la structure du cerveau humain. En effet, le cerveau est divisé en plusieurs parties et chaque une d'elles est responsable de tâches bien précises. L'algorithme SOM fait ainsi, évoluer un ensemble de solutions (paramètres) à un problème donné, dans l'optique de trouver les meilleurs regroupements des descripteurs les plus corrélés, et représenter par la suite chaque groupement par un seul paramètre (un seul neurone).

En ce qui concerne la classification, de nombreux modèles ont été implémentés pour la classification des signaux EEGs et des signaux ECGs. Parmi les classificateurs proposés, ceux qui ont donné les meilleurs résultats et les plus utilisés sont les SVMs [11, 12, 13]. Ils fournissent, une bonne robustesse vis à vis de la dimensionnalité, de bonnes performances en généralisation, et une insensibilité au sur-apprentissage. Les SVMs ont été initialement conçus pour des problèmes binaires et leurs extensions aux problèmes multiclassés (M-SVMs) est toujours un champ de recherche actif. Nous pouvons trouver deux approches, l'approche indirecte qui implique seulement des SVMs bi-classes et l'approche directe qui tient compte de toutes les classes simultanément. La première traite le problème partiellement et couvre trois méthodes principales : un-contre-un (1-c-1), un-contre-tous (1-c-tous) et Codes Correcteurs d'Erreurs (ECOC). A l'opposé, la seconde approche traite le problème d'apprentissage des SVMs comme étant un problème d'optimisation globale. On trouve quatre grands modèles des M-SVMs directes sur lesquels la communauté scientifique est actuellement concentrée : le modèle de Weston et Watkins (WW) [14], le modèle de Crammer et Singer (CS) [15], le modèle de Lee et al. (LLW) [16], et la M-SVM à coût quadratique (M-SVM<sup>2</sup>) [17], modèle in-

troduit par Guermeur et Monfrini. Cependant, tous les systèmes proposés pour la classification des signaux EEGs et des signaux ECGs, sont basés sur des M-SVMs impliquant des méthodes de décomposition, et à notre connaissance personne n'a introduit l'utilisation des M-SVMs directes.

Comme deuxième contribution de cette étude, nous avons décidé d'étudier l'apport des quatre modèles des M-SVMs (WW, GS, LLW, et M-SVM<sup>2</sup>) dans la problématique de la classification des signaux EEGs et des signaux ECGs. Aussi, une implementation des M-SVMs indirectes basées sur l'approche 1-c-1 et l'approche 1-c-tous a été effectuée à titre comparative avec les quatres methodes des M-SVMs directes.

D'un autre coté, généralement en IA, il est connu que dans la majorité des applications de reconnaissance des formes, la sortie d'un classifieur doit être munie d'un certain degré de fiabilité. Pour cette raison, un bon classifieur est un discriminateur qui est capable d'exprimer ses sorties en terme de probabilités à posteriori. Toutefois, le principe standard des SVMs ne leur permet pas d'estimer une telle mesure. A fin de palier à ce problème, Platt [18] proposa une méthode de post-traitement des sorties d'une unique SVM en probabilité a posteriori. Cette méthode a été généralisée un peu plus tard par d'autres auteurs [19, 20, 21] pour des problèmes multiclassés.

Comme troisième contribution de ce travail, nous avons post-traité les sorties des M-SVMs afin de générer des estimations de probabilités à posteriori des classes. Ensuite, nous avons fixé un seuil de confiance comme suit : si la probabilité associée à la sortie gagnante du classifieur est supérieure à 0.75, la décision est retenue. Sinon, la décision est rejetée. Il est plus instructif de prendre une décision si nous savons qu'il y a 90 % de chance qu'une description appartienne à une classe donnée, et il est préférable de s'abstenir de faire une fausse prédiction. A noter, que nous avons aussi étudié la possibilité d'implémenter les M-SVMs directes dans des systèmes temps réel. Nous avons donc étudié le temps de calcul des différentes approches directes, à savoir, le temps d'entraînement et le temps de prédiction.

### 1.3 Structure de la thèse

Cette thèse se compose de six chapitres et chaque chapitre fournit des informations importantes sur la thématique de recherche étudiée. Le reste est structuré comme suit :

- Le chapitre 2 expose les informations relatives au cerveau humain, les principes fondamentaux des signaux EEGs, et le concept des systèmes BCIs.
- Le chapitre 3 fournit des descriptions sur le coeur, les principes fondamentaux des signaux ECGs, et les caractéristiques des arythmies cardiaques considérées dans cette étude.
- Le chapitre 4 présente en premier lieu quelques notions et définitions qui sont les fondements de base des M-SVMs. Nous considérons ensuite la première approche mise en oeuvre pour traiter des problèmes multi classes au moyen des méthodes de décomposition. Puis, nous exposons les principaux modèles de l'approche directe, en mettant en évidence leurs spécificités par le biais du modèle générique. En fin, nous décrirons les démarches suivies pour post-traiter en terme de probabilités, les différentes sortie des M-SVMs directes et indirectes
- Le chapitre 5 expose une nouvelle structure hybride basée sur deux étages, avec apprentissage supervisé et non supervisé. Cette approche est validée sur la base de données EEG de Keirn et Aunon [22], pour distinguer entre cinq tâches mentales différentes : Relaxation (RelT), Multiplication (MulT), Lettre (LetT), Rotation (RotT), et Comptage (CompT). Le premier étage utilise une carte SOM, pour le regroupement et la selection des variables, tandis que le deuxième étage utilise les M-SVMs probabilistes (WW, LLW, CS, M-SVM<sup>2</sup>, 1-c-1, et 1-c-tous) pour la classification finale. Les performances des M-SVMs directes et des M-SVMs indirectes ont été par la suite discutées.
- Le chapitre 6 propose un système basé sur la DWT et les M-SVMs probabilistes directes et indirectes pour la classification des battements normaux

(N) et cinq arythmies cardiaques : Bloc de Branche Droit (RBBB), et Bloc de Branche Gauche (LBBB), Extrasystole Ventriculaire (PVC), Extrasystole Auriculaire (APB), et la classe des enregistrements spécifiques au port des pacemaker (PB). Cette approche est validée sur la base de données MIT-BIH [23]. Le but principale de ce chapitre est de comparer les performances des deux approches des M-SVMs et déduire l'apport des M-SVMs directes dans le traitement des signaux ECGs.



# Chapitre 2

## Introduction à l'électrophysiologie cérébrale

Ce chapitre présente les connaissances de base relatives aux signaux ElectroEncephaloGrphiques (EEG) afin d'introduire certaines terminologies et des informations connexes à cette recherche. La section 2.1 présente brièvement les caractéristiques anatomiques et neurophysiologiques du cerveau humain. Cette section se concentre particulièrement sur les fonctions des principales parties du cerveau et sur la création des courants électrochimiques captés par des électrodes de cuir chevelu formant le signal EEG. Secondairement, dans la section 2.2, une étude sur la nature et les caractéristiques du signal EEG est réalisée. Finalement, la section 2.3 fournit les informations nécessaires relatives aux principes des interfaces cerveau machine (BCI) notamment leurs structures et leurs applications dans la vie réelle.

### 2.1 Le cerveau humain

Le commandant suprême du corps humain est le cerveau. Il constitue la partie centrale du système nerveux et régit les fonctions des différents organes dans le corps. En premier lieu, nous expliquons dans cette section les structures ana-

tomiques du cerveau et leurs fonctions. Ensuite, nous nous concentrons sur la physiologie électrique du cerveau. Cela va nous permettre de comprendre : comment, pourquoi et où le cerveau génère des signaux électriques qui peuvent être capturés et transformés en signal EEG.

### 2.1.1 Structures anatomiques de l'encéphale et leurs fonctions

Anatomiquement le cerveau fait partie de l'encéphale qui peut être divisé en trois parties majeures (voir figure 2.1) : le cerveau, le cervelet et le tronc cérébral [24] :

**Le cerveau :** est la partie la plus grande et la plus importante de l'encéphale humain, il est situé au-dessus du cervelet et du tronc cérébral et est généralement associé à des fonctions cérébrales tel que les pensées, les émotions et les mouvements. La couche la plus externe du cerveau : le cortex cérébral est un tissu neuronal dense plissé par des circonvolutions cérébrales. Le cerveau est composé de deux hémisphères, l'hémisphères droit et gauche. Chaque hémisphère peut être divisé en quatre lobes : frontal, pariétal, occipital et temporal [25]. Ces lobes sont responsables d'une variété de fonctions corporelles :

- Le lobe frontal est associé avec la personnalité, la génération des émotions, la résolution de problèmes, le développement moteur, le raisonnement et la planification.
- Le lobe pariétal est responsable de la sensation (par exemple la douleur, le toucher), la compréhension sensorielle, la reconnaissance, la perception des stimuli, l'orientation et le mouvement.
- Lobe occipital est responsable du processus visuel.
- Lobe temporal est impliqué dans le traitement de la parole et la mémoire.

**Le cervelet :** est situé dans le bas postérieur de la tête et est également divisé en deux hémisphères. Il contient plus de la moitié de la totalité des neurones de l'encéphale et en représente la deuxième plus grande structure. Le cervelet est l'une des zones sensibles du cerveau qui est responsable de la coordination et du contrôle des mouvements volontaires, la perception sensorielle, la posture et la régulation de l'équilibre.

**Le tronc cérébral :** est situé dans la partie la plus basse de la tête et relie l'encéphale à la moelle épinière. Le tronc cérébral contrôle les fonctions vitales de l'organisme, y compris, la respiration, la conscience, les mouvements des yeux et de la bouche, le rythme cardiaque, la pression artérielle, et relais des messages sensoriels (douleur, chaleur, bruit, faim..).

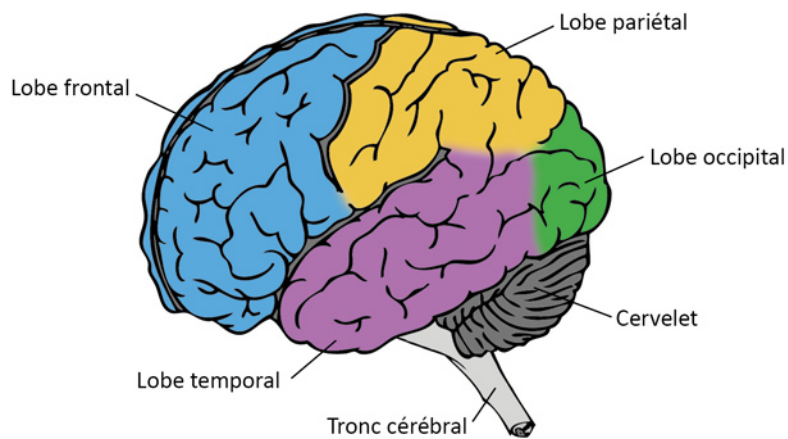


FIGURE 2.1 – Les régions anatomiques de l'encéphale.

## 2.1.2 La neurophysiologie du cerveau humain

### 2.1.2.1 Le neurone

Le cerveau humain est composé d'environ 100 milliards de cellules nerveuses appelées neurones. Les neurones partagent les mêmes caractéristiques et contiennent les mêmes organites que les autres cellules, mais la spécificité électrochimique leur permet de transmettre des signaux électriques et de transférer des messages les uns aux autres sur une longue distance. Un neurone est organisé en trois parties principales : le corps cellulaire (soma), l'axone et les dendrites [26, 25], comme schématisé sur la figure 2.2. Le noyau cellulaire est le coeur de la cellule, il lui

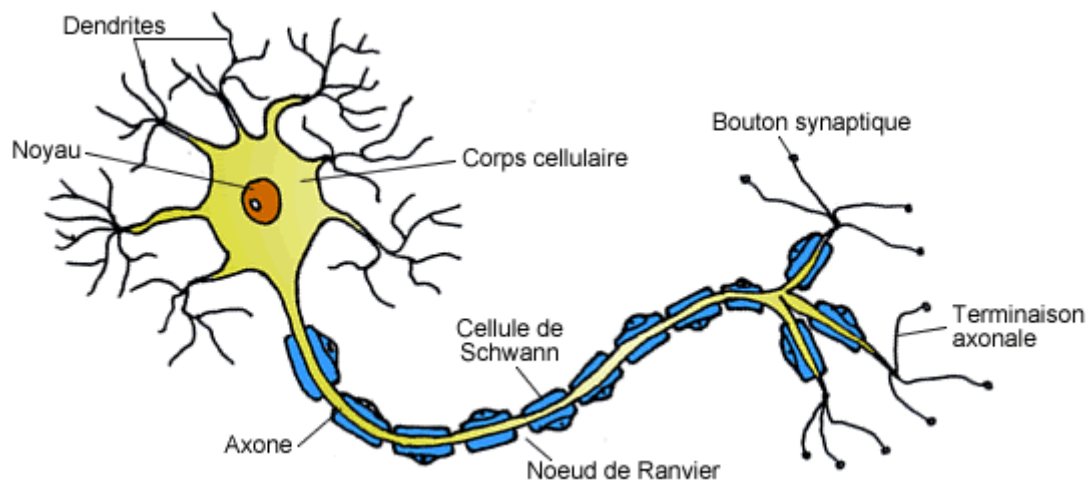


FIGURE 2.2 – Structure simplifiée d'un neurone.

donne des instructions et régis son fonctionnement. L'axone est une longue portion du neurone qui relie le noyau de son propre neurone à la dendrite d'un autre. La dendrite est une courte section du neurone avec de nombreux récepteurs pour des neurotransmetteurs qui peuvent être envoyés par des terminaison axonal d'un autre neurone.

### 2.1.2.2 L'activité électrique neuronale

À travers le lien axone-dendrite, les neurones peuvent communiquer entre eux. Cette communication est rendue possible grâce à deux potentiels : le Potentiel d'Action (AP) et le Potentiels PostSynaptique (PSP).

L'AP est un événement électrique généré par des canaux à ions voltages dépendants situés le long du neurone. La modification rapide de la composition en ions de part et d'autre de la membrane cytoplasmique du neurone par les canaux ioniques, provoque une vague de dépolarisations qui travers le long l'axone en direction des dendrites du neurone suivant c'est le signal électrique neuronal [27]. Une fois que les terminaisons axonales reçoivent le signal électrique, elles libèrent dans la synapse (espace qui relie les extrémités axonales d'un neurone et les dendrites d'un autre neurone) un neurotransmetteur, ce dernier va se lier à des récepteurs spécifiques fixés aux dendrites du neurone post synaptique.

Quant aux PSP, ce sont les potentiels résultants de la somme des potentiels excitateurs et inhibiteurs pré-synaptiques présents dans le noyau (soma) cellulaire. Une fois le PSP atteint le niveau seuil de conduction pour le neurone post-synaptique, un nouveau AP est déclenché [27, 28, 25]. Le système de communication interneurone est décrit dans la figure 2.3.

### 2.1.3 Origine de l'électrogenèse cérébrale

Lorsque les neurones sont activés au moyen d'un gradient de concentration électrochimique, des flux de courant locales sont produits (AP, PSP).

Les diverses manifestations électriques globales enregistrées sur le scalp par des électrodes, proviennent des activités électriques des cellules nerveuses sous-jacentes. Il y a une variété de neurones : bipolaires, multipolaires, unipolaires et pyramidaux. Cependant, les générateurs principaux sont les neurones pyramidaux du cortex. Ces cellules, sont, de part leurs activités et celles des synapses, la source de courants extracellulaires oscillant en permanence entre leur soma et leurs dendrites, sièges des PSP. Les neurones pyramidaux sont reconnaissable grâce à la forme

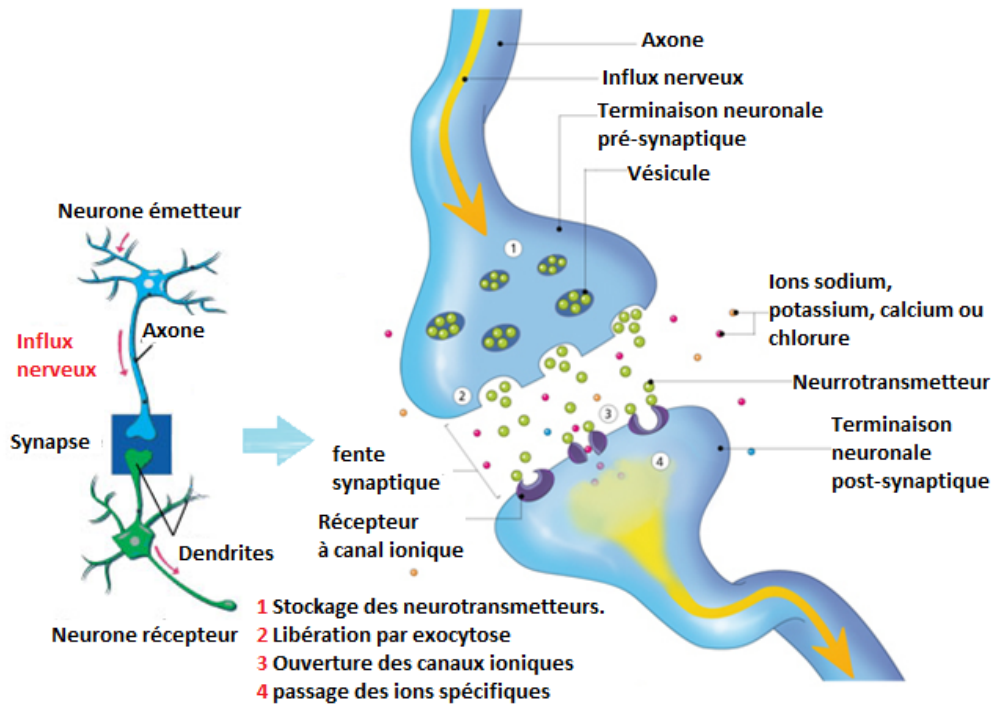


FIGURE 2.3 – Système de communication interneurone.

pyramidale de leurs soma et grâce à leurs différents arbres dendritiques apicaux et basaux (voir figure 2.4 et figure 2.5).

Au niveau du néocortex cérébral, les neurones pyramidaux sont organisés en "parallèles" (figure 2.4), perpendiculaire à la surface corticale et parallèle les uns par rapport aux autres [29]. Il excite plusieurs types de cellules pyramidales différentes les unes des autres [30] (voir figure 2.5), leur forme diffère selon la couche cérébrale dans laquelle elles sont situées (sachant que les cellules pyramidales des couches III et V sont les principaux générateurs de l'EEG).

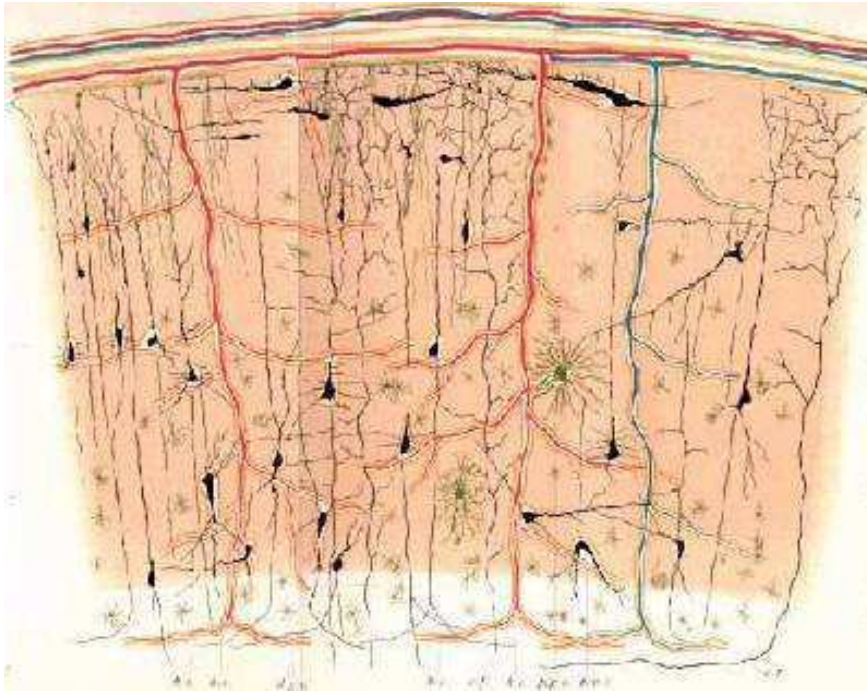


FIGURE 2.4 – Neurones pyramidaux. Organisation en palissades [1, 2].

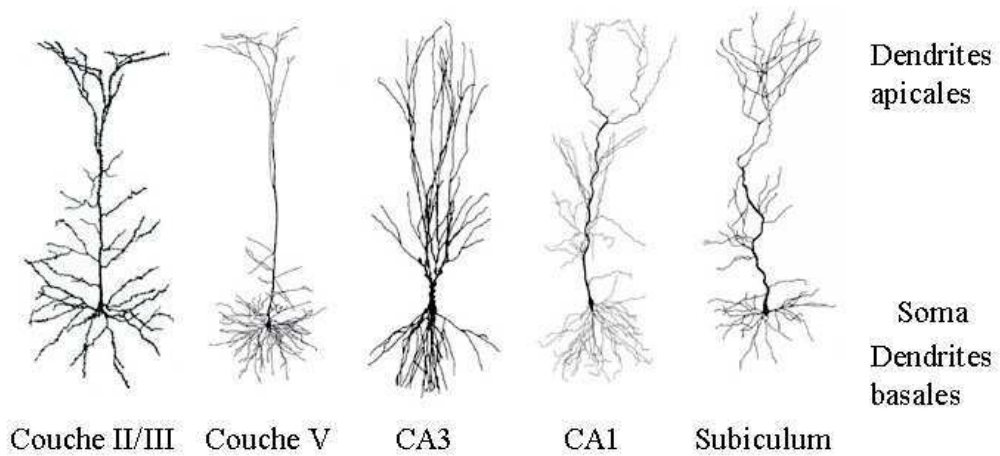


FIGURE 2.5 – Différences entre les neurones pyramidaux [3, 2].

L'activité électrique des neurones n'ait recueillie en surface que si des milliers de cellules sont actives en même temps. Cela s'explique par l'existence de phénomènes de synchronisation de certaines populations de neurones présentant une similitude architectonique et fonctionnelle. Ainsi, la sommation des potentiels électriques générés au niveau du cortex se prolonge vers la cuire chevelue ou elle pourra être captée par des électrodes.

## 2.2 L'électroencéphalographie

L'électroencéphalographie est l'enregistrement de l'activité électrique du cerveau.

- D'après [31], l'histoire de l'électroencéphalographie fut commencée en 1875, quand Richard, un jeune assistant de physiologie de la Royal Infirmary de Liverpool détecta en ouvrant la boîte crânienne d'un singe, la présence de courants électriques attestée par des oscillations du galvanomètre. Aussi, le biologiste observe, une variation significative du courant électrique localisée au niveau de la région occipitale du cortex dé que la rétine est soumise à un stimulus lumineux.
- Le premier EEG a été inventé par Hans Berger en 1929 [32]. Berger, un neuropsychiatre de l'Université de Jena en Allemagne, a utilisé le terme allemand "elektrenkephalogramm" pour décrire le graphique représentant les courants électriques générés dans le cerveau. Il suggéra que ses courants électriques changeaient en fonction de l'état fonctionnel du cerveau telles que le sommeil, l'anesthésie et l'épilepsie. Ce fut l'idée révolutionnaire qui a aidé à créer une nouvelle branche de la science médicale appelée neurophysiologie.
- En 1939, Grass permit de visualiser les activités électriques sur papier. Les bases des principes de l'examen furent posées dès 1945 et sont toujours appliquées aujourd'hui.



### 2.2.1 l'enregistrement électroencéphalographie (EEG)

L'EEG représente la transcription sous forme d'un tracé des variations dans le temps des potentiels électriques recueillis sur la boîte crânienne en différents points du scalp. C'est un test facilement accessible et non traumatique qui fournit des informations sur l'activité du cerveau. L'EEG est largement exploité par les médecins et scientifiques dans l'exploration du fonctionnement cérébral ainsi que dans le diagnostic des troubles neurologiques tels que l'épilepsie, les tumeurs cérébrales, les traumatismes crâniens, les troubles du sommeil, la démence, les troubles du comportement, le retard de langage ect. On trouve aussi des applications dans la surveillance de l'anesthésie au cours d'une intervention chirurgicale [33, 34].

Au cours de la réalisation du EEG, un certain nombre de petits disques appelés électrodes sont placées à différents endroits sur la surface du cuir chevelu avec des colles temporaires. Ensuite, chaque électrode est reliée à un amplificateur (un amplificateur par paire d'électrodes) et à une machine d'enregistrement EEG. Enfin, les signaux électriques provenant du cerveau sont convertis en lignes ondulées sur papier ou sur un écran d'ordinateur formant l'EEG. En fonction de leur utilisation, les EEGs peuvent utiliser 1 ou jusqu'à 256 électrodes dans le cas des enregistrements EEGs multicanaux. Une paire d'électrodes représente habituellement un canal. Chaque canal produit un signal lors d'un enregistrement EEG.

Il existe plusieurs types d'EEGs, selon l'endroit où les électrodes qui captent le signal électrique sont placées : on parle d'électroencéphalogramme classique lorsque le potentiel électrique est recueilli sur la boîte crânienne en différents points du scalp. On réserve le nom d'électrocorticogramme au tracé obtenu lorsque des électrodes sous durales sont placées directement sur le cortex et l'on parle de stéréoelectroencéphalographie lorsque les électrodes sont implantées directement dans les structures cérébrales (lors d'interventions neurochirurgicales). L'amplitude d'un signal EEG est généralement comprise entre 1 à 100 mV chez un adulte normal, et entre 10 à 20 mV lorsqu'il est mesurées avec des électrodes sous-durales. Du fait de la non uniformité architecturale du cerveau et de l'organisation fonctionnelle du

cortex cérébral, le profil EEG varie en fonction de l'emplacement des électrodes.

### 2.2.2 Les capteurs

Les électrodes doivent être placées d'après des règles strictes. Depuis 1957, le système le plus couramment utilisé est le système international 10-20 [35]. La ligne de départ de ce système est celle qui réunit le nasion et l'inion en passant par le vertex. Cette ligne est divisée en 6 parties : 10% de la longueur sont portés au-dessus du nasion pour former le plan frontal et 10% au-dessus de l'inion pour le plan occipital. Le reste est divisé en 4 parties égales représentant chacune 20% de la longueur totale.

L'emplacement de chaque électrode est désigné par une lettre renvoyant au lobe dans lequel l'électrode est fixée : F, T, C, P et O pour Frontal, Temporal, Central, Pariétal et occipital respectivement (Z se réfère à une électrode placée sur la ligne médiane), et un chiffre pour déterminer l'emplacement au niveau du lobe. Pour l'hémisphère droit on utilise des chiffres pairs et pour hémisphère gauche on utilise des chiffres impairs. La figure 2.6 représente l'emplacement des différentes électrodes selon le système international 10-20.

Un signal de tension EEG représente une différence de potentiel entre deux électrodes, ainsi les ondes obtenues sur l'EEG diffèrent selon la combinaison en paire des électrodes ou le montage des électrodes. On appelle "montage" la combinaison qui fait correspondre à deux électrodes un amplificateur et un système d'inscription, c'est à dire une voie (ou dérivation). Plusieurs types de montages peuvent être utilisés :

- Montage bipolaire : Une paire d'électrodes représente habituellement un canal. Chaque canal (ou onde) représente la différence entre deux électrodes adjacentes [36]. Par exemple, le canal (FP1 -F3) représente la différence tension entre l'électrode et l'électrode Fp1 et F3. Le canal suivant dans la le montage est le (F3- C3), qui représente la différence de tension entre F3 et C3, et ainsi de suite sur l'ensemble du réseau d'électrodes.

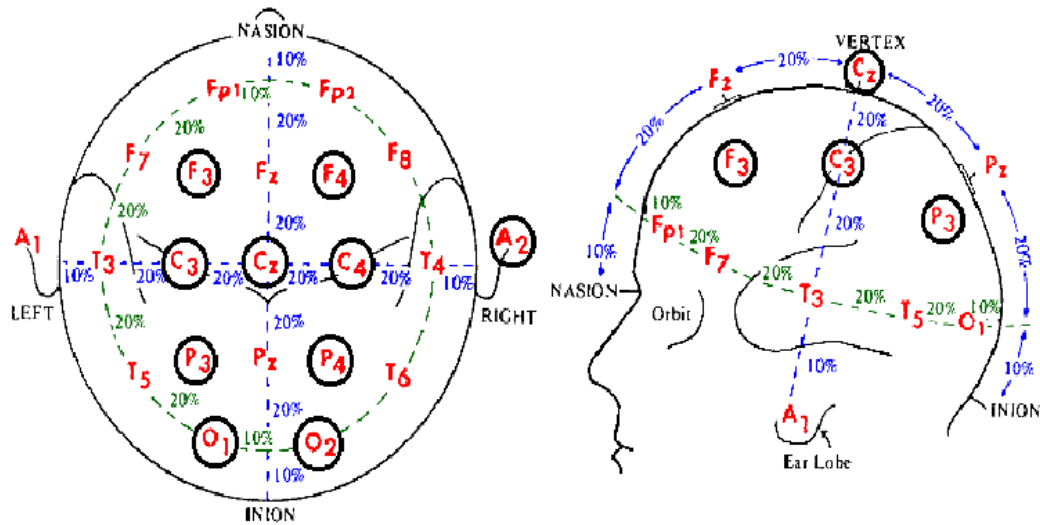


FIGURE 2.6 – Emplacement des électrodes sur le scalp selon le système international 10-20.

- Le montage monopolaire : chaque couple d'électrodes est constitué d'une électrode du scalp "active" et d'une référence. [36] Sur le plan théorique, la méthode monopolaire consiste à enregistrer le potentiel "absolu" d'un point du scalp, c'est à dire la différence entre le potentiel du point étudié et le potentiel nul. Le potentiel nul est celui d'un point à l'infini ou par approximation situé loin du point enregistré. La référence la plus satisfaisante serait donc une électrode extra-céphalique, par exemple de type sterno-épineuse, mais, en pratique, elle n'est utilisable que dans un cas sur cinq en raison des artefacts EEG. Certains auteurs préfèrent choisir une électrode céphalique, sur le menton ou le nez, l'électrode Z placée sur la ligne médiane, ou encore un couple d'électrodes placées sur chaque mastoïde ou aux deux lobes des oreilles, mais l'inconvénient est une diffusion de l'activité électrique à partir des régions cérébrales voisines.

Il est possible d'employer aussi une référence électronique obtenue en liant plusieurs électrodes et donnant ainsi un potentiel moyen qui sert de réfé-

rence. Cet artifice dit référence moyenne de Wilson, constitué par l'ensemble des électrodes reliées entre elles au travers d'une résistance élevée, est souvent utilisé bien qu'ayant l'inconvénient de privilégier, parfois, l'activité d'un capteur prédominant (recueillant un signal de grande amplitude). [36]

Le montage Laplcaien : Chaque canal correspond à la différence entre une électrode et une moyenne pondérée des électrodes environnantes. [36].

La compréhension de l'activité cérébrale par les signaux EEGs se fait par l'étude des caractéristiques morphologiques et l'examen des bandes de fréquences associées à différents activités mentales ou états de conscience. Les bandes de fréquences peuvent être divisées en cinq catégories. Dans la section suivante, nous examinerons les modèles les plus courants de signaux EEGs dans les situations où les individus sont dans un état de vigilance, sommeil, souffrant d'un trouble cérébral et dans un état émotionnel expérimenté.

### 2.2.3 Nature et fréquences des signaux EEGs

La fréquence des oscillations des signaux cérébraux (ou rythme cérébrale) est l'une des données les plus importantes pour l'évaluation des anomalies de l'EEG, et pour comprendre les comportements fonctionnels dans la recherche cognitive. Un rythme cérébral désigne une oscillation dans une bande de fréquence donnée résultant de l'activité électrique cohérente d'un grand nombre de neurones du cerveau. Ces ondes sont de très faible amplitude, elles sont de l'ordre du microvolt (chez l'être humain) et ne suivent pas toujours une sinusoïde régulière. Les activités électriques cérébrales rythmiques sont classées selon leur fréquence en [36] :

- Ondes Delta ( $\delta$ ) : Oscille avec une fréquence comprise entre 0,5 et 4 Hz. Caractérisée par la plus grande amplitude et la plus large longueur d'onde. Correspond à une activité cérébrale minimale, dans le sommeil profond par exemple, ou dans le cas d'une altération cérébrale importante.
- Onde thêta ( $\theta$ ) : Oscille entre 4,5 et 8 Hz, généralement avec des amplitudes supérieures à 20  $\mu$ V. On les observe principalement chez l'enfant,

- l'adolescent et le jeune adulte. Elles caractérisent également certains états de somnolence, d'hypnose, de méditation, de stress émotionnel, en particulier la frustration ou la déception ainsi que la mémorisation d'information.
- Onde alpha ( $\alpha$ ) : Oscille entre 8 et 13 Hz. Ce type d'onde apparaissent généralement au niveau du lobe occipital quand un sujet a les yeux fermés ou dans un état de relaxation.
  - Onde beta ( $\beta$ ) : Avec une fréquence entre 13 et 30 Hz, apparaissent le plus souvent dans le lobe frontal dans les situations d'activité cérébral intense : concentration, résolution de problèmes et anxiété.
  - Onde gamma ( $\gamma$ ) : Avec une fréquence supérieure à 30 Hz apparaissent lors de la réalisation des différentes taches cognitives et motrices.

La figure 2.7 représente ces 5 classes de rythme cérébrale. Les signaux électriques de l'EEG dont l'origine est non-cérébrale sont appelés artefacts. Les données EEG sont dans la plus part des cas contaminées par ces signaux parasites. L'amplitude d'artefacts peut être relativement plus importante par rapport à celle des signaux corticaux. De ce fait un grand nombre de systèmes de filtration d'artefacts ont été mis en point et permettent une lecture plus aisée des données, Cependant l'interprétation correcte des tracés EEGs reste complexe et requiert une expérience suffisante.

## 2.3 Aperçu sur le concept d'interface cerveau-machine (BCI) "machine Brain Computer Interfaces"

Une BCI peut être définie comme une communication et un contrôle via des canaux inhabituels qui ne dépendent pas des voies de commande conventionnelles du cerveau via les nerfs périphériques et les muscles [37]. Les messages et les instructions envoyés par le biais d'une BCI sont encodés suivant l'activité cérébrale de l'utilisateur. La Technologies BCI fournit une interface et permet une communication directe entre le cerveau et un ordinateur [38]. L'objectif ultime d'une BCI

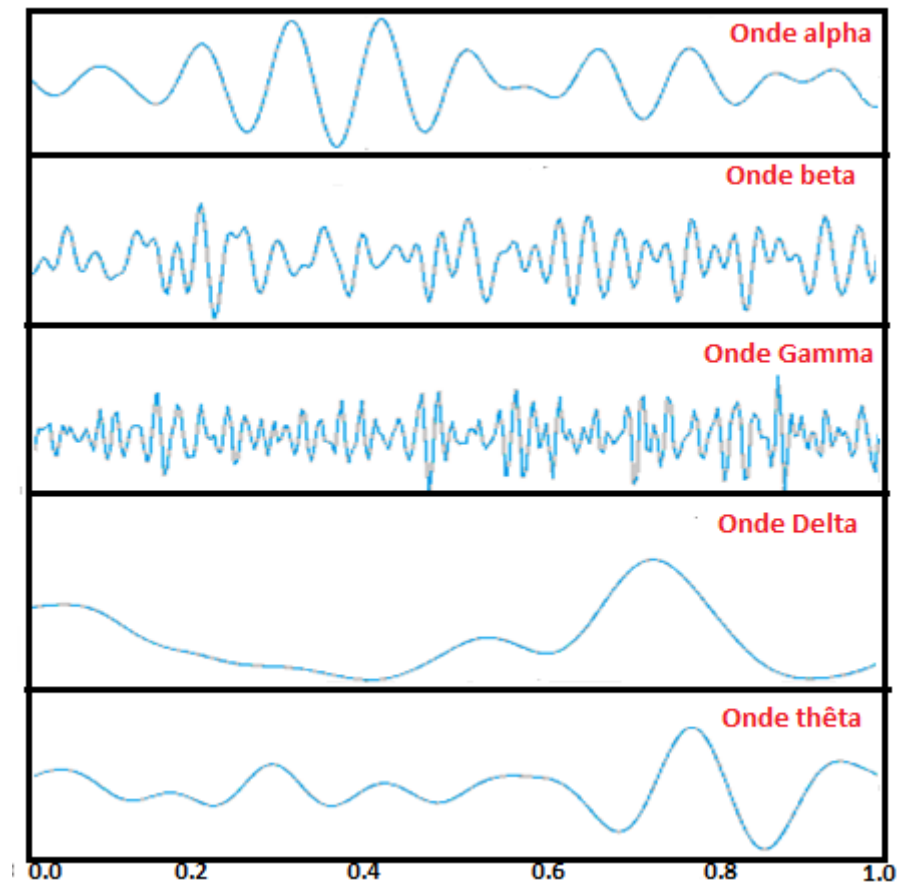


FIGURE 2.7 – Les 5 classes de rythme cérébrale.

est de mettre à disposition des humains une voie de communication alternative, permettant la transmission directe des messages du cerveau par le traitement de son activité électrique.

Une BCI est une technologie qui offre une nouvelle façon de communiquer et de manipuler une machine en utilisant seulement la pensée. Des électrodes sont placées sur la tête d'un utilisateur pour mesurer les signaux EEGs. L'utilisateur, en imaginant une tâche spécifique telle un mouvement des membres ou la composition d'un mot, affecte le rythme électrique de son cerveau. L'ordinateurs détecte et classe ces modèles EEGs associés à différentes tâches afin de les exécuter, exemple :

contrôler une application informatique ou commander un fauteuil roulant.

Contrairement à toutes les autres interfaces, les BCIs ne nécessitent aucun mouvement musculaire et peut-être le seul moyen de communication possible pour les personnes avec un handicap moteur complet. De même, les BCI peuvent également assurer le contrôle à d'autres groupes de sujets souffrant de troubles moteurs moins graves qui souhaitent manipuler une orthèse ou un fauteuil roulant par exemple [39], Les utilisateurs en bonne santé peuvent trouver un grand intérêt dans l'exploitation des BCIs notamment dans certaines situations où les moyens conventionnels de communications sont difficiles, peu pratiques ou inadéquates. [40]

### 2.3.1 Architecture des systèmes BCIs

Généralement, un système BCI est organisé en une succession de 6 étapes formant un circuit fermé : mesure de l'activité cérébrale, prétraitement, extraction des caractéristiques, classification et traduction en une commande et rétroaction [41], représentée sur la figure 2.8.

- Mesure de l'activité cérébrale : La mesure de l'activité du cerveau est une étape délicate et primordiale pour un tel système. Les intentions humaines modulent les signaux électriques du cerveau qui sont mesurés à l'aide de divers types d'électrodes, puis ces signaux sont numérisés. Dans ce projet, nous utilisons les signaux EEGs comme reflet de l'activité cérébrale.
- Prétraitement : le prétraitement vise à simplifier les opérations de traitement ultérieures, en améliorant la qualité du signal sans perte d'information. Dans cette étape, les signaux enregistrés sont traités par débruitage et filtrage afin d'améliorer la qualité de l'information portée par les signaux. [42].
- Extraction des caractéristiques : Les schémas cérébraux utilisés dans BCIs sont spécifiés par certaines caractéristiques. L'extraction de données vise à décrire les signaux par quelques valeurs pertinentes appelées "caractéristiques" [42].

- Classification : L'étape de classification attribue une classe à un ensemble de caractéristiques extraites à partir des signaux. Cette classe correspond au type d'état mentale identifié. Cette étape peut également être désignée : "la traduction des caractéristiques".
- Traduction en une commande : Une fois que l'état mental est identifié, une commande lui est associée afin de contrôler une application donnée tel qu'un ordinateur ou un automate.
- Rétroaction : Enfin, cette étape permet à l'utilisateur des retours perceptifs sur l'action exécutée. Cela offre un meilleur contrôle par l'utilisateur de ses activités cérébrales. L'objectif est d'augmenter les performances.

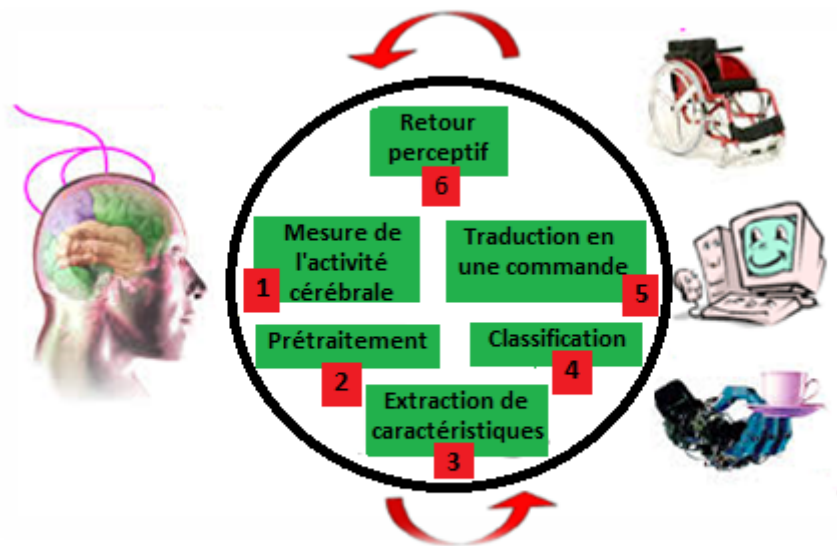


FIGURE 2.8 – Architecture d'une interface cerveau machine.

Une BCI ne peut détecter et classifier que des modèles de signaux spécifiques associés à des tâches ou des événements particuliers. Pour produire ces modèles cérébraux l'utilisateur doit suivre une stratégie mentale du système BCI. La stra-



tégie mentale est le fondement de toute communication cerveau machine.

La stratégie mentale détermine ce que l'utilisateur doit faire pour produire des signaux cérébraux que la BCI peut interpréter. Les stratégies mentales les plus courantes sont l'Imagerie Motrice (MI) et l'Attention Sélective (SA) ciblée. La MI consiste à imaginer un mouvement sans réellement l'effectuer. D'autre part, les BCIs basées sur la SA exigent des stimuli externes fournis par le système BCI. Les stimuli peuvent être auditives ou somatosensoriels.

### **2.3.2 Application des interfaces cerveau machine**

Les champs possibles d'applications des BCIs sont très larges. Ces systèmes ont été validés dans de nombreux domaines, y compris les dispositifs d'orthographe (clavier virtuel), certains jeux vidéo, les neuroprothèses (pour les patients tétraplégiques), la rétroaction biologique (réduction des crises d'épilepsie, renforcement de la réadaptation après un AVC, traitement des troubles l'attention) [43]. La plupart de ces applications sont exécutées sur des ordinateurs conventionnels qui hébergent un système BCI et son application. Ainsi, Un nombre croissant de systèmes permettent le contrôle des dispositifs plus sophistiqués, y compris les orthèses, prothèses, et bras robotiques [44].

*Résumé :*

*Le cerveau est le commandant suprême de l'organisme. Il fait partie de l'encéphale et présente une organisation anatomique parfaitement caractérisée en deux hémisphères gauche et droit. Chacun des hémisphères est divisé fonctionnellement en 4 lobes à activité cérébral spécifique : frontal, pariétale, occipital, et temporal. Le tissu cérébral (le cortex cérébrale) est constitué de milliards de neurones qui communiquent entre eux via des signaux électriques générés par des APs. Le profil électrique ainsi produit est modifié en fonction de l'état physiologique et pathologique du cerveau et selon l'intensité de l'activité cérébrale. Ses signaux électriques peuvent être captés et interprétés sous forme d'un tracé : l'EEG. Le recueil de ces impulsions électriques de l'ordre du mV, nécessite des dispositifs mettant en jeu des amplificateurs et des électrodes spéciales placées sur le scalp en des point précis suivant une standardisation internationale. L'EEG fait partie des examens d'exploration de l'état cérébrale les plus utilisés. Il est exploité par les scientifiques notamment pour l'étude du fonctionnement du cerveau mais aussi pour le diagnostic des atteintes mentales. L'EEG trouve aussi application dans certains domaines révolutionnaire tel que les IBCs qui offre une voie de communication et de contrôle direct entre le cerveau et la machine, un espoir pour les handicapés moteur complet et partiel. Dans le chapitre 5 nous allons essayer de traiter ses signaux EEGs et d'introduire un nouveau système hybride à apprentissage supervisé et non supervisé dont l'objectif est d'améliorer les performances des systèmes BCIs.*

## Chapitre 3

# Introduction à l'électrophysiologie cardiaque

Le coeur est un organe qui a pour fonction principale de faire circuler le sang dans tout l'organisme. Son fonctionnement repose sur une succession d'événements mécaniques et électriques complexes. L'enregistrement ElectroCardioGraphiques (ECG) s'est révélé comme étant une technique primordiale pour l'enregistrement en temps réel de l'activité électrique du coeur.

Dans ce chapitre, nous présenterons quelques notions fondamentales de l'électrophysiologie cardiaque afin de préciser la ligne directrice des travaux exposés. Ce chapitre couvrira en premier lieu une courte description sur l'anatomie du coeur et son fonctionnement. L'accent sera mis ensuite sur l'ECG : sa genèse, son enregistrement, et son analyse. Enfin, nous énumérons brièvement les caractéristiques des arythmies cardiaques considérées dans cette étude.

### 3.1 Anatomie et fonctionnement du cœur humain

Le cœur est la pompe musculaire située dans le thorax et responsable de la circulation du sang dans la totalité des organes de l'organisme. Il est capable de faire circuler 4 à 5 litres de sang, en permanence, depuis la naissance jusqu'à la mort.

Le cœur mesure une quinzaine de centimètres et pèse de 300 à 350 grammes. Il est divisé en deux parties fonctionnant simultanément nommées cœur gauche et cœur droit, Chaque moitié du cœur comprend une oreillette et un ventricule. (Voir figure 3.1)

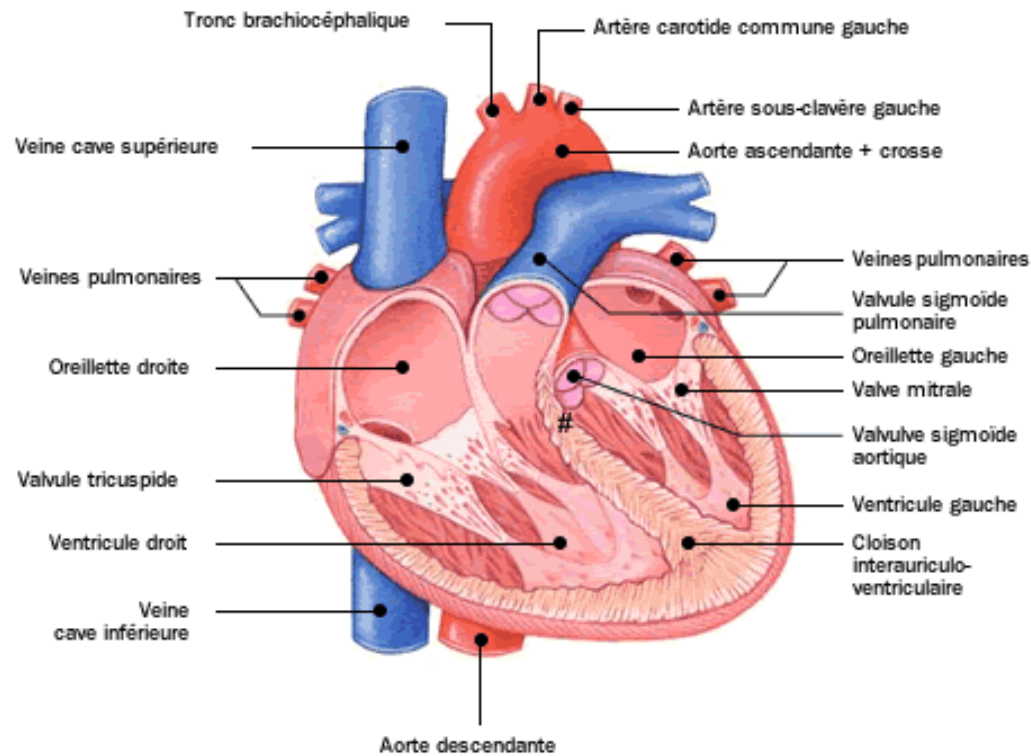


FIGURE 3.1 – Anatomie du cœur humain.[4]

L'oreillette droite reçoit le sang veineux (appauvri en oxygène) provenant des organes périphériques grâce à trois veines (la veine cave supérieure, la veine cave

inférieure et le sinus coronaire). Le sang est alors envoyé vers le ventricule droit, qui va ensuite envoyer le sang vers les poumons à travers l'artère pulmonaire où il sera réoxygéné.

Ensuite, le coeur gauche reçoit le sang oxygéné provenant des poumons grâce à la veine pulmonaire. Le sang est donc à nouveau dans l'oreillette, et grâce à une contraction du ventricule gauche, le sang repart vers l'aorte qui va le redistribuer vers les autres organes.

Une pulsation cardiaque s'effectue en deux étapes : la diastole, pendant laquelle le muscle est relâché et le sang s'introduit ; la systole, au cours de laquelle les oreillettes et les ventricules se contractent.

Dans la systole auriculaire, une diminution de la pression dans chaque ventricule provoque l'ouverture des valvules tricuspide et mitrale, puis le sang envahit les ventricules. De plus, chaque oreillette se contracte et le sang qui en sort finit de remplir les ventricules.

Durant la systole des ventricules, qui commence juste après celle des oreillettes, les ventricules se contractent, les valvules mitrale et tricuspide se ferment, les valvules aortiques et pulmonaires s'ouvrent et le sang est éjecté dans les artères. Quand le coeur se relâche, les valvules aortiques et pulmonaires se ferment et une nouvelle diastole commence.

## 3.2 L'engendrement du signal électrique cardiaque

Le travail du myocarde est dû à la contraction des cellules cardiaques, cette dernière est provoquée par la propagation d'une impulsion électrique le long des fibres musculaires cardiaques. Le coeur est en fait constitué de cellules qui produisent et propagent des impulsions électriques, ainsi que des cellules qui répondent à ces impulsions par une contraction. Pendant la systole (période d'activité) et la diastole (période de repos), les cellules cardiaques sont le siège des échanges d'ions : le sodium  $\text{Na}^+$ , le potassium  $\text{K}^+$ , le calcium  $\text{Ca}^{2+}$  et le chlorure  $\text{Cl}^-$ . Ce qui conduit à la variation de potentiels au cours d'un cycle cardiaque (potentiel d'action AP).

L'AP comprend 5 phases (figure 3.2) :

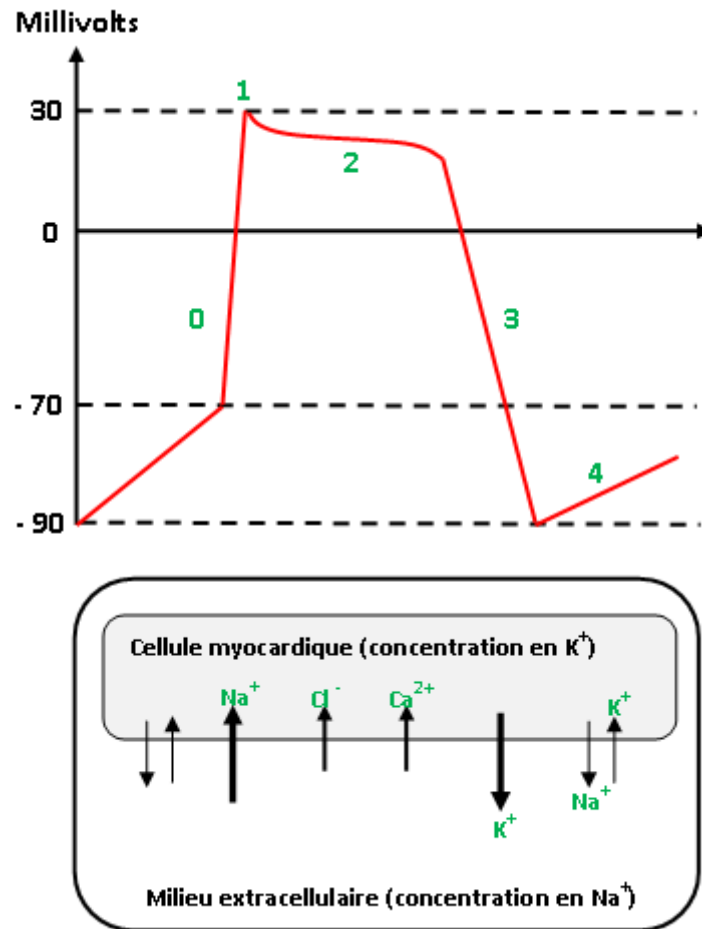


FIGURE 3.2 – Potentiel cellulaire et mouvement ionique. Les cinq phases représentent : la phase 0 la dépolarisation de la cellule, la phase 1 la repolarisation initiale, la phase 2 la repolarisation lente (à phase en plateau), la phase 3 à la repolarisation rapide, et la phase 4 le potentiel de repos.

- phase 0 : la cellule se dépolarise brusquement par l'ouverture des canaux sodiques qui font entrer massivement les ions  $\text{Na}^+$  à l'intérieur de la cellule. La polarisation membranaire est donc inversée, l'intérieur est positif et l'extérieur devient négatif.

- phase 1 : c'est la phase de repolarisation initiale. Les canaux sodiques se referment, et des ions chlorures  $\text{Cl}^-$  entrent à l'intérieur de la cellule. Elle est brève et rapide.
- phase 2 : c'est la phase en plateau. Une nouvelle dépolarisation est cette fois-ci maintenue par l'entrée des ions  $\text{Ca}^{2+}$  par les canaux calciques.
- phase 3 : la cellule se repolarise, son potentiel membranaire reprend sa valeur de repos. Cela est causé par l'ouverture du canal potassique permettant la sortie des ions  $\text{K}^+$  et la fermeture du canal calcique.
- phase 4 : c'est la phase de potentiel de repos. Lors de cette phase, l'intérieur de la membrane cellulaire est chargé négativement par rapport à l'extérieur.

Le résultat de l'addition des variations de potentiels est un vecteur électrique qui constitue le signal cardiaque. Par conséquent, la répétition de ces phénomènes donne le rythme cardiaque.

### 3.3 L'électrocardiographie

L'électrocardiographie permet au médecin d'obtenir à l'aide de l'ECG, un tracé des variations du potentiel électrique en fonction du temps. Le potentiel électrique généré par le myocarde est connu depuis les travaux de Carlo Matteucci en 1842. En 1878 John Burden Sanderson et Frederick Page réalisent les premières expérimentations, et détectent à l'aide d'un électromètre capillaire les phases QRS et T, qui ont été définies par la suite. En même temps, il a été démontré par Augustus Waller [45] et E. Marey [46], que l'activité électrique du cœur peut être suivie à partir de la peau. En 1907, Augustus Waller publie le premier ECG humain. En 1895, les cinq ondes (P, Q, R, S et T) ont été mises en évidence par Willem Einthoven [47]. En 1901, il utilise le galvanomètre à cordes, et en 1906, il expose les premières classifications des ECGs pathologiques. En 1924, Willem Einthoven, a été récompensé par un prix Nobel pour ces recherches sur l'électrocardiographie. Les dérivations précordiales et les dérivations frontales, ont été respectivement mises en usage pour le diagnostic médical, à partir de 1932 et 1942. Ce qui a per-

mis la réalisation du premier tracé sur 12 voies, par Emanuel Goldberger.

Notons qu'une conférence internationale transatlantique a eu lieu en 1938, afin de fixer la position des dérivations précordiales, V1 à V6.

Dans le reste du chapitre, nous exposons tout d'abord les principales méthodes d'enregistrements de l'ECG, et par la suite nous nous attarderons sur son analyse.

### 3.3.1 L'enregistrement (ECG)

L'ECG est le tracé papier de l'activité électrique dans le coeur. Ce tracé est recueilli, soit par la méthode invasive en plaçant des électrodes à la surface du coeur (enregistrement endocavitaire), ou bien, par des électrodes placées à la surface de la peau, on parle alors de la méthode non-invasive. cette dernière est la plus utilisée, car c'est une technique rapide ne prenant que quelques minutes, indolore et dénuée de tout danger. (Note : L'ECG fera référence à l'ECG non-invasive)

La technique non invasive est subdivisée en deux types d'enregistrements de l'activité cardiaque : l'ECG de repos et l'ECG d'effort. l'ECG de repos est enregistré par le biais des électrodes placées sur un patient qui est allongé. Tandis que l'ECG d'effort est enregistré durant un exercice dynamique, soit sur un cyclo-ergomètre, soit sur un tapis roulant. Cet examen est utile pour déterminer des pathologies qui ne peuvent survenir qu'à l'effort. La méthode d'exploration de ces deux techniques, emploient l'ECG à 12 dérivations réalisé par commutation de la position des électrodes périphériques.

A noter, qu'il existe des méthodes d'enregistrement n'utilisons pas le système des 12 dérivations tel que "le Holter", qui représente un enregistrement continu sur 24 heures. Aussi, il existe la technique dite "le mapping", qui utilise un nombre d'électrodes qui varie entre 20 et 200, placées à la surface du corps pour enregistrer l'activité cardiaque.



### 3.3.1.1 Le dipôle électrique

La compréhension du fonctionnement des dérivations, repose principalement sur les notions exposées dans la section 3.2.

En fait, l'ECG est une représentation graphique de l'activité électrique du coeur liée aux variations des APs. Ce dernier, correspond à la propagation d'une zone de dépolarisation (positive) dans un myocarde repolarisé (négatif). Ce qui est équivalent à tracer le vecteur exposant la différence de potentiel entre le front positif et le front négatif. ce qui donne :

- Une ligne isoélectrique (un enregistrement nul), quand le coeur est repolarisé (négatif) ou complètement dépolarisé (positif).
- Un signal positif, à l'approche du front positif.
- Un signal négatif, à l'éloignement du front positif.

### 3.3.1.2 Système des 12 dérivations de l'ECG

Les différences de potentiel sont reportées entre deux électrodes placées à la surface du corps. On appelle ligne de dérivation (ou dérivation) en électrocardiographie, la droite théorique passant par deux électrodes d'enregistrement. Un système de dérivation est un ensemble cohérent de dérivations, choisi de manière à étudier la totalité du potentiel électrique du coeur. On peut trouver plusieurs systèmes de dérivations, mais en général en utilise les 12 dérivations standards, a savoir :

**Dérivations bipolaires :** Ces trois dérivations forment le triangle d'Einthoven. En théorie, ce triangle est équilatéral et le coeur en occupe le centre :

- I ou DI = bras droit, bras gauche.
- II ou DII = bras droit, jambe gauche.
- III ou D III = bras gauche, jambe gauche.

**Dérivations unipolaires des membres :** Le pôle positif de l'électrocardiogramme est relié à une électrode placée à la surface du corps. Tandis que le pôle négatif de l'électrocardiogramme est relié à une électrode neutre (borne centrale de Wilson).

- aVR = bras droit borne centrale.
- aVL = bras gauche borne centrale.
- aVF = jambe gauche borne centrale.

La ligne de chacune de ces trois dérivations passe par le centre géométrique du triangle et l'un des sommets. (Voir figure 3.3)

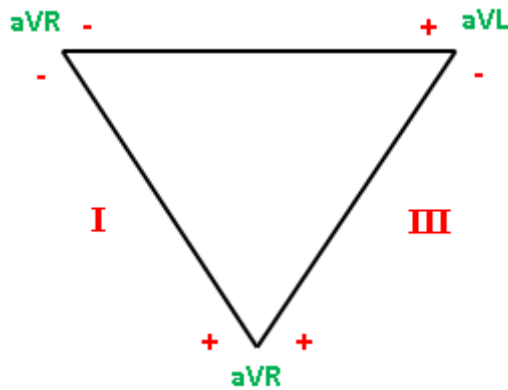


FIGURE 3.3 – Triangle d'Einthoven.

Les dérivations bipolaires et unipolaires des membres permettent l'étude de l'activité électrique cardiaque dans le plan frontal.

**Dérivations précordiales des membres :** A fin d'explorer l'activité électrique cardiaque dans un plan horizontal, 6 électrodes sont placées sur le thorax, comme illustré par la figure 3.4. Le pôle positif de l'appareil est relié à l'électrode exploratrice, et le pôle négatif à la borne centrale.

- V1 = 4ème espace intercostal droit au bord droit du sternum
- V2 = 4ème espace intercostal gauche au bord gauche du sternum

- V3 = entre V2 et V4.
- V4 = 5ème espace intercostal gauche, sur la ligne médio-claviculaire gauche.
- V5 = sur la ligne axillaire antérieure à la même hauteur que V4.
- V6 = sur la ligne axillaire moyenne à la même hauteur que V4.

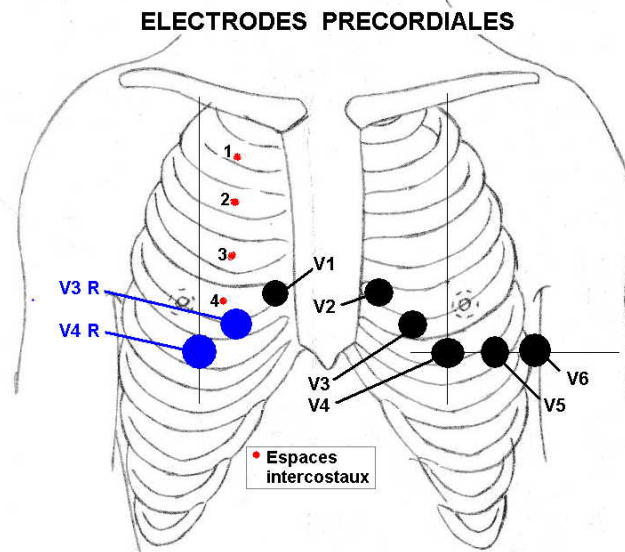


FIGURE 3.4 – Dérivations précordiales des membres.

### 3.3.2 Analyse de L'ECG

L'ECG est réalisé grâce à des électrodes placées en des points bien précis permettant de définir des dérivations standards. Il permet l'enregistrement de l'activité électrique du coeur. Par conséquent, le suivi du battement cardiaque. Chaque phase du battement est constitué d'un ensemble de traces électriques particulières (ondes et intervalles), qui peuvent être facilement reconnues par un oeil expert.

### 3.3.2.1 Ondes, intervalles et segments d'un signal ECG

L'activité électrique cardiaque normale prend naissance dans le noeud sinusal. En fait, l'impulsion initiale vient du sinus et elle n'est pas visible sur l'ECG. L'onde électrique se propage par la suite dans les oreillettes, qui se contractent entraînant la dépolarisation auriculaire, celle-ci est traduite par l'onde P. L'impulsion arrive alors au noeud auriculo ventriculaire, où se produit la courte pause qui se traduit sur l'ECG par un petit segment plat. Puis elle arrive aux ventricules entraînant leur dépolarisation, ce qui dessinent sur l'ECG une succession de trois ondes (Q, R et S) appelé complexe QRS. L'onde T correspond au courant de repolarisation des cellules constituant les ventricules. Cette onde succède au complexe QRS après retour à la ligne isoélectrique, pendant cette phase le coeur est mécaniquement inactif. On peut aussi constater l'onde U qui suit l'onde T, et dont l'origine est discutée. (Voir figure 3.5)

Les intervalles et segments d'un tracé ECG représentent des paramètres importants, pour l'évaluation de la ou la non normalité, du temps écoulé entre deux événements électriques.

- L'intervalle P-R, correspond au délai entre la dépolarisation de l'oreillette et celle du ventricule. Il représente le temps de conduction auriculo-ventriculaire.
- L'intervalle Q-T, correspond au temps de systole ventriculaire, depuis l'excitation des ventricules jusqu'à la fin de leur relaxation.
- Le segment S-T, correspond au début de la repolarisation ventriculaire.
- L'intervalle R-R, sépare les sommets de deux ondes R successives.

L'ensemble des ondes, intervalles et segments sont illustrés par figure 3.6.

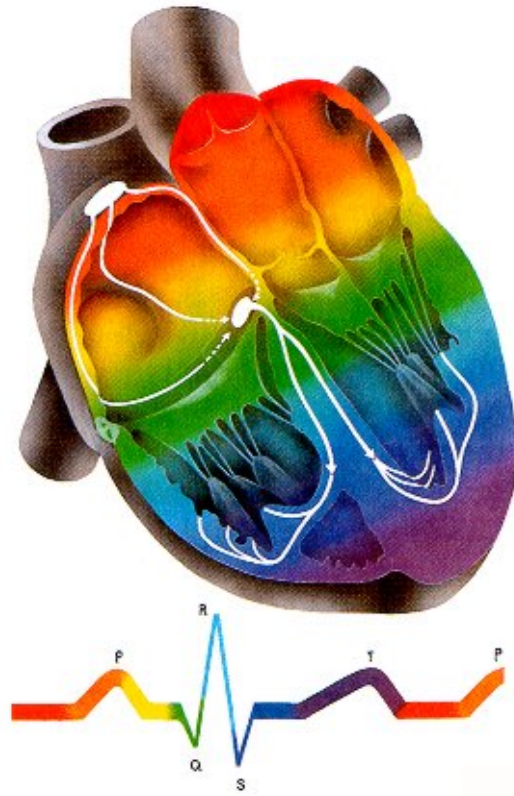


FIGURE 3.5 – Processus de propagation du potentiel d'action dans le coeur et tracé ECG généré. L'onde P, représente la contraction des deux oreillettes. Le complexe QRS, représente la contraction des ventricules. L'onde T, représente la repolarisation des ventricules.

### 3.3.2.2 Caractéristiques d'un ECG sein

L'analyse de la ou la non normalité d'un tracé ECG, doit être méthodique et systématique et doit comporter l'étude de [48] :

- La fréquence et le rythme cardiaque.
- Les onde P, T, et complexe QRS : morphologie, amplitude, durée, et axe électrique.
- L'onde U : morphologie, amplitude, durée.
- L'intervalle PR : durée

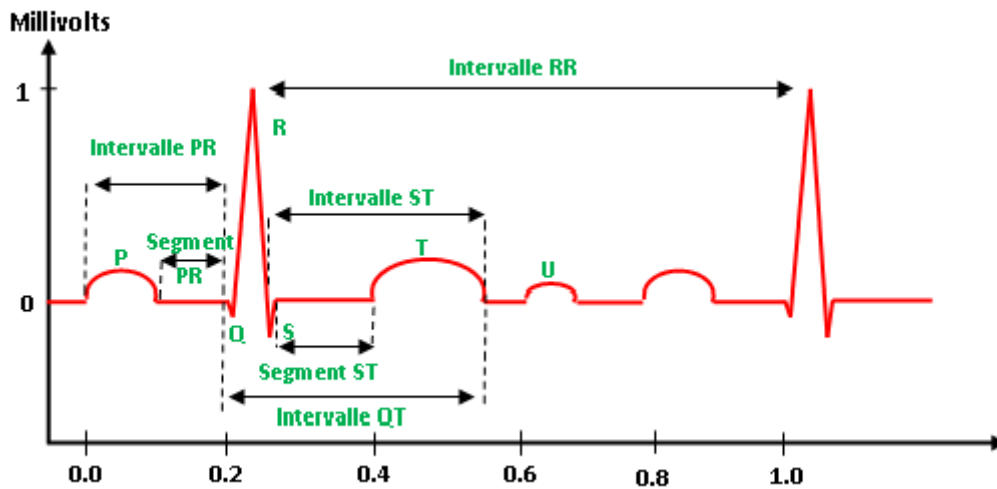


FIGURE 3.6 – Ondes, intervalles et segments d'un signal ECG

- Les intervalles ST et QT : morphologie, durée.

**La fréquence et le rythme cardiaque :** Le rythme cardiaque normal est le rythme sinusal, dont les principales caractéristiques sont :

- Chaque onde P suivie d'un complexe QRS et inversement.
- Une fréquence cardiaque comprise entre 60 et 100 bpm.

La fréquence cardiaque est toujours référencée à la fréquence ventriculaire. Elle est donc calculée en comptant le nombre de complexes QRS par minute. La méthode la plus simple, est de diviser le chiffre de 300 par le nombre de grands carrés séparant deux RR consécutifs. Une fréquence inférieure à 60 bpm est signe d'une bradycardie, tandis qu'une fréquence supérieure à 100 bpm est signe d'une tachycardie.

**Onde P :**

- L'onde P est négative en aVR et positive DI, DII, DIII, et V6.
- L'axe moyen en dérivation standard entre  $+30^\circ$  et  $+70^\circ$ .

- Sa durée et son amplitude sont inférieures ou égales à 0,12 sec et à 2.5 mm respectivement.
- Biphasique avec une composante dominante positive ou monophasique avec une déflexion positive en V1.

### **Complexes QRS**

- La durée est inférieure ou égale à 0,12 sec. et l'amplitude de la plus grande déflexion en dérivation standard excède 5 mm de hauteur.
- L'axe moyen en dérivation standard entre  $-30^\circ$  et  $+90^\circ$ .
- L'onde R croît de V1 à V6; en V1-V2 l'image observée est du type (RS, Q ou QR); en V5-V6 on observe une image du type QRS; et en V3-V4 on enregistre une image transitionnelle (R et S ont la même amplitude).

### **Onde T :**

- L'amplitude de T oscille souvent entre 1 et 4 mm, et doit être au moins 2 mm en hauteur en dérivation où elle est plus grande.
- Sa durée est de 0,20 à 0,25 secondes, l'analyse de sa durée est comprise dans l'analyse de la durée de l'intervalle QT.
- Positive en DI, DII et de V4 à V6, et négative en aVR. Elle est normalement positive mais occasionnellement négative en DIII et aVF.
- L'axe moyen en dérivation standard entre  $0^\circ$  et  $+70^\circ$ .

### **Onde U :** Elle peut ne pas exister, mais, si elle est présente, elle devra :

- être plus grande en V2 et V4.
- positive en V1 et V2 et en dérivation avec une onde T positive.
- Son amplitude représente 25% de l'amplitude de l'onde T, mais peut atteindre 2 mm en V2 et V4.

**Intervalle PR :**

- Sa durée est comprise entre 0,12 et 0,24 sec.. Il diminue si la fréquence cardiaque augmente et augmente avec l'âge.

**Intervalle ST :** C'est la portion séparant le complexe QRS de l'onde T. Le segment ST va du point J, fin du complexe QRS, au début de l'onde T.

- Le segment est généralement isoélectrique.
- La déflexion, si elle existe, ne doit pas dépasser 1mm en DI, DII, DIII et aVF.
- Possibilité de l'élévation jusqu'à 4 mm dans les dérivations V1-V3.

**Intervalle QT :** La durée normale de l'intervalle QT varie avec la fréquence cardiaque, Selon la formule de Bazett :

$$QT_{corrige'} (QTc) = \frac{QT_{observe'}}{\sqrt{intervalleRR (ensecondes)}}. \quad (3.1)$$

Ou QT est le temps entre le début du complexe QRS à la fin de l'onde T.

La valeur tolérée du QTc doit être inférieure à 440 millisecondes.

### 3.3.3 Arythmies cardiaques étudiées

On définit comme arythmie, Tout battement compromettant l'activité électrique ou la fonction de pompage du coeur. En fait, l'arythmie se présente lorsque l'impulsion électrique initiale prend naissance ailleurs que dans le noeud sinusal, dans les ventricules par exemple, ou que l'onde électrique ne suit plus les voies normale de propagation.

Dans cette thèse, nous avons considéré quatre arythmies cardiaques les plus fréquentes : extrasystole ventriculaire (PVC), extrasystole auriculaire (APB), bloc de branche gauche (LBBB), bloc de branche droit (RBBB), en plus du profil électrocardiographique spécifique au port d'un pacemaker (PB), et la classe des battements normaux.



### 3.3.3.1 L'extrasystole

Elle représente une contraction prématurée du coeur provoquée par des stimulations électriques qui naissent d'un foyer ectopique et non du noeud sinusal.

**Extrasystole ventriculaire (PVC)** Elle prend son origine dans les ventricules. Ce trouble du rythme cardiaque est le plus fréquent des arythmies et son diagnostic repose sur les deux éléments suivants : l'existence d'un complexe QRS prématuré et élargi (plus de 0,12 seconde et différent des complexes QRS de base), l'absence d'onde P prématurée avant le complexe QRS. Aussi, on peut constater un rythme normal et régulier, dû à la pause compensatrice qui suit l'extrasystole.

Une PVC peut survenir, entre deux contractions normales, après deux révolutions cardiaques normales (extrasystole bigéminée), ou après trois battements cardiaques normales (extrasystole trigéminée), ou en salves. (Voir exemple d'une PVC trigéminée figure 3.7)

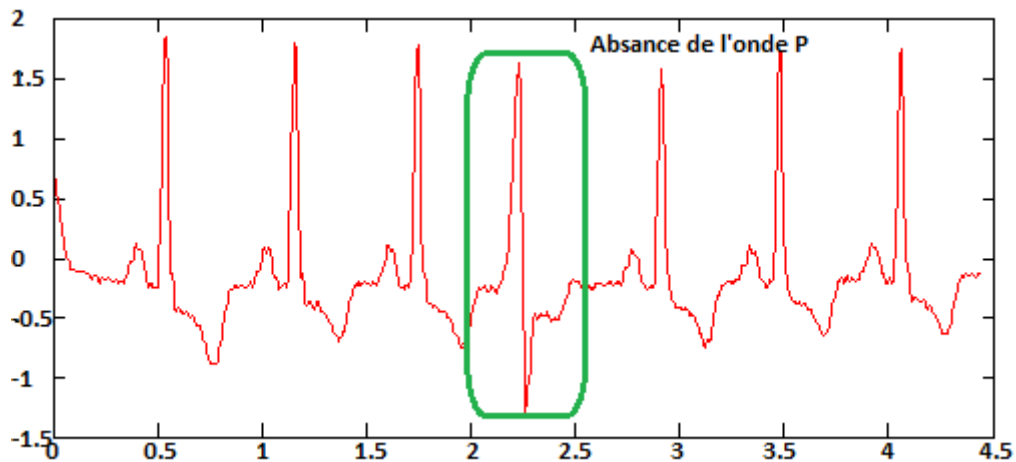


FIGURE 3.7 – Extrasystole ventriculaire.

**Extrasystole auriculaire (APB)** C'est une systole dont le point de départ de l'influx se situe dans un point quelconque de l'oreillette. Elle est caractérisée

par une onde P prématurée qui peut être de morphologie différente (on peut la confondre avec l'onde T du complexe précédent), suivie, ou non, d'un complexe QRS. En présence du QRS, on parle d'une APB avec aberration de conduction. Dans ce cas, le complexe est souvent identique aux QRS normal, mais parfois anormal, élargi. De même, on peut avoir un QRS absent, on parle alors d'une APB bloquée.

De même que les PVCs les APBs peuvent être isolées, en couplet, bigémées, trigémées, ou en salve. (Voir l'exemple d'une APB bloquée figure 3.8)

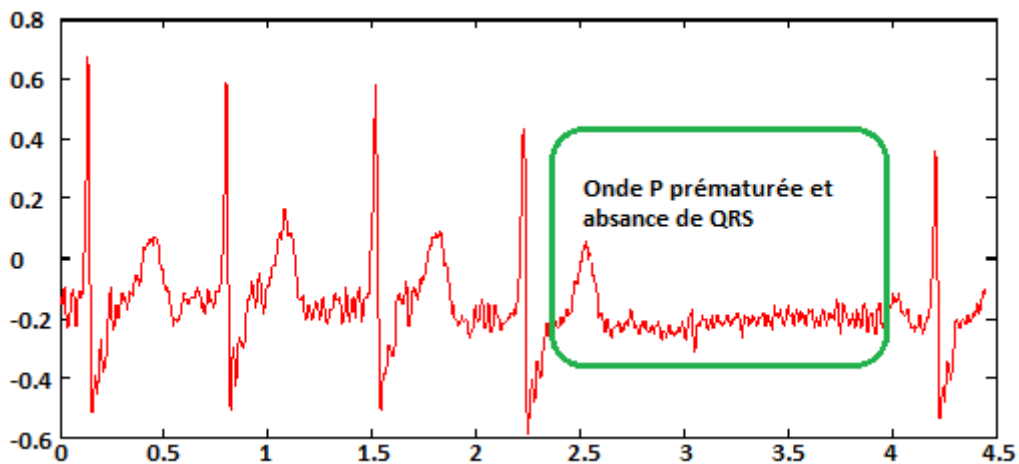


FIGURE 3.8 – Extrasystole auriculaire.

### 3.3.3.2 Bloc de branche

On parle de bloc de branche, lorsque la conduction cardiaque est interrompue ou ralentie dans les branches du faisceau de His, au niveau des ventricules cardiaques. En conséquence, un asynchronisme de contraction ventriculaire est engendré (détérioration de la capacité des ventricules à pomper le sang et diminution de la quantité de sang éjectée par les ventricules), d'où l'élargissement du complexe QRS. On fait référence à un bloc de branche complet si le QRS est élargi à plus de 0,12 sec., et à un bloc de branche incomplet si la durée de QRS est entre 0,09

sec. et 0,12 sec..

**Bloc de branche gauche (LBBB)** L'existence d'un LBBB, l'interruption de la transmission du courant dépolarisant dans la branche gauche du faisceau de His, l'activation du ventricule gauche se fait à travers le ventricule droit et donc nécessairement après ce dernier. Ce qui induit à une contraction désynchronisée des ventricules, traduite sur l'ECG sous la forme d'un QRS large (durée du complexe QRS  $> 0.120$  sec.).

En V4, V5 et V6, on a habituellement un QRS : croché de l'onde R ( rR' ou RR'), ou aspect (rsR' ou RsR'), ou avec un plateau. On s'attend aussi à trouver en cas de LBBB des segments ST et des ondes T décalées approximativement de moins de 5 mm. (Voir l'exemple d'un LBBB figure 3.9)

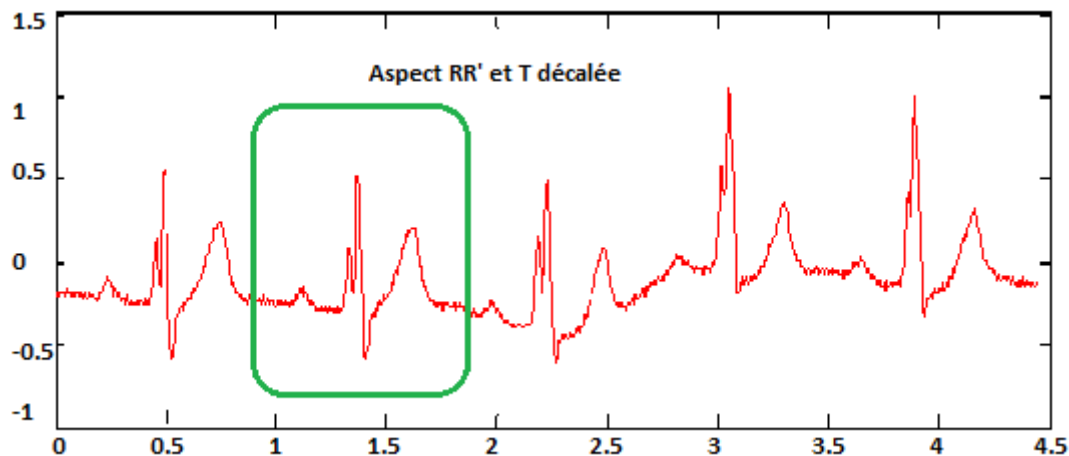


FIGURE 3.9 – Bloc de branche gauche.

**Bloc de branche droit (RBBB)** L'existence d'un RBBB, signifie que le ventricule droit n'est pas activé par les impulsions électriques progressant au travers la branche droite du faisceau de His. Cependant, le ventricule gauche, est normalement activé par la branche gauche. Ces impulsions progressent au travers du ventricule gauche vers le ventricule droit, activant finalement celui-ci (l'activation

du ventricule droit est retardée, impliquant ainsi un complexe QRS élargi). Sur l'ECG on constate, une onde T inversée en V1, un aspect rSR' ou RR' en V1 et V2, et une onde S large et profonde. (Voir l'exemple d'un RBBB figure 3.10)

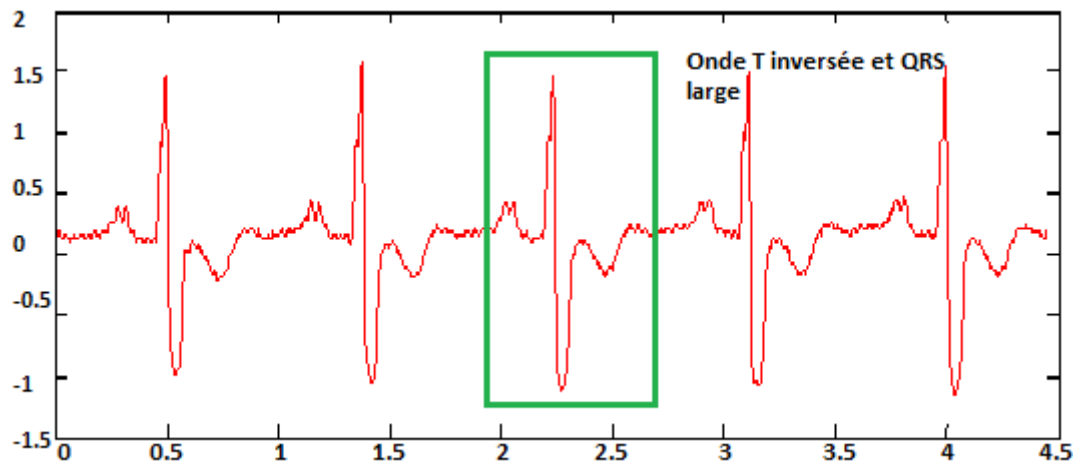


FIGURE 3.10 – Bloc de branche droit.

### 3.3.3.3 pacemaker (PB)

Le PB est un stimulateur cardiaque, il est destiné à générer des impulsions électriques à fin d'assurer un rythme cardiaque suffisamment élevé pour les patient dont le rythme est trop lent. La présence d'un PM se traduit sur un tracé ECG par l'existence de spikes avant l'activité atriale dans le cas d'une stimulation auriculaire, et/ou précèdent l'activité ventriculaire si la stimulation est ventriculaire.

*Résumé :*

*Le coeur constitue le moteur du système cardio-vasculaire, dont le rôle est de pomper le sang oxygène et le faire distribuer aux organes. Le coeur est divisé en deux parties coeur gauche et coeur droit, chaque moitié comprend une oreillette et un ventricule. À chaque battement une impulsion électrique traverse l'ensemble des cellules du coeur, engendrant les contractions successives des oreillettes et des ventricules et pompage du sang vers les poumons et vers le reste du corps.*

*La propagation de cette impulsion électrique peut être suivie en temps réel grâce à L'ECG, qui est obtenue en plaçant des électrodes à la surface du corps dans des points bien précis. Ce signal est susceptible à un ensemble de modifications (modifications au niveau des ondes, intervalles, et segments) en cas de pathologie due à des anomalies de la commande de l'influx électrique ou de sa propagation par exemple. Ce qui fait de l'ECG un moyen d'exploration efficace pour le suivi de l'état du coeur. Dans la suite de la thèse, le chapitre 6 présente en détails la méthodologie adoptée pour l'extraction des paramètres constituant un signal ECG en utilisant la Transformée par Ondelettes Discrètes (DWT). Aussi, nous exposerons et nous discuterons les résultats obtenus après application des Machines à Vecteurs de Support Multiclasses (M-SVMs) directes et indirectes.*

## Chapitre 4

# Machines à Vecteurs de Support Multi Classes Probabilistes

Les Machines à Vecteurs de Support (SVMs) sont actuellement une approche très puissante pour des problèmes de régression et de classification. Ce sont des modèles discriminants dont le procédé est de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. Plusieurs études ont établi leur supériorité par rapport aux méthodes de discrimination tels que les réseaux de neurones. En effet, les SVMs offrent : une bonne performance en généralisation, une robustesse vis à vis de la dimensionnalité (peut traiter des problèmes mettant en jeu de grands nombres de descripteurs des données), et une insensibilité au sur-apprentissage.

Les SVMs sont conçus pour traiter des données étiquetées par rapport à deux classes d'appartenance. Pour la plupart des modèles d'apprentissage, l'extension au cas multi classes est facile et semble parfois même naturelle ; en revanche, la supériorité des SVMs ne cache pas la grande difficulté de leur adaptation aux

problèmes de discrimination multi classes (M-SVMs). Depuis la première extension proposée par Vapnik [49], plusieurs chercheurs se sont attachés à utiliser les SVMs dans des applications à plusieurs classes. Les approches utilisées jusqu'à nos jours sont réparties en deux catégories : La première catégorie de méthodes, qu'on qualifie d'indirectes, permet d'aborder le problème à catégories multiples comme la fusion de problèmes de calcul de plusieurs classifieurs binaires. Tandis, que la deuxième catégorie de méthodes, qu'on qualifie de directe, essaye de résoudre un problème d'optimisation général. Cette dernière catégorie couvre principalement quatre grands modèles à savoir : le modèle de Weston et Watkins (WW) [14], le modèle de Crammer et Singer (CS) [15], le modèle de Lee et al. (LLW) [16], et la M-SVM à coût quadratique (M-SVM<sup>2</sup>) [17]. En 2012 Guermeur développe un modèle générique des M-SVMs qui regroupe les spécificités des quatres modèles des M-SVMs directes [50].

D'autre part, dans la plus part des applications de reconnaissance des formes, la decision finale du classifieur doit être munie d'une mesure de confiance, telle que la probabilité a posteriori; cependant, à la base, les SVMs ne sont pas conçus pour fournir de telles mesures. Par consequent, Platt [18] propose une méthode de post-traitement très simple, qui a permis de représenter les sorties des SVMs sous forme de probabilités. Cette Méthode de post-traitement a été quelques années plus tard approuvée [19] et améliorée par plusieurs auteurs [19, 20, 21].

Dans ce chapitre, nous présenterons dans un premier lieu quelques notions et définitions qui sont les fondements de base des M-SVMs. Nous considérons ensuite la première approche mise en oeuvre pour traiter des problèmes multi classes au moyen des méthodes de décomposition. Puis, nous exposerons les principaux modèles de l'approche directe, en mettant en évidence leurs spécificités par le biais du modèle générique. En fin, nous décriverons les démarches suivies pour post-traiter en terme de probabilités, les différentes sorties des M-SVMs directes et indirectes.

## 4.1 Cadre théorique et notions de base

Nous nous plaçons dans un problème de classification à  $Q$ -categories, avec  $3 \leq Q < +\infty$ . Chaque objet est représenté par sa description  $x$  appartenant à l'ensemble de descriptions  $\mathcal{X}$  et l'ensemble  $\mathcal{Y}$  des catégories  $y \in \llbracket 1, Q \rrbracket$ . L'ensemble  $\mathcal{X}$  des descriptions et l'ensemble  $\mathcal{Y}$  des catégories sont deux espaces probabilistes, et le lien  $\mathcal{X} \times \mathcal{Y}$  est caractérisé par une mesure de probabilité  $P$  inconnue. Le problème d'apprentissage consiste donc à sélectionner une fonction  $g$  à partir d'un ensemble de fonctions  $\mathcal{G}$  sur  $\mathcal{X}$ ; de manière à affecter les descriptions aux différentes catégories avec une probabilité d'erreur minimale aussi identique que possible de celle de la règle de décision de Bayes (risque).

Soit  $d_g$  la règle de décision (de  $\mathcal{X}$  vers  $\mathcal{Y} \cup \{*\}$ ) associée à  $g$ , et  $(X, Y)$  un couple aléatoire à valeurs dans  $\mathcal{X} \times \mathcal{Y}$  distribué suivant  $P$ . Idéalement, la fonction objectif à minimiser sur  $\mathcal{G}$  est le risque  $P(h(X) \neq Y)$ . Dans la pratique, puisque  $P$  est inconnu, le risque ne peut pas être utilisé comme fonction objective. On a donc recours, à une procédure qui s'appuie sur l'utilisation d'un échantillon d'apprentissage  $\mathbf{z}_m = ((X_i, Y_i))_{1 \leq i \leq m}$  constitué de copies de  $(X, Y)$  indépendantes, cet échantillon va contribuer à construire un estimateur empirique du risque. Ce dernier, représente l'espérance par rapport à  $P$  d'une fonction de perte  $l$ , définie sur  $\mathcal{Y} \times \mathcal{G}(\mathcal{X})$  par :  $l(y, g(x)) = \mathbb{1}_{g_y(x) \leq \max_{k \neq y} g_k(x)}$ . Le risque de  $g$  se calcule alors comme suit :

$$R(g) = \mathbb{E}_{(X, Y)} \left[ \mathbb{1}_{\{g_y(x) \leq \max_{k \neq y} g_k(x)\}} \right] = \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{1}_{\{g_y(x) \leq \max_{k \neq y} g_k(x)\}} dP(x, y).$$

Sachant que Pour une fonction donnée  $g \in \mathcal{G}$ , sa règle de décision  $d_g$  correspondante est définie comme suit :

$$\forall i \in \mathcal{X}, \begin{cases} \text{si } \exists k \in \llbracket 1, Q \rrbracket : g_k(x) > \max_{l \neq k} g_l(x), \text{ then } d_g = k \\ \text{sinon } d_g = *. \end{cases}$$



avec  $*$  désigne une catégorie fictive introduite pour traiter les cas d'égalité (compté comme une réponse erronée et contribue au risque).

Les SVMs appartiennent à la famille des machines à noyau (*Kernel machines*) [51, 52], ce qui implique que la classe des fonctions sur lesquelles elles opèrent sont induites par un noyau symétrique semi-défini positif (ou noyau de Mercer). La connexion entre noyau de Mercer et l'espace de Hilbert à noyau reproduisant est assurée par le théorème de Moore-Aronszan. Dans ce qui suit, nous donnons les définitions concernant : Le noyau symétrique semi-défini positif, l'espace de Hilbert à noyau reproduisant, théorème de Moore-Aronszan, et la classe des fonctions. Aussi, nous donnerons une petite définition sur les SVMs à marge douce et dure.

**Definition 1** (*Le noyau symétrique semi-défini positif ou noyau de Mercer, Définition 2 [53], Définition 1 [17]*) : la fonction  $\kappa$  sur  $\mathcal{X}^2$  est appelée noyau de Mercer, si elle est symétrique et

$$\forall n \in \mathbb{N}^*, \forall (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n, \forall (a_i)_{1 \leq i \leq n} \in \mathbb{R}^n, \sum_{i=1}^n \sum_{j=1}^n a_i a_j \kappa(x_i, x_j) \geq 0$$

**Definition 2** (*L'espace de Hilbert à noyau reproduisant, Définition 1 [53], Définition 2 [17]*) : soit l'espace d'Hilbert  $(\mathbf{H}, \langle \cdot, \cdot \rangle_{\mathbf{H}})$  des fonctions à valeurs réelles sur  $\mathcal{X}$ . La fonction  $\kappa$  sur  $\mathcal{X}^2$  est un noyau de reproduction de  $\mathbf{H}$  s'il satisfait les deux conditions suivantes :

1.  $\forall x \in \mathcal{X}, \kappa_x = \kappa(x, \cdot) \in \mathbf{H}$  ;
2.  $\forall x \in \mathcal{X}, \forall h \in \mathbf{H}, \langle \cdot, h \rangle_{\mathbf{H}} = h(x)$ . (*propriété de reproduction*)

L'espace d'Hilbert qui possède un noyau de reproduction est appelé L'espace de Hilbert à noyau reproduisant (RKHS).

**Theorem 1** (*théorème de Moore-Aronszan, Theoreme 3 [53], Theoreme 1[17]*) : Soit  $\kappa$  un noyau symétrique semi-défini positif. Il existe un unique  $(\mathbf{H}, \langle \cdot, \cdot \rangle_{\mathbf{H}})$  avec  $\kappa$  comme noyau reproduisant.

**Definition 3** (*Les classes des fonctions  $\overline{\mathcal{H}}$  et  $\mathcal{H}$ , Définition 3 [17]*) : Soit  $\kappa$  un noyau de Mercer et  $(\mathbf{H}_{\kappa}, \langle \cdot, \cdot \rangle_{\mathbf{H}_{\kappa}})$  est RKHS correspondant. Alors,  $\overline{\mathcal{H}}$  représente

l'espace de Hilbert des fonctions à valeurs vectorielles définies comme suit :  $\overline{\mathcal{H}} = \mathbf{H}_\kappa^Q$  et  $\overline{\mathcal{H}}$  est muni du produit scalaire  $\langle \cdot, \cdot \rangle_{\overline{\mathcal{H}}}$  donné par :

$$\forall(\overline{h}, \overline{h}') \in \overline{\mathcal{H}}^2, \quad \overline{h} = (\overline{h}_k)_{1 \leq k \leq Q}, \quad \overline{h}' = (\overline{h}'_k)_{1 \leq k \leq Q}, \quad \langle \overline{h}, \overline{h}' \rangle_{\overline{\mathcal{H}}} = \sum_{i=1}^Q \langle \overline{h}_k, \overline{h}'_k \rangle_{\mathbf{H}_\kappa}.$$

Soit  $\{1\}$  est l'espace des fonctions constantes à valeurs réelles sur  $\mathcal{X}$ .

$$\mathcal{H} = \overline{\mathcal{H}} \oplus \{1\}^Q = (\mathbf{H}_\kappa \oplus \{1\})^Q.$$

Pour un noyau donné  $\kappa$ , soit la carte de fonction  $\phi$  de  $\mathcal{X}$  dans l'espace de représentation (feature space)  $\mathbf{H}_\kappa$  définie par :

$$\forall x \in \mathcal{X}, \quad \phi(x) = \kappa_x.$$

On peut conclure à partir de la Définition 3 et les propriétés de reproduction, que les fonctions  $h$  de  $\mathcal{H}$  peuvent s'écrire comme suit [17] :

$$h(\cdot) = \overline{h}(\cdot) + b = (\overline{h}_k(\cdot) + b_k)_{1 \leq k \leq Q} = (\langle \overline{h}_k, \phi(\cdot) \rangle_{\mathbf{H}_\kappa} + b_k)_{1 \leq k \leq Q}$$

ou  $\overline{h} = (\overline{h}_k)_{1 \leq k \leq Q} \in \overline{\mathcal{H}}$  et  $b = (b_k)_{1 \leq k \leq Q} \in \mathbb{R}^Q$ .

**Definition 4** Si une SVM est entraînée sous la contrainte  $\xi = 0$ , elle est appelée SVM de marge dure. Sinon, elle est appelée SVM de marge douce.

## 4.2 Les M-SVMs basées sur les méthodes de décomposition

La première approche présentée dans cette étude pour l'extensions des SVMs au cas multiclassés et l'approche directe. Cette approche est basée principalement sur la fusion des SVM bi-classes.

Soit  $\mathcal{X}$  un ensemble non vide (l'espace de description), et  $\mathcal{Y} = (y_i)_{1 \leq i \leq m} \in \{-1, 1\}$  est le vecteur des étiquettes des exemples de formation. Soit  $\kappa$  un noyau symétrique semi-défini positif sur  $\mathcal{X}^2$ , et  $\mathbf{H}_\kappa, \mathcal{H}_\kappa$  deux classes de fonctions induites par  $\kappa$  selon la définition 3. Ainsi, étant donné un ensemble de formation  $\mathbf{z}_m = \{(x_i, y_i) : 1 \leq i \leq m\}$ , les SVMs binaires sont des discriminateurs qui pour une fonction donnée  $g$ ; la règle de décision associé est comme suit :

$$\forall x \in \mathcal{X} \begin{cases} g(x) < 0 \Leftrightarrow d_g = -1 \\ g(x) = 0 \Leftrightarrow d_g = * \\ g(x) > 0 \Leftrightarrow d_g = 1 \end{cases},$$

avec  $*$  désigne une catégorie fictive introduite pour traiter les cas d'égalité. Une SVM binaire est un modèle discriminant à grande marge obtenu par la résolution d'un problème de programmation quadratique convexe de la forme suivante :

**problem 1** (*Apprentissage d'une SVM binaire, formulation primale, dans [54]*)

$$\begin{aligned} \min_{h \in \mathcal{H}, b \in \mathbf{R}} & \left\{ \frac{1}{2} \|\bar{h}\|_{\mathcal{H}}^2 + C \sum_{i=1}^m \xi_i \right\} \\ \text{s.t.} & \begin{cases} \forall i \in \llbracket 1, m \rrbracket, y_i h(x_i) \geq 1 - \xi_i. \\ \forall i \in \llbracket 1, m \rrbracket, \xi_i \geq 0. \end{cases} \end{aligned}$$

Où  $C$  représente un paramètre positif libre qu'on doit fixé, il constitue une balance entre les deux termes de la fonction objective : la marge et les erreurs permises (la maximisation de la marge et la minimisation de l'erreur de classification).

Résoudre le probleme 1 revient à résoudre le Wolf dual d'une SVM binaire en utilisant des multiplicateurs de Lagrange  $\alpha_i$  et  $\beta_i$ .

**Wolf dual du problème 1** Soit  $(\alpha_i)_{1 \leq i \leq m} \geq 0$  le vecteur des multiplicateurs de Lagrange associés aux contraintes de bonne classifications. Soit  $(\beta_i)_{1 \leq i \leq m} \geq 0$  est le vecteur des multiplicateurs de Lagrange associés aux contraintes de non négativité des variables d'écart.

$$L(h, b, \xi, \alpha, \beta) = \frac{1}{2} \|\bar{h}\|_{\bar{H}}^2 + C \sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i (y_i (\bar{h} + b) - 1 + \xi_i) - \sum_{i=1}^m \beta_i \xi_i.$$

A l'optimum, les dérivées par rapport aux variables :  $\bar{h}$  (équation 4.1),  $b$  (équation 4.2), et  $\xi_i$  (équation 4.3) s'annulent.

$$\nabla_{\bar{h}} L(h, b, \xi, \alpha, \beta) = \bar{h} - \sum_{i=1}^m \alpha_i y_i \kappa_{x_i}. \quad (4.1)$$

$$\frac{\partial}{\partial b} L(h, b, \xi, \alpha, \beta) = \sum_{i=1}^m y_i \alpha_i. \quad (4.2)$$

$$\frac{\partial}{\partial \xi} L(h, b, \xi, \alpha, \beta) = C - \alpha_i - \beta_i. \quad (4.3)$$

A noter aussi, que le produit des contraintes aux multiplicateurs s'annule. Par conséquent, Les conditions d'optimalité de Karush, Kuhn et Tucker qui permettent de caractériser la solution du problème primal  $(h^*, b^*, \xi^*)$  et les multiplicateurs de lagrange  $\alpha_i^*$  et  $\beta_i^*$  associés s'écrivent :

$$\left\{ \begin{array}{l} \bar{h}^* = \sum_{i=1}^m \alpha_i^* y_i \kappa_{x_i}. \\ \sum_{i=1}^m y_i \alpha_i^* = 0. \\ C = \alpha_i^* + \beta_i^*. \\ \alpha_i^* (y_i (\bar{h}^* + b) - 1 + \xi_i^*) = 0. \\ \beta_i^* \xi_i^* = 0. \end{array} \right. \quad (4.4)$$

En substituant l'ensemble des déductions exposé dans la formule (4.4), dans le lagrangien on obtient le problème dual suivant :

**problem 2** (*Apprentissage d'une SVM binaire à marge douce, formulation duale, dans [54]*)

$$\begin{aligned} & \max_{\alpha} \left\{ -\frac{1}{2} \alpha^T H \alpha + I_m^T \alpha \right\} \\ \text{s.t.} & \begin{cases} y^T \alpha = 0 \\ \forall i \in \llbracket 1, m \rrbracket, 0 \leq \alpha_i \leq C. \end{cases} \end{aligned}$$

ou  $H = (h_{i,j})_{1 \leq i,j \leq n} \in M_{m,m}(\mathbb{R})$  est une matrice hessien avec  $h_{i,j} = y_i y_j \kappa_{x_i}$

Avec de légères modifications, le calcul ci-dessus peut être adapté pour exprimer le Wolfe Dual du problème d'apprentissage d'une machine à marge dure. Nous obtenons alors :

**problem 3** (*Apprentissage d'une SVM binaire à marge dure, formulation duale, dans [54]*)

$$\begin{aligned} & \max_{\alpha} \left\{ -\frac{1}{2} \alpha^T H \alpha + I_m^T \alpha \right\} \\ \text{s.t.} & \begin{cases} y^T \alpha = 0 \\ \forall i \in \llbracket 1, m \rrbracket, 0 \leq \alpha_i. \end{cases} \end{aligned}$$

L'extension des SVMs au cas multi classes (M-SVMs) en utilisant l'approche indirecte, consiste à fusionner plusieurs classifieurs binaires. Les résultats obtenus par chaque classifieur binaire sont ensuite combinés pour donner le résultat final. L'approche indirecte inclut trois méthodes principales : un-contre-un, un-contre-tous, et les codes correcteurs d'erreurs.

**Un-contre-un** : dans l'approche un-contre-un (1-c-1)  $C_Q^2 = Q(Q-1)/2$  classifieurs sont construits, chacun vise à distinguer entre deux classes  $(k, l)$  (avec  $1 \leq k < l \leq Q$ ). Après la résolution du problème 1 et l'obtention  $C_Q^2$  fonction de décision, le classifieur multi-classes est construit en utilisant la stratégie "Maxwin" [55] : si un classifieur attribue un exemple  $x_i$  à la classe  $k$ , un vote est ajouté à la classe  $k$ , en même temps qu'un vote est diminué de la classe  $l$ .

**Un-contre-tous :** l'approche un-contre-tous (1-c-tous) consiste à utiliser un classifieur binaire par catégorie. Le classifieur  $k$  permet de faire la distinction entre la classe  $k$  et toutes les autres. Après avoir résolu le problème 1, la décision finale est prise à l'aide de la stratégie "winner-takes-all" [55] : la classe sélectionnée est celle associée au classifieur qui retourne la valeur la plus élevée.

**Les Codes correcteurs d'erreurs :** Une autre approche couramment employée dans le cadre de la combinaison des SVMs est donnée par l'approche basée sur les Codes CORrecteurs d'Erreurs (ECOC), Elle consiste à associer à chaque classe  $k \in \mathcal{Y}$  des mots binaires de même taille  $N$  (les mots codes), correspondant à une ligne d'une matrice de codage  $M$ , telle que  $M(l, k) = (-1, 1)$  ou  $M(l, k) = (1, 0)$ . Ainsi, on obtient une matrice  $M$  dont chaque ligne définit un mot codes et dont Chaque colonne définit un problème de discrimination binaire. Après avoir estimé des modèles SVMs binaires sur tous les jeux de données induits par les colonnes de la matrice  $M$ , nous obtenons  $l$  fonctions de décision :  $h_l, l \in \llbracket 1, N \rrbracket$ . Une nouvelle observation  $x_i$  est ainsi affectée à la classe qui correspond à la ligne de la matrice  $M$  la plus proche du vecteur :  $h_1(x), \dots, h_N(x)$  au sens de la distance de Hamming.

### 4.3 Les M-SVMs directes

A l'opposé de l'approche indirecte qui est basée sur les méthodes de décomposition, l'approche directe considère toutes les classes simultanément et traite le problème d'apprentissage des M-SVMs comme étant un unique problème d'optimisation. On trouve quatre grands modèles directs : le modèle de Weston et Watkins (WW) [14], le modèle de Crammer et Singer (CS) [15], le modèle de Lee et al. (LLW) [16], et la M-SVM à coût quadratique (M-SVM<sup>2</sup>) [17], modèle introduit par Guermeur et Monfrini. En 2012 Guermeur développe un modèle générique de M-SVM qui comprend tous les quatre modèles des M-SVM [50].

### 4.3.1 Le modèle générique

Soit  $\mathcal{X}$  un ensemble non vide (l'espace de description), et  $\mathcal{Y}$  est l'ensemble des catégories  $\in \llbracket 1, Q \rrbracket$  avec  $3 \leq Q < +\infty$ . Soit  $\kappa$  un noyau symétrique semi-défini positif sur  $\mathcal{X}^2$ , soit  $\mathbf{H}_{\kappa, Q}, \mathcal{H}_{\kappa, Q}$  deux classes de fonctions induites par  $\kappa$  selon la définition 3. Soit  $P_{\mathbf{H}_{\kappa, Q}}$  l'opérateur de projection orthogonale de  $\mathbf{H}_{\kappa, Q}$  sur  $\mathcal{H}_{\kappa, Q}$ . Pour  $m \in \mathbb{N}^*$ , Soit  $\mathbf{z}_m = ((x_i, y_i))_{1 \leq i \leq m} \in (\mathcal{X} \times \llbracket 1, Q \rrbracket)^m$ ,  $\mathbf{y}_m = (y_i)_{1 \leq i \leq m}$ , et  $\xi \in \mathbb{R}^{Qm}(\mathbf{y}_m)$ . Une M-SVM à  $Q$ -catégories est un modèle discriminant à grande marge obtenu par la résolution d'un problème de programmation quadratique convexe de la forme :

**problem 4** (*Apprentissage d'une M-SVM, formulation primale, dans [50]*)

$$\min_{h \in \mathcal{H}_{\kappa, Q}, \xi \in \mathbb{R}^{Qm}(Y_m)} \left\{ \lambda \left\| P_{\mathbf{H}_{\kappa, Q}} h \right\|_{\mathbf{H}_{\kappa, Q}}^2 + \|M\xi\|_p^p \right\}$$

$$s.t. \begin{cases} \forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus \{y_i\}, & K_1 h_{y_i}(x_i) - h_k(x_i) \geq K_2 - \xi_{ik} \\ \forall i \in \llbracket 1, m \rrbracket, \forall (k, l) \in (\llbracket 1, Q \rrbracket \setminus \{y_i\})^2, & K_3(\xi_{ik} + \xi_{il}) = 0 \\ \forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus \{y_i\}, & (2 - p)\xi_{ik} \geq 0 \\ (1 - K_1) \sum_{k=1}^Q h_k = 0 \end{cases}$$

Ou  $\lambda \in \mathbb{R}_+^*$ ,  $(K_1, K_3) \in \{0, 1\}^2$ ,  $K_2 \in \mathbb{R}_+^*$ ,  $M \in M_{cm, cm}(\mathbf{z}_m)$  est une matrice de rang  $(Q - 1)m$ , et  $p \in \{1, 2\}$ . Si  $p = 1$ , alors  $M$  est une matrice diagonale.

**Wolf dual du problème 4, dans [50]** .

Soit  $\alpha = (\alpha_{ik})_{1 \leq i \leq m, 1 \leq k \leq Q} \in \mathbb{R}^{Qm}(d_m)$  le vecteur des multiplicateurs de Lagrange associés aux contraintes de bonne classification. Soit  $\beta$  est le vecteur des multiplicateurs de Lagrange associés aux contraintes de non négativité des variables d'écart. Si  $K_3 = 0$ , alors  $\beta = (\beta_{ik})_{1 \leq i \leq m, 1 \leq k \leq Q} \in \mathbb{R}^{Qm}(d_m)$ , sinon  $\beta = (\beta_i)_{1 \leq i \leq m} \in \mathbb{R}^{Qm}(d_m)$ .  $\gamma \in \mathbf{H}_\kappa$  et  $\delta \in \mathbb{R}$  sont des multiplicateurs de Lagrange associés respectivement aux contraintes  $\sum_{k=1}^Q \bar{h}_k = 0$  et  $\sum_{k=1}^Q b_k = 0$ .

$$L(h, \xi, \alpha, \beta, \gamma, \delta) =$$

$$\|M\xi\|_p^p + \lambda \|P_{\mathcal{H}} h\|_{\mathcal{H}}^2 - \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik} \{K_1 h y_i(x_i) - h_k(x_i) - k_2 + k_3 \xi_i + (1 - k_3) \xi_{ik}\}$$

$$- (2-p) \sum_{i=1}^m \{k_3 \beta_i \xi_i + (1 - k_3) \sum_{k \neq y_i} \beta_{ik} \xi_{ik}\} - (1 - K_1) \langle \gamma, \sum_{k=1}^Q \bar{h}_k \rangle_{\mathcal{H}_\kappa} - (1 - k_1) \delta \sum_{k=1}^Q b_k.$$

Par application de la propriété de reproduction,

$$\forall i \in \llbracket 1, Q \rrbracket, \quad \nabla_{\bar{h}_k} L(h, \xi, \alpha, \beta, \gamma, \delta) = 2\lambda \bar{h}_k - k_1 \sum_{\{i: y_i=k\}} \sum_{i \neq k} \alpha_{ik} \kappa_{x_i}$$

$$+ \sum_{\{i: y_i \neq k\}} \alpha_{ik} \kappa_{x_i} - (1 - k_1) \gamma.$$

Ainsi, à l'optimum,

$$\forall k \in \llbracket 1, Q \rrbracket, (1 - K_1) \gamma^* =$$

$$2\lambda \bar{h}_k^* - k_1 \sum_{\{i: y_i=k\}} \sum_{i \neq k} \alpha_{il}^* \kappa_{x_i} + \sum_{\{i: y_i \neq k\}} \alpha_{ik}^* \kappa_{x_i} - (1 - k_1) \gamma. \quad (4.5)$$

L'addition de l'ensemble des indexes donne,

$$(1 - K_1) \gamma^* = \frac{2\lambda}{Q} \sum_{k=1}^Q \bar{h}_k^* + \frac{(1 - k)}{Q} \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \kappa_{x_i}.$$



Une conséquence directe de cette équation est que même lorsque  $k_1 = 1$

$$\sum_{k=1}^Q \bar{h}_k^* = 0.$$

Par substitution dans (4.5), nous obtenons

$$\begin{aligned} \forall k \in \llbracket 1, Q \rrbracket, \quad \bar{h}_k^* &= \frac{1}{2\lambda} \left\{ k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il}^* \kappa_{x_i} + (1-k_1) \frac{1}{Q} \sum_{i=1}^m \sum_{i \neq y_i} \alpha_{il}^* \kappa_{x_i} - \sum_{\{i, y_i \neq k\}} \alpha_{ik}^* \kappa_{x_i} \right\} \\ &= \frac{1}{2\lambda} \sum_{i=1}^m \sum_{i \neq y_i} \left\{ k_1 \delta_{y_i, k} + (1-k_1) \frac{1}{Q} - \delta_{k, l} \right\} \alpha_{il}^* \kappa_{x_i} \end{aligned}$$

On tenant compte du fait que  $(\alpha_{iy_i})_{1 \leq i \leq m} = 0_m$ , ceci se simplifie à :

$$\forall k \in \llbracket 1, Q \rrbracket, \quad \bar{h}_k^* = \frac{1}{2\lambda} \sum_{i=1}^m \sum_{i=1}^Q \left\{ k_1 \delta_{y_i, k} + (1-k_1) \frac{1}{Q} - \delta_{k, l} \right\} \alpha_{il}^* \kappa_{x_i} \quad (4.6)$$

$$\forall k \in \llbracket 1, Q \rrbracket, \quad \frac{\partial}{\partial b_k} L(h, \xi, \alpha, \beta, \gamma, \delta) = -k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il} + \sum_{\{i, y_i \neq k\}} \alpha_{ik} - (1-K_1)\delta.$$

Ainsi, à l'optimum

$$\forall k \in \llbracket 1, Q \rrbracket, \quad (1-K_1)\delta^* = -k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il}^* + \sum_{\{i, y_i \neq k\}} \alpha_{ik}^*. \quad (4.7)$$

Une sommation sur tous les indices  $k$  donne :

$$(1-K_1)\delta^* = \frac{1-K_1}{Q} \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^*.$$

Par substitution dans (4.7), en prenant en compte le fait que  $(\alpha_{iy_i})_{1 \leq i \leq m} = 0_m$  on aura

$$\forall k \in \llbracket 1, Q \rrbracket, \quad \sum_{i=1}^m \sum_{i=1}^Q \left\{ k_1 \delta_{y_i, k} + (1-k_1) \frac{1}{Q} - \delta_{k, l} \right\} \alpha_{il}^*. \quad (4.8)$$

Pour calculer le gradient de la fonction lagrangienne par rapport au vecteur  $\xi$ , nous distinguons les quatre cas possibles correspondant aux valeurs  $(k_3, p)$ .

$$\begin{aligned}
 &— (k_3 = 0) \wedge (p = 1) \\
 &\forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \frac{\partial}{\partial \xi_{ik}} L(h, \xi, \alpha, \beta, \gamma, \delta) = m_{ik,ik} - \alpha_{ik} - \beta_{ik}.
 \end{aligned}$$

En conséquence, nous obtenons

$$\forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \alpha_{ik}^* + \beta_{ik}^* = m_{ik,ik} \quad (4.9)$$

$$\begin{aligned}
 &— (k_3 = 0) \wedge (p = 2) \\
 &\forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \frac{\partial}{\partial \xi_{ik}} L(h, \xi, \alpha, \beta, \gamma, \delta) = 2(M^T M \xi)_{ik} - \alpha_{ik}.
 \end{aligned}$$

Ainsi,

$$\forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, 2(M^T M \xi^*)_{ik} = \alpha_{ik}^*. \quad (4.10)$$

A noter que puisque  $M \in M_{Qm, Qm}(\mathbf{z}_m)$ , Nous avons aussi

$$\forall i \in \llbracket 1, m \rrbracket, (M^T M \xi^*)_{iy_i} = 0,$$

Ainsi (3.6) peut être étendu à :

$$2M^T M \xi^* = \alpha^*. \quad (4.11)$$

$$\begin{aligned}
 &— (k_3 = 1) \wedge (p = 1) \\
 &\forall i \in \llbracket 1, m \rrbracket, \frac{\partial}{\partial \xi_{ik}} L(h, \xi, \alpha, \beta, \gamma, \delta) = \sum_{k \neq y_i} m_{ik,ik} - \sum_{k \neq y_i} \alpha_{ik} - \beta_i.
 \end{aligned}$$

On constatera que

$$\forall i \in \llbracket 1, m \rrbracket, \sum_{k \neq y_i} m_{ik,ik} = \sum_{k \neq y_i} \alpha_{ik}^* + \beta_i^*. \quad (4.12)$$

$$\begin{aligned}
 &— (k_3 = 1) \wedge (p = 2) \\
 &\forall i \in \llbracket 1, m \rrbracket, \frac{\partial}{\partial \xi_{ik}} L(h, \xi, \alpha, \beta, \gamma, \delta) = 2 \sum_{k \neq y_i} (M^T M \xi)_{ik} - \sum_{k \neq y_i} \alpha_{ik}.
 \end{aligned}$$

Ainsi

$$\forall i \in \llbracket 1, m \rrbracket, 2 \sum_{k \neq y_i} (M^T M \xi^*)_{ik} = \sum_{k \neq y_i} \alpha_{ik}^*. \quad (4.13)$$

A l'optimum, les termes de la fonction lagrangienne impliquant le vecteur  $b$  disparaissent, à savoir

$$-\sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 b y_i^* - b_k^*\} = 0 \quad (4.14)$$

Aussi

$$-\sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 b y_i^* - b_k^*\} = \sum_{k=1}^Q b_k^* \{-k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il}^* + \sum_{\{i, y_i \neq k\}} \alpha_{ik}^*\};$$

et en raison de (4.7), le côté droit de cette équation peut être réécrit comme suit :

$$\sum_{k=1}^Q b_k^* \{-k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il}^* + \sum_{\{i, y_i \neq k\}} \alpha_{ik}^*\} = \sum_{k=1}^Q b_k^* (1 - K_1) \delta^* = 0.$$

Par application de (4.6),

$$\lambda \|P_{\overline{\mathcal{H}}} h\|_{\overline{\mathcal{H}}}^2 = \frac{1}{4\lambda} \sum_{k=1}^Q \sum_{i=1}^m \sum_{j=1}^m \sum_{i=1}^Q \sum_{n=1}^Q \{k_1 \delta_{y_i, k} + (1 - K_1) \frac{1}{Q} - \delta_{k, l}\} \\ \{k_1 \delta_{y_j, k} + (1 - K_1) \frac{1}{Q} - \delta_{k, n}\} \kappa(x_i, x_j) \alpha_{il}^* \alpha_{jn}^*$$

A partir de,

$$\lambda \|P_{\overline{\mathcal{H}}} h\|_{\overline{\mathcal{H}}}^2 = \frac{1}{4\lambda} \sum_{k=1}^Q \sum_{i=1}^m \sum_{j=1}^m \sum_{i=1}^Q \sum_{n=1}^Q \{k_1 \delta_{y_i, k} + (1 - K_1) \frac{1}{Q} - \delta_{k, l}\} \\ \{k_1 \delta_{y_j, k} + (1 - K_1) \frac{1}{Q} - \delta_{k, n}\} \kappa(x_i, x_j) \alpha_{il}^* \alpha_{jn}^* = \\ K_1 \delta_{y_i, y_j} - k_1 \delta_{y_i, n} + (1 - K_1) \frac{1}{Q} - (1 - K_1) \frac{1}{Q} - k_1 \delta_{y_j, l} - (1 - K_1) \frac{1}{Q} + \delta_{l, n} = \\ K_1 (\delta_{y_i, y_j} - \delta_{y_i, n} \delta_{y_j, l}) - (1 - K_1) \frac{1}{Q} + \delta_{l, n},$$

l'expression du pénaliseur à l'optimum se simplifie a,

$$\lambda \|P_{\overline{\mathcal{H}}}h\|_{\overline{\mathcal{H}}}^2 = \frac{1}{4\lambda} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^Q \sum_{l=1}^Q \{K_1(\delta_{y_i, y_j} - \delta_{y_i, l} \delta_{y_j, k}) - (1 - K_1) \frac{1}{Q} + \delta_{k, l}\} \kappa(x_i, x_j) \alpha_{ik}^* \alpha_{jl}^* \quad (4.15)$$

En faisant usage de la propriété de reproduction, on obtient

$$\sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 \bar{h}_{y_i}^*(x_i) - \bar{h}_k^*(x_i)\} = \sum_{k=1}^Q \langle \bar{h}_k^*, k_1 \sum_{\{i, y_i=k\}} \sum_{i \neq k} \alpha_{il}^* \kappa_{x_i} - \sum_{\{i, y_i \neq k\}} \alpha_{ik}^* \kappa_{x_i} \rangle$$

Par application de (4.5) on obtient

$$\begin{aligned} \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 \bar{h}_{y_i}^*(x_i) - \bar{h}_k^*(x_i)\} &= \sum_{k=1}^Q \langle \bar{h}_k^*, 2\lambda \bar{h}_k^* - (1 - K_1) \delta^* \rangle_{\mathcal{H}_k} \\ &= 2\lambda \|P_{\overline{\mathcal{H}}}h\|_{\overline{\mathcal{H}}}^2 - \langle (1 - K_1) \sum_{k=1}^Q \bar{h}_k^*, \delta^* \rangle_{\mathcal{H}_k} \end{aligned}$$

Puisque  $(1 - K_1) \sum_{i=1}^Q \bar{h}_k^*$  en fonction des contraintes du problème 1, et puisque  $\sum_{i=1}^Q \bar{h}_k^* = 0$  cela simplifie en,

$$\sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 \bar{h}_{y_i}^*(x_i) - \bar{h}_k^*(x_i)\} = 2\lambda \|P_{\overline{\mathcal{H}}}h\|_{\overline{\mathcal{H}}}^2 \quad (4.16)$$

Soit  $H \in M_{Qm, Qm}(\mathbb{R})$  une matrice a terme générale :

$$\forall (i, j, k, l) \in \llbracket 1, m \rrbracket \times \llbracket 1, m \rrbracket \times \llbracket 1, Q \rrbracket \times \llbracket 1, Q \rrbracket,$$

$$h_{ij, jl} = \{k_1(\delta_{y_i, y_j} - \delta_{y_i, l} - \delta_{y_j, k}) - (1 - k_1) \frac{1}{Q} + \delta_{k, l}\} \kappa(x_i, x_j). \quad (4.17)$$

La combinaison de (4.15), (4.16) et (4.17) donne :

$$\lambda \|P_{\overline{\mathcal{H}}}h\|_{\overline{\mathcal{H}}}^2 - \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_1 \bar{h}_{y_i}^*(x_i) - \bar{h}_k^*(x_i)\} = -\frac{1}{4\lambda} \alpha^{*T} H \alpha^*. \quad (4.18)$$

Soit  $\mathbf{1}_{Qm}$  est un vecteur de  $\mathbb{R}_{Qm}$  dont les composantes sont toute égales à 1. Etant donné (4.14) et (4.17), à l'optimum,

$$L(h^*, \xi^*, \alpha^*, \beta^*, \gamma^*, \delta^*) = -\frac{1}{4\lambda} \alpha^{*T} H \alpha^* + K_2 \mathbf{1}_{Qm}^T \alpha^* + \mathbf{J}(\xi^*)$$

Avec

$$\mathbf{J}(\xi^*) =$$

$$\|M\xi^*\|_p^p - \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{k_3 \xi_i^* + (1 - k_3) \xi_{ik}^*\} - (2 - p) \sum_{i=1}^m \{k_3 \beta_i^* \xi_i^* + (1 - k_3) \sum_{k \neq y_i} \beta_{ik}^* \xi_{ik}^*\}$$

Ainsi, pour obtenir l'expression de la fonction objectif dual, il reste à exprimer  $\mathbf{J}(\xi^*)$  en tant que fonction de  $\alpha^*$ . À cette fin, nous distinguons les cas où  $p = 1$  et  $p = 2$ .

$$\begin{aligned} & \text{--- } (p = 1) \\ \mathbf{J}(\xi^*) &= \sum_{i=1}^m \sum_{k \neq y_i} m_{ik,ik} \{k_3 \xi_i^* + (1 - k_3) \xi_{ik}^*\} - \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik} \{k_3 \xi_i^* + (1 - k_3) \xi_{ik}^*\} \\ & \quad - \sum_{i=1}^m \{k_3 \beta_i^* \xi_i^* + (1 - k_3) \sum_{k \neq y_i} \beta_{ik}^* \xi_{ik}^*\} \\ &= K_3 \sum_{i=1}^m \left\{ \sum_{k \neq y_i} m_{ik,ik} - \sum_{k \neq y_i} \alpha_{ik}^* - \beta_i^* \right\} \xi_i^* + (1 - k_3) \sum_{i=1}^m \sum_{k \neq y_i} \{m_{ik,ik} - \alpha_{ik}^* - \beta_{ik}^*\}. \end{aligned}$$

L'application de (4.9) et (4.12) dans la partie droite de cette équation donne :

$$\mathbf{J}(\xi^*) = 0$$

$$\begin{aligned} & \text{--- } (p = 2) \\ \mathbf{J}(\xi^*) &= \xi^{*T} M^T M \xi - \sum_{i=1}^m \sum_{k \neq y_i} \alpha_{ik}^* \{K_3 \xi_i^* + (1 - K_3) \xi_{ik}^*\}. \end{aligned}$$

Ici, on distingue une fois de plus les cas  $k_3 = 0$  et  $k_1 = 0$ .

$$\text{--- } (k_3 = 0) \wedge (p = 2)$$

Soit  $N = M^T M$ . Par définition de  $M$ , les lignes et les colonnes de  $N$  dont les indices sont ceux des variables d'écart sont égaux au vecteur nul. Soit  $\bar{N} \in M_{(Q-1)m, (Q-1)m}(\mathbb{R})$  la sous-matrice de  $N$  obtenue en supprimant ces lignes et colonnes. Une fois de plus par définition de  $M$ ,  $\bar{N}$  est régulière. Soit  $N^{(-1)} \in M_{Qm, Qm}(\mathbb{R})$  la matrice déduite de  $\bar{N}^{-1}$  en ajoutant à  $\bar{N}^{-1}$  les lignes et les colonnes de zéro mentionnées ci-dessus. Compte tenu de (4.11), par construction,

$$\frac{1}{2} N^{(-1)} \alpha^*. \quad (4.19)$$

Ainsi,

$$\mathbf{J}(\xi^*) = \xi^{*T} M^T M \xi^* = \frac{1}{2} \alpha^{*T} \xi^* - \alpha^{*T} \xi^* = \frac{1}{4} \alpha^{*T} N^{(-1)} \alpha^*.$$

$$— (k_3 = 1) \wedge (p = 2)$$

Soit  $\check{\alpha} = (\sum_{k \neq y_i} \alpha_{ik})_{1 \leq i \leq m}$  et  $\check{\xi} = (\xi_i)_{1 \leq i \leq m}$ . Les vecteurs  $\check{\alpha}^*$  et  $\check{\xi}^*$  sont deduits de  $\check{\alpha}$  et  $\check{\xi}$  en remplaçant  $\alpha_{ik}$  par  $\alpha_{ik}^*$  et  $\xi_i$  par  $\xi_i^*$ . Alors,

$$M \xi^* = \check{M} \check{\xi}^*;$$

ou  $\check{M} = (\check{m}_{ik,j})_{1 \leq i, j \leq m, 1 \leq k \leq Q} \in M_{Qm, Qm}(\mathbb{R})$  est la matrice deduite de  $M$  comme suit

$$\forall (i, j, k) \in \llbracket 1, m \rrbracket \times \llbracket 1, m \rrbracket \times \llbracket 1, Q \rrbracket, \check{m}_{ik,j} = \sum_{l=1}^m \check{m}_{ik,jl}.$$

En outre, (4.13) peut être réécrite comme suit :

$$2\check{M}^T \check{M} \check{\xi}^* = \check{\alpha}^*.$$

Par définition de  $M$ , la matrice  $\check{N}$  egale à  $\check{M}^T \check{M}$  est régulière, de sorte que :

$$\mathbf{J}(\xi^*) = \check{\xi}^{*T} \check{N} \check{\xi}^* - \check{\alpha}^{*T} \check{\xi}^* = \frac{1}{2} \check{\alpha}^{*T} \check{\xi}^* - \check{\alpha}^{*T} \check{\xi}^* = -\frac{1}{4} \check{\alpha}^{*T} \check{N}^{-1} \check{\alpha}^*.$$

En regroupant les choses, nous obtenons l'expression suivante pour la fonction objective du Wolfe dual du Problème 1 :

$$\mathbf{J}_{M-SVM.d}(\alpha) = -\frac{1}{4}\left\{\left(\frac{1}{\lambda}H + (1 - K_3)(p - 1)N^{(-1)}\right)\alpha + K_3(p - 1)\check{\alpha}^T \check{N}^{-1}\check{\alpha}\right\} + K_2\mathbf{I}_{Qm}^T\alpha. \quad (4.20)$$

Dans le cas ou  $p = 1$ , les contraintes d'inégalité sont déduites (4.9) et (4.12), en fonction de la valeur de  $K_3$ . Dans le cas ou  $p = 2$ , nous obtenons simplement.

$$\forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus \{y_i\}, \quad \alpha_{ik} \geq 0.$$

Les contraintes d'égalité sont déduites (4.8). Notez que nous pouvons tirer profit du fait que

$$\begin{aligned} & \sum_{k=1}^Q \sum_{i=1}^m \sum_{l=1}^Q \{K_1\delta y_{i,k} + (1 - K_1)\frac{1}{Q} - \delta_{k,l}\}\alpha_{il} = 0 \\ & \begin{cases} \sum_{k=1}^Q \sum_{i=1}^m \sum_{l=1}^Q \{K_1\delta y_{i,k} + (1 - K_1)\frac{1}{Q} - \delta_{k,l}\}\alpha_{il} = 0 \\ \forall k \in \llbracket 1, Q - 1 \rrbracket, \sum_{i=1}^m \sum_{l=1}^Q \{K_1\delta y_{i,k} + (1 - K_1)\frac{1}{Q} - \delta_{k,l}\}\alpha_{il} = 0 \end{cases} \\ & \implies \sum_{i=1}^m \sum_{l=1}^Q \{K_1\delta y_{i,Q} + (1 - K_1)\frac{1}{Q} - \delta_{Q,l}\}\alpha_{il} = 0 \end{aligned}$$

Pour reduire leur nombre a  $Q - 1$ , alors, le Wolfe Dual du prolème 1 est,

**problem 5** (*Apprentissage d'une M-SVM de marge douce, formulation du dual*)

$$\begin{aligned} & \max_{\alpha} \mathbf{J}_{M-SVM.d}(\alpha) \\ & s.t. \begin{cases} \forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \quad 0 \leq (1 - K_3)(2 - p)\alpha_{ik} \leq (2 - p)m_{ik,ik}. \\ \forall i \in \llbracket 1, m \rrbracket, \quad 0 \leq K_3(2 - p) \sum_{k \neq y_i} \alpha_{ik} \leq (2 - p) \sum_{k \neq y_i} m_{ik,ik}. \\ \forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \quad (p - 1)\alpha_{ik} \geq 0. \\ \forall k \in \llbracket 1, Q - 1 \rrbracket, \sum_{i=1}^m \sum_{l=1}^Q \{K_1\delta y_{i,k} + (1 - K_1)\frac{1}{Q} - \delta_{k,l}\}\alpha_{il} = 0 \end{cases} \end{aligned}$$

Ou  $\mathbf{J}_{M-SVM.d}(\alpha)$  est donnée par l'équation (4.20).

Avec de légères modifications, le calcul ci-dessus peut être adapté pour exprimer le Wolfe Dual du problème d'apprentissage d'une machine a marge dure. Nous obtenons alors :

**problem 6** (*Apprentissage d'une M-SVM de marge dure, formulation du dual*)

$$\max_{\alpha} \left\{ -\frac{1}{4\lambda} \alpha^T H \alpha + K_2 \mathbf{1}_{Q^m}^T \alpha \right\}$$

$$s.t. \begin{cases} \forall i \in \llbracket 1, m \rrbracket, \forall k \in \llbracket 1, Q \rrbracket \setminus y_i, \alpha_{ik} \geq 0. \\ \forall k \in \llbracket 1, Q - 1 \rrbracket, \sum_{i=1}^m \sum_{l=1}^Q \{K_1 \delta_{y_i, k} + (1 - K_1) \frac{1}{Q} - \delta_{k, l}\} \alpha_{il} = 0 \end{cases}$$

Les valeurs de  $b^* = (b^*)_{1 \leq k \leq Q}$  sont déduites des conditions complémentaires de Karush-Kuhn-Tucker (KKT). Dans le cas ou  $K_1 = 0$ , les valeurs des composantes de ce vecteur sont obtenues individuellement, et il suffit de vérifier qu'ils ne satisfont pas  $\mathbf{1}_Q^T b^* = 0$ . Dans le cas  $K_1 = 1$ , les conditions de complémentarité KKT nous fournissent juste des valeurs pour les différences entre les valeurs des composants. Cela signifie que ces dernières valeurs ne sont connues qu'à une constante additive. Ce degré de liberté supplémentaire peut être utilisé pour appliquer la contrainte  $\mathbf{1}_Q^T b^* = 0$ . Ceci termine la discussion sur la raison pour laquelle la contrainte  $\sum_{k=1}^Q h_k = 0$  apparaît dans le problème 1 seulement pour les cas ou  $K_1 = 0$ .

### 4.3.2 Caractérisation des quatre principales M-SVMs

Les M-SVMs directes incluent quatre grands modes à savoir : le mode de WW, le mode de CS, le mode de LLW, et la M-SVM<sup>2</sup> introduite par Geurmeur. Avant l'introduction de la M-SVM<sup>2</sup>, il y avait essentiellement une unique définition unificatrice des trois premières M-SVMs directes (WW, CS, LLW), qui a été introduite indépendamment (avec de légères différences), par plusieurs chercheurs. Selon [50], la définition standard d'une M-SVM directe peut être formulée comme suit :



**Definition 5** (*Definition standard d'une M-SVM directe*) Soit  $\mathcal{X}$  un ensemble non vide (l'espace de description), et  $\mathcal{Y}$  est l'ensemble des catégories  $\in \llbracket 1, Q \rrbracket$  avec  $3 \leq Q < +\infty$ . Soit  $\kappa$  un noyau symétrique semi-défini positif sur  $\mathcal{X}^2$ , et  $\mathcal{H}_{\kappa, Q} = (\mathbf{H}_{\kappa, Q} \oplus \{1\})$ . Pour  $m \in \mathbb{N}^*$ , Soit  $\mathbf{z}_m = ((x_i, y_i))_{1 \leq i \leq m} \in (\mathcal{X} \times \llbracket 1, Q \rrbracket)^m$ . Une M-SVM à Q-catégories est un modèle discriminant à grande marge obtenu par la résolution d'un problème de programmation quadratique convexe de la forme :

**problem 7** (*Apprentissage d'une M-SVM, formulation primale*)

$$\min_{h \in \mathcal{H}_{\kappa, Q}} \left\{ \sum_{i=1}^m l_{M-SVM}(y_i, h(x_i)) + \lambda \sum_{k=1}^Q \|\bar{h}_k\|_{\mathbf{H}_{\kappa}}^2 \right\}$$

$$s.t. \quad \sum_{k=1}^Q h_k = 0$$

ou  $l_{M-SVM}$  est la fonction de perte.

Ainsi, a partir du problème 7, les fonction de perte des trois premiers modèles des M-SVMs directes (WW, CS, LLW) sont comme suit :

$$l_{WW}(y, h(x)) = \sum_{k \neq y} (1 - h_y(x) + h_k(x))_+$$

$$l_{CS}(y, h(x)) = (1 - h_y(x) + \max_{k \neq y} h_k(x))_+$$

$$l_{LLW}(y, h(x)) = \sum_{k \neq y} (h_k(x) + \frac{1}{Q-1})_+$$

Sachant que la fonction  $(\cdot)_+$  retournant son argument si celui-ci est positif, et 0 dans le cas contraire.

Le modèle le plus récent des M-SVMs directes est le modèle M-SVM<sup>2</sup> [17]. Contraire aux trois anciens modèles, sa définition ne peut pas être fondée sur une spécification du problème 7. elle résulte d'une spécification du problème 4. C'est

précisément la M-SVM directe dont  $p = 2$  et  $(k_t)_{1 \leq t \leq 3} = (0, \frac{1}{Q-1}, 0)^T$ , avec la matrice  $M$  étant remplacée par la matrice  $M^{(2)}$  de terme général :

$$m_{ik,jl}^{(2)} = (1 - \delta_{y_i,k}) (1 - \delta_{y_j,l}) \left( \delta_{k,l} + \frac{\sqrt{Q} - 1}{Q - 1} \right) \delta_{i,j},$$

Soit  $N^{(2)} = M^{(2)T} M^{(2)}$ , son terme général est :

$$n_{ik,jl}^{(2)} = (1 - \delta_{y_i,k}) (1 - \delta_{y_j,l}).$$

soit  $I_{Qm}(\mathbf{z}_m)$  la matrice de  $M_{Qm,Qm}(\mathbf{z}_m)$  dont le terme  $m_{ik,jl}$  est respectivement :

$$m_{ik,jl} = \delta_{i,j} \delta_{k,l} (1 - \delta_{y_i,k}).$$

La formulation des quatre modèles de M-SVMs (WW, CS, LLW, M-SVM<sup>2</sup>) comme des instances du modèle générique, correspond à appliquer les valeurs des hyperparamètres représentées dans le tableau 4.1 suivant [50] :

TABLE 4.1 – les hyperparamètres des quatres M-SVMs

M-SVM	$M$	$H$	$p$	$k_1$	$K_2$	$K_3$
WW	$I_{Qm}(\mathbf{d}_m)$	$(\delta_{y_i,y_j} - \delta_{y_i,l} - \delta_{y_j,k} - \delta_{k,l})\kappa(x_i, x_j)$	1	1	1	0
CS	$\frac{1}{Q-1} I_{Qm}(\mathbf{d}_m)$	$(\delta_{y_i,y_j} - \delta_{y_i,l} - \delta_{y_j,k} - \delta_{k,l})\kappa(x_i, x_j)$	1	1	1	1
LLW	$I_{Qm}(\mathbf{d}_m)$	$(\delta_{k,l} - \frac{1}{Q})\kappa(x_i, x_j)$	1	0	$\frac{1}{Q-1}$	0
M-SVM <sup>2</sup>	$M^2$	$(\delta_{k,l} - \frac{1}{Q})\kappa(x_i, x_j - \frac{1}{2C} \delta_{i,j})$	2	0	$\frac{1}{Q-1}$	0

## 4.4 Post-traitement des sorties des M-SVMs

Généralement dans la majorité des applications de reconnaissance, le classifieur ne contribue qu'à une petite partie de la décision finale. Il est donc essentiel que la sortie du classificateur soit une mesure de confiance calibrée, comme la probabilité à posteriori. Cependant, bien que les SVMs standards ne fournissent pas de telles

probabilités, une méthode de post-traitement simple pour caractériser les sorties d'une SVM unique en probabilités postérieures a été proposée par Platt [18] et approuvée par Lin et co-auteur [19].

Dans cette section nous détaillerons la méthodologie adoptée pour post-traiter en terme de probabilités, les différentes sorties des M-SVMs (directes et indirectes) considérées dans cette thèse.

#### 4.4.1 Les M-SVMs indirectes

Dans cette étude nous nous sommes intéressés et nous avons implémenté les deux stratégies de décomposition les plus utilisées à savoir : 1-c-1 et 1-c-tous. La méthode de décomposition basée sur les ECOC n'a pas été implémentée pour la simple raison qu'elle est principalement basée sur les deux techniques (1-c-1 et 1-c-tous).

**L'approche 1-c-tous** La stratégie 1-c-tous est basée sur la construction d'une SVM par classe, qui est formée pour distinguer les échantillons d'une classe des échantillons de toutes les classes restantes. Habituellement, la classification d'un exemple inconnu est faite en fonction de la sortie maximale entre toutes les SVMs. L'approche la plus intuitive pour estimer la probabilité a posteriori avec la technique 1-c-tous consiste à réinterpréter séparément les résultats de chaque SVM en terme de probabilités en utilisant la méthode proposée par Platt [18], qui utilise une sigmoïde supplémentaire :

$$\forall k \in \llbracket 1, Q \rrbracket, P(k | h_k(x)) = \frac{1}{1 + \exp(A_k h_k(x) + B_k)} \quad (4.21)$$

Où les valeurs de  $A_k$  et  $B_k$  sont déterminées par un apprentissage appliquant le principe de maximum de vraisemblance.

$$-\sum_{i=1}^m t_i \log(p_i) + (1 - t_i) \log(1 - p_i), \quad (4.22)$$

ou  $p_i$  représente la sortie de la sigmoïde et  $t_i$  est la probabilité ciblée. Pour résoudre ce problème d'optimisation, Platt [18] propose d'utiliser l'algorithme de Levenberg Marquardt. Cependant, Lin et co-auteur [19] ont montré qu'il y avait quelques problèmes avec cet algorithme et ils ont proposé d'utiliser un autre algorithme de minimisation basé sur la méthode de Newton avec la recherche de ligne de retour en arrière. (Note : pour les lecteurs qui veulent plus de détails sur l'algorithme de Newton et la méthode Levenberg Marquardt, ils peuvent se référer aux bibliographies suivantes [19])

Cependant, rien ne garantit que :

$$\sum_{k=1}^Q P(k | h_k(x)) = 1. \quad (4.23)$$

Pour cette raison, il semble préférable de normaliser les probabilités comme suit [20] :

$$\forall k \in \llbracket 1, Q \rrbracket, \quad P(k | x) = \frac{P(k | h_k(x))}{\sum_{l=1}^Q P(l | h_l(x))} \quad (4.24)$$

**L'approche 1-c-1** La stratégie 1-c-1 consiste à construire une SVM pour chaque paire de classes. Ainsi, pour un problème avec  $Q$  classes,  $C_Q^2$  SVMs sont formées pour distinguer les échantillons d'une classe des échantillons d'une autre classe. Habituellement, le classement d'un exemple inconnu se fait selon le vote maximal, où chaque SVM vote pour une classe. Cependant, l'estimation des probabilités à posteriori avec la technique 1-c-1 pose problème. En effet, après avoir estimé la sortie de chaque SVM en terme de probabilités en utilisant la fonction sigmoïde (voir équation 4.21), la tâche consiste à exprimer les probabilités a posteriori globales  $P(k | x)$  comme fonctions des probabilités postérieures locales  $P(k | h_{k,l}(x))$ , où  $h_{k,l}(x)$  dénote la sortie de la SVM formée pour distinguer la classe  $k$  de la classe  $l$ . Différentes méthodes ont été proposées dans la littérature [56, 57, 58], dans cette étude nous avons choisi d'utiliser la méthode proposée par Price et co-auteur [56].

Ils ont considéré que pour chaque classe  $k$  :

$$\forall k \in \llbracket 1, Q \rrbracket, \quad \sum_{l=1, l \neq k}^Q P(k, l | x) - (Q - 2)P(k | x) = 1, \quad (4.25)$$

ou  $k, l$  dénote l'union des classes  $k$  et  $l$ , alors en utilisant :

$$P(k | h_{k,l}(x)) \simeq \frac{P(k | x)}{P(k, l | x)}, \quad (4.26)$$

il est possible de dériver l'expression ci-dessous :

$$P(k | x) = \frac{1}{\sum_{l=1, l \neq k}^Q \frac{1}{P(k | h_{k,l}(x))} - (Q - 2)}. \quad (4.27)$$

Cependant, comme rien ne garantit que la somme de toutes les probabilités  $P(k | x) = 1$ , il faut diviser chaque estimation comme suit :

$$\forall k \in \llbracket 1, Q \rrbracket, \quad P(k | x) = \frac{P(k | x)}{\sum_{l=1}^Q P(l | x)} \quad (4.28)$$

#### 4.4.2 les M-SVMs directes

Selon [21], pour obtenir des M-SVMs directes probabilistes, il convient d'étendre la solution bi-classes de Platt [18] au cas multi catégories, conformément à la fonction softmax suivante :

$$\forall k \in \llbracket 1, Q \rrbracket, \quad P(k | x) = \frac{\exp(h_k(x))}{\sum_{l=1}^Q \exp(h_l(x))} \quad (4.29)$$

*Résumé :*

*Les SVMs constituent une approche de classification et de regression très puissante. En effet, les SVMs ont fait leurs preuves sur plusieurs bases de données standards, due à : une bonne généralisation, une insensibilité au sur-apprentissage, et une robustesse vis à vis de la dimensionnalité.*

*Les SVMs sont principalement conçus pour les problèmes de classification binaires, et généralement leur extension pour des problèmes multiclassés (M-SVMs) implique seulement des SVMs bi-classes (l'approche indirecte), qui traitent le problème d'apprentissage partiellement. Ces dernières années, la communauté scientifique se concentre sur quatre grands modèles des M-SVMs qui considèrent toutes les classes simultanément (l'approche directe), et qui traitent le problème d'apprentissage des SVMs comme étant un seul problème d'optimisation : le modèle de WW, le modèle de CS, le modèle de LLW, et la M-SVM<sup>2</sup>.*

*D'autre part, les probabilités a posteriori constituent une mesure de confiance très importante pour les sorties d'un classifieur. En revanche, même si les SVMs standards ne fournissent pas une telle mesure, plusieurs post-traitements des sorties des SVMs ont été proposés pour leurs permettre d'estimer des probabilités à posteriori. Dans la suite de la thèse, nous implémenterons les M-SVMs probabilistes directes et indirectes pour la détermination des tâches mentales et la détection des pathologies cardiovasculaires. Ainsi, nous introduirons l'utilisations des quatre modèles de l'approche directe dans les deux problématiques, et nous deduirons par la suite leur avantage par rapport aux M-SVMs indirectes.*

## Chapitre 5

# Application 1 : Hybridation des SVMs multi-classes et les cartes de kohonen pour l'interprétation de la pensée.

Les Interfaces Cerveau-Machine (BCIs) offrent une opportunité de communication entre le cerveau et l'environnement autour d'une personne ayant un handicap moteur sévère [6, 5]. Il existe plusieurs techniques (invasives et non invasives) d'enregistrement de l'activité cérébrale, et avec lesquelles nous pouvons déterminer les tâches mentales : ElectroEncephaloGramme (EEG), Imagerie par Résonance Magnétique (IRM), Imagerie par Résonance Magnétique fonctionnelle (fMRI), Magneto EncéphaloGraphie (MEG), Spectroscopie Proche InfraRouge (NIRS), Tomoscintigraphie par Emission de Positons (PET), et Signal Optique Lié à l'Événement (EROS). Parmi ces techniques, l'EEG est le plus utilisé en raison de sa capacité à enregistrer les signaux cérébraux de façon non invasive et son faible coût.

Cependant, la mise en oeuvre des BCIs nécessite un bon système de traitement du signal, dont les performances dépendent principalement de la technique utilisée pour sélectionner les meilleures caractéristiques, et de la méthode de classification appliquée pour effectuer la discrimination entre les différentes catégories. Dans ce chapitre, nous proposerons une nouvelle structure hybride basée sur deux étages, avec apprentissage supervisé et non supervisé. Cette approche est validée sur la base de données EEG de Keirn et Aunon [22], qui consiste sur des enregistrements EEGs de 7 sujets effectuant cinq tâches mentales différentes, à savoir : Relaxation (RelT), Multiplication (MulT), Lettre (LetT), Rotation (RotT), et Comptage (CompT). Le premier étage utilise les Cartes Auto Organisatrices de Kohonen (SOMs), il permet de regrouper les fonctions non pertinentes redondantes et de sélectionner les meilleurs descripteurs. Le deuxième étage utilise les Machines à Vecteurs de Support Multiclasses (M-SVMs) probabilistes pour la classification finale ( voir figure 5.1). Les deux méthodes des M-SVMs (directe et indirecte) sont considérées dans cette étude, dans le but de comparer leurs performances en généralisation et de déduire l'apport des M-SVMs directes dans la problématique de la classification des signaux EEGs.

Dans le présent chapitre, nous donnerons en premier lieu, une description de la base de données utilisée. Puis nous effectuerons un état de l'art, dans lequel nous exposerons certains modèles proposés dans la littérature pour l'interprétation des tâches mentales en utilisant des signaux EEGs. En troisième lieu, nous expliquerons la méthodologie adoptée pour construire le premier ensemble de descripteurs. Ensuite, nous présenterons les détails de notre structure hybride proposée. Enfin, nous exposerons et nous discuterons les résultats obtenus.

## 5.1 Description de la base de données

La base de données EEG utilisée dans cette étude a été assemblée par Keirn [22], et distribuée en ligne par Anderson [59]. Elle se compose de 325 enregistrements EEGs de sept sujets (âgés de 21 à 48 ans) effectuant cinq tâches mentales distinctes,



CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

---

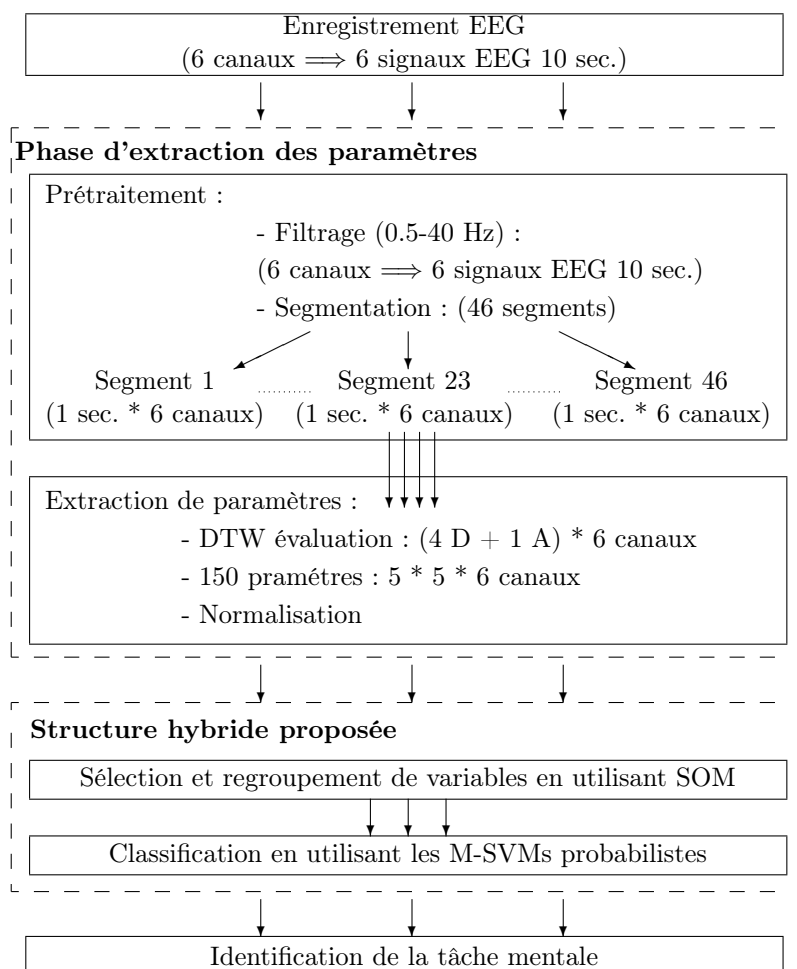


FIGURE 5.1 – Diagramme général de la méthode proposée pour un enregistrement EEG.

à savoir :

- La Tâche de Relaxation (RelT), les sujets ont été invités à se détendre et ne penser à rien en particulier.
- La Tâche de Multiplication (MulT), les sujets avaient pour consigne de résoudre un problème non trivial de multiplication (eg.  $28 \times 62$ ). Les sujets ne doivent pas vocaliser, ni faire des mouvements quand ils résolvent le problème.

- La Tâche Lettre (LetT), les sujets avaient pour instruction de composer mentalement une lettre à une autre personne sans vocaliser.
- La Tâche Rotation (RotT), les sujets ont eu une durée de 30 sec. pour voir un objet 3D en rotation. Après avoir retiré l'objet, ils ont été invités à imaginer que l'objet est en rotation autour d'un axe.
- La Tâche de Comptage (ComT), il a été demandé à tous les sujets d'imaginer un tableau noir et de visualiser les numéros en cours d'introduction séquentielle dans le tableau.

Un électro-cap a été utilisé pour obtenir chaque enregistrement EEG à partir de six canaux : C3, C4, P3, P4, O1 and O2, selon le système 10-20. Les signaux obtenus pour chaque tâche ont une durée d'enregistrement de 10 sec., et une fréquence d'échantillonnage de 250 Hz.

Nous avons utilisé tous les sujets de la base de données, sauf le sujet 4 en raison de l'invalidation des signaux dans plusieurs essais (Trials) comme rapporté par Liang et al. [60]. Chaque sujet a participé à (1 session, 2 sessions, ou 3 sessions) de cinq trials par session. Chaque trial contient cinq enregistrements, et chaque enregistrement représente l'une des cinq tâches mentales.

Nous notons l'existence d'un septième canal nommé ElectroOculoGraphie(EOG). Il aide à localiser les clignements des yeux, en effectuant un traitement sur le signal qu'il enregistre. Dans [61, 60, 11], les auteurs ont proposé des schémas de détection pour localiser et pour effectuer une correction d'artefact oculaire. Cependant, ce canal n'a pas été utilisé dans ce travail, parce que les effets de clignotement des yeux est faible dans la région centrale et occipitale où se trouvaient les six autres canaux [7].

## 5.2 Etat de l'art

On peut trouver de nombreux modèles proposés dans la littérature pour l'interprétation des tâches mentales utilisant des signaux EEGs. Ci-après, nous avons sélectionné ceux dont la base de données est similaire à la notre :

Keirn et Aunon [61] ont proposés une méthode basée sur des caractéristiques AutoRégressif (AR) et un classifieur Bayésien Quadratique (BQ). Ils ont démontré la nécessité d'utiliser les bandes ( $\delta$ ,  $\theta$ ,  $\alpha$ , and  $\beta$ ) et les caractéristiques AR pour améliorer les résultats. D'abord, les signaux EEGs ont été divisés en des segments de 1/4 sec., et 36 AR caractéristiques ont été estimées pour chaque segment. Puis, un classifieur BQ a été proposé pour classer les différentes combinaisons paires de tâches pour cinq sujets. Une précision moyenne de 84,6 % a été obtenue sur l'ensemble des tâches paires et sur l'ensemble des cinq sujets.

Palaniappan [62] a subdivisé les signaux EEGs en 20 segments d'une longueur de 1/2 sec.. Ensuite, il a calculé comme caractéristiques de chaque segment, le Spectre de Puissance (PS) et les Ratios d'ASymétrie (ASR), relative au sub bandes  $\delta$ ,  $\theta$ ,  $\beta$ , and gamma ( $\gamma$ ). Les composants obtenus ont été présentés comme des entrées d'un classifieur de type Réseaux de Neurones d'Elman (ENN). Il conclut de cette étude, que les performances ont été améliorées par l'ajout de la bande  $\gamma$ . Aussi, un taux moyen de : Sujet 1 : 79.8%, Sujet 2 : 69.1%, Sujet 3 : 70.2%, Sujet 4 : 80.3% a été réalisé sur l'ensemble des tâches paires.

Dans [60], Liang *et al.*, les auteurs ont subdivisé les signaux EEGs en des segments de 1/2 sec. avec un chevauchement de 1/4 sec. entre les segments adjacents, et compter pour chaque segment les mêmes AR caractéristiques utilisées par Keirn et Aunon [61]. Ensuite, une Machine à Apprentissage Extrême (ELM) a été proposée comme une méthode de classification de cinq tâches mentales. En plus, les auteurs ont réalisé un moyennage sur les sorties du classifieur à travers 20 segments consécutifs. Les résultats ont indiqué une amélioration significative du taux de test, et des taux moyens entre 57.37% et 86.70% ont été obtenus.

Zhiwie et Minfen [11] ont proposé une méthode utilisant l'Entropie (energie) de la Transformée par Ondelettes Discrètes (DTW) et les SVMs. Ils ont divisé les signaux EEGs en des segments de 1 sec. qui se chevauchent par une durée de 0.8 sec., puis ils ont calculé l'énergie de chacune des sub bandes  $\delta$ ,  $\theta$ ,  $\alpha$ , and  $\beta$ . Ensuite, un classifieur de type SVM a été formé pour discriminer entre deux classes, trois classes, quatre classes, et cinq classes. Le système proposé atteint pour 5-classes un taux de classification moyen de : (76.3%) pour le sujet 1 et (68.5%) pour le sujet 2. Il a été montré à partir de l'étude, que DTW peut être une bonne méthode pour l'extraction de caractéristiques et que la SVM peut être une bonne méthode pour discriminer entre les différentes tâches.

Diez et al. [63] ont appliqué leur méthode d'extraction des caractéristiques proposées, sur des segments EEGs d'une durée de 1 sec.. Cette méthode était basée sur la décomposition modale empirique (EMD) et l'estimation de six caractéristiques : moyenne quadratique (RMS), la variance, l'entropie de Shannon, Valeur de la complexité de Lempel-Ziv, la fréquence moyenne, et la fréquence maximale. Ensuite, la réduction de la dimensionnalité a été réalisée sur la base du paramètre Wilks Lambda (WL). Deux classifieurs ont été utilisées pour la discrimination entre les tâches : les réseaux de neurones artificiels (ANN) et l'analyse discriminante linéaire (LDA). Avec ces méthodes, un taux de 91,17% a été obtenu sur l'ensemble des tâches paires et en utilisant LDA classifieur. Ces résultats concluent l'efficacité des EMD pour traiter des signaux non stationnaires et non linéaires tel que l'EEG.

Tolic and Jovic [7] ont présenté une méthode basée sur la DWT et ANN pour classer, deux, quatre, et cinq tâches mentales relatives au sujet 1 et sujet 2. Ils ont décomposé chaque signal EEG en des segments de durée 1/2 sec., et ils ont appliqué la DWT pour extraire les différentes bandes de fréquences. Les énergies des différentes bandes obtenues, ont été ensuite calculées comme caractéristiques. Les vecteurs des descripteurs résultant ont été présentés comme entrée d'un ANN, qui a atteint pour cinq classes un taux de 90,75 % associé au sujet 1.

Dans [64] les auteurs, ont tous d'abord subdivisé les signaux ECG en segments

de 1 sec. qui se chevauchent par une durée de 0.8 sec.. Puis, ils ont appliqué la Ttransformée de stockwell (ST) pour obtenir cinq bandes de fréquences, ces bandes sont présentées sous forme de matrice de dimension  $n * m$ , sachant que les lignes représentent les fréquences et les colonnes représentent le temps. La dérivation standard est par la suite calculée, puis la moyenne des racines carrées de la dérivation standard (MSR) a été utilisée pour obtenir 30 descripteurs (5 descripteurs \* 6 channels). Les auteurs ont implémenté quatre classifieurs nommés : LDA, Analyse Discriminante Quadratic (QDA), k plus proches voisins (kNN), et SVM, pour la discrimination entre 2 classes, 3 classes, 4 classes, et 5 classes. Les auteurs ont souligné que les meilleurs résultats de classification de 5 classes ont été obtenus par le discriminateur LDA avec un taux globale de classification de 93.71 %.

Les résultats obtenus de différentes études résumées dans cette section sont donnés dans le tableau 5.10.

### 5.3 Méthode proposée pour l'extraction des paramètres des signaux EEGs

Dans cette section, nous décrivons la méthode d'extraction des paramètres qui comprend l'étape du prétraitement et l'étapes d'estimation des descripteurs. Dans l'étape de prétraitement, on élimine les artefacts et on subdivise chaque signal EEG en 46 segments d'une durée de 1 sec. chacun. Ensuite, dans l'étape d'estimation des paramètres, on utilise le DWT pour extraire de chaque segment, cinq principales sous bandes qui caractérisent les signaux EEGs :  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$ , puis on calcule pour chaque sous-bande et chaque canal EEG cinq descriptions, pour fournir un vecteur de caractéristiques de 150 composants.

### 5.3.1 Prétraitement

Nous appliquons d'abord en utilisant une fenêtre de Hamming, un filtre passe bande à Réponse Impulsionnelle Finie (FIR) (de type 1 et d'ordre 49) avec une fréquence de coupure entre  $[0.5, 40]$  Hz, pour inclure toute les cinq sous-bandes et supprimer les artefacts de sources internes et externes (les mouvements musculaires, l'interférence du secteur du réseau électrique, les mouvements d'électrodes, etc.). Ensuite, chaque signal EEG est divisé en 46 segments représentant le contenu d'une fenêtre glissante de 1 sec. avec un chevauchement de 0,8 sec. entre les segments adjacents [11, 64]. Ainsi, le premier segment correspond à une fenêtre située entre  $[1, 250]$  échantillons, le segment suivant entre  $[51, 300]$  échantillons, etc. figure 5.2 illustre les résultats obtenus après filtrage et segmentation.

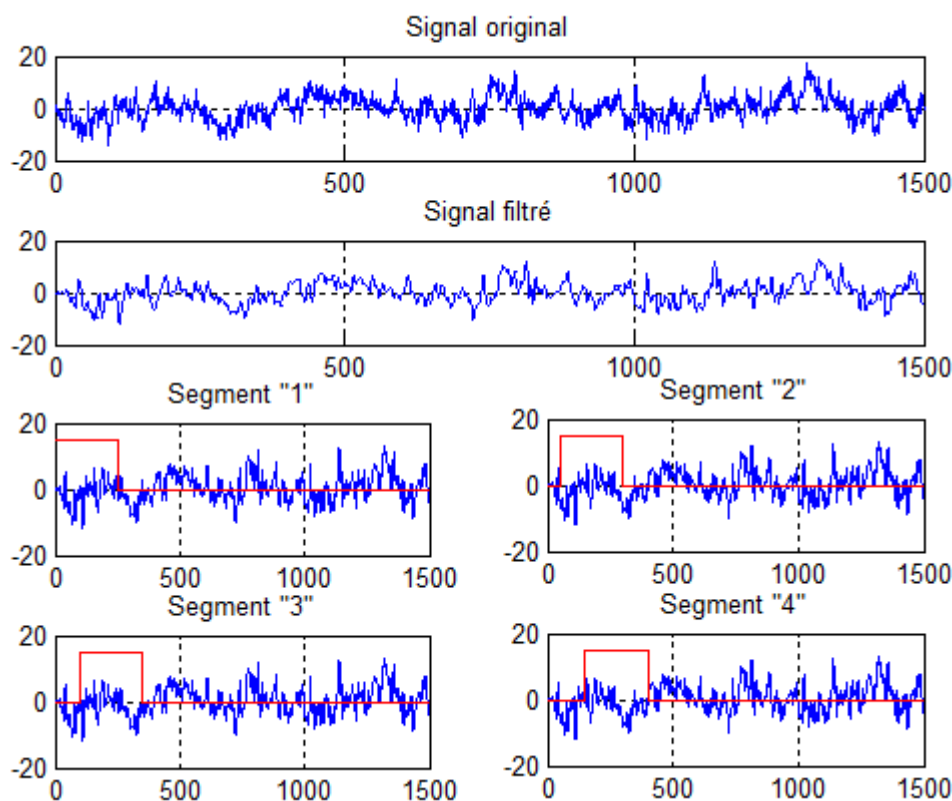


FIGURE 5.2 – Resultat du filtre passe bande et de la segmentation.

### 5.3.2 Estimation des paramètres

Habituellement, l'estimation des paramètres d'un signal EEG ce fait en deux étapes. La première étapes consiste à extraire de chaque signal EEG les cinq principales sous bandes :  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$ . Tandis que la deuxième étape consiste à exprimer les cinq sous bandes à l'aide de descripteurs.

On peut trouver de nombreuses approches intéressantes proposées dans la littérature pour la détermination des ondes ( $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$ ) à partir des signaux EEGs. Elles comprennent : DWT [7, 8, 9, 65], EMD [63], Transformée de Fourier rapide (FFT) [66], La transformée de Fourier à Court Terme (STFT) [9], AR [61, 60], et Transformée de Fourier Discrète (DFT) [67]. Parmi les approches mentionnées, la DWT a été utilisée avec succès dans l'extraction de caractéristiques des signaux EEGs, car elle fournit une approche mathématique rigoureuse qui lui donne la capacité de représenter les signaux non-stationnaires dans le domaine temporel et dans le domaine fréquentiel. De plus, les coefficients à chaque étage de la DWT sont presque stationnaires et de faible auto-corrélation, ce qui offre de meilleures caractéristiques des signaux. (Pour les lecteurs qui veulent avoir plus de détails sur la DWT, ils peuvent se référer à ces bibliographies [68, 69].)

Chaque étage de la DWT se compose de deux filtres numériques : passe-haut et sa version miroir passe-bas, et fournit respectivement le détail (D) et l'approximation (A). L'approximation à l'étage  $s$  contribue à calculer le détail et l'approximation de l'étage  $s + 1$ , simplement, on fait passer ces coefficients à travers des filtres passe-bas et passe-haut.

Dans cette étude, nous utilisons 5 étages de décomposition et l'ondelette de Daubechies 4 (db4), pour subdiviser chaque signal en cinq bandes de fréquences différentes (voir figure. 5.3) : A5 ( $\delta$  : 0.5-3.9 Hz), D5 ( $\theta$  : 3.9-7.8 Hz), D4 ( $\alpha$  : 7.8-15.3 Hz), D3 ( $\beta$  : 15.3-31.5 Hz) et D2 ( $\gamma$  : 31.5-40 Hz) selon les deux equations suivantes :

— Approximation

$$\left[ 0, \frac{F_s}{2^{(n+1)}} \right], \quad (5.1)$$

— détail

$$\left[ \frac{F_s}{2^{(n+1)}}, \frac{F_s}{2^n} \right]. \quad (5.2)$$

Ou  $F_s$  représente la frequence d'échantillonnage 250 Hz, et  $n$  est le nombre de niveau de décomposition egale a 5.

Pour ce qui est de l'extraction des caracteristiques des sous bandes ( $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$ ), on peut trouver une variété de caractéristiques. Dans cette étude nous extrayons des sous-bandes (D2, D3, D4, D5 et A5), cinq caractéristiques statistiques ci dessous et qui representent les descripteurs du signal :

— L'énergie du signal : L'énergie d'un signal est calculé comme suit :

$$E_x = \frac{1}{N} \sum_{n=1}^N (x(n)^2). \quad (5.3)$$

Ou,  $x(n)$  est le signal, et  $N$  est le nombre d'échantillons.

— La variance du signal : La variance d'un signal est déterminée par l'équation mathématique suivante :

$$\sigma_x^2 = \frac{1}{N} \sum_{n=1}^N (x(n) - \bar{x})^2. \quad (5.4)$$

Ou,  $x(n)$  est le signal,  $\bar{x}$  est la moyenne des échantillons du signal, et  $N$  est le nombre d'échantillons.



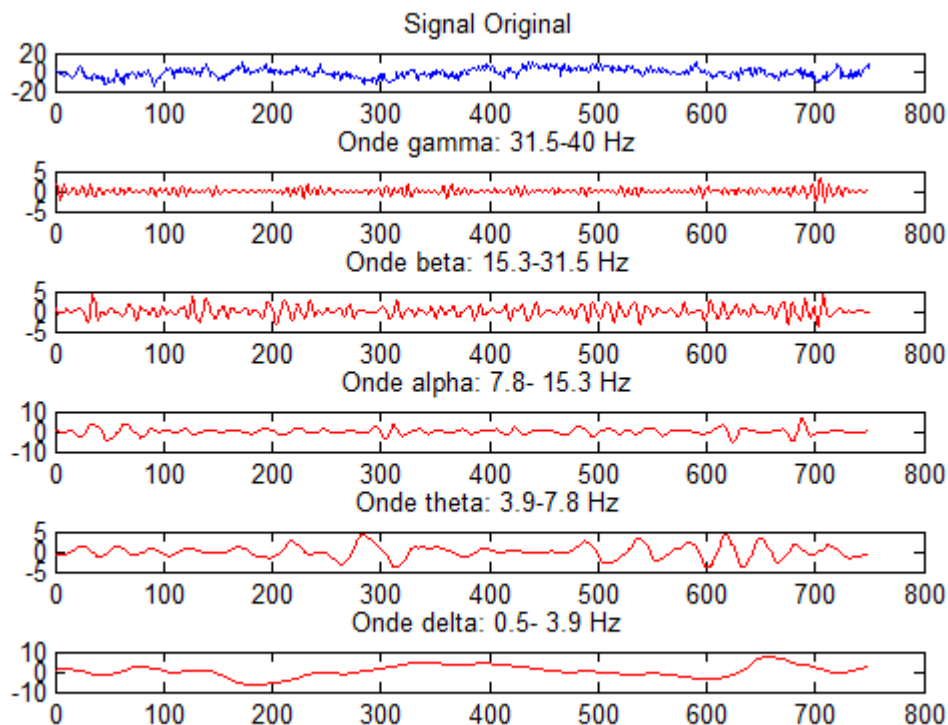


FIGURE 5.3 – Les coefficients des détails (D2, D3, D4 and D5) et de la cinquième approximation A5 pour un segment donné.

- La variance de la fonction d'auto-correlation d'un signal : La fonction d'autocorrélation représente une métrique pour mesurer la similitude entre un signal  $x(i)$  et sa version décalée dans le temps. Sa variance est calculée en fonction de l'équation suivante :

$$\sigma(R_{xx}(l)) = \sigma\left(\sum_{n=i}^{N-|k|-l} [x(n) - x(n-l)]^2\right). \quad (5.5)$$

Ou,  $l$  est l'indice de décalage temporel.  $i = l$ ,  $k = 0$  pour  $l \geq 0$ , et  $i = 0$ ,  $k = 1$  pour  $l < 0$ .

- Min ( $x(n)$ ) : représente la plus petite amplitude.
- Max ( $x(n)$ ) : représente la plus grande amplitude.

Comme mentionné dans la section 4.1, pour chaque tâche, l'activité électrique du cerveau est réfléchiée par six électrodes C3, C4, P3, P4, O1 and O2. Par conséquent, nos vecteurs caractéristiques seront composés de 150 coefficients (6 canaux \* 5 bandes de fréquences \* 5 paramètres). D'autre part, et dans le but de normaliser tous les descripteurs au même niveau  $[-1, +1]$ , un processus de normalisation est appliqué comme suit :

$$\tilde{x}_{ij} = \text{tansig} \left( \frac{(x_{ij} - \bar{x}_j)}{\sigma_{x_j}} \right). \quad (5.6)$$

Où  $\bar{x}_j$  et  $\sigma_{x_j}$  sont respectivement la moyenne et l'écart-type du  $j^{eme}$  paramètre des vecteurs de caractéristiques, à travers chaque sujet et chaque session. *tansig* est une fonction de transfert sigmoïde tangente hyperbolique.

## 5.4 Structure hybride proposée

Nous rapportons dans cette section, la méthodologie adoptée pour construire notre système hybride, qui est basé sur la carte SOM et les M-SVMs probabilistes [65] (Figure. 5.4).

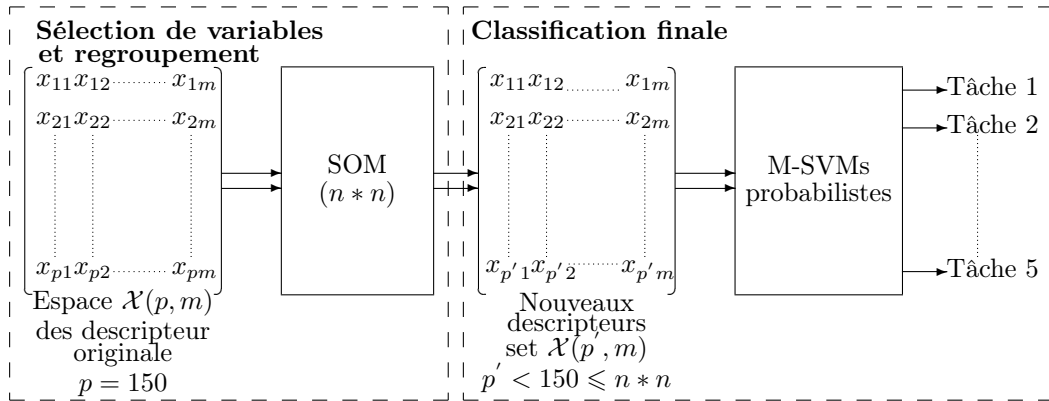


FIGURE 5.4 – Structure hybride proposée.

### 5.4.1 Sélection et regroupement des paramètres en utilisant une SOM

En général, une observation est caractérisée par un ensemble de descripteurs qui ne sont pas tous informatifs. En effet, certains d'entre eux peuvent être insignifiants, corrélés ou non pertinents pour la tâche de classification. Par conséquent, la sélection des paramètres pertinents présente un intérêt majeur qui permet de réduire le volume d'informations à traiter et par conséquent à réduire la complexité des algorithmes de classification et également d'améliorer leur performance de généralisation. Le processus de sélection de variables est caractérisé par les algorithmes de sélection et le critère d'évaluation.

- Les algorithmes de sélection peuvent être classés en deux catégories, les algorithmes myopes, qui estiment la pertinence d'un descripteur sans tenir compte des autres paramètres, et les algorithmes contextuels, qui prennent en compte les interactions (corrélations) entre tous les descripteurs.
- Le critère d'évaluation permet d'évaluer l'aptitude du sous-ensemble sélectionné pour discriminer entre les classes. Plusieurs méthodes sont associées à ce critère et les plus utilisées sont Filter et Wrapper méthodes. La première approche fonctionne indépendamment de l'algorithme de classification (la sélection est effectuée avant l'apprentissage). Cependant, la seconde approche utilise une méthode de classification pour tester la pertinence des variables sélectionnées.

Dans ce contexte, certaines approches ont été proposées pour sélectionner les paramètres pertinents des signaux EEGs. Ils comprennent : les Algorithmes Génétiques (GA) [70], l'Analyse en composantes principales (PCA) [71], et la mesure de WL [63], etc.

Dans cette étude, nous proposons une nouvelle alternative de sélection des paramètres en utilisant l'algorithme SOM. Ce choix repose principalement sur le fait que l'algorithme SOM a la particularité de regrouper des ensembles de grande dimension d'une manière non supervisée. Cette approche peut être consi-

dérée comme contextuelle car elle prend en compte les interactions (corrélations) entre tous les paramètres, et Wrapper parce que nous utilisons un classificateur de type Perceptron Multi Couche (MLP) pour évaluer à chaque fois le sous-ensemble de paramètres sélectionnés. (Pour les lecteurs qui veulent avoir plus de détails sur les SOMs et les MLPs, ils peuvent se référer à ces bibliographies [72, 73, 74].)

**Apprentissage non supervisé d'une SOM :** Le système simulé est une SOM bidimensionnelle avec  $n * n$  neurones artificiels :

- chaque vecteur  $X1$  représente la ligne de l'espace de description  $\mathcal{X}(p, m)$  (avec  $m$  le nombre d'échantillons de l'ensemble d'apprentissage, et  $p$  le nombre de paramètres = 150) et il est relié à chaque neurone de la carte.
- Les connexions externes du neurone  $i \in [1, n * n]$  avec le vecteur d'entrée  $X1$  sont matérialisées par un vecteur de poids synaptiques  $V_i = [v_{i1}, v_{i2}, \dots, v_{im}]$ .
- Aussi, chaque neurone de la carte est lié à tous les autres neurones. Les poids des connexions internes  $W_{ir}$  du neurone  $i$  avec ses voisins  $r(t)$  sont donnés par la fonction appelée "chapeau mexicain".

Au cours de la phase d'apprentissage, le réseau s'adapte à différents vecteurs d'entrée en adaptant les poids des connexions externes de la carte comme suit :

- En premier lieu, la distance euclidienne entre le vecteur d'entrée et chaque vecteur de poids est calculé en fonction de l'équation suivante :

$$d_i(t) = \|X1(t) - V_i(t)\| \quad (5.7)$$

- Ensuite, le neurone  $c$  qui donne la distance minimale par rapport au vecteur d'entrée est sélectionné en tant que gagnant et est caractérisé par la même distance :

$$d_c = \min(d_i(t)) \quad , i \in [1, n * n] \quad (5.8)$$

- Enfin, les poids du neurone gagnant et ses voisins  $r(t)$  sont mis à jour comme suit :

$$V_c(t + 1) = V_c(t)\alpha(t) \|X1(t) - V_i(t)\| \quad (5.9)$$

$$V_i(t + 1) = V_i(t)\alpha(t) \|X1(t) - V_i(t)\|_{i \in r(t)} \quad (5.10)$$

Avec  $\alpha(t)$  le taux d'apprentissage utilisé pour moduler le nouveau poids ;  $\alpha(t)$  et  $r(t)$  diminuent avec  $t$ .

A la fin du processus de formation, on obtient une carte qui s'est auto-organisée en  $(n * n)$  neurones. Chaque neurone représente un ou plusieurs composants. Les paramètres les plus appropriés (qui sont indépendantes de l'autre) sont représentés individuellement par un seul neurone. Cependant, les paramètres corrélés sont regroupés dans un même neurone. Ainsi, nous pouvons extraire les paramètres qui constituent chaque neurone de la carte obtenue. En outre, il existe des neurones qui ne contiennent pas de descripteurs (les neurones morts), ces neurones sont ignorés et éliminés de la carte. La carte résultante représentera alors le nouvel ensemble d'apprentissage (Figure 5.5).

Les paramètres de chaque nouvel exemple sont tout d'abord regroupés dans des combinaisons similaires à celles de la carte obtenue dans l'étape d'apprentissage. Ensuite, le nouveau vecteur réduit est calculé en faisant la moyenne des valeurs des paramètres de chaque neurone. Par conséquent, on obtient un vecteur  $\mathcal{X}(p', m)$  avec  $p' < 150 \leq n * n$ .

### 5.4.2 Classification par les M-SVMs probabilistes

De nombreux modèles ont été mis en oeuvre dans le cadre de la classification des signaux EEGs et les plus courants sont : ANN [60, 7, 63, 75, 76], SVM [64, 60, 9], BQ [61], LDA [64], kNN [64, 66] et ELM [60]. Parmi les classifieurs proposés, ceux qui ont donné les meilleurs résultats et les plus utilisés sont les SVMs. En effet, ils offrent une bonne généralisation, d'excellentes performances à la malédiction de la dimensionnalité, et l'insensibilité au surentraînement. Les SVMs ont été initialement conçus pour des problèmes de classification binaires et leur extension aux cas

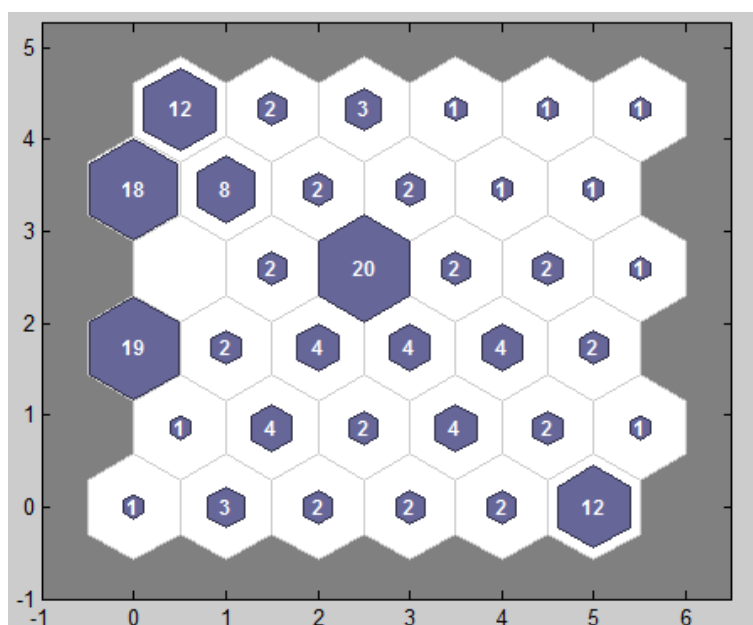


FIGURE 5.5 – Résultat du clustering et la sélection des variables en utilisant la carte SOM.

multi classes (M-SVM) inclut deux approches : la méthode directe et la méthode indirecte. Cependant, à notre connaissance, tous les systèmes proposés sont basés sur des M-SVMs impliquant des méthodes de décomposition, et personne n'a pensé à l'utilisation des M-SVMs directes dans la problématique de traitement des signaux EEGs. Dans cette étude nous introduisons l'utilisation des M-SVMs directes pour la classification des signaux EEGs. Aussi, deux méthodes de décomposition impliquant des SVMs binaires sont également mises en œuvre pour comparer leurs performances avec les M-SVMs directes : un-contre-un (1-c-1) et un-contre-tous (1-c-tous).

Dans nos expériences, nous avons utilisé le package des M-SVMs directes (M-SVMpack) [77], qui nous a permis de mettre en œuvre tous les différents modèles : le modèle de Weston et Watkins (WW), le modèle de Crammer et Singer (CS), le modèle de Lee et al. (LLW), et la M-SVM à coût quadratique (M-SVM<sup>2</sup>). Aussi, pour construire des classifieurs SVMs binaires, nous utilisons le même package et

nous ajustons les paramètres du modèle CS (le choix du modèle de CS est justifié du fait qu'il a donné les meilleurs résultats dans le cas des M-SVMs directes : voir section résultats).

## 5.5 Résultats

Comme mentionné dans la section 4.1, nous avons utilisé tous les sujets de la base de données à l'exception du sujet 4. Chaque sujet a été évalué séparément, session par session, car l'état mental et le raisonnement des personnes exécutant la même tâche peuvent être différents.

Afin d'illustrer les résultats, cinq procédures de validation croisée ont été réalisées à l'aide de cinq trials distincts. Nous avons, effectué l'apprentissage sur trois trials (15 tâches \* 46 segments = 690 segments), effectué la validation sur le quatrième trial (5 tâches \* 46 segments = 230 segments), et effectué le test sur le cinquième trial (5 tâches \* 46 segments = 230 segments).

### 5.5.1 Sélection des paramètres

En général, pour trouver les meilleurs hyperparamètres pour toute méthode, il est recommandé d'effectuer la sélection du modèle. Cela se fait sur les données de validation. Ensuite, le meilleur ensemble des hyperparamètres obtenus est utilisé pour construire le modèle pour les essais futurs.

Pour le modèle SOM, le principal paramètre qui doit être optimisé est le paramètre lié à la taille de la carte. Ainsi, pour trouver la meilleure carte (en utilisant l'ensemble de l'apprentissage), nous avons procédé comme suit :

- En premier lieu, cinq ( $n * n$ ) topologies ont été testées (avec  $n \in [4, 8]$ ) et chaque topologie a été répétée vingt fois, pour un total de 100 époques pour chaque ensemble de validations.
- Ensuite, pour chaque époque, nous modifions l'ensemble de validations en fonction de la carte obtenue (l'ensemble d'apprentissage résultant).

- Puis, un réseau MLP est utilisé pour évaluer la performance de chaque topologie (méthode Wrapper). Le MLP est formé sur l'ensemble d'apprentissage résultant et testé sur l'ensemble de validation modifié.
- Enfin, la meilleure carte choisie permettra de générer le nouveau ensemble de test. Ainsi, les ensembles de validation et d'apprentissage (de la meilleure carte sélectionnée) ajoutés à l'ensemble de test résultant seront utilisés au deuxième étage.

En ce qui concerne le modèle SVM, les paramètres qui doivent être optimisés sont ceux liés au paramètre de pénalité  $C$  et au noyau  $\gamma$ . Nous notons que trois types de noyaux sont disponibles dans le MSVMpack (Linéaire, Polynomiale, et Gaussien/RBF). Ici, nous avons retenu le noyau RBF qui est le plus utilisé. Ainsi, pour les M-SVMs avec le noyau RBF, deux paramètres sont optimisés (le paramètre  $C$  et le noyau  $\gamma$ ) sur les cinq procédures de validation croisée, comme suit :

- En premier lieu, pour chaque niveau de validation nous testons,  $15 \times 15 = 225$  différentes combinaisons de  $(C, \gamma)$ ,  $\gamma = [2^4, 2^3, 2^2, \dots, 2^{-10}]$  et  $C = [2^{12}, 2^{11}, 2^{10}, \dots, 2^{-2}]$  [55].
- Puis, pour chaque combinaison, la performance de validation est calculée, en effectuant la formation sur l'ensemble d'apprentissage et en effectuant le test sur l'ensemble de validation.
- En fin, dans la phase de prédiction, nous effectuons l'apprentissage sur le même ensemble de formation en utilisant la paire de  $(C, \gamma)$  qui donne le meilleur taux de validation, et nous testons sur l'ensemble de test.

Pour les M-SVMs basés sur l'approche 1-c-1 et 1-c-tous, nous avons considéré pour chaque étage de validation croisée et pour tous les problèmes binaires, la même paire de  $(C, \gamma)$  [55].



### 5.5.2 Évaluation des systèmes de classification implémentés

Les tables 5.1, 5.2, 5.3, 5.4, 5.5, 5.6 exposent les résultats obtenus avec la structure hybride proposée basée sur les SOM et les M-SVMs probabilistes (directes et indirectes) comme suit :

- Le taux de test par tâche, il représente la moyenne des taux obtenus pour chaque tâche au cours des cinq etages de validation croisée.
- Le taux de test par session, il représente la moyenne des taux de test de toutes les tâches mentales par session.
- Le taux de rejet par session, il représente la moyenne des taux de rejets cumulés sur les cinq étages de validation croisée.
- Aussi, pour associer un degré de fiabilité du système proposé, nous avons utilisé la mesure statistique définie ci-dessous [78]. Cette mesure permet de donner un interval de taux de classement correct pour les futurs jeux de tests, avec un taux de confiance  $\alpha\%$ .

$$I(\alpha, N) = \frac{T + \frac{Z_\alpha^2}{2N} \pm Z_\alpha \sqrt{\frac{T(1-T)}{N} + \frac{Z_\alpha^2}{4N^2}}}{1 + \frac{Z_\alpha^2}{N}} \quad (5.11)$$

Ou  $N$  est la taille de l'ensemble de test,  $T$  est le taux de classification correcte, et  $Z_\alpha = 1.96$  for  $\alpha = 95\%$ .

- D'autre part, nous avons post-traité les sorties des M-SVMs afin de générer des estimations de probabilité à posteriori des classes (voir chapitre 4, section 4.4). Ensuite, si la probabilité associée à une classe est supérieure à 0,75, la tâche est considérée comme correctement classée. Sinon, la tâche est affectée à la classe de rejet. Il est plus instructif de prendre une décision si nous savons qu'il y a 90 % de chance que la tâche appartienne à une classe donnée.

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMS  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

TABLE 5.1 – Résultats de la classification de la structure hybride proposée (%) (sujet1).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Reject	Rejet + Taux
<b>Session 1</b>								
WW	84.78	90	99.56	86.95	87.82	89.82 [85.23- 93.09]	6.17	95.99
CS	86.08	89.56	98.69	92.17	86.95	<b>90.69</b> [86.23-93.80]	6	96.69
LLW	82.60	88.69	97.82	86.95	84.78	88.16 [83.34- 91.72]	6.52	94.68
MSVM <sup>2</sup>	84.78	90.43	98.69	87.82	84.78	89.3 [84.64- 92.66]	6.60	95.9
1-c-1	79.13	88.69	97.39	85.21	82.60	86.6 [81.59- 90.40]	5.91	92.51
1-c-tous	83.47	85.65	96.95	87.39	82.60	87.21 [82.27- 90.92]	5.82	93.03
<b>Session 2</b>								
WW	96.52	98.69	94.34	85.65	85.21	92.08 [87.85- 94.91]	6.26	98.34
CS	97.82	99.56	94.78	86.52	87.82	<b>93.3</b> [89.30- 95.87]	5.91	99.21
LLW	95.21	97.39	93.47	83.91	84.34	90.86 [86.43- 93.94]	6	96.86
MSVM <sup>2</sup>	96.08	99.56	93.04	84.78	86.08	91.90 [86.94- 94.29]	6.17	98.07
1-c-1	88.69	97.39	99.13	81.30	81.73	89.64 [85.02- 92.94 ]	6.08	95.72
1-c-tous	93.47	96.52	94.34	82.60	83.91	90.16 [85.62- 93.37]	5.56	95.72

TABLE 5.2 – Résultats de la classification de la structure hybride proposée (%) (sujet2).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Reject	Rejet + Taux
<b>Session 1</b>								
WW	69.56	82.60	90.43	96.95	91.73	86.25 [81.20- 90.10]	7.56	93.81
CS	70.86	84.34	88.26	97.82	92.60	<b>86.77</b> [80.92- 89.89]	7.39	94.16
LLW	67.39	77.39	88.69	96.52	87.39	83.47 [78.12- 87.71]	7.56	91.03
MSVM <sup>2</sup>	70.43	83.47	89.56	95.65	91.30	86.08 [81.01- 89.96]	7.73	93.81
1-c-1	64.78	77.82	87.39	94.34	92.17	83.30 [77.94- 87.56]	6.52	89.82
1-c-tous	66.08	73.47	88.26	95.17	85.65	81.12 [75.56- 85.65]	6.86	87.98

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMS  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

TABLE 5.3 – Résultats de la classification de la structure hybride proposée (%) (sujet3).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Reject	Reject + Taux
<b>Session 1</b>								
WW	69.13	86.08	77.82	66.95	73.47	74.69 [68.69- 79.87]	19.04	93.73
CS	68.69	88.26	77.82	67.82	72.17	<b>74.95</b> [68.97- 80.10]	18.86	93.81
LLW	66.52	75.21	74.78	64.34	60.43	68.25 [61.97- 73.92]	19.04	87.29
MSVM <sup>2</sup>	70.00	86.52	76.52	65.65	72.60	74.25 [68.23- 79.47]	18.95	93.2
1-c-1	61.73	78.26	69.13	60.43	68.26	67.56 [61.26- 73.27]	19.21	86.77
1-c-tous	65.65	76.08	73.91	64.78	59.13	67.91 [61.62- 73.60]	19.13	87.04
<b>Session 2</b>								
WW	88.26	78.26	87.39	78.69	77.82	82.08 [67.60- 86.49]	14.52	96.6
CS	89.13	79.56	85.65	80.43	78.26	<b>82.60</b> [77.17- 86.95]	14.52	97.12
LLW	85.65	64.78	83.91	75.65	73.47	76.69 [70.81- 81.68]	14.34	91.03
MSVM <sup>2</sup>	87.82	79.13	86.52	77.82	77.39	81.73 [76.22- 86.18]	14.52	96.25
1-c-1	80.86	71.73	80.43	74.78	74.34	76.42 [70.52- 81.44]	12.43	88.85
1-c-tous	93.91	65.65	81.30	74.34	71.73	75.38 [96.42- 80.50]	12.60	87.98

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMS  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

TABLE 5.4 – Résultats de la classification de la structure hybride proposée (%) (sujet5).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Rejet	Rejet + Taux
<b>Session 1</b>								
WW	91.30	63.47	90.43	88.69	66.52	80.08 [74.44- 84.72]	17.30	97.38
CS	93.47	63.47	91.30	88.26	68.69	<b>81.03</b> [75.46- 85.57]	17.65	98.68
LLW	83.47	63.04	86.08	76.95	66.08	75.12 [69.15- 80.26]	17.65	92.77
MSVM <sup>2</sup>	92.17	64.34	88.26	82.60	66.08	78.69 [72.94- 83.48]	17.73	96.42
1-c-1	87.39	58.69	83.47	78.26	63.91	74.34 [68.32- 79.55]	17.73	92.07
1-c-tous	82.17	61.30	87.39	76.08	66.52	74.69 [68.69- 79.87]	17.30	91.99
<b>Session 2</b>								
WW	82.60	66.52	89.56	74.34	64.34	75.47 [69.52- 80.58]	17.91	93.38
CS	86.52	67.39	88.26	76.08	66.95	<b>77.04</b> [71.18- 82]	17.82	94.86
LLW	76.52	64.78	88.26	66.08	64.78	72.08 [65.95- 77.47]	17.47	89.55
MSVM <sup>2</sup>	83.47	65.65	88.69	73.47	66.08	75.47 [69.52- 80.58]	18.08	93.55
1-c-1	79.13	63.04	83.91	69.56	63.91	71.91 [65.77- 77.32]	16.69	88.6
1-c-tous	74.34	63.04	88.69	66.95	64.34	71.47 [65.31- 76.91]	16.26	87.73
<b>Session 3</b>								
WW	86.95	84.34	95.21	83.91	67.82	83.64 [78.31- 87.86]	10	93.64
CS	86.08	86.95	95.65	86.52	65.21	<b>84.08</b> [78.79- 88.24]	10.17	94.25
LLW	88.69	76.52	90.43	76.52	69.13	80.25 [74.62- 84.87]	10.08	90.33
MSVM <sup>2</sup>	86.08	86.08	94.34	83.47	66.95	83.38 [78.02- 87.63]	10.17	93.55
1-c-1	83.04	81.30	90	76.08	65.21	79.12 [73.41- 83.87]	8.86	87.98
1-c-tous	86.08	75.21	89.13	77.82	68.26	79.3 [73.60- 84.03]	10	89.3

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMS  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

TABLE 5.5 – Résultats de la classification de la structure hybride proposée (%) (sujet6).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Reject	Rejet + Taux
<b>Session 1</b>								
WW	94.34	92.60	96.08	84.34	93.04	92.08 [87.85- 94.91]	7.56	99.64
CS	95.21	93.47	96.08	85.65	92.17	<b>92.51</b> [88.36- 95.25]	6	98.51
LLW	95.21	83.91	96.52	86.52	88.26	90.08 [85.53- 93.30]	8	98.08
MSVM <sup>2</sup>	95.21	91.73	95.65	85.21	91.30	91.82 [87.55- 94.71]	7.73	99.55
1-c-1	90	87.82	94.34	84.78	87.39	88.86 [84.13- 92.30]	7.56	96.42
1-c-tous	90.86	82.60	97.82	85.65	89.13	89.21 [84.53- 92.59]	8.26	97.47
<b>Session 2</b>								
WW	91.30	95.65	89.13	83.91	95.65	91.12 [87.90- 94.95]	6.26	97.38
CS	91.73	97.82	86.08	88.26	96.08	<b>91.99</b> [87.75- 94.84]	6	97.99
LLW	90	83.47	87.39	92.17	90.43	88.69 [83.94- 92.16]	9.39	98.08
MSVM <sup>2</sup>	90.43	97.39	86.95	83.91	95.21	90.77 [86.33- 93.87]	7.04	97.81
1-c-1	86.08	92.60	85.21	83.04	94.78	88.34 [84.54- 91.87]	7.56	95.9
1-c-tous	88.26	82.17	89.13	88.69	86.52	86.95 [81.98- 90.70]	8.78	95.73

TABLE 5.6 – Résultats de la classification de la structure hybride proposée (%) (sujet7).

Session	RT	MT	LT	RT	CT	Taux + interval de confiance de 95%	Reject	Rejet + Taux
<b>Session 1</b>								
WW	95.65	88.26	96.95	88.26	89.13	91.65 [87.35- 94.57]	6	97.65
CS	96.52	89.56	97.82	88.69	89.13	<b>92.34</b> [88.16- 95.12]	5.91	98.25
LLW	93.47	86.08	95.56	81.30	83.91	88.06 [83.23- 91.63]	8.95	97.01
MSVM <sup>2</sup>	96.08	87.82	97.82	86.08	87.39	91.03 [86.63- 94.08]	7.82	98.85
1-c-1	86.08	88.26	95.21	81.30	86.52	87.47 [82.56- 91.14]	8.43	95.9
1-c-tous	90.43	85.65	94.34	82.17	85.65	87.64 [82.75- 91.28]	8.52	96.16

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

---

Selon les resultats exposés, nous avons constaté :

- Le taux moyen de généralisation se situe entre 67.56% et 93.13%.
- Les M-SVMs directes généralisent mieux que les M-SVMs indirectes.
- Les taux de généralisation des M-SVMs directes sont aproximativement similaires, cependant, le modèle de CS surpasse les trois autres modèles.

Aussi, nous avons constaté a partir des intervals de confiance obtenus, que même dans le pire des cas, le système proposé (basé sur les SOMs et les M-SVMs) est capable avec une confiance de 95% de donner de bons resultats de classification pour des ensembles de tests futurs.

TABLE 5.7 – Resultats relatifs au : taux de la classe rejet + taux globale de classification.

Taux + rejet	Modèle WW	Modèle GS	Modèle LLW	Modèle M-SVM <sup>2</sup>	1-c-1	1-c-tous
<b>Moyenne</b>	96.14	96.70	93.33	96.08	91.86	91.83
<b>SD ±</b>	2.18	2.04	3.79	2.30	3.65	3.97

A partir du tableau 5.7, nous avons remarque que sur l'ensemble des differents sujets, la moyenne du : taux de la classe rejet + taux globale de classification, est comprise entre : 91.83% SD ± 3.97 et 96.70% SD ± 2.04 (Ou SD represente la Deviation Standard). Cette observation prouve que le nombre de tâches mal classées est reste faible, même si notre système n'est pas capable de bien classer une tâche donnée (il est préférable de dire je ne sais pas, que de faire une fausse prediction).

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

---

En dernier lieu et dans le but d'étudier la possibilité d'implémenter les M-SVMs directes dans des systemes temps réel, à savoir les BCIs, nous avons étudié le temps de calcul des differentes approches directes. Le temps d'apprentissage (en utilisant 3 trials) et le temps de prédiction (en utilisant 1 trial) de l'ensemble des M-SVMs directes sont rapportés dans le Tableau 5.8 et 5.9 comme suit :

- Le temps d'apprentissage par session, il représente la moyenne des differents temps d'apprentissage cumulés au cours des cinq etages de validation croisée.
- Le temps de prédiction par session, il représente la moyenne des differentes temps de prédiction cumulés au cours des cinq etages de validation croisée.

TABLE 5.8 – Temps d'apprentissage (en sec.) des M-SVMs directes probabilistes.

Sujets	Modèle WW	Modèle CS	Modèle LLW	Modèle M-SVM <sup>2</sup>
<b>1 Session1</b>	0.236	0.456	0.301	0.488
<b>Session2</b>	0.324	0.498	0.386	0.554
<b>2 Session1</b>	0.425	0.617	0.453	0.620
<b>3 Session1</b>	0.432	0.609	0.496	0.940
<b>Session2</b>	0.408	0.589	0.456	0.668
<b>5 Session1</b>	0.472	0.567	0.489	0.608
<b>Session2</b>	0.420	0.608	0.476	0.688
<b>Session3</b>	0.221	0.406	0.298	0.464
<b>6 Session1</b>	0.387	0.538	0.404	0.574
<b>Session2</b>	0.204	0.412	0.278	0.428
<b>7 Session1</b>	0.325	0.417	0.339	0.486
<b>Moyenne</b>	0.350	0.519	0.397	0.592
<b>SD ±</b>	0.094	0.085	0.082	0.142

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

---

TABLE 5.9 – Temps de prédiction (en sec.) des M-SVMs directes probabilistes.

Sujets	Modèle WW	Modèle CS	Modèle LLW	Modèle M-SVM <sup>2</sup>
<b>1 Session1</b>	0.013	0.017	0.009	0.012
<b>Session2</b>	0.012	0.015	0.013	0.010
<b>2 Session1</b>	0.022	0.014	0.017	0.020
<b>3 Session1</b>	0.014	0.011	0.013	0.016
<b>Session2</b>	0.008	0.009	0.007	0.004
<b>5 Session1</b>	0.026	0.015	0.020	0.023
<b>Session2</b>	0.019	0.013	0.010	0.008
<b>Session3</b>	0.013	0.010	0.007	0.014
<b>6 Session1</b>	0.012	0.013	0.011	0.017
<b>Session2</b>	0.012	0.005	0.006	0.004
<b>7 Session1</b>	0.016	0.018	0.017	0.020
<b>Moyenne</b>	0.015	0.012	0.011	0.013
<b>SD ±</b>	0.005	0.003	0.004	0.006

On peut voir à partir des tableaux 5.8 et 5.9 que le temps de tests cumulé est compris entre  $0.011 \text{ SD} \pm 0.004 \text{ sec.}$  (relatif au modèle directe de LLW) et  $0.015 \text{ SD} \pm 0.005 \text{ sec.}$  (relatif au modèle directe de WW), et le temps d'entraînement cumulé est compris entre  $0.350 \text{ SD} \pm 0.094 \text{ sec.}$  (relatif au modèle directe de WW) et  $0.592 \text{ SD} \pm 0.142 \text{ sec.}$  (relatif au modèle directe de M-SVM<sup>2</sup>). Ceci conclut que les M-SVMs directes nécessitent un temps de calcul raisonnable et approprié pour les applications temps réel.



## 5.6 Discussion

Pour prouver la contribution de l'approche proposée pour améliorer les performances des systèmes BCIs, il est intéressant de comparer nos résultats avec ceux des méthodes détaillées dans la section 4.2. (Voir Table 5.10).

- Dans [61] cinq sujets ont été considérés et dix combinaisons des tâches paires ont été testées par sujet. Les résultats ont été représentés sous forme, de taux de classification globale de toutes les combinaisons des cinq sujets considérés.
- Palaniappan [62] et Diez et al. [63], ont respectivement considéré quatre et cinq sujets et dix combinaisons des tâches paires ont été menées en chaque sessions. Les résultats ont été représentés pour chaque sujets sous forme de, taux de classification globale de toute les combinaisons paires effectuées au cours des defferentes sessions. Aussi, les auteurs ont reporté la moyenne des differents taux globales obtenus pour chaque sujet.
- Pour [60, 11, 7, 64] et notre méthode, cinq taches ont été testés simultanément. Les résultats ont été représentés sous forme, du meilleur taux global relatif aux sessions executées par sujet, et la moyenne des taux globales des sessions selectionnées. (Deux sujets ont été utilisés dans [11, 7], sept sujets ont été utilisés dans [64], et six sujets ont été utilisés dans [60] et notre méthode).

Cependant, pour que notre comparaison soit complètement correcte, il est nécessaire de prendre en consideration le nombre de classes et le nombre de sujets considérés dans les différentes études. Par conséquence, nous focalisons notre comparaison seulement avec les méthodes qui considerent 5 classes.

- Liang et al. [60] obtiennent un taux de classification entre 57.37% et 86.70%, et un taux global de 71.71%. le systeme de clasification proposé est capable de donner un taux de classification entre 82.60% et 93.3%, et un taux globale de 88.60%.

CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs  
MULTI-CLASSES ET LES CARTES DE KOHONEN POUR  
L'INTERPRÉTATION DE LA PENSÉE.

TABLE 5.10 – Les résultats des travaux précédants (%).

Methode	nombre classe	S1	S2	S3	S4	S5	S6	S7	Taux
<b>AR+BQ [61]</b>	2	–	–	–	–	–	–	–	84.16
<b>(PS+ASR)+ENN [62]</b>	2	79.80	69.10	70.20	80.30	–	–	–	74.85
<b>AR+ELM [60]</b>	5	86.70	75.99	57.37	–	62.79	72.62	74.81	71.71
<b>WPE+SVM [11]</b>	5	76.30	68.50	–	–	–	–	–	72.4
<b>EMD+LDA [63]</b>	2	95.00	92.45	80.93	92.73	85.13	94.61	97.36	91.17
<b>DWT+BPNN [7]</b>	5	90.75	73.30	–	–	–	–	–	82.02
<b>ST+LDA [64]</b>	5	97.97	88.78	84.72	97.72	95.12	97.32	94.32	93.71
<b>Our Method</b>	5	93.30	86.77	82.60	–	84.08	92.51	92.34	88.6

- Zhiwie and Minfen [11], obtiennent un taux de classification de 76.30% pour le sujet 1 et un taux de classification de 68.50% pour le sujet 2. Aussi, Tolic and Jovic [7] obtiennent un taux de 90.75%, 73.30% pour le sujet 1 et 2, respectivement. La méthode proposée surpasse ces résultats avec un taux de 93.3% pour le sujet 1 et 86.77% pour le sujet 2.
- Dans [64] les auteurs ont obtenus des taux de classification superieurs a ceux qu'ont a obtenus dans cette applications. Ceci peut se justifier par le fait qu'on a utilisé des classifieurs probabilistes qui sont munis d'une mesure de confiance calibrée (si la probabilité associée à une classe est supérieure à 0.75, la tâche est considérée comme correctement classée. Sinon, la tâche est affectée à la classe de rejet). De plus, les auteurs ont utilisé 10 niveaux de validation croisée, ce qui implique la presence des mêmes fragments de segments dans les exemples d'apprentissage et les exemples de test ; l'ensemble de ces fragments peuvent énormément contribuer à l'augmentation des taux de classification.

L'ensemble de ces résultats prouve que le système proposé peut être un excellent modèle pour améliorer les performances des systèmes BCIs.

## CHAPITRE 5. APPLICATION 1 : HYBRIDATION DES SVMs MULTI-CLASSES ET LES CARTES DE KOHONEN POUR L'INTERPRÉTATION DE LA PENSÉE.

---

### *Résumé :*

*Dans ce chapitre, nous avons proposé et implémenté une nouvelle structure hybride à deux étages pour effectuer la discrimination entre cinq tâches différentes : RelT, MulT, LetT, RotT, et ComT, prises de la base de données de Aunon et Keirn.*

*Dans le premier étage, nous avons proposé une nouvelle alternative de sélection de variables basée sur les cartes SOMs à apprentissage non supervisé. Les cartes reçoivent comme entrées les 150 paramètres statistiques obtenus à partir de la DWT.*

*Dans le deuxième étage, nous avons utilisé les M-SVMs probabilistes à apprentissage supervisé pour effectuer la classification finale. Les M-SVMs reçoivent comme entrées les vecteurs de paramètres, qui ont été réduits par les cartes SOMs. Les deux approches des M-SVMs ont été considérées dans cette étude, à savoir, l'approche directe et l'approche indirecte.*

*Les résultats ont démontré que les M-SVMs directes sont plus performants en généralisation que les M-SVMs indirectes, et que le modèle de CS surpasse les trois autres modèles directs. D'un autre côté, il a été déduit que le post traitement des sorties des M-SVMs ont significativement diminué le nombre de fausses alertes.*

*Aussi, il a été conclu que les M-SVMs directes nécessitent un temps de calcul raisonnable et approprié pour les applications temps réel. L'ensemble de ces observations concluent à l'efficacité de la méthode proposée pour améliorer les performances des systèmes BCIs.*

## Chapitre 6

# Application 2 : Classification des signaux ECGs pour la détection des pathologies cardiovasculaires

Les pathologies cardiovasculaires constituent la première cause de mortalité dans le monde. La détection précoce de ces maladies réduit énormément le taux de décès. En conséquence, plusieurs techniques d'enregistrement de l'activité cardiaque ont été mises en place. Afin de fournir des informations significatives et utiles sur les aspects fonctionnels du cœur : l'ElectroCardioGramme (ECG), la radiographie thoracique, l'échocardiographie, la tomodensitométrie ("scanner"), l'Imagerie par Résonance Magnétique (IRM), le cathétérisme cardiaque, les examens du sang, et rarement la biopsie du muscle cardiaque. Parmi elles, l'ECG est de loin la technique la plus utilisée en raison de sa capacité à enregistrer les signaux cardiaques de façon non invasive et en raison de son coût réduit. Cependant, dans certains cas, les symptômes de la maladie ne peuvent pas survenir tout le temps,

mais surviennent à certains intervals inattendus au cours de la journée. En conséquence, pour effectuer des diagnostics rigoureux et précis, il est indispensable d'opérer une étude du signal ECG sur plusieurs heures. Ceci implique la nécessité d'un traitement automatique des signaux ECGs pour aider les cardiologues dans l'analyse des enregistrements longs.

Par ailleurs, l'étude de ces signaux ECGs est difficile, en effet ces signaux sont contaminés par des artefacts (perturbation de la ligne d'alimentation, les mouvements d'électrodes et les mouvements musculaires, etc.). Il est donc nécessaire, pour assurer de bonnes performances du diagnostic, d'avoir un système qui inclut : une bonne méthode d'extraction de paramètres, et une bonne méthode de classification. Ainsi, dans le but d'avoir de bonnes performances en généralisation, nous proposons dans ce chapitre un système d'aide au diagnostics des pathologies cardiovasculaires (figure 6.1). Ce système est basé sur la Transformée par Ondelettes Discrètes (DWT) et des techniques de classification récemment développées, à savoir, les Machines à Vecteurs de Support Multiclasses (M-SVMs) directes probabilistes. L'approche des M-SVMs indirectes probabilistes a aussi été implémentée. Le but principal de ce chapitre est de comparer les performances des deux approches des M-SVMs et déduire l'apport des approches directes dans le traitement des signaux ECGs. Cette étude est validée sur des enregistrements reportés de la base de données MIT-BIH [23], pour la discrimination de cinq arythmies cardiaques : Bloc de Branche Droit (RBBB), et Bloc de Branche Gauche (LBBB), Extrasystole Ventriculaire (PVC), Extrasystole Auriculaire (APB), la classe des enregistrements spécifiques au port des pacemaker (PB), en plus du battement normal (N).

Dans le présent chapitre, nous donnerons en premier lieu, une description de la base de données utilisée. Puis nous effectuerons un état de l'art, dans lequel nous exposerons certains modèles proposés dans la littérature pour la classification des pathologies cardiaques en utilisant des signaux ECGs. En troisième lieu, nous expliquerons en détail la chaîne de traitement que nous proposons pour ca-

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTÉCTION DES PATHOLOGIES CARDIOVASCULAIRES

racteriser et extraire l'information utile d'un battement cardiaque. Ensuite, nous présenterons la méthode de classification utilisée. Enfin, nous exposerons et nous discuterons les résultats obtenus.

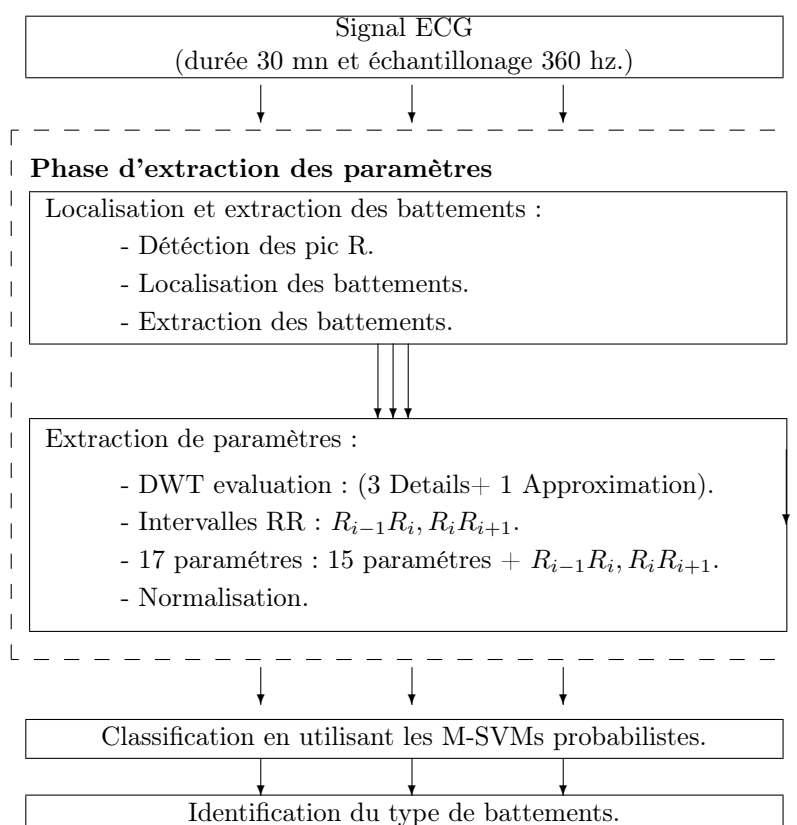


FIGURE 6.1 – Diagramme général de la méthode proposée pour l'identification de la classe des battements.

## 6.1 Description de la base de données

Pour réaliser ce travail, nous avons utilisé une base de données universelle disponible gratuitement sur internet. C'est la base "MIT BIH" [23] qui contient plusieurs enregistrements issus de patients différents. Ces signaux ont été obtenus par le laboratoire d'arythmies du « Beth Israël Hospital ». Cette base contient 49 enregistrements numérotés de (100 à 234). Quelques numéros d'enregistrements sont manquants. Chaque enregistrement a une durée de 30 min. et une fréquence d'échantillonnage de 360 Hz et se compose de trois fichiers. Par exemple pour le patient 105 on trouve :

- Un fichier 105.dat, qui contient les échantillons digitalisés du signal.
- Un fichier texte 105.heg, qui décrit le code du patient, format du stockage, fréquence de prélèvement, nombre et types des signaux.
- Un fichier 105.atr, qui contient les étiquettes des battements (normaux, pathologiques).

Nous avons utilisé 23 enregistrements (table 6.1) de cette base de données, et nous avons examiné six types de battements, à savoir : Bloc de Branche Droit (RBBB), et Bloc de Branche Gauche (LBBB), Extrasystole Ventriculaire (PVC), Extrasystole Auriculaire (APB), battement normal (N) et la classe des enregistrements spécifiques au port des pacemaker (PB). Tableau 6.1

## 6.2 Etat de l'art

On peut trouver de nombreux travaux intéressants proposés dans la littérature pour la détection des pathologies cardiovasculaires par le biais des signaux ECGs. Ci-après, nous avons sélectionné les travaux suivants :

Zhao et Zhang [79], proposent une approche pour une reconnaissance automatique fiable de six types de battements différents, à savoir : N, RBBB, LBBB, PVC, APB, et PB. Après avoir segmenté les signaux ECGs en des battements de 300 échantillons, les auteurs ont proposé une nouvelle méthode d'extraction des para-

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

---

TABLE 6.1 – Patients et nombre de battements utilisés dans cette étude.

Type	Enregistrement	Total
<b>N</b>	103, 113, 115, 123, 220, 234	1200*6
<b>LBBB</b>	109, 111, 207, 214	1200*4
<b>RBBB</b>	118, 124, 212, 231	1200*4
<b>PVC</b>	119	400
	221	300
	200,233	800*2
<b>APB</b>	209	300
	222	200
	232	1200
<b>PB</b>	107, 217	1200*2
		23200

mètres. Cette technique est basée sur l'application de deux méthodes différentes la DWT et la modélisation AutoRégressif (AR). La DWT a été utilisée pour extraire les coefficients de la transformée en fonction des caractéristiques de chaque segment ECG. Simultanément, la modélisation AR est également appliquée pour obtenir les structures temporelles des formes d'onde ECG. En phase de classification, un discriminateur de type SVM est utilisé pour séparer entre les six classes. À partir de simulations réalisées, une précision globale de classification de 99,68 % est obtenue sur l'ensemble de 6 types de battements cardiaques.

In [10], Yu et Chen ont proposé un système de classification des battements ECGs basé sur la DWT et les Réseaux de Neurones Artificiels (ANNs) probabilistes pour discriminer entre six types de pathologies cardiovasculaires : N, RBBB, LBBB, PVC, APB, PB. Les auteurs, ont tout d'abord subdivisé les différents signaux ECGs en segments de 64 échantillons de telle manière à contenir le complexe QRS de chaque battement. Dans la phase d'extraction de paramètres, deux niveaux de DWT ont été appliqués à chaque segment. Ensuite, trois ensembles



de caractéristiques statistiques des signaux décomposés, ainsi que l'énergie et les intervals RR du signal original sont exploités pour caractériser les signaux ECGs. Dans la phase de généralisation, les auteurs ont implémenté un classifieur de type ANN probabiliste. Les résultats obtenus exposent une précision prometteuse de 99,65 %.

In [13], les auteurs ont proposé un système intelligent qui intègre l'analyse en composantes indépendantes (ICA), pour la discrimination entre huit types de battements ECGs différents : N, RBBB, LBBB, APB, PVC, PB, Flutter ventriculaire (VFW), et Battement d'Échappement Ventriculaire (VEB). Après la segmentation des signaux ECGs en battements de 200 échantillons ; chaque échantillon est prétraité en supprimant tout d'abord la valeur moyenne pour éliminer l'effet de décalage, puis en divisant sur l'écart type. En phase d'extraction des paramètres, trois fonctions de contraste pour le calcul des composantes indépendantes ont été étudiées. Aussi les auteurs ont proposé une nouvelle stratégie de sélection pour la réorganisation des composantes indépendantes, l'algorithme de JADE, ce dernier a été comparé avec trois autres algorithmes : OWSL, OWLS, et OWFP. En phase de généralisation, Deux classifieurs, ANN et SVM, ont été utilisés et leurs effets dans le classement des battements ECGs ont été comparés. D'après les résultats obtenus, un taux globale de 98.7 % a été obtenu avec, la fonction  $\logcosh(\cdot)$ , l'algorithme OWSL, et le classifieur SVM.

Martis et co-auteur [12], ont automatiquement classé cinq types de battements ECGs : N, RBBB, LBBB, APB, PVC. Les signaux ECGs ont d'abord été débruités en utilisant une technique de débruitage à base d'ondelettes. Après le débruitage, les complexes QRS des différents signaux ont été localisés grâce à l'algorithme de Pan Tompkins. Après détection des complexes QRS, les signaux ECGs ont été subdivisés en segments de 200 échantillons représentant chacun un battement cardiaque. Les auteurs ont par la suite, proposé et comparé les performances de trois approches d'extraction de paramètres. La première approche utilise les composantes principales des battements ECGs, la seconde utilise les composantes prin-

principales des signaux d'erreur du modèle de Prédiction Linéaire (PL), alors que la troisième approche utilise les composantes principales des coefficients de DWT comme caractéristiques. Les caractéristiques provenant de deux approches ont été classées indépendamment en utilisant les ANNs et une SVM à Moins carré (LS-SVM) (Least Square SVM). Le meilleur résultat a été obtenu en utilisant la première approche d'extraction des paramètres et les LS-SVMs, avec un taux de classification de 98.11 %.

Dans [80], les auteurs ont proposé un système automatique d'aide au diagnostic de cinq pathologies cardiovasculaires : LBBB, RBBB, PVC, PB, et la fusion du PB et N (PB+N), en plus de la classe N. Tout d'abord, un filtre passe bandes (type db3) a été implémenté par application d'un filtre passe haut (fréquence de coupure 12 Hz) en cascade avec un filtre passe bas (fréquence de coupure 5 Hz). Les fréquences de coupure ont été choisies de telle manière à inclure toutes les ondes du complexe QRS, dont la fréquence est comprise entre 5 Hz et 15 Hz. Par la suite, les signaux ECGs ont été segmentés en battement contenant chacun 200 échantillons, ces derniers ont été utilisés par la suite comme descripteurs des battements cardiaques. En phase de généralisation, un classifieur binaire de type (ANN) a été utilisé pour discriminer entre une classe et toutes les autres classes restantes. Les auteurs ont insisté sur le fait que le nombre d'exemples d'apprentissage des différentes catégories doit être équiprobable (d'où le terme : Balanced Trained Neural Network). Cette contrainte d'apprentissage a significativement augmenté les résultats, et a permis d'avoir un taux de généralisation globale de 99.49 %, avec un taux de généralisation de 100 % obtenu pour trois classes.

Casas et al [81], comparent dans leur travaux trois méthodes de classification pour discriminer entre la classe N, la classe PVC et d'autres types de battements. Après l'extraction des battements, les auteurs ont tout d'abord pré-traité chacun d'eux en lui supprimant la valeur moyenne. Une fois les battements normalisés, chacun a été divisé en quatre vecteurs, puis chacun des vecteurs était soumis à une procédure de traitement pour être ensuite caractérisé par 20 descripteurs. Ainsi

chaque battement a été caractérisé par 80 descripteurs. En phase de classification, trois classifieurs ont été testés : SVM, ANN, et l'algorithme de Regression Logistique (ALG). Les auteurs ont souligné que les meilleurs résultats ont été obtenus avec un classifieur SVM, avec un taux globale de 97.3 % et un taux de 99.9 % pour la classe PVC.

Les résultats obtenus de différentes études résumées dans cette section sont donnés dans le tableau 6.11.

### 6.3 Méthode d'extraction des paramètres des signaux ECGs proposée

De manière générale, l'analyse d'un signal ECG est habituellement composée de deux étapes : la première qui est l'extraction des paramètres consiste à exprimer le signal original à l'aide de descripteurs, en effectuant une transformation du signal de l'espace d'enregistrement à l'espace des descripteurs. La seconde étape s'effectue dans ce deuxième espace : c'est l'analyse de la valeur des descripteurs pour déduire les propriétés recherchées et réaliser le diagnostic. Il est évident alors que plus la première étape est effectuée rigoureusement ou, en d'autres termes, plus les descripteurs sont pertinents pour la propriété recherchée, plus la seconde étape se voit simplifiée et fiable.

Les paramètres associés à un battement cardiaque peuvent être extraits par le biais de différentes approches : l'approche directe [81, 80], l'approche temporelle [82, 83, 84], l'approche fréquentielle [85, 86, 84], l'approche statistique [87], et l'approche temporelle et fréquentielle [79, 10, 84, 88, 89, 90]. Dans la suite de cette section, nous décrivons la chaîne de traitement que nous proposons pour l'extraction des paramètres d'un battement ECG. Cette chaîne de traitement se décompose principalement en deux étapes, à savoir : la localisation et l'extraction des battements, et l'estimation des paramètres.

### 6.3.1 Localisation et extraction des battements

Dans cette partie nous définissons la méthodologie adoptée pour la localisation d'un battement cardiaque par le biais de l'algorithme de Pan et Tompkins [91], ainsi que la méthodologie adoptée pour l'extraction d'un battement une fois que ce dernier soit détecté.

**Détection des battements :** l'algorithme de Pan et Tompkins [91], a été choisi pour détecter les battements. Il retourne pour chaque complexe détecté, la position du pic R située au niveau où la dérivée du battement présente l'énergie la plus élevée. Les différentes étapes qui le constituent, sont :

- Calcul et lissage des dérivées, cette étape permet de mettre en évidence les fortes pentes qui distinguent les complexes QRS des composantes ECG de basses fréquences, telles que les ondes P et T. Pour le calcul des dérivées premières  $Y_1$  et secondes  $Y_2$ , la méthode de dérivées à point central a été choisie, en raison de son insensibilité au bruit par rapport à la méthode de différence entre deux points successifs. Les deux dérivées sont ensuite lissées par la méthode de lissage par moyenne mobile sur trois points, l'opération de base pour ce calcul est définie par l'équation 6.3.

$$Y_1 = x(n+1) - x(n-1). \quad (6.1)$$

$$Y_2 = x(n+2) - 2x(n) + x(n-2). \quad (6.2)$$

$$Y_{liss} = 1/4[y(n-1) + 2y(n) + y(n+1)]. \quad (6.3)$$

Avec  $Y_{liss1}$  lissage de la dérivée première et  $Y_{liss2}$  lissage de la dérivée seconde.

- Mise au carré des dérivées, Nous avons ajouté un processus non linéaire qui est l'élevation au carré des points des deux dérivées lissées point à point, cette opération va nous permettre de faire ressortir les valeurs les plus hautes, à savoir, les valeurs des complexes QRS.

$$y(n) = y_{liss1}^2 + y_{liss2}^2 \quad (6.4)$$

- Fenêtre d'intégration, le signal obtenu  $y(n)$  est passé à travers une fenêtre d'intégration de  $N$  points (equation 6.5), la sortie de l'intégrateur peut être utilisée pour détecter des complexes QRS, mesurer les intervals RR ou déterminer la durée des complexes QRS. Dans cette étude nous avons obtenu  $Z(n)$  en utilisant la convolution de  $y(n)$  par une fenêtre rectangulaire  $W(n)$  (equation 6.6).

$$Z(n) = [1/n] \sum x(n - (N - 1)). \quad (6.5)$$

$$Z(n) = conv(w(n), y(n)). \quad (6.6)$$

- Seuillage, On calcule les positions des complexes QRS grâce à un seuil  $s$  le seuil est originalement fixé à 35% de la valeur maximal de  $Z(n)$ .
- Localisation des pics R, Localisation des pics R qui représente les instants de présence des extremes des complexes QRS, figure 6.2.

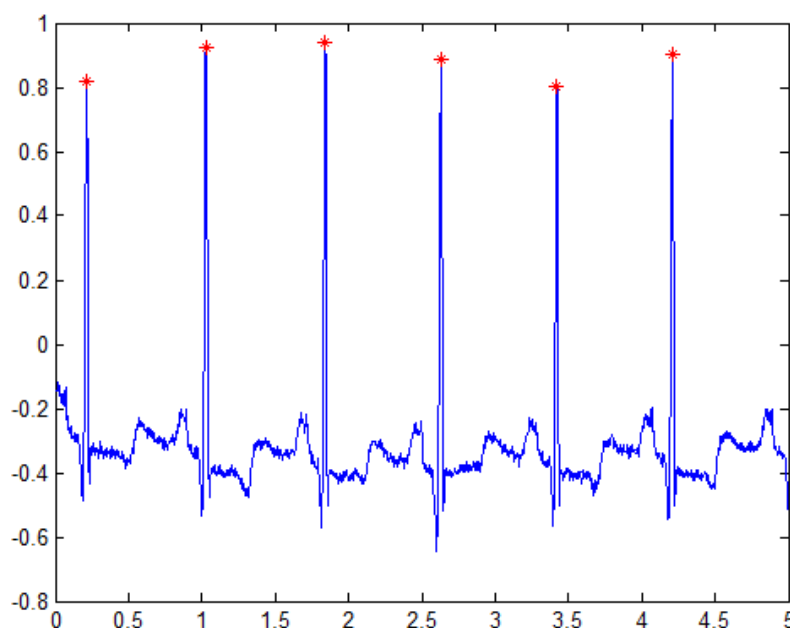


FIGURE 6.2 – Localisation des pics R des complexes QRS du signal EEG 100.

**Extraction des battements :** Une fois les battements détectés, une fenêtre temporelle, dont la taille est fixée de manière à contenir un battement entier, est ouverte autour de chacun d’entre eux. Dans cette étude les battements seront présentés sous forme d’un vecteur de 360 échantillons centrés sur le pic R, ainsi on est sûr que toutes les ondes du battement seront incluses et la morphologie du battement est préservée (figure 6.3)

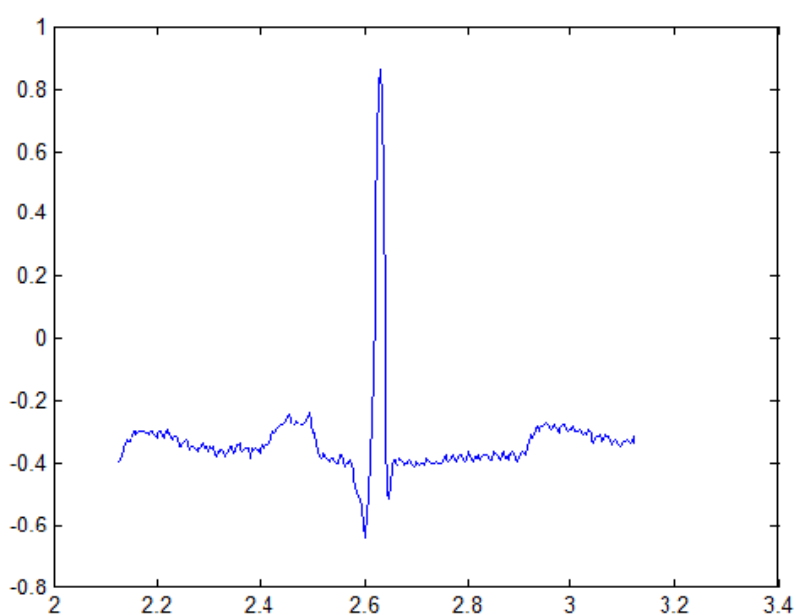


FIGURE 6.3 – Battement échantillonné sur 360 Hz centré sur le pic R (ECG 100)

### 6.3.2 Estimation des paramètres

L’ECG est un signal bio-électrique qui enregistre l’activité électrique du coeur dans le temps et la fréquence. Par conséquent, l’étude des signaux ECGs nécessite des méthodes capables de fournir des informations dans le domaine temporel et dans le domaine fréquentiel. Par conséquent, la DWT a été sélectionnée, car elle a

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

---

la capacité de représenter les signaux non-stationnaires tels que les signaux ECGs dans ces deux domaines. Nous avons appliqué pour chaque battement cardiaque cinq niveaux de décomposition, en utilisant l'ondelette Daubechies 2 (db2). Nous avons ensuite sélectionné les trois derniers détails (D3, D4, et D5) et la dernière approximation (A5), car le signal ECG est compris entre 0,5 Hz et 40 Hz et l'énergie des coefficients d'ondelettes est essentiellement concentrée dans les sous-bandes inférieures. Les largeurs des sous-bandes (D3, D4, D5, et A5) d'un signal ECG obtenues par application de la DWT (selon les équations 5.1 et 5.2) sont énumérées dans le tableau 6.2, aussi, La figure 6.4 illustre Les coefficients des 5 niveaux de décomposition et la dernière approximation de la DWT pour un battement normal.

TABLE 6.2 – Les gammes de fréquences des sous-bandes (D3, D4, D5, et A5) relatives à un signal ECG échantillonné à 360 hz.

sous-bandes	La fréquence des sous-bandes (Hz)
D3	22.5-45
D4	11.25-22.5
D5	5.625-11.25
A5	0-5.625

Nous avons extrait des sous-bandes (D1, D2, D3, et A5) et le signal d'origine, trois caractéristiques statistiques [10] pour caractériser chaque battement de coeur, à savoir : la variance et la variance de la fonction d'autocorrélation d'un signal (les équations mathématiques sont définies dans le chapitre 5), amplitude relative (équation 6.7).

$$r(x) = \frac{Min(x(n))}{Max(x(n))} \quad (6.7)$$

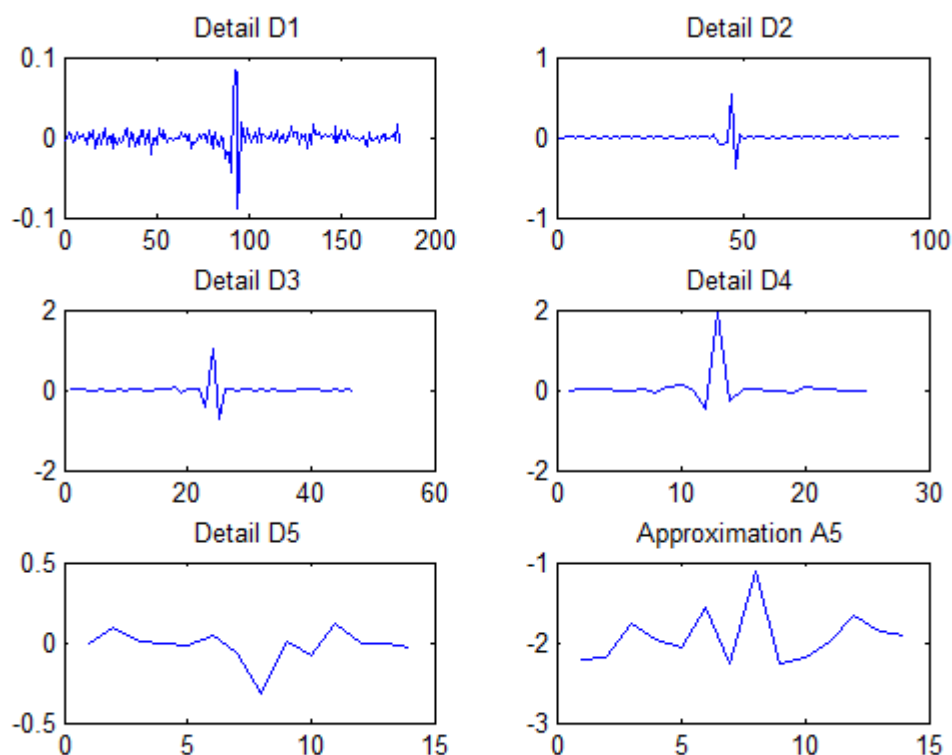


FIGURE 6.4 – Les coefficients des 5 niveaux de décomposition et la dernière approximation de la DWT pour un battement normal (ECG 100).

Certaines arythmies, comme PVC et APB, sont caractérisées par des intervalles  $RR$  plus courts que les autres types de signaux. Par conséquent, nous exploitons l'intervalle  $RR$  comme étant un autre paramètre, qui est défini comme le temps relatif séparant deux pics R adjacents  $R_{i-1}R_i$  et  $R_iR_{i+1}$ . Les intervalles  $RR$  sont calculés une fois que les pics R ont été détectés à l'aide de la méthode de Pan et Tompkins [91].



## 6.4 Classification des signaux ECGs par les M-SVMs probabilistes

Le but de l'étape de classification est d'assigner automatiquement une classe au vecteur de caractéristiques précédemment extrait. Cette classe représente la catégorie à laquelle appartient un battement cardiaque. Ainsi, la nécessité d'avoir un bon système de classification joue un rôle important dans la performance du diagnostic. Beaucoup de méthodes ont été proposées au cours de ces dernières années pour la détection et l'identification assistée par ordinateur des arythmies cardiaques à partir des signaux ECGs. Parmi ces méthodes on peut citer : les ANNs [92, 10, 90], les SVMs [12, 13, 88, 89], les Mélanges d'Experts [87], etc.. Comme pour les signaux EEGs, l'application des SVMs a connu un grand succès, dans la problématique de la détection automatique des pathologies cardiovasculaires. Neanmoins, personne n'a pensé à tester l'apport des M-SVMs directes. Dans cette étude nous introduisons l'utilisation des quatre modèles des M-SVMs directes probabilistes (modèle WW, modèle CS, modèle LLW, et le modèle de la M-SVM<sup>2</sup>) pour la classification des signaux ECGs, nous avons aussi implémenté deux méthodes de décomposition probabiliste : 1-c-1 et 1-c-tous.

Pour les expériences de cette application, nous utilisons le même package M-SVM (M-SVMpack) [77] pour l'implémentation des approches directes des M-SVMs. De même, pour la construction des classifieurs binaires nous avons ajusté les paramètres du modèle CS, du fait qu'il surpasse les autres modèles en taux de généralisation.

## 6.5 Résultats

Les performances de notre système sont calculées en fonction de quatre procédures de validation croisée. Dans chaque jeu de validation croisée, 1/4 des 23200 battements sont choisis aléatoirement pour le test, 1/4 des 23200 battements sont choisis aléatoirement pour la validation, et les 2/4 restant sont sélectionnés pour constituer les données de formation.

### 6.5.1 Sélection des paramètres

A fin de trouver les meilleurs hyper paramètres pour les modèles M-SVMs implémentés avec le noyau RBF, nous effectuons comme pour la première application la sélection du modèle et nous optimisons deux paramètres : le paramètre  $C$  et le noyau  $\gamma$ . Pour chaque niveau de validation nous testons,  $15 \times 15 = 225$  différentes combinaisons de  $(C, \gamma)$ ,  $\gamma = [2^4, 2^3, 2^2, \dots, 2^{-10}]$  et  $C = [2^{12}, 2^{11}, 2^{10}, \dots, 2^{-2}]$ .

De même, pour les M-SVMs basées sur l'approche 1-c-1 et 1-c-tous, nous avons considéré pour chaque étage de validation croisée et pour tous les problèmes binaires, la même pair de  $(C, \gamma)$ .

### 6.5.2 Évaluation des systèmes de classification implémentés

Les tableaux 6.3, 6.4, 6.5, 6.6, 6.7, 6.8 exposent : la matrice de confusion entre classes, rejet par classe, et les faux négatifs ; relatives aux différents classifieurs M-SVMs probabilistes implémentés. Aussi, le tableau 6.9 expose pour chaque un des modèles directes et indirectes : le taux par classe, le taux global, rejet global, ainsi que les faux négatifs globaux. Les intervalles de confiance obtenus après application de la métrique statistique de Bennani et co-auteurs (voir équation 5.11), sont également exposé dans le tableau 6.9. A noter que :

- Le taux de test par classe, il représente pour chaque classe, la moyenne des taux obtenus au cours des quatre étages de validation croisée.

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

---

- Le taux de global, il représente pour chaque modèle, la moyenne des taux de tests des différentes classes.
- Le rejet, il représente pour chaque classe l'ensemble des rejets cumulés au cours des quatre étages de validation croisée.
- Le rejet global, il représente l'ensemble des rejets cumulés sur les six classes.
- Les Faux Négatifs (FN), ils représentent pour chaque classe, l'ensemble des exemples mal classés et non rejetés cumulés au cours des quatre étages de validation croisée.
- Les faux négatifs globaux, ils représentent l'ensemble des FN cumulés sur les six classes.

A noter aussi, que les sorties de nos M-SVMs sont post-traitées afin de générer des estimations de probabilités à posteriori des classes (voir chapitre 4, section 4.4). Ensuite, si la probabilité associée à une classe est supérieure à 0.75, la tâche est considérée comme correctement classée. Sinon, la tâche est affectée à la classe de rejet.

TABLE 6.3 – Matrice de confusion 1-c-1.

	<b>N</b>	<b>LBBB</b>	<b>RBBB</b>	<b>PVC</b>	<b>APB</b>	<b>PB</b>	<b>Rejet</b>	<b>FN</b>
<b>N</b>	7155	6	8	4	2	0	25	20
<b>LBBB</b>	7	4766	6	1	1	0	19	15
<b>RBBB</b>	3	7	4769	1	3	0	17	14
<b>PVC</b>	2	0	0	2275	5	0	18	7
<b>APB</b>	5	1	1	7	1676	0	10	14
<b>PB</b>	0	0	0	0	0	2397	3	0

CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS  
POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

TABLE 6.4 – Matrice de confusion 1-c-tous.

	N	LBBB	RBBB	PVC	APB	PB	Rejet	FN
N	7170	4	6	0	1	1	18	12
LBBB	4	4772	9	0	2	0	13	15
RBBB	3	8	4771	1	0	0	17	12
PVC	3	1	0	2278	7	0	11	11
APB	2	0	0	7	1678	0	13	9
PB	0	0	0	0	0	2399	1	0

TABLE 6.5 – Matrice de confusion WW.

	N	LBBB	RBBB	PVC	APB	PB	Rejet	FN
N	7193	0	2	0	0	0	5	2
LBBB	2	4786	4	0	1	0	7	7
RBBB	1	4	4790	0	1	0	4	6
PVC	1	0	0	2287	4	0	8	5
APB	1	0	0	2	1693	0	4	3
PB	0	0	0	0	0	2400	0	0

TABLE 6.6 – Matrice de confusion CS.

	N	LBBB	RBBB	PVC	APB	PB	Rejet	FN
N	7194	1	1	0	0	0	4	2
LBBB	0	4792	3	0	0	0	5	3
RBBB	3	4	4786	0	1	0	7	7
PVC	1	0	1	2290	4	0	4	6
APB	0	0	0	3	1692	0	5	3
PB	0	0	0	0	0	2400	0	0

CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS  
POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

TABLE 6.7 – Matrice de confusion LLW.

	N	LBBB	RBBB	PVC	APB	PB	Rejet	FN
N	7185	2	4	0	1	0	8	7
LBBB	3	4782	7	1	0	0	7	11
RBBB	2	4	4786	0	0	0	8	6
PVC	1	0	0	2287	4	0	7	5
APB	2	0	0	5	1687	0	6	7
PB	0	0	0	0	0	2400	0	0

TABLE 6.8 – Matrice de confusion M-SVM<sup>2</sup>.

	N	LBBB	RBBB	PVC	APB	PB	Rejet	FN
N	7189	1	3	0	1	0	6	5
LBBB	1	4785	2	0	0	0	12	3
RBBB	0	4	4787	0	1	0	8	5
PVC	0	0	0	2290	2	0	8	2
APB	1	0	0	3	1687	0	9	4
PB	0	0	0	0	0	2400	0	0

TABLE 6.9 – Resultat de classification des M-SVMs (%).

Modèle	N	LBBB	RBBB	PVC	APB	PB	Taux	Rejet	FN
<b>1-c-1</b>	99.37	99.29	99.35	98.91	98.58	99.87	99.22 [99.09-99.32]	92	70
<b>1-c-tous</b>	99.58	99.41	99.39	99.04	98.70	99.95	99.34 [99.22-99.43]	73	59
<b>WW</b>	99.90	99.70	99.79	99.43	99.58	100	99.73 [99.65-99.78]	28	23
<b>CS</b>	99.91	99.83	99.70	99.56	99.52	100	99.75 [99.67-99.80]	25	21
<b>LLW</b>	99.79	99.62	99.70	99.43	99.23	100	99.62 [99.53-99.69]	36	36
<b>M-SVM<sup>2</sup></b>	98.84	99.68	99.72	99.56	99.23	100	99.67 [99.58-99.73]	43	19

Selon ces résultats, on peut observer que les meilleures précisions moyennes sont obtenues en utilisant les M-SVMs directes. Une fois de plus nous avons constaté que le modèle de CS surpasse les autres M-SVMs directes, néanmoins cette différence en généralisation reste raisonnable. Aussi, nous avons conclu à partir des intervalles de confiance obtenus, que même dans le pire des cas, les M-SVMs directes sont capables avec une confiance de 95 % de donner de bons résultats de classification pour des ensembles de tests futurs.

D'autre part, un bon système de classification doit avoir un nombre réduit de FN. Donc, pour examiner plus précisément l'effet des M-SVMs directes dans la classification des signaux ECGs, on a ajouté les FN comme un autre indice de performance. Nous avons observé que le nombre de battements mal classés par les M-SVMs directes est faible en comparaison avec les autres modèles indirectes. Ces observations confirment que les approches directes généralisent mieux que les approches indirectes.

Il est aussi très intéressant pour une méthode de discrimination de mesurer son temps de calcul. En effet, ce dernier est un facteur primordial pour étudier la faisabilité d'introduire une méthode de classification dans des systèmes temps réel. Le tableau 6.10 expose le temps d'entraînement ainsi que le temps de prédiction des quatre modèles des M-SVMs directes comme suit :

- Le temps d'apprentissage, il représente la moyenne des différents temps d'apprentissage cumulés au cours des quatre étapes de validation croisée.
- Le temps de prédiction, il représente la moyenne des différents temps de prédiction cumulés au cours des quatre étapes de validation croisée.

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

TABLE 6.10 – Temps d’entraînement et de prédiction (en sec.) des M-SVMs directes probabilistes.

Temp	Modèle WW	Modèle CS	Modèle LLW	Modèle M-SVM <sup>2</sup>
<b>Apprentissage</b>	1.912	2.718	1.876	2.578
<b>Test</b>	0.806	0.787	0.718	0.796

On peut voir à partir du tableau que le temps de tests cumulés est compris entre 0.718 sec. (relative au modèle directe de LLW) et 0.806 sec. (relative au modèle directe de WW), et le temps d’entraînement cumulé est compris entre 1.876 sec. (relative au modèle directe de LWW) et 2.718 sec. (relative au modèle directe de CS). A partir de ces resultats on constate que le modèle le plus long, qui est le modèle de CS est capable de traiter 7196 battements par minute (sachant qu’en moyenne le nombre de battements par minute est inclu entre 80 et 100 battements par minute). Ceci conclut que les M-SVMs directes necessitent un temps de calcul raisonnable, et peuvent ainsi etre embarqués dans des systemes temps réel d’aide au diagnostique des pathologies cardiovasculaires. A noter qu’en générale, on ne prend pas en consideration, le temps necessaire pour l’apprentissage et l’ajustement des paramètres des systemes réel. Ce qui est le plus important, est le temps necessaire pour reconnaitre une certaine pathologie.

### 6.6 Discussion

Le tableau 6.11 résume, les resultats des travaux exposés dans la section 6.2, et les resultats que nous avons obtenus en utilisant la M-SVM directe de CS. Il est donc intéressant de comparer notre approche avec les méthodes résumées dans le tableau 6.11. Neanmoins, nous ne pouvons pas considérer que cette comparaison est tout à fait juste, parce qu’il ya plusieurs facteurs à prendre en considération par exemple : type et nombre de classes, nombre d’exemples, etc.

Les resultats sont reportés sous forme de taux de tests des classes suivantes :

N, RBBB, LBBB, PVC, APB, PB, et autre ; et le taux de classification global qui représente la moyenne des taux de tests des différentes classes.

On note que la classe "Autre" [80, 81, 13] constitue le taux de tests moyens, relatif aux classes différentes de celles considérées dans cette étude. Aussi, on note que les auteurs dans [12] ont juste exposé dans leurs résultats le pourcentage du taux global.

- Dans [80] les auteurs ont utilisé 300 battements ECGs pour la discrimination entre six classes (N, LBBB, RBBB, PVC, PB, et la fusion du PB et N). Tandis que Martis et co-auteur [12] ainsi que Yu et co-auteur [13], ont considéré respectivement 34989 battements et 4900 battements pour la discrimination entre 8 classes (N, LBBB, RBBB, PVC, APB, PB, VFW, et VEB). Aussi, Casas et co-auteur [81] ont considéré 11239 battements pour la discrimination entre la classe N, PVC, et l'ensemble des pathologies restantes de la base Mit-Bih.
- Dans les travaux exposés dans [79, 10], les auteurs ont considéré respectivement 3403 battements et 11600 battements respectivement pour effectuer la discriminations entre six types de classes : N, LBBB, RBBB, PVC, APB, et PB. De même dans nos travaux, nous avons considéré 23200 battements pour effectuer la classification des mêmes classes.



CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS  
POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

---

TABLE 6.11 – Les résultats des travaux précédents (%)

Méthode	Nombre d'ECG	N	LBBB	RBBB	PVC	APB	PB	Autre	Taux
DWT/AR+SVM [79]	3403	100	98.66	100	99.66	100	100	–	99.68
DTW+NN [10]	11600	99.86	99.04	99.42	97.65	98.82	100	–	99.65
ICA+SVM [13]	4900	99.92	99.57	100	98.08	97.72	99.97	91.15	98.70
PCA+SVM [12]	34989	–	–	–	–	–	–	–	98.11
ECG+ANN [80]	300	100	100	100	99.00	–	99.77	98.18	99.49
ECG+SVM [81]	11239	93.3	–	–	99.9	–	–	99.0	97.4
Methode proposée	23200	99.91	99.83	99.70	99.56	99.52	100	–	99.75

A partir de ces taux, on a observé que nos résultats sont comparables avec ceux proposés au préalable dans la littérature. De même, si on concentre notre comparaison avec les travaux exposés dans [79, 10], on observe que les résultats de nos travaux sont légèrement meilleurs. Ces observations prouvent que les M-SVMs directes (en particulier le modèle de CS), peuvent être d'excellents modèles pour le diagnostic des maladies cardiaques.

## CHAPITRE 6. APPLICATION 2 : CLASSIFICATION DES SIGNAUX ECGS POUR LA DÉTECTION DES PATHOLOGIES CARDIOVASCULAIRES

---

### *Résumé :*

*Dans ce chapitre, nous avons proposé et implémenté une nouvelle approche basée sur la DWT et les M-SVMs directes pour discriminer entre cinq arythmies : PVC, APB, BBD, BBG, PB, plus ECG normal. Deux autres méthodes des M-SVMs indirectes ont été utilisées pour comparer leurs performances avec ceux de l'approche directe introduite : 1-contre-1, 1-contre-tous. Les M-SVMs ont été validées sur des enregistrements ECGs pris à partir de la base de données MIT-BIH.*

*Les résultats indiquent que les M-SVMs directes donnent de meilleures précisions moyennes par rapport aux M-SVMs indirectes, sachant que les meilleurs taux ont été obtenus par le modèle direct de CS. Aussi, il a été conclu à partir du post-traitement des sorties des M-SVMs en terme de probabilités a posteriori, que le nombre des FN a considérablement diminué. De plus il a été démontré à partir des intervalles de temps de calcul archivés par les M-SVMs directes, qu'ils peuvent être embarqués dans des systèmes temps réel d'aide au diagnostic des pathologies cardiaques. L'ensemble de ces observations confirment l'efficacité des M-SVMs directes dans la problématique des classifications des signaux ECGs.*

# Conclusion générale

A l'issue de nos travaux, nous voulons effectuer un retour sur la problématique et les objectifs à atteindre.

Le cerveau est le commandant suprême de l'organisme. L'état physiologique et pathologique du cerveau ainsi que les différentes activités cérébrales sont représentées sous forme de signaux. Ces signaux peuvent être enregistrés par plusieurs techniques. Parmi elles, l'EEG est de loin la technique la plus utilisée en raison de sa capacité à enregistrer les signaux cérébraux de façon non invasive et en raison de son coût réduit. Il est exploité par les scientifiques notamment pour l'étude du fonctionnement du cerveau mais aussi pour le diagnostic des atteintes mentales. L'EEG trouve aussi application dans certains domaines révolutionnaires tel que les BCIs. En effet, les BCIs offrent une opportunité de communication directe entre le cerveau et l'environnement, un espoir pour les handicapés moteur complet et partiel.

D'un autre côté, le coeur constitue aussi un organe très important pour la survie d'un être humain. Néanmoins, ce muscle peut être affecté par plusieurs pathologies, ces pathologies constituent la première cause de mortalité dans le monde. La détection précoce de ces altérations cardiaques peut considérablement réduire l'incidence des décès. Devant un tel défi, plusieurs outils performants d'exploration cardiaque ont été mis en place. Parmi les examens cardiologiques possibles, l'ECG est le plus couramment effectué, il s'est révélé être un outil d'exploration et de diagnostic puissant et irremplaçable. Son intérêt réside dans sa réalisation

à partir d'un appareillage simple, peu coûteux et d'une innocuité totale pour le patient. Cependant l'ECG n'est performant que lorsqu'il est enregistré sur une longue durée, trop contraignante pour les médecins. L'analyse et l'interprétation automatique des signaux ECGs est donc nécessaire notamment pour assurer une surveillance continue des malades, et pour préparer le travail du cardiologue lors de l'analyse des enregistrements de longue durée. C'est un domaine privilégié dans les applications de l'informatique biomédicale.

Les objectifs principaux de cette thèse étaient, tout d'abord, la proposition d'une nouvelle alternative de sélection et de regroupement de variables; cette alternative a été appliquée aux signaux EEGs. Puis, l'introduction de l'utilisation des techniques de classification récemment développées, pour l'amélioration des performances de classification : des systèmes BCIs, et des systèmes experts pour la détection des pathologies cardiovasculaires. Enfin, l'objectif final était, le post traitement des sorties des M-SVMs en terme de probabilités, afin de les munir d'une mesure de fiabilité calibrée; ainsi que, l'étude de la possibilité d'implémenter les M-SVMs directes dans des systèmes temps réel.

Pour atteindre les objectifs que nous nous sommes fixés, nous avons procédé comme suit :

- Nous avons implémenté une technique d'extraction des caractéristiques des signaux ECGs et EEGs basée sur la DWT, nous avons extrait à partir des sub-bandes obtenues des caractéristiques statistiques. Cependant, Les signaux EEGs sont obtenus à partir de plusieurs canaux, et les vecteurs caractéristiques sont construits par concaténation des descriptions qui ont été obtenues à partir des multiples canaux utilisés. Ce qui se traduit par un grand vecteur de caractéristiques, tandis que le nombre d'échantillons de l'ensemble de la formation est limité. Pour pallier à ce problème, nous avons proposé une nouvelle alternative de sélection de variables en utilisant les cartes SOMs. En effet l'algorithme SOM a la particularité de regrouper des ensembles de grandes dimensions d'une manière non supervisée.

- Nous avons introduit et étudié l'apport des quatre modèles des M-SVMs (le modèle de WW, le modèle de CS, le modèle de LLW et la M-SVM<sup>2</sup>) dans la problématique de la classification des signaux ECGs et des signaux EEG. Deux autres méthodes de M-SVMs impliquant des méthodes de décomposition ont été utilisées afin de comparer leurs performances avec celles des approches directes introduites : l'approche 1-c-1 et l'approche 1-c-tous.
- Nous avons post traité les sorties des M-SVMs directes et des M-SVMs indirects afin de générer des estimations de probabilités à posteriori des classes. Puis, nous avons fixé un seuil de confiance ; la décision d'un classifieur est retenue, si sa probabilité est supérieure à 0.75. Sinon, la décision est rejetée. Aussi, dans le but de savoir si les M-SVMs directes peuvent être embarqués dans des processus temps réel, nous avons considéré leurs temps de calcul : temps d'apprentissage et le temps de test.

Les signaux étudiés dans cette thèse sont : ceux de la base de données EEG de Keirn et Aunon pour la discrimination entre cinq tâches mentales (RelT, MulT, LetT, RotT, et ComT), et ceux de la base de données MIT-BIH pour la classification des cinq pathologies cardiaques (LBBB, RBBB, PVC, APB, et PB) plus la classe normale.

A partir de cette étude, il a été tout d'abord reporté que l'approche directe des M-SVMs généralise mieux que l'approche indirectes, et que le modèle de CS surpasse les trois autres modèles indirects. En second lieu, il a été déduit que le post traitement des sorties des différentes M-SVMs en terme de probabilités a considérablement amélioré les performances de classification, surtout du point de vu des fausses alertes. En outre, nos classifieurs rejettent toutes les décisions dont la probabilité est inférieure à 0.75 (il est préférable de dire je ne sais pas, que de faire une fausse prédiction). En troisième lieu, il a été démontré que les M-SVMs directes nécessitent un temps de calcul raisonnable et approprié pour les applications temps réel. En fin, il a été conclu à partir de l'ensemble des taux de généralisation obtenus, l'efficacité de la méthode de compression et de sélection

des variables proposées et appliquées aux signaux EEGs.

Pour conclure, nous considérons, de notre point de vue, que notre travail ouvre de nombreuses perspectives, à savoir :

- Enrichir les bases de données. Notamment, pour le diagnostic automatique des atteintes mentales (par exemple épilepsie), et pour la détection d'autres pathologies cardiaques.
- Fusion des quatre modèles des M-SVMs directes, et effectuer une comparaison des M-SVMs avec d'autres techniques de classification.
- Essayer d'autres techniques d'extraction de paramètres ainsi que d'autres techniques de compression de données des signaux EEGs, et effectuer une comparaison avec la méthode proposée dans cette thèse.
- Envisager des implémentations en pratique réelle.

# Bibliographie

- [1] J. De Oliveira. *Brain and Mind Electronic Magazine on Neuroscience*. Consciousness, 5, 1998.
- [2] R.A. Salido Ruiz. *Problemes inverses contraints en EEG : applications aux potentiels absolus et a l'influence du signal de reference dans l'analyse de l'EEG*. Traitement du signal et de l'image. Universite de Lorraine, France., 2012.
- [3] N. Spurston. Pyramidal neuron. *Scholarpedia*, 4(5) :6130, 2009.
- [4] <http://www.doctissimo.fr/html/sante/atlas/fiches-corps-humain/coeur.htm>.
- [5] S. Vaid, P. Singh, and C. Kaur. Eeg signal analysis for bci interface : A review. *IEEE Conference on Advanced Computing and Communication Technologies*, pages 143–147, 2015.
- [6] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, Peckham P.H, G. Schalk, E. Donchin, C.J. Robinson L.A. Quatrano, and T.M. Vaughan. Brain-computer interface technology : A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8 :164–173, 2000.
- [7] M. Tolic and F. Jovic. Classification of wavelet transformed eeg signals with neural network for imagined mental and motor tasks. *International Journal of Neural Systems*, 45(1) :130–138, 2013.
- [8] E. Estrada, H. Nazeran, G. Sierra, F. Ebrahimi, and S.K. Setarehdan. Wavelet-based eeg denoising for automatic sleep stage classification. In *CO-NIELECOMP'11*, pages 295–298, 2011.

- [9] A. Ahmadi, R. Jafari, and J.J. Hart. Light-weight single trial eeg signal processing algorithms : Computational profiling for low power design. In *IEMBS'11*, pages 4426–4430, 2011.
- [10] S-N. Yu and Y-H. Chen. Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Journal Pattern Recognition Letters*, 28(10) :1142–1150, 2007.
- [11] L. Zhiwei and S. Minfen. Classification of mental task eeg signals using wavelet packet entropy and svm. In *ICEMI'07*, pages 3–907–3–910, 2007.
- [12] R.J. Martis, U.R. Acharya, K.M. Mandana, A.K. Ray, and C. Chakraborty. Application of principal component analysis to eeg signals for automated diagnosis of cardiac health. *Expert Systems with Applications*, 39 :11792–11800, 2012.
- [13] S-N. Yu and K-T. Chou. Evaluation of three machine learning algorithms as classifiers of premature ventricular contractions on eeg beats. *Expert Systems with Applications*, 36 :2088–2096, 2009.
- [14] J. Weston and C. Watkins. Multi-class support vector machines. Technical Report CSD-TR-98-04, Royal Holloway, University of London, Department of Computer Science, 1998.
- [15] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2 :265–292, 2001.
- [16] Y. Lee, Y. Lin, and G. Wahba. Multicategory support vector machines : Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, 99(465) :67–81, 2004.
- [17] Y. Guermeur and E. Monfrini. A quadratic loss multi-class svm for which a radius-margin bound applies. *INFORMATICA*, 22(1) :73–96, 2011.



- 
- [18] J.C. Platt. Probabilities for SV machines. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, chapter 5, pages 61–73. The MIT Press, Cambridge, MA, 2000.
- [19] H.T. Lin, C.J. Lin, and R.C. Weng. A note on platt’s probabilistic outputs for support vector machines. Technical report, University of Taiwan, 2003.
- [20] O. Milgram, M. Cheriet, and R. Sabourin. *"One Against One" or "One Against All" : Which One is Better for Handwriting Recognition with SVMs ?*. Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France), 2006.
- [21] Y. Guermeur and F. Thomarat. Estimating the class posterior probabilities in protein secondary structure prediction. *PRIB*, pages 261–271, 2011.
- [22] Z.A. Keirn. *Alternative Modes of Communication Between Man and Machines*. Masters thesis, Electrical Engineering Departement, Purdue University, USA, 1998.
- [23] <http://ecg.mit.edu>.
- [24] F.J. Gray. *Anatomy for the medical clinician*. first edition, Shannon Books Pty Ltd, Victoria, Australia., 2002.
- [25] D. Purves, G.J. Augustine, D. Fitzpatrick, L.C. Katz, A.S. Lamantia, and J.O. McNamara. *Neuroscience, Sinauer associates*. third edition, Inc. Publishers, Sunderland, Massachusetts, USA., 2004.
- [26] N.R. Carlson. *Foundations of physiological psychology*. 5th ed., Boston, Mass. London : Allyn and Bacon., 2002.
- [27] H.L. Atwood and W.A. MacKay. *Essentials of neurophysiology*. oronto, Philadelphia : B.C. Decker., 1989.
- [28] N.R. Carlson. *Structure and Functions of the Nervous System.*, volume 5. Foundations of physiological psychology. Boston, Mass. London : Allyn and Bacon., 2002.

- 
- [29] F. Lopes da Silva. Electrical potentials. In *Encyclopedia of the human brain*, pages 147–167. V.S. Ramachandran, 2002.
- [30] N. Spurston. Pyramidal neurons : Dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3) :206–221, 2008.
- [31] A. Crouzeix. *Méthodes de localisation des générateurs de l'activité électrique cérébrale à partir de signaux électro et magnéto encéphalographiques*. Processus Mentaux et Activation Cérébrale, INSERM U280, LYON, France., 2001.
- [32] T.F. Collura. History and evolution of electroencephalographic instruments and techniques. *Clinical Neurophysiology*, 10(4) :476–504, 1993.
- [33] N. Hazarika, J.Z. Chen, A.C.Tsoi, and A.Sergejew. Classification of eeg signals using the wavelet transform. *Signal Process*, 59(1) :61–72, 1997.
- [34] H. Adeli, Z. Zhou, and N. Dadmehr. Analysis of eeg records in an epileptic patient using wavelet transform. *J. Neurosci. Methods*, 123(1) :69–87, 2003.
- [35] H.H. Jasper. Report of the committee on methods of clinical examination in electroencephalography : The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10(2) :371–375, 1958.
- [36] B.J. Fisch. *EEG premier : Basic principles of digital and analog EEG*. (third edition), Elsevier publication., 1999.
- [37] J.R. Wolpaw N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113 :767–791, 2002.
- [38] T.M. Vaughan, W.J. Heetderks, L.J. Trejo, W.Z. Rymer, M. Weinrich, M.M. Moore, A. Kubler, B.H. Dobkin, N. Birbaumer, E. Donchin, E.W. Wolpaw, and J.R. Wolpaw. Brain-computer interface technology : a review the second international meeting. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2) :94–109, 2003.

- 
- [39] B. Graimann, B. Allison C. Mandel, T. Luth, D. Valbuena, and A. Gra-ser. *Non-invasive brain-computer interfaces for semi-autonomous assistive devices*. Robust Intelligent Systems, Springer London Ltd., pp. 113-137, 2008.
- [40] B.Z. Allison, E.W. Wolpaw, and J.R. Wolpaw. Brain computer interface systems : progress and prospects. *In. E.Poll (Ed.), British Review of Medical Device*, 4(4) :463–474, 2007.
- [41] S.G. Mason and G.E. Birch. A general framework for brain-computer interface design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(1) :70–85, 2003.
- [42] A. Bashashati, M. Fatourechi, R.K. Ward, and .E. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2) :R35–57, 2007.
- [43] B. Blankertz, G. Dornhege, M. Krauledat, K. Muller, and G. Curio. The non-invasive berlin brain-computer interface : fast acquisition of effective performance in untrained subjects. *Neuroimage*, 37 :539–550, 2007.
- [44] M. Vellistte, S. Perel, M.C. Spalding, A.S. Whitford, and A.B. Schwartz. Cortical control of a prosthetic arm for self-feeding. *Nature.*, 453 :1098–1101, 2008.
- [45] A. Waller. *An introduction to human physiology*. Longmans Green, 1893.
- [46] J.M. Rawles, G.R. Pai, and S.R. Reid. A method of quantifying sinus arrhythmia : parallel effect of respiration on p-p and p-r intervals. *Clin Sci (Lond).*, 76 :103–108, 1989.
- [47] W. Einthoven. *An introduction to human physiologyThe galvanometric registration of the human electrocardiogram, likewise a review of the use of capillary-electrometer in physiology*. Cardiac Classic : Willius FW-CV Mosby, 1941.
- [48] M. Kom. *Electrocardiographie haute amplification*. Thèse de doctorat d’Etat, Université de Ydé I ; spécialité Physique Appliquée., 2001.

- [49] C. Cortes and V.N. Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.
- [50] Y. Guermeur. A generic model of multi-class support vector machine. *International Journal of Intelligent Information and Database Systems*, 6(6) :555–577, 2012.
- [51] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, 2004.
- [52] B. Schölkopf and A.J. Smola. *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, Cambridge, MA, 2002.
- [53] A. Berline and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, Boston, 2004.
- [54] L. Yufeng. Fisher consistency of multicategory support vector machines. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, pages 289–296, 2007.
- [55] C-W. Hsu and C-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13 :415–425, 2002.
- [56] D. Price, S. Knerr, L. Personnaz, and G. Dreyfus. Pairwise neural network classifiers with probabilistic outputs. *Neural Information Processing Systems*, pages 1109–1116, 1995.
- [57] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26 :451–471, 1998.
- [58] T. Hamamura, H. Mizutani, and B.Irie. A multiclass classification method based on multiple pairwise classifiers. *International Conference on Document Analysis and Recognition*, pages 809–813, 2003.
- [59] C.W. Anderson, 1989. <http://www.cs.colostate.edu/eeg/main/data/>.

- [60] N. Liang, P. Saratchandran, G. Huang, and N. Sundararajan. Classification of mental tasks from eeg signals using extreme learning machine. *International Journal of Neural Systems*, 16(1) :29–38, 2006.
- [61] Z.A. Keirn and J.I. Aunon. A new mode of communication between man and his surroundings. *IEEE Transactions on Biomedical Engineering*, 37(12) :1209–1214, 1990.
- [62] R. Palaniappan. Utilizing gamma band to improve mental task based brain-computer interface designs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(3) :299–303, 2006.
- [63] P.F. Diez, V. Mut, E. Laciari, A. Torres, and E. Avila. Application of the empirical mode decomposition to the extraction of features from eeg signals for mental task classification. In *EMBC'09*, pages 2579–2582, 2009.
- [64] H. Hariharan, V. Vijejan, R. Sindhu, P. Divakar, A. Saidatul, and Z. Yacob. Classification of mental tasks using stockwell transform. *ELSEVIER, Computers and Electrical Engineering*, 40 :1741–1749, 2014.
- [65] M. Hendel, A. Benyettou, and F. Hendel. Hybrid self organizing map and probabilistic quadratic loss multi-class support vector machine for mental tasks classification. *Informatics in Medicine Unlocked (ELSEVIER)*, 4 :1–9, 2016.
- [66] A. Majkowski, M. Kolodziej, and R.J. Rak. Implementation of selected eeg signals processing algorithms in asynchronous bci. In *MeMeA'12*, pages 1–3, 2012.
- [67] H.A. Shedeed. A new method for person identification in a biometric security system based on brain eeg signal processing. In *WICT'11*, pages 1205 – 1210, 2011.
- [68] A.N. Akansu, W.A. Serdijn, and I.W. Selesnick. Emerging applications of wavelets : A review. *PhysicalCommunication*, 3 :1–18, 2010.
- [69] J. Olkkonen. *Discrete walvelet transforms theory and applications*. Janeza Trdine 9, 51000 Rijeka, Croatia, 2011.

- [70] S.D. Katebia M. Sabetia, R. Boostania and G.W. Priceb. Selection of relevant features for eeg signal classification of schizophrenic patients. *ELSEVIER, Biomedical Signal Processing and Control*, 2(2) :122–134, 2007.
- [71] C.W. Anderson, S.V. Devulapalli, and E.A. Stolz. Eeg signal classification with different signal representation. In F. Girosi, J. Makhoul, E. Manolakos, and E. Wilson, editors, *IEEE Service Centre, Piscataway, NJ*, volume V, pages 475–483, 1995.
- [72] C.M. Bishop. *Neural network for pattern recognition*. Oxford University Press, 1995.
- [73] T. Kohonen. Physiological interpretation of the self-organizing map. *Neural Network*, 6 :895–905, 1993.
- [74] T. Kohonen. *Self-organizing map*. Second edition, Springer, 1997.
- [75] S. Nasehi and H. Pourghassem. Mental task classification based on hmm and bpnn. In *CSNT'13*, pages 210–214, 2013.
- [76] R. Chai, S.H. Ling, G.P. Hunter, Y. Tran, and H.T. Nguyen. Brain-computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization. *IEEE Journal of Biomedical and Health Informatics.*, 8(5) :1614–1624, 2014.
- [77] F. Lauer and Y. Guermeur. MSVMpack : a multi-class support vector machine package. *Journal of Machine Learning Research*, 12 :2269–2272, 2011. <http://www.loria.fr/~lauer/MSVMpack>.
- [78] Y. Bennani and F. Bossaert. Predictive neural networks for traffic disturbance detection in the tele-phonenetwork. *IEEE Computational Engineering in System Applications.*, pages 1085–1090, 1996.
- [79] Q. Zhao and L. Zhang. Ecg feature extraction and classification using wavelet transform and support vector machines. *Neural Networks and Brain*, 4 :1089–1092, 2005.

- [80] S.K. Dash and G.S. Rao. Robust multiclass ecg arrhythmia detection using balanced trained neural network. *Neural Networks and Brain*, 5 :186–191, 2016.
- [81] M.M. Casas, R.L. Avitia, M.A. Reyna, and A. Cardenas. Evaluation of three machine learning algorithms as classifiers of premature ventricular contractions on ecg beats. *Global Medical Engineerin*, 1 :36–42, 2016.
- [82] P. De Chazal. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7) :1196–1206, 2004.
- [83] Y.H Hu, S.H Palreddy, and W Tompkins. A patient adaptable ecg beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, 44 :891–900, 1997.
- [84] T. Liu, Y. Si, D. Wen, M. Zang, and L. Lang. Dictionary learning for vq feature extraction in ecg beats classification. *Expert Systems With Applications*, 53 :129–137, 2016.
- [85] L. Khadra, A. Al-Fahoum, and S. Binajjaj. A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques. *IEEE Transactions on Biomedical Engineering*, 52(11) :1840–1845, 2005.
- [86] K.. Minami, H. Nakajima, and T. Toyoshima. Real-time discrimination of ventricular tachyarrhythmia with fourier- transform neural network. *IEEE Transactions on Biomedical Engineering*, 46 :179–185, 1999.
- [87] S. Osowski and T.H Linh. Ecg beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, 48 :1265–1271, 2001.
- [88] M. Hendel, A. Benyettou, and F. Hendel. Heartbeats arrhythmia classification using quadratic loss multi-class support vector machines. *International Review on Computers and Software*, 11(1) :49–55, 2016.
- [89] M. Hendel and A. Benyettou. Heartbeats arrhythmia classification using probabilistic multi-class support vector machines : A comparative study. *International Review on Computers and Software*, 11(2) :1293–1297, 2016.

- [90] M. Hendel, A. Benyettou, F. Hendel, and H. Khelil. Automatic heartbeats classification based on discrete wavelet transform and on a fusion of probabilistic neural networks. *Journal of Applied Sciences*.
- [91] and W.J. Tompkins J.éPan. A real-time qrs detection algorithm. *IEEE Trans Biomed Eng*, 32(3) :230–236, 1985.
- [92] S-N. Yu and K-T. Chou. Integration of independent component analysis and neural networks for ecg beat classification. *Expert Systems with Applications*, 34 :2841–2846, 2008.