

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



École Nationale Polytechnique
Département d'Electronique
Laboratoire Signal et Communication



Projet de fin d'études
En vue de l'obtention du diplôme
D'Ingénieur d'Etat en Electronique

Etudié par :
ACHAIBOU Amina

Thème :

**Contribution au développement de
différents systèmes de classification de
signaux biomédicaux**

Soutenu devant le jury composé de :

TERRA Zidane
Hicham BOUSBIA-SALAH
Assya BOUSBIA- SALAH
ADNANE Mourad

président du jury
promoteur
co-promoteur
examineur

Juin 2014

Dédicaces

*Je dédie ce travail à mes parents,
à mes frères,
à tous les membres de ma famille,
et à tous mes amis.*

ملخص:

في هذا العمل، اقترحنا طريقة للتصنيف التلقائي لإشارات التخطيط الدماغي إلى مجموعتين (أشخاص مرضى وآخرون عاديون). وفي هذا السياق، مجموعة من الميزات الإحصائية استخرجت من معاملات الموجات (DWT) ولقد تم استعمال تحليل التمايز الخطي (LDA) لتخفيض حجم المعلومات. في الأخير استخدمنا آلات إشعاعات الدعم وشبكة العصبونات الاصطناعية للتصنيف.

الكلمات المفتاحية: إشارات التخطيط الدماغي (EEG) تحويل الموجات المتقطع (DWT) استخراج المكونات الخصوصية ، تحليل التمايز الخطي ، آلة دعم الأشعة (SVM) وشبكة العصبونات الاصطناعية للتصنيف (FFBPNN)

Résumé :

Dans ce travail, nous avons proposé plusieurs méthodes de classification automatique de signaux EEG en cas normaux et cas épileptiques. Dans ce cadre, des caractéristiques statistiques ont été extraites à partir des coefficients de (DWT). L'Analyse Discriminante Linéaire (ADL) a été utilisée pour réduire la dimension des données. Enfin, on a utilisé les machines à vecteurs de supports (SVM) et les réseaux de neurones pour la classification.

Mots clés : Électroencéphalogramme (EEG), transformée en ondelette(DWT), extraction du paramètre Caractéristique, Analyse discriminante linéaire (LDA), machine à vecteurs de support (SVM), réseaux de neurones (FFBPNN).

Abstract:

In this work, we have proposed different methods of automatic classification of EEG signals that are classified into two classes: normal and epileptic subjects. In this study, statistical characteristics were computed from (DWT) coefficients. The linear discriminant analysis (LDA) were used for dimensionality reduction. Finally, support vector machine (SVM) and neural network have been used to classify EEG signals.

Keywords: electroencephalogram (EEG), wavelet transform (WT), feature extraction, linear discriminant analysis (LDA), Support Vector Machine (SVM), neural network (FFBPNN).

Table des matières

Liste des figures	i
Liste des tableaux	ii
Chapitre 01: Contexte médical	3
1.1. Introduction	3
1.2. Notion générale sur le cerveau	3
1.2.1. Anatomie de cerveau.....	3
1.2.2. La physiologie cérébrale	4
1.2.3. Activité éeectique cérébral.....	6
1.3. Conditions de mesure et Méthodes d'exploration de l'activité électrique	8
1.4. Electroencéphalogramme (EEG).....	10
1.4.1. Historique.....	10
1.4.2. Définition	11
1.4.3. Les différents rythmes du signal EEG ($\alpha, \beta, \gamma, \theta, \Delta, \mu$) :.....	12
1.5. Procédure d'enregistrement des signaux EEG	14
1.5.1. Electrodes.....	14
1.5.2. Montage et système d'emplacement.....	15
1.5.3. L'acquisition des signaux EEG.....	18
1.6. Conclusion	19
Chapitre 02: Outils de traitement des signaux EEG (Les méthodes d'analyse temps fréquence)...20	
2.1. Introduction	20
2.2. La transformée de Fourier :	20
2.3. La transformée de Fourier fenêtrée (STFT) :	22
2.3.1. Définition et principe de la STFT :	22

2.3.2. Principe d'incertitude et problème posé par la STFT :	24
2.4. Transformée en ondelette :	27
2.4.1. Historique :	27
2.4.2. Définition :	27
2.4.3. Résolution de problème de la STFT :	28
2.4.4. Principe de la transformée en ondelettes.....	29
2.4.5. Les différents types d'ondelettes	31
2.4.6. Transformée en ondelette discrète	31
2.4.6.1 Concept de base de la DWT.....	34
2.4.6.2 Decomposition : coefficients d'approximation et de details	35
2.4.6.3 Decomposition avec sous échantillonnage	38
2.4.6.4 Decomposition multi-niveau	39
2.4.6.5 le nombre de niveaux de la decomposition.....	40
2.4.7 Domaines d'application des onedelletes	41
2.4.8 Conclusion :.....	41
Chapitre 03:Extractions des paramètres caractéristiques et Analyse des données.....	42
3.1. Introduction	42
3.2. L'extraction des paramètres caractéristiques.....	42
3.2.1. Paramètres caractéristiques Non-statistiques.....	44
3.2.2. Paramètres caractéristiques statistiques	44
3.3. Réduction de la dimension par analyse discriminantes linéaire	45
3.3.1.Le modèle de base de l'analyse discriminante linéaire (LDA) (point de vue intuitif)	46
3.3.2. Le modèle de base de l'analyse Discriminante Linéaire (LDA) (point de vue mathématique).....	46
3.3.3. Application de l'Analyse discriminante pour deux groupes.....	50

3.3.4. Analyse discriminante multiple	52
3.4. Conclusion	52
Chapitre 04: Méthode de classification Machine à vecteurs support (SVM) et Réseaux de neurones	53
4.1. Introduction :	53
4.2. Notion de classification :	53
4.3. Notion d'apprentissage :	54
4.4. Classification par la machine à vecteurs support (SVM) :	55
4.4.1. Concept de la machine à vecteurs support (SVM) :	55
4.4.2. Cas des données linéairement séparables :	57
4.4.3. Cas des données non linéairement séparables :	59
4.4.4. Cas d'utilisation d'un classificateur non linéaire :	62
4.5. Classification par les réseaux de neurones artificiels :	63
4.5.1. Fonctionnement général d'un réseau de neurone formel.	63
4.5.2. Architectures des réseaux de neurones.	65
4.5.3. Les réseaux « feedforward » à rétro propagation.....	66
4.6. Les différentes applications de RNA	69
4.7. Les paramètres de performance	70
4.8. Conclusion	72
Chapitre 05: Résultats et évaluation des performances	73
5.1. Introduction :	73
5.2. Système automatique de classification des signaux EEG :	73
5.3. Signaux EEG :	75
5.4. Application de la DWT aux signaux EEG :	76
5.5. Extraction des paramètres caractéristiques :	78
5.6. Réduction de dimensions par LDA :	82

5.7. Classification par SVM :	83
5.8. Classification par les réseaux neurones :	89
5.9. Conclusion :	94
Conclusion générale	95
Bibliographie.....	97

Liste des figures

Figure 1.1 : Représentation des deux hémisphères de cerveau à gauche et des différents lobes qui le constituent à droite	4
Figure 1.2 : Modèle d'un neurone biologique.....	5
Figure 1.3 : Connexion entre deux neurones.....	5
Figure 1.4 : Représentation d'une cellule pyramidale.....	6
Figure 1.5 : Activité électrique du neurone.....	7
Figure 1.6 : Déclenchement d'un potentiel post-synaptique (Image de la synapse adaptée d'après [4])	8
Figure 1.7 : Enregistrement EEG	9
Figure 1.8 : Schéma comparatif des résolutions temporelles et spatiales et du caractère invasif des différentes techniques d'imagerie cérébrale (EEG, IRM, IRMf, MEG, SEEG, TDM, TEMP, TEP) [6]	10
Figure 1.9 : Exemple d'un enregistrement d'EEG	11
Figure 1.10 : Les rythmes corticaux [6]	13
Figure 1.11 : Photographie d'électrode sous forme de coupe.....	14
Figure 1.12 : Photographie d'électrode d'Ag/AgCl	16
Figure 1.13 : Système de positionnement des électrodes 10–20 (21 électrodes).....	18
Figure 1.14 : Acquisition du signal EEG	18
Figure 2.1 : Un signal et son spectre de Fourier.....	20
Figure 2.2 : Illustration de la transformée de fourier fenêtrée [20]	22
Figure 2.3 : Découpage en atomes temps-fréquence : (résolution temps fréquence par la TF fenêtrée).....	23
Figure 2.4 : Résultat de la STFT appliqué au signal représenté en figure 2.2 en utilisant des fenêtres de tailles différentes [20]	23
Figure 2.5 : Application la STFT a une sinusoïde.....	24
Figure 2.6 : Illustration de problème de résolution de le STFT	26

Figure 2.7 : Découpage en atomes temps-fréquence : (résolution temps fréquence par la WT)	28
Figure 2.8 : La différence entre la STFT et la WT	28
Figure 2.9 : La fonction analysante dans la TF est l'onde sinusoïdale qui est à la gauche et la fonction analysante dans la WT est l'ondelette qui à droite	29
Figure 2.10 : Illustration de la transformée en ondelettes [20]	30
Figure 2.11 : La transformé en ondelette du signal représenté en (fig 2.7) [20]	30
Figure 2.12 : Principe de la transformée en ondelette	31
Figure 2.13 : Quelques familles d'ondelette	32
Figure 2.14 : Spectre des ondelettes orthogonales	35
Figure 2.15 : La fonction d'échelle ϕ et la fonction d'ondelette ψ de l'ondelette mère de Daubechies db4	36
Figure 2.16 : La décomposition par banc de filtre	36
Figure 2.17 : Duplication de nombres d'échantillons lors de la décomposition standard	37
Figure 2.18 : Décomposition a un seul niveau d'un signal non stationnaire	37
Figure 2.19 : Décomposition par banc de filtre avec sous échantillonnage	37
Figure 2.20 : Décomposition d'un signal non stationnaire à trois niveaux	38
Figure 2.21 : Arbre de décomposition en ondelettes à cinq niveaux	39
Figure 2.22: Pavage temps fréquence correspondant à chacun des niveaux de décomposition	40
Figure 3.1 : Exemple de décomposition de signal EEG d'un sujet normal	43
Figure 3.2 : Exemple de décomposition de signal EEG d'un sujet épileptique	43
Figure 3.3 : Les entrées et sorties du processus de transformation par LDA	46
Figure 3.4 : Tableau centré réduit en LDA	47
Figure 3.5 : Processus de la Discrimination linéaire	49
Figure 3.6 : Séparation des nuages (droite de Fisher)	50
Figure 4.1 : Exemple de classification binaire linéaire et non linéaire dans \mathbb{R}^2 .	56

Figure 4.2 : Hyperplan linéaire séparant avec la marge Δ les exemples positifs des exemples négatifs	56
Figure 4.3 : les différents hyperplans séparateurs	57
Figure 4.4 : Espace de redescription	60
Figure 4.5 : La transformation linéaire des données permet une séparation linéaire dans un nouvel espace. Adapté de [52].	60
Figure 4.6 : (a) représente le neurone biologique et (b) représente la structure de base d'un neurone artificiel	63
Figure 4.7 : Illustration de fonctionnement de neurone biologique [53]	65
Figure 4.8 : Les réseaux les plus utilisés	66
Figure 4.9 : Architecture d'un réseau de neurones multicouches.	68
Figure 4.10 : La courbe ROC	72
Figure 5.1 : Structure du système de classification automatique de signaux EEG	74
Figure 5.2 : Représentation temporelle des signaux EEG.....	75
Figure 5.3 : Décomposition sans décimation de deux signaux EEG du Set A et Set B.....	77
Figure 5.4 : Décomposition avec décimation de deux signaux EEG du Set A et Set B	77
Figure 5.5 : Représentation d'énergie des différents niveaux	80
Figure 5.6 : Représentation d'entropie de niveaux 4 et 5.....	80
Figure 5.7 : Représentation d'écart type de niveaux 4 et 5	81
Figure 5.8 : Représentation des valeurs minimales de niveau 4 et 5	81
Figure 5.9 : Représentation des valeurs maximales de niveaux 4 et 5.....	82
Figure 5.10 : Réduction de dimension par LDA (paramètre d'énergie)	82
Figure 5.11 : Les performances de classification par SVM avec LDA et sans LDA.....	83

Liste des tableaux

Tableau 4.1 : Analogie entre les neurones biologiques et artificiels	64
Tableau 4.2 : Table de diagnostic (confusion)	71
Tableau 5.1 : Gamme de fréquence.....	76
Tableau 5.2 : Les valeurs des paramètres caractéristiques	78
Tableau 5.3 : Performances de classification par SVM.....	85
Tableau 5.4 : Les performances de classification par FFBPNN pour les trois groupes	90
Tableau 5.5 : Classification par SVM	94

Introduction générale

L'épilepsie est une maladie chronique caractérisée par la répétition de manifestations cliniques paroxystiques. Ces crises se traduisent par une modification brutale de la conscience et du comportement du malade. Elles sont la conséquence d'un dysfonctionnement cérébral généralisé ou localisé dû à une décharge électrique anormale. L'épilepsie peut être d'origine génétique ou lésionnelle. Toutes les lésions cérébrales, tumeurs, dysplasie, infections,... peuvent provoquer des crises d'épilepsie.

L'électroencéphalogramme (EEG), est un examen permettant d'enregistrer l'activité électrique produite par les neurones du cerveau. C'est un examen totalement indolore et non invasif.

L'EEG est utile dans le bilan des malaises, pour le diagnostic et particulièrement dans le suivi de l'épilepsie, mais également pour d'autres pathologies neurologiques comme la maladie d'Alzheimer, troubles du sommeil...

A cet effet, nous avons contribué à l'élaboration d'algorithmes d'aide au diagnostic, par une identification des signaux EEG en vue d'une classification automatique.

Les informations relatives à l'état du cerveau se reflètent par son activité électrique en fonction du temps et la fréquence ; par conséquent, les études exigent des méthodes capables de décrire la variation qualitative du signal en temps et en fréquence. La transformée en ondelettes est une méthode puissante qui permet d'effectuer une analyse des structures locales d'un signal avec un zoom en fonction de l'échelle considérée.

Les facteurs de translation et de dilation sont utilisés dans la transformé en ondelette afin d'obtenir une famille d'éléments temps-fréquence.

L'extraction des paramètres pertinents uniquement, constitue une étape primordiale dans le diagnostic et/ou la classification des signaux EEG. Pour cela, on peut utiliser plusieurs paramètres pour caractériser les différentes décompositions ; dans cette étude nous avons opté pour les paramètres statistiques suivants : la variance, écart type, la gamme, le minimum, le maximum, entropie, énergie, ...ces valeurs pertinentes vont décrire les signaux EEG, l'étape qui suit cette extraction est la réduction de dimension pour adapter la méthode qui convient à nos données. Et à la fin, vient l'étape de classification.

Dans le cadre de ce projet de fin d'études, on s'est intéressés à l'analyse discriminante de Fisher des signaux EEG en vue d'une classification automatique.

On a utilisé dans un premier temps deux banques de données différentes : la première à partir des signaux EEG normaux (set A) et la deuxième à partir des signaux EEG pathologiques (set E).

Les paramètres représentant les signaux de ces deux banques de données seront prétraités par application de la transformée en ondelettes.

La pertinence de ces paramètres sera établie grâce à la conception d'un système exploitant la méthode de l'analyse de discrimination de Fisher des deux banques de données par application de la classification par SVM et FFBPNN.

Ce mémoire de fin d'études est structuré comme suit :

Le premier chapitre de ce mémoire décrit le contexte et la description de l'épilepsie (définition, classification, traitement) et de l'électroencéphalographie (EEG) et ses activités épileptiques.

Le second chapitre est consacré à l'étude de l'une des principales techniques d'analyse temps-fréquence : la transformée en ondelette et la description succincte des ondelettes utilisées, pour l'analyse des signaux électroencéphalographiques.

Dans le troisième chapitre on présentera la méthode de réduction de données : l'analyse discriminante linéaire (LDA).

Dans le quatrième chapitre on présentera la méthode de classification automatique supervisée telle que Machine à Vecteurs de Support (SVM) et celle de réseau de neurone (FFBPNN).

Le cinquième chapitre illustre les tests et résultats obtenus : le calcul des performances et l'évaluation des résultats de classification.

Chapitre 01

Contexte médical

1.1 Introduction

Le cerveau est le centre de contrôle de l'être humain qui a une forte activité déroulante en son sien. Cette activité est caractérisée par des flux électriques et peut être observée à la surface du crâne à l'aide des électrodes qui permettent de mesurer le potentiel électrique en point donné. Ce qu'on appelle l'électroencéphalogramme (EEG) est ainsi obtenu. Ce dernier est une méthode d'exploration fonctionnelle du cerveau qui peut identifier les patients qui ont des crises infra cliniques et surtout reste très utile dans le diagnostic et la prise en charge des états de mal épileptiques (EME).

Dans ce chapitre on présentera un panorama général du contexte médical en commençant par la physiologie de cerveau, ensuite les méthodes d'acquisition et d'exploration de son activité électrique.

1.2 Notion générale sur le cerveau

1.2.1 Anatomie de cerveau

Le cerveau est l'étage le plus élevé dans la hiérarchie fonctionnelle du système nerveux central, il désigne le tissu que l'on trouve dans le crâne, et il possède deux moitiés relativement symétriques appelées des hémisphères, l'un à droite et l'autre à gauche (figure 1.1(a)). Toute la couche externe du cerveau, la substance grise qui présente des circonvolutions complexes en périphérie, est appelée le cortex cérébral. Le cortex de chaque hémisphère est divisé en quatre lobes (figure 1.1(b)) : le lobe pariétal, le lobe occipital, le lobe frontal et le lobe temporal dont l'appellation est directement inspirée du nom des os crâniens qui les recouvrent [1] ; et on trouve que ces lobes sont impliqués dans différentes fonction [2] :

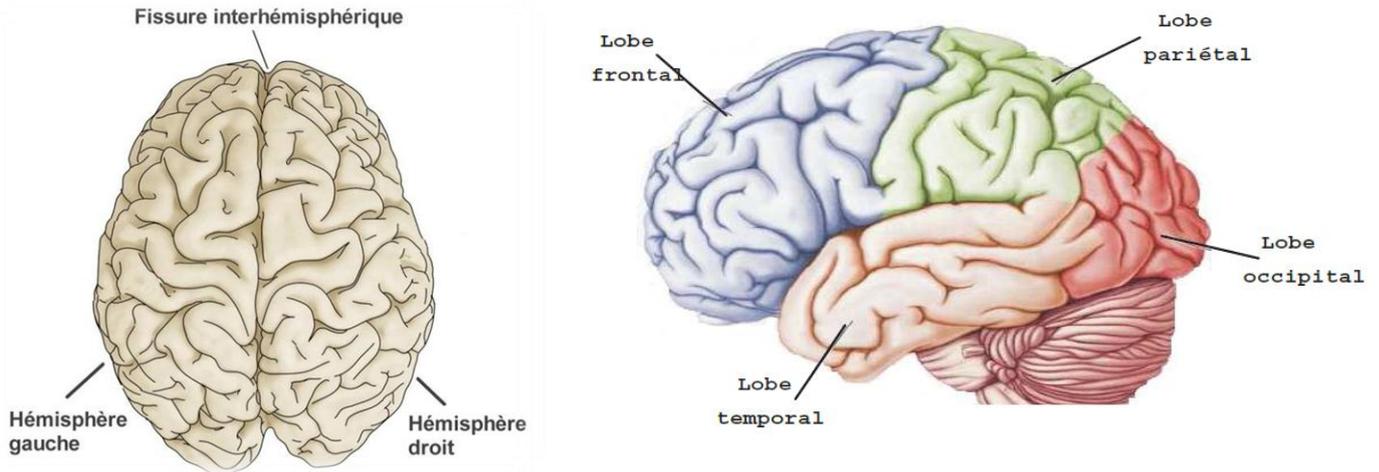


Figure 1.1 : représentation des deux hémisphères de cerveau à gauche et des différents lobes qui le constituent à droite.

- Les lobes occipitaux, localisés à l'arrière du cortex, sont spécialisés dans la vision.
- Les lobes pariétaux sont situés à l'avant des lobes occipitaux. Ils sont impliqués dans le sens du toucher, dans la détection des mouvements dans l'environnement et la localisation des objets dans l'espace.
- Les lobes frontaux sont impliqués dans le mouvement, l'attention, la planification, les compétences sociales, la pensée abstraite, la mémoire et certains aspects de la personnalité.
- Les lobes temporaux sont importants dans l'audition, le langage et la reconnaissance visuelle des objets.

1.2.2 La physiologie cérébrale

Les cellules constituant l'unité fonctionnelle du système nerveux sont les neurones. Le nombre total de neurones du cerveau humain est estimé de 100 milliards (10^{11}) cellules [3]. Elles sont en charge de traitement de l'information au sein du cerveau, elles assurent la génération et la transmission d'un signal bioélectrique que l'on nomme l'influx nerveux.

Les neurones ont une organisation cellulaire de base qui ressemble à celle des autres cellules. Ils s'en distinguent toutefois nettement par leur spécialisation pour la communication intercellulaire. Elles sont divisées en plusieurs régions ayant chacune une fonction propre (figure 1.2).

Les neurones sont caractérisées par :

- Un corps cellulaire contenant le noyau et la plupart des organites responsables de l'entretien de la cellule,
- Un long prolongement cellulaire (l'axone) qui est issu de la cellule, et s'étendant parfois sur une longue distance et ainsi servant à transmettre des signaux à d'autres cellules,

- de nombreux prolongements courts (les dendrites) qui accroissent la surface disponible pour des connexions avec les axones d'autres neurones,
- Des jonctions spécialisées (les synapses) qui relient entre les axones de différentes cellules neuronales, permettant ainsi une communication cellulaire directe voir (figure 1.3).

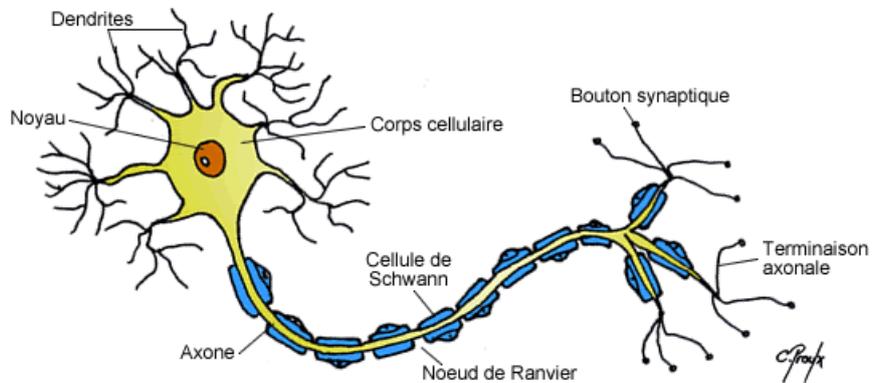


Figure 1.2 : modèle d'un neurone biologique

Les neurones ne sont pas indépendants les uns des autres. Ils établissent entre eux des liaisons et forment des chaînes de neurones constituant un fantastique réseau. Elles sont des cellules excitables qui reçoivent et transmettent des signaux de nature électrochimique grâce à leur propriété physiologiques qui se présentent dans la conductivité, leur capacité de répondre aux stimulations et de les convertir en impulsions nerveuses.

Chaque neurone peut recevoir des informations de nombreux autres neurones et en transmettre à une multitude d'autres. Ce transfert d'information d'un neurone à l'autre s'effectue au niveau des **synapses**.

La synapse est l'unité structurale et fonctionnelle de connexions entre deux neurones, il assure la transmission de l'influx nerveux.

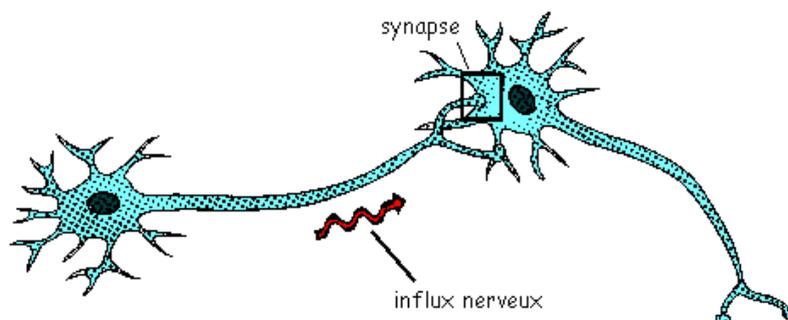


Figure 1.3 : Connexion entre deux neurones

En se fondant sur les mécanismes de transmission qu'elles utilisent, on peut en distinguer deux types

- La **synapse chimique** qui utilise des neurotransmetteurs pour transmettre l'information,
- La **synapse électrique** où le signal est transmis électriquement par l'intermédiaire d'une jonction communicante.

1.2.3 Activité électro-cérébrale

Il existe plusieurs sortes de neurones, la plupart du temps, ce sont surtout les cellules pyramidales qui génèrent l'activité électrique du cerveau. En effet, ces cellules ont une forme très asymétrique avec une dendrite apicale située dans la couche superficielle du cortex et à l'opposé le soma plus profondément enfoncé [5] (figure 1.4).

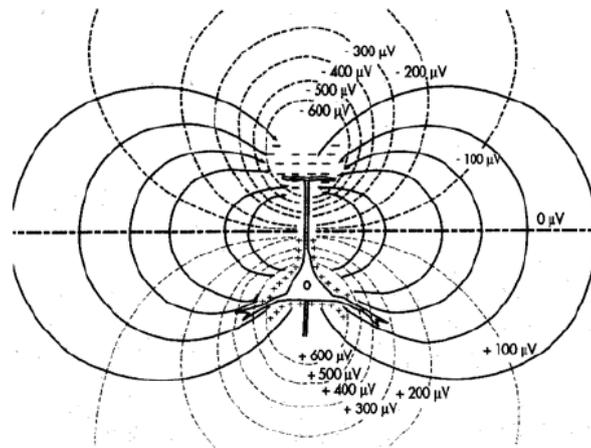


Figure 1.4 : Représentation d'une cellule pyramidale. L'excitation d'une cellule pyramidale par ses dendrites apicales crée une électronégativité apicale par rapport au soma qui devient électropositif. Les lignes en pointillé représentent les lignes iso-potentielles. Cette différence de potentiels produit un courant extracellulaire indiqué par les lignes continues qui sont perpendiculaires aux lignes de potentiels [5]

Au moment de la transmission synaptique, la dendrite apicale devient électronégative alors que le soma (l'autre pôle) reste positif. Le tout peut alors être représenté sous la forme d'un dipôle électrique.

On a vu dans le paragraphe précédent que la connexion neuronale se fait selon les deux types de synapses existant donc deux mécanismes sont à l'œuvre pour cette communication :

- la transmission de l'influx nerveux le long des fibres nerveuses par le potentiel d'action,
- et la transmission synaptique qui fait le lien entre deux neurones.

La membrane des neurones possède la propriété d'être excitable ; cette caractéristique est expliquée du fait qu'à l'état de repos, il existe une différence de potentiel entre l'extérieur et l'intérieur d'environ -60 à -70 mV due à une différence de concentration ionique [4]. Les ions

potassium (K^+) sont plus concentrés à l'intérieur de la membrane du neurone, alors que les ions sodium (Na^+) sont plus concentrés à l'extérieur.

Les neurones échangent des influx nerveux véhiculés par la « dépolarisation » de leur membrane (figure 1.5) : lors d'une stimulation de la membrane, les ions sodium traversent Celle-ci à l'emplacement de la stimulation, puis cette réaction se propage le long de la membrane ce qui crée le potentiel d'action du neurone qui est une onde biphasique (dépolariation puis hyperpolarisation de la membrane) d'amplitude $1 \times 10^{-13} \text{mA}$, elle est sur une portion d'axone étendue assimilée à un dipôle électrique orienté.

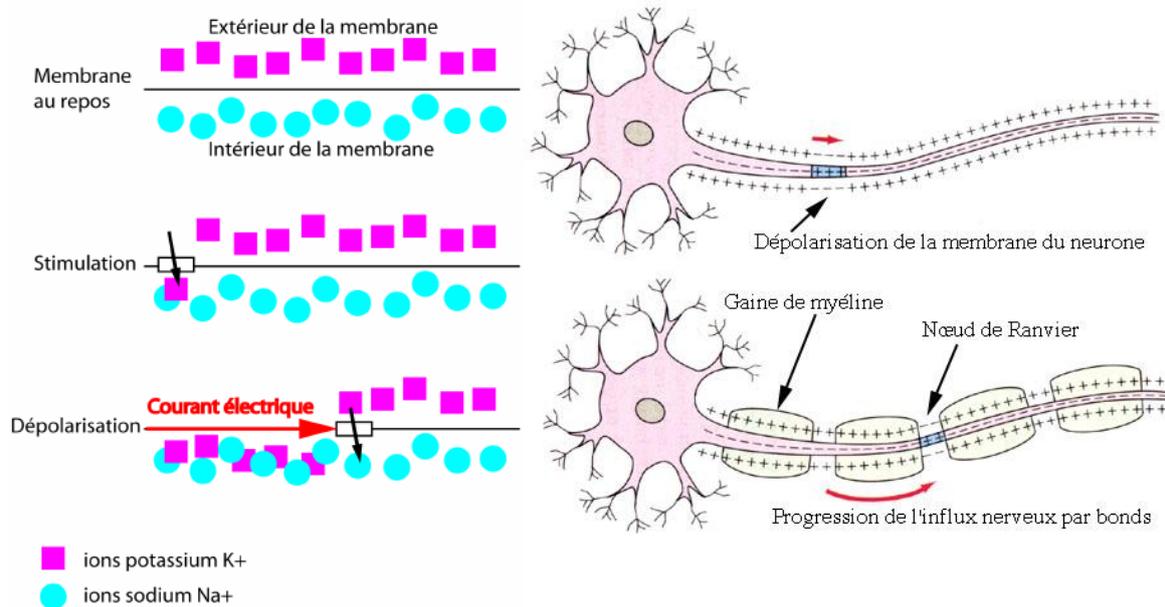


Figure 1.5 : Activité électrique du neurone. À gauche, mécanisme de la dépolarisation ; à droite propagation du courant électrique (ou potentiel d'action) dans le neurone

L'arrivée de l'influx nerveux au niveau de la terminaison axonique déclenche la libération de neurotransmetteurs. Ces neurotransmetteurs se fixent au niveau de la membrane post-synaptique. Selon le type de neurotransmetteurs, une dépolarisation ou une hyperpolarisation se produit, donnant respectivement un PPSE (potentiel post-synaptique excitateur) ou un PPSI (potentiel post-synaptique inhibiteur). Le PPSE peut être la source d'un potentiel d'action sur le second neurone [4], il est donc activateur et aussi dépolariise localement la membrane d'environ 25mV; par contre le PPSI inhibe l'apparition de potentiel d'action, augmente la différence de potentiel et il hyper-polarise la membrane (figure 1.6).

Si la membrane dépasse le seuil critique de dépolarisation, un potentiel d'action est initié. Les PPSI empêchent le déclenchement d'un potentiel d'action alors que les PPSE le favorisent. Les potentiels post-synaptiques se propagent le long des dendrites jusqu'au péricarion. C'est à ce niveau que sont générés les potentiels d'actions qui se propageront ensuite le long de l'axone vers d'autres synapses [4]. Le rôle du neurone est donc d'intégrer ces potentiels post-synaptiques pour moduler sa production propre de potentiels d'action.

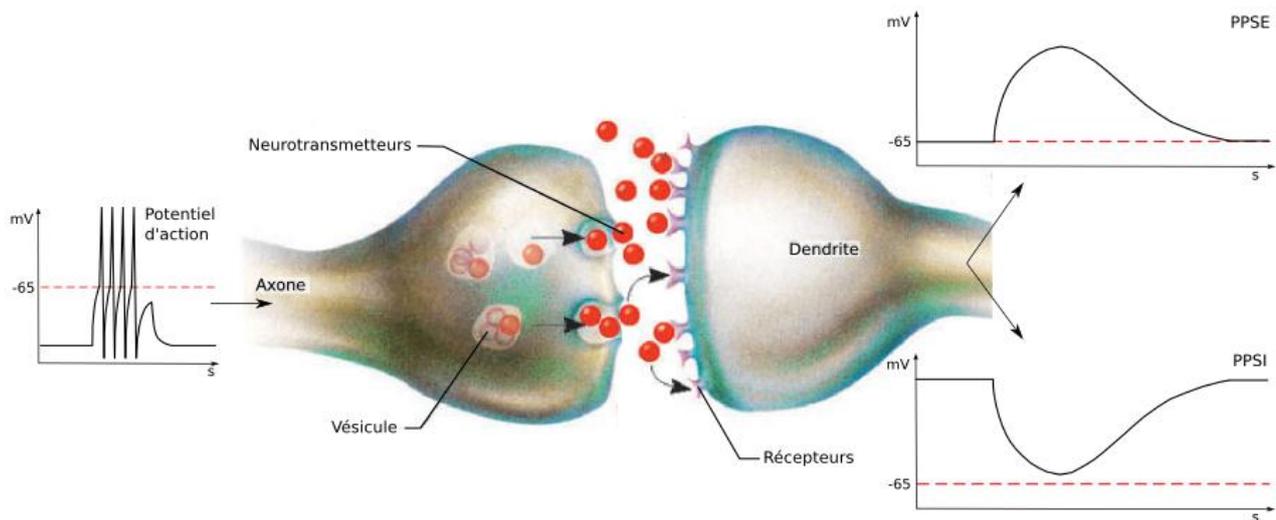


Figure 1.6 : déclenchement d'un potentiel post-synaptique (Image de la synapse adaptée d'après [4])

1.3 Conditions de mesure et Méthodes d'exploration de l'activité électrique

Les échanges d'ions chargés électriquement produisent un courant électrique qui peut être capté si une électrode est placée à l'extérieur du neurone. Si plusieurs neurones synchronisés sont dans le voisinage de l'électrode, l'électrode enregistre l'ensemble des activités de ces neurones. La condition pour laquelle cette activité puisse être enregistrée est l'alignement des cellules, pour les cellules qui ne sont pas alignées leurs signaux sont annuler mutuellement [5]. Par exemple, les cellules pyramidales du cortex sont organisées de façon verticale (en colonnes parallèles) : leurs potentiels peuvent donc se sommer (Figure 1.7). Les cellules de l'amygdale, par contre, sont mal alignées les unes par rapport aux autres. De plus, les électrodes de surface captent principalement les courants perpendiculaires au scalp. Les replis du néocortex dans de nombreuses directions ne produisent donc pas tous des courants détectables en surface. L'activité mesurée correspond donc essentiellement à l'activité post synaptique des cellules pyramidales corticales alignées perpendiculairement au scalp, avec une résolution spatiale de l'ordre du centimètre.

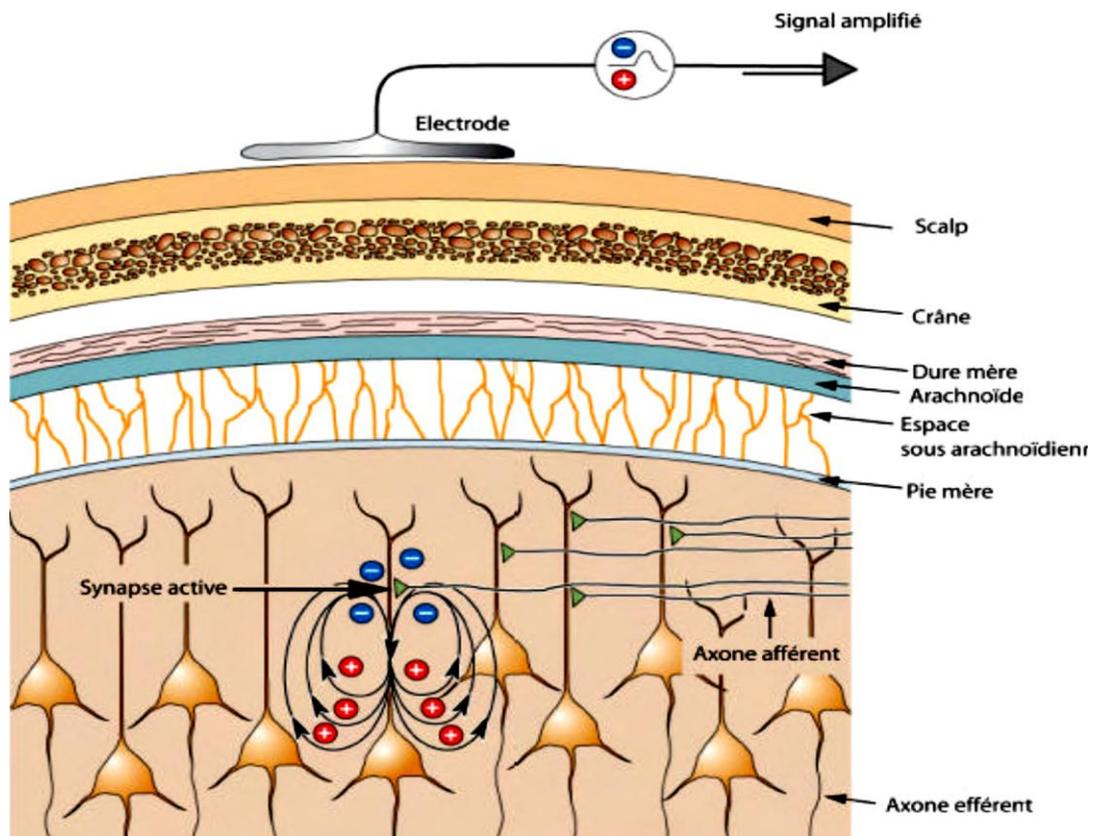


Figure 1.7 : Enregistrement EEG. L'activité électrique des neurones est enregistrée par une électrode placée sur le scalp. Si les neurones sont orientés parallèlement, comme c'est le cas pour les cellules pyramidales du cortex représentées ici (en colonnes verticales parallèles).

Il existe plusieurs types de modalités pour explorer les activités cérébrales, la figure 1.8 [6] illustre la variabilité des résolutions temporelles et spatiales parmi les méthodes d'exploration cérébrales (techniques d'imagerie et techniques électro-physiologiques). En principe ces méthodes peuvent être groupées en deux classes : l'imagerie et l'exploration électro-physiologique. Nous citons plus particulièrement les méthodes électro-physiologiques, et on s'intéresse surtout à l'électroencéphalogramme (EEG), qui est le point de départ de notre travail. Les autres méthodes sont les suivantes : La magnétoencéphalographie (MEG), l'EEG invasive qui est la stéréo électroencéphalographie (SEEG) et l'électro-corticographie (ECoG).

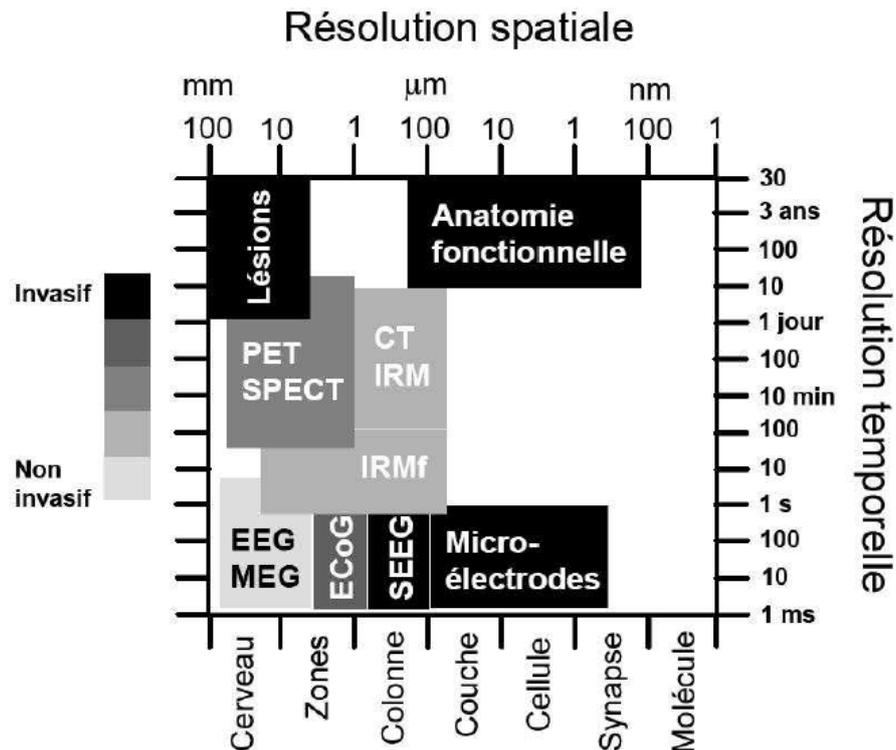


Figure 1.8: Schéma comparatif des résolutions temporelles et spatiales et du caractère invasif des différentes techniques d'imagerie cérébrale (EEG, IRM, IRMf, MEG, SEEG, TDM, TEMP, TEP)[6]

1.4 Electroencéphalogramme (EEG)

1.4.1 Historique

L'invention de l'électroencéphalographie est généralement attribuée au physiologiste allemand Hans Berger, qui commença à l'étudier dans les années 1920 et enregistra le premier signal d'activité cérébrale en 1929. Ses travaux furent repris et complétés par le britannique Edgar Douglas Adrian, qui obtient en 1932 le prix Nobel de physiologie. L'EEG ne s'est vraiment développé et répandu dans la pratique médicale courante, en particulier dans la pratique épiléptologique, que dans les années 1950.

Historiquement les courbes d'EEG étaient tracées sur des rouleaux de bandes de papier millimétré afin d'être relues ensuite par les médecins neurologues pour y déceler d'éventuels signes. Aujourd'hui, le signal est numériquement converti et traité par ordinateur.

Le développement de l'EEG en clinique date de l'apparition des amplificateurs différentiels au début des années 1960. Les EEG sont actuellement utilisés afin de diagnostiquer des foyers épileptiques, des tumeurs cérébrales, des lésions. Ils permettent aussi de trouver l'origine de migraines, de problèmes d'étourdissements, ou de somnolence.

1.4.2 Définition

L'électroencéphalogramme (EEG) est un examen qui enregistre l'activité électrique du cerveau en utilisant une technique non invasive pour mesurer les courants générés par les cellules nerveuses (neurones). Il est considéré le moyen le plus utilisé pour mettre en évidence l'activité électrique du cerveau, et particulièrement l'activité épileptique proprement dite. Comparé aux autres techniques d'investigation, L'EEG fournit des informations en temps réel et avec une très bonne résolution temporelle, de l'ordre de la milliseconde (figure 1.8). Il demeure incontournable pour le diagnostic et la classification des épilepsies. Berger a mis cela en évidence chez l'être humain en 1924 [7].

L'EEG est le résultat d'une différence de potentiel mesurée entre deux points. L'enregistrement entre deux points se fait grâce à des électrodes placées sur le cuir chevelu.

La (figure 1.9) présente un exemple d'un enregistrement d'EEG. Ces potentiels varient au cours du temps et forment des rythmes. La morphologie des tracés dépend de l'état d'activation des neurones enregistrés et cet état d'activation va se traduire par des variations de fréquence et d'amplitude des ondes enregistrées.

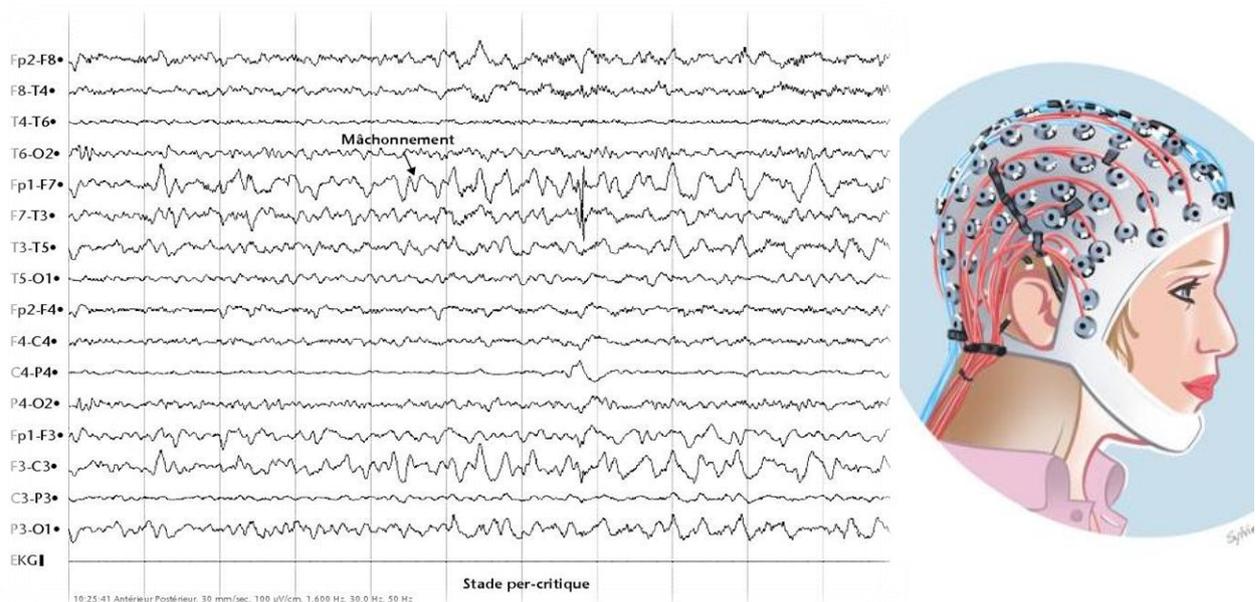


Figure 1.9 : Exemple d'un enregistrement d'EEG

1.4.3 les différents rythmes du signal EEG (α , β , γ , θ , Δ , μ) :

Un rythme cérébral désigne une oscillation électromagnétique dans une bande de fréquences donnée, résultant de de l'activité électrique cohérente d'un grand nombre de neurones du cerveau telle que observé en électroencéphalographie (EEG).

Les caractéristiques des rythmes cérébraux dépendent de l'individu selon l'état psychologique, pathologique ou normal, et son enregistrement nous permet de caractériser des maladies neurologiques, telle que l'épilepsie. Ces rythmes cérébraux sont classés selon leur fréquence en 5 groupes : delta (0.5–4Hz), thêta (4–8Hz), alpha (8–13Hz), bêta (13–30Hz) et gamma (supérieur à 30Hz) [8].

- **Les ondes alpha** : ont une fréquence de 8 à 13 Hz et une amplitude qui varie entre 30 jusqu'à 50 μ V. Elles se rencontrent principalement dans les régions occipito-pariétales [4]. Elles apparaissent dans l'EEG de presque tous les individus normaux en état de veille ou au repos les yeux fermés. Elles disparaissent complètement pendant le sommeil. Elles caractérisent un état de conscience apaisé que l'on retrouve particulièrement lors d'une méditation légère ou d'une séance de relaxation [9].

- **Les ondes bêta** : ont une fréquence de 14 à 30 Hz, d'amplitude faible inférieur à 20 μ V, occupe les régions moyennes des deux hémisphères souvent de manière asynchrone. Elles apparaissent généralement lorsque le système nerveux est actif et sont donc associées à la stimulation sensorielle et à l'activité mentale. Elles sont présentes particulièrement en période d'activité intense, de concentration ou d'anxiété. Comme elle peuvent être masqués par les rythmes alpha qui sont plus énergétique [9].

- **Les ondes thêta** : ont une fréquence de 4 à 7 Hz et une amplitude de 10 à 50 μ V. Elles se forment suite à l'interaction entre les populations du cortex et de l'hippocampe [10]. Elles se rencontrent surtout sur les régions temporales dans certaines fonctions cognitives et lors de l'endormissement [11].

- **Les ondes gamma** : ces ondes est un type d'occurrence plus rare qui ont une amplitude très faible et une fréquence varient entre 20 Hz et 60 Hz, elles sont associée à des fonctions motrices et cognitives (perception, raisonnement, langage, mémoire). la détection de ces rythmes peut être employée pour la confirmation de certaines pathologies.

- Les ondes delta

Les ondes delta se situent dans la gamme de 0,5 à 4 Hz, avec des amplitudes variables. Elles sont principalement associées à un sommeil profond, et à l'état de veille.

- **Les ondes de types μ** : ou rythme en arceaux, sont une particularité du rythme alpha [4]. Elles sont retrouvées dans les basses fréquences de l'alpha sont observées principalement dans les zones centrales lors de l'exécution d'un mouvement. Elles sont considérées comme étant caractéristiques du rythme alpha moteur.

La figure 1.10 illustre les principaux rythmes de l'activité cérébrale.

Pour les rythmes delta et thêta (les ondes lentes), peuvent témoigner d'un processus pathologique lésionnel ou fonctionnel sous-jacent. Dans les épilepsies partielles, elles sont fréquentes et souvent localisées dans la même région que le foyer épileptique inter-critique (les ondes thêta et delta sont largement présentes lors des variations du niveau de vigilance et, dans ces conditions, leur apparition n'est pas pathologique) [12].

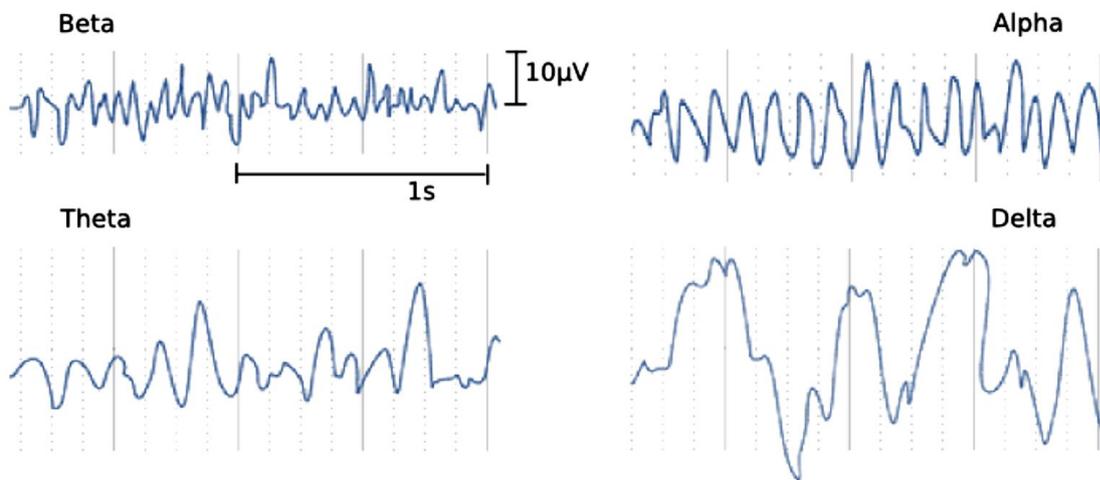


Figure 1.10 : les rythmes corticaux [6]

1.5 Procédure d'enregistrement des signaux EEG

Différentes procédures d'enregistrement existent, nous avons choisi de nous focaliser sur la technique la plus employée commençant tout d'abord de définir les différentes électrodes utilisées et leur système d'emplacement.

1.5.1 Electrodes

Les électrodes les plus couramment utilisées en électroencéphalographie sont, d'une part, les électrodes faites de métal noble (ou platine), en forme de coupe (Figure 1.11), que l'on remplit de gel. Et d'autre part, les électrodes jetables d'Ag/AgCl plus pratiques puisqu'elles contiennent déjà leur propre gel électrolytique et qu'elles sont auto-adhésives (Figure 1.12).

Les électrodes sont des senseurs électrochimiques qui effectuent la transduction d'un transport d'ions en un flot d'électrons électrolytique et que l'on attache sur la peau à l'aide de papier adhésif.

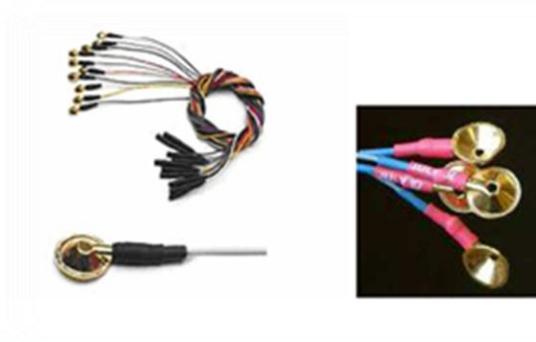


Figure 1.11 : photographie d'électrode sous forme de coupe

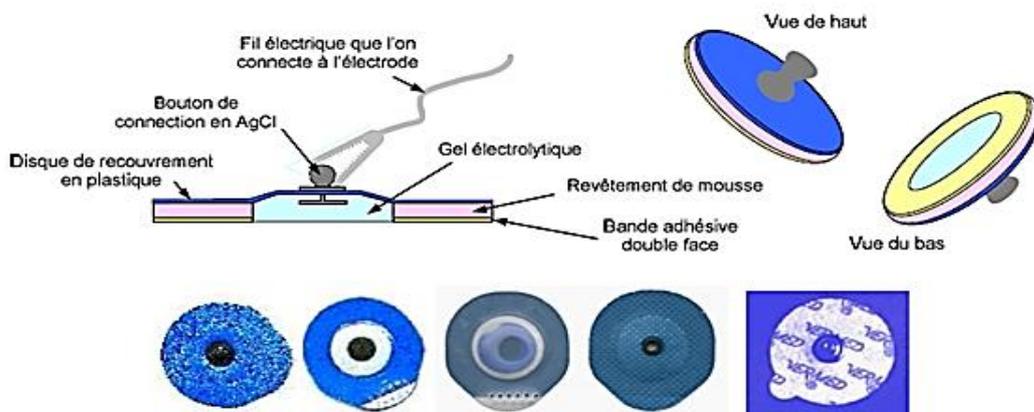


Figure 1.12 : photographie d'électrode d'Ag/AgCl.

Les électrodes sont de 3 types [13] :

- **L'électrode tampon** : il a la forme d'un petit dôme de 2 cm² à la base, en argent chlorure recouvert d'un tampon de tissu imbibé d'eau saline, est simplement placée sur le scalp après dégraissage du cuir chevelu et application d'une pâte conductrice, elle est maintenue par un casque de caoutchouc épousant le crâne du sujet.
- **L'électrode cupule** : il a la forme d'un disque en argent de 5 mm de diamètre dont la cupule centrale permet l'application d'une pâte conductrice, elle est collée sur le crâne sur un support imbibé de collodion. Ces électrodes sont recommandées pour tout enregistrement de longue durée et lorsque la probabilité d'enregistrement d'une crise épileptique est forte.
- **L'électrode aiguille** : il a la forme d'une aiguille hypodermique en acier inoxydable ou argentée, introduite juste sous la peau, parallèlement à la surface. Ces électrodes sont à usage unique et réservées aux situations d'urgence, en réanimation, au bloc opératoire.

1.5.2 Montage et système d'emplacement

a. Système d'emplacement

Pour assurer un positionnement spatial identique selon un repère anatomique propre à chaque patient, des normes ont été élaborées en routine clinique. Ces normes dépendent de la résolution spatiale souhaitée. En ce qui concerne les travaux présentés dans ce mémoire, dix-neuf électrodes de surface, voire 21 semblent indispensables en routine. L'emplacement des électrodes est standardisé par une nomenclature internationale appelée 10 – 20 [14]. Aujourd'hui avec l'amélioration des électrodes, des technologies d'enregistrement, le nombre d'électrodes peut considérablement être augmenté pour atteindre parfois 64, 128 voire 256 (EEG à "haute résolution – EEGHR"), mais toutes découlent du système 10-20. Ce système fixe de façon proportionnelle l'emplacement des électrodes par rapport à des repères osseux fixes (figure 1.13).

Les positions sont définies en divisant les lignes crâniennes des plans médian et transversal, en segments de 10 et 20 pour cents de leur longueur (figure 1.13).

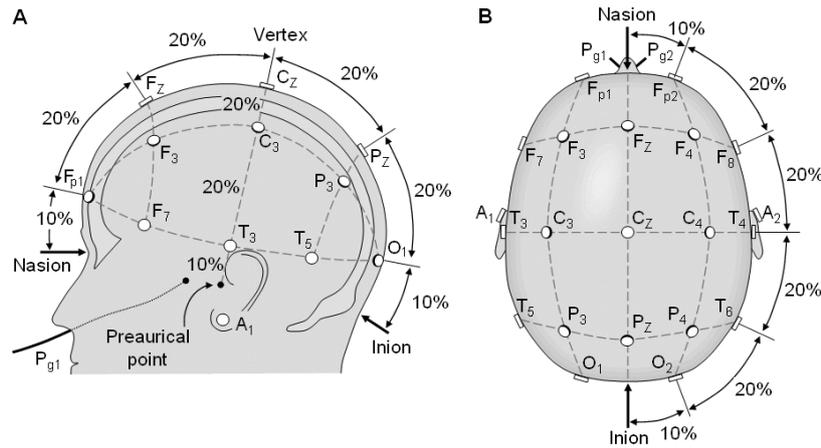


Figure 1.13 : Système de positionnement des électrodes 10–20 (21 électrodes)

Chaque électrode porte un nom précis, les chiffres pairs de chaque électrode indiquant le côté droit, et les chiffres impairs le côté gauche. Les lettres obéissent à la logique suivante : A=lobe de l'oreille, C=central, P=pariétal, F=frontal, Fp=frontal polaire, O=occipital, T=temporal et la lettre z indique l'axe central.

b. Les montages

Un montage EEG correspond à la combinaison de couples d'électrodes. Deux types de montages principaux peuvent être distingués : les montages référentiels et les montages bipolaires. Il peut exister aussi un troisième type de montage, moins utilisé qui est le montage Laplacien.

- Montage référentiel

Dans les montages référentiels, l'enregistrement de chaque électrode est fait par rapport à une électrode de référence. En théorie, la différence de potentiel correspond à la valeur absolue du signal recueilli sous la seule électrode active. En pratique, l'électrode de référence doit être choisie pour être la plus inactive possible, mais la référence parfaitement nulle n'existe pas et il existe toujours une contamination bioélectrique cérébrale ou extra-cérébrale de la référence. Le problème du choix l'électrode de référence est encore un problème d'actualité [13].

Une solution usuelle en clinique est de faire appel à un deuxième type de montage référentiel, consistant à utiliser une référence virtuelle (référence moyenne). La référence est représentée par le potentiel de référence moyen défini par la moyenne des potentiels de l'ensemble. En supposant une répartition symétrique des électrodes sur la surface de la tête, les sources intracérébrales, modélisées comme des dipôles électriques, devraient engendrer sur la surface

de la tête des potentiels de somme nulle, ce qui fait que cette référence moyenne est supposée nulle [15]. Cette méthode est particulièrement intéressante quand il n'y a pas de potentiels de très grande amplitude (artéfacts oculaires) ou de potentiels diffus, dans le cas contraire la polarité risque d'être faussée par les perturbations, ce qui peut amener à des conclusions erronées (une référence moyenne perturbée par les artéfacts par exemple). De plus, en réalité la tête n'est pas entièrement couverte par des électrodes de mesure, ce qui fait que l'hypothèse de base (répartition symétrique) est très approximative et donc que le signal obtenu en moyennant les mesures (référence moyenne) n'est pas strictement nul.

- **Le montage bipolaire**

Les montages bipolaires sont des connexions entre deux électrodes "actives", c'est-à-dire différentes de la référence. La différence de potentiel correspond à la somme algébrique des signaux recueillis sous chaque électrode active (ce qui conduit implicitement à l'élimination de la référence). Il y a principalement deux types de montages bipolaires : le montage bipolaire antéropostérieur (ou longitudinal) qui explore d'avant en arrière et le montage bipolaire transverse qui explore de droite à gauche.

Le montage référentiel et le montage bipolaire sont complémentaires, car ils ne s'interprètent pas de la même manière. Dans un montage référentiel, c'est l'amplitude de l'enregistrement qui renseigne sur l'activité cérébrale alors que dans un montage bipolaire, c'est plutôt l'inversion de phase qui permettra de déduire la localisation du dipôle.

Le montage bipolaire représente une mesure locale de l'activité sous-jacente. Une autre différence est que les premiers sont plutôt utilisés pour le diagnostic d'une crise très focale alors que les seconds sont préférés pour des crises plus étendues.

Un enregistrement se réalise toujours sur un montage référentiel avec possibilité ultérieure de recombinaison de montages bipolaires ou référentiel moyen. Lors de la lecture du tracé, pour l'interprétation, les cliniciens choisissent le ou les montages les plus informatifs. Le choix du montage dépend de la pathologie sous-jacente et des informations suspectées à la lecture d'un premier montage. On peut ainsi comparer une même séquence (par exemple une crise épileptique) selon différents montages.

- **Le montage Laplacien** : Montage utilisable en EEG numérisé, consistant en une transformation mathématique utilisant la dérivée seconde spatiale. On peut approcher la valeur Laplacienne du

potentiel en utilisant la valeur moyenne des électrodes de voisinage comme référence, et ce pour chaque électrode [16].

1.5.3 L'acquisition des signaux EEG

L'obtention d'un signal EEG comporte plusieurs étapes en utilisant un matériel d'acquisition. il est composé de.

- Un bonnet d'électrodes pour capter l'activité électrique,
- une têtère, boîte d'entrée des connecteurs reliés aux électrodes du casque,
- un amplificateur afin d'amplifier les signaux EEG qui sont de l'ordre du microvolt,
- un filtre analogique coupe-bande de 50Hz permettant la suppression des parasites dus au courant électrique,
- un convertisseur analogique numérique, il faut alors respecter le théorème de Shannon pour le choix de la fréquence d'échantillonnage. La période d'échantillonnage temporel est classiquement entre 0.5ms et 2ms, c'est-à-dire 2000 à 500 points par seconde et par électrode, ce qui d'après Shannon permet d'enregistrer des signaux de fréquence de 1000 à 250Hz,
- un système de visualisation et d'écriture sur disque.

L'ensemble de ces éléments est représenté sur la (figure 1.14).

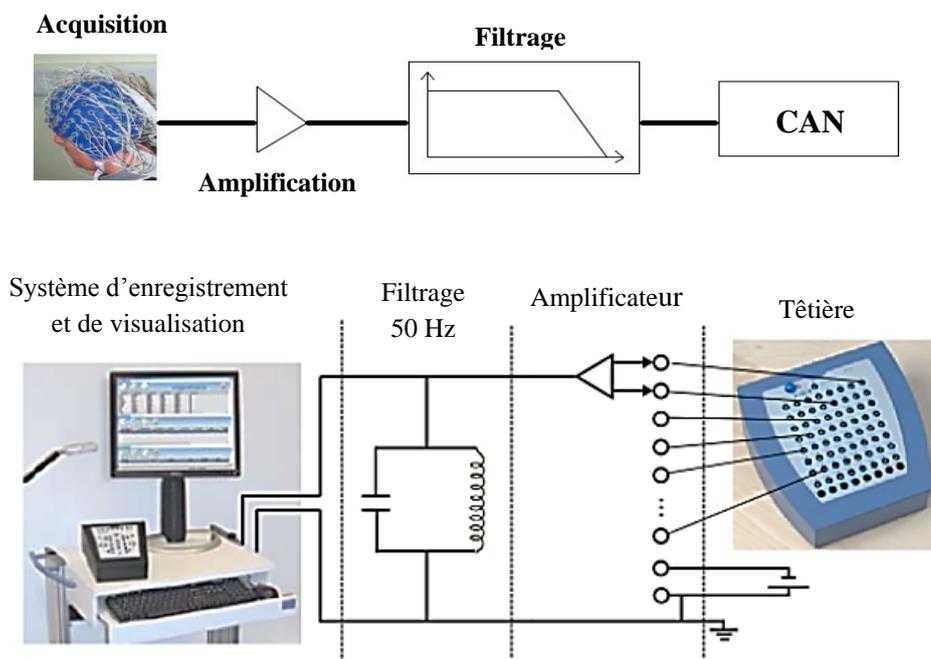


Figure 1.14 : Acquisition du signal EEG.

1.6. Conclusion

L'électroencéphalogramme enregistre les variations de potentiels électriques à la surface du scalp. Ces potentiels traduisent l'activité cérébrale. Ils sont utilisés dans le traitement de l'information extraite par le cerveau afin d'étudier le fonctionnement cérébral, il est donc obligatoire de se pencher sur l'analyse du signal cérébrale. La suite de ce manuscrit concernera l'étude de la synchronisation neuronale à travers l'utilisation des techniques d'analyse des signaux. On mettra en rapport autant que possible ces techniques avec les techniques actuelles de traitements des signaux EEG présentées dans la littérature.

Chapitre 02

Outils de traitement des signaux EEG (Les méthodes d'analyse temps fréquence)

2.1 Introduction

L'étude de l'activité cérébrale nécessite une analyse quantitative des oscillations codant le transfert de l'information d'une population neuronale à une autre. Ces oscillations, enregistrées au moyen d'EEG, sont des événements transitoires, imprévisibles même statistiquement. Une analyse des variations fréquentielles permettrait de décrire ces événements. Dans ce chapitre, on s'intéressera à l'analyse la plus adéquate qui répond à cette problématique.

2.2 La transformée de Fourier :

L'observation d'un signal $s(t)$ au cours du temps nous permet de connaître son début et sa fin et de constater ses éventuelles variations qualitatives. Cependant, il est moins évident d'avoir une idée sur ses périodicités (i.e. ses fréquences), d'où l'utilisation de la transformée de Fourier (TF) afin de décomposer le signal en une fréquence fondamentale accompagnée de ses harmoniques qui composent le spectre du signal. Chaque fréquence correspond à une fonction sinusoïdale dépendante du temps et nous donne une information sur la régularité globale du signal (figure 2.1).

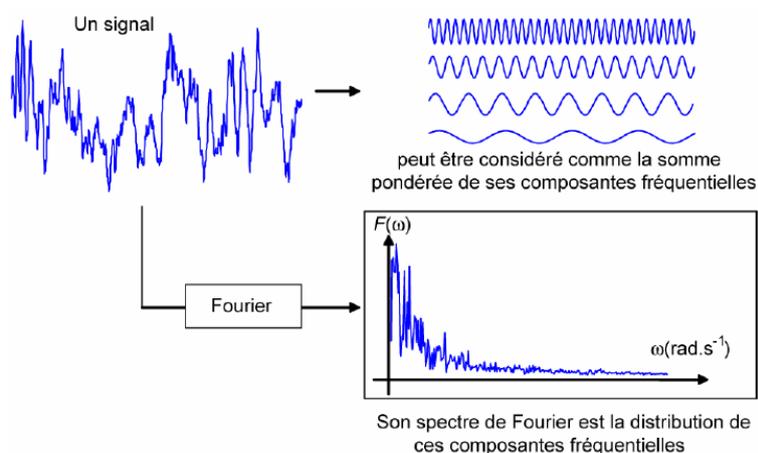


Figure 2.1 : Un signal et son spectre de Fourier. La transformée de Fourier fait passer le signal de l'espace temporel à l'espace des pulsations: $f(t) \rightarrow F(\omega)$. Elle indique la répartition de l'énergie selon les fréquences.

Soit un signal $x(t)$, si $x(t)$ est une fonction du temps t , absolument intégrable sur \mathbb{R} , sa transformée de Fourier notée $X(w)$, fonction de la pulsation w (en rad.s^{-1}), est donnée par :

$$X(w) = TF(x(t)) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (2.1)$$

Cette transformée est inversible par :

$$x(t) = TF^{-1}(X(w)) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(w)e^{j\omega t} dw \quad (2.2)$$

L'énergie E de signal est conservée d'après l'égalité de PERSOVAL Eq (2.3), telle que l'énergie de la représentation fréquentielle est égale à l'énergie de la représentation temporelle.

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega \quad (2.3)$$

Les signaux que nous traitons sont des signaux échantillonnés, dont nous voudrions connaître les composantes fréquentielles. Désignons par la fréquence d'échantillonnage f_e , et supposons que l'on cherche à obtenir des informations fréquentielles jusqu'à la fréquence f_{max} . Si la fréquence d'échantillonnage respecte le critère de SHANON:

$$f_e > 2 f_{max} \quad (2.4)$$

Souvent, l'information qui ne peut pas être distinguée dans le domaine temporel pourrait être facilement visible dans le domaine fréquentiel ce que montre les équations (2.1) et (2.3) ; La TF a permis de faire de grandes avancées sur la description des signaux EEG [17]. Mais elle montre vite ses limites dès lors que l'on sort du cadre rigoureux de sa définition : le domaine des signaux stationnaires d'énergie finie. Dans l'analyse de Fourier, tous les aspects temporels (début, fin, durée d'un évènement), bien que présents dans la phase, deviennent illisibles dans le spectre.

En particulier, la transformée de Fourier d'un signal EEG ne permet pas de définir l'enchaînement des oscillations neuronales, mais simplement les différents types d'oscillations mises en jeu. Or une analyse à la fois en temps et en fréquence est souhaitée, pour associer à une oscillation sa fréquence d'une part et sa date de début et de fin d'autre part.

La TF n'est pas donc l'outil adéquat pour les signaux non stationnaires comme l'EEG, à une exception : elle peut être utilisée pour les signaux non stationnaires si on ne s'intéresse qu'aux composantes spectrales qui existent dans le signal et non aux instants où elles apparaissent.

Alors que le but le plus désiré par le traitement des signaux EEG est de savoir à quel moment débute la crise épileptique, c'est-à-d. préciser quelles composantes spectrales apparaissent et à quels instants.

L'étude de signaux non stationnaires nécessite donc soit des transformées introduisant l'aspect temps- fréquence ou des méthodes spécifiques comme la distribution d'énergie. Fourier a développé une autre extension de la TF pour résoudre ce problème ; il s'agit de la transformée de Fourier fenêtrée (STFT).

2.3 La transformée de Fourier fenêtrée (STFT) :

2.3.1 Définition et principe de la STFT :

L'idée de cette technique est de donner une version "locale" de la transformation de Fourier, adaptée à des situations non stationnaires comme dans le cas du signal EEG [18]. Dans les années 1940, le physicien britannique Dennis Gabor découvre la première forme de représentation temps fréquence. Il obtient une analyse temporelle en découpant arbitrairement le signal en plage de longueur limitée. Chaque plage centrée autour d'un paramètre de localisation en temps, est alors étudiée séparément des autres par l'analyse traditionnelle de Fourier, ce qui revient à décomposer le signal sur des fonctions élémentaires ; qui dérivent toutes d'une même fonction fenêtre $W(t)$ par translation en temps [19] (figure 2.2) :

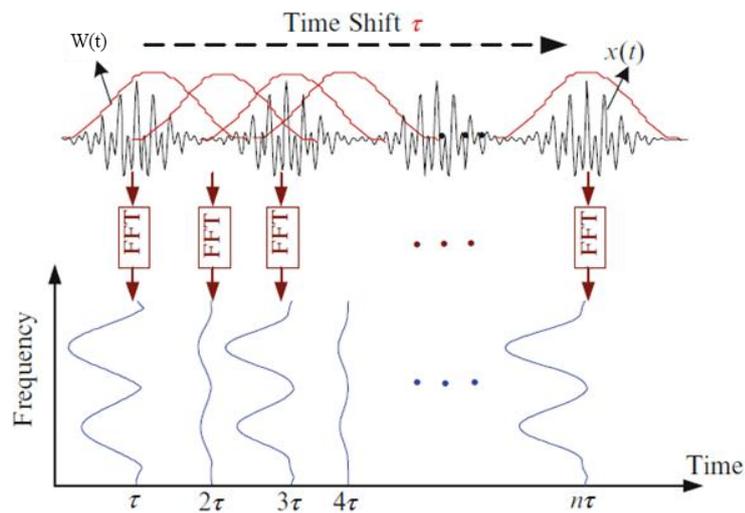


Figure 2.2 : illustration de la transformée de fourier fenêtrée [20]

La différence entre la STFT et la transformée de Fourier usuelle (TF) est l'ajout d'une fenêtre glissante $W(t)$ qui se déplace dans le temps d'un intervalle τ . La STFT permet donc de produire une analyse du contenu fréquentiel d'un signal de façon quasi instantanée, c'est-à-dire pour une portion du signal allant de t à $t+\tau$. Elle décompose le signal d'abord exprimé en fonction du

temps en une multitude d'atomes, ou coefficients, dans le domaine temps-fréquence. Tel que montré à la (Figure 2.3), la largeur temporelle de l'atome dépend de la fenêtre $W(t)$ et elle est toujours constant, peu importe la fréquence.

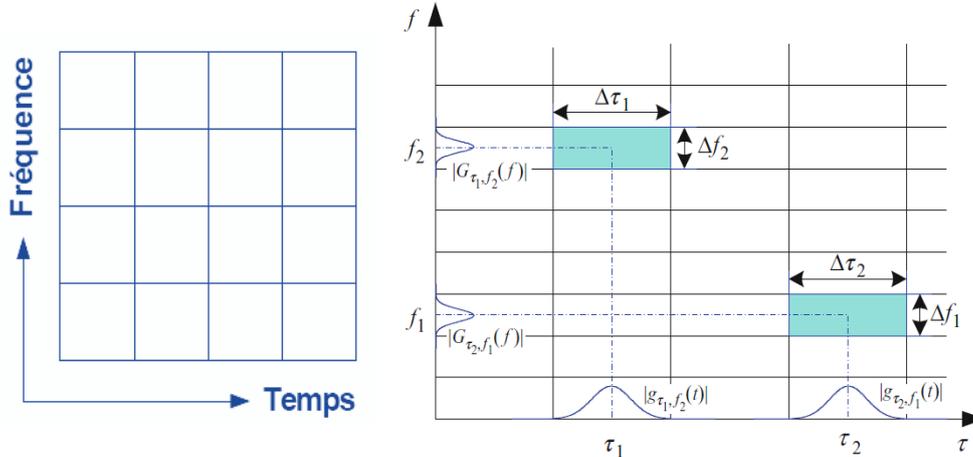


Figure 2.3 : découpage en atomes temps-fréquence : (résolution temps fréquence par la TF fenêtrée)

La transformée de Fourier à fenêtre glissante d'un signal $x(t) \in L^2$ génère une fonction à deux variables réelles, $X(f, \tau)$ qui s'écrit sous la forme suivante [21] :

$$X(f, \tau) = \int_{-\infty}^{+\infty} x(t) W^*(t - \tau) e^{-j\omega t} dt \quad (2.5)$$

Où la fonction $W(t - \tau)$ est la fenêtre qui permet de localiser $x(t)$ autour de τ avant de prendre sa transformée de fourier dans l'intervalle $[0, T]$ telle que T est la largeur de la fenêtre. Le résultat obtenu correspond donc à une représentation temps-fréquence du signal figure 2.4.

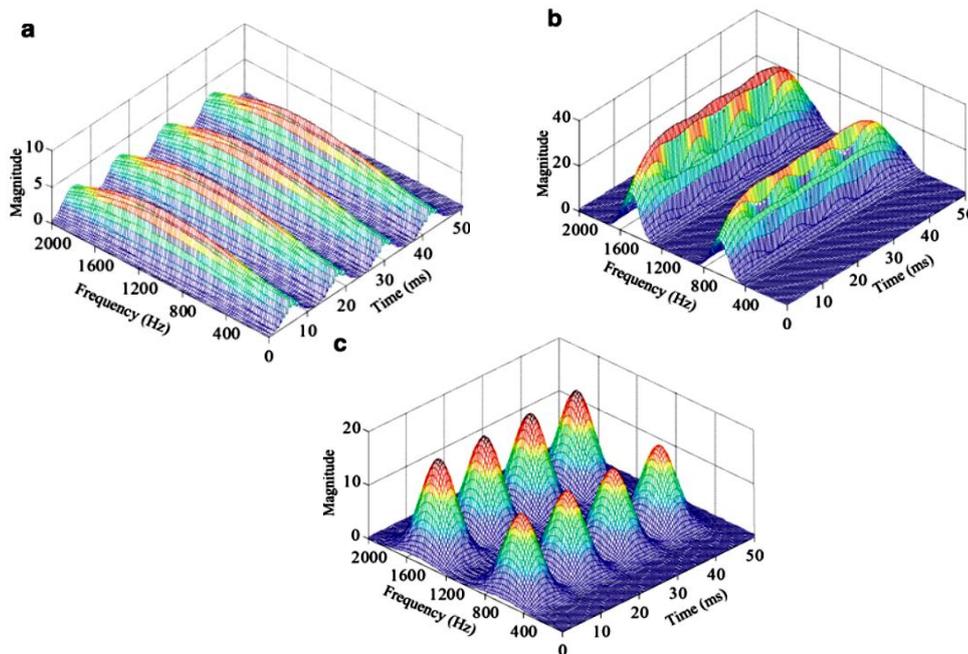


Figure 2.4 : Résultat de la STFT appliqué au signal représenté en figure 2.2 en utilisant des fenêtres de tailles différentes [20].

2.3.2 Principe d'incertitude et problème posé par la STFT :

L'équation (2.5) montre que, fenêtrant un signal veax dire que on le multiplie par une fenêtre rectangulaire (ou de Hamming, etc). Ceci correspond dans le domaine fréquentiel à effectuer un produit de convolution de leur transformée de Fourier. Ainsi, si on l'applique pour une sinusoïde on obtient les résultats suivants :

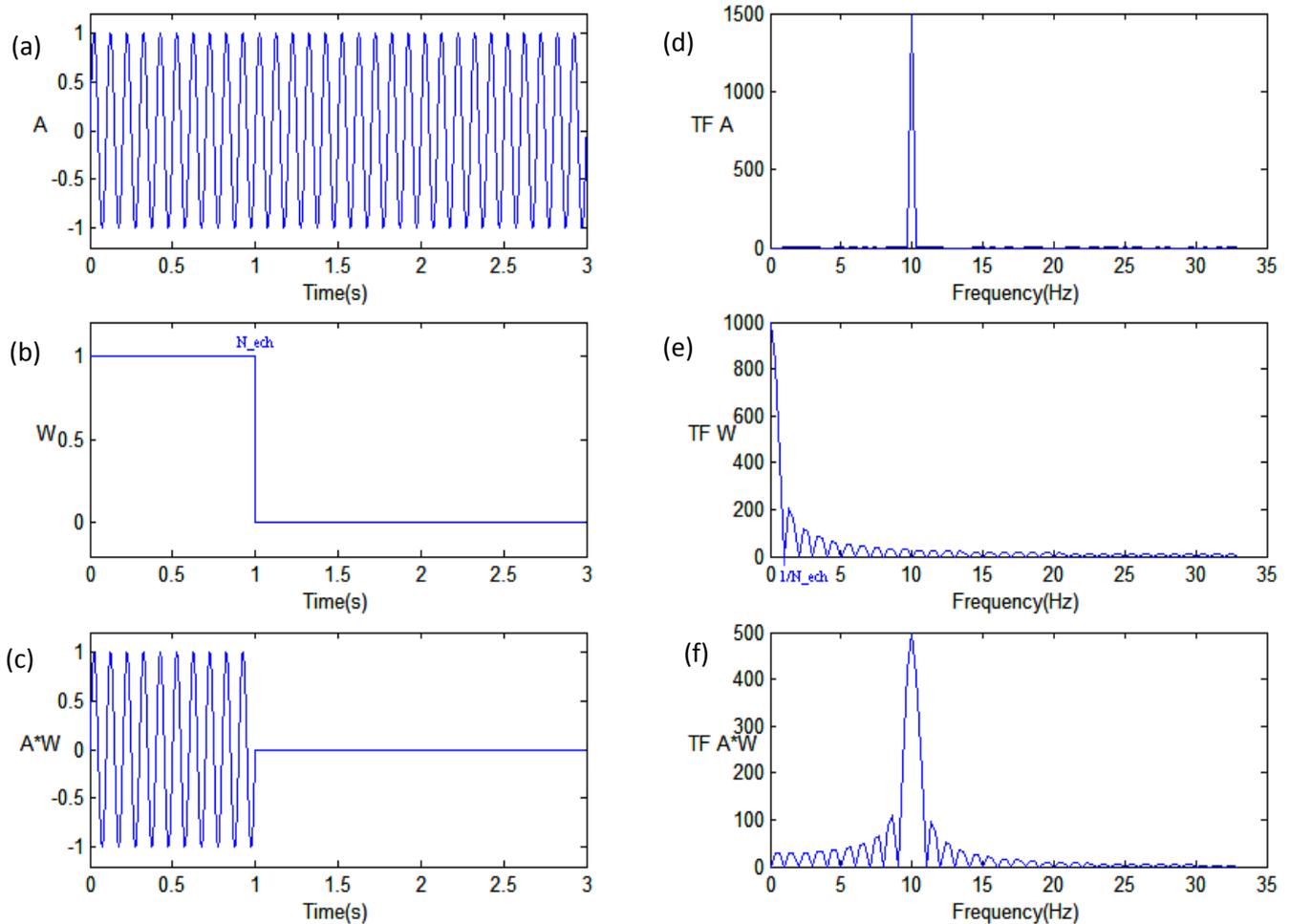


Figure 2.5 : application de la STFT a une sinusoïde

Les figures (a, b, c) représentent, respectivement, le signal sinusoïdal, la fenêtre rectangulaire d'observation, et le signal observé dans le domaine temporel. Leurs spectres sont représentés par les figures (d, e, f).

A partir de cet exemple on remarque que lorsqu'on cherche à bien localiser le pic on aura une perte de résolution dans le domaine fréquentiel puisque "le pic s'élargit". De ce fait, nous ne savons plus exactement quelles composantes de fréquence existent dans le signal, mais plutôt quelles sont les bandes de fréquence qui existent.

Si nous comparons la STFT à la TF dans ce cas on peut dire que cette perte de résolution est due au fait que la fonction $e^{j\omega t}$ n'est plus multipliée et intégrée de $-\infty$ à $+\infty$ mais de $-T/2$ à $T/2$.

Pour obtenir une résolution fréquentielle parfaite, il nous faudrait une fenêtre de longueur infinie, mais alors, nous aurions le même problème qu'au départ au sujet de la stationnarité du signal. Ce problème est en fait lié au principe d'incertitude d'Heisenberg exprimant que les résolutions en temps et en fréquence ne peuvent être arbitrairement petites en même temps car :

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad \text{avec} \quad \begin{cases} \Delta t^2 = \frac{\int t^2 |w(t)|^2 dt}{\int |w(t)|^2 dt} \\ \Delta f^2 = \frac{\int f^2 |W(f)|^2 df}{\int |W(f)|^2 df} \end{cases} \quad (2.6)$$

Donc La fenêtre ne doit pas être trop grande pour que le signal fenêtré soit stationnaire et que la résolution temporelle soit correcte ; Mais elle ne doit pas être trop petite non plus pour que les lobes correspondant à la TF de la fenêtre ne soient pas trop larges et pour que la résolution fréquentielle soit correcte.

Pour un autre exemple d'un signal constitué de 4 sinusoides on peut illustrer encore mieux le problème de la résolution de la STFT. Telle qu'on applique des fenêtres de différentes largeurs

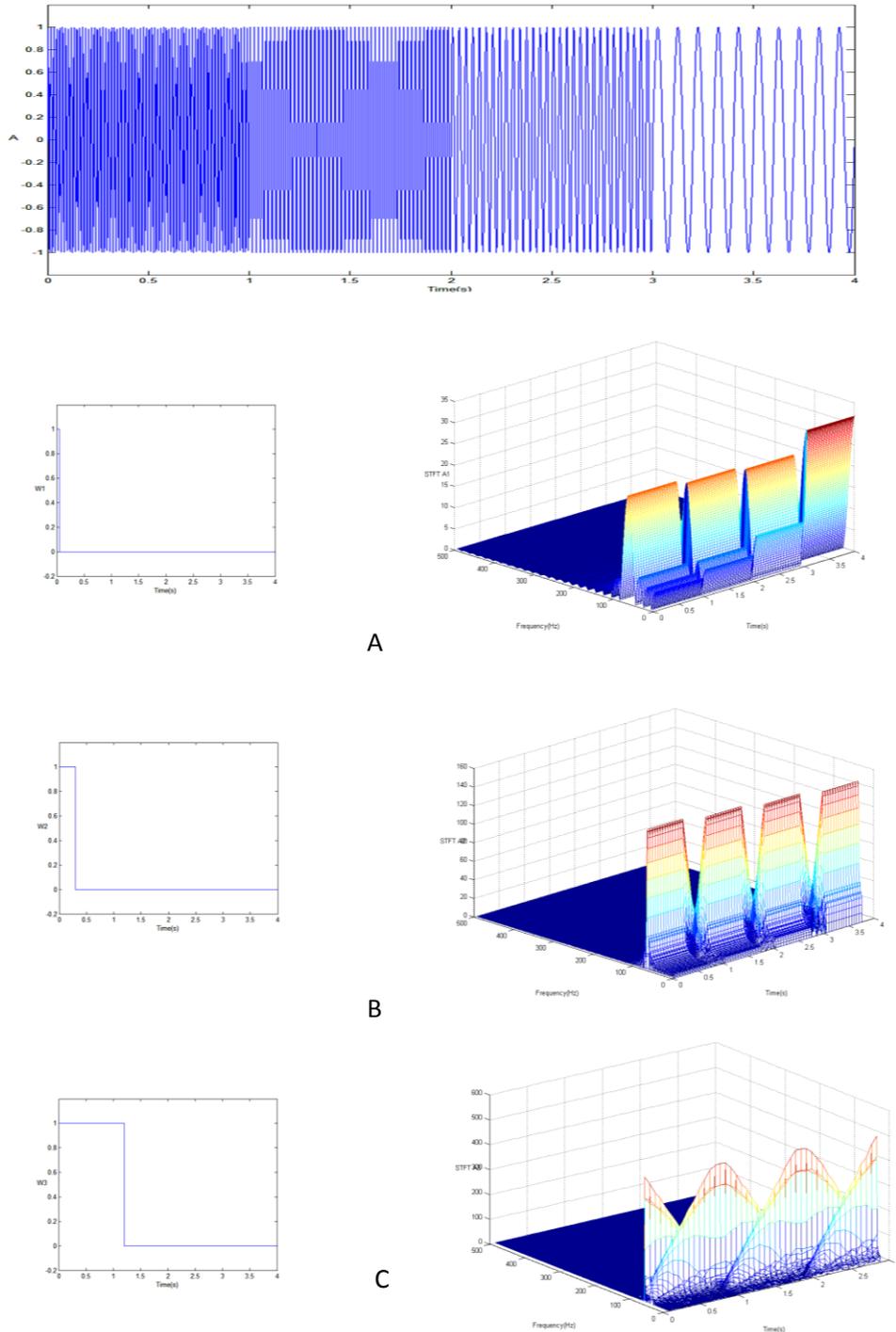


Figure 2.6 : illustration de problème de résolution de le STFT (application des fenêtres analysantes de différentes largeurs (A,B ,C))

On voit bien que la fenêtre glissante adaptée à l'analyse d'un signal précis est en même temps incapable d'analysé un autre ; or que on a besoin d'une méthode qui ne privilégie aucune échelle particulière, mais qui généralise à toutes les échelles l'analyse locale.

Donc en résumé le problème posé par la technique STFT a un rapport avec la largeur de la fenêtre employée, cette fenêtre est de longueur finie donc elle ne couvre qu'une partie de signal, de ce fait on ne peut pas connaître les composantes exactes de fréquence existant mais on peut connaître seulement une bande de ces fréquences, ce qui provoque une résolution fréquentielle faible. Et si on pense à augmenter la largeur pour éviter ce problème on va se retrouver dans un autre qui est l'absence de la résolution temporelle, c-à-d si on emploie une fenêtre de longueur infinie, on retombe sur la transformée de Fourier, qui donne la résolution fréquentielle parfaite, mais aucune information de temps. Idem si on veut obtenir une stationnarité il faut utiliser une fenêtre assez étroite, dans laquelle le signal est stationnaire ; telle que plus la fenêtre est plus étroite, plus la résolution en temps est meilleure, mais on aura une faible résolution en fréquence.

Afin de résoudre le dilemme de la résolution, une autre transformée, introduite dans le paragraphe suivant, a été développée.

2.4 - Transformée en ondelette :

2.4.1 Historique :

La première apparition de la technique de transformation en ondelette (WT : Wavelet Transform) était en 1910 par Haar ; l'idée de cette technique est héritée de la transformée de Fourier en remplaçant la sinusoïde par une famille de translation et dilatation d'une même fonction, appelée "ondelette".

Le premier papier traitant la Transformée en ondelettes continue (CWT : continuous wavelet transform) fut apparu en 1984 publié par Grassmann et Morlet. Puis en 1986, des ondelettes discrètes ont été construites par Daubechies, menant aux bases d'ondelettes et aux bases orthogonales par Lemarié et Meyer. En 1989, Mallat a donné le premier formalisme de l'analyse multi-résolution qui a mené à l'algorithme rapide de la transformée en ondelettes. Depuis l'apparition de cet outil mathématique, un grand nombre d'applications a été développé utilisant les CWT à l'aide des ondelettes orthogonales [22].

2.4.2 Définition :

La transformée en ondelettes permet d'analyser efficacement des signaux où se combinent des phénomènes d'échelles très différentes. C'est un outil mathématique qui découpe les données, les fonctions ou les opérations en composantes fréquentielles suivant une résolution adaptée à l'échelle du temps, elle a été développée pour palier quelques défauts de résolution de la Transformée de Fourier fenêtrée (STFT).

2.4.3 Résolution de problème de la STFT :

L'analyse en ondelettes a apporté une solution au problème d'adaptation de la taille de fenêtre d'analyse aux caractéristiques locales du signal : petite fenêtre lorsque le signal varie rapidement (hautes fréquences) et plus grande fenêtre lorsque ses variations sont lentes (basses fréquences) (voir figure (2.7)). La méthode consiste à la décomposition du signal sur une base de signaux élémentaires appelés Ondelettes, obtenus par dilatation et décalage d'une ondelette de base figure (2.10). En modifiant ainsi, par dilatations ou contractions successives, la taille de l'ondelette analysante, il est possible de détecter des détails localisés plus facilement qu'avec la TF.

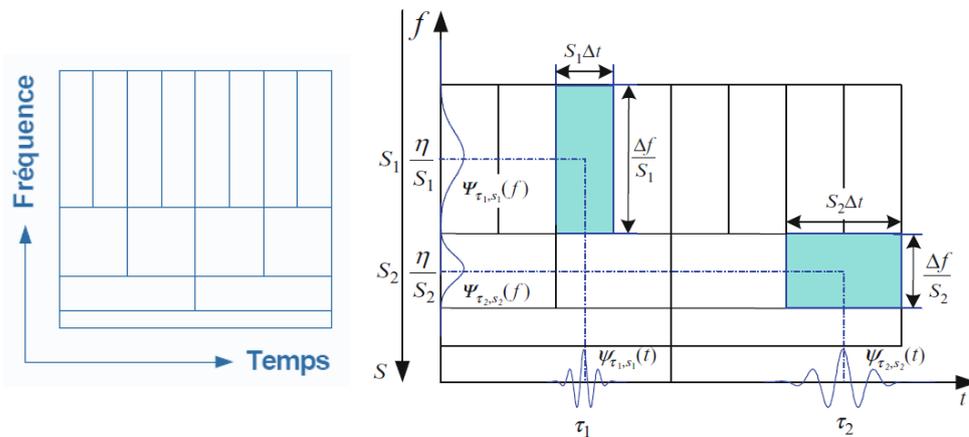


Figure 2.7 : découpage en atomes temps-fréquence : (résolution temps fréquence par la WT)

La principale différence entre la STFT et la WT est le découpage atomique du plan temps fréquence : pour la STFT, la largeur temporelle de l'atome dépend de la fenêtre $W(t)$ et elle est constante, peu importe la fréquence, tandis que le découpage par WT, produit des atomes plus larges dans le temps lorsque la fréquence est basse et inversement. Ainsi, les transitoires rapides d'un signal sont prises en compte par des atomes de haute fréquence qui ont une résolution temporelle très fine, alors que les mouvements plus lents, de basse fréquence, peuvent bénéficier d'une représentation fréquentielle plus précise figure (2.8).

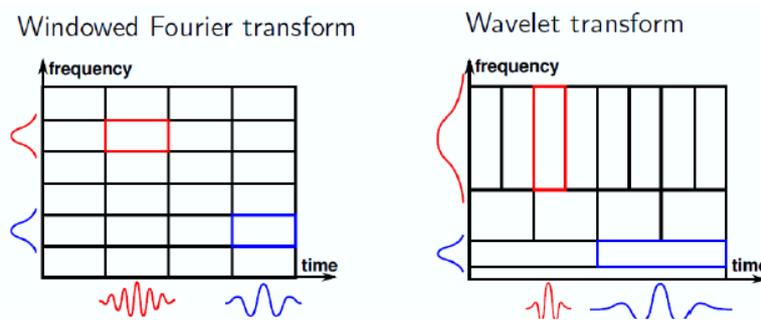


Figure 2.8: la différence entre la STFT et la WT

L'utilisation de la WT est devenue populaire et efficace, non seulement pour l'analyse des signaux de la parole ou d'images mais aussi pour diverses applications de traitement de signaux géologiques ou biomédicaux, et en particulier, l'électroencéphalogramme grâce à sa décomposition fréquentielle suivant une résolution adaptée à l'échelle du temps.

2.4.4 Principe de la transformée en ondelettes

La transformée en ondelette qui est une extension de l'analyse de Fourier, remplace la sinusoïde de la TF par une famille de translation et dilatation d'une même fonction. Cette fonction, nommé ondelette ψ (ou ondelette mère) figure (2.9).

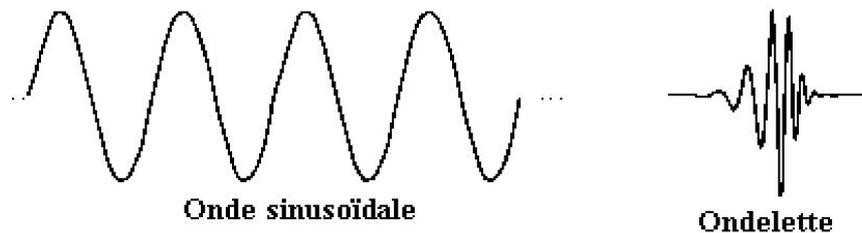


Figure 2.9 : La fonction analysante dans la TF est l'onde sinusoïdale qui est à la gauche et la fonction analysante dans la WT est l'ondelette qui à droite.

La fonction d'analyse compare le signal de versions décalées et compressée (ou dilatées) d'une ondelette ; la compression d'une fonction est dénommé dilatation ou mise à l'échelle et correspond à la notion physique de l'échelle.

En comparant le signal de l'ondelette à différentes échelles et positions (figure 2.10), on obtient une fonction de deux variables (figure 2.11). La représentation bidimensionnelle d'un signal à une dimension est redondante.

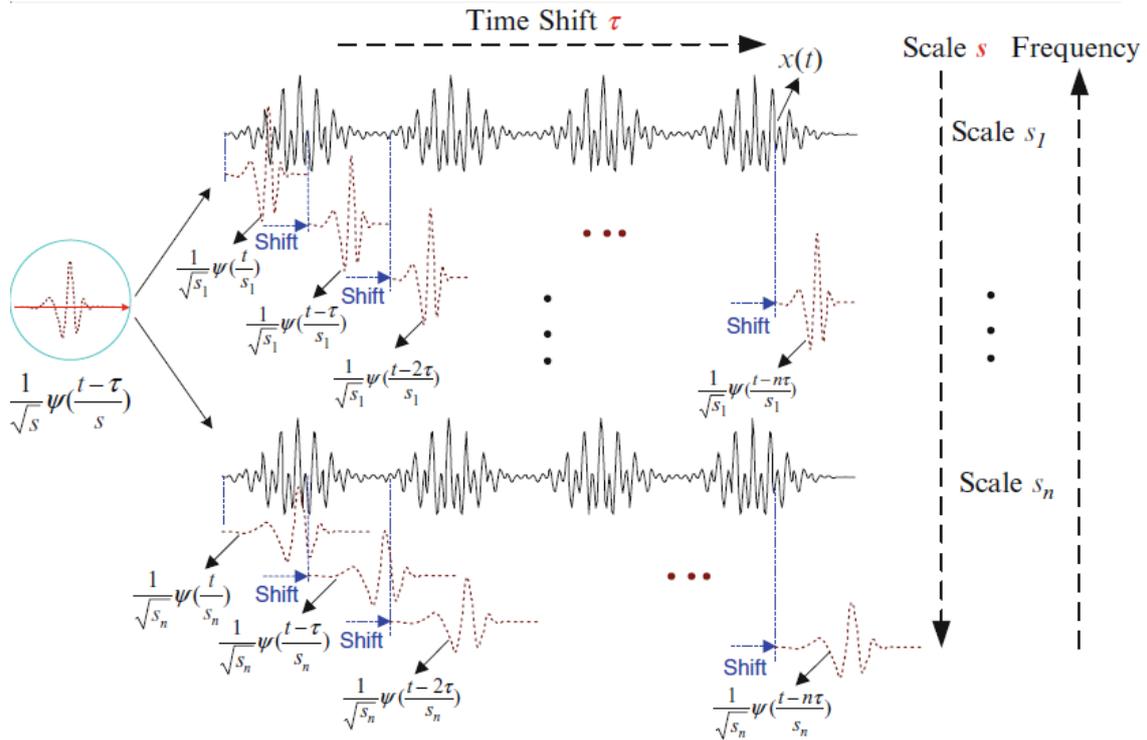


Figure 2.10 : illustration de la transformée en ondelettes [20]

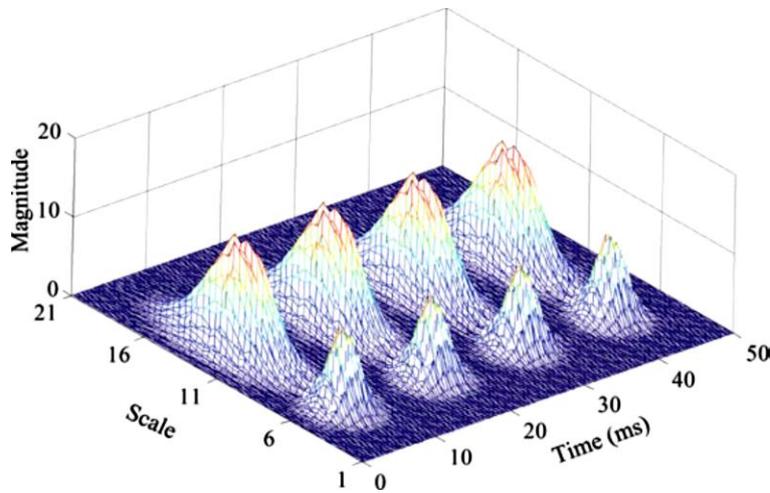


Figure 2.11 : la transformée en ondelette du signal représenté en (fig 2.7) [20]

Les valeurs des coefficients de la transformée sont obtenues par le produit du signal avec la fonction analysante et par intégration de ce produit (figure 2.12).

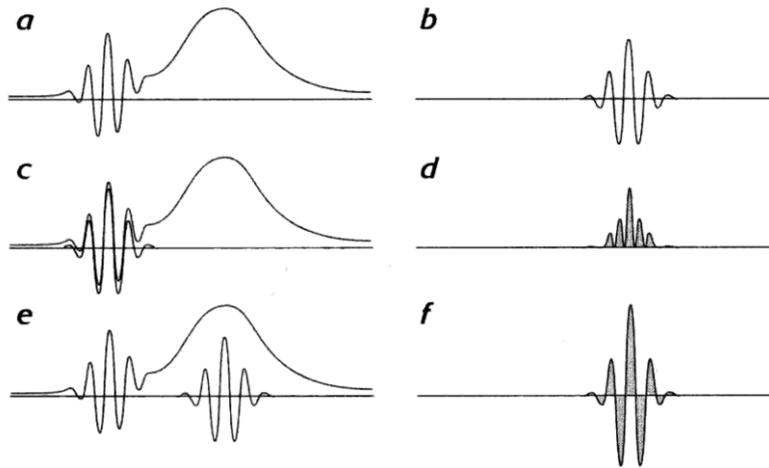


Figure 2.12 : principe de la transformée en ondelette

- (a) : signal à analyser,
- (b) : ondelette analysante,
- (c) : positionnement de l'ondelette au début du signal,
- (d) : résultat du produit du signal avec l'ondelette,
- (e) : déplacement de l'ondelette par τ ,
- (f) : résultat du produit du signal avec l'ondelette après déplacement.

Principe de la transformée en ondelette : La transformation d'une ondelette du signal (a) compare une ondelette (b) aux divers morceaux du signal (c et e). Le produit d'un morceau du signal et de l'ondelette donne une courbe (d et f) ; l'aire située sous cette courbe est égale au coefficient d'ondelette (en grisé). Un morceau de signal qui ressemble à l'ondelette (c) donne un coefficient important car le produit de l'ondelette et du signal est positif. Un morceau qui change lentement (e) donne un faible coefficient car les valeurs négatives de l'intégrale (f) compensent presque les valeurs positives. Ainsi, les ondelettes font ressortir les variations du signal. Adapté d'après [23].

2.4.5 Les différents types d'ondelettes

Il existe une infinité de fonction d'ondelette car toute fonction oscillante localisée est une ondelette mère possible. Toutefois, elles ne possèdent pas toutes des propriétés intéressantes. Pour cette raison de nombreux spécialistes en analyse d'ondelettes ont regroupée les familles d'ondelettes qui possèdent certaines propriétés remarquables. Parmi ces familles, les ondelettes de Haar sont les plus simples, mais elles ne sont pas bien localisées. Ingrid Daubechies a construit aussi des ondelettes à support compact qui permettent d'utiliser des filtres de taille finie, aussi une autre famille d'ondelettes est la famille des ondelettes Splines dont la réponse

fréquentielle est bien localisée. Les différentes familles d'ondelettes sont utilisées selon leurs propriétés en fonction du problème à résoudre.

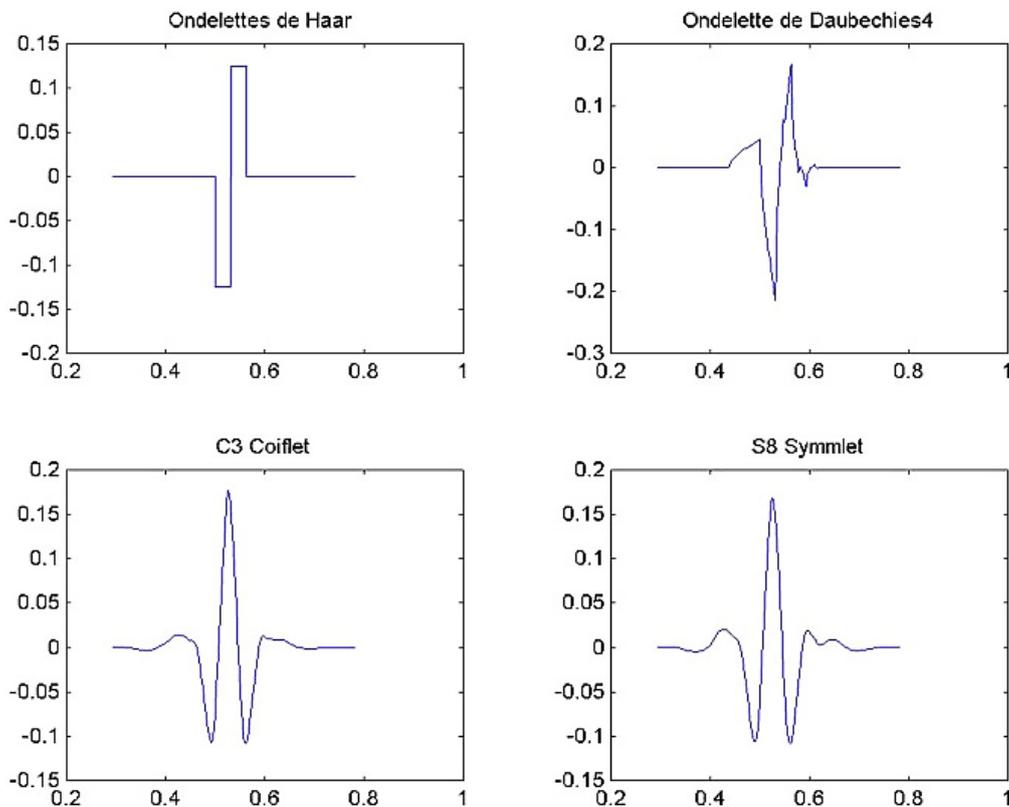


Figure 2. 13 : Quelques familles d'ondelette

Le choix de l'ondelette analysante est fait en testant les différentes ondelettes qui existent en se basant sur l'erreur représentant la différence entre le signal original et le signal de synthèse, et donc, cette erreur doit être calculée à chaque test effectué pour désigner le bon choix. Bien évidemment, l'ondelette choisie sera celle qui représente la plus faible erreur de reconstruction en respectant les particularités du signal.

Dans le cadre d'analyse des signaux EEG plusieurs recherches ont été effectuées concernant le choix de l'ondelette analysante ; on trouve souvent dans la littérature que les ondelettes de Coiflet et Daubechies sont les plus utilisées.

Dans ce travail, notre choix s'est porté sur l'ondelette Daubechies d'ordre 4 (db4) utilisée par [24],[25],[26],[27] et donnant de bons résultats. **Transformée en ondelette continue :**

La transformée en ondelettes continue d'un signal $x(t)$ prend la forme de l'Equation (2.7) [28]:

$$T_x^\psi (s, \tau) = \int_{-\infty}^{+\infty} x(t) \psi_{s,\tau}(t) dt \quad (2.7)$$

Ou s est le paramètre d'échelle (dilatation / contraction), τ est le paramètre de position (translation) et $\psi_{s,\tau}(t)$ est l'ondelette analysante ou ce qu'en appelle fonction de fenêtrage. Elle est localisée en temps et en fréquence, oscillante, de dimension finie et appartient à $L^2(\mathbb{R})$, elle est de moyenne nulle, normée: $\|\psi\| = 1$, centrée au voisinage de $t = 0$ et d'intégrale nulle car elle oscille en prenant des valeurs positives et négatives (Equation(2.8)).

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (2.8)$$

Par translation et dilatation de l'ondelette, les atomes de la transformée en ondelette sont définis pour toute échelle $s \in \mathbb{R}^{+*}$ et toute position $\tau \in \mathbb{R}$, un atome de la transformée est défini par l'équation (2.9).

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (2.9)$$

La dilatation change l'étendue temporelle de l'ondelette ($s > 1$ augmente l'étendue temporelle, $s < 1$ la réduit), tandis que la translation localise l'ondelette dans le temps (Figure 2.12(d)).

Mise à part les propriétés de dilatation-contraction et de translation, la CWT est caractérisée dans le plan temps-échelle par une fenêtre dont la largeur diminue lorsqu'on se focalise sur les structures de petite échelle (haute fréquence) ou s'élargit lorsqu'on s'intéresse au comportement à grande échelle (basse fréquence). Cette capacité d'adaptation en fonction de l'échelle d'analyse lui a valu la dénomination de « microscope mathématique » dont le grossissement est donné par le choix de la fonction analysante [29].

L'emploi de la CWT souffre de la redondance, ainsi l'utilisation des algorithmes rapides sont nécessaires pour appliquer cette Transformation.

2.4.6 Transformée en ondelette discrète :

La transformée en ondelettes discrète (DWT) est produite pour surmonter le problème de redondance de la CWT qui mobilise une grande quantité de ressource de calcul. Elle est considérablement plus simple à implémenter que la CWT.

La DWT fournit suffisamment d'information, tant pour l'analyse que pour la reconstruction du signal original dans un temps de calcul notablement réduit. Dans ce qui suit, nous introduisons les concepts de base de la DWT, ses propriétés et les algorithmes utilisés pour la calculer ainsi son application à l'analyse de signale EEG.

2.4.6.1 concepts de base de la DWT

Pour mettre en œuvre la transformée en ondelette sur des signaux numériques comme les échantillons EEG qu'en va traiter plus tard, une discrétisation des paramètres de résolution s et de position τ est nécessaire. Cette discrétisation du plan temps-fréquence correspond à la construction d'une grille non uniforme définie par les paramètres de l'équation 2.10 telle que τ_0 dépend de l'ondelette choisie.

$$\begin{cases} s = s_0^j \\ \tau = k \cdot \tau_0 s_0^j \end{cases} \text{ avec } s_0 > 1 \text{ et } \tau_0 > 0 \quad j, k \in \mathbb{N} \quad (2.10)$$

Les ondelettes sont définies alors par :

$$\psi_{m,n} = \frac{1}{\sqrt{s_0^m}} \psi^* \left(\frac{1}{s_0^m} t - n\tau_0 \right) \quad (2.11)$$

Et donc la transformée en ondelettes discrète aura la forme suivante :

$$DWT_x^\psi = \frac{1}{\sqrt{s_0^j}} \int x(t) \psi \left(\frac{1}{s_0^j} t - k\tau_0 \right) dt \quad (2.12)$$

L'idée essentielle derrière ce schéma de discrétisation est la suivante : Une fois choisi, un grossissement (i.e. s_0) étudie le processus à une position donnée, ensuite nous nous déplaçons à une autre position. Si le grossissement est important (analyse des détails à petite échelle), les déplacements se font par petits incréments. Au contraire, si le grossissement est faible (analyse à plus grande échelle), les déplacements peuvent être plus rapides. Ceci est automatiquement accompli si l'incrément de déplacement est inversement proportionnel au grossissement.

Pour avoir une analyse beaucoup plus efficace et tout aussi précise on choisit les échelles et les positions basées sur des puissances de deux, ce qui donne une transformation dyadique c'est-à-dire on prend $\tau_0 = 2$ et $s_0 = 1$, dans ce cas le principe de la décomposition discrète en ondelettes est alors de construire des ondelettes ψ qui génèrent des bases orthonormées de $L^2(\mathbb{R})$

par translation et par dilatation [30]. En plus de ce choix pour que la DWT soit envisageable la condition que les ondelettes analysantes soit orthogonal doit être vérifié.

L'orthogonalité signifie que l'information capturée par une ondelette est totalement décorrélée de celle capturée par une autre. Cela permet de ne garder que l'information nécessaire et suffisante afin d'assurer la réversibilité. Une ondelette mère orthogonale et sa fonction d'échelle permettent d'effectuer une analyse multi-résolution orthogonale : les espaces de détail et d'approximation sont alors orthogonaux, c'est-à-dire que la projection des vecteurs de la base de l'un des espaces sur l'autre donne zéro. Chaque ondelette et sa fonction d'échelle associée sont également orthogonales [6].

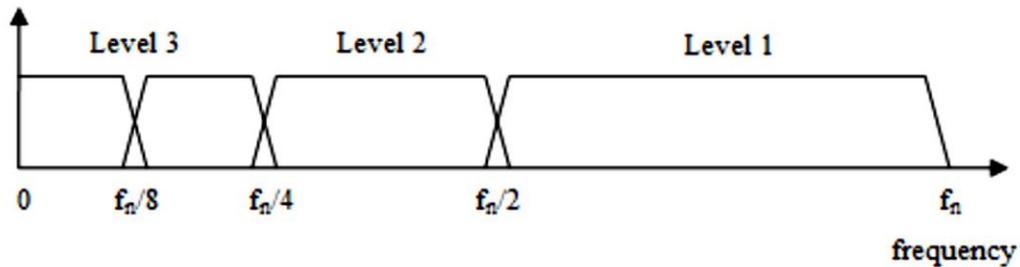


Figure 2.14 : Spectre des ondelettes orthogonales

Après avoir fixé les paramètres précédents l'équation 2.12 devient :

$$DWT(j, k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{+\infty} x(t) \psi \left(\frac{t-2^j k}{2^j} \right) dt \quad (2.13)$$

2.4.6.2 Décomposition : Coefficients d'Approximations et de Détails :

La DWT analyse le signal à différentes bandes de fréquence avec différentes résolutions en décomposant le signal par une approximation grossière et une information détaillée [31].

Les deux principaux algorithmes employés pour calculer une étape de décomposition de DWT sont le banc des filtres et le schéma de Lifting [22]. Pour l'analyse des signaux EEG, on s'intéresse au premier algorithme qui décompose le signal par banc de filtre. Dans ce cas la DWT est calculée dans une base de deux fonctions fondamentales, appelées fonctions d'étalonnage (ou ce qu'on appelle fonction d'échelle) et des fonctions d'ondelette, qui sont associées à des filtres passe-bas et passe-haut, respectivement (figure 2.15) .

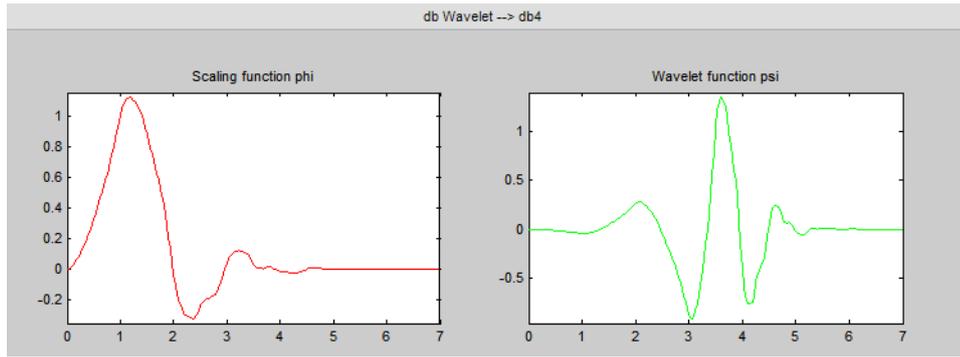


Figure 2.15 : La fonction d'échelle φ et la fonction d'ondelette ψ de l'ondelette mère de Daubechies db4

La fonction d'ondelette ψ représente le filtre passe haut, qui définit les détails, sa fonction correspondante a pour expression :

$$\psi_{j,k}(t) = 2^{j/2}g(2^j t - k) \quad (2.14)$$

Les approximations sont définis par la fonction d'échelle φ qui représente le filtre passe bas, sa fonction correspondante est :

$$\varphi_{j,k}(t) = 2^{j/2}h(2^j t - k) \quad (2.15)$$

Le processus de filtrage du signal original (s) donne deux types de signaux : les approximations qui sont les hautes-échelles (les composantes basses fréquences du signal S) et les détails qui sont les basses-échelles (les composantes hautes fréquences) figure (2.16);

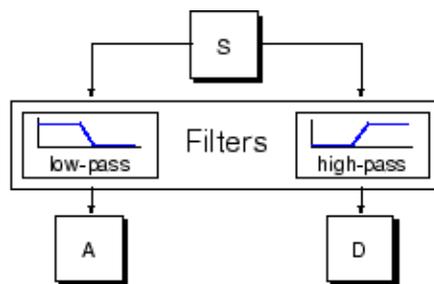


Figure 2.16 : La décomposition par banc de filtre

Les coefficients d'approximation cA_i et de détail cD_i obtenus lors de ces filtrages forment donc la transformée [32]. Au i ième niveau, ils sont décrits comme définis dans les équations suivantes :

$$cA_i(k) = \frac{1}{\sqrt{M}} \sum x(t) \cdot \varphi_{j,k}(t) \quad (2.16)$$

$$cD_i(k) = \frac{1}{\sqrt{M}} \sum x(t) \cdot \psi_{j,k}(t) \quad (2.17)$$

Où $t = 0, 1, 2, \dots, M - 1, j = 0, 1, 2, \dots, J - 1, k = 0, 1, 2, \dots, 2^j - 1, J = \log_2(M)$, et M c'est la longueur du signal tel que $M = 2^J$. Le taux d'échantillonnage k et la résolution j précisent les positions et les largeurs sur l'axe des x de fonctions, respectivement. Les hauteurs (ou amplitudes) de fonctions dépendent de $2^{j/2}$ valeur.

Le Calcul des coefficients d'ondelettes à toutes les échelles génère beaucoup de données, telle que si on effectue cette opération sur un signal numérique réel, on se retrouve avec des données deux fois plus importantes que celles avant traitement. Supposons par exemple que le signal S original se compose de 1000 échantillons de données ; alors les signaux résultants auront chacun 1000 échantillons donc ce qui donne en total 2000 échantillons pour une décomposition de 1^{er} niveau figure (2.17).

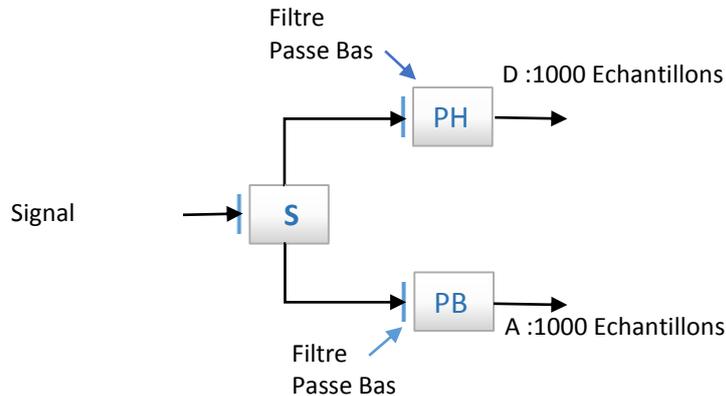


Figure 2.17 : duplication de nombres d'échantillons lors de la décomposition standard

La figure 2.18 illustre le processus de la décomposition d'un signal non stationnaire :

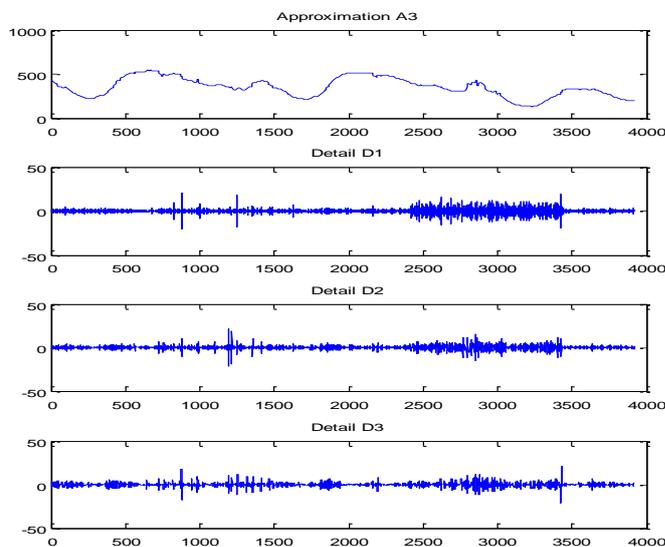


Figure 2.19 : décomposition a un seul niveau d'un signal non stationnaire

2.4.6.3 Décomposition avec sous échantillonnage :

Les signaux A et D sont intéressants, mais on obtient une dimension plus grande que celle de signal original S . Pour réduire la dimension, on utilise l'algorithme de Mallat qui est à base de banc de filtres suivi d'une étape de décimation. Cette étape permet de garder un seul point sur deux dans chacun des deux échantillons pour obtenir l'information complète. Il s'agit de la notion de sous-échantillonnage. On produit donc deux séquences (vecteur) appelées Coefficient d'Approximation (cA) et Coefficient de Détail (cD) figure (2.20).

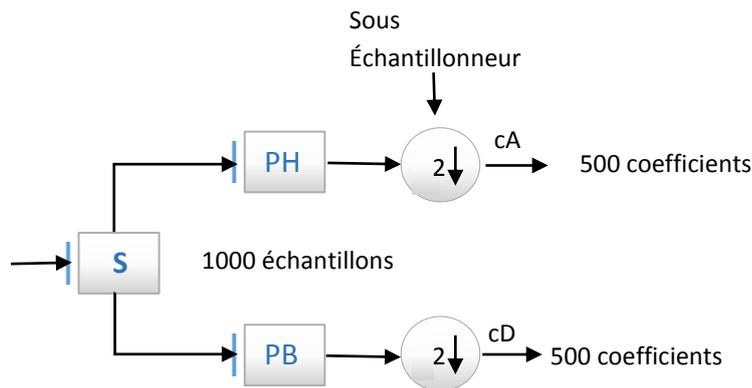


Figure 2.20 : Décomposition par banc de filtre avec sous échantillonnage

Si la taille du signal originale est n , et le nombre d'échantillons constituant chaque filtre associé à l'ondelette mère est égale à $2N$, les signaux D et A sont de longueur $2N + n - 1$ et les coefficients de cA et cD sont de longueur $L_{cD,cA}$ qui est donnée par :

$$L_{cA,cD} = \frac{2N+n-1}{2} \quad (2.15)$$

Le sous-échantillonnage d'un signal correspond à la réduction du taux d'échantillons (le sous-échantillonnage par un facteur n réduit n fois le nombre d'échantillons).

L'opération filtrage et celle du sous-échantillonnage peut être répétée pour avoir des conséquences avantageuses pour la décomposition du signal ; pour chaque niveau de décomposition on obtient :

- La moitié des échantillons (opération de sous-échantillonnage) \Rightarrow la moitié de la résolution temporelle.
- La moitié de la bande de fréquence (opération du filtrage) \Rightarrow le double de la résolution fréquentielle [31].

La structure de la décomposition en ondelette est utilisée pour avoir le maximum d'énergie dans chaque sous-bande

L'exemple représenté sur la figure (2.21), nous illustre le processus de décomposition avec décimation :

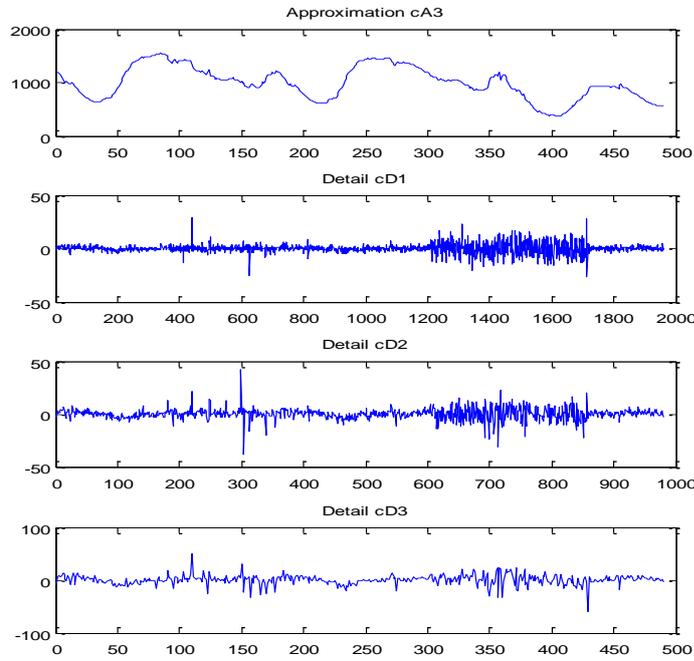


Figure 2.21 : décomposition d'un signal non stationnaire à trois niveaux

2.4.6.4 Décomposition multi-niveaux :

La décomposition multi-niveaux est la décomposition du signal en différentes bandes de fréquence elle est obtenue par les filtrages successifs (figure 2.22).

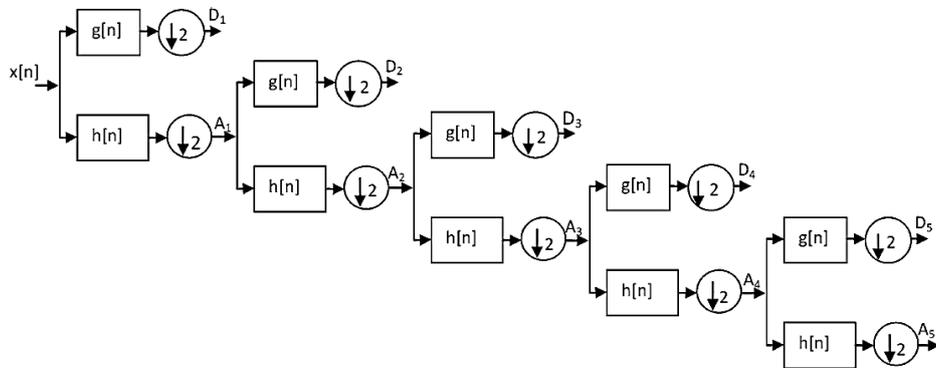


Figure 2.22 : Arbre de décomposition en ondelettes à cinq niveaux.

Dans notre travail, une décomposition à 5 niveaux a été effectuée. Le signal original $s[n]$ est d'abord passé par un filtre passe haut demi-bande $g[n]$ et un filtre passe bas $h[n]$, puis à chaque niveau, on filtre seulement la partie basse de la bande de fréquence, tandis qu'on laisse la partie haute intacte. Pour les signaux EEG l'information est contenue principalement dans la partie basse de la bande ; par conséquent, on divise la bande basse en plusieurs sous-bandes. Le pavage correspondant à chacun des niveaux de décomposition est montré sur la figure (2.23).

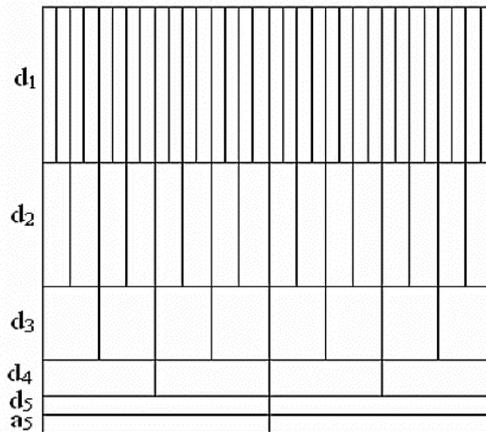


Figure 2.23 : Pavage temps fréquence correspondant à chacun des niveaux de décomposition.

La représentation à multi niveau représente ainsi le signal avec le minimum de coefficients nécessaires, soit autant de coefficients que de points d'échantillonnage, sans pertes d'informations : pour les basses fréquences, moins de coefficients que pour les hautes fréquences sont obtenus, avec un rapport de puissance de 2 entre chaque étape de décomposition [6].

2.4.6.5 Le Nombre de niveaux de la décomposition :

Dans la pratique, on doit sélectionner un nombre suffisant de niveaux basé sur la nature du signal, ou sur un critère approprié tel que l'entropie.

Si la taille du signal à décomposer est n , le nombre maximum de décomposition

$N = \log_2(n)$, N_{max} égale la partie entière de N , par exemple pour $n=4097$,

$$N = \log_2(4097) = 12.0004 \Rightarrow N_{max} = 12$$

Le nombre d'échantillons du filtre passe haut associé à la fonction ondelette mère et du filtre passe-bas associé à la fonction d'échelle est de $4(2N)$, pour les différents coefficients les tailles sont calculées par :

$$L_{CA,CD} = \frac{2N + n - 1}{2}; L_{CA1,CD1} = 499, L_{CA2,CD2} = 251, L_{CA3,CD3} = 127, \dots$$

2.4.7 Domaines d'application des ondelettes :

Les applications utilisant les ondelettes sont diverses ; aujourd'hui on peut utiliser l'analyse par ondelette dans les domaines suivants :

- Compression des données de haute résolution (e.g.images).
- Filtrage (denoising).
- Lissage (smoothing).
- Extraction des propriétés caractéristiques (Feature Extraction).
- Détection des discontinuités (discontinuity detection).
- Estimation des distributions de probabilités (distribution estimation).
- Analyse des données (data analysis) (par exemple : biomédicales, financières. . .).
- Télécommunication (par exemple : codage de sources et canaux (Source and Channel Coding)).
- Astronomie (par exemple : distances dans l'univers, galaxies formant de structures hiérarchiques à différent niveaux de l'échelle).
- Analyse de séries temporelles pour des prévisions de marché boursier.
- Fractals : ajustement parfait pour analyser les fractals.
- Analyse de turbulences : Analyse d'écoulement turbulent des fluides de basse viscosité coulant aux vitesses élevées.

2.4.8 Conclusion :

Une des raisons du succès de la transformée en ondelette est son implantation matérielle efficace de sa version discrète. La méthode la plus répandue pour l'implanter est l'utilisation de banc de filtre sous forme d'une structure pyramidale, technique qui provient des travaux de Mallat. La DWT représente ainsi le signal avec le minimum de coefficients nécessaires qui vont être utilisé pour calculer les paramètres caractéristique des signaux EEG d'où l'introduction du prochain chapitre.

Chapitre 03

Extractions des paramètres caractéristiques et Analyse des données

3.1 Introduction

L'utilisation des signaux EEG pour identifier les tâches mentales et détecter les différentes pathologies en particulier l'épilepsie repose sur deux points clés : l'extraction des paramètres caractéristiques et la classification [33].

On s'intéresse dans ce chapitre, plus particulièrement, à l'extraction qui vise à décrire les signaux EEG par quelques valeurs pertinentes appelées « caractéristiques », cette étape permet à faire une classification utile des signaux EEG.

Dans ce chapitre, nous définissons ces paramètres caractéristiques extraits des signaux EEG et une méthode de réduction qui sera utilisé afin d'enlever les informations redondantes et par conséquent, réduire la dimension des données à traiter pour améliorer les performances des méthodes de classification.

3.2 L'extraction des paramètres caractéristiques

La DWT est utilisée pour le prétraitement et la décomposition de ces signaux en sous-bandes en introduisant des filtres passe-bas et passe-haut à plusieurs niveaux en vue d'extraire les coefficients d'ondelette représentés par les paramètres d'approximation et de détail du signal. Ces coefficients extraits fournissent une représentation compacte du signal à différentes bandes de fréquence dans le temps et dans le domaine fréquentiel (voir figure (3.1) et figure (3.2)).

L'étape qui suit cette décomposition consiste à extraire et calculer un ensemble de caractéristiques qui sont des fonctions soit statistiques ou non statistiques elles doivent être calculés pour chaque niveau (chaque sous bandes) [34].

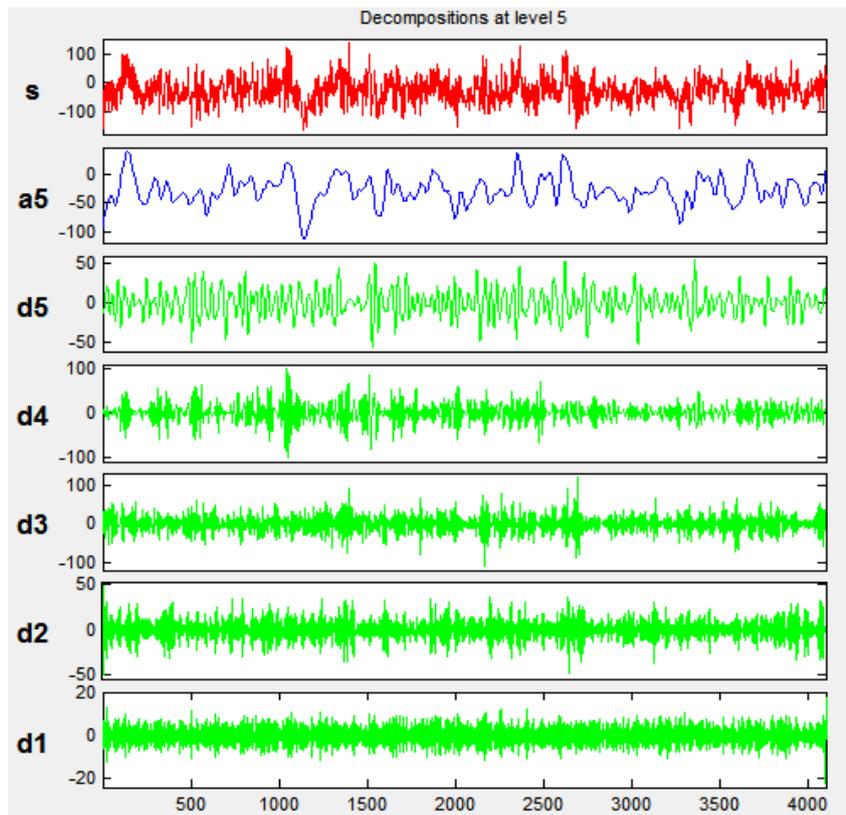


Figure 3.1 : Exemple de décomposition de signal EEG d'un sujet normal

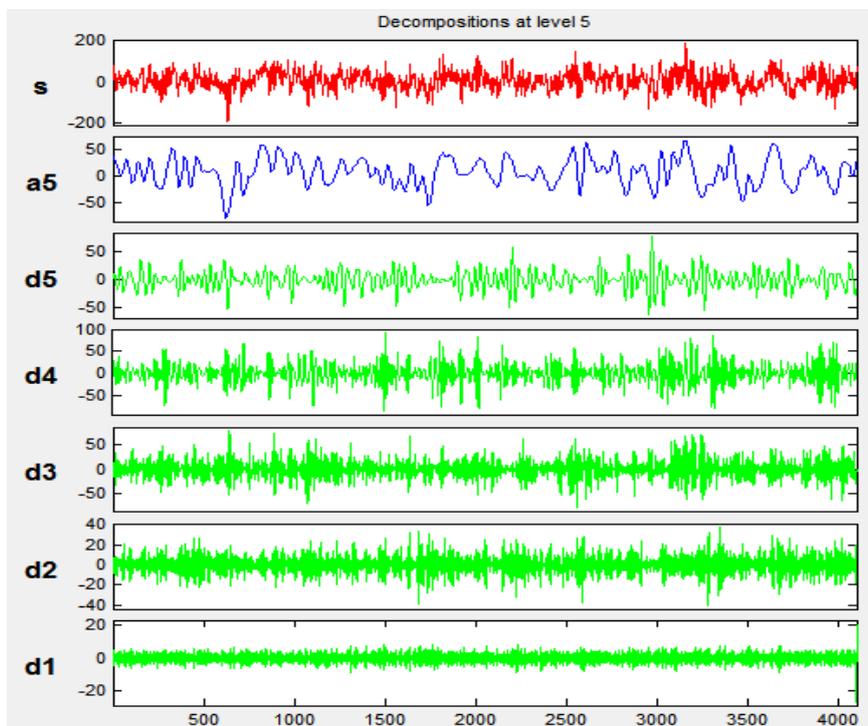


Figure 3.2 : Exemple de décomposition de signal EEG d'un sujet épileptique

3.2.1 Paramètres caractéristiques Non-statistiques

- **L'énergie** : L'énergie du signal est définie comme étant la somme du carré du module des valeurs de l'échantillon de tous les signaux. L'énergie s'écrit sous la forme suivante : C'est la somme quadratique des valeurs du signal ou des coefficients.

$$E = \sum_{n=0}^{N-1} |X|^2 \quad (3.1)$$

- **L'entropie** : Correspond à la quantité d'information contenue dans un signal. L'équation de l'entropie d'un signal X est défini comme suit :

$$e(X) = - \sum_{i=0}^{N-1} X^2 \log_2(X) \quad (3.2)$$

Ces vecteurs de caractéristiques, calculées pour les bandes de fréquence A5 et D1 à D5, ont été utilisés pour la classification.

3.2.2 Paramètres caractéristiques statistiques

- **La variance** : la variance est définie comme une mesure de la dispersion d'un ensemble de points de données X_i autour de leur valeur moyenne, elle est défini comme suit :

$$y = \frac{\sum (X - \mu)^2}{n} \quad (3.3)$$

Où : μ est la moyenne de l'ensemble X_i , n est le nombre d'échantillon.

- **La densité spectrale de puissance (DSP)** : C'est est la quantité d'énergie par unité de fréquence. L'intégrale de la DSP sur une bande de fréquence donnée calcule la puissance moyenne du signal sur cette bande de fréquence. La DSP est calculée par le module au carré de la transformée de Fourier de la série temporelle du signal :

$$S = |X|^2 / T \quad (3.4)$$

Où : X est la transformée de Fourier d'un signal $x(t)$

Les valeurs maximum et minimum estimées à partir de DSP de chaque sous-bande du signal peuvent être considérées comme des caractéristiques pour la classification.

- **Le maximum** : Traduit la plus grande valeur qu'atteint le signal.
- **Le minimum** : Traduit la plus petite valeur qu'atteint le signal.
- **La gamme** : Traduit la distance entre la valeur maximale et minimale du signal.

- **L'écart-type** : (en anglais : standard deviation) est une notion mathématique définie en probabilités et appliquée en statistique. L'écart type est une mesure de la dispersion d'une variable aléatoire réelle [35]

Dans cette étude ces paramètres statistiques ont été utilisés pour représenter la distribution temps-fréquence des signaux EEG

$$\text{Ect} = \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.5)$$

Pour chaque niveau de décomposition on calcule ces paramètres ce qui donne pour chaque paramètre une matrice (n.m) telle que n est le nombre de niveau de décomposition et m est le nombre des signaux EEG prit en traitement ce qui va générer des données à une très grande dimension, pour cela on utilise une des méthodes de réduction utilisée dans la littérature pour réduire la dimension de ces données.

3.3 Réduction de la dimension par analyse discriminantes linéaire

La réduction de dimension est généralement réalisée préalablement avant la classification pour contrer le fléau de la dimension afin d'améliorer les performances de la classification.

Pour cela il existe plusieurs méthodes de réduction, parmi ces méthodes il y a ceux qui sont linéaires comme l'analyse discriminante linéaire (LDA : Linear Discriminant Analysis) et l'analyse en composante principale (PCA : Principal Component Analysis) et d'autre qui sont non linéaires comme Isomap (Isometric Mapping) et LLE (Locally Linear Embedding).

La méthode proposée dans ce travail est la technique d'analyse discriminante linéaire qui offre un cadre formel intéressant sous espace discriminant de dimension réduite.

3.3.1 Le modèle de base de l'analyse discriminante linéaire (LDA) (point de vue intuitif)

La technique LDA s'applique lorsque les classes des individus sont connues. Elle consiste à chercher un espace vectoriel de faible dimension qui maximise la variance interclasse.

La LDA se base sur l'analyse discriminante de Fischer qui consiste dans un premier temps à projeter les données de l'espace initial sur des axes discriminants qui maximisent le rapport de la variance interclasse sur la variance intra-classe, puis de choisir entre les combinaisons linéaires des variables celle qui maximise l'homogénéité de chaque classe. Cela se fait par une technique d'optimisation à base de vecteurs propres pour trouver l'enrobage de faible dimension de points,

de telle sorte que chaque point est toujours décrit avec la même combinaison linéaire de ces voisines ; car les p variables ne sont certes pas toutes nécessaires pour décrire chacune des classes mais l'ensemble des variables est le plus souvent nécessaire pour discriminer les classes les unes par rapport aux autres.

3.3.2 Le modèle de base de l'analyse Discriminante Linéaire (LDA) (point de vue mathématique)

En terme mathématique, cela revient à trouver les vecteurs propres de la matrice de covariance formée par les différents individus de notre base d'apprentissage [36]

- Donnée et notation :

Le point de départ de la LDA est une matrice X de données observées (individus \times variables) dont les éléments sont identifiés dans une (et une seule) des k classes possibles et le point d'arrivée est une autre matrice Y qui est l'image de X obtenu par une transformation optimale. On dispose donc un ensemble d'apprentissage à n observations réparties dans des groupes d'effectifs n_i . Notons X l'ensemble qui regroupe toutes les observations dans l'espace initiale $X = (X_1, \dots, X_j)$. La discrimination linéaire de cette espace va générer un autre ensemble dans un autre espace qu'on note Y telle que, $Y = (y_1, \dots, y_k)$. On note aussi m_k les centres de gravité des nuages de points conditionnels, S_k leurs matrices de variance-covariance.

Les entrées et sorties de cette technique sont représentées sur la figure 3.3.

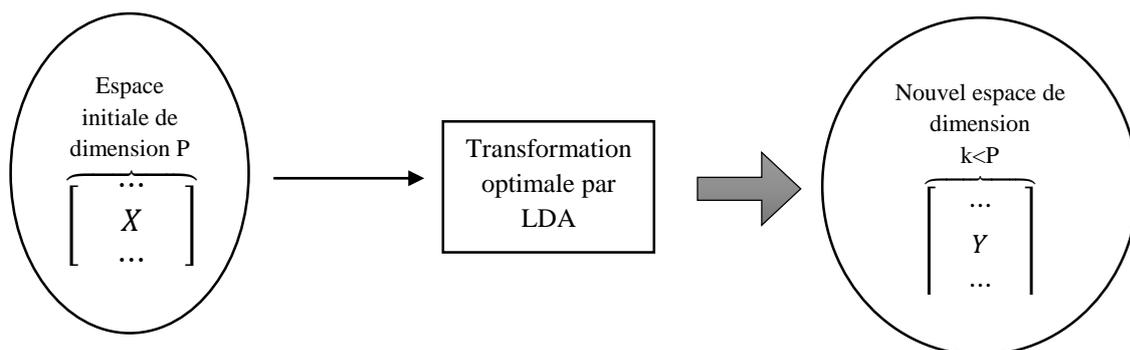


Figure 3.3 : Les entrées et sorties du processus de transformation par LDA

- Centrée réduite :

Considérons le tableau de données X composé de n individus décrits par p variables, dont le terme générale x_{ij} est une valeur centrée réduite. Les n individus sont répartis en q classes d'affectation connues à priori figure (3.4).

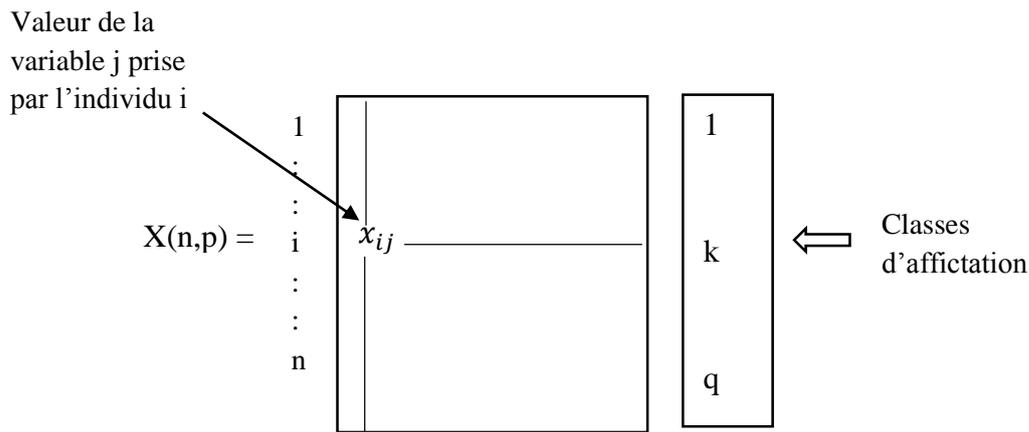


Figure 3.4 : Tableau centré réduit en LDA

Cette méthode permet de représenter graphiquement un ensemble d'individus par un nuage de points scindés en k sous-nuages déjà identifiés, Ces individus sont décrits par p variables numériques donc il est pratique de standardiser (soustraire de la valeur observée la moyenne afin de centrer puis diviser l'écart obtenu par l'écart-type afin de réduire) [37].

- Décomposition de la covariance :

En utilisant le théorème de Huyghens on peut décomposer la matrice de covariances de la population de taille n en deux matrices différentes, l'une donnant la variabilité dans chacune des k classes et l'autre la variabilité entre les k classes :

Donc à partir de la matrice X du tableau centré-réduit, la matrice d'inertie ou de corrélation, aussi appelée la matrice de variance total du nuage S , est calculée selon la relation (3.6). On a les différentes dispersions :

$$S = X^T X \quad (3.6)$$

- La dispersion à l'intérieur d'un groupe est décrite par la matrice de variance covariance. Nous pouvons en déduire (à un facteur près) la dispersion intra-groupe

$$S = \frac{1}{n} \sum_k n_k \times S_k \quad (3.7)$$

- L'éloignement entre les groupes, entre les centres de gravité des groupes, est traduit par la matrice de variance covariance intergroupes (à un facteur près)

$$S_b = \frac{1}{n} \sum_k n_k (m_k - m)^T (m_k - m) \quad (3.8)$$

Où m est le centre de gravité du nuage de points global.

- La dispersion totale du nuage est obtenue par la matrice de variance covariance totale. En vertu du théorème d'Huygens (qui est la généralisation multi dimensionnelle de la formule de décomposition de la variance) :

$$S = S_b + S_w \quad (3.9)$$

- **Critère de Fischer :**

Le premier axe factoriel sera donc défini par le vecteur directeur w_1 tel que l'on maximise la quantité

$$\frac{w_1^T S_b w_1}{w_1^T S_w w_1} \quad (3.10)$$

La variance interclasses sur ce premier axe factoriel Z_1 sera maximum ; La matrice de projection « P » est obtenue en maximisant le critère de Fisher généralisé.

Les facteurs discriminants u_1, u_2, \dots, u_{q-1} sont les vecteurs propres de $S - IS_b$ et ont pour valeurs propres respectives $\lambda_1, \lambda_2, \dots, \lambda_{q-1}$ telles que $\lambda_1 < \lambda_2 < \dots < \lambda_{q-1}$. Ces valeurs propres correspondent au pouvoir discriminant du facteur discriminant correspondant, c'est-à-dire, les $q - 1$ axes factoriels discriminants. Ainsi, par exemple, dans le cas d'une analyse avec trois groupes, l'ensemble des individus pourra être représenté dans le plan défini par les deux axes factoriels discriminants obtenus : c'est alors la meilleure représentation plane possible du nuage de points telle que les groupes soient les plus séparés possibles (voir figure (3.5)).

- **Évaluation :**

L'évaluation se situe à deux niveaux : évaluer le pouvoir discriminant d'un axe factoriel discriminant d'un ensemble d'axes factoriels. L'idée sous-jacente est de pouvoir déterminer le nombre d'axes suffisants pour distinguer les groupes d'observations dans le nouveau système de représentation.

Bien entendu, ces évaluations n'ont de sens que si les groupes sont discernables dans l'espace de représentation initial. Il nous faut donc, dans un premier temps, apprécier dans quelle mesure les centres de gravité des groupes sont distincts. En d'autres termes, il s'agit de vérifier si la

part de S_b dans l'équation $S = S_b + S_w$ est assez importante pour que cela vaille la peine de la décomposer par la suite [37].

- **Rapport de Corrélation (discriminante) :**

Une autre manière de rapporter l'importance d'un axe est de calculer le rapport de corrélation. Il s'appuie sur la formule de décomposition de la variance. Pour un axe factoriel Z , il est égal au rapport :

$$\eta(Z, Y) = \frac{\sum_k n_k (\bar{Z}_k - \bar{Z})^2}{\sum_k \sum_i (Z_{ik} - \bar{Z})^2} \quad (3.11)$$

(La somme des carrés intergroupes divisée par la somme des carrés totale, représente une observation).

Un axe sera d'autant plus intéressant qu'il présente un rapport de corrélation élevé [32].

- **Projection :**

Pour projeter des observations supplémentaires dans le plan factoriel, les logiciels fournissent les équations des fonctions discriminantes. Il suffit de les appliquer sur la description de l'individu à classer pour obtenir ses coordonnées dans le nouveau repère [39].

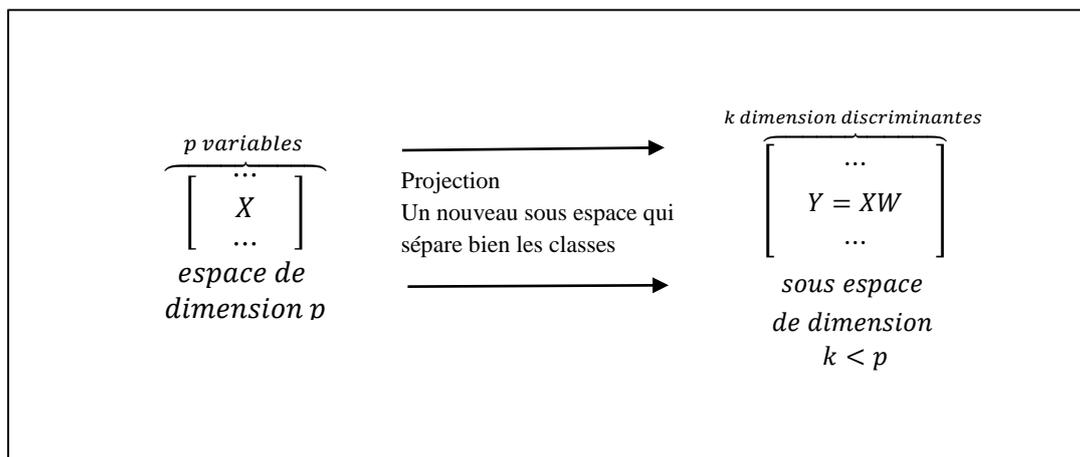


Figure 3.5 : processus de la Discrimination linéaire

Les points représentatifs des individus seront projetés sur des axes factoriels orthogonaux entre eux, dans un espace vectoriel à p dimensions. Ces axes ne correspondent pas à telle ou telle variable observée mais à une combinaison linéaire de ces différentes variables comme le montre la figure (3.6).

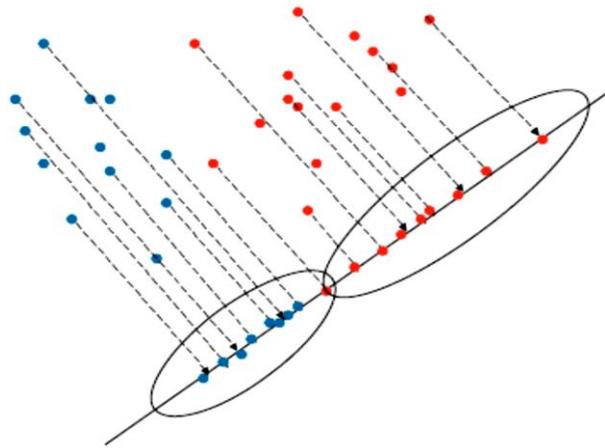


Figure 3.6 : séparation des nuages (droite de Fisher)

3.3.3 Application de l'Analyse discriminante pour deux groupes

Dans ce cas la LDA va se traduire par une projection de points sur un axe gradué. Donc le but est de déterminer un axe qui passe par les deux groupes de sorte que les projetés de chaque groupe sur l'axe doivent être aussi distinctes que possible.

La question qui se pose alors : comment trouver l'axe qui recevra d'un côté les projetés de la classe A et d'un autre côté les projetés de la classe B ? Figure (3.6).

Deux possibilités sont offertes mais, elles ne donnent pas les mêmes résultats : soit trouver l'axe sur lequel les barycentres de chaque classe auront les projections les plus éloignées possibles, soit déterminer l'axe sur lequel les projections des observations intra-classes seront les plus rapprochées. En fait, c'est une combinaison de ces deux possibilités qui sera réalisée.

Le problème revient à maximiser l'inertie interclasse projetée sur l'axe tout en minimisant l'inertie intra-classe projetée. L'idée est alors de maximiser le rapport de la première sur la seconde, ce qui est un excellent compromis. Donc, le nuage global est centré sur son barycentre par lequel passe notre axe. On démontre que l'axe qui sépare le mieux les deux groupes, également nommé droite de Fisher, correspond au vecteur propre associé à la valeur propre de $W^{-1}S_b$. La matrice W s'interprète comme une moyenne pondérée des matrices de variances-covariances intra-classe tandis que S_b est la matrice interclasse qui synthétise les distances entre barycentres

On dispose comme entrée deux groupes $D_1 = \{x_1, \dots, x_{n_1}\}$ $D_2 = \{X_1, \dots, X_{n_2}\}$. Les matrices de covariance des deux groupes sont :

$$\begin{cases} S_1 = \sum_{x \in D_1} (x - m_1)(x - m_1)^T \\ S_2 = \sum_{x \in D_2} (x - m_2)(x - m_2)^T \end{cases} \quad (3.12)$$

Où m_1 et m_2 sont les moyennes de chaque groupe elle est calculée par la formule suivante :

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x_i \quad (3.13)$$

La matrice de variance intra classe est S_w telle que :

$$S_w = S_1 + S_2 \quad (3.14)$$

La matrice de variance inter classe est S_b :

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (3.15)$$

En sortie on veut trouver 1-q projection linéaire w qui maximise la séparabilité entre les deux classes.

- La projection des données de D_1 et D_2 va générer deux autres groupes y_1 et y_2 :

$$\begin{cases} y_1 = w^T D_1 \\ \text{et} \\ y_2 = w^T D_2 \end{cases} \quad (3.16)$$

- La projection de centre de gravité de chaque groupe est :

$$\tilde{m}_i = w^T m_i \quad (3.17)$$

- La projection des données intra classes :

$$\tilde{S}_w = w^T S_w w \quad (3.18)$$

- La projection des données entre classes :

$$\tilde{S}_b = w^T S_b w \quad (3.19)$$

- La discrimination linéaire de Fisher est donc :

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{S}_1 - \tilde{S}_2} = \frac{|\tilde{S}_b|}{|\tilde{S}_w|} = \frac{w^T S_b w}{w^T S_w w} \quad (3.20)$$

Pour trouver l'hyperplan sur le quel en aura une projection optimal cela revient à trouver l'optimum de l'équation (3.9) (vecteur propre w et valeur propre correspondant).

Telle que $J(w) = \frac{w^T S_b w}{w^T S_w w}$ est maximisé lorsque $S_b w = \lambda S_w w$ ce qui correspond à la projection optimal.

La formulation de Lagrange permet de trouver le vecteur w recherché en annulant la dérivée partielle de de l'équation suivante :

$$L(w, \lambda) = w^T S_b w - \lambda (w^T S_w w - c) / w^T S_w w = c \neq 0 \quad (3.21)$$

On calcule donc $\frac{\partial L(w,\lambda)}{\partial w}$:

$$\frac{\partial L(w,\lambda)}{\partial w} = S_b w - \lambda S_w w \quad (3.22)$$

Ce qui donne : $S_b w^* = \lambda S_w w^*$ et donc $S_w^{-1} S_b w = \lambda w$

$S_b w$ doit être la direction la plus longue entre les deux centres de gravité m_1 et m_2 et donc la solution est $w = S_w^{-1}(m_1 - m_2)$

3.3.4 Analyse discriminante multiple

Dans ce cas on a plusieurs classes :

- La variance intra classes est comme le cas des deux classes: $S_w = \sum_{i=1}^c S_i$
- La variance entre classes est : $S_b = \sum_{i=1}^c (m_i - m)(m_i - m)^T$
- La variance totale : $S_t = \sum_x (x - m)(x - m)^T = S_w + S_b$
- L'objectif de l'analyse discriminantes dans ce cas MDA est de trouver w qui maximise

$$J(w) = \frac{|\tilde{S}_b|}{|\tilde{S}_w|} = \frac{|w^T S_b w|}{|w^T S_w w|}$$

La solution est obtenue par G-EVD : $S_b w_i = \lambda_i S_w w_i$

Où w_i sont les vecteurs propres correspond aux valeurs propres les racines de polynôme caractéristique : $|S_b - \lambda_i S_w| = 0$;

Pour chaque λ_i on trouve w_i apartire de $w_i(S_b - \lambda_i S_w) = 0$

3.4 Conclusion :

L'extraction des paramètres caractéristiques des données a pour but de décrire les signaux EEG et de réduire leur dimension. L'analyse discriminante linéaire ne garde que les informations utiles, ce qui rend leur manipulation plus flexible, afin de gagner en temps et en espace mémoire lors du développement de notre système de classification.

Chapitre 04 :

Méthode de classification Machine à vecteurs support (SVM) et Réseaux de neurones

4.1 Introduction :

Dans le présent chapitre, on donnera un aperçu sur les méthodes de classification utilisées dans le domaine du traitement des signaux EEG, et on développera les deux méthodes choisies qui sont la classification par Machine à Vecteurs Support (SVM) et par Réseaux de Neurones (RNA) ; ensuite on parlera des critères de performances sur lesquels on s'est basé dans le chapitre final pour évaluer les classificateurs proposés.

4.2 Notion de classification :

La classification est l'une des techniques les plus utilisées en médecine. Elle constitue le principe de base de plusieurs systèmes d'aide au diagnostic.

La classification est une opération de structuration qui vise à organiser un ensemble d'observation en groupes homogènes et contrastés afin de faciliter l'analyse des informations et d'effectuer des prédictions [40]. Il s'agit donc dans notre cas d'affecter un ensemble d'objets (enregistrement des encéphalogrammes) à une classe selon un descripteur de celle-ci. Cette description est effectuée grâce à des propriétés ou des conditions typiques propres aux classes. Les objets sont alors classés suivant qu'ils vérifient ou non ces conditions ou propriétés.

La classification est effectuée grâce à des propriétés ou des conditions typiques propres aux classes. Les objets sont alors classés suivant qu'ils vérifient ou non ces conditions ou propriétés. Il existe une multitude de méthodes de classification qui sont utilisées pour classer les signaux EEG parmi ces méthodes, on cite :

- ANN : Artificial Neural Network [41]
- ANFIS : Adaptive Neuro-Fuzzy Inference System [42]
- KNN: K-Nearest Neighbors classification algorithm [43]
- RBFNN : Radial Basis Function Neural Network [44]
- RNN : Recurrent Neural Network [45]

- SVM : Machine à Vecteurs Support [46]

La classification des sujets épileptiques et normaux en fonction de certaines paramètres qui caractérisent leurs signaux EEG se base sur des règles d'apprentissage, c'est la notion qu'on va détailler dans le paragraphe suivant.

4.3 Notion d'apprentissage :

- **L'apprentissage artificiel**, est une branche de l'intelligence artificielle qui peut être mise à contribution pour appréhender les problèmes de prédictions [41]. On peut classer les méthodes d'apprentissage artificiel en deux grandes familles : l'apprentissage non supervisé et l'apprentissage supervisé.

- Dans **l'apprentissage non supervisé**, l'algorithme va lui-même, sans information supplémentaire, catégoriser les variables d'entrées. Ce type d'apprentissage permet l'élaboration d'une représentation interne de l'espace des données d'entrée en identifiant une structure statistique sous-jacente des variables sous une forme plus ou moins simple.

- **L'apprentissage artificiel supervisé**, de façon générale consiste à créer un modèle de prédiction (classification ou prévision) à partir d'une base d'apprentissage comprenant les exemples d'entrée ainsi que les sorties désirées associées. Les paramètres du modèle vont ainsi s'adapter en comparant à chaque fois les sorties obtenues et les sorties désirées, d'où l'appellation supervisé [47].

Une fois le modèle obtenu par une base d'apprentissage, l'utilisation d'une base de test, comprenant des nouveaux exemples non utilisés pendant l'apprentissage, permet de mesurer les performances de la méthode.

Pour la classification des signaux EEG, classiquement, elle est réalisée par la deuxième famille.

- **L'algorithme d'apprentissage** permet de "prédire" une valeur cible étant donnée une ou des valeurs d'entrées. Dans le cas où cette valeur cible est discrète (dans un ensemble fini), la tâche réalisée par l'algorithme est appelée classification supervisée puisqu'il s'agit de trouver la classe correspondant à un exemple donné en entrée.

Soit $E = \{(X(t), y(t)) \in R^d \times R / 0 \leq t \leq N\}$ l'ensemble de données qui peut être une base d'apprentissage ou une base de test. $X(t)$ est le vecteur d'entrée à l'instant t , $X(t) \in R^d$. $y(t)$ représente la valeur cible correspondant à $X(t)$ et N le nombre d'exemples dans la base. Le but de l'algorithme d'apprentissage est de trouver une fonction f , la représentation mathématique du modèle obtenu, qui soit le plus proche possible de la fonction cible \tilde{f} . On notera que \tilde{f} reste

toujours inconnue et dans de nombreux cas hypothétique [48]. Le système pourra être évalué à la fin par l'erreur MSE :

$$MSE = \frac{1}{N} \sum_{t=0}^N (\hat{y}(t) - y(t))^2 \quad (4.1)$$

Où $\hat{y}(t)$ est la valeur prédite c'est-à-dire $f(X(t)) = \hat{y}(t)$ l'approximation de la valeur $y(t)$.

Pour la mise en œuvre de cet algorithme plusieurs méthodes ont été développées, dans ce travail, on s'intéressera à deux types de méthodes : les réseaux de neurones et les supports vecteurs machines.

4.4 Classification par la machine à vecteurs support (SVM) :

Les machines à vecteurs de support ou séparateurs à vaste marge ont été introduites en 1995 par Cortes et Vapnik [40], [49]. Elles sont utilisées dans de nombreux problèmes d'apprentissage : reconnaissance de forme, catégorisation de texte ou encore diagnostic médical. Elles constituent une méthode de classification supervisée particulièrement bien adaptée pour traiter des données de grande dimension. Par rapport aux techniques classiques d'apprentissage.

4.4.1 Concept de la machine à vecteurs support (SVM) :

Les SVM sont une classe d'algorithmes d'apprentissage supervisé, elles sont basées sur les mêmes principes d'apprentissage que les réseaux de neurones. Toutefois, elles possèdent l'avantage d'être plus simples à configurer.

Le modèle du classificateur est construit à partir d'un ensemble d'apprentissage de N exemples étiquetés (x_i, y_i) avec $x_i \in \mathcal{R}^p$ et $y_i \in \{1; -1\}$ selon la classe (p représente la dimension des vecteurs d'entrée ou encore le nombre de caractéristiques dans les exemples d'entrées). Il est à remarquer que la version originale des SVM traite uniquement le cas binaire à deux classes, mais des extensions existent pour la multi-classe. L'apprentissage permet de construire la fonction f de décision appelée aussi hyperplan séparateur

$$f(x) = \text{signe}(\langle w, x \rangle + b) \quad (4.2)$$

Avec $w \in \mathcal{R}^p$ et b les paramètres permettant de déterminer dans quelle partie de l'hyperplan se trouve l'exemple x . Cet hyperplan séparant peut être linéaire ou non linéaire. La figure 4.1 présente le cas non linéairement séparable et linéairement séparable.

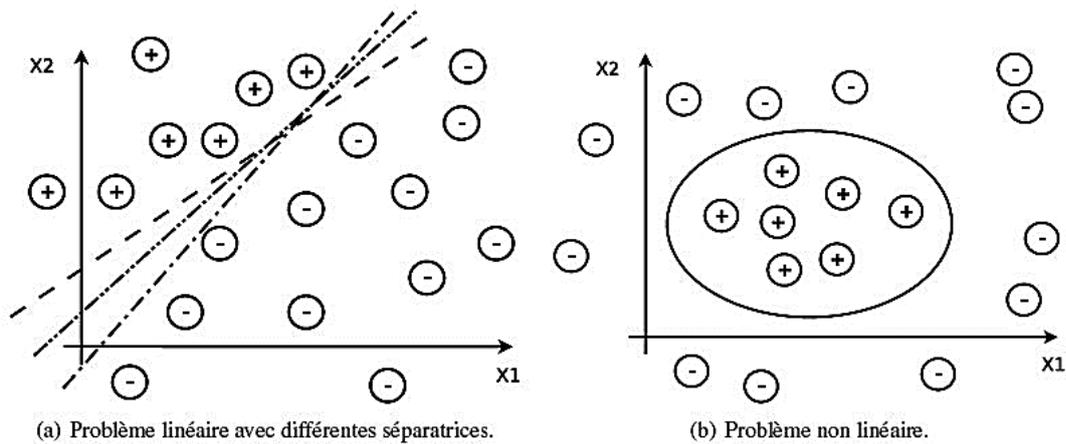


Figure 4.1 : Exemple de classification binaire linéaire et non linéaire dans \mathbb{R}^2 .

La SVM repose donc sur deux notions principales : celle de fonction séparatrice et celle de marge maximale.

La marge Δ est la distance entre la frontière de séparation et les échantillons les plus proches appelés vecteurs de support. Elle est définie par la distance minimale entre les deux points des différentes classes figure (4.2).

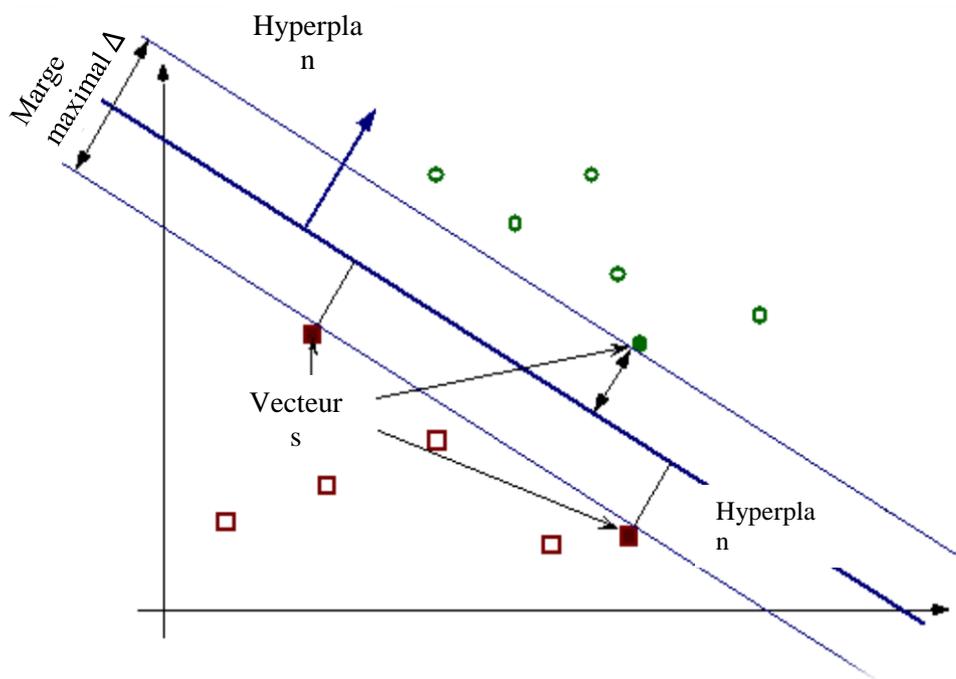


Figure 4.2 : Hyperplan linéaire séparant avec la marge Δ les exemples positifs des exemples négatifs

Dans un problème linéairement séparable, les SVM trouvent une séparatrice qui maximise cette marge. Dans le cas d'un problème non linéaire on utilise une fonction noyau pour projeter les données dans un espace de plus grande dimension où elles seront linéairement séparables.

4.4.2 Cas des données linéairement séparables

- L'hyperplan séparateur

La méthode générale de construction de l'hyperplan optimal (HO) qui sépare des données appartenant à deux classes différentes linéairement séparables :

Soit $H: (w \cdot x) + b$ l'hyperplan qui satisfait les conditions suivantes :

$$\begin{cases} w \cdot x_i + b \geq 1 & \text{si } y_i = 1 \\ w \cdot x_i + b \leq -1 & \text{si } y_i = -1 \end{cases} \quad (4.3)$$

Ce qui est équivalent à :

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{pour } i = 1, \dots, m \quad (4.5)$$

Dans ce cas il existe une infinité d'hyperplans séparateurs comme nous pouvons le voir sur la figure (4.3).

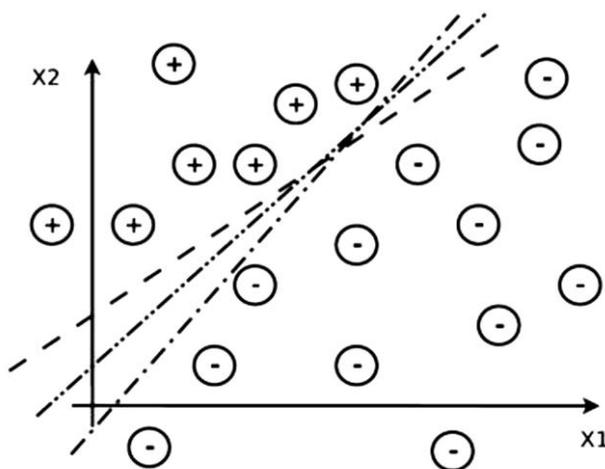


Figure 4.3 : les différents hyperplans séparateurs

L'idée des SVM dans ce cas est de choisir l'hyperplan optimum selon la théorie de Vapnik [49] c'est-à-dire l'hyperplan qui donnera la règle qui se généralisera le mieux à d'autres données que celles de l'ensemble d'apprentissage. Afin de déterminer ce qui caractérise le meilleur hyperplan.

- La recherche de l'hyperplan optimale

Maximiser la marge Δ est équivalent à maximiser la somme des distances des deux classes par rapport à l'hyperplan. Ainsi la marge a l'expression mathématique suivante :

$$M = \min_{x_i|y_i=1} \frac{w \cdot x + b}{\|w\|} - \max_{x_i|y_i=-1} \frac{w \cdot x + b}{\|w\|} = \frac{2}{\|w\|} \quad (4.6)$$

Trouver HO revient donc à maximiser $\frac{2}{\|w\|}$, ce qui est équivalent à minimiser $\frac{\|w\|^2}{2}$ sous la contrainte (4.5). Ceci est un problème de minimisation d'une fonction objectif quadratique avec contraintes linéaires. Ce problème peut se résoudre en utilisant les principes de Fermat, de Lagrange ou de Kuhn-Tucker [50], [51].

En appliquant le principe de Kuhn-Tucker on est amené à rechercher un point (w_0, b_0, α) . Le lagrangien correspondant à notre problème est

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^m \alpha_i [(y_i (x_i \cdot w) + b) - 1] \quad (4.7)$$

Le lagrangien doit être minimal par rapport à w et b et maximal par rapport à $\alpha \geq 0$.

$L(w, b, \alpha)$ est minimal par rapport à b :

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^m \alpha_i y_i = 0 \quad (4.8)$$

$L(w, b, \alpha)$ est minimal par rapport à w :

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Leftrightarrow w - \sum_{i=1}^m \alpha_i x_i y_i = 0 \quad (4.9)$$

$L(w, b, \alpha)$ est maximal par rapport à $\alpha \geq 0$.

En remplaçant (4.8) et (4.9) dans le lagrangien (4.7) on aura :

$$L(w, b, \alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (4.10)$$

Ainsi le problème est maximisé $L(w, b, \alpha)$ sous la contrainte :

$$\sum_{i=1}^m \alpha_i y_i = 0; \quad \alpha_i \geq 0 \quad (4.11)$$

Soit la solution $\alpha^0 = (\alpha_1^0, \dots, \alpha_m^0)$. D'après le théorème de Kuhn-Tucker une condition nécessaire et suffisante pour que α^0 soit optimal est :

$$\alpha_i^0 [y_i [(w_0 \cdot x) + b_0] - 1] = 0 \quad \text{pour } i = 1, \dots, m \quad (4.12)$$

Ce qui veut dire que : $\alpha_i^0 = 0$ ou $y_i [(w_0 \cdot x) + b_0] = 1$

Les vecteurs supports SV sont définis comme étant tout vecteur x_i tel que :

$y_i [(w_0 \cdot x) + b_0] = 1$. Ce qui est équivalent à :

$$VS = \{x_i / \alpha_i > 0\} \text{ pour } i = 1, \dots, m \quad (4.13)$$

Ainsi, w_0 et b_0 sont facilement calculés :

$$w_0 = \sum_{VS} \alpha_i^0 y_i x_i \quad (4.14)$$

$$b_0 = -\frac{1}{2} [w_0 \cdot x^* (1)] + [w_0 \cdot x^* (-1)] \quad (4.15)$$

La fonction de classement $class(x)$ est définie par

$$class(x) = sign[(w_0 \cdot x) + b_0] = sign[\sum_{x_i \in VS} \alpha_i^0 y_i (x_i \cdot x) + b_0] \quad (4.16)$$

Si $class(x)$ est inférieur à 0, x est de la classe -1 sinon il est de la classe 1.

4.4.3 Cas des données non linéairement séparables

- La fonction noyau :

Les problèmes de classification sont souvent non linéaires. Pour résoudre ce point, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de redescription ». En effet, intuitivement, plus la dimension de l'espace de redescription est grande plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma de la figure suivant :

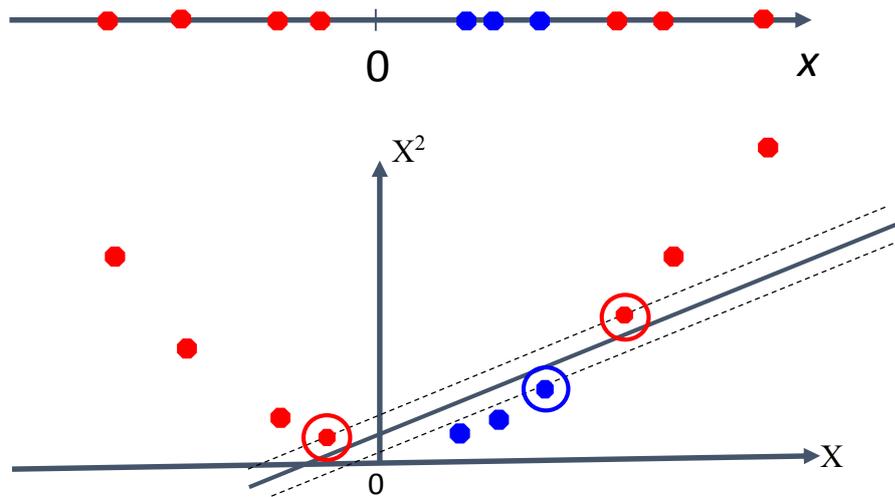


Figure 4.4 : Espace de redescription

L'application de la transformation non linéaire (\cdot) aux vecteurs d'entrée x_i tel que $x_i \in \mathbb{R}$ et $\Phi(x_i) \in \mathbb{R}^2$ va conduire à passer d'un produit scalaire dans l'espace d'origine $x_i \cdot x_j$ à un produit scalaire $\Phi(x_i) \cdot \Phi(x_j)$ dans l'espace de redescription (voir la figure 4.5 Φ). L'astuce est d'utiliser une fonction noyau notée K qui évite le calcul explicite du produit scalaire dans l'espace de redescription. Les fonctions noyaux doivent satisfaire le théorème de Mercer.

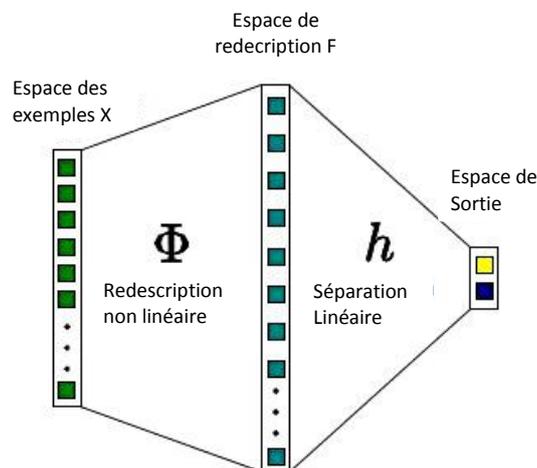


Figure 4.5 : La transformation linéaire des données permet une séparation linéaire dans un nouvel espace. Adapté de [52].

Nous avons alors l'égalité suivante :

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (4.17)$$

Il existe de nombreuses fonctions noyau prédéfinies, les deux les plus usitées sont le noyau gaussien (équation IV.19) et le noyau polynomial (équation IV.20) :

$$K_\gamma(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (4.18)$$

$$K_{\gamma,d,r}(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d \quad (4.19)$$

Les noyaux gaussien sont des noyaux dits de type radiale (fonction à base radial abrégé RBF), indiquant qu'ils dépendent de la distance entre deux exemples.

- L'hyperplan séparateur :

L'hyperplan séparateur se réécrit avec la fonction noyau sous la forme suivante :

$$h(x) = \sum_{i=1}^p \alpha_i^* y_i K_\gamma(x, x_i) + b \quad (4.20)$$

L'hyperplan optimal est celui qui satisfait les conditions suivantes :

- la distance maximale entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale

Pour formaliser ces conditions, des variables de pénalité non-négative ($\xi_i, i = 1, \dots, m$) appelées encore variables d'écart sont introduites. Ces variables transforment l'inégalité (4.5) comme suit :

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad \text{pour } i = 1, \dots, m \quad (4.17)$$

L'objectif est de minimiser la fonction suivante :

$$\psi(w, E) = \frac{1}{2} w \cdot w + C \sum_{i=1}^m \xi_i \quad (4.18)$$

Où C est un paramètre de régularisation. Il permet de définir l'importance aux erreurs. Cela mène à un problème légèrement différent de celui du cas des données linéairement séparables, à savoir maximiser le lagrangien donné par l'équation (4.10) par rapport à α_i sous les conditions suivantes :

$$\sum_{i=1}^m \alpha_i y_i = 0 \text{ avec } 0 \leq \alpha_i \leq C \text{ pour } i = 1, \dots, m \quad (4.19)$$

Le calcul de la normale w_0 , du biais b_0 et de la fonction de classification $\text{class}(x)$ reste exactement le même que pour les données linéairement séparables.

4.4.4 Cas d'utilisation d'un classificateur non linéaire :

Les classificateurs SVM à noyau utilisent le même principe de l'hyperplan optimal pour calculer une frontière entre les nuages des points. Cependant l'idée principale est cette fois de construire une surface de séparation linéaire dans un nouvel espace, l'espace des caractéristiques, qui correspond à une surface non-linéaire dans l'espace d'entrée.

Soit l'ensemble D' l'image de l'ensemble D , défini par la transformation Φ .

$$D' = \{(\Phi(x_i), y_i) \in \mathbb{R}^p * \{-1,1\} \text{ pour } i = 1, \dots, m \mid p \geq n\} \quad (4.20)$$

En construisant un hyperplan optimal dans l'espace des caractéristiques suivant la technique expliquée dans le paragraphe précédent, la fonction de classement sera :

$$\text{class}(x) = \text{sign} \left[\sum_{x_i \in VS} \alpha_i^0 y_i (\Phi(x_i) \cdot \Phi(x)) + b_0 \right] \quad (4.21)$$

Pour que le coût de calcul reste pratiquement inchangé et que le nombre de paramètres libres du système n'augmente pas, il faut que la fonction Φ satisfasse la condition suivante :

$$\Phi(u) \cdot \Phi(v) = K(u, v) \quad (4.22)$$

C'est à dire que le produit scalaire dans l'espace des caractéristiques va être représenté comme un noyau de l'espace d'entrée.

Plusieurs noyaux ont été utilisés par les chercheurs, en voici quelques-uns :

- Le noyau linéaire : $K(u, v) = u \cdot v$
- Le noyau polynomial : $K(u, v) = [(u \cdot v) + 1]^d$

Où d est le degré du polynôme à déterminer par l'utilisateur.

- Le noyau RBF (Radial Basis Function) : $K(u, v) = \exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right)$

Où σ est à déterminer.

Ayant défini le noyau, la fonction du classement (5. 37) devient :

$$\text{class}(x) = \text{sign}[\sum_{x_i \in VS} \alpha_i^0 y_i K(x_i, x) + b_0] \quad (4.23)$$

4.5 Classification par les réseaux de neurones artificiels :

La deuxième technique de classification appliquée est le réseau de neurones artificiels (RNA) qui est caractérisé par sa propriété d'approximation universelle, sa vitesse de décision et sa bonne capacité de généralisation.

4.5.1 fonctionnement général d'un réseau de neurone formel.

Un réseau de neurones artificiel est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des neurones biologiques. Le neurone est l'unité de traitement de base dans un réseau de neurones artificiel. Un neurone artificiel est composé des éléments suivants : les nœuds d'entrée, les connections synaptiques dont les forces sont représentées par des poids et la fonction d'activation φ qui relie toutes les entrées à la sortie de neurone.

La figure (4.6) présente le schéma d'un neurone biologique et la structure d'un neurone formelle

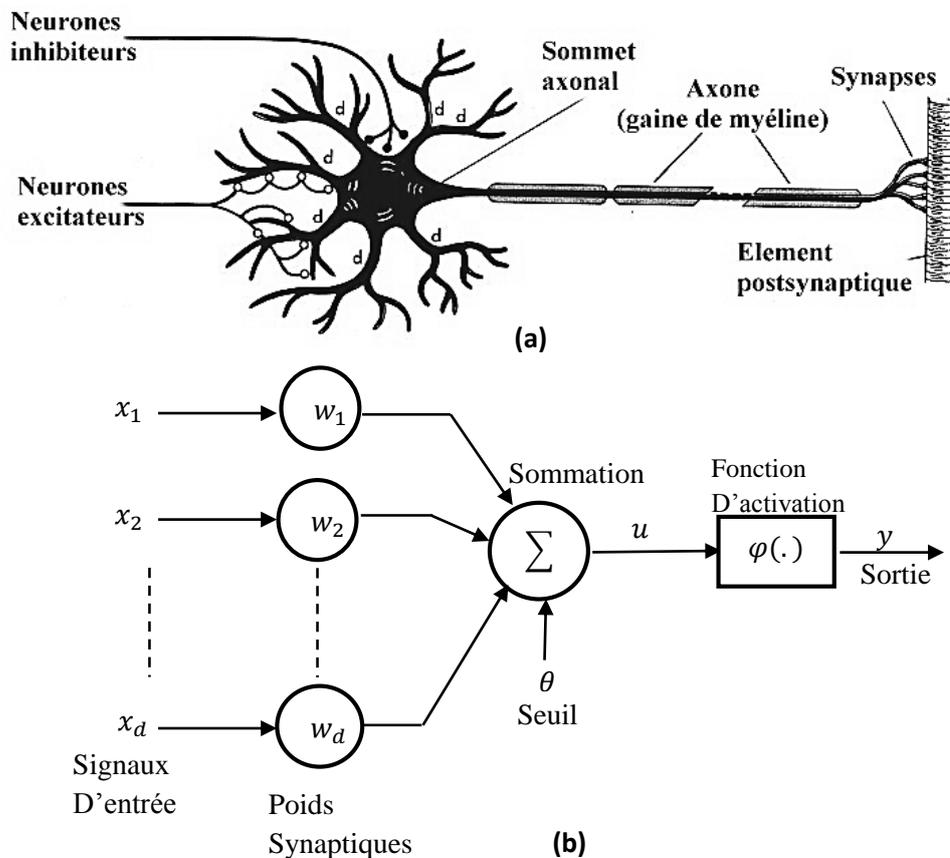


Figure 4.6 : (a) représente le neurone biologique et (b) représente la structure de base d'un neurone artificiel.

Le tableau suivant illustre la similitude existant entre le neurone biologique et le neurone formelle.

Tableau 4.1. Analogie entre les neurones biologiques et artificiels

Neurone biologique	Neurone artificiel	Représentation symbolique
Soma	Neurone	
Dendrite	Intrant (Input)	x_i, X (intranet net)
Axone	Extrant (output)	Y
Synapse	Poids	W_i

Dans le réseau de neurones, les neurones sont fortement connectés entre eux et travaillent en parallèles. Ce neurone formel reçoit et émet des signaux binaires (0/1). La somme pondérée des signaux d'entrée, constituant l'activation du neurone, se transforme en sortie après son passage par une fonction de seuillage ou de transfert. Et donc la somme de toutes les entrées est comparée à un seuil θ . Si ce seuil est dépassé, le neurone s'active, sinon il ne transmet aucun signal [53].

Chaque neurone est une fonction non linéaire paramétrée telle que la fonction de transfert de chaque neurone est la suivante :

$$y = \varphi(u) \quad (4.24)$$

Avec
$$u = \sum_{i=1}^d x_i w_i - \theta = XW - \theta \quad (4.25)$$

Où X et W sont les vecteurs d'entrée et de poids respectivement, φ est la fonction d'activation.

$X = [x_1, x_2, x_3, \dots, x_d]$: Le vecteur d'entrée

$W = [w_1, w_2, w_3, \dots, w_d]$: Le vecteur de poids

Les poids des connexions sont généralement estimés par minimisation d'un critère de moindres carrés, qui permet d'estimer les probabilités a posteriori.

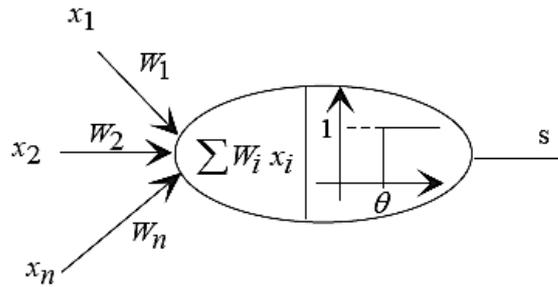


Figure 4.7 : illustration de fonctionnement de neurone biologique [53]

La fonction d'activation φ peut-être :

- La fonction signe :
$$\varphi(u) = \begin{cases} 1 & \text{si } u > 0 \\ -1 & \text{si } u \leq 0 \end{cases}$$
- La fonction sigmoïde de paramètre λ :
$$\varphi(u) = \frac{1}{1+e^{-\lambda u}}$$

Il y a aussi différentes autres fonctions qui peuvent être utilisées.

4.5.2 Architecteurs des réseaux de neurones.

Plusieurs types de réseaux de neurones existent. Selon la connectivité, On distingue deux grandes familles : les réseaux non bouclés et les réseaux bouclés [54].

La figure suivante représente les réseaux les plus utilisés

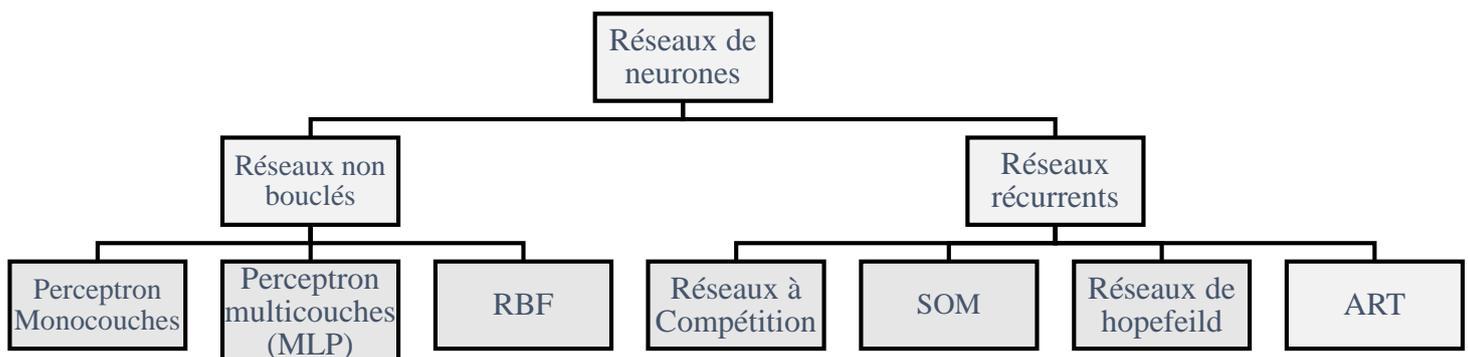


Figure 4.7 : Les réseaux les plus utilisés

La variante du réseau de neurones artificiels, qui va être appliquée dans ce travail, inclue la structure non bouclée ou ce qu'on appelle propagation directe "feedforward", à savoir le modèle de Perceptron Multicouches (MLP). Cette méthode a été utilisée par [55], [56].

4.5.3 Les réseaux « feedforward » à rétro propagation

Le FFNN est un réseau qui comprend trois variantes (MLP, PM, RBF), l'information dans ce type circule dans une seule direction, de l'entrée vers la sortie, il n'autorise pas le retour en arrière de l'information. Le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé, pour cette raison, le FFNN est aussi qualifié que d'autres types de réseaux. L'algorithme d'apprentissage le plus utilisé pour les FFNN ainsi que leurs variantes est l'algorithme de la Rétro-Propagation du gradient, en anglais Back-Propagation (BP)[57]. C'est un algorithme qui est souvent utilisé dans la majorité d'expériences dont le principe est de rétro-propager le gradient de l'erreur de la couche de sortie vers l'entrée.

On trouve souvent que la variante MLP est, en générale, la plus utilisée pour les applications biomédicales telles que la classification.

4.5.3.1 L'algorithme de rétro propagation :

C'est un algorithme qui permet à un réseau multicouche d'apprendre par mises à jour des poids des différentes couches. Comme on va considérer des réseaux avec plus d'une unité de sortie, on va redéfinir l'erreur E :

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in \text{exemples}} \sum_{k \in \text{sorties}} (\text{cible}_{kd} - \text{sortie}_{kd})^2 \quad (4.26)$$

De la même manière que nous avons considérée E comme une surface d'erreur pour les unités linéaires, nous allons chercher dans le cas de la rétro-propagation à minimiser la surface d'erreur avec notre nouvelle définition. Dans le cas de l'unité linéaire, nous n'avons qu'un seul extremum qui était le minimum global. Avec des unités sigmoïdes, nous allons avoir plusieurs extrema. Malheureusement, on n'est pas sûr que l'algorithme converge vers le minimum global, il peut converger vers un minimum local. En pratique cependant, la rétro-propagation donne de bons résultats.

On va décrire l'algorithme pour un réseau de sigmoïdes à deux couches, mais il est possible de l'étendre à un nombre quelconque de couches [58].

- **Créer un réseau à 2 couches de neurones**
 - Initialiser les poids aléatoirement avec de petites valeurs. L'entrée de l'unité i dans l'unité j est notée x_{ij} et les poids de l'unité i à l'unité j sont notés w_{ij} .
- Pour chaque e dans l'exemple :
- Calculer la sortie de chaque neurone du réseau
- Pour chaque unité k dans les neurones de sortie :
- Calculer le terme d'erreur $\delta_k = sortie_k(1 - sortie_k)(cible_k - sortie_k)$
- Pour chaque unité h dans neurones cachés :
- Calculer le terme d'erreur $\delta_h = sortie_h(1 - sortie_h) \sum_{k \in sorties} w_{kh} \cdot \delta_k$
- Pour chaque poids du réseau w_{ij} :
- Mettre à jour les poids $w_{ij} = w_{ij} + \eta \cdot \delta_j \cdot x_{ji}$

Cet algorithme est spécifique aux réseaux de neurones à 1 couche cachée avec des unités sigmoïdes mais on peut facilement l'adapter à un nombre quelconque de couches cachées et pour des unités à fonction d'activation différentiables [58].

4.5.3.2 La variante Perceptron multicouches

Les réseaux Perceptron Multicouche (Multi Layer Perceptron : MLP) ont été utilisés pour résoudre des problèmes difficiles par apprentissage supervisé à partir d'un algorithme de "retro-propagation de l'erreur" qui se base sur la règle d'apprentissage par correction d'erreur. Le processus de rétro propagation d'erreur consiste à effectuer deux passages à travers les différentes couches du réseau : un passage direct et un passage rétrograde.

Pendant le passage direct, les poids synaptiques du réseau sont fixes, et en rétrograde les poids synaptiques sont ajustés à partir d'une règle de correction de l'erreur. Ces ajustements continuent jusqu'à ce que la réponse du réseau s'approche de la réponse désirée au sens de l'ensemble d'apprentissage ou de la fonction.

a. Architecture du réseau :

Considérons un Perceptron Multicouches formé d'une couche de L entrées, d'une couche de S neurones cachés et d'une couche de M neurones de sortie. Le réseau est entièrement connecté, c.à.d. chaque neurone d'une couche est relié à l'ensemble des neurones de la couche suivante.

Entrées : $X_n; n = 0, 1, \dots, L; X_0 = 1$

Sorties : $y_k; k = 1, \dots, M.$

Sorties internes : $Z_j : j = 1, \dots, S; Z_0 = 1$

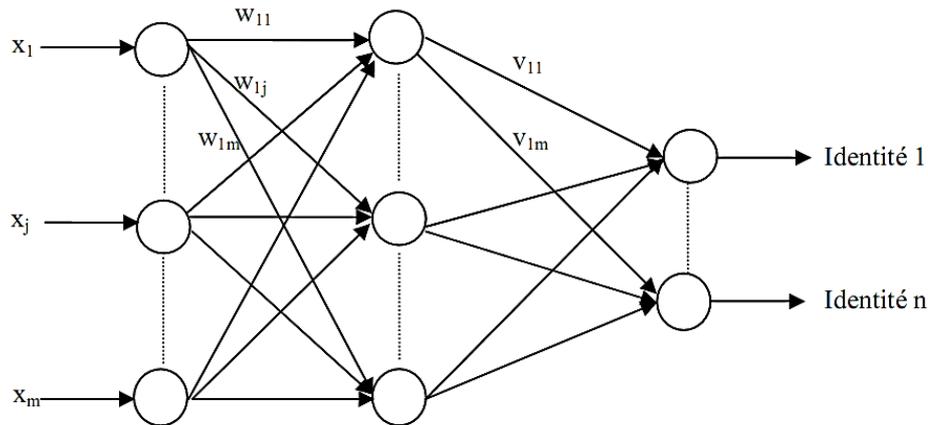


Figure 4.8 : Architecture d'un réseau de neurones multicouches.

b. Equations de fonctionnement :

Connexions entrées- neurones cachés : $u_j = \sum_{n=0}^L w_{jn} x_n ; j = 1, \dots, S$

Fonctions d'activation : $Z_j = \Phi(u_j)$

Connexion neurones cachés-sorties : $v_k = \sum_{j=0}^S w_{kj} Z_j ; k = 1, \dots, M$

Fonction d'activation : $y_k = \Phi(v_k)$

c. Stratégie d'apprentissage

L'erreur commise par le réseau, définie comme la différence entre la sortie attendue et la sortie du réseau, va permettre d'adapter les poids de la structure afin de minimiser cette erreur. Plusieurs algorithmes existent afin de réaliser cet apprentissage. On peut citer par exemple la rétro propagation du gradient.

L'erreur au carré entre les sorties désirées et les sorties calculées par le réseau est :

$$E = \frac{1}{2} \sum_{k=1}^M (d_k - y_k)^2$$

Pour la classification des signaux EEG, le nombre d'entrées de la première couche de réseau dépend du nombre de paramètres extraits pour chaque signal. Plus le nombre de paramètres augmente, plus la complexité de calcul augmente. Pour ce qui sont des sorties de réseau, leur nombre dépend de nombre de classes à déterminer.

Dans la classification des signaux EEG avec les réseaux de neurones, chaque signal est représenté par un vecteur colonne qui contient les différents paramètres caractéristiques, ce qui donne une matrice qui sera envoyé à l'entrée de réseau.

4.6 Les différentes applications de RNA

Les réseaux de neurones sont utilisés dans plusieurs domaines, citons la classification, la reconnaissance de formes, l'identification et la commande de processus. Le choix d'utiliser tel ou tel type de réseau de neurones dépend de l'application mais aussi des capacités de traitement du processeur sur lequel ils s'exécutent ; sans tarder voici quelques domaines [59] :

- Reconnaissance des formes.
- Traitement de la parole.
- Détection d'anomalies en médecine (qui est l'axe de notre travail).
- Traitement dépendant du temps (prédiction, d'identification et de commande de processus).

4.7 Les paramètres de performance

Pour évaluer les performances des deux classificateurs proposés, on s'est intéressé à un ensemble de paramètres qui sont les plus répandus lors des recherches effectuées dans le domaine de la classification des signaux biomédicaux, [60],[61].

- **Précision** : rapport des vrais positifs(TP) et vrais négatifs(TN). Elle mesure le taux de classification correcte des cas normaux et épileptiques, Elle est donnée par la formule suivante :

$$précision = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

- **Sensibilité** : rapport des vrais positif (TP), elle mesure le taux de classification correcte des cas normaux. Elle est donnée par la formule suivante :

$$Sensibilité = \frac{TP}{TP + FN} \times 100\%$$

- **Spécificité** : rapport des vrais négatifs (TN), elle mesure le taux de classification correcte des cas épileptiques. Elle est donnée par la formule suivante :

$$Spécificité = \frac{TN}{TN + FP} \times 100\%$$

La sensibilité de test mesure sa capacité à donner un résultat positif lorsqu'une hypothèse (normale ou épileptique) est vérifiée. Elle s'oppose à la spécificité, qui mesure la capacité de test à donner un résultat négatif lorsque cette hypothèse n'est pas vérifiée.

- **Taux d'erreur** : mesure l'erreur commise lors de la classification des cas normaux et épileptiques. Il est donné par la formule suivante :

$$\text{Taux d'erreur} = \left(1 - \frac{TP + TN}{TP + FN + FP + TN}\right) \times 100\%$$

- **TP** (true positive) : vrai positif signifie que le cas normal est classé dans la classe des normaux. représente le nombre d'individus malades avec un test positif
- **TN** (true negative) : vrai négatif signifie que le cas épileptique est classé dans la classe des épileptiques.
- **FP** (false positive) : faux positif signifie que le cas épileptique est classé dans la classe des normaux.
- **FN** (false negative) : faux négatif signifie que le cas normal est classé dans la classe des épileptiques.

Table 4.2 : Table de diagnostic (confusion)

		Classe réelle	
		Classe +	Classe -
Classe prédite	Classe +	TP (vrai positifs)	FP (faux positifs)
	Classe -	FN (faux négatifs)	TN (vrais négatifs)

Les paramètres ci-dessus permettent d'évaluer la capacité de classification des deux classificateurs proposés, mais on souhaite encore aborder un autre point qui est la comparaison des deux classificateurs, pour cela on utilise un autre paramètre de performance qui est la courbe ROC.

- La courbe Roc

La mesure la plus couramment utilisée pour comparer deux classificateurs est le taux d'erreur (ou de façon équivalente, le taux de classification correcte) obtenue sur le jeu de validation. Dans le cas de la classification binaire, des outils permettent à la fois de décrire les caractéristiques des classificateurs étudiés et de les comparer entre eux. Les courbes ROC et «rappel-précision » remplissent cette tâche [62].

La courbe ROC « receiver operating characteristic » un outil graphique permet de représenter la capacité d'un test à discriminer entre la population des malades et des non-malades. Elle représente en ordonnée la proportion de tests positifs parmi la population malade (la sensibilité) en fonction de la proportion de tests positifs parmi la population non-malade (complément de la spécificité ou $1 - \text{spécificité}$, en abscisse) voir la figure (4.8),

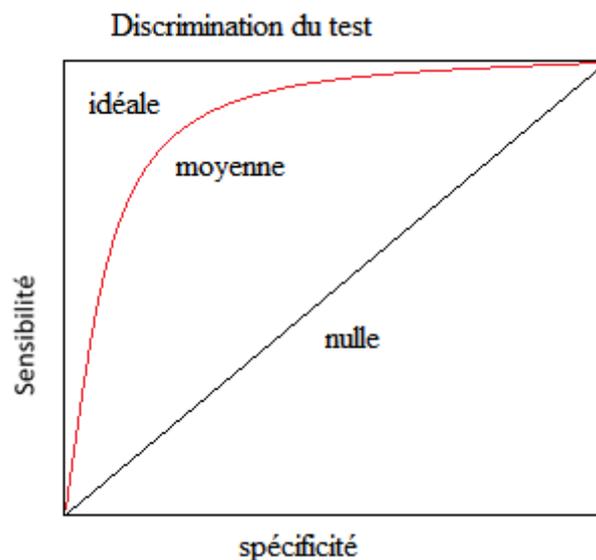


Figure 4.8 : La courbe ROC

4.8 Conclusion :

Dans le présent chapitre on a présenté les concepts d'apprentissage et de classification introduite pour les deux méthodes SVM et RNA. On a abordé tout d'abord la méthode SVM du point de vue intuitif et mathématique. On a vu que l'avantage de cette méthode est la possibilité de l'utiliser pour des données de grandes dimensions, mais elle est applicable pour des tâches de classification à deux classes, ensuite on a vu les bases des réseaux de neurones après avoir donné un aperçu sur la similitude neurone biologique/neurone artificiel, on s'est intéressé plus particulièrement à l'étude des architectures de type perceptron multicouches

entraînées avec l'algorithme de la rétro propagation du gradient. Les performances de la classification neuronale, qui est l'objet de ce travail sont exposées en détail au chapitre 5.

Chapitre 05 :

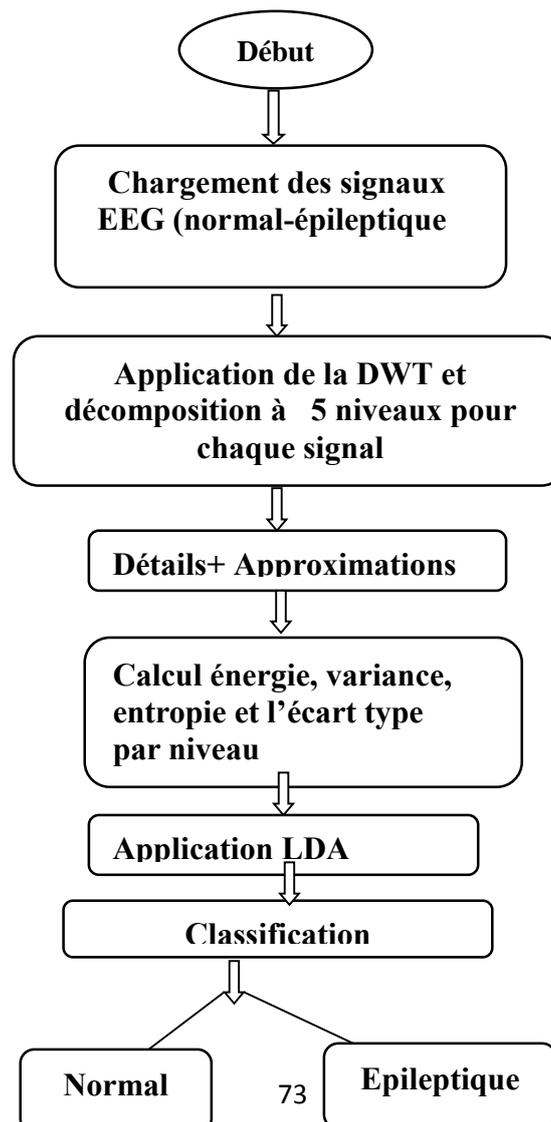
Résultats et évaluation des performances

5.1 Introduction :

Dans ce présent chapitre on va mettre en évidence et en pratique toutes les notions citées précédemment afin de réaliser un classificateur automatique qui représente une aide au diagnostic de la pathologie épileptique, on trouve dans cette partie les résultats obtenus par la classification appliquée sur des données réelles, ainsi l'évaluation et la vérification de la fiabilité de système de classification conçu. Pour cela, on évalue les résultats de chaque étape de développement de ce système de la classification.

5.2 Système automatique de classification des signaux EEG :

Le processus de réalisation : est représenté par l'organigramme suivant



Le prétraitement (analyse temps fréquence): pour ce faire on a utilisé la transformée en ondelettes discrète (DWT) pour la décomposition des signaux EEG afin d'avoir les coefficients constituant les signaux de la base de donnée utilisée (set A, B, C,D,E) [44]. Le choix de la DWT a été fait selon la nature de ces signaux qui sont aléatoire et non stationnaire.

L'extraction des paramètres : l'étape qui suit le prétraitement et l'extraction des paramètres caractéristiques de tous les coefficients calculé par la DWT comme l'énergie, l'entropie, l'écart type, le max, min, gamme,... Et cela est pour décrire les signaux EEG, afin de pouvoir les classifier par la suite.

Réduction de la dimension : on applique après l'extraction une réduction de donnée par LDA sur les paramètres calculé précédemment pour éviter les redondances et les informations inutiles dans le but de réduire la dimension.

Classification : une fois la réduction est réalisée nous utiliserons ces résultats pour faire une classification par les deux méthodes SVM et FFBNP sur trois groupes :

- Groupe1 : composé de Set A \rightarrow Set E,
- Groupe2 : composé de Set (A+B) \rightarrow Set (C+E),
- Groupe3 : composé de Set (A+B) \rightarrow Set (D+C),
- Groupe4 : composé de Set (A+B) \rightarrow Set (D+E),

Le groupe 1 a été employée juste pour évaluer la méthode de réduction LDA , et pour déterminer les bons paramètres pour la fonction SVM (le choix de la fonction Kernel)

On discutera par la suite les performances de notre classificateur en se basant sur trois paramètres qui sont : la précision, la sensibilité et la spécificité.

Pour cela nous avons exploité la plateforme Matlab (R2013a) avec ces différentes fonctions incluant d'autres fonctions disponibles dans la Wavelet Toolbox, Bioinformatics Toolbox.

Au final on affichera les différentes simulations effectuées et on discutera les résultats obtenus. Un récapitulatif des résultats de ce classificateur sera donné pour pouvoir l'évaluer et déterminer les paramètres menant à une classification performante.

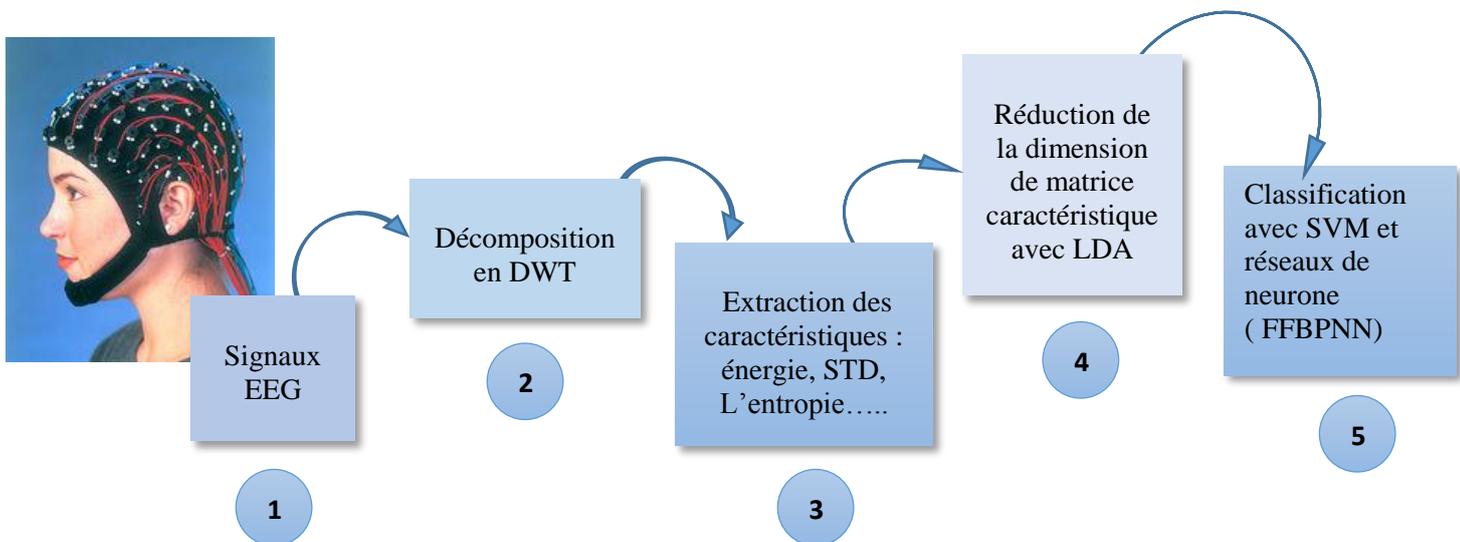


Figure 5.1 : Structure du système de classification automatique de signaux EEG

5.3 Signaux EEG :

On a utilisé la base de données mise à la disposition du public [44], comprenant 5 ensembles (dénotés A, B, C, D, E), chacun d'eux contient 100 signaux d'une durée de 23,6s. Les ensembles A et B ont été pris à partir d'enregistrements EEG de surface de cinq volontaires sains avec les yeux ouverts et fermés, respectivement. L'ensemble E correspond aux signaux présentant des crises épileptiques. Les ensembles A et B ont été enregistrés par voie extra crânienne, tandis que les ensembles C (la commissure hippocampique de l'hémisphère opposé du cerveau), D (la zone épileptogène) et E ont été enregistrés par voie endocrânienne.

Tous les signaux EEG ont été enregistrés avec le système d'amplification à 128 canaux, en utilisant une moyenne de référence commune. Les données ont été numérisées à 173.61 échantillons par seconde en utilisant une résolution de 12 bits. La bande passante est de 0,53 à

40 Hz (12 dB / oct). La figure (5.2) représente les signaux EEG obtenus à partir des Set A, B, C, D et E

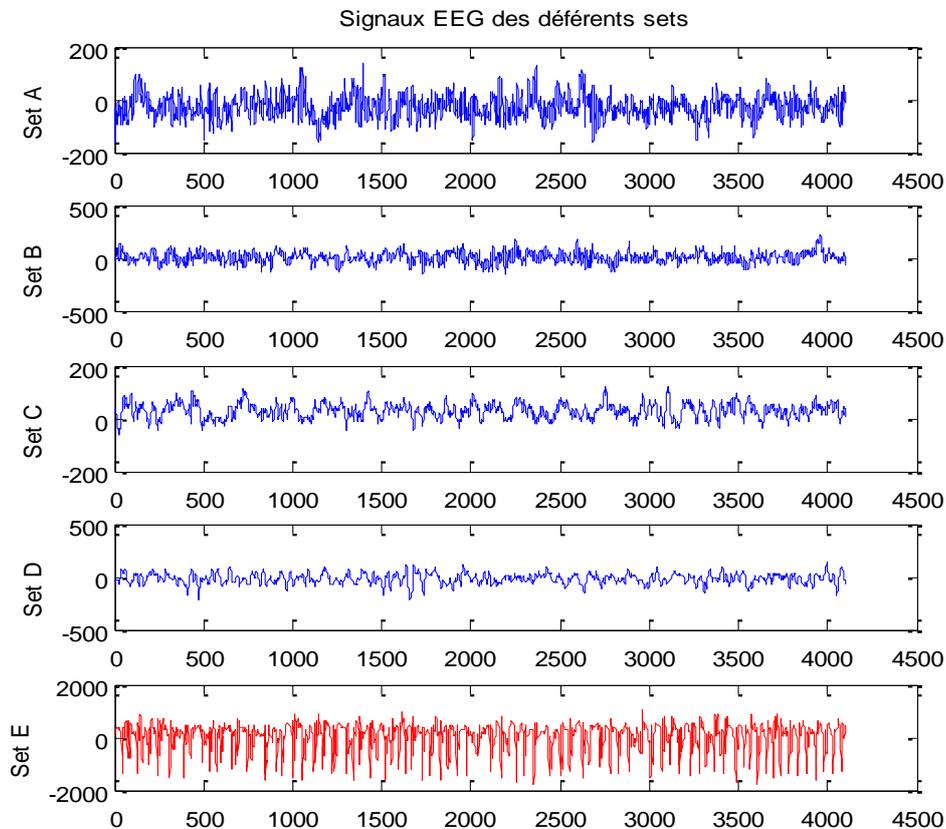


Figure 5.2 : Représentation temporelle des signaux EEG.

5.4 Application de la DWT aux signaux EEG :

La sélection des ondelettes appropriées et le nombre de niveaux de décomposition est important dans l'analyse des signaux EEG :

- Le choix d'ondelette analysante :

Dans le cadre d'analyse des signaux EEG plusieurs recherches ont été effectuées concernant le choix de l'ondelette analysante ; on trouve souvent dans la littérature que les ondelettes de COIFLET et de Daubechies sont les plus utilisées.

Dans ce travail on applique la Daubechies d'ordre 4 'db4' qui a été utilisée par [24], [25], [26], [27] car elle donne des bons résultats.

- Le nombre des niveaux de décomposition :

Le nombre de niveau de décomposition est choisi en fonction des composantes de fréquences dominantes du signal. Dans notre étude comme les signaux EEG n'ont pas de composantes de

fréquence utiles au-dessus de 80 HZ [63] on s'arrête au niveau 5. Par conséquent, le signal sera décomposé en cinq (05) coefficients d'approximation A1 jusqu'à A5 et Cinq (05) coefficients de détail D1 jusqu'à D5. Le tableau 2.1 résume les gammes de fréquences de chaque sous-bande.

Tableau 5.1 : Gamme de fréquence

Niveau de décomposition	Fréquences(Hz)
D1	43.4 – 86.8
D2	21.7 – 43.4
D3	10.8 – 21.7
D4	5.4 – 10.8
D5	2.7 – 5.4
A5	0 – 2.7

- Décomposition sans décimation

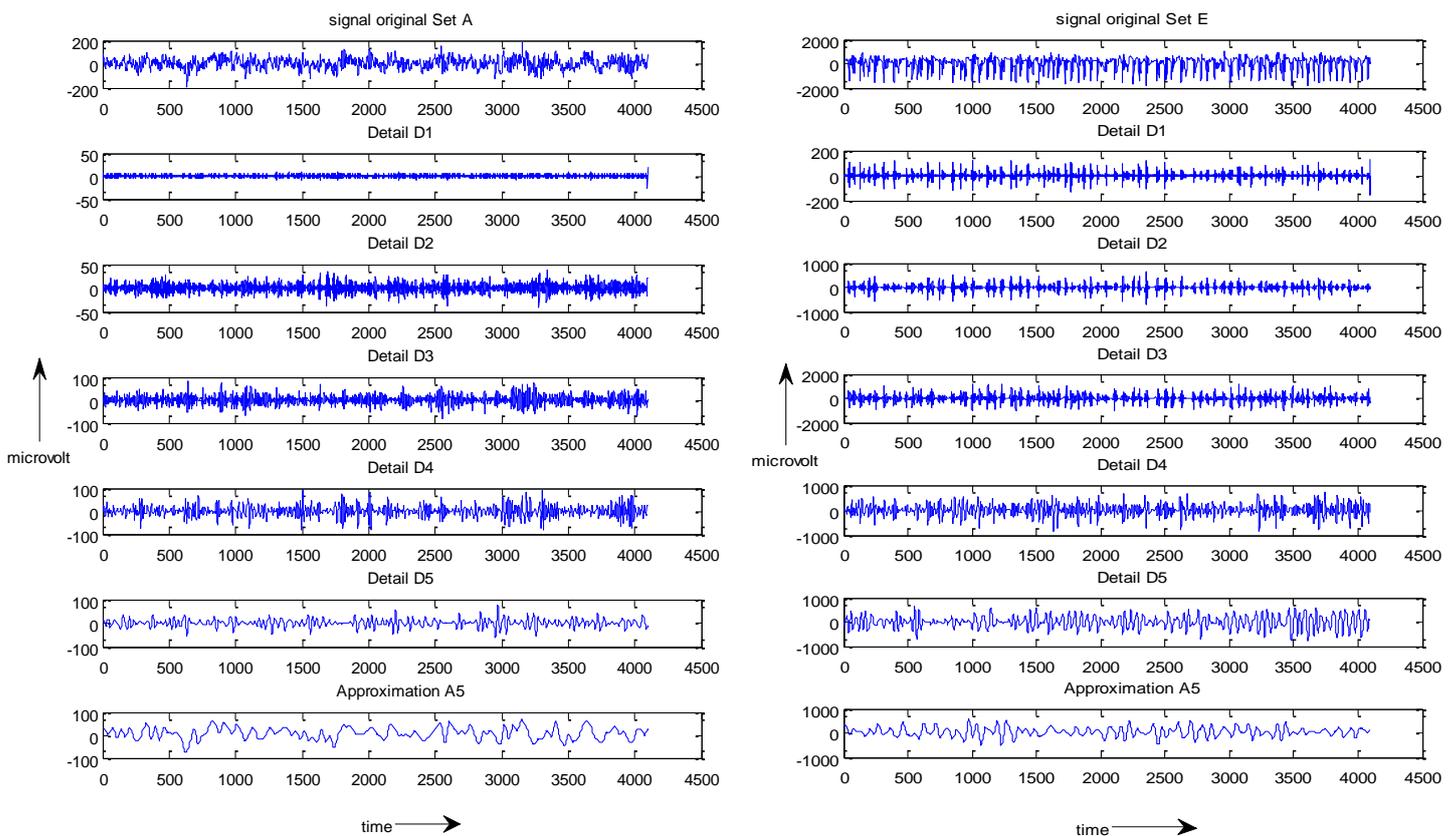


Figure5.3 : Décomposition sans décimation de deux signaux EEG du Set A et Set B

- **Décomposition avec décimation**

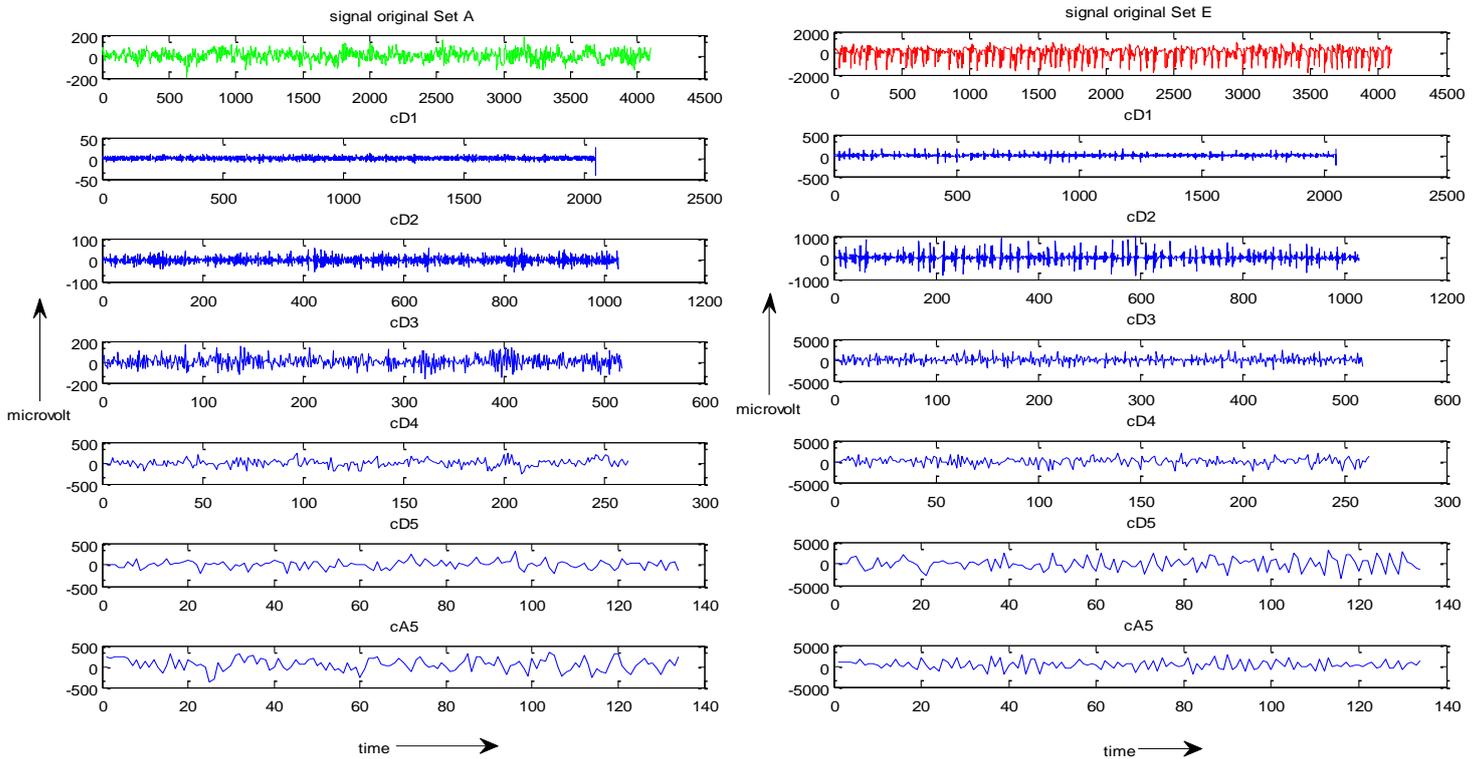


Figure 5.4 : Décomposition avec décimation de deux signaux EEG du Set A et Set B

Interprétation : Dans le cas de la décomposition sans décimation, on note que le filtrage passe bas classique enlève l'information à haute fréquence, mais laisse l'échelle inchangée et la dimension des différents niveaux de décomposition est la même, ce qui donne des données redondantes de grande dimension, voir figure (5.3).

Pour le cas de la décomposition avec décimation, on note qu'à chaque niveau de décomposition l'échelle augmente et la dimension diminue. Sachant que le filtrage passe bas demi-bande enlève la moitié des fréquences sans aucune perte d'information, et le sous-échantillonnage après filtrage n'affecte pas la résolution [31], on a utilisé cette méthode pour le prétraitement des signaux EEG. Les résultats obtenus sont représentés par la figure (5.4).

5.5 Extraction des paramètres caractéristiques :

Dans cette section, on a extrait les paramètres caractéristiques à savoir le Maximum (Max), le Minimum (Min), la Gamme (Gam), l'Écart-type (Ect), l'Entropie (Ent), et l'Énergie (Eng) avec lesquels on va manipuler les signaux EEG pour pouvoir les classer par la suite.

Les résultats obtenus sont représentés dans le tableau suivant

Tableau 5.2: les valeurs des paramètres caractéristiques

Set A						
	ENT	ENG	STD	MAX	MIN	MEAN
A5	-8.7560 10 ⁷	59,9258	158,8707	194,9115	-647,3119	-176,4743
D1	-2.4393 10 ⁵	8,99629	5,3713	24,4836	-32,98385	0,1407
D2	-2.9239 10 ⁶	12,4377	20,4368	71,4144	-72,45226	0,0495
D3	-1.6693 10 ⁷	14,7520	59,8772	245,0704	-218,5969	-0,4460
D4	-1.4924 10 ⁷	3,41693	77,3279	295,6630	-195,56859	2,9024
D5	-1.0969 10 ⁷	0,47123	92,1916	223,1826	244,7397	-0,5571
Set B						
A5	-5.0801 10 ⁷	41,2373	180,5769	914,8598	-430,7591	22,7268
D1	-2.2097 10 ⁵	11,3155	5,0325	51,5042	-38,8225	-0,0155
D2	-2.4283 10 ⁶	22,8510	18,8561	53,7659	-77,3319	0,0664
D3	-2.0284 10 ⁷	20,6889	65,3856	282,9209	-223,1707	0,6687
D4	-2.4296 10 ⁷	3,4209	96,6689	232,4171	-304,5847	-3,1217
D5	-1.1855 10 ⁷	0,4861	95,3339	255,5380	-195,5858	-1,3556
Set C						
A5	-6.1114 10 ⁷	79,7694	122,1227	459,5171	2563,3856	161,3153
D1	-3.4557 10 ⁴	11,6345	2,4796	22,6795	168,7899	-0,0414
D2	-1.9494 10 ⁵	5,07345	6,5288	20,3231	991,5181	-0,0082
D3	-1.2499 10 ⁶	2,6996	18,9194	117,4891	2467,7895	0,4017
D4	-2.7745 10 ⁶	0,6389	36,5070	108,0211	2122,7629	0,5722
D5	-7.5829 10 ⁶	0,1839	77,3661	197,4476	3050,4916	3,6893
Set D						
A5	-6.9268 10 ⁷	51,4233	184,5852	284,7733	-580,1922	-106,8945
D1	-1.9889 10 ⁴	31,5502	2,0408	14,6616	-9,8824	0,0284
D2	-2.4661 10 ⁵	13,3828	7,1274	30,5633	-21,9864	-0,0324
D3	-2.7216 10 ⁶	3,1281	26,7010	96,1519	-94,4963	-0,6297
D4	-1.4995 10 ⁷	0,4429	77,6900	234,3669	-195,1415	3,0688
D5	-4.1857 10 ⁷	0,0724	167,2334	424,6647	-718,3162	-0,7513
Set E						
A5	-2.3239 10 ⁹	16,4993	1050,4916	2563,3856	-2109,6311	296,8814
D1	-1.6616 10 ⁷	26,6827	30,3811	168,7899	-231,0194	-0,3855
D2	-5.9563 10 ⁸	19,6371	217,6710	991,5181	-827,7920	0,0420
D3	-4.3917 10 ⁹	31,9161	770,2641	2467,7895	-2201,3306	5,6766
D4	-2.6963 10 ⁹	5,0676	850,0801	2122,7629	-2333,7795	22,3452
D5	-3.8797 10 ⁹	0,1969	1388,2996	3050,4916	-3330,2753	-29,4267

Le but de ces calculs est de chercher à identifier les caractéristiques du signal EEG qui permettent de réaliser la meilleure classification possible (EEG normal ou pathologique), pour

Cela on a représenté ces paramètres pour les différents niveaux afin de pouvoir choisir ceux qui donnent une bonne classification.

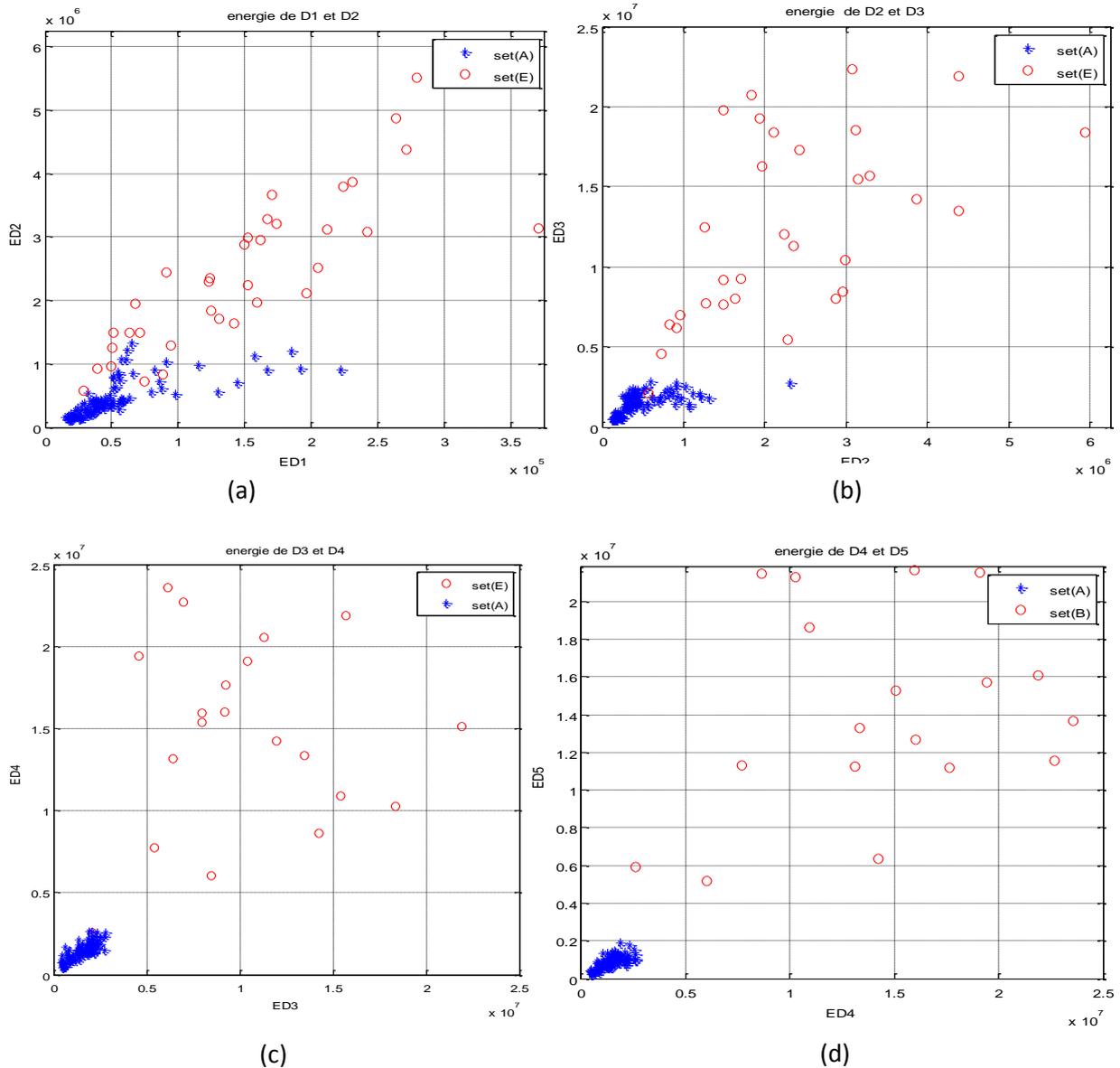


Figure 5.5 : représentation d'énergie des différents niveaux. (a) : énergie de niveau 1 et niveau 2, (b) : énergie de niveau 2 et 3, (c) : énergie de niveau 4 et 5.

On a effectué ces représentations pour le reste des paramètres et on a remarqué que les niveaux 4 et 5 donnent de bonnes séparations entre la population épileptique et normale.

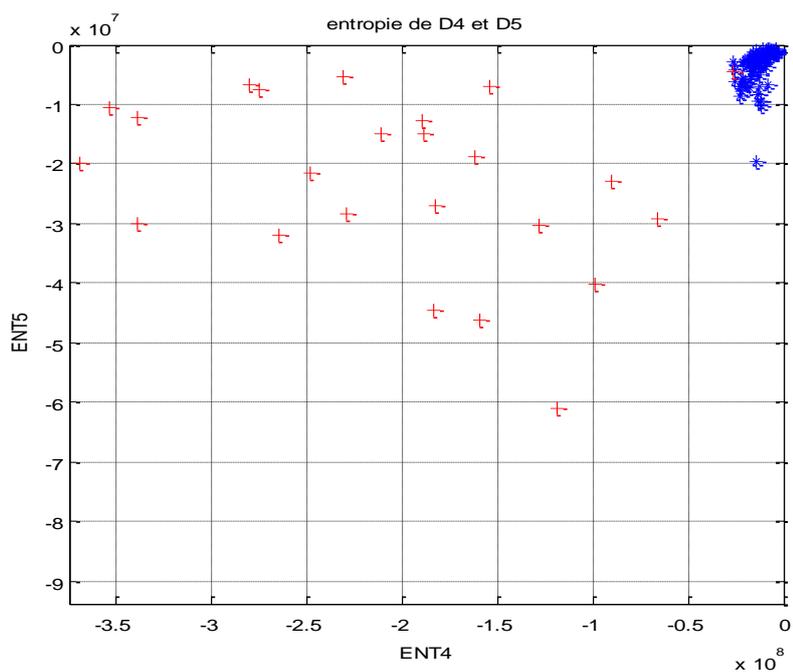


Figure 5.6 : représentation d'entropie de niveaux 4 et 5

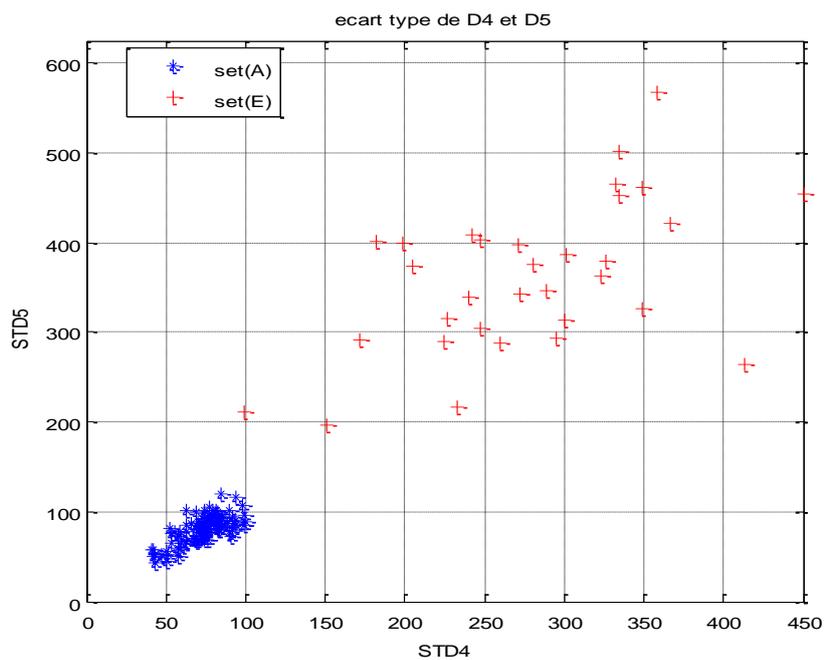


Figure 5.7 : représentation d'écart type de niveaux 4 et 5

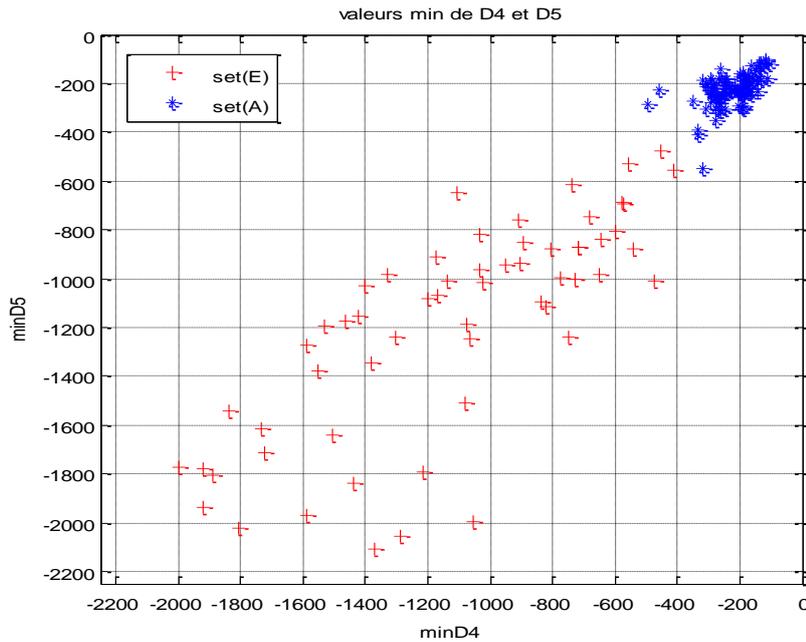


Figure 5.8 : représentation des valeurs minimales de niveau 4 et 5

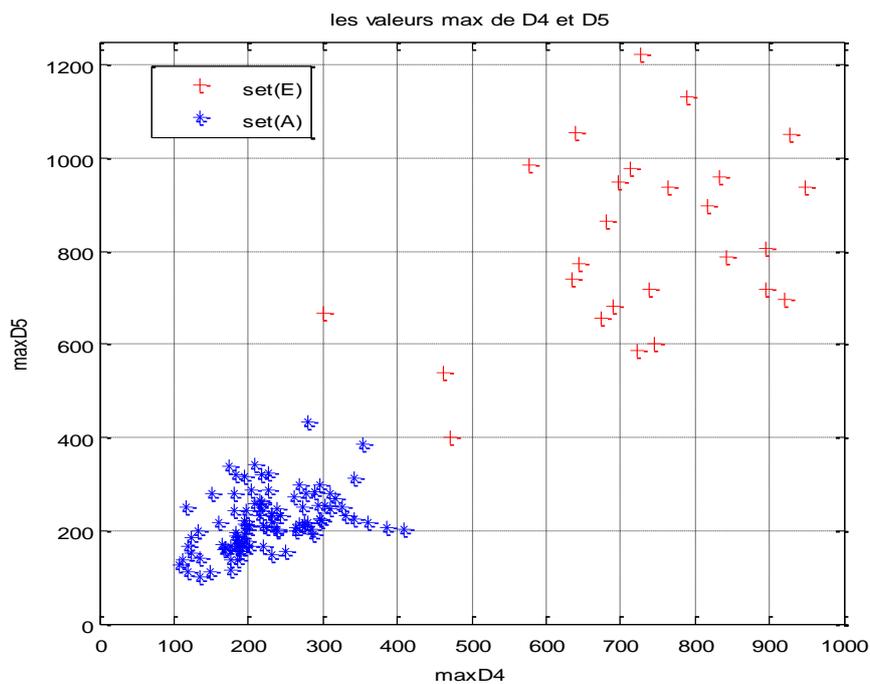


Figure 5.9 : représentation des valeurs maximales de niveaux 4 et 5

D'après ces représentations on note que ces paramètres donnent de bonnes séparations, donc sont nécessaires pour la classification.

5.6 Réduction de dimensions par LDA :

Dans ce travail, on a proposé une méthode de réduction dans les sous-espaces discriminants de Fisher. Le résultat de la réduction de dimension est représenté par les figures suivantes (5.6) :

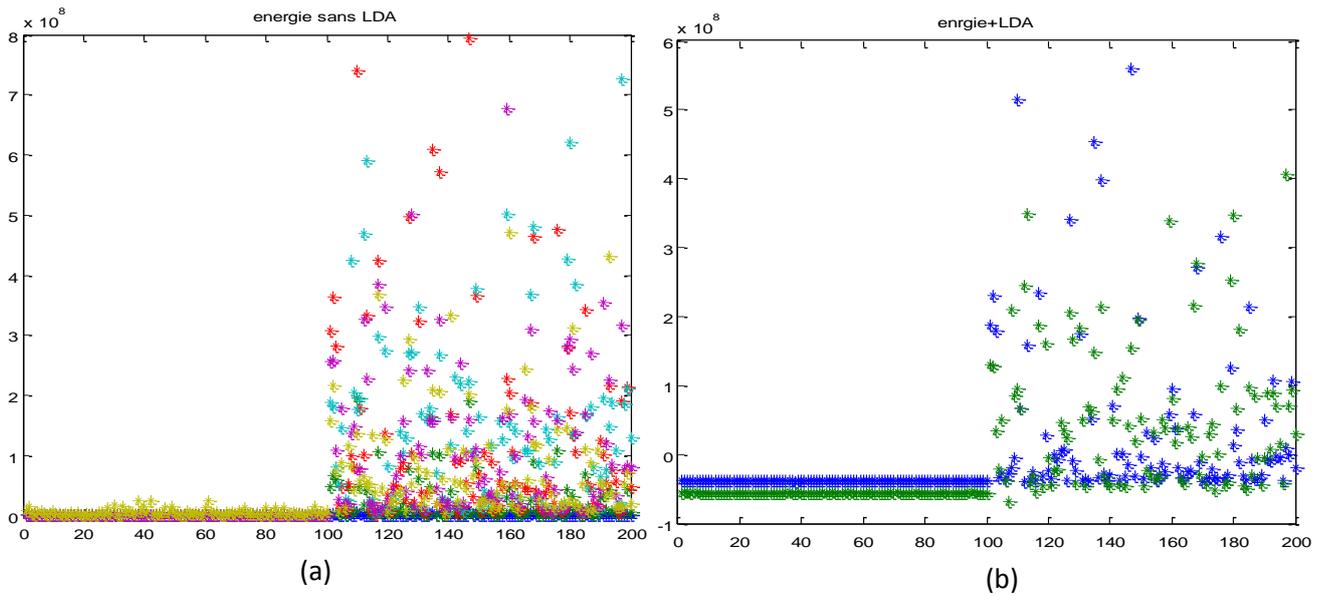


Figure 5.6 : Réduction de dimension par LDA (paramètre d'énergie). La figure (a) représente le paramètre d'énergie pour les différents niveaux sans réduction, (b) représentation d'énergie avec réduction par LDA.

Afin d'évaluer l'efficacité de cette méthode dans le problème de classification, on va comparer les résultats de classification de quelques paramètres caractéristiques par SVM avec LDA et sans LDA. Les résultats trouvés sont représentés dans l'histogramme suivant figure (5.7).

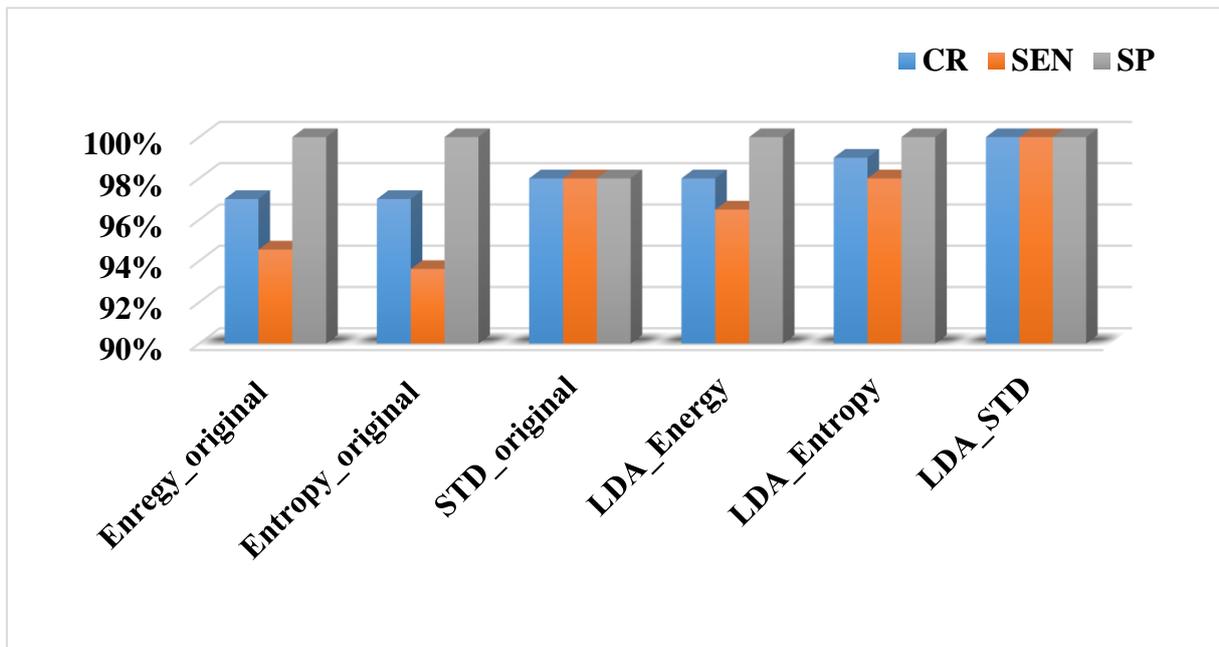


Figure 5.7 : les performances de classification par SVM avec LDA et sans LDA

On remarque que le pourcentage de la classification correcte dans le cas où on a utilisé la réduction par LDA est plus élevé par rapport au cas où on a classifié directement les paramètres originaux.

De plus cette approche s'est avérée plus performante que les méthodes traditionnelles pour une classification automatique d'après [64].

LDA s'est révélée être peu sensible à l'augmentation de la dimension et aux données bruitées. De plus, le nombre de paramètres à estimer est indépendant de la dimension originale des données.

5.7 Classification par SVM :

- **La validation croisée :**

La validation croisée est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. En fait, il y a au moins trois techniques de validation croisée :

- « Holdout methode »
- « K-fold cross-validation »
- « Leave-one-out cross validation.»

La première méthode est très simple, il suffit de diviser l'échantillon de taille n en échantillons d'apprentissage (>60% de l'échantillon) et échantillons de test. Le modèle est bâti sur l'échantillon d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant l'erreur quadratique moyenne.

Dans la seconde, on divise k fois l'échantillon, puis on sélectionne un des k échantillons comme ensemble de validation et les $(k-1)$ autres échantillons constitueront l'ensemble d'apprentissage. On calcule comme dans la première méthode l'erreur quadratique moyenne. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les $(k-1)$ échantillons qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. La moyenne des k erreurs quadratiques moyennes est enfin calculée pour estimer l'erreur de prédiction.

La troisième méthode est un cas particulier de la deuxième méthode ou $k=n$, c'est-à-dire que l'on apprend sur $(n-1)$ observations puis on valide le modèle sur la n ème observation et l'on répète cette opération n fois [47].

- **la fonction noyau :**
 - quadratique : noyau quadratique
 - polynôme : noyau polynomiale
 - MLP : Perceptron multicouches noyau (MLP)
 - RBF : Fonction gaussienne de base radiale

Dans ce travail on a pris la fonction noyau polynomiale avec une validation croisé de type « holdout »,

L'ordre du polynôme a été déterminé après plusieurs essais, en comparant au fur et à mesure les performances pour différents ordres.

L'ordre du polynôme pour la classification avec les paramètres (énergie, entropie, écart type) a été fixé à 11. Pour les paramètres restants on a fixé l'ordre à 9.

Tous ces essais ont été effectués sur les données de groupe 1.

La classification avec les différents paramètres sont représentées par les figures ci-dessous.

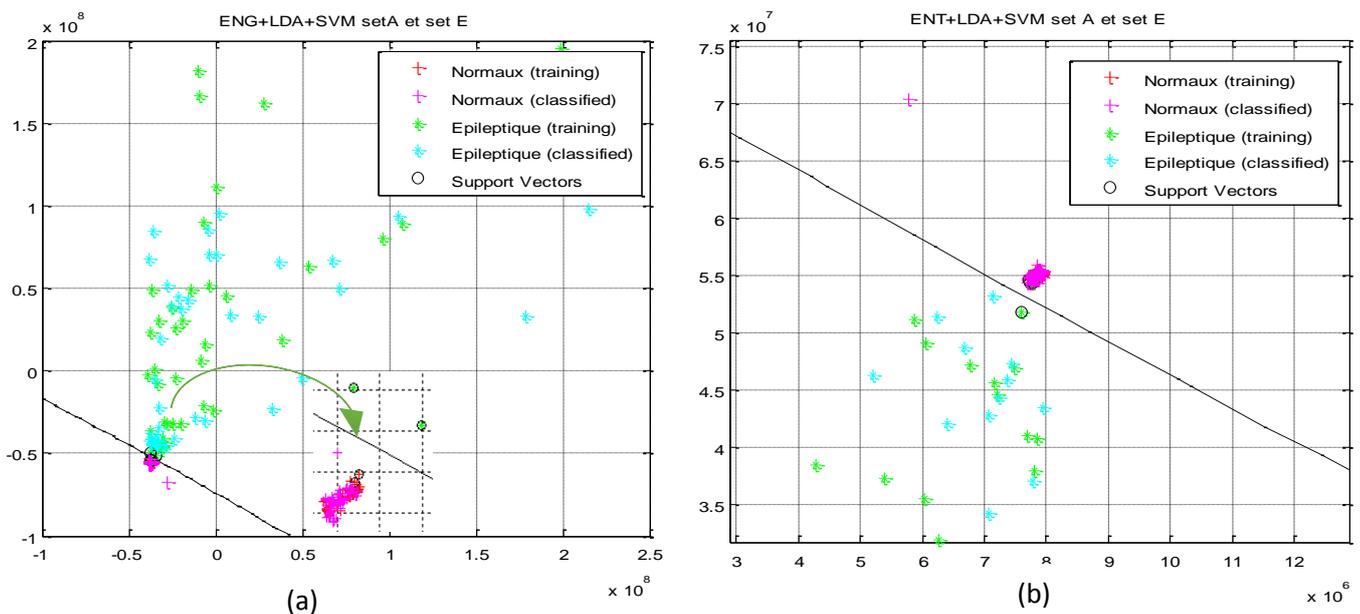


Figure 5.8 : classification entre set A et set E par SVM on utilisant les paramètres non statistiques (a) énergie, (b) entropie

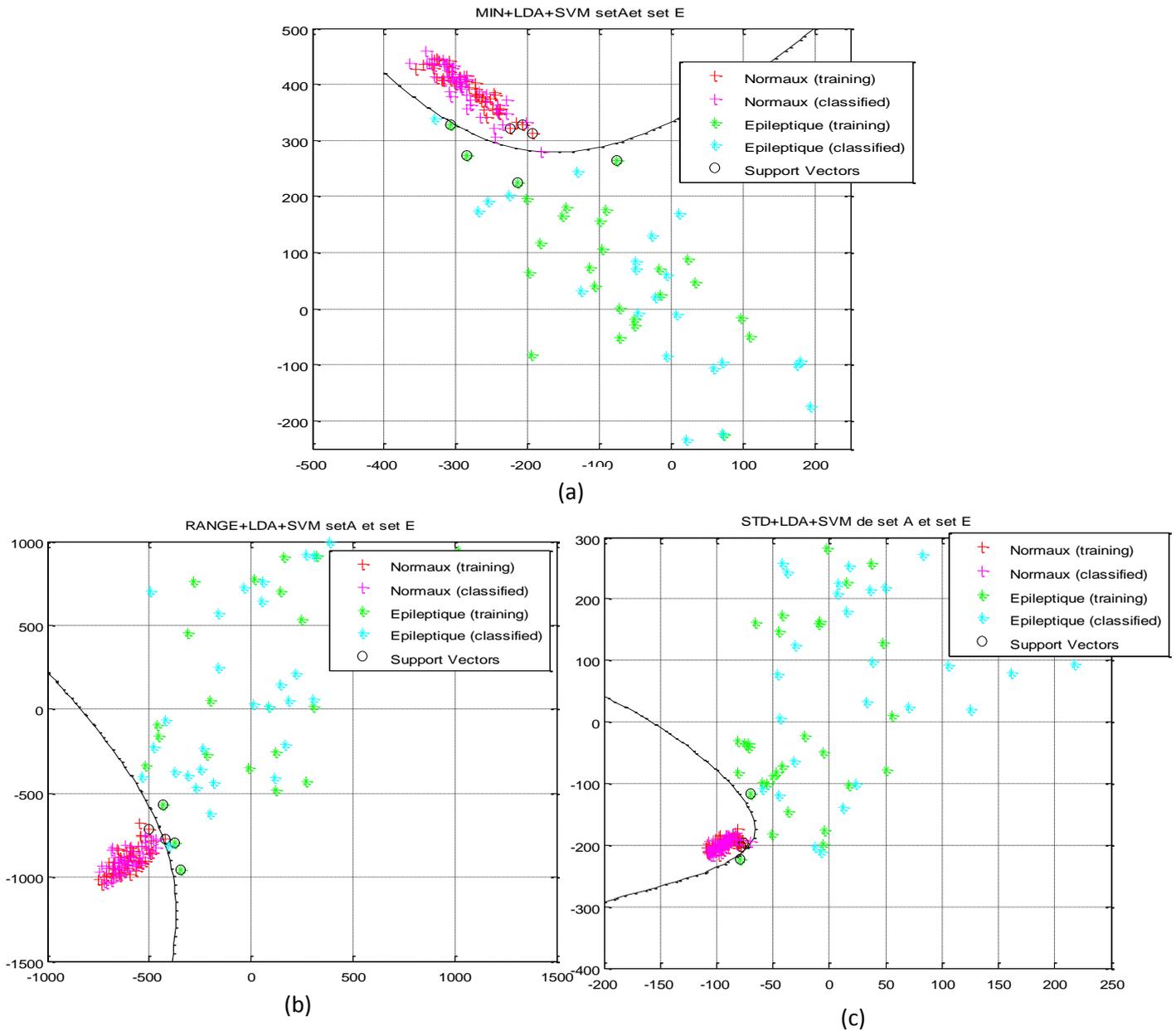
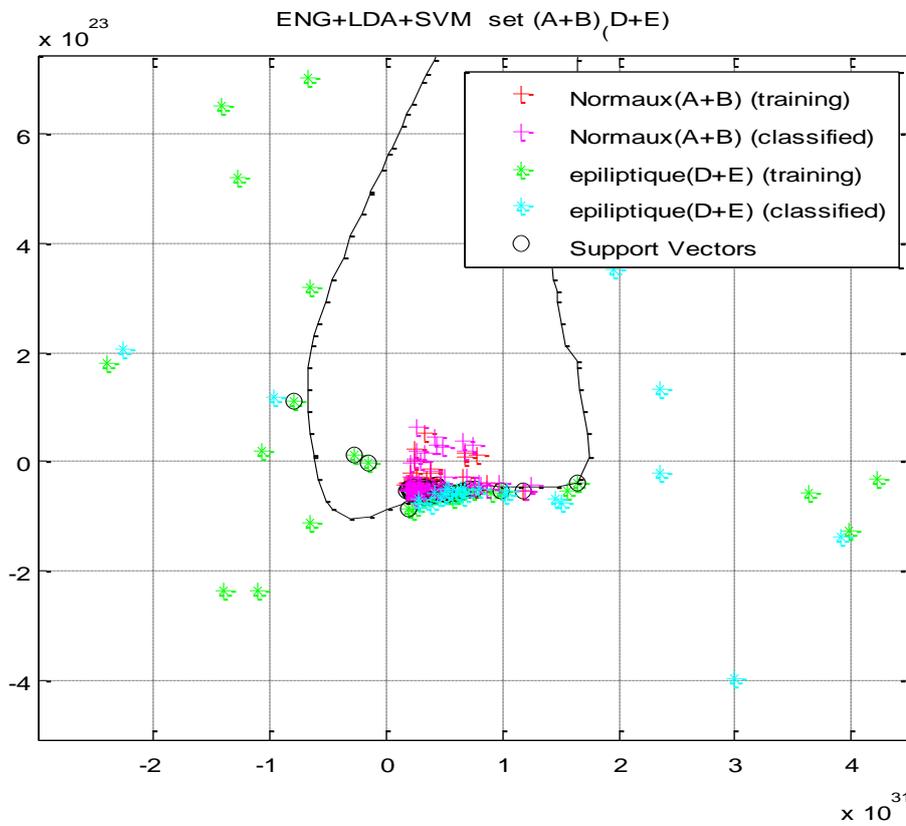
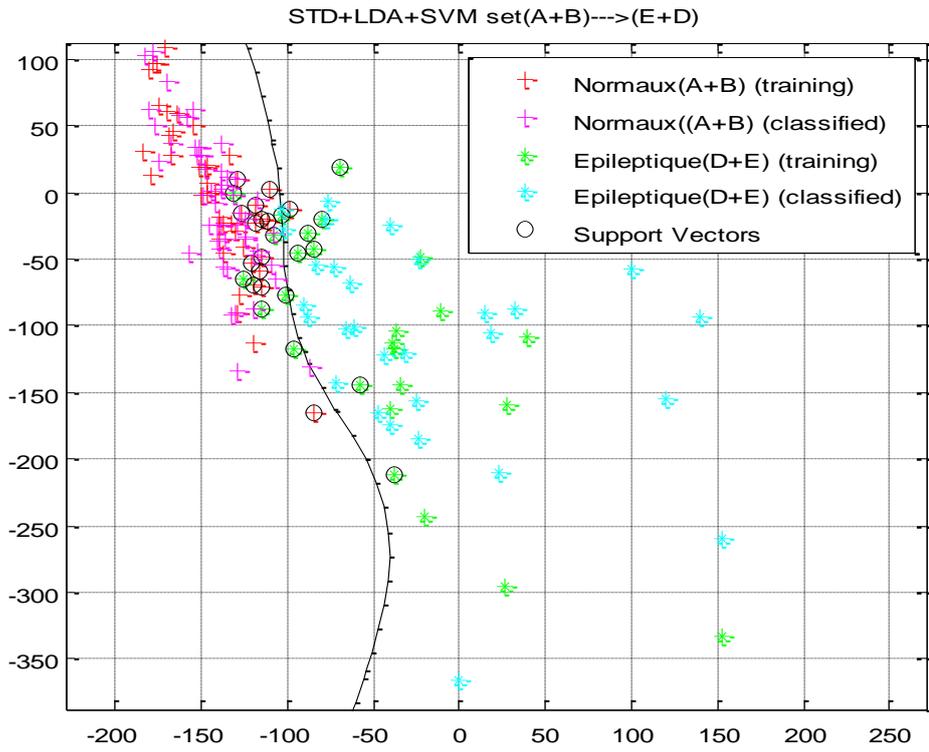


Figure 5.9 : classification entre set A et set E par SVM on utilisant les paramètres statistiques.

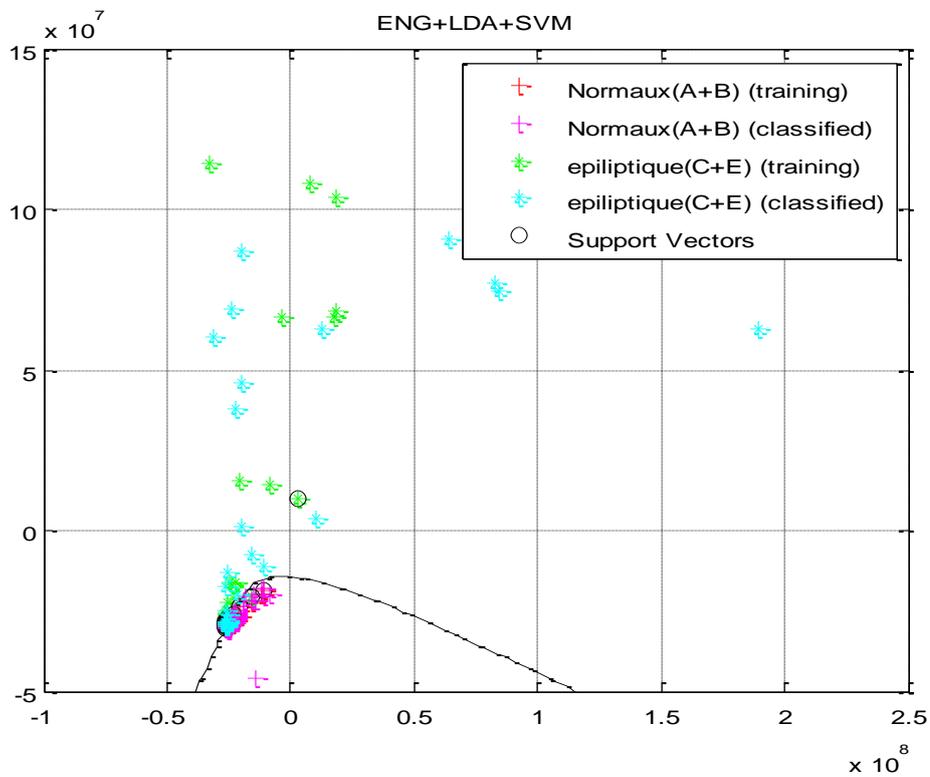
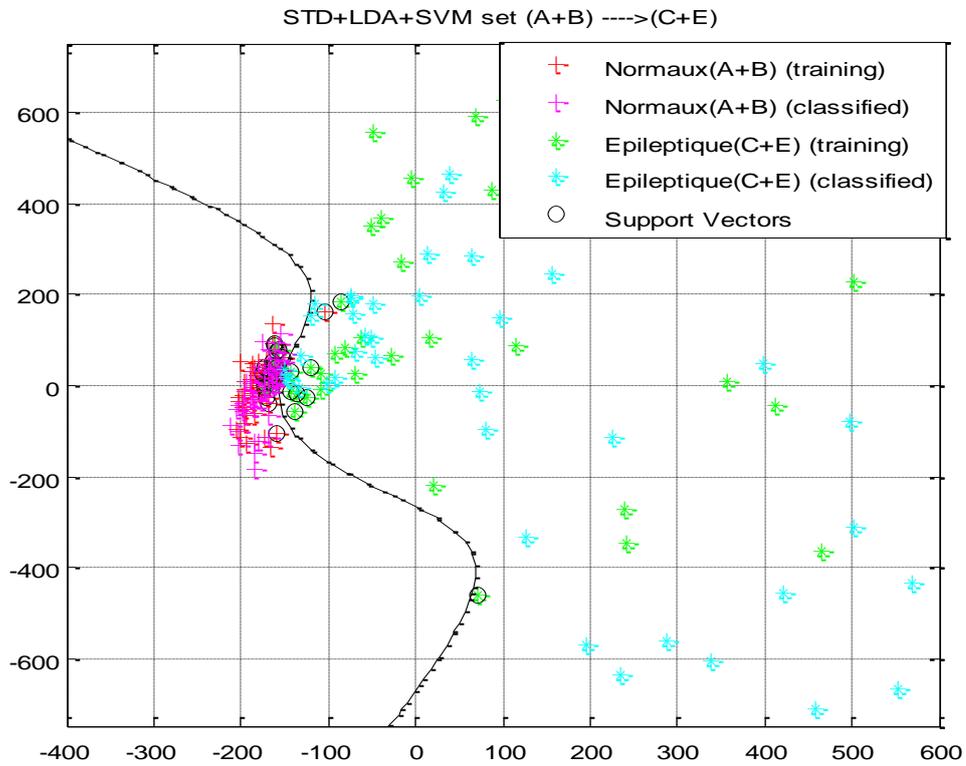
(a) : minimum, (b) :gamme, (c) : écart type

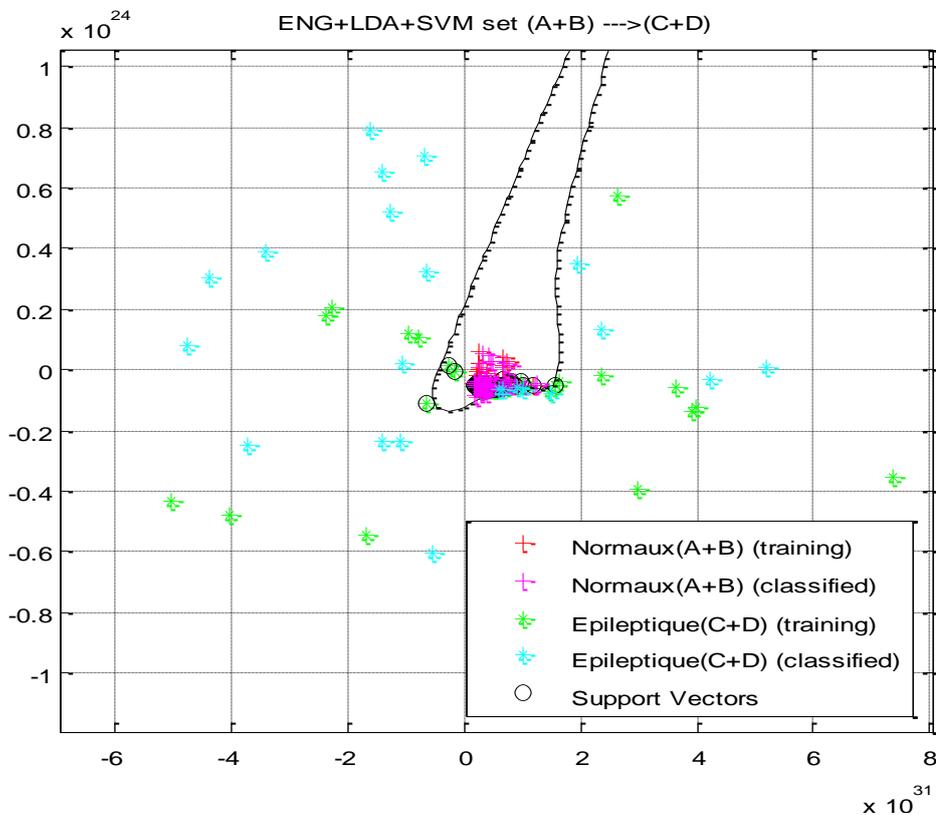
Après avoir choisi les paramètres de la fonction SVM en classifiant les groupes cités précédemment :

Groupe 4: (A+B)→(E+D)



Groupe 2 : (A+B) → (C+E)



Groupe 3 : (A+B) → (C+D)**5.8 Classification par les réseaux neurones :**

Le réseau de neurone utilisé est le feed forward à rétro-propagation (FFBPNN)

Le principal problème lié à son utilisation est le choix des fonctions d'activation des couches cachées et la fonction linéaire a été utilisée pour la couche de sortie en plus de la définition du nombre de neurones cachés.

Dans certaines applications de classification, on utilise :

- la fonction tangente hyperbolique pour la couche cachée
- la fonction linéaire pour la couche de sortie.

De bons résultats ont été obtenus avec ou sans utilisation de la durée.

Pour spécifier la structure du réseau, il est nécessaire de choisir

- le nombre de couches
- le nombre de neurones sur chaque couche.
- Le nombre d'entrées du réseau

- De même que le nombre de neurones sur la couche de sortie est fixé par les spécifications du problème que l'on veut résoudre avec ce réseau.

Pour notre application, on a pris les paramètres suivantes :

La fonction d'activation utilisée pour la couche cachée est la fonction tangente hyperbolique

$$f(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}}$$

La fonction d'activation utilisée pour la couche de sortie : la fonction linéaire

$$f(n) = n$$

Les réseaux perceptron multicouche ont été utilisés pour résoudre des problèmes difficiles par apprentissage supervisé à partir d'un algorithme de rétro propagation de l'erreur qui se base sur la règle d'apprentissage par correction d'erreur.

On a pris le nombre des neurones de la couche caché 10, 6 pour la couche d'entrée, et 2 pour la couche de sortie.

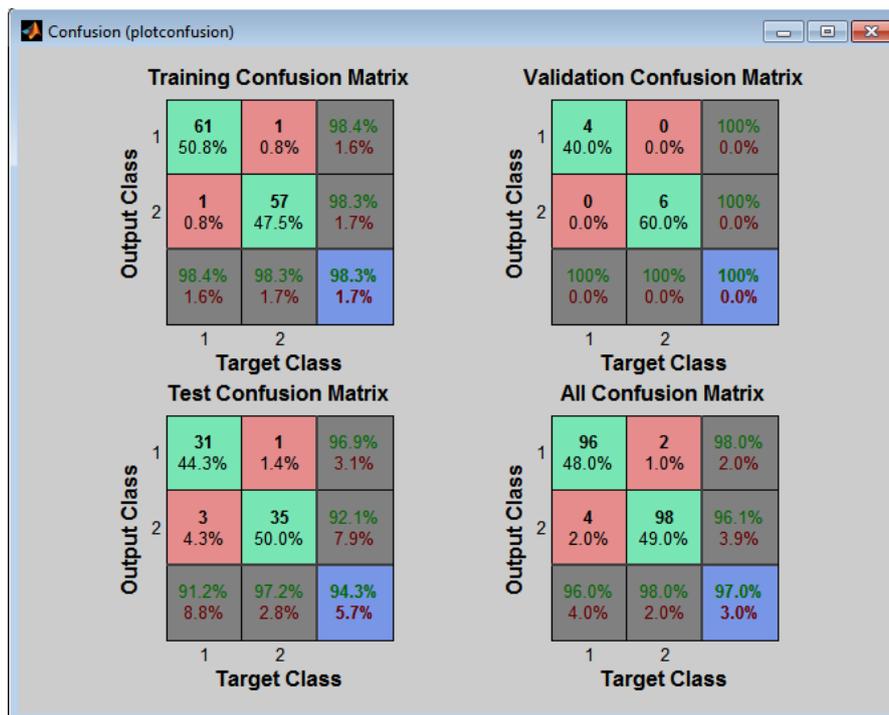
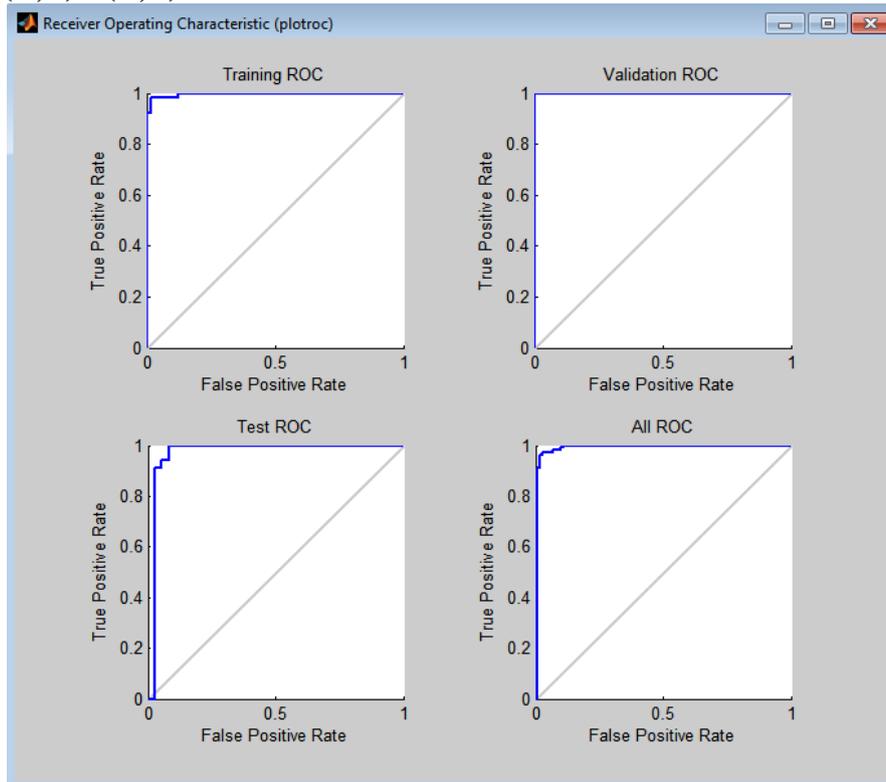
Les résultats de classification par FFBPNN sont mesurés par les matrices de confusion et les courbes rocs sont représentés ci-dessous et les résultats sont récapitulés dans le tableau.

Telle que la matrice de confusion, définie dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). Les données utilisées pour chacun de ces groupes doivent être différentes.

Un des intérêts de la matrice de confusion est qu'elle montre rapidement si le système parvient à classifier correctement.

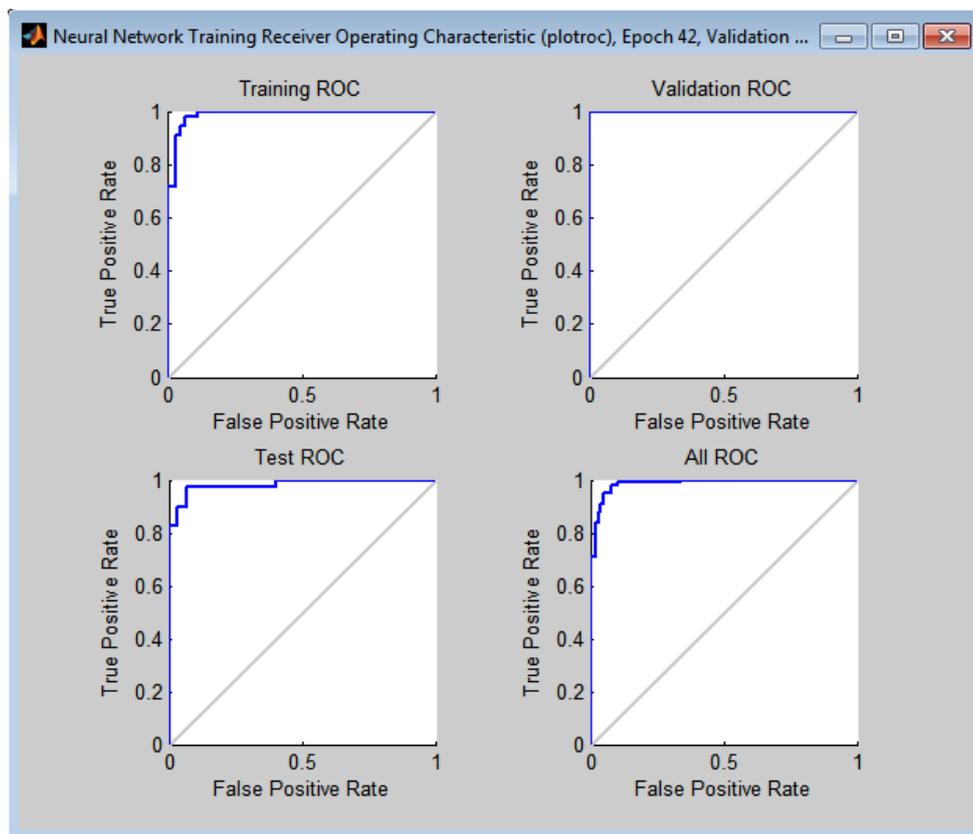
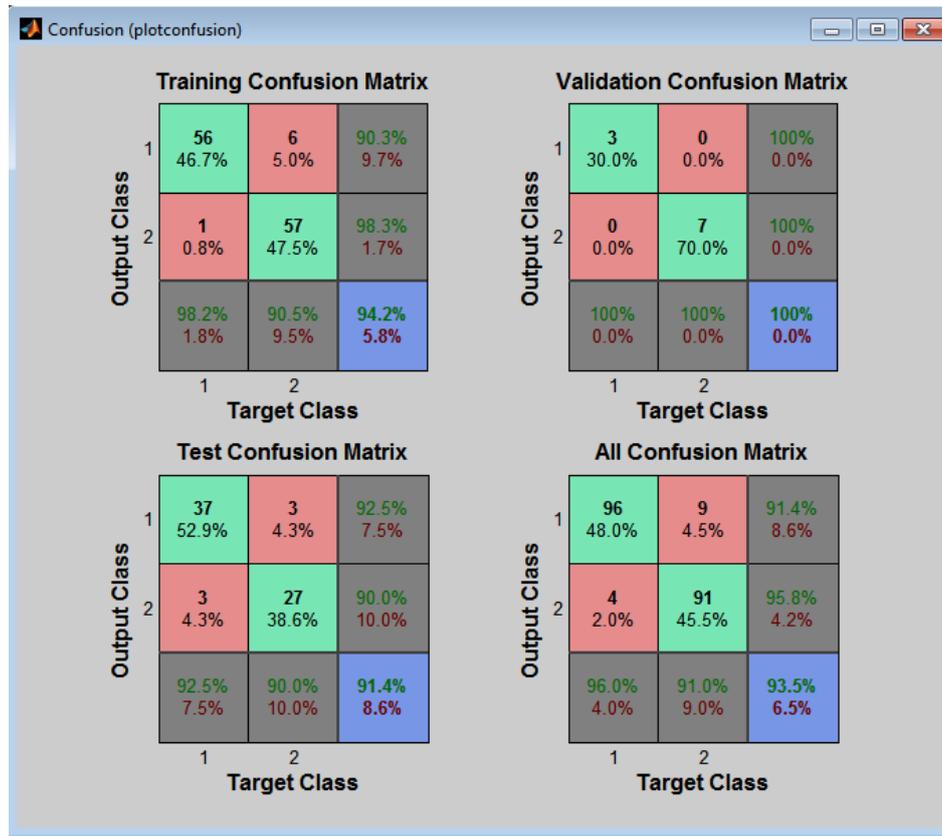
Les résultats de la classification de la méthode FFBPNN sont représentés dans les graphes qui suivent : Ces résultats sont obtenus on utilisant le paramètre STD, et la classification est représentée selon les groupes choisie auparavant.

Groupe2 : (A,B)->(C,E)



Pour le groupe 2 on a eu une classification de 97.00 % avec un taux d'erreur de 3.0%.

Group3 : (A,B)->(C,D)



Pour le groupe 2 on a eu une classification de 93.5% avec un taux d'erreur de 6.5%

Groupe4 : (A+B)→(D+E)



Pour le groupe 2 on a eu une classification de 96.5% avec un taux d'errer de 3.5%

En comparant les deux méthodes de classification, il est bien clair d'après les résultats obtenus que la classification par FFBPNN est mieux que celle faite par SVM.

5.9 Conclusion

Dans ce chapitre, on extrait différents paramètres des coefficients issus de la décomposition en DWT, et on a effectué une réduction de dimension par LDA qui a donné de bons résultats de réduction, ensuite on classifie par deux méthodes SVM et FFBPNN, et d'après les résultats trouvés il s'avère que le classificateur FFBPNN est meilleur que celui qui se base sur la méthode SVM.

Conclusion générale

L'évolution que connaissent l'acquisition, le prétraitement et la classification de signaux d'enregistrements EEG, signaux non-stationnaires, a rendu leur exploitation complexe ; ce qui redirige les recherches vers différentes techniques afin d'en extraire les informations utiles.

Plusieurs méthodes d'analyse, de prétraitement et de classification existent dans la littérature. Elles visent à extraire et à utiliser uniquement la structure intrinsèque de ces données qui contient juste l'information utile pour les traitements postérieurs, en vue d'une classification automatique pour une aide au diagnostic dans le domaine biomédical. Il s'agira, dans le cas présent, d'un prétraitement par transformation en ondelettes discrètes et en paquets d'ondelette. On propose des solutions à ces problèmes en déterminant les caractéristiques statistiques extraites à partir des coefficients de la transformée d'ondelette.

Les aspects théoriques et fondamentaux des ondelettes ont été abordés. Ces ondelettes constituent le principal outil mathématique qu'on utilise pour le prétraitement des signaux EEG par la DWT. Il était donc de première importance de bien comprendre ces fonctions et d'étudier leurs propriétés.

Une réduction de dimension a été nécessaire afin de représenter nos données sur un espace plus petit. Pour cela, un algorithme d'analyse discriminante linéaire de Fisher a été élaboré. Le choix de la transformée en ondelette est dicté par les caractéristiques du signal non-stationnaire d'enregistrement EEG.

Les machines à vecteur de support constituant une méthode de classification supervisée particulièrement bien adaptée pour le traitement de nos données ont été exploitées afin de classer les enregistrements des signaux EEG en cas normaux et en cas épileptiques.

Finalement, les performances de mesures ont été discutées en se basant sur l'exactitude, la sensibilité et la spécificité.

Il est à noter que les algorithmes qu'on a développé pour la classification de l'épilepsie sont exploitables par n'importe quelle application de diagnostic des signaux biomédicaux. Il suffit de changer la banque de données d'apprentissage.

Ce travail s'inscrit dans le cadre d'un domaine de recherche en pleine expansion : la neuroscience. Du point de vue pédagogique, ce projet m'a permis de découvrir le domaine du traitement et de la classification des signaux électroencéphalographies. Le développement des différents logiciels alloués à cette application, était une occasion pour maîtriser le logiciel MATLAB 2013a.

Les résultats obtenus avec l'algorithme suivi sont très prometteurs. On envisage en perspective des possibilités d'optimisation et d'implémentation matérielle sur une carte DSP et/ou FPGA.

Bibliographie :

- [1] KOLB, B. et WHISHAW, I. (2002). Cerveau et comportement. De Boeck Université.
- [2] WESTEN, D. et GARITTE, C. (2000). Psychologie : Pensée, cerveau et culture. De Boeck Université.
- [3] WILLIAMS, R.W. et HERRUP, K. (1988). The control of neuron number. Annual Review of Neuroscience, 11:423–453.
- [4] Montaine BERNARD,(2002), Méthodologie d'analyse des synchronisations neuronales dans les signaux EEG à l'aide de graphes d'informations temps-fréquence, UNIVERSITÉ DE POITIERS.
- [5] Matthieu Caparos, (2006), Analyse automatique des crises d'épilepsie du lobe temporal à partir des EEG de surface, Institut National Polytechnique de Lorraine.
- [6] ROMO V´AZQUEZ, 2010, Contribution à la détection et à l'analyse des signaux EEG épileptiques : débruitage et séparation de sources.
- [7] BERGER, H. (1929). Über das elektrenkephalogramm des menschen. European Archives of Psychiatry and Clinical Neuroscience, 87(1):527–570.
- [8] SANEI, S. et CHAMBERS, J. (2007). EEG signal processing. Wiley-Interscience.
- [9]: <http://www.ovarium.com/blog/2009/08/12/alpha-beta-theta-et-delta-voici-vos-ondes-cerebrales>
- [10] R. MILLER : Corticohippocampal interplay and the representation of contexts in the brain. Springer Verlag, 1991.
- [11] K. SASAKI, T. TSUJIMOTO, S. NISHIKAWA, N. NISHITANI et T. ISHIHARA: Frontal mental theta wave recorded simultaneously with magnetoencephalography and electroencephalograph. Neurosci Res, 26(1):79–81, Sep 1996.
- [12] VESPIGNANI, H. (1998). L'EEG de la technique à la clinique. Novartis.
- [13] A. Vespignani and S. Zapperi, Physical Review E 57, 6345 (1998).
- [14] JASPER, H. (1958). The ten-twenty electrode system of the International Federation for Electroencephalography. Electroencephalography and Clinical Neurophysiology, 10:371-375.
- [15] PERRIN, F., PERNIER, J., BERTRAND, O. et ECHALLIER, J. (1989). Spherical splines for scalp potential and current density mapping. Electroencephalography and Clinical Neurophysiology, 72(2):184–187.
- [16] GUÉRIT, J., MAUGUIÈRE, F. et PLOUIN, P. (2002). Guide pratique de neurophysiologie clinique. Recommandations de la Fédération internationale de neurophysiologie clinique, Traduction française.
- [17] D. O.WALTER : Spectral analysis for electroencephalograms : mathematical determination of neurophysiological relationships from records of limited duration. Exp. Neurol, 8:155–181, 1963.
- [18] D. Gabor: prix nobel de physique en 1971, inventeur de l'holographie (article original publié en 1948), « ANALYSE DE FOURIER LOCALE EN TRAITEMENT DES SIGNAUX »

- [19] Yves Meyer, Stéphane Jaffard et Olivier Rioul, 1987, Pour la science « l'analyse par Ondelettes ».
- [20] Robert X. Gao | Ruqiang Yan, « Wavelets Theory and Applications for Manufacturing», Springer New York Dordrecht Heidelberg London, 2011.
- [21] J.Salmon, Février 2009, Le Pennec, « Fourier et représentations temps/fréquences ».
- [22] Mohamed Ben MESSAOUD, Dr. Ing.ENSIS,200, « Traitement des Électrocardiogrammes en Vue de Diagnostic des Pathologies Cardiaques Suivi de Applications des Systèmes Adaptatifs avec Modèle de Référence aux Systèmes Électromécaniques », Université de Sfax
- [23] B. B. HUBBARD : Ondes et ondelettes : La saga d'un outil mathématique. Sciences D'avenir. Belin. Pour la Science, Paris, 1995.
- [24]A.Subasi, M.I.Gursoy, "EEG Signal classification using PCA, ICA, LDA and support vector machines",Expert Systems with Applications,vol:37,pp:8659-8666,2010.
- [25]T.Gandi, B.K.Panigrahi, S.Anand, "A comparative study of wavelet families for EEG signal classification",Neurocomputing, vol:74, pp:3051-3057,2011.
- [26]I.Omerhodzic, S.Avdakovic, A.Nuhanovic, K.Dizdarevic,"Energy Distribution of EEG Signals:EEG Signal Wavelet-Neural Network classifier".
- [27]M.K.Arjmandi,M.Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features,linear discriminant analysis and support vector machine",Biomedical Signal processing and control,vol:7,pp:3-19,2012.
- [28] Frédéric Truchetet, 1998, Paris, « Ondelettes pour le signal numérique »
- [29] Y. MEYER : Les ondelettes. In Contributions to nonlinear partial differential equations, Vol. II (Paris, 1985), volume 155 de Pitman Res. Notes Math. Ser., pages 158–171.Longman Sci. Tech., Harlow, 1987.
- [30] [Mal99]
- [31]Tarik ALANI, Introduction aux ondelettes (wavelets), 2001.
- [32] mathworks. wavelet toolbox,user's guide,discrete wavelet transform.
<http://www.mathworks.com/help/toolbox/wavelet/gs/f3-998127.html>, 2012.
- [33] D.J. McFarland, C. W. Anderson, K.-R. Muller, A. Schlogl, and D.J. Krusienski. Bci meeting 2005-workshop on bci signal processing: feature extraction and translation. IEEE transactions on Neural Systems and Rehabilitation Engineering, 14(2):135 – 138, 2006.
- [34] Wang. D; Miao. D; Xie. Cl. Best basis-based wavelet packet entropy feature extraction and hierarchical eeg classification for epileptic detection. *Expert Systems with Applications*, vol. 38, no. 11, pp. 14314-14320, 2011.
- [35] Wikipedia. <http://fr.wikipedia.org/wiki/Ecart-type>.
- [36] CHELLAPPA, R.WILSON C. L, and SIROHEYS: « Human and machine recognition of faces»: A survey, Proceedings of the IEEE, p. 705-740, 1995.
- [37] <http://aparate.elth.ucv.ro/POPA/DIVERSE/statist%20mat/Analiza%20discriminanta>
- [38] Sophie FUCHS, « Conception et Réalisation de Méthodes de Détection de Polluants Gazeux Atmosphériques à l'aide d'un Nez Electronique Portable », L'UNIVERSITE DE METZ, 2009.

[39] <http://www.jybaudot.fr/Factorielles/afdlin.html>

[40] BOTTOU, L. et CHIH-JEN, L. (2007). Support vector Machine Solvers, in Large Scale Kernel Machines. MIT Press.

[41] Subasi Alkan, Koklukaya. Automatic seizure detection in eeg using logistic regression and artificial neural network. Journal of Neuroscience Methods, vol. 148, no. 2, pp. 167-76, 2005.

[42] Ubeyli Guler. Adaptive neuro-fuzzy inference system for classification. Journal of Neuroscience Methods, vol. 148, no. 2, pp. 113-121, 2005.

[43] Bishop, C. M. (2006). Pattern Recognition and Machine Learning, volume 4

[44] Sahin Ogulata Erol Aslan, Bozdemir. A radial basis function neural network model for classification of epilepsy using eeg signals. The Journal of Medical Systems, vol.32, no. 5, 2008.

[45] Homan Dashei Wunsch Petrosian, Prokhorov. Recurrent neural network based prediction of epileptic seizures in intra and extracranial eeg. *Neurocomputing*, vol.30, no. 1-4, pp. 201-218, 2000.

[46] Mourad Adnane ; Zhongwei Jiang ; Zhonghong Yan. Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram. Expert Systems with Applications, vol. 39, pp. 1401-1413, 2012.

[47] Nilsson, N. J. (2004). Introduction to Machine Learning, volume 56. The MIT Press. de Information science and statistics. Springer.

[48] A.CHERIF, « Réseaux de neurones, SVM et approches locales pour la prévision des séries temporelles », UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS,2013.

[49] C.Cortes and V.Vapnik. Support vector networks.Machine Learning 20:273--297, 1995. V.Vapnik. The Nature of Statistical Learning Theory}, Springer.

[50] Ciarlet P., "Introduction à l'analyse numérique matricielle et à l'optimisation", Masson, nouvelle édition, 1994.

[51] Fletcher R., "Practical methods of optimization", John wiley, 2nd edition, 1987.

[52] CORNUÉJOLS, A., MICLET, L. et KODRATOFF, Y. (2002). Apprentissage artificiel. Eyrolles.

[53] N. Martaj, M. Mokhtari, MATLAB R2009, SIMULINK et STATEFLOW pour Ingénieurs, Chercheurs et Etudiants, DOI 10.1007/978-3-642-11764-0_17, Springer-Verlag Berlin Heidelberg 2010.

[54]. G. Dreyfus, J.-M. Martinez, M. Samuelides M. B. Gordon, F. Badran, S. Thiria «Apprentissage statistique» Eyrolles 2004.

[55]Y.Kumar,M.L.Dewal,R.S.Anand, “ Relative wavelet Energy and wavelet Entropy Based epileptic Brain signals classification”, the society of medical &biological engineering and springer,vol 2,pp147-157, 2012.

[56]Allahabad,”classificationof EEG Signals under Different Mental Tasks Using Wavelet Transform and neural Network with One step Secant Algorithm” International Journal of Scientific Engineering and technology,vol2,pp :256-259

[57] Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1986).Learning internal representations by error propagation. Parallel distributed processing : explorations in the microstructure of cognition, 1:318–362.

- [58]. Antoine Cornuéjols - Laurent Miclet Avec la participation d'Yves Kodratoff ;« Apprentissage artificiel Concepts et algorithms » ;Eyrolles 2003.
- [59] Antoine Cornuéjols - Laurent Miclet Avec la participation d'Yves Kodratoff ;« Apprentissage artificiel Concepts et algorithms » ;Eyrolles 2003.
- [60] A.Subasi, M.I.Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines", *Expert Systems with Applications*, vol. 37, pp.8659-8666, 2010.
- [61] L .Guo*, D.Rivero, A.Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks", *Journal of Neuroscience Methods*, vol.193 ,pp.156–163, 2010.
- [62] C.BOUVEYRON. « Modélisation et Classification des Données de Grande Dimension Application à l'Analyse d'images ». UNIVERSITÉ JOSEPH FOURIER – GRENOBLE 1. 2006.
- [63] Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*, 123, 69–87.
- [64] Charles Bouveyron, Camille Brunet, 'Classification automatique dans les sous-espaces discriminants de Fisher', SAMOS-MATISSE, CES, UMR CNRS 8174 – Université Paris 1.
- [65] <http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>