

Sommaire

<i>Remerciements</i>	iv
<i>Dédicaces</i>	v
<i>Dédicaces</i>	vi
<i>Liste des figures</i>	vii
<i>Liste des tableaux</i>	viii
<i>Liste des abréviations</i>	ix
<i>Introduction générale</i>	1
<i>Chapitre I :Apprentissage Artificiel</i>	4
<i>I.1. Introduction</i>	5
<i>I.2. Phase d'apprentissage</i>	5
<i>I.3.Types d'apprentissage</i>	5
<i>I.3.1. Apprentissage supervisé</i>	5
<i>I.3.2. Apprentissage non supervisé</i>	6
<i>I.3.3. Apprentissage semi supervisé</i>	7
<i>I.3.4. Apprentissage par renforcement :</i>	8
<i>I.3.5. Apprentissage « en ligne » et apprentissage « hors-ligne »</i>	9
<i>I.4. Phase de test</i>	9
<i>I.4.1. Taux de classification</i>	9
<i>I.4.2. La Sensibilité</i>	10
<i>I.4.3. La Spécificité</i>	10
<i>I.5. La validation croisée</i>	10
<i>I.6. Conclusion</i>	10
<i>Chapitre II : Etat de l'art</i>	11
<i>II.1. Introduction</i>	12
<i>II.2. Mono classification</i>	12
<i>II.2.1. Machine à Vecteurs de support (SVM)</i>	13
<i>II.2.2. Réseaux de Neurones (RNA)</i>	13

II.2.3. <i>K plus proches voisins (KNN)</i>	14
II.2.4. <i>Arbre de décision (ADD)</i>	14
II.3. <i>Multi classification</i>	15
II.4. <i>Méthodes de sélection</i>	16
II.5. <i>Conclusion</i>	16
Chapitre III : <i>Mono et Multi Classification</i>	17
III.1. <i>Introduction</i>	18
III.2. <i>Mono Classification</i>	18
III.2.1. <i>Les machines à vecteur de support (SVM)</i>	18
III.2.1.1. <i>Principes de fonctionnement général de la technique SVM</i>	18
III.2.1.2. <i>Séparation linéaire et non linéaire</i>	20
III.2.2. <i>Les réseaux de neurones (RNA)</i>	21
III.2.2. 1. <i>Neurone formel</i>	21
III.2.2. 2. <i>Perceptron multicouche</i>	23
III.2.2. 3. <i>L'algorithme de rétro-propagation du gradient</i>	24
III.2.3. <i>Les K plus proche voisins (KNN)</i>	25
III.2.3.1. <i>Algorithme des KNN</i>	26
III.2.4. <i>Arbre de décision (ADD)</i>	26
III.2.4.1. <i>Construction d'un arbre de décision</i>	27
III.3. <i>Multi classification</i>	28
III.3.1. <i>Introduction</i>	28
III.3.2. <i>Combinaison Homogène</i>	30
III.3.2.1. <i>Processus de sélection de variables</i>	30
III.3.3. <i>Combinaison Hétérogène</i>	32
III.3.3.1 <i>Vote majoritaire (VM)</i>	32
III.4. <i>Conclusion</i>	38
Chapitre IV : <i>Résultats et Discussion</i>	39
IV.1. <i>Introduction</i>	40
IV.2. <i>l'hypothyroïdie</i>	40
IV.2.1. <i>Définition de la thyroïde</i>	40
IV.2.2. <i>Ensemble hypothalamus/hypophyse/thyroïde</i>	41

IV.2.3. Définition de l'hypothyroïdie	41
IV.2.4. Comment détecter l'hypothyroïdie ?	42
IV.2.5. Les facteurs de risque : [75]	42
IV.2.6. Les types de l'hypothyroïdie	42
IV.3. Base de données :	43
IV.3.1. Description des paramètres de la base de données	43
IV.3.2. Prétraitement de la base	44
IV.4. Critère d'évaluation	45
IV.5. Expérimentation 1 : mono-Classification	45
IV.5.1. Les paramètres des différents classifieurs utilisés	46
IV.5.2. Sélection de variables	47
IV.5.2.1. Sélection Relief	47
IV.5.2.2. Sélection Rank	47
IV.6. Expérimentation 2 : Multi-Classification	50
IV.6.1. Vote majoritaire	50
IV.6.2. Théorie de Dempster-Shafer	50
IV.6.3. Combinaison homogène	50
IV.6.4. Combinaison hétérogène	52
IV.6.4.1. Combinaison hétérogène (BDD sans sélection de variables)	52
IV.6.4.2. combinaison hétérogène (BDD avec sélection de variables Relief)	54
IV.6.4.3. combinaison hétérogène (BDD avec sélection de variables Rank)	55
IV.8. Conclusion	56
Conclusion & Perspectives	57
Références	59

Remerciements

On remercie dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Mr M.A.Chikh**, on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

Nos remerciement s'adresse à **Mr A. Bechar** pour son aide pratique et son soutien moral et ses encouragements.

Nous sommes conscientes de l'honneur que nous a fait **Mr H. Bechar** en étant président du jury et **Mme Y.Benchaib** d'avoir accepté d'examiner ce travail.

On remercie aussi **Mr A. Lazouni** et **Melle S. Bouchikhi** pour leurs aides et leurs encouragements.

Nos remerciement s'adresse également à tous nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenue de près ou de loin principalement à **Mr N.Abbas** et **Mr B.Hadjadji** et **Mr M.Messabihi**

Dédicaces

A mon père Mohammed

Tu représentes pour moi l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir ; papa que j'adore

A ma très chère mère Fatma

Affable, honorable, aimable : Tu représentes pour moi la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur ; maman que j'adore.

Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.

A mon très cher mari Mohammed

Mon mari à qui je tiens à exprimer ma profonde gratitude et tout mon respect pour son soutien, sa disponibilité et sa patience. Merci à ma belle famille en témoignage de ma reconnaissance envers leur soutien et leurs encouragements

A mes chères sœurs et frères

Aux personnes dont j'ai bien aimé la présence dans ce jour, à tous mes sœurs Wissem, Meriem, Malika, Zoubida, Noriya, à mon frère Adlen. Je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements. Je vous exprime à travers ce travail mes sentiments de fraternité et d'amour

A mes chères amies

Aux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés, et qui m'ont accompagnaient durant mon chemin d'études supérieures, mes aimables amies, Merieme, Sihem, Sanaa.

A mon binôme Imane et toute la famille BASSAID.

Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

Asma.

Dédicaces

A ma très chère mère Atika

Affable, honorable, aimable : Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.

A mon père Ali

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation.

A mes très cher frères Sofiane, Imade

Mes chers frères présentent dans tous mes moments d'examens par son soutien moral et ses belles surprises sucrées. Je vous souhaite un avenir plein de joie, de bonheur, de réussite et de sérénité. Je vous exprime à travers ce travail mes sentiments de fraternité et d'amour.

A mon fiancé Chérif

Mon fiancé a qui je tiens à exprimer ma profonde gratitude et tout mon respect pour son soutien, sa disponibilité et sa patience. Merci à ma belle famille en témoignage de ma reconnaissance envers leur soutien et leurs encouragements

A mes très chère sœurs Sarah, Nabila, Amina, Rabiaa

En témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous. Malgré la distance, vous êtes toujours dans mon cœur. Je vous remercie pour votre affection si sincère. Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite.

A mon binôme Asma et toute la famille BELAIDI.

Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

Imane.

Liste des figures

Figure III.1: Différents hyperplans possibles ainsi que les vecteurs de supports	19
Figure III.2: Maximisation de la marge.....	19
Figure III.3: Choix de l'hyperplan optimal.....	20
Figure III.4 : Transformation d'espace.....	20
Figure III.5 : présentation d'un neurone formel.....	21
Figure III.6 : Fonction d'activation les plus utilisées.....	22
Figure III.7: Architecture d'un perceptron multicouche.....	24
Figure III.8 : Fonctionnement de KNN.....	26
Figure III.9 : Architecture d'un arbre de décision.....	27
Figure III.10 : Combinaison séquentielle des classifieurs.....	29
Figure III.11 : Combinaison parallèle des classifieurs.....	29
Figure III.12 : Combinaison hybride des classifieurs.....	30
Figure III.13 : principe de la sélection de variables.....	30
Figure IV.1: Schéma représentatif de la thyroïde.....	40
Figure IV.2 : Schéma représentatif de la procédure de fonctionnement du cerveau afin de contrôler la thyroïde.....	41
Figure IV.3 : Expérimentation 1 : Mono-classification.....	46
Figure IV.4 : Présentation du TC, SE, SP de l'expérimentation 1 : mono classification	49
Figure IV.5 : Expérimentation 2 : Combinaison homogène.....	50
Figure IV.6: Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Homogène.....	51
Figure IV.7 : Expérimentation 2 : Combinaison hétérogène (BDD sans sélection).....	52
Figure IV.8 : Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Hétérogène (BDD sans sélection).....	53
Figure IV.9 : Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Hétérogène(BDD Relief).....	54
Figure IV.10 : Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Hétérogène(BDD Rank).....	55

Liste des tableaux

Tableau IV.1 : Les TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec Différents attributs obtenu par l'algorithme Relief.....	47
Tableau IV.2 : Les TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec Différents attributs obtenu par l'algorithme Rank.....	48
Tableau IV.3 : Résultats de l'Expérimentation 1: Mono classification.....	48
Tableau IV.4 : Expérimentation 2 : Combinaison homogène.....	51
Tableau IV.5 : Expérimentation 2 : Combinaison hétérogène (BDD sans sélection)...	53
Tableau IV.6: Expérimentation 2 : Combinaison hétérogène (BDD avec Relief).....	54
Tableau IV.7: Expérimentation 2 : Combinaison hétérogène (BDD avec Rank).....	55

Liste des abréviations

ADD : Arbre de décision.

BDD : Base de données.

Bel : Croyance.

BetP : Probabilité pignistique.

CART : Classification And RegressionTree.

CC : Taux de classification correcte.

Cr : Crédibilité.(BeliefFunction).

D_{SmT} : Paradoxal de Dezert-Smarandache.

DST : Théorie de Dempster shafer.

Fn : Faux négatif.

Fp : Faux positif.

IA : Intelligence artificiel.

IC : Indice de confiance.

ID3 : Inductive DecisionTree.

K : Le conflit.

KNN :K-nearest neighbor (K plus prochevoisins).

LDA : Analyse Discriminante linéaire.

m: La masse.

Max : Maximum.

Min : Minimum.

Moy : Moyenne.

n:Le nombre de classes simples considérées.

Pls : Plausibilité.

PMC: Perceptron Multi-Couches.

Q : Communalité.

RBF : Réseaux de neurones à base de fonctions radiales.

RNA : Réseau de neurones.

$r_{\theta S}$: Le score du sous-ensemble.

$\bar{r}_{\theta i}$: La moyenne arithmétique des corrélations.

\bar{r}_{ij} : La moyenne des k^2 intercorrélations entre variables.

SAD : Système d'aide au diagnostic.

SE : Sensibilité.

SMA : Système multi agent.

SP : Spécificité.

SVM : Machine à vecteur de support.

S3VM : Machine à vecteur de support semi supervisé.

TBF : Le modèle de croyances transférables (TBF) de Smets.

TC : Taux de classification.

T3 : Triiodothyronine.

T4 : Tétraiodothyronine.

TSH : La thyroestimuline.

VM : Vote majoritaire.

vn : Vrai négatif.

vp : Vrai positif.

UCI : Machine Learning Respository.

Θ : Cadre de discernement.

θ_i : La classe.

\oplus : La somme orthogonale.

Introduction générale

L'ampleur et la complexité croissante des connaissances médicales, telles que les données physiopathologiques ou les moyens diagnostiques et thérapeutiques, obligent le médecin à gérer toujours plus d'informations pour soigner un patient. Le développement de systèmes d'aide à la décision, simulant le raisonnement médical, nécessite de modéliser cette pratique.

L'approche proposée dans ce travail consiste à effectuer une classification automatique de l'hypothyroïdie qui est une maladie génétique très fréquente et en croissance remarquable dans le monde, elle touche beaucoup plus les femmes et les personnes âgées de plus de 50 ans. Cette maladie peut entraîner des complications sur la santé de patient comme par exemple le manque d'énergie, la fatigue, perte de cheveux et gain de poids... Cette pathologie peut être diagnostiquée par des tests sanguins, mais elle ne se guérit pas, elle se contrôle très bien par la prise quotidienne de médicament.

Nous pouvons constater dans la littérature qu'il n'existe pas beaucoup de recherches qui sont concentrées sur la reconnaissance automatique de cette maladie, c'est pour cette raison que dans ce projet de fin d'études, nous nous sommes intéressé à la réalisation d'un système intelligent capable d'aider les médecins dans leurs routines cliniques et plus particulièrement au diagnostic de la maladie d'hypothyroïdie en utilisant quatre classifieurs supervisés : machine à vecteur de support (SVM), réseaux de neurones (RNA), arbre de décision (ADD), et finalement k plus proche voisin (KNN). Ce choix a été dicté par leur puissance dans le domaine de classification, qui a été approuvé et éprouvée par plusieurs chercheurs dans la littérature, ainsi que leurs facilités d'implémentation.

Pour réaliser ce système de classification, nous avons utilisé une base de données (BDD) de l'hypothyroïdie puis nous avons effectué un prétraitement sur cette base ensuite nous avons appliqué chaque classifieur indépendamment des autres sur cette dernière à fin d'effectuer une détection automatique de cette pathologie.

Puisque toutes les méthodes de classification ont leurs points forts et leurs points faibles et le domaine médical est très sensible aux erreurs, il nécessite beaucoup de précision et de certitude, nous avons proposé d'effectuer une combinaison homogène et hétérogène de classifieurs par la méthode Dempster-shafer (DST) et la méthode du vote majoritaire (VM) afin d'améliorer leurs performances et obtenir des réponses plus fiables et des décisions bien précises.

Le manuscrit de cette thèse est divisé en quatre chapitres :

Chapitre I : Apprentissage artificiel

Nous présenterons dans ce chapitre les fondements théoriques des différentes techniques d'apprentissage artificiel.

Chapitre II : Etat de l'art

Nous présenterons dans ce chapitre les différents travaux de recherche dans le domaine médicale qui sont réalisés par les différentes techniques de classification supervisée ainsi par les différentes techniques de combinaison de classifieurs et de sélection de variables.

Chapitre III : Mono et Multi Classification

Nous présenterons dans ce chapitre les fondements théoriques des différentes techniques de classification à savoir SVM, KNN, RNA, ADD ainsi le principe de fonctionnement de processus de sélection de variables et un panorama de base théorique des différentes méthodes de combinaison utilisée qui sont la théorie de DST et le VM.

Chapitre IV : Résultats et discussion

Nous présenterons dans ce chapitre le contexte médical de l'hypothyroïdie puis nous détaillerons la BDD utilisée dans ce travail ensuite nous discuterons les résultats obtenus par les différents classifieurs proposés, ainsi les résultats obtenus par les différents algorithmes de sélection de variables et les différentes méthodes de combinaison de classifieurs.

Nous terminerons notre thèse par une conclusion générale tout en proposant plusieurs pistes et perspectives de recherche dans le domaine.

*Chapitre I : Apprentissage
Artificiel*

1.1. Introduction

L'intelligence artificielle est sans doute l'un des domaines de recherches scientifiques qui a connu une forte avancée ces dernières années. Souvent abrégée par IA. Marvin Lee Minsky a défini l'IA comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains, car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuelle, l'organisation de la mémoire et le raisonnement critique ». L'IA utilise des ordinateurs (avec des programmes informatiques adéquats) ou de processus électroniques élaborés afin d'imiter le comportement humain. Donc, il s'agit de reproduire l'intelligence humaine mais avec l'utilisation des machines, d'où la qualification artificielle. Les domaines traités par l'IA en général couvrent les sciences cognitives, la représentation et l'acquisition des connaissances, la robotique, la classification, la vision par ordinateur, la reconnaissance de formes, la modélisation des raisonnements,.... [1]

La classification qui est l'un des axes de recherche de l'IA, consiste à trouver une application de l'ensemble des objets à classer, décrits par les variables descriptives choisies, dans l'ensemble des classes. L'algorithme ou la procédure qui réalise cette application est appelé classifieur et il doit passer par deux phases : une phase d'apprentissage et une phase de test.

1.2. Phase d'apprentissage

L'apprentissage, c'est un processus calculatoire qui doit être capable d'amener à une certaine prédiction et à une certaine généralisation.

Cornuéjols, L. Miclet ont défini l'apprentissage comme : " un programme possède des capacités d'apprentissage si ses potentialités de comportement sur les données se modifient en fonction de ses performances au fur et à mesure qu'il traite les données". [60]

1.3. Types d'apprentissage

Il existe différents types d'apprentissages :

1.3.1. Apprentissage supervisé

L'apprentissage supervisé « supervised learning » est vue comme : "un programme informatique où il peut apprendre à partir de l'expérience E relativement à une classe de tâches T et à une mesure de performance P, si sa performance à traiter les tâches de T, comme mesurée par P, augmente avec l'expérience E." [2].

Ce type est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des « exemples ». Où L'apprentissage automatique désigne l'ensemble des changements

dans un système qui lui permettent de réaliser une même tâche, ou des tâches similaires, de manière plus efficace ou plus efficiente au cours du temps [2].

Le principe d'apprentissage supervisé consiste à déterminer une nouvelle sortie Y à partir d'une nouvelle entrée X , connaissant un ensemble de données $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. Lorsque les Y_i prennent des valeurs discrètes, donc il s'agit d'un problème de classification. Par exemple en classification binaire, on cherche à attribuer à X une étiquette 0 ou 1, tandis que des Y_i ont des valeurs réelles, nous plaçons dans le cadre de la régression.

On trouve deux grandes familles d'apprentissage supervisé en fonction du langage des hypothèses manipulé : les approches symboliques et les approches numériques. Les approches symboliques construisent des hypothèses dans des langages directement compréhensibles par les experts du domaine traité, comme les arbres de décision ou les bases de règles. Les approches numériques utilisent des langages de représentation moins directement interprétables comme les réseaux de neurones ou les réseaux bayésiens.

L'apprentissage supervisé est utile soit pour prédire, soit pour expliquer. Il est utile pour prédire quand l'hypothèse apprise a pour but de servir à classer correctement de nouveaux exemples non encore classés. Il est utile pour expliquer quand on s'intéresse au contenu de l'hypothèse apprise pour comprendre ce qui relie les exemples à leur classe

Plusieurs méthodes de classification supervisée sont utilisées dans la littérature comme :

Réseau de neurones, Méthode des k plus proches voisins, Arbre de décision, Machine à vecteurs de support...Etc.

1.3.2. Apprentissage non supervisé

Dans le domaine informatique, l'apprentissage non supervisé « clustering » est une méthode d'apprentissage automatique. Il s'agit pour un classifieur de diviser un groupe hétérogène de données, en sous-groupes de manière que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données considérées comme différentes se retrouvent dans d'autres groupes distincts ; l'objectif étant de permettre une extraction de connaissance organisée à partir de ces données

Ce type d'apprentissage cherche des régularités parmi un ensemble d'exemples, sans être nécessairement guidé par l'utilisation qui sera faite des connaissances apprises. Par exemple, le clustering cherche à grouper des exemples de manière à ce que les exemples au sein d'un même groupe se ressemblent suffisamment, et que les exemples de groupes différents soient suffisamment différents [2].

L'idée principale de la classification non supervisée est de classer les données en k groupes satisfaisant les deux conditions suivantes :

- Chaque groupe contient au moins une donnée et chaque donnée appartient à un et un seul groupe.
- La seconde condition impose que deux groupes ne peuvent avoir de données en commun et que les k groupes contiennent toutes les données.

Plusieurs techniques qui s'occupent de ce genre d'apprentissage on site : le K_means, carte de kohonen, classification hiérarchique ascendante et descendante...

Remarque :

Inconvénient de la classification supervisée :

- la disponibilité d'un grand nombre de données acquises de manière automatique grâce aux nouvelles technologies .
- l'étiquetage manuel pour les fins d'algorithmes d'apprentissage de la formation → « la classification » est souvent un processus lent, coûteux, et source d'erreurs
- les données étiquetées sont couteuses et sont disponible en peu, car elle nécessite la présence d'un expert qui doit analyser toutes les données et les classifier. Cette opération devient impossible dans le cas d'une base de données volumineuse.

Inconvénient de la classification non supervisée :

- l'application de différents algorithmes non-supervisés sur les mêmes données donneront des regroupements différents.
- de plus, ces algorithmes sont souvent sensibles aux points aberrants. Il est donc important de réaliser une phase de prétraitement pour éliminer l'impact de ce type de points.
- Citons aussi le problème du choix du nombre de clusters. Lorsque l'on souhaite découvrir de nouvelles connaissances et les regrouper en clusters, un problème est de savoir combien de clusters sont nécessaires pour bien représenter ces connaissances.
- Ces dernières limites ont encouragé l'introduction du paradigme de l'apprentissage semi-supervisé comme outil fiable de classification

1.3.3. Apprentissage semi supervisé

C'est une méthode d'apprentissage où l'on dispose à la fois d'un (petit) ensemble de données étiquetées, et d'un (grand) ensemble de données non-étiquetées [61]

Nous présentons ci dessous quelque méthode utilisée dans la littérature :

a. L'auto apprentissage " self training " :

La méthode de classification par auto-apprentissage est un simple modèle de classification semi- supervisée, caractérisée par le fait que la procédure d'apprentissage utilise ses propres prévisions pour réapprendre elle-même. Une stratégie commune est de définir un seuil de confiance afin de décider quels exemples seront retenus pour être réappris. Elle est utilisée dans des tâches réelles

b. Le co-apprentissage (co-training):

C'est un algorithme d'apprentissage qui est une extension de la technique d'auto-apprentissage. Il est utilisé quand on dispose de petites quantités de données libellées avec une grande quantité des données non-libellées pour augmenter ses performances. Le co-apprentissage part du principe que s'il existe "deux vues" d'un même objet, indépendantes entre elles, alors deux classifieurs entraînés avec des vues différentes devront attribuer la même étiquette à l'objet

c. Modèle génératif :

C'est la plus ancienne des techniques pour l'apprentissage semi-supervisé. L'idée de base consiste à considérer un modèle comme un mélange de distributions, par exemple un mélange de gaussiennes. Ainsi, avec les données étiquetées et non étiquetées, on caractérise les différentes composantes du modèle de distribution supposé. Puis avec un seul élément étiqueté de chaque composante, on complète les caractéristiques du mélange.

d. Les méthodes à base de graphes :

S'appuient sur l'hypothèse que les exemples non-étiquetés doivent recevoir les mêmes étiquettes que leur(s) (proche(s)) voisin(s) étiqueté(s) [62]

e. Famille s3vm :

Parmi les techniques d'apprentissages semi-supervisé, machine à vecteurs de support semi supervisé (s3vm) constitue l'approche la plus populaire et qui a une solide base théorique héritant la notion de « large-margin » de SVM supervisées.

1.3.4. Apprentissage par renforcement :

Sutton et Barto en 1998 a défini l'apprentissage par renforcement comme: « L'apprentissage par renforcement désigne toute méthode adaptative permettant de résoudre un problème de décision séquentielle » [3].

Le terme adaptatif dans la définition de Sutton signifie que : « On part d'une solution inefficace, et qui sera améliorée progressivement en fonction de l'expérience du système au cours du temps. »

L'apprentissage par renforcement consiste à apprendre quoi faire, comment associer des actions à des situations, afin de maximiser quantitativement une récompense. On ne dit pas à l'apprenant quelle action faire, mais au lieu de cela, il doit découvrir quelles actions donnant le plus de récompenses en les essayant. Dans le cas le plus intéressant, des actions peuvent affecter non seulement les récompenses immédiates, mais aussi la situation suivante, et par là, les récompenses à plus long terme. Ces deux propriétés « recherche par essai-erreur et récompense à long terme » sont les deux caractéristiques les plus importantes de l'apprentissage par renforcement. Ce type d'apprentissage est différent de l'apprentissage supervisé. Car le classifieur n'a pas de superviseur à sa disposition, il interagit avec l'environnement qui lui donne un retour quantitatif sur les valeurs de ses actions.

1.3.5. Apprentissage « en ligne » et apprentissage « hors-ligne »

Il existe deux modes principaux d'apprentissage. Le premier, dit apprentissage « en-ligne », consiste à modifier les paramètres d'apprentissage immédiatement après la présentation d'un objet ou d'un exemple. C'est généralement le cas pour l'apprentissage par renforcement qui nécessite une réponse de l'environnement. L'avantage de ce mode, c'est leur rapidité, mais nécessite plus de données d'apprentissage.

Dans le second mode principal d'apprentissage : « hors-ligne » ou « off-line », le classifieur est entraîné à partir d'une base d'apprentissage (généralement, une base d'exemples). Ce type d'apprentissage consiste à n'effectuer l'adaptation des paramètres que lorsque l'ensemble des objets d'apprentissage ont été présentés au classifieur. Cette méthode permet de mieux apprendre, car c'est fait à partir d'un ensemble d'objets, plutôt qu'à partir d'un seul.

1.4. Phase de test

Cette phase permet l'affectation d'un nouvel objet à l'une des classes, au moyen d'une règle de décision intégrant les résultats de la phase d'apprentissage [63].

Plusieurs régimes expérimentaux ont été utilisés dans la littérature afin d'établir les performances d'un système de classification comme : les taux de classification (TC), les taux d'erreurs (TE), la sensibilité (SE) et la spécificité (SP).

1.4.1. Taux de classification

Le Taux de Classification (TC) est un paramètre significatif pour l'évaluation d'un classifieur :

$$TC = \frac{\text{Nombre de patients bien classés}}{\text{Nombre total des patients}} \quad (1)$$

Le taux de classification correcte est défini par :

$$CC = \frac{vp(i) + vn(i)}{vp(i) + vn(i) + fp(i) + fn(i)} \quad (2)$$

1.4.2. La Sensibilité

La SE est définie par la fréquence des tests positifs chez les malades.

$$SE(i) = \frac{vp(i)}{vp(i) + fn(i)} \quad (3)$$

1.4.3. La Spécificité

La SP est définie par la fréquence des tests négatifs chez les non-malades.

$$SP(i) = \frac{vn(i)}{vn(i) + fp(i)} \quad (4)$$

1.5. La validation croisée

C'est une méthode utilisée pour l'évaluation de la fiabilité des résultats obtenus par un classifieur donné. Cette méthode est basée sur une technique d'échantillonnage. Son principe de fonctionnement est résumé dans le paragraphe suivant [8]:

- les données disponibles sont divisées en K blocs disjoints, tout en essayant de prendre des blocks de la même taille.
- le système de classification est ensuite entraîné en utilisant les données de K-1 blocs. Le test du système est obtenu en utilisant les données du bloc restant.
- l'entraînement et le test sont répétés K fois (K expériences) pour que tous les blocs puissent servir de test. Le taux d'erreur final du système sera la somme des erreurs commises par toutes les K expériences.

1.6. Conclusion

Ces méthodes sont souvent combinées pour obtenir diverses variantes d'apprentissage. L'utilisation de tel ou tel algorithme dépend fortement de la tâche à résoudre (classification, estimation de valeurs, etc).

Dans le chapitre suivant nous présentons un état de l'art détaillé des travaux concernant l'aide au diagnostic médical.

Chapitre II : Etat de l'art

II.1. Introduction

Les travaux de recherche dans le domaine médical ont mobilisé durant ces dernières années une large communauté de chercheurs [4–5]. La recherche dans ce domaine s'appuie principalement sur la prise en compte des cas réels, c'est-à-dire des cas qui représentent la situation des patients.

Le domaine médical a su très rapidement intégrer des outils d'aide au diagnostic [8], et la santé a constitué un enjeu très motivant pour les recherches dans ce domaine. En effet, il ne s'agit plus de rechercher les causes d'une maladie chez un patient, mais plutôt les causes d'une défaillance d'un système vivant.

Un modèle de classification est un modèle capable d'identifier la nature d'un problème en examinant les symptômes observés. Les sorties d'un tel modèle représentent le diagnostic. [8]

Globalement, le but principal d'un modèle de classification est de détecter une ou plusieurs pathologies le plutôt possible, et être capable de donner un Indice de Confiance (IC) associé au diagnostic effectué, ou le cas échéant de donner des remèdes (solutions) appropriés aux pathologies, c'est-à-dire offrir une certaine interprétabilité. L'IC permet de déterminer la qualité des résultats donnée par le classifieur. En effet, plus l'IC est élevé, plus le diagnostic du classifieurs est pertinent. [8]

Dans ce travail, nous proposons un modèle de classification basé sur une combinaison de plusieurs classifieurs pour extraire la solution la plus optimale.

En effet plusieurs classifieurs peuvent donner une décision partiellement identique ; dans ce cas-là, une solution proposée peut être une combinaison de plusieurs réponses des classifieurs pour donner la meilleure décision. Pour cela, différentes techniques d'IA ont été utilisées et appliquées à des BDD médicale ces techniques de classification sont les RNA, SVM, KNN, ADD.

Dans ce chapitre, nous présentons un état de l'art complet des approches utilisées dans les systèmes de classification et plus spécifiquement celles qui concernent le domaine médicale, à savoir RNA, SVM, KNN, ADD et différentes techniques de combinaison telle que la technique de DST et du VM.

II.2. Mono classification

La majorité des travaux dans l'état de l'art utilisent en particulier des mono classifieurs

Un classifieur à base de réseau bayésien a été réalisé par l'équipe d'Aline Conseil et al. [21] Pour le diagnostic de l'hypothyroïdie. Leur objectif était de réaliser un système automatique pour diagnostiquer l'hypothyroïdie, ils ont essayé de faire plusieurs structures de réseau bayésiens pour déterminer le plus fiable avec un TC élevé et bien sûr en regardant d'autres critères, en ce qui concerne le TC était de 99%.

II.2.1. Machine à Vecteurs de support (SVM)

Les machines à vecteurs de support, ou Séparateur à Vaste Marge (SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVMs sont une généralisation des classifieurs linéaires. Ils ont été développés dans les années 1990 par Valdimir Vapnik grâce à leur capacité de travailler avec des données de grande dimension, leur faible nombre d'hypers paramètre à régler, le fait qu'ils soient bien fondés théoriquement et leur pouvoir de généralisation.

Parmi les travaux réalisés avec cette technique, nous avons :

En 2012, Miquel et Marc [10] ont testé les performances d'une implémentation personnalisée du SVM sur trois ensembles de données médicales (Wisconsin Breast Cancer, Pima Diabetes, Heart StatLog) à l'aide de l'outil matlab. Dans leur travail, ils ont montré la différence entre les cas de marges dures et de marges souples ainsi la différence entre les types de noyaux afin d'obtenir plusieurs gains pour mesurer les performances de système global.

Plus récemment un travail présenté par A.Lazouni [8] intitulé « Aide Au Diagnostic Pour Un Médecin anesthésiste réanimateur » ce travail est basé sur le test de cinq différentes techniques de classification sur une nouvelle base de données collectée localement (1200 patients). Ces techniques sont les RNA, RBF, SVM, KNN, ADD ensuite un système de vote majoritaire a été utilisé dans le but d'obtenir la meilleure précision possible. Les résultats obtenus sont satisfaisants et prometteurs, ils montrent la cohérence et la robustesse de son système proposé.

Dans la phase de classification, ils ont réalisé une combinaison de plusieurs classifieurs SVM afin de comparer les résultats et d'améliorer les performances du système global.

II.2.2. Réseaux de Neurones (RNA)

Pour les Réseaux de Neurones (RNA), le diagnostic de défaillances est essentiellement vu comme un problème de reconnaissance et de classification [33]. En effet, l'une des tâches principales à effectuer en diagnostic médical est la reconnaissance et la classification des données, et les RNA sont des outils bien adaptés pour cette tâche.[8].

Dans le même contexte M.Olivier [13] a réalisé un système dont le but est de détecter les cellules anormales ou suspectes qui auraient pu échapper à la vigilance du pathologiste. Ce système traite des images couleur de la cytologie des séreuses sur lesquelles il doit reconnaître les cellules anormales avec un taux maximum.

Aussi, un autre travail conçu par R. Benali, M. A. Chikh [14] dans le but de reconnaissance des extrasystoles ventriculaires par la combinaison des RNA et la logique floue. Au cours de ce travail, ils ont conçu et implémenté un classifieur neuro-flou (CLNF) pour la reconnaissance des extrasystoles ventriculaires. Les performances de la

classification ont été évaluées par le calcul de la SE, SP et TC. Avec l'utilisation d'une BDD médicale réelle (MIT-BIH) pour valider les résultats.

II.2.3. K plus proches voisins (KNN)

La technique des K-plus proches voisins (k NearestNeighbor) (KNN) est connue comme une technique de classification supervisée simple mais robuste et permet de donner des bons résultats en classification des données médicales.

La méthode des K-plus proches voisins (k NearestNeighbor) a été beaucoup utilisée dans le domaine d'aide au diagnostic médical dans différentes spécialités à savoir la détection du diabète, du cancer du sein, de rétinopathie diabétique en ophtalmologie, et du traitement du signal comme le EEG par exemple et d'autres ...

En 2010, E.Mathieu-Dupas [15] a présenté le fondement théorique de la méthode k plus proche voisin ainsi, il a appliqué cette technique pour résoudre des problèmes de diagnostique d'une pathologie complexe.

F.Chamroukhi[16] dans son projet il s'est basé sur la classification supervisée par Les K-plus proches voisins, il a développé une bibliothèque implémentant un algorithme de classification supervisée, dite aussi discrimination de données brutes en utilisant la méthode de KNN pour prédire les classes de nouvelles données à partir de données étiquetées (données d'apprentissage).

II.2.4. Arbre de décision (ADD)

L'arbre de décision (ADD) est un outil utilisé dans différents domaines : sécurité, fouille de données, médecine, etc. Sa popularité est due à sa lisibilité, sa rapidité d'exécution, le peu d'hypothèses nécessaires qu'il propose, et son interprétabilité que souhaiterait avoir tous les médecins a priori, expliquent sa popularité actuelle. L'apprentissage par ADD se situe dans le cadre de l'apprentissage supervisé, où la classe de chaque objet dans la base est donnée.

L'approche ADD a été largement étudiée et appliquée dans le domaine de classification supervisée. En trouvant dans les travaux de R. Quinlan le principe et le développement de cette méthode [17–18].

Parmi les projets réalisés par cette méthode, on trouve :

N.settoui et K.Douibi [19] qui ont effectué un travail pour la détection automatique de l'hypothyroïdie par l'intermédiaire de classifieur ADD « CART », elles ont obtenu un très bon TC et le nombre de cas où le classifieur s'est trompé est minime ainsi les résultats donnés sont interprétables et les règles de l'arbre construit correspondent exactement aux mêmes paramètres qu'utilise le médecin pour déclarer si la personne est hypothyroïdienne ou pas.

L'objectif de travail qui est intitulé « Classification des tumeurs du sein par un classifieur Bayésien Naïf et les arbres de décision » [20] est de construire un classifieur Bayésien Naïf et ADD qui aide à la classification des tumeurs du sein en utilisant des symptômes cliniques et des caractéristiques extraites de l'image mammographie. La classification des tumeurs a été faite par l'algorithme de l'arbre de décision ID3 (Inductive DecisionTree). Après Djebbar& al ont validé les deux classifieurs sur une base réelle des images (MIAS) et les résultats obtenus sont satisfaisants.

II.3. Multi classification

Dans la majorité des cas, les techniques classiques de classification telle que les RNA,SVM, KNN, ADD donnent des réponses partiellement similaire à celle données par l'expert, mais l'utilisation de plusieurs classifieurs au même temps puis fusionné les réponses de chaque classifieurs par les différentes techniques de combinaisons comme la théorie de Dempster-Shafer et le vote majoritaire peut améliorer les performances de la classification.

Parmi les travaux réalisés par ces techniques de fusion,nous trouvons :

Les travaux proposés par N.Abbas[23] qui portent essentiellement sur le développement et la mise en œuvre de divers schémas de combinaison des classifieurs SVM en utilisant la DST et paradoxal de Dezert-Smarandache(DSmT). Les deux théories DST et DSmT permettent de traiter les conflits entre les réponses des classifieurs en tentant de sélectionner les meilleures réponses. L'objectif du travail dans cette thèse était de développer deux applications principales pour évaluer l'utilisation efficace de ces deux théories qui sont respectivement la vérification de signature manuscrite et la reconnaissance des chiffres manuscrits. L'auteur dans son travail avait proposé quatre solutions pour améliorer la reconnaissance de l'écriture manuscrite.

Une autre réalisation par Barhoumi et Zagrouba [24] qui ont mis le point sur la théorie de l'évidence comme étant une technique performante de fusion de l'information multi-sources afin d'accéder à une information plus fiable. Les auteurs ont réalisé une application médicale de reconnaissance de mélanome,c'est-à-direun système qui décide le type de la lésion maligne ou bénigne, à partir d'une image dermatologique couleur. Chaque élément d'information de ce système est décrit en termes de ses sources d'imperfection ainsi des techniques permettant de réduire les effets de ces imperfections.

Aussi belghabri& al [26] ont évalué la performance d'un ensemble de trois classifieurs, chaque classifieur est testé individuellement sur un ensemble de données d'apprentissage afin de réduire les erreurs de reconnaissance ou de classification, une stratégie de combinaison puissante basée sur la DST est utilisée pour fusionner les trois classifieurs.

Alors que les résultats de la reconnaissance des classifieurs individuels sont comparés à ceux obtenus après la fusion des sorties des classifieurs par la DST, leproduit,

le maximum, le minimum, la moyenne, la méthode de Bayes et le VM les résultats montrent des améliorations significatives des performances.

II.4. Méthodes de sélection

Les techniques de sélection de dimension consistent à rechercher des directives informatives et d'éliminer les directives qui ne contiennent que du bruit. Parmi ces méthodes, on trouve les deux techniques : ReliefF et Rank : [30]

H.Senoussi a effectué dans son travail [29] une sélection de données pour l'apprentissage des RNA, ADD et les KNN par plusieurs techniques de sélection telle que : relief, CFS, mRMR, FCBF, INERACT et elle a appliqué ces méthodes pour résoudre les problèmes du diagnostic industriel et les résultats obtenus sont très prometteurs.

Un travail présenté par N.Settoui, A.Hafa [30] dont l'objectif était de réduire le nombre de variables parmi lesquelles peu sont informatives et les autres constituent essentiellement du bruit. Ainsi, elles ont montré l'aspect théorique et algorithmique des différentes techniques de sélections parmi les : ReliefF, MI, mRMR et fisher. Afin de valider l'efficacité et les résultats des différentes techniques de sélection, elles ont évalué les quatre méthodes de sélection avec plusieurs classifieurs de natures différentes tel que RNA, RBF, KNN, analyse Discriminante linéaire (LDA), ADD et les résultats finales sont satisfaisant pour la technique de sélection ReliefF.

H.Stoppiglia a présenté deux techniques de sélection qui sont « Rand » et « Rank » elle a montré la différence entre ces deux dernières et leurs manières d'utilisation ainsi, elle a appliqué et testé leurs performances. Ces deux méthodes ont été utilisées avec succès dans certains nombres d'applications. [31]

Un travail présenté par Novakovic & al qui ont réalisé une comparaison entre plusieurs méthodes de sélection sur deux jeux de données réels [32]. Ils ont testé six méthodes de sélection parmi les : algorithmes « rank » sur quatre algorithmes d'apprentissage supervisé à savoir, IB1, Naïve Bayes, C4.5 arbre de décision et le réseau de RBF. Les résultats montrent que la sélection pourrait être importante pour la précision de la classification.

II.5. Conclusion

Dans ce chapitre, nous avons présenté un état de l'art des différentes approches de classification utilisées dans notre système à savoir les RNA, SVM, KNN, ADD ainsi, on a montré les différentes techniques de combinaisons telles que la DST et le VM.

Dans le chapitre suivant nous allons présenter d'abord le principe de fonctionnement de chaque classifieur puis le principe de fonctionnement des méthodes de combinaisons.

***Chapitre III : Mono et Multi
Classification***

III.1. Introduction

Dans le domaine médical, les classifieurs sont omniprésents, leur objectif est de reconnaître différents types de pathologies (classes) d'un patient donné. L'introduction des techniques numériques dans ce domaine a permis d'utiliser des méthodes de classification pour détecter les différentes pathologies dont chacune est basée sur un principe donné, et cela, d'aider les médecins dans leurs routines médicales. C'est pour cela que nous avons proposé dans ce travail de développer et d'implémenter un système capable de fournir une aide pour la reconnaissance de l'hypothyroïdie et nous présenterons essentiellement dans ce chapitre les différentes techniques de mono et multi classification utilisées.

III.2. Mono Classification

Dans cette section, nous allons vous présenter le principe de fonctionnement des différents classifieurs qui sont en mode supervisé, à savoir les SVM, RNA, KNN, et ADD, tout en soulignant certains de leurs avantages et de leurs inconvénients.

III.2.1. Les machines à vecteur de support (SVM)

Les SVMs qui sont aussi appelés : séparateurs à vaste marge (support Vector Machine) présentent une technique de classification dont l'apprentissage est supervisé. Cette technique a été introduite par le russe Vladimir Vapnik en 1995 [34].

Les SVMs reposent sur une théorie mathématique solide. Elles ont été à l'origine utilisées pour la classification binaire et la régression. Aujourd'hui, elles sont utilisées par de nombreux chercheurs et dans différents domaines tels que l'aide au diagnostic médical qui fait l'objet de notre étude, le marketing, la biologie, etc. La raison du succès de cette technique est son adaptation rapide aux BDD à grandes dimensions (big data), son nombre réduit de paramètre à régler, et ses bons résultats en pratique. [8]

La technique SVM, permet de résoudre des problèmes linéairement et non linéairement séparable. Elle repose sur deux notions clés : celle de la maximisation de la marge et celle de la fonction noyau. Dans la section suivante, les principes de fonctionnement général des deux cas (linéairement et non linéairement séparable) vont être présentés.

III.2.1.1. Principes de fonctionnement général de la technique SVM

Le but de l'approche SVM est de trouver une bonne reconnaissance ainsi qu'une bonne séparation entre différentes classes. Le classifieur linéaire qui effectue une séparation de données entre les classes est appelé hyperplan.

Sur la figure III.1, on détermine un hyperplan séparant deux classes (cercle bleu, et carré rouge).

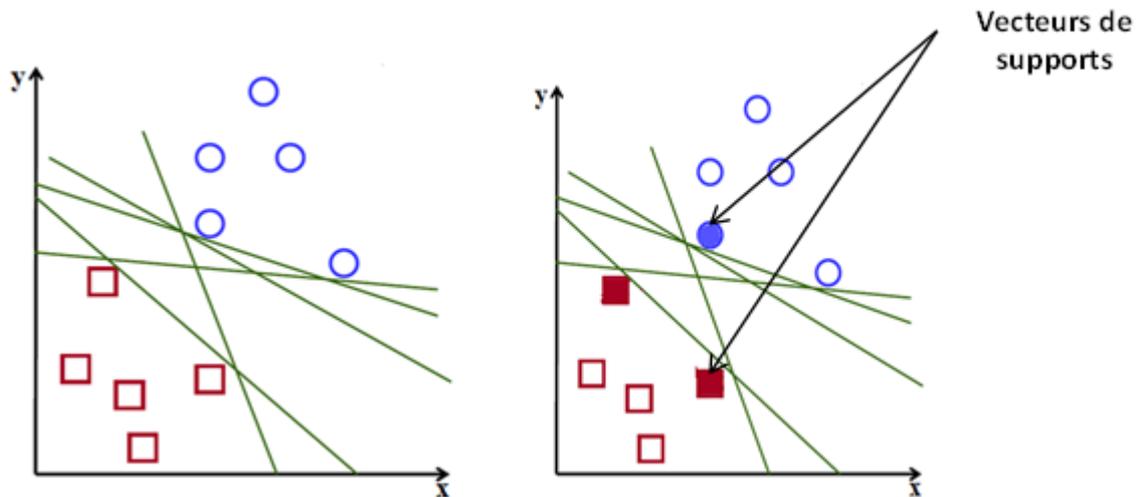


Figure III.1: Différents hyperplans possibles ainsi que les vecteurs de supports

Les échantillons les plus proches des deux classes sont appelés les vecteurs de support. Ces derniers sont utilisés pour la détermination de l'hyperplan (partie droite de la figure). Comme la montre la figure ci-dessus, il existe une multitude d'hyperplans valides, c'est-à-dire qui assurent une séparation entre les deux classes.

La principale particularité des SVMs est que l'hyperplan choisi doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe au milieu des deux classes. Intuitivement, cela revient à chercher l'hyperplan le plus sûr, c'est-à-dire celui qui passe au milieu des deux classes et qui offre le maximum de précision pour une bonne classification. Formellement, cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale. On appelle cette distance « marge » entre l'hyperplan et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge d'où le nom des séparateurs à vaste marge (Voir figure III.2) [8]

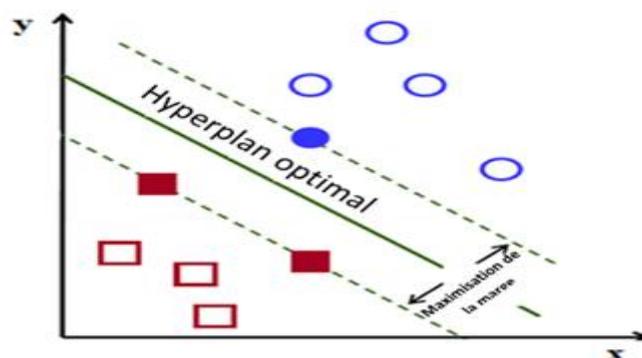


Figure III.2: Maximisation de la marge

La maximisation de la marge est réalisée dans le but d'obtenir une précision maximale lorsqu'on classe un nouvel exemple (période de test). Dans la figure III.3 la partie gauche nous montre qu'avec un hyperplan optimal, un nouvel exemple (cercle en jaune)

reste bien classé, par contre dans la partie droite où il y a une marge plus petite, le nouvelexemple se voit mal classé.

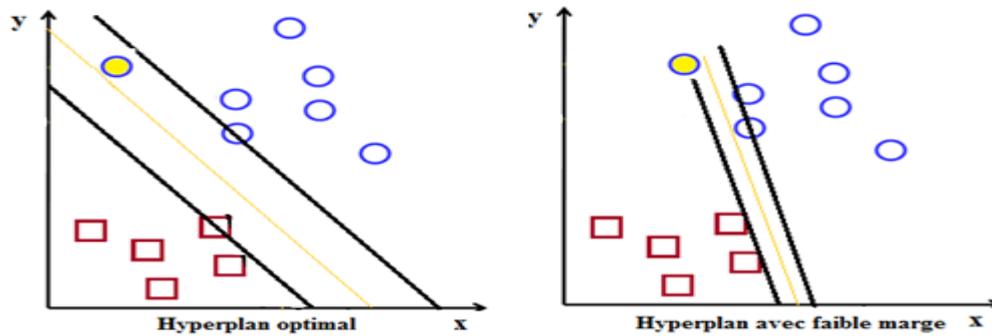


Figure III.3: Choix de l'hyperplan optimal

III.2.1.2. Séparation linéaire et non linéaire

Il existe deux cas de séparation, les cas linéairement séparables et les cas non linéairement séparables. Le premier cas est le plus simple à réaliser mais dans la majorité des cas pratiques, très peu de problèmes peuvent être séparés d'une manière linéaire. Pour surmonter les inconvénients du deuxième cas, l'idée des SVMs est de changer l'espace des données (figure III.4). La transformation non-linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc procéder à un changement de dimension. Cette nouvelle dimension est appelée «espace de redescription». En effet, plus la dimension de l'espace de redescription est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. [8]

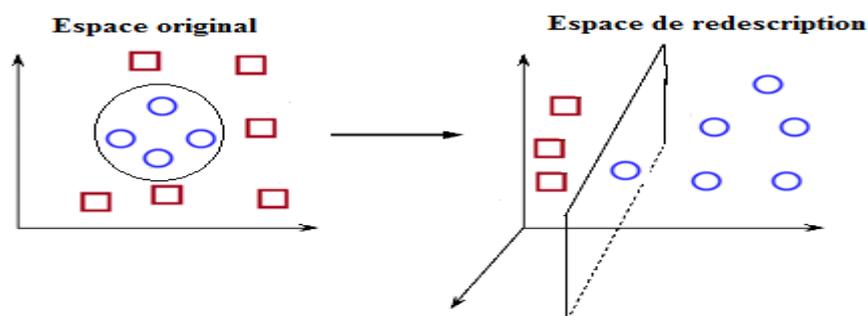


Figure III.4 : Transformation d'espace

Nous avons donc transformé un problème non linéairement séparable dans l'espace de représentation à deux dimensions à un cas linéairement séparable dans un espace de plus grande dimension qui est l'espace de redescription (Featurespace). Cette transformation est réalisée via une fonction noyau (fonction Kernel). En pratique, quelques familles de fonctions noyaux paramétrables sont connues et il revient à l'utilisateur d'effectuer des tests pour déterminer celle qui convient le mieux à son application. Nous pouvons citer les exemples de noyaux suivants : Linéaire, Polynomiale, Gaussien, et Laplacien.

-Le noyau linéaire : $K(x_i, x_j) = x_i^T \cdot x_j$ (5)

-Le noyau polynomial : $K(x_i, x_j) = (x_i^T \cdot x_j + 1)^d$ (6)

-Le noyau gaussien : $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$ (7)

-Le noyau Laplacien : $k(x, y) = \exp\left(-\frac{\|x-y\|}{\sigma}\right)$ (8)

Dans cette partie, nous avons présenté de manière simple le concept et le fonctionnement général de la technique SVM.

III.2.2. Les réseaux de neurones (RNA)

Les RNA sont des modèles de calcul utilisés en informatique dont le but est de résoudre des problèmes de classification. Ces derniers sont schématiquement inspirés du fonctionnement des neurones biologiques.

L'étude de RNA a apparu au début des années 1940 par les travaux des deux neurologues Warren Mc Culloch et Walter Pitts en 1943 et a été étendue aux problèmes de classification et reconnaissance des formes par Rosenblatt en 1962. [8]

Dans la section qui suit, quelques définitions relatives à la théorie des réseaux de neurones vont être présentées.

III.2.2. 1. Neurone formel

Un neurone artificiel ou formel est une fonction mathématique non-linéaire.. Son fonctionnement s'inspire directement du mécanisme du neurone biologique, en cherchant à imiter son processus d'apprentissage. Il est généralement constitué de plusieurs entrées et d'une seule sortie. Chacune de ces entrées « x_1, \dots, x_n », est associée à un poids appelé poids synaptique ou coefficients numérique « W_1, \dots, W_n ». Ces derniers sont ajustés lors de la phase d'apprentissage.

Comme la montre la figureIII.5, le neurone formel calcule la somme pondérée des entrées qu'il perçoit auquel se rajoute un seuil qui est fixé au début par l'utilisateur. Ensuite, il applique à la valeur obtenue une fonction non-linéaire qu'on appelle fonction d'activation. Enfin, la valeur finale est affectée en sortie du neurone.

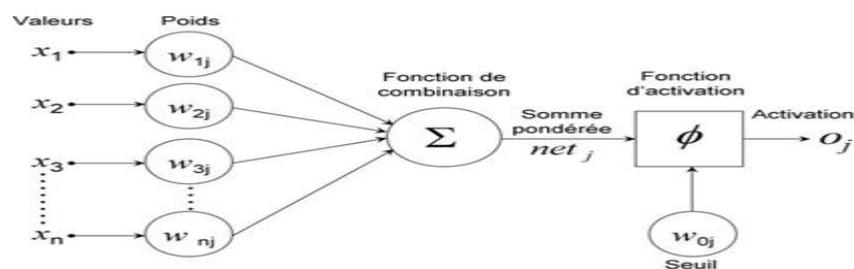


Figure III.5 : présentation d'un neurone formel

Le choix de la fonction d'activation dépend de l'utilisateur ainsi que du problème étudié. Ce choix est très important car il influe sur les propriétés du neurone formel. Il existe cependant plusieurs fonctions d'activation (Figure III.6) tel que : [29]

- a) Tout ou rien.
- b) Fonction signe.
- c) Plus ou moins à seuil.
- d) Fonction affine.
- e) Saturation.
- f) Sigmoide.
- g) Fonction arctangente.
- h) Fonction radiale de base de type gaussien.

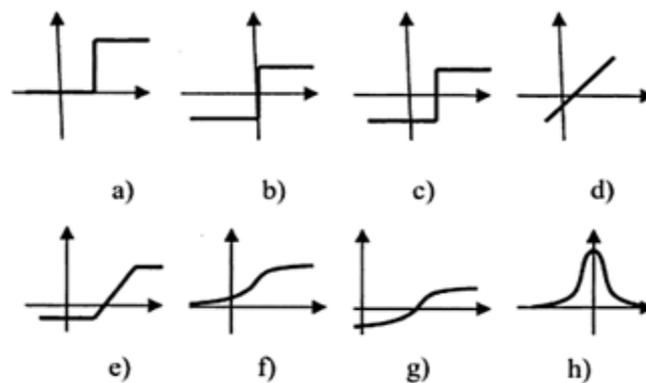


Figure III.6 : Fonction d'activation les plus utilisées

Dans notre travail, nous avons utilisé la fonction d'activation sigmoïde.

Considérons le cas d'un neurone formel à n entrées auxquelles sont affectées n grandeurs numériques « x_1, \dots, x_n », pour définir la règle de calcul. Un poids synaptique noté « w_1, \dots, w_n » est associé à chaque entrée, et la sommation pondérée est obtenue de la façon suivante :

$$w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n = \sum_{i=1}^n w_i \cdot x_i \quad (9)$$

Le neurone commence à devenir actif lorsque la sommation pondérée dépasse un seuil noté w_0 . Il est additionné à cette grandeur, pour ensuite être transformé par la fonction d'activation notée ϕ :

$$\phi \left(w_0 + \sum_{i=1}^n w_i \cdot x_i \right) \quad (10)$$

En général, l'expression de la valeur de sortie est simplifiée en ajoutant une entrée fictive x_0 fixée à la valeur 1, ce qui donne une fonction précédente comme suit :

$$\phi \left(\sum_{i=0}^n w_i \cdot x_i \right) \quad (11)$$

Un RNA se compose de plusieurs couches (couches d'entrée, couches cachées et couches de sortie) qui contiennent des neurones connectés de façon à ce que la sortie d'un neurone puisse être l'entrée d'un ou plusieurs autres neurones

Les couches d'entrée et de sortie constituent l'interface avec l'extérieur. La couche d'entrée reçoit les variables en entrée du modèle, et la couche de sortie renvoie les résultats attendus. Les neurones internes du réseau constituent les couches cachées.

Chacune des couches effectue un traitement indépendamment des autres afin de transférer le résultat de son analyse à la couche suivante. L'information, donc, se propage de couche en couche, de l'entrée vers la sortie toute en passant par la couche cachée. Cette dernière est variable selon la nature du modèle. [8]

En fonction du sens et de l'orientation des différentes connexions définissant la structure du réseau, la valeur obtenue en sortie du neurone par la fonction d'activation est alors transmise comme entrée aux neurones suivants qui lui sont reliés. En fonction de l'algorithme d'apprentissage, l'information peut également être propagée en arrière (back propagation).

III.2.2. 2. Perceptron multicouche

Le Perceptron multicouche PMC est un modèle du réseau de neurones formel. Il a été inventé en 1957 par Frank Rosenblatt.

C'est un réseau de type feedforward, c'est-à-dire que l'information peut circuler dans un seul sens de la couche d'entrée vers la couche de sortie toute en passant par la couche cachée (le mode back propagation n'est pas possible).

Ce modèle intègre une ou plusieurs couches cachées entre l'entrée et la sortie. À l'exception des couches d'entrée et de sortie, chaque neurone est connecté avec tous ceux de la couche suivante et tous ceux de la précédente. Par contre, il n'existe pas de liaison entre des neurones issus de la même couche (Figure III.7).

L'avantage de ce modèle par rapport à celui de neurone formel (Perceptron monocouche), est qu'il peut résoudre les problèmes non linéairement séparables ainsi que d'autres problèmes plus complexes comme ceux du domaine biomédical. C'est pour cela

que le perceptron multicouche a été beaucoup utilisé par différents chercheurs dans le domaine d'aide au diagnostic médical.

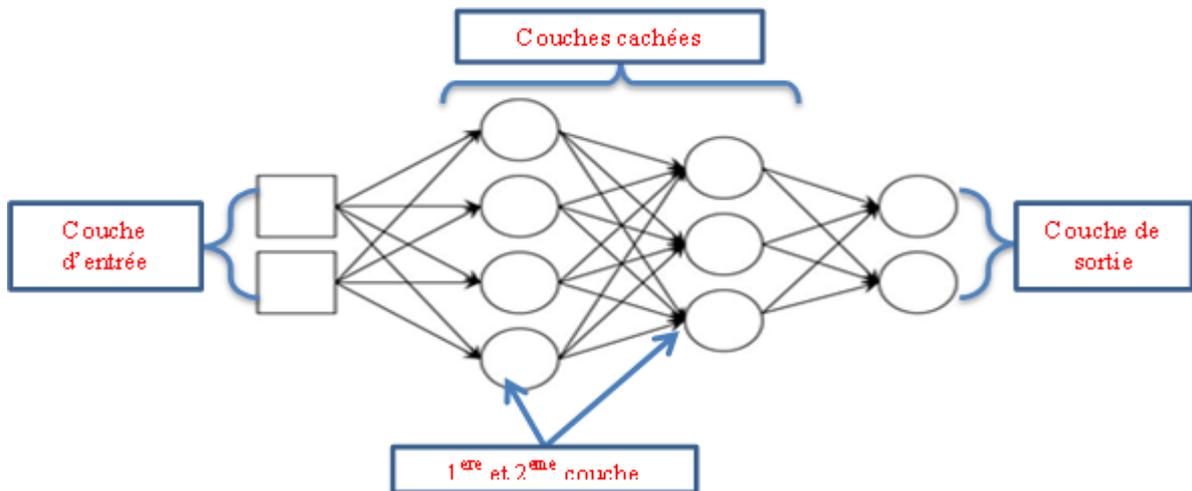


Figure III.7: Architecture d'un perceptron multicouche

Il existe des modèles autres que le PMC tel que les réseaux de Kohonen et les réseaux à Fonctions de Base Radiale (RBF). La différence entre les PMC et les RBF est que la fonction d'activation des RBF est une gaussienne contrairement aux sigmoïdes dans les PMC [36].

III.2.2. 3. L'algorithme de rétro-propagation du gradient

L'algorithme de rétro-propagation du gradient (Back propagation) est une méthode qui permet de calculer le gradient de l'erreur pour chaque neurone d'un PMC. Cela est effectué de la dernière couche vers la première. Cet algorithme itératif a pour objectif de faire converger le modèle vers un minimum d'erreur entre les valeurs désirées et celles calculées.

Pour chaque neurone du réseau, le gradient de l'erreur est calculé afin de mesurer la contribution de chacun des poids synaptiques à l'erreur commise. Les poids synaptiques qui contribuent à engendrer une erreur importante se verront modifiés de manière plus significative que les poids qui ont engendré une erreur minimale.

L'algorithme de rétro-propagation du gradient d'erreur est résumé en sept étapes comme suite :

1. Effectuer un apprentissage du réseau de neurones.
2. Comparaison entre la sortie du réseau calculée avec la sortie désirée.
3. Calculer l'erreur en sortie de chaque neurone du réseau.
4. Calculer pour chacun des neurones la valeur de sortie.
5. Ajustement du poids de chaque connexion vers l'erreur la plus faible.

6. Sanctionner les neurones qui ont commis des grandes erreurs.
7. Recommencer à partir de l'étape 4, pour les neurones sanctionnés jusqu'à l'obtention de l'erreur la plus petite. [8]

III.2.3. Les K plus proche voisins (KNN)

La méthode des k Plus Proches Voisins KNN est une technique de classification non-paramétrique et simple à implémenter. Son fonctionnement est en mode supervisé. Elle a été introduite en 1951 par les deux chercheurs Fix et Hodges aux états unis (USA) [37]. C'est la première méthode de classification qui propose une analyse de similitude entre les données en utilisant des distances.

L'algorithme réalise un calcul de distance entre tous les individus. Suite à cela, chaque objet est classé dans le groupe où se trouvent ses K plus proches voisins qui sont des individus de l'ensemble des données d'entraînement.

Alors que le nombre de k doit être déterminé par l'utilisateur. En effet, une grande valeur de k réduit l'effet du bruit sur la classification, par contre une faible valeur de k permet de mieux visualiser les frontières entre classes. Un bon choix pour k peut être réalisé grâce à des techniques heuristiques, par exemple, la validation-croisée. [29]

Et pour la distance ; son choix est primordial au bon fonctionnement de la méthode, il existe différents types de distances telles que :

- Distance euclidienne.
- Distance de Hamming.
- Distance corrélation.
- Distance de Manhattan.
- Distance cosinusoidale.

KNN est une technique robuste et capable de produire des résultats de haute performance, même pour des applications complexes. Elle a été beaucoup utilisée dans le domaine d'aide au diagnostic médical.

Cependant, cette méthode devient inefficace quand la taille de l'ensemble d'apprentissages est importante. En effet, à chaque nouvelle observation à classer, cette méthode mesure toutes les distances entre cette observation et celles de toutes les observations de l'ensemble d'apprentissages. Elle nécessite, dans ce cas, des ressources mémoires importantes et un temps de calcul élevé qui dépend principalement de la distance utilisée. Ainsi, avec la méthode des KNN, les prévisions reposent sur un ensemble d'exemples prototypes qui servent à prévoir de nouvelles données (c'est-à-dire, inconnues), et cela, sur la base du VM des K plus proches prototypes. [8]

III.2.3.1. Algorithme des KNN

Le principe intuitif de l'algorithme des KNN est que les objets les plus proches ont le plus de chances d'appartenir à la même catégorie (figure III.8).

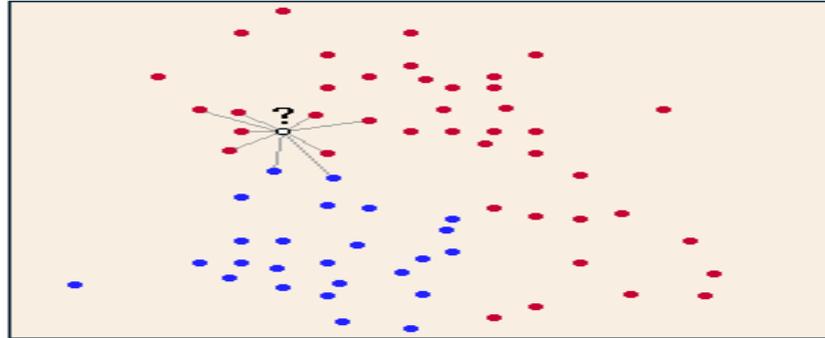


Figure III.8 : Fonctionnement de KPP

L'algorithme standard se résume en deux étapes clés qui sont :

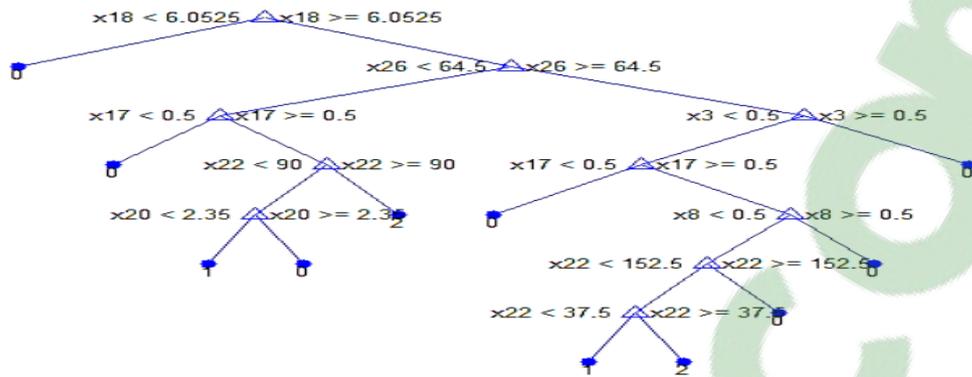
1. Déterminer $N_k(x)$, l'ensemble des k plus proches voisins de x
2. Choisir la classe de x sur la base d'un vote majoritaire dans $N_k(x)$

III.2.4. Arbre de décision (ADD)

Un arbre de décision est un outil puissant d'aide à la décision et à l'exploration de données. Il permet de modéliser simplement, graphiquement et rapidement un phénomène mesuré plus ou moins complexe. Cette technique est connue par son interprétabilité, sa rapidité d'exécution et le peu d'hypothèses nécessaires a priori, et sa vaste utilisation dans différentes applications tel que le domaine médical. Il est composé de :

- une racine, qui est le point de départ.
- des nœuds, où sont réalisés des tests et
- des feuilles, qui contiennent l'information cherchée.

Le parcours d'un arbre, ce fait, par des tests successifs, de la racine à une feuille et constitue une règle de décision. (Figure III.9). [8]



La figure III.9 : Architecture d'un arbre de décision

III.2.4.1. Construction d'un arbre de décision

III.2.4.1.1. Bases de règles

Un arbre est un système de règles de type particulier. En effet, il y a autant de règles que de feuilles et autant de tests que de nœuds avec des différences :

- les règles sont mutuellement exclusives.
- une règle est entièrement évaluée.
- les tests des attributs sont réalisés dans un ordre donné. [8]

III.2.4.1.2. Prétraitements

Nous attachons un test à chaque nœud et une réponse à chaque feuille, ce qui définit la spécification des arbres de décision. Avec les avantages suivants :

- suite de tests mono variés.
- pas de problème de combinaison de variables, de différentes natures.
- hiérarchisation de variables intégrée.
- rapidité d'exécution.
- génération de règles logiques simples de classification. [8]

Il existe une multitude de modèles d'arbres de décision tels que : l'ADD classique, l'algorithme CART (Classification And RegressionTree) ; celui-ci a été développé en 1984 par les chercheurs L. Breiman, et al. [40] ID3 et C4.5 ont été développés par Ross Quinlan en 1975 [41], etc. Dans notre travail nous, avons utilisé l'algorithme CART.

Les avantages offerts par les méthodes utilisant ADD sont : leur rapidité et, surtout, leur pouvoir d'interprétation des règles de décision. La clarté des règles de décision rend possible le dialogue homme-machine. En outre, ces méthodes ne font aucune hypothèse sur les données (méthodes non-paramétriques). Par ailleurs, elles ont l'inconvénient d'être extrêmement complexes si le nombre d'attributs et de classes augmentent.

Dans cette section, nous avons présenté des fondements théoriques des différentes techniques de classification supervisée qui sont les SVMs, RNA, KNN, et ADD, tout en montrant certains de leurs domaines d'application ainsi que leurs avantages et inconvénients.

Dans la section suivante, nous allons détailler les méthodes de combinaison utilisées dans ce travail

III.3. Multi classification

III.3.1. Introduction

L'idée principale derrière la combinaison de classifieurs est l'augmentation de performances. Cette augmentation de performance peut se traduire par plus de fiabilité dans les réponses, moins de rejet, ou bien les deux en même temps. En effet, de plus en plus, les chercheurs se rendent compte qu'il n'existe aucune méthode pouvant manifester une supériorité sur les autres dans tous les problèmes et toutes les situations. Toutes les méthodes ont leurs points forts et leurs points faibles. Si aucune d'entre elles ne peut satisfaire entièrement les exigences d'une application envisagée, l'utilisation simultanée de plusieurs méthodes en même temps peut éventuellement permettre d'en cumuler les avantages sans en cumuler les inconvénients. [44]

Différents schémas de combinaison de classifieurs ont été proposés dans la littérature selon les critères de combinaison. Parmi ces schémas, on trouve : la combinaison séquentielle, parallèle et hybride.

a. La combinaison séquentielle :

Un ou plusieurs classifieurs sont exécutés et les résultats obtenus par ce ou ces classifieurs sont utilisés pour modifier l'exécution d'autres classifieurs. L'ordre d'exécution des classifieurs est important et s'il est changé, le résultat final peut être différent. La figure III.10 montre une représentation de la combinaison séquentielle de classifieurs. [45].

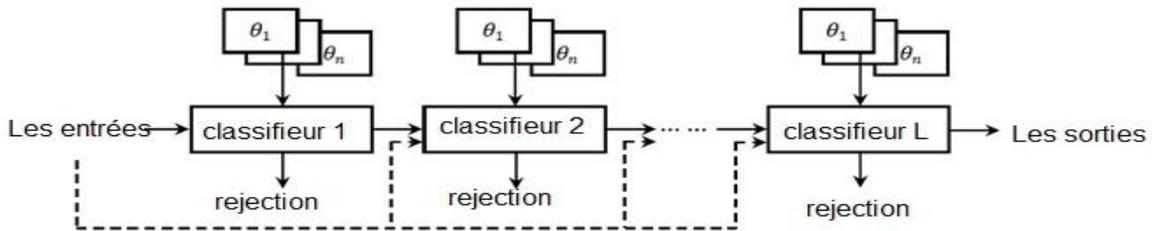


Figure III.10 : Combinaison séquentielle des classifieurs

Où θ_i est l'ensemble des échantillons de la $i^{\text{ème}}$ classe, $i = 1, \dots, n$.

b. La combinaison parallèle :

Cette méthode conduit dans un premier temps différents classifieurs à fonctionner indépendamment les uns des autres, puis combine leurs réponses respectives. L'ordre d'exécution des classifieurs n'est pas impliqué dans cette démarche. La figure III.11 montre une représentation de la combinaison en parallèle de classifieurs. [23]

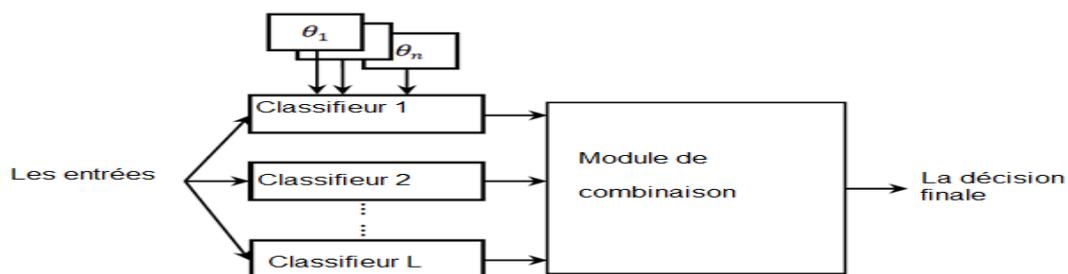


Figure III.11 : Combinaison parallèle des classifieurs

c. La combinaison hybride :

L'approche hybride combine les deux architectures parallèles et séquentielles afin de profiter pleinement de chaque classifieur utilisé. La figure III.12 montre un exemple de combinaison hybride dans lequel un classifieur en série est combiné avec deux classifieurs en parallèle. Ce type d'approche permet de générer de nombreux programmes de coopération qui peuvent devenir rapidement complexes à optimiser. Il illustre les deux aspects de la combinaison: d'une part, la réduction de l'ensemble des classes possibles, d'autre part la recherche d'un consensus entre les classifieurs afin de parvenir à une décision unique. [23]

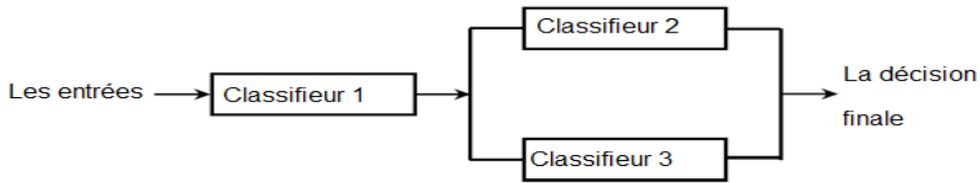


Figure III.12 : Combinaison hybride des classifieurs

Dans notre travail, nous avons utilisé la combinaison parallèle et nous avons effectué deux scénarios de cette dernière qui sont une combinaison homogène et une combinaison hétérogène.

Dans cette section, nous allons vous présenter le principe des deux méthodes de combinaison utilisées qui sont la DST et le VM ainsi le principe de fonctionnement des différentes techniques de sélection de variables à savoir l’algorithme Relief et l’algorithme Rank

III.3.2. Combinaison Homogène

Ce type de combinaison se définit par l’application d’un seul classifieur sur des entrées différentes comme par exemple des bases de données obtenu par différent méthodes de sélection de variables, ensuite les sorties vont être combiné via les méthodes de combinaison tel que la DST et le VM pour l’obtention d’une décision finale.

Dans ce qui suit nous allons vous présentez le processus de quelques méthodes de sélection de variables.

III.3.2.1. Processus de sélection de variables

Chaque phénomène est caractérisé par un ensemble de variables. Ces variables ne sont pas toute informatives. En effet, certaines d’entre elles peuvent être peu significatives, corrélées ou non-pertinentes .La sélection des variables pertinentes (figure III.13) présente, un intérêt majeur qui permet non seulement de réduire le volume de l’information à traiter et par conséquent de réduire le temps de calcul et la complexité des algorithmes de classification, aussi, il permet d’améliorer les performances de généralisation de ces derniers.

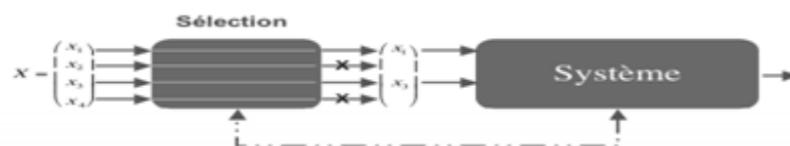


Figure III.13 : principe de la sélection de variables

Nous citons les deux différentes techniques réalisées dans ce travail et qui sont : la méthode de sélection par l'algorithme de « ReliefF » et par l'algorithme « Rank ».

III.3.2.1.1. Méthode Relief

C'est un algorithme qui donne une liste ordonnée de toutes les variables suivant leur pondération par un critère de distance ainsi, il permet de choisir un échantillon d'instances à partir de l'ensemble d'apprentissages. Le nombre d'instances et le seuil de sélection de variables sont fixés par l'utilisateur.

La meilleure variable est celle qui discrimine les plus proches voisins appartenant à des classes différentes et qui comporte des valeurs similaires pour ceux de mêmes classes. Le nombre de variables retenues dépend des seuils retenus par l'utilisateur.

Algorithme de la méthode de sélection Relief [30]:

1. Initialiser les poids.
2. Tirer aléatoirement une donnée X_i
3. Trouver les k plus proches voisins de X_i ayant les même étiquettes (hits),
4. Trouver les k plus proches voisins X_i ayant une étiquette différente de la classe X_i (misses)
5. Pour chaque caractéristique mettre à jour les poids

$$W_d = W_d - \sum_{j=1}^k \frac{\text{diff}(x_i, d_i, \text{hits}_j)}{m * k} + \sum_{c \neq \text{class}(x_i)} \left(\frac{p(c)}{1 - p(\text{class}(x_i))} \right) \sum_{j=1}^k \frac{\text{diff}(x_i, d_i, \text{misses}_j)}{m * k} \quad (12)$$

6. La distance utilisée est définie par :

$$\text{diff}(x_i, d_i, x_j) = \frac{|x_i d - x_j d|}{\max(d) - \min(d)} \quad (13)$$

Relief est capable de travailler avec des variables bruitées et corrélées et demande un temps d'exécution linéaire par rapport au nombre de variables et au nombre d'instances. Cet algorithme est également capable de traiter des données nominales et continues. L'inconvénient majeur est qu'il n'apporte pas d'amélioration avec des variables redondantes, il génère souvent un sous-ensemble de taille non-optimale en présence de variables non répété. Ce problème peut être résolu par une recherche ultérieure

approfondie dans le sous-ensemble de variables obtenues par Relief. Aussi l'un des inconvénients de Relief est qu'il ne travaille que sur des classes binaires [46]

III.3.2.1.2. Méthode Rank

Pour la sélection d'un sous-ensemble de variables, un score pour chaque variable est attribué en fonction des corrélations variable-classe et des corrélations variable-variable selon la formule suivante : [47]

$$r_{\theta S} = \frac{k\bar{r}_{\theta i}}{\sqrt{k + k(k-1)\bar{r}_{ij}}} \quad (14)$$

Avec $r_{\theta S}$ le score du sous-ensemble de cardinal k , $\bar{r}_{\theta i}$ la moyenne arithmétique des corrélations entre θ et les variables i , et \bar{r}_{ij} la moyenne des k^2 intercorrélations entre variables. Cette équation exprime que le score du sous-ensemble augmente si les variables sont fortement corrélées avec θ et diminue si elles sont fortement corrélées entre elles.

De plus, cette formule est valable dans le cadre de scores normalisés (i.e. de variance unitaire). L'idée est de dire qu'un bon sous-ensemble est constitué de variables hautement corrélés avec la classe (pour ne pas garder les indépendances), et peu corrélées entre elles (pour éviter la redondance). [47]

La corrélation ou la dépendance entre deux variables peut être définie de plusieurs façons. Utiliser, le coefficient de corrélation statistique, le test de Fisher, le test du Chi-2, t-statistic, F-statistic et l'entropie de Shannon. [30].

III.3.3. Combinaison Hétérogène

Ce type de combinaison peut être défini par l'application de plusieurs classifieurs sur la même base de données et les sorties de ces derniers vont être combiné via les méthodes de combinaison comme le VM et la DST pour obtenir à la fin une décision finale.

Dans la suite, nous allons détailler les deux méthodes de combinaison utilisées dans ce travail et qui sont le VM et la DST :

III.3.3.1 Vote majoritaire (VM)

Le vote majoritaire consiste à faire voter de manière indépendante plusieurs classifieurs et à retenir la réponse majoritaire, correspondant normalement à la meilleure solution. Ceci est issu du fait que les classifieurs pris individuellement sont incapables d'assurer une bonne reconnaissance pour tous les chiffres. Suivant la méthode qui les anime, ils sont très performants pour certains et moins pour d'autres [48].

III.3.3.1.1 Principe du vote majoritaire : [49]

Le principe du vote est la méthode de fusion d'informations la plus simple à mettre en œuvre. Plus qu'une approche de fusion, le principe du vote est une méthode de combinaison particulièrement adaptée aux décisions de type symbolique. Notons $S_j(x) = i$ le fait que la source S_j attribue la classe C_i à l'observation x . Nous supposons ici que les classes C_i sont exclusives. À chaque source nous associons la fonction indicatrice :

$$M_i^j(x) = \begin{cases} 1 & \text{si } S_j(x) = i, \\ 0 & \text{sinon.} \end{cases} \quad (15)$$

La combinaison des sources s'écrit par :

$$M_k^E(x) = \sum_{j=1}^m M_k^j(x) \quad (16)$$

Pour tout k . L'opérateur de combinaison est donc associatif et commutatif. La règle du VM consiste à choisir la décision prise par le maximum de sources, c'est-à-dire le maximum de M_k^E . Cependant, cette règle simple n'admet pas toujours de solutions dans l'ensemble des classes $D = \{C_1, \dots, C_n\}$. En effet, par exemple si le nombre de sources m est pair et que $m/2$ sources décident C_{i_1} et $m/2$ autres sources disent C_{i_2} , ou encore dans le cas où chaque source affecte à x une classe différente. [49]

La règle la plus employée est la règle du vote majoritaire absolu qui s'écrit :

$$E(x) = \begin{cases} k & \text{si } \max_k M_k^E(x) > \frac{m}{2} \\ n + 1 & \text{sinon.} \end{cases} \quad (17)$$

À partir de cette règle, il a été démontré [64] que plusieurs résultats prouvant que la méthode du vote permet d'obtenir de meilleures performances que toutes les sources prises séparément, sous des hypothèses d'indépendance statistique des sources et de même probabilité, et ceci est d'autant plus vrai que m est impaire.

III.3.3.2. Théorie de l'évidence de Dempster- Shafer (DST)

III.3.3.2.1. Définition

La théorie de l'évidence de Dempster-Shafer (ou Dempster-Shafer Theory), appelée également théorie du raisonnement plausible, est souvent présentée et considérée comme une généralisation de la théorie bayésienne des probabilités subjectives. Cette théorie a été introduite par Dempster puis développée par Shafer. Elle est basée sur un fondement mathématique solide, permet de donner des solutions satisfaisantes en termes de modélisation et en termes de complexité liées aux problèmes de la fusion de données [68].

Elle a été utilisée pour la fusion de données dans divers domaines tels que, la télédétection pour des applications militaires et civiles, la médecine pour des applications de segmentation d'images IRM, la robotique pour des applications de localisation de robots mobiles dans un environnement partiellement connu

III.3.3.2.2. Présentation de la théorie

III.3.3.2.2. 1. Cadre de discernement

C'est un ensemble Θ incluant tous les singletons possibles d'un contexte précis. Les hypothèses θ_i de l'ensemble Θ sont supposées exclusives dans la DST. Aussi, l'ensemble Θ est considéré comme étant exhaustif.

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$$

On appelle cadre de discernement exclusif, un cadre de discernement dont les éléments sont, deux à deux, disjoints, i.e. $\theta_i \cap \theta_j = \emptyset, \forall i \neq j$.

On appelle cadre de discernement exhaustif, un cadre de discernement qui contient l'objet recherché. Si l'objet recherché ne se trouve pas dans le cadre de discernement, nous parlons alors d'un cadre de discernement non-exhaustif. [42]

L'ensemble puissance de Θ , que nous notons 2^Θ , est l'ensemble de tous les sous-ensembles de Θ .

$$2^\Theta = \{\emptyset, \theta_1, \theta_2, \dots, \theta_N, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \dots, \Theta\}$$

III.3.3.2.2. 2. Définition des fonctions de masse

Soit Θ un cadre de discernement. Une fonction $m : 2^\Theta \rightarrow [0,1]$ est appelée une fonction de masse si et seulement si elle respecte les axiomes suivants :

$$m(\emptyset) = 0 \quad (18)$$

$$m(A) \in [0,1] \quad \forall A \subseteq \Theta \quad (19)$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \quad (20)$$

III.3.3.2.2. 3. Les étapes de la théorie de Dempster-Shafer

Cette théorie se décompose en trois étapes fondamentales :

a. Estimation de masses :

Il existe plusieurs modèles pour l'estimation des fonctions de masse parmi ces modèles, nous citons : le modèle de probabilité inférieure et supérieure de Dempster, le

modèle d'estimation par une distance, le modèle de croyances transférables (TBF) de Smets, le modèle consonant de Dubois et Prade, et le modèle d'Appriou[68]

Dans ce travail nous avons utilisé le modèle Appriou[72] modifié par N.Abbas [23] qui est défini uniquement pour deux classes.

Avant de faire une estimation de masses il faut d'abord passer par ces deux étapes :

- Calculer la fonction de décision pour chaque classifieur.
- Faire une normalisation de la fonction de décision entre 0 et 1.

Ensuite les masses de chaque classifieur vont être calculé selon ces formules [] :

$$m_{id}(classe_1) = \frac{(1 - \beta_{id})h_d(classe_1)}{1 + h_d(classe_1)} \quad (21)$$

$$m_{id}(classe_2) = m_{id}(\overline{classe_1}) = \frac{1 - \beta_{id}}{1 + h_d(classe_2)} \quad (22)$$

$$m_{id}(classe_2 \cup classe_1) = \beta_{id} \quad (23)$$

$$m_{id}(classe_2 \cap classe_1) = 0 \quad (24)$$

où:

$m_{id}(classe_1)$: représente la masse de la classe 1.

$m_{id}(classe_2)$: représente la masse de la classe 2.

β_{id} : est un paramètre fixé par l'utilisateur.

$h_d(classe_1), h_d(classe_2)$: sont obtenues par l'algorithme de normalisation de la fonction de décision.

a.1. Critères de la DST :

Nous avons vu que la modélisation de Dempster-Shafer était très riche. Pour exprimer de manière plus explicite, notre connaissance du problème, quatre fonctions ont été établies, dérivées de la fonction de masses. Ce sont les fonctions de crédibilité, croyance, de plausibilité et de communalité [73]

a.1.1. Crédibilité (cr) :

La crédibilité (BeliefFunction) correspond à la quantité d'informations qui est tout entière contenue dans le sous-ensemble considéré. Il contient l'ensemble des connaissances créditant la véracité de ce sous-ensemble. [73]Par définition :

$$Cr(A) = \sum_{B \subseteq A} m(B) \quad (25)$$

a.1.2. Fonction de croyance (Bel):

La croyance accordée à un sous-ensemble A de Θ représente la confiance totale dont le sous-ensemble A peut bénéficier [73]. La fonction de croyance,

Bel : $2^\Theta \rightarrow [0,1]$, peut être définie à partir de la fonction de masse par :

L'équation de croyance de A est notée Bel(A), est représentée par la somme des masses des ensembles inclus dans A.

$$Bel(A) = \sum_{B_i \subseteq A} m(B_i) \quad (26)$$

a.1.3. Fonction de plausibilité (pl) :

La plausibilité d'un sous-ensemble A correspond à la quantité d'information ne discréditant pas A, c'est-à-dire toute l'information contenue dans les sous-ensembles ayant une intersection avec A: [73]elle, est représentée par la somme des masses des ensembles qui ont une intersection non-nulle avec A.

$$Pl(A) = 1 - Bel(\bar{A}) \quad (27)$$

a.1.4. Communalité (Q)

La Communalité d'un sous-ensemble de A est définie par la relation suivante[73]:

$$Q(A) = \sum_{A \subseteq B} m(B) \quad (28)$$

Cela correspond à la quantité d'information de tous les sous-ensembles contenant A. La communalité n'a pas une interprétation physique aussi simple que la crédibilité ou la plausibilité, mais elle est parfois utilisée dans le calcul de combinaison de jeu de masses dans le cadre de la fusion.

b. Combinaison évidentielle :

En présence de plusieurs capteurs ou de plusieurs informations provenant d'un même capteur, il devient intéressant de combiner les connaissances de chaque source pour en extraire une connaissance globale et d'y appliquer une règle de décision. De manière générale, une règle de combinaison entre deux fonctions de masse dans la théorie de l'évidence est notée par \oplus et est appelée somme orthogonale. Elle est définie par:[42]

$$\left\{ \begin{array}{l} (m_1 \oplus m_2)(\emptyset) = 0 \quad (29) \\ m(B) = (m_1 \oplus m_2)(B) = \frac{\sum_{A_1 \cap A_2 = B} m_1(A_1) \cdot m_2(A_2)}{1 - K}, B \neq \emptyset \quad (30) \\ \text{avec } K = \sum_{A_1 \cap A_2 \neq \emptyset} m_1(A_1) \cdot m_2(A_2) \quad (31) \end{array} \right.$$

Ou K est appelé le conflit entre les deux sources m_1 et m_2 , Si K est proche de 0, les deux fonctions de masse ne sont pas en conflit, mais si le conflit K est proche de 1, les deux fonctions de masse se contredisent et si $K=1$ la règle de Dempster n'est pas définie. [42]

c. *Processus de décision :*

Contrairement à la théorie Bayésienne où le critère de décision est très souvent le maximum de vraisemblance, la théorie de l'évidence propose de nombreuses règles. Les plus utilisées sont le maximum de crédibilité, le maximum de plausibilité, les règles basées sur l'intervalle de confiance, le maximum de probabilité pignistique [74] et la décision par maximum de vraisemblance.

c.1. *Maximum de crédibilité :*

On choisit l'hypothèse simple possédant la crédibilité maximum. C'est l'hypothèse la plus souvent choisie, en raison de sa simplicité de mise en œuvre. [73]

c.2. *Maximum de plausibilité :*

Cette position correspond à un comportement plus prudent, qui trouve son utilité en cas de conflit entre les sources. Par contre, lors d'une absence de conflit, toutes les plausibilités sont égales à 1, donc cette règle n'est pas adaptée à tous les cas, et ne permet pas de conclure. [73]

c.3. *Règles basées sur l'intervalle de confiance [Cr, Pl] :*

L'intervalle de confiance [Cr, Pl] exprime la connaissance que l'on a sur une hypothèse, bien mieux que chaque notion prise séparément:

En effet, un intervalle de confiance de [1, 1] certifie que cette hypothèse est réellement sûre, alors qu'un intervalle de confiance de [0,0] certifie que cette hypothèse est impossible. [73]

On répartit uniformément la masse des hypothèses composées sur chacune des hypothèses simples la composant. Ensuite on applique un critère classique de maximum a posteriori provenant de la théorie Bayésienne de la décision. [73].

III.3.3.2.3. Avantages et inconvénients de la théorie de Dempster-Shafer

Avantages :

- la théorie de Dempster-Shafer est une extension de l'approche probabiliste à la gestion de l'incertain.
- la théorie de l'évidence peut apporter des informations complémentaires et permet de mettre en évidence d'une manière différente.

Inconvénients :

- le comportement peut aussi s'avérer très sensible par moment
- cette combinaison exige une grande quantité de calculs par rapport aux méthodes classiques

III.4. Conclusion

Dans ce chapitre, nous avons d'abord présenté le principe de fonctionnement des différentes techniques de classification supervisée qui sont les machines à vecteurs de support, les réseaux de neurones, les K- plus proches voisins, et les arbres de décision ensuite nous avons détaillé les deux scénarios de combinaison qui sont la combinaison homogène et hétérogène tout en montrant le processus de sélection de variables et des fondements théoriques de la méthode d'évidence et du vote majoritaire.

Dans le chapitre suivant nous allons présenter les résultats expérimentaux obtenus pour chaque classifieur à part, ainsi que les résultats de la combinaison proposé.

*Chapitre IV : Résultats et
Discussion*

IV.1. Introduction

Après la présentation du principe de fonctionnement des différentes approches de classification ainsi le fondement théorique des différentes méthodes de combinaison dans le chapitre précédent, nous présenterons dans ce chapitre l'implémentation de ces méthodes en utilisant la base de données hypothyroïdie.

En effet, dans ce chapitre, nous présenterons les résultats expérimentaux obtenus suites aux tests effectués pour chaque classifieur à part, ainsi que les résultats du prototype de combinaison proposé dans sa globalité. C'est pour cela nous décomposant ce chapitre en plusieurs sections. Dans la première, nous allons présenter le contexte médical de la maladie d'hypothyroïdie, pour la deuxième section, nous allons détailler la base de données utilisée dans ce travail, la troisième section résume les résultats obtenus suite au test de chaque classifieur indépendamment, pour la quatrième section, nous allons discuter sur les résultats de la combinaison homogène et pour la dernière section, nous allons également discuter les résultats de la combinaison hétérogène.

IV.2. l'hypothyroïdie

IV.2.1. Définition de la thyroïde

La glande thyroïde est un organe en forme de papillon situé juste au-dessous du larynx, sous la pomme d'Adam (Figure IV.1). Généralement, elle n'est pas palpable. Elle produit deux hormones:

- a. La thyroxine, aussi appelée tétraiodothyronine, ou T4, parce qu'elle contient quatre atomes d'iode.
- b. La triiodothyronine, ou T3, qui porte trois atomes d'iode.

La T3 et la T4 sont aussi appelées hormones thyroïdiennes :[75]

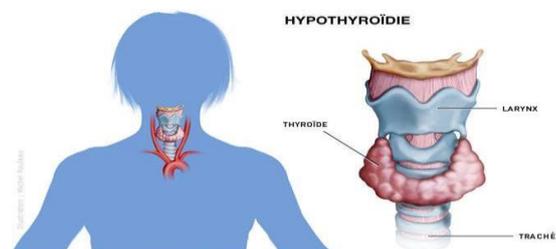


Figure IV.1: Schéma représentatif de la thyroïde

Nous citons le fonctionnement de cet organe:

- elles régulent le métabolisme de base des cellules de notre corps.
- elles contrôlent l'énergie musculaire, l'humeur, la température du corps.

- elles augmentent le rythme cardiaque, etc.
- elles ont un rôle dans l'utilisation et la transformation des glucides, des lipides et des protides.
- chez l'enfant, elles sont indispensables à la croissance et au développement, en particulier du système nerveux central et des os.

Remarque : La thyroïde est sous la dépendance d'une autre glande, l'hypophyse, située à la base du cerveau. Cette glande produit, la thyroïdostimuline (ou TSH) qui régule la sécrétion des hormones thyroïdiennes.

IV.2.2. Ensemble hypothalamus/hypophyse/thyroïde

Deux autres organes produisent des hormones qui exercent une action sur la glande thyroïde (Figure IV.2) :

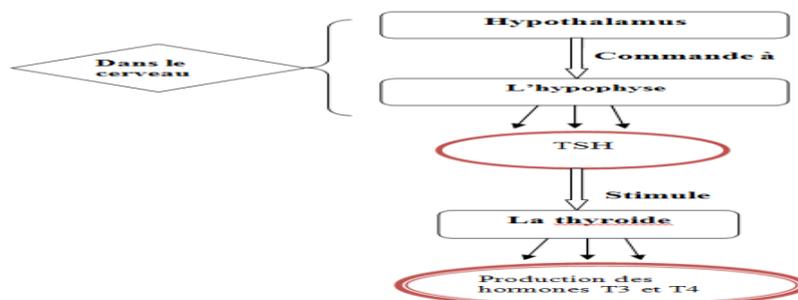


Figure IV.2 : Schéma représentatif de la procédure de fonctionnement du cerveau afin de contrôler la thyroïde

L'hypophyse : produit une hormone qui stimule la thyroïde, la thyrotrophine, la TSH. Sous l'action de celle-ci, la T3 et la T4 sont synthétisées par la fixation d'atomes d'iode à la tyrosine. La sécrétion de la TSH est régie par la thyroïdolibérine, la TRH de l'hypothalamus. L'hypothalamus : produit la TRH. La libération de cette dernière dépend, entre autres facteurs, des concentrations sanguines de TSH, de T3 et de glucose ainsi que de l'activité métabolique de l'organisme.[75]

Remarque :

- Si la quantité d'hormones T3 et T4 sécrétée est grande donc on a l'hyperthyroïdie
- Si la quantité d'hormones T3 et T4 sécrétée est petite donc on a l'hypothyroïdie

IV.2.3. Définition de l'hypothyroïdie

L'hypothyroïdie peut se définir comme étant une sécrétion insuffisante d'hormones thyroïdiennes T3 et T4 par la glande thyroïde. Ce défaut de sécrétion peut être dû à une anomalie de la glande elle-même.[22]

IV.2.4. Comment détecter l'hypothyroïdie ?

On peut détecter une hypothyroïdie en mesurant le taux de TSH dans le sang, et on le trouve élevé, car l'hypophyse réagit au manque d'hormones thyroïdiennes (T3 et T4) en sécrétant d'avantage TSH en tentant de stimuler la thyroïde pour qu'elle produise plus d'hormones.

IV.2.5. Les facteurs de risque :[75]

- les femmes et les personnes âgées de plus de 50 ans sont les plus touchées.
- les personnes qui ont des antécédents familiaux de la maladie de la thyroïde
- les femmes qui ont enfanté au cours de l'année
- la prise de certains médicaments
- la carence alimentaire en iode.
- les tumeurs de l'hypophyse.

IV.2.6. Les types de l'hypothyroïdie

On distingue plusieurs types d'hypothyroïdies. Un diagnostic précis permet une meilleure prise en charge, c'est-à-dire la mise en place d'un traitement adapté.

a. Hypothyroïdie primaire :

En cas d'hypothyroïdie primaire, la thyroïde ne produit pas suffisamment d'hormones.

- on retrouve alors une TSH élevée et un taux d'hormones T3 et T4 significativement bas.
- D'un point de vue morphologique, la glande thyroïde est de petite taille.

b. Hypothyroïdie secondaire :

Contrairement à la primaire, l'hypothyroïdie secondaire donne à voir une thyroïde ayant conservé une taille normale.

À l'origine de cette pathologie thyroïdienne se trouve l'hypophyse qui ne joue pas correctement son rôle en ne stimulant pas suffisamment la glande thyroïde.

Le taux de TSH est donc particulièrement bas et il s'accompagne d'un taux tout aussi bas d'hormones thyroïdiennes T3 et T4.

c. Hypothyroïdie compensée :

Est caractérisée par un niveau élevé de TSH avec les niveaux normaux ou seulement légèrement réduits d'hormones thyroïdiennes [19]

IV.3. Base de données :

La maladie hypothyroïdie est une maladie qui est en croissance rapide, c'est pour cette raison que nous avons effectué un travail qui permet de détecter cette dernière automatiquement. Comme base de données, nous avons utilisé une base obtenue à partir de site web UCI « machine Learning Respository »¹ et qui contient 3772 individus qui sont répartis en quatre classes et 29 descripteurs.

IV.3.1. Description des paramètres de la base de données

***Age (entier)** : représente l'âge du patient.

***Sex (F,M)** : femme ou homme.

***on thyroxine (f,t)** : hormone contenant de l'iode, sécrétée par la glande thyroïde.

***query on thyroxine (f,t)** : indique si le patient prend la thyroxine

***on antithyroidmedication (f,t)** : si elle est traitée par un médicament ?

***sick (f,t)** : si le patient est malade ou non

***pregnant (f,t)** : grossesses : il existe une augmentation de volume de la thyroïde de 10 à 20 % durant toute grossesse, augmentation plus importante en cas de carence iodée. Beaucoup de goitres ne régressent pas après l'accouchement

***thyroidsurgery (f,t)** : est-ce que le malade effectue un traitement chirurgical ?

***I131 treatment (f,t)** : c'est un type de traitement par radio iode.

***queryhypothyroid (f,t)** : Est-ce que le patient a une hypothyroïdie, c'est-à-dire sécrétion insuffisante d'hormones thyroïdiennes.

***queryhyperthyroid (f,t)** : Est-ce que le patient a une hyperthyroïdie.

***Lithium (f,t)** : Le lithium peut provoquer l'hypo- ou l'hyperthyroïdie chez 10 % des patients traités par lithium.

***Goitre (f,t)** : est une augmentation du volume de l'ensemble de thyroïde.

***Tumor (f,t)** : le cancer de la thyroïde est relativement rare. Il peut se présenter sous différentes formes et avoir des causes multiples. Il concerne essentiellement les personnes plutôt jeunes (3 fois plus de femmes que d'hommes).

***Hypopituitary (f,t)** : la présence de hypo pituitaire ou non

***Psych(f,t)** : les troubles psychiques qui accompagnent l'hypothyroïdie

¹<http://archive.ics.uci.edu/ml/>

***TSH measured(f,t) ;TSH (Real)**:l'hormone TSH stimule la thyroïde pour qu'elle produise les hormones thyroïdiennes, dont la T3 et la T4

***T3measured(f,t) ;T3 (Real)**:principal hormones sécrétées par la thyroïde

***TT4measured (t,f) ;TT4 (Real)**:thyroxine totale (T4T) hormone iodé sécrétée par la thyroïde

***T4U measured (t,f) ; T4U (Real)** :taux de thyroxine utilisé.

***FTI measured (t,f) ;FTI (Real)**:index libre de thyroxine. Pour savoir combien, de hormone thyroïdienne dans le sang.

***TBG measured(f) ; TBG (Real)**:toutes les valeurs de cet attribut dans la base de données sont nulles donc cet attribue n'a pas d'impact sur la classification car n'apporte pas une différence entre les classes.

***Referral source (SVHC,other,SVI,STMW,SVHD)**:il n'apporte aucune information intéressante concernant la maladie de la thyroïde, il s'agit de l'endroit où les données ont été récoltées.

***Class:**

- **Negative** : Le patient n'a pas une hypothyroïdie
- **Compensatedhypothyroid**: Le patient est malade.
- **Primary-hypothyroid** : Indique que le patient est malade son anomalie se situe au niveau de la glande.
- **secondary-hypothyroid**:indique que le patient est malade.

IV.3.2. Prétraitement de la base

Avant d'utiliser cette base de données, nous avons d'abord passé par une étape de prétraitement, on se basant sur l'article de [19] et nous avons remplacé les valeurs des différents attributs par le codage suivant :

***Pour l'attribut « sex » :**

F=femme=2

M=homme=1

***Pour les autres attributs qui contient les valeurs « t,f »**

f=false=0

t=true=1

***Pour l'attribut « classe » :**

1=Non malade

2=malade

En observant le jeu de données, nous avons constaté que certaines variables étaient non significatives donc nous avons effectué les modifications suivantes :

– Suppression de l'attribut « referral source » : il n'apporte aucune information intéressante concernant la maladie de la thyroïde, puisqu'il s'agit de l'endroit où les données ont été récoltées.

– suppression de la variable « TBG measured » car toutes les valeurs de cet attribut dans la base de données sont nulles donc cet attribut n'a pas d'impact sur la classification car n'apporte pas une différence entre les classes.

– supprimer la variable « TBG » car toutes les données sont manquantes et donc on conclut que ce paramètre n'est pas pris en considération lors de l'étude et recueil de données de base de données.

– nous avons remplacé les données manquantes, par la moyenne des deux données suivantes et précédentes.

Puisque notre système de combinaison de classifieur ne fonctionne qu'avec une BDD à deux classes et puisque cette dernière contient quatre classes, on a réalisé ces modifications :

– nous avons supprimé la classe « secondary-hypothyroid » : il n'y a que deux individus qui appartiennent à cette classe.

-Nous avons discuté avec Docteur Meliani qui est un médecin spécialiste en endocrine et avec Docteur Bassaid qui est un médecin résident en chirurgie A au CHU Tlemcen sur les valeurs des descripteurs des individus qui ont la classe primary-hypothyroid et ils nous ont autorisé de rassembler les deux classes « compensated-hypothyroid, et primary-hypothyroid » en une seule classe dite malade.

IV.4. Critère d'évaluation

Dans notre travail, nous avons utilisé la validation croisée avec 5 blocs et nous avons également utilisé des différentes méthodes scalaires à savoir les taux de classification, sensibilité et spécificité afin d'évaluer les performances de notre système.

IV.5. Expérimentation 1 : mono-Classification

En premier lieu, nous avons effectué une étude expérimentale de quatre classifieurs de type RNA, KNN, SVM, ADD d'une manière indépendante afin d'étudier la performance de chaque classifieur (Figure IV.3) d'une part.

D'autre part, effectuant une sélection de variables par deux algorithmes différents (Relief et Rank) dans le but d'extraire les descripteurs les plus pertinents. La figure IV.3 illustre cette architecture:

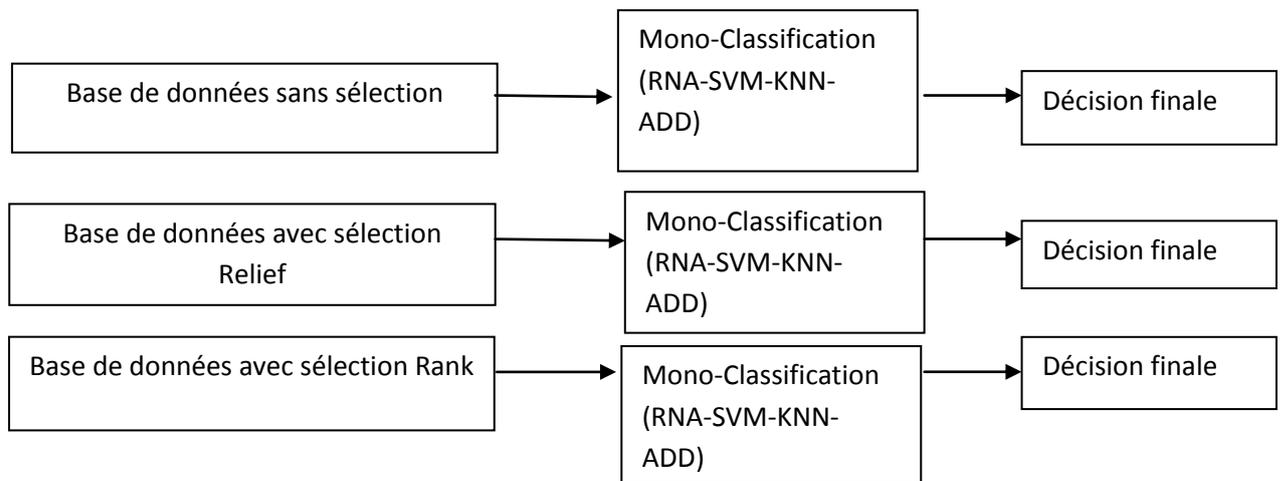


Figure IV.3 : Expérimentation 1 : Mono-classification

IV.5.1. Les paramètres des différents classifieurs utilisés

Dans nos expérimentations, nous avons utilisé le paramétrage suivant :

- **Pour les RNA :**
 - ❖ Nombre époques=150,
 - ❖ Nombre de neurones de couche d'entrée :
 - BDD sans sélection=26
 - BDD avec sélection Relief=20
 - BDD avec sélection Rank=20
 - ❖ Nombre de couches cachées=2
 - ❖ Nombre de neurones dans la couches cachées=[10,15]
 - ❖ Nombre de neurones dans la couche de sortie=2
 - ❖ Fonction d'activation=Sig, Gaus
- **Pour SVM :**
 - ❖ Type de fonction noyau : la fonction gaussienne
- **Pour KNN :**
 - ❖ Nombre de k voisins=5
 - ❖ Distance utilisée=distance euclidienne.
- **Pour ADD :**
 - Paramètres par défaut.

IV.5.2. Sélection de variables

IV.5.2.1. Sélection Relief

Dans cette partie, nous avons étudié l'algorithme de sélection Relief, nous avons utilisé un nombre de voisins qui égale à 5 et nous avons obtenu comme résultats les 26 attributs ordonnés du plus pertinent au moins pertinent.

Dans nos expérimentations, nous avons choisi les 10, 15 et 20 descripteurs les plus pertinents.

Le Tableau IV.1 montre les résultats des TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec différents attributs obtenu par l'algorithme Relief :

	10	15	20
SVM	80,00	88,02	95,00
KNN	84,51	90,32	93,76
RNA	87,14	89,74	93,80
ADD	90,24	91,45	95,65

Tableau IV.1 : Les TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec Différents attributs obtenus par l'algorithme Relief

Nous remarquons que le meilleur taux de classification correspond à la base qui contient 20 attributs donc nous avons décidé d'utiliser cette base pour les étapes suivante c'est-à-dire dans les étapes de combinaison

Les 20 descripteurs sélectionnés sont : 'Age', 'Sex', 'query on thyroxine', 'sick', 'pregnant', 'thyroidsurgery', 'queryhypothyroid', 'queryhypothyroid', 'lithium', 'tumor', 'hypopituitary', 'psych', 'TSH', 'T3', 'TT4', 'TT4', 'T4U', 'T4U measured', 'FTI measured', 'FTI'

IV.5.2.2. Sélection Rank

Dans cette partie, nous avons étudié l'algorithme de sélection Rank, nous avons utilisé un nombre d'indice qui égale 10, 15 et 20.

Le Tableau IV.2 montre les résultats des TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec différents attributs obtenu par l'algorithme Rank:

	10	15	20
SVM	89,41	91,45	96,25
KNN	90,25	91,47	95,36
RNA	90,74	92,54	93,96
ADD	93,56	94,48	97,18

Tableau IV.2 : Les TC du test des différents classifieurs (RNA, KNN, SVM, ADD) sur la BDD avec Différents attributs obtenus par l’algorithme Rank

Nous remarquons que le meilleur taux de classification correspond à la base qui contient 20 attributs donc nous avons décidé d’utiliser cette sous base pour les étapes suivant c’est-à-dire dans les étapes de combinaison

Les 20 descripteurs sélectionnés sont : ‘Sex’, ‘on thyroxine’, ‘on antithyroidmedication’, ‘sick’, ‘pregnant’, ‘queryhypothyroid’, ‘queryhypothyroid’, ‘goitre’, ‘hypopituitary’, ‘psych’, ‘TSH measured’, ‘TSH’, ‘T3 measured’, ‘T3’, ‘TT4’, ‘TT4measured’, ‘T4U’, ‘T4U measured’, ‘FTI measured’, ‘FTI’

Remarque :

Nous avons discuté avec des médecins et ils nous ont confirmés que les descripteurs (TSH, T3, FTI) sont les plus pertinents et les plus informatives, ils permettent la détection de l’hypothyroïdie.

En plus de ces descripteurs, il y a d’autres qui sont également importants comme (Age, Goitre, on thyroxine, on antithyroidmedication), nous remarquons que la plus part de ces derniers sont obtenus par la méthode Rank et non pas par la méthode Relief donc nous pouvons conclure que l’algorithme de sélection Rank est meilleure que l’algorithme Relief selon les explications des médecins

Le Tableau IV.3 illustre les résultats de sélection (Relief et Rank) et aussi sur la base complète obtenus en appliquant les 4 types de classifieurs.

Mono classifieur	Sans Sélection %			Sélection Relief %			Sélection Rank %		
	TC	SE	SP	TC	SE	SP	TC	SE	SP
RNA	95,07	95,04	95,35	95,00	94,87	95,17	96,25	96,18	96,58
SVM	94,84	94,69	94,91	93,76	93,55	93,89	95,36	95,24	95,78
KNN	93,89	93,78	93,92	93,80	93,75	93,94	93,96	93,87	93,99
ADD	99,69	99,65	99,83	95,65	95,61	95,80	97,18	97,09	97,34

Tableau IV.3 : Résultats de l’expérimentation 1 : Mono classification

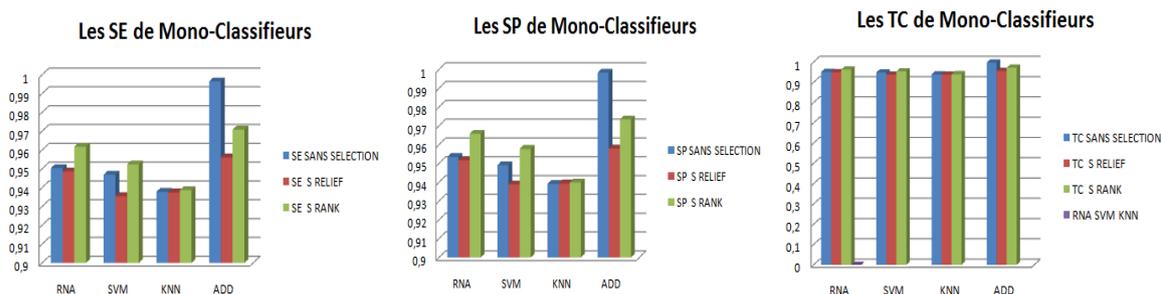


Figure IV.4 : Présentation du TC, SE, SP de l'expérimentation 1 : mono classification

Lors de l'application des différents classifieurs vus précédemment. A pour objectif d'avoir les résultats meilleurs.

Le tableau IV.3 présente les différents critères d'évaluation qui illustrent les performances des quatre différentes techniques de classification (RNA, SVM, KNN, ADD) appliquées aux bases de données (base sans sélection de variables, base avec sélection de variables par Relief, puis par le Rank).

Nous constatons qu'un bon apprentissage et puis par la suite, une bonne classification a été globalement obtenu par les quatre différents classifieurs. En effet, le taux de connaissance de RNA varie entre 95.07% et 96. Le meilleur TC est obtenu par la méthode de sélection Rank, et pour la SE et SP les résultats sont satisfaisantes.

Celui de SVM avec base sans sélection de variables, les critères d'évaluation prennent des valeurs dans l'intervalle 94% pour les trois mesures (TC, SE, SP) et avec la base des attributs sélectionnés par Relief, les trois valeurs sont dégradées et atteignent 93% mais ces derniers sont améliorées avec la technique de sélection Rank dont les trois mesures prennent un TC et une SE,SP de 95 %.

Le classifieur KNN a réussi une reconnaissance acceptable que ce soit avec la base sans sélection de variables ou avec sélection la reconnaissance varie respectivement entre 93.89% 93.80% et 93.96 % (base sans sélection de variables, base avec sélection Relief, puis Rank) et de SE et de SP de 93%.

Le classifieurs d'arbre de décision (Cart) également donné une bonne reconnaissance d'hypothyroïdie. En effet, il montre une excellente classification de BDDsans sélection de variables avec un TC de 99.69. Nous remarquons une diminution de performance en utilisant les techniques de sélection.

En mono classification, la méthode de sélection Rank a montré son efficacité en termes de performance avec l'utilisation de RNA, SVM, KNN. Par contre, la meilleure reconnaissance a été obtenue par les arbres de décision sans sélection grâce leurs interprétabilité et leurs rapidité d'exécution.

IV.6. Expérimentation 2 : Multi-Classification

Dans cette partie, nous avons effectué une combinaison de multi classifieurs en deux scénarios et les techniques de vote majoritaire et la DST ont utilisé pour la décision finale.

IV.6.1. Vote majoritaire

Nous avons choisi d'effectuer la combinaison entre trois classifieurs non pas deux ou quatre afin de tenir compte le problème de nombre pair de classifieur. C'est-à-dire si nous avons deux classifieurs et ces derniers nous donnent deux sorties différentes alors quelle sortie doit-on choisir ? Et aussi pour voir quel classifieur peut fausser les résultats et quel classifieur permet de donner les meilleurs résultats, pour cela nous avons utilisé trois classifieurs.

IV.6.2. Théorie de Dempster-Shafer

Avant d'appliquer la DST, il faut d'abord calculer les masses de chaque sortie de différents classifieurs.

IV.6.3. Combinaison homogène

En combinaison homogène plusieurs classifieurs de même type sont combinés entre eux. Dans nos expérimentations, nous avons utilisé la sélection de variables afin de différencier les entrées de chaque classifieurs, deux techniques de sélection ont été utilisées et à la fin, une décision est prise par la DST et le vote majoritaire

La figure IV.5 illustre le mécanisme de cette combinaison :

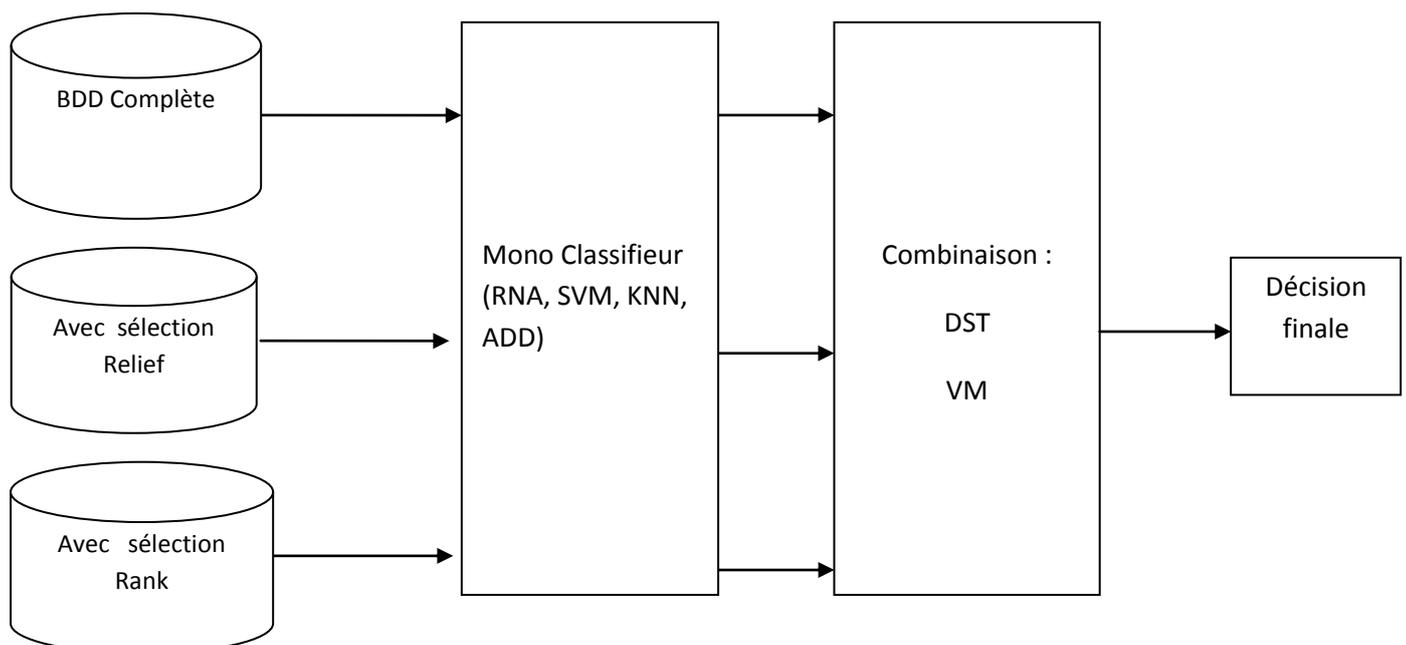


Figure IV.5 : Expérimentation 2 : Combinaison homogène

Combinaison Homogène	VM%			DST%		
	TC	Se	Sp	TC	Se	Sp
RNA	96,96	96,87	97,01	97,50	97,37	97,68
KNN	93,98	93,78	94,08	94,97	94,86	95,00
SVM	95,80	95,69	95,99	96,51	96,41	96,68
ADD	99,70	98,91	99,14	99,83	99,01	99,27

Tableau IV.4 : Expérimentation 2 : Combinaison homogène

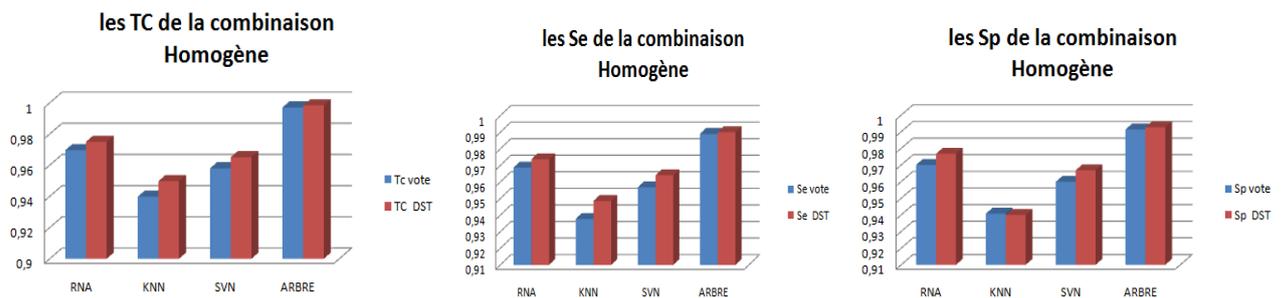


Figure IV.6: Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Homogène

Notre contribution est concentrée sur la réalisation d'un processus qui permet de combiner les différentes décisions provenant de différents classifieurs pour avoir des réponses plus proches au médecin et plus convaincantes, c'est pour cette raison, nous avons appliqué deux mécanismes différents pour réaliser cette fusion d'information la première, c'est le vote majoritaire et une deuxième celle de la théorie de Dempster-Shafer.

Nous avons obtenue pour le RNA avec le VM un TC élevé de 96.96% et uneSE de 96.87%, et uneSPde 97.01%.

Celui de la DST les trois mesures sont améliorées significativement nous avons obtenue respectivement 97.50%, 97.37% et 97.68% de TC, de SE et de SP.

Pour bien comprendre comment et quels sont les individus et logiquement les attributs qui ont causé un conflit pour les classifieurs et par la suite, le VM est tombé dans la même erreur, nous avons réalisé une étape qui permet d'extraire les individus avec leurs attributs qui ont été corrigé par la DST. Après nous avons discuté avec un médecin spécialiste Dr Meliani et un médecin résident en chirurgie A Dr Bassaid sur ces attributs.

Une question primordiale été posé pour-quoi les classifieurs ont échoué explicitement dans ces cas ?? La réponse, c'était que les normes de décision sont proches, c'est-à-dire il peut que le classifieur été embarrassé entre les valeurs des attributs. Cette situation était expliquée par des médecins que c'est un cas infra clinique, c'est-à-dire que le patient ne souffre pas aux effets d'hypothyroïdie, mais le bilan sanguin montre une légère

perturbation. Donc il se peut que deux médecins donnent deux avis différents celle est malade et l'autre est non malade.

Pour le classifieur KNN avec le VM les trois mesures d'évaluation (TC, Se, Sp) prennent des valeurs de 93.98% et 93.78% ,94.08% respectivement.

Ces résultats sont aussi améliorés par la DST qui a marqué un TC de 94.97% et une SE de 94.86% et de SP de 95%.

Citons aussi pour les SVM le VM nous a permis d'avoir des résultats satisfaisants avec un TC de 95.80%, une SE de 95.69%, et 96.68 de SP. Mais pour la DST les mesures sont devenues encore meilleurs, ils ont varié entre 96.51% de TC et 96.41% de Se, 96.68% Sp.

Finalement, l'arbre de décision, c'est lui qui permet de donner dans la première contribution de très bon taux de classification et par la suite, c'est lui aussi qui permet de donner avec le vote majoritaire un TC élevé 99.70%, une SE de 98.91%, et 99.14% de SP. Et la DST, elle permet dans tous les cas d'améliorer les taux par rapport au VM nous avons touché cet avantage dans l'arbre de décision, les trois mesures prennent les valeurs de TC et Se, Sp de 99.83%,99.01%, 99.27% respectivement.

IV.6.4. Combinaison hétérogène

Dans cette partie, nous avons appliqué chaque classifieur indépendamment des autres sur la même base de données, c'est-à-dire sur la base de données sans sélection puis sur la base avec sélection Relief ensuite sur la base avec sélection Rank après nous avons choisi de combiner les sorties des trois classifieur obtenue à partir de leurs application sur la même entrées avec la méthode du VM ensuite avec la méthode de DST.

IV.6.4.1. Combinaison hétérogène (BDD sans sélection de variables)

La Figure IV.7 illustre les expérimentations de combinaison hétérogène.

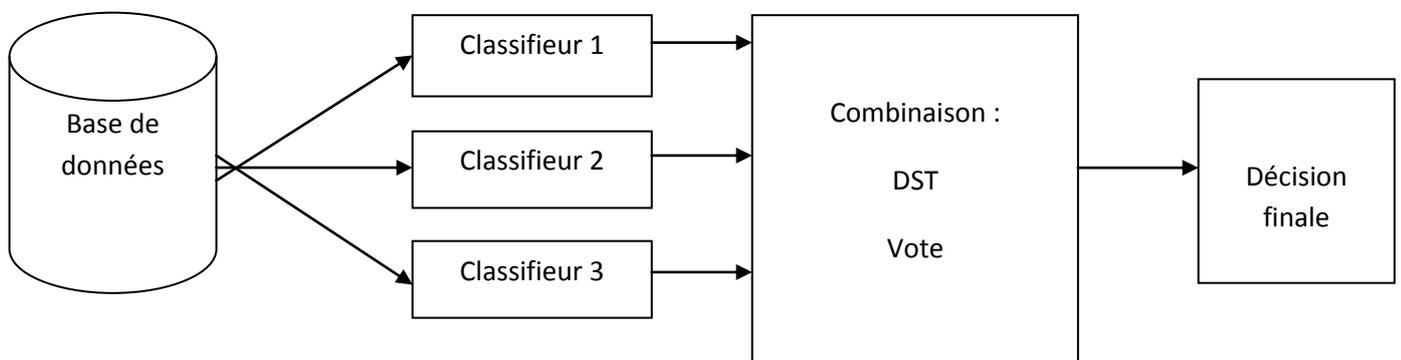


Figure IV.7 : Expérimentation 2 : Combinaison hétérogène (BDD sans sélection)

Combinaison Hétérogène	Sans sélection%					
	VM			DST		
	TC	SE	SP	TC	SE	SP
RNA_SVM_ARBRE	99,76	99,70	99,82	99,85	99,72	99,89
RNA_SVM_KNN	95,70	95,63	95,79	96,53	96,48	96,60
RNA_ARBRE_KNN	99,72	96,69	99,79	99,76	99,68	99,88
KNN_SVM_ARBRE	99,78	99,71	99,84	99,89	99,76	99,92

Tableau IV.5 : Expérimentation 2 : Combinaison hétérogène (BDD sans sélection)

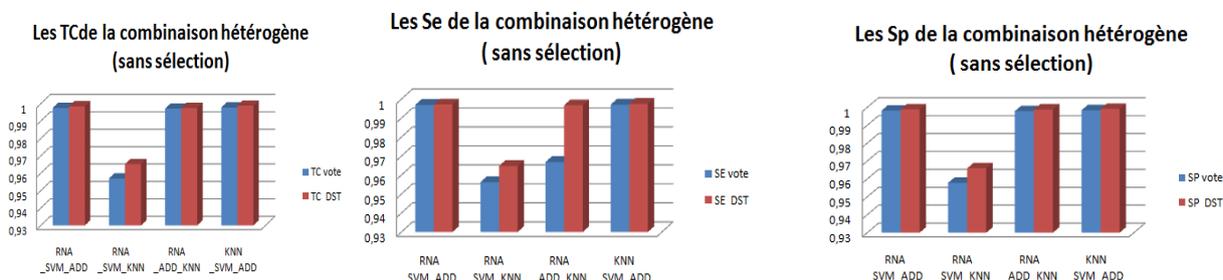


Figure IV.8 : Présentation du TC, SE, SP de l'expérimentation 2 : Combinaison Hétérogène (BDD sans sélection)

La Figure IV.8 Présente les performances à savoir le TC, SE, SP des deux méthodes de combinaison : VM et DST ces dernières méthodes nous ont permis de fusionner les sorties des trois différents classifieurs sur la base de données sans sélection de variables.

Nous remarquons à partir de ces résultats que le taux de classification de la méthode du VM varie entre 95.70% et 99.78% avec une SE qui atteint 99.71% et une SP de 99.84% et nous constatons que la combinaison « KNN_SVM_ARBRE » est la meilleure combinaison avec les meilleurs taux de reconnaissance de la pathologie de l'hypothyroïdie.

Et pour la méthode de DST, nous constatons que : ses valeurs de TC varie entre 96.53% et 99.89%, ses valeurs de SE atteint 99.76% et pour ses valeurs de SP atteint 99.92% et la combinaison « KNN_SVM_ARBRE » donne les meilleurs valeurs.

Lors de l'utilisation de la base de données sans sélection de variables, la combinaison « KNN_SVM_ARBRE » est la meilleure que ce soit avec la méthode du VM ou avec la méthode du DST, elle donne un taux très élevé de classification, de sensibilité et de spécificité.

IV.6.4.2.combinaison hétérogène (BDD avec sélection de variables Relief)

Combinaison Hétérogène	Relief %					
	Vote			DST		
	TC	SE	SP	TC	SE	SP
RNA_ SVM_ ARBRE	95,80	95,76	95,88	97,10	97,04	97,25
RNA_ SVM_ KNN	95,13	95,08	95,39	95,56	95,47	95,69
RNA_ ARBRE_ KNN	95,82	95,85	95,90	97,89	97,75	97,92
KNN_ SVM_ ARBRE	95,70	95,67	95,84	97,67	97,57	97 ,75

Tableau IV.6: Expérimentation 2 : Combinaison hétérogène (BDDavecRelief)

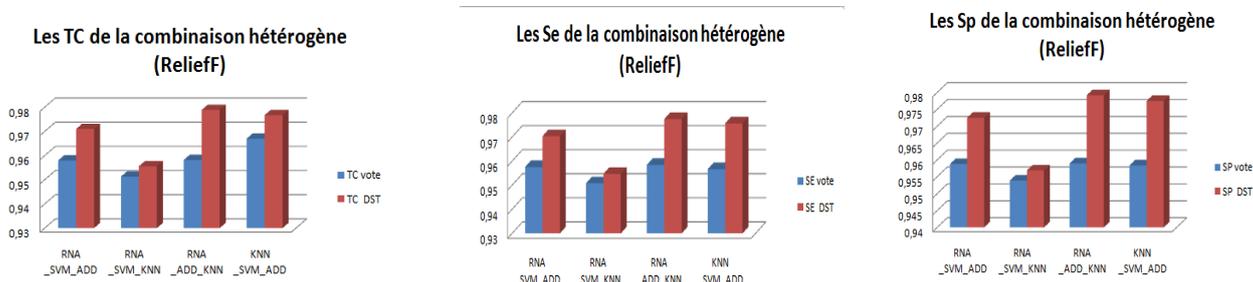


Figure IV.9 : Présentation du TC, SE, SP de l’expérimentation 2 :CombinaisonHétérogène (BDD Relief)

La Figure IV.9 Présente les différents résultats de TC, SE et SP qui permet d’évaluer les performances des deux méthodes de combinaison : VM et DST ces dernières méthodes nous ont permis de fusionner les sorties des trois classifieurs sur la base de données avec sélection de variables Relief.

A partir de ces résultats, nous constatons que le TC de la méthode du VM atteint 95.82% avec un SE qui varie entre 95.08% et 95.85% et une SP de 95.90% et nous remarquons aussi que la meilleure combinaison est celle de « RNA_ ARBRE_ KNN».

Et pour la méthode de DST, nous remarquons que : ses valeurs de TC varie entre 95.56% et 97.89%, ses valeurs de SE atteint 99.75% et pour ses valeurs de SP atteint 99.92% et la combinaison « RNA_ ARBRE_ KNN» donne les meilleurs résultats.

Donc la combinaison « RNA_ ARBRE_ KNN» est la meilleure que ce soit avec la méthode du vote majoritaire ou avec la méthode du DST, elle donne un taux très élevé de sensibilité et de spécificité et elle permet une bonne reconnaissance de l’hypothyroïdie.

IV.6.4.3.combinaison hétérogène (BDD avec sélection de variables Rank)

Combinaison Hétérogène	Rank%					
	VM			DST		
	TC	SE	SP	TC	SE	SP
RNA_ SVM_ ARBRE	97,50	97,45	97,64	99,01	98,99	99,17
RNA_ SVM_ KNN	96,30	96,24	96,47	97,32	97,24	97,48
RNA_ ARBRE_ KNN	97,41	97,36	97,56	99,00	98,97	99,14
KNN_ SVM_ ARBRE	97,28	97,18	97,37	98,98	98,95	99,02

Tableau IV.7: Expérimentation 2 : Combinaison hétérogène (BDD avec Rank)

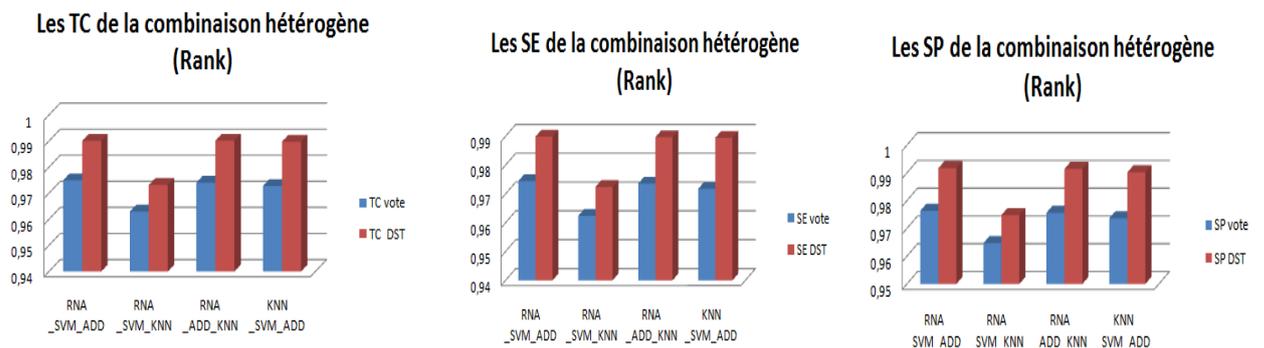


Figure IV.10 : Présentation du TC, SE, SP de l'expérimentation 2 :CombinaisonHétérogène(BDD Rank)

La Figure IV.10Présente les différents critères d'évaluation de taux de classification, de sensibilité et de spécificité qui permet d'évaluer les performances des deux méthodes de combinaison : VM et DST ces dernières méthodes nous ont permis de fusionner les sorties des trois classifieurs sur la base de données avec sélection de variables Rank.

Nous remarquons à partir de ce tableau que le TC de la méthode du VM varie entre 96.30% et 97.50% avec un SE qui varie entre 96.24% et 97.45% et une SP de 97.64%et nous constatons que la meilleure combinaison est « RNA_ SVM_ ARBRE».

Pour la méthode de DST, nous remarquons que : ses valeurs de taux de classification varient entre 97.02% et 99.01%, ses valeurs de sensibilité atteint 98.99% et pour ses valeurs de spécificité atteint 99.17% et la combinaison « RNA_ SVM_ ARBRE» donne les meilleurs valeurs.

Donc la combinaison « RNA_ SVM_ ARBRE» est la meilleure que ce soit avec la méthode du vote majoritaire ou avec la méthode du DST, elle donne un taux très élevé de sensibilité et de spécificité et elle permet une bonne détection de l'hypothyroïdie.

IV.8. Conclusion

L'ensemble des résultats obtenus par les différents classifieurs (RNA, SVM, K-NN et Arbre de décision) ainsi que le vote majoritaire et la DST à travers la sélection de variable et sans sélection ont été convaincants.

Pour l'expérimentation mono classifieur nous constatons que dans le cas de sélection de variable et le cas sans sélection, le classifieur arbre de décision est meilleur, il a effectué une bonne classification de l'hypothyroïdie avec un taux qui atteint 99.96%.

Pour les différentes expérimentations nous trouvons que la méthode de sélection de variables Rank donne des meilleurs résultats que la méthode de sélection Relief et donc les explications et les déclarations des médecins concernant les attributs les plus pertinents ont été bien vérifiées.

Nous remarquons que la méthode du DST donne des meilleurs taux que la méthode du vote majoritaire alors pour savoir pourquoi il y a cet écart nous avons tiré les individus qui représentent cette différence avec leurs descripteurs et leurs classes et nous avons discuté avec le Docteur Meliani qui est un spécialiste en endocrine et avec Docteur Bassaid qui est un médecin résident sur ces individus et leurs avis ont été identiques donc nous avons comparé leurs réponses avec nos réponses et nous avons trouvé les mêmes résultats.

Le vote majoritaire a échoué pour quelques individus parce que parfois les valeurs des attributs sont très proches de la norme par exemple un patient qui a une valeur de TSH égale à 6 et la norme du TSH égale à 5 donc le vote a classé ce patient comme étant non malade et la DST a classé ce patient comme étant malade et dans la réalité c'est un patient malade qui a une légère hypothyroïdie

De même cette situation est expliquée par les médecins que ces individus présentent des cas infra cliniques c'est-à-dire le patient ne souffre pas aux effets de l'hypothyroïdie mais le bilan sanguin montre une légère perturbation.

Et donc nous avons conclu que la DST est la meilleure méthode de combinaison parce que c'est une méthode probabiliste qui repose sur des fondements mathématiques solides, elle permet de combiner les sorties réelles des fonctions de décision contrairement au VM qui combine les étiquettes.

Finalement la DST a amélioré significativement les résultats et elle nous a donné un système robuste fiable qui permet la prédiction de la maladie.

Conclusion & Perspectives

Le développement des systèmes d'aide au diagnostic médical présente aujourd'hui un domaine de recherche très motivant. En effet, beaucoup de chercheurs travaillant dans le domaine d'intelligence artificielle essayent de proposer des systèmes automatiques intelligents capables d'aider les médecins dans leurs routines cliniques.

Dans de nombreuses spécialités, on observe la présence de multiples informations, provenant de différentes sources, qui représentent le même phénomène. Afin d'améliorer la connaissance disponible sur le phénomène observé, il est souvent nécessaire de combiner ou intégrer ces informations, aussi bien de point de vue quantitatif (c'est-à-dire plus d'informations) par l'utilisation de la complémentarité des sources que de point de vue qualitatif (c'est-à-dire de donner une information plus fiable) en exploitant la redondance des sources.

Notre démarche consiste à développer en premier temps des classifieurs qui vont apprendre et classer les individus de la maladie hypothyroïdie puis appliqué deux différents algorithmes de sélection tel que « Relief », et « Rank » pour qu'on puisse avoir les paramètres les plus pertinents de la maladie. Différentes techniques de classification à l'apprentissage supervisé ont été testées dans le prototype proposé, à savoir les machines à vecteur de support, les arbres de décision, les réseaux de neurones, ainsi que les K plus proche voisin.

Comme une deuxième partie dans notre PFE nous avons proposé de fusionner les sorties de ces classifieurs pour améliorer le taux de reconnaissance de la maladie hypothyroïdie, une première contribution basée sur le principe de vote majoritaire pour se rapprocher au maximum de l'avis du médecin puis la seconde consiste à appliquer la fameuse théorie de Dempster-Shafer pour réaliser des combinaisons homogènes et hétérogènes afin d'obtenir des taux de connaissance satisfaisants.

Les expérimentations réalisées avec les différents classifieurs et les différentes techniques de combinaison proposés ont permis d'évaluer les performances ainsi elles ont amélioré significativement les résultats donc le système conçu semble un système robuste plus cohérent.

Les perspectives futures sont en premier lieu faire tester les différents algorithmes de classification sur plusieurs bases de données et faire combiner ces classifieurs par d'autres méthodes de combinaison homogène et hétérogène comme la somme, le maximum, le minimum, la moyenne, le produit, la méthode de Bayes...

Puisque au début de notre PFE on a commencé de travailler sur la technique des machines à vecteurs de support semi supervisé (S3VM) mais on n'a pas trouvé un algorithme clair à implémenter donc on propose comme deuxième perspective de ce travail d'introduire par exemple le classifieur (S3VM) et faire combiner leurs sorties pour obtenir plus de réponses fiables.

Références

- [1] M. Adonkon, "Apprentissage semi-supervisé pour les svm et leurs variantes montreal", thèse de doctorat, l'école de technologie supérieure, 2009.
- [2] S. Mustière, "Apprentissage Supervisé pour la Généralisation Cartographique", thèse de doctorat, université de Paris VI 2001.
- [3] <http://wwwia.lip6.fr/~tollaris/ARTICLES/THESE/node7.html#SECTION02213010000000000000>, date de consultation:12/01/2015 .
- [4] K. Blakrishnan, V.Honavar, "Intelligent diagnosis systems ", Ph.D. thesis, Université Iowa state USA,Journal of Intelligent Systems, 1997.
- [5] A.Wolf, CH .Barbosa, EC .Monteiro, M .Velsaco, "Multiple mlp neural Networks applied on the determination of segment limits in eeg signals ", in International work conference on artificial and natural NN, 2003 June.
- [8] M.A Lazouni, "Aide au Diagnostic Pour Un Médecin Anesthésiste Réanimateur», thèse de doctorat université Abou BekrBelkaid, 2014.
- [10] M.P Nieto, M.A.G.Gonzalo, "Supervised classification exercice",UniversitéPolitecnica de Catalunya, Master in Artificial Intelligence, decembre 2012.
- [13] M. LEZORAY Olivier, "Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones : application à la classification de cellules en cytologie des séreuses ",Doctorat de l'université de CAEN,2000.
- [14] R. Benali, M. A. Chikh, "Reconnaissance des extrasystoles ventriculaires par la combinaison des réseaux de neurones et la logique floue", Université Abou BekrBelkaïd.
- [15] E.M Dupas, "Algorithme des k plus proches voisins pondérés et application en diagnostic", Responsable Plate-forme bioinformatique &biostatistique, Juin 2010.
- [16] F.CHAMROUKHI, "Classification supervisée : Les K-plus proches voisins ", Université du Sud Toulon – Var, 2013.
- [17] J.R Quinlan, "Induction of decision trees", Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007.
- [18] J.R Quinlan, M. Kaufmann, "Programs for Machine Learning",Department of Computer Science, Johns Hopkins University, Baltimore, 1994.
- [19] ,K.Douibi, N.Settouti, "Classification Automatique de l'hypothyroïdie", nbr page=16, Université Abou Bekr Belkaïd,2013.
- [20] A.Djebbar, H.Djellali, H.F Merouani, "Classification des tumeurs du sein par un classifieur Bayésien Naïf et les arbres de décision", Computer Science Department, LRI laboratory, SRF equip, Computer Science Department, LRS Laboratory, ,Badji Mokhtar University.
- [21] Aline CONSEIL, Violaine FERANDEZ, Nicolas FERRAFIAT, and Olivier MONTANGUON. Réseaux bayésiens pour le diagnostic de l'hypothyroidie. Technical report, Master 2 GIS MIS, D2006.
- [22] A.BENAOUDA,"Evaluation de l'activité des paramètres biochimiques chez une population atteinte d'hypothyroïdie", Université Abou BekrBelkaid Tlemcen, 2014.

- [23] N.ABBAS, "SVM Classifier Combination for Handwritten Recognition",Thèse de doctorat,2014.
- [24] W.Barhoumi, E.Zagrouba, B.Solaiman, F.Ghorbel, "FUSION DE L'INFORMATION PAR LA THÉORIE DE L'ÉVIDENCE : Application en Diagnostic du Mélanome".
- [26] M.A Bagheri , Q. Gao, S.Escalera, "Logo recognition based on the Dempster-Shafer Fusion of Multiple Classifiers", Faculty of Computer Science, Dalhousie University, Halifax, Canada, Computer Vision Center, Campus UAB
- [29] SENOUSI Hafidaa, "Sélection de Données pour l'Apprentissage des Réseaux de Neurones, Arbres de Décision et les k-Plus Proches Voisins",Thèse de doctorat, Université des sciences et de la technologie d'Oron mohamed boudiaf,2014.
- [30] N.Settouti, A.Hafa "Approche Filtre pour la sélection des gènes pertinents des données biopuces du Cancer du Côlon", Université de Tlemcen, nbr de page =25, 2013.
- [31] H. Stoppiglia, "Ranking a Random Feature for Variable and Feature Selection",Journal of Machine Learning Research 3,2003.
- [32] J.NOVAKOVIĆ, P.STRBAC, D.BULATOVIĆ, "Toward optimal feature selection using ranking methods and classification algorithm",Faculty of Computer Science,2011.
- [33] S .Haykin, "Neural networks a comprehensive foundation ", in Second ED prentice-hall,Mc Master University Hamilton, Ontario,Canada, 1999.
- [34] Vladimir N .Vapnik, "The nature of statistical learning theory", Second edition, Springer-Verlag New York, Inc, New York, NY, USA, 1995.
- [36] S .Tuffery, "Data mining et statistique decisionnelle l'intelligence dans les bases de donnees", Technip, 2005.
- [37] Fix E and Hodges J L, "Discriminatory analysis. non-parametricdiscrimination",in Randolph Field, TX, 4 Technical Report, 1951.
- [40] L .Breiman, J .Friedman, R .Olshen, and C .Stone, "Cart: Classification and regression trees", in Wadsworth International, 1984.
- [41] J .Ross Quinlan, "Induction of decision trees", Machine Learning, pp. 81–106, 1986.
- [42] M.C FLOREA, "Combinaison d'informations hétérogènes dans le cadre unificateur des ensembles aléatoires : approximations et robustesse",Université Laval,2007.
- [44] S. Chitroub, A. Houacine et B.Sansal , "Combinaison de Classifieurs, Une Approche pour l'Amélioration de la Classification d'Images",nbr page=8.
- [45] F. Kimura ,M. Shridhar, "Handwritten numerical recognition based on multiple algorithms", *Pattern Recognition*, Vol. 24 (10), pp. 969-983, The University of Michigan-Dearborn, Dearborn, Michigan, U.S.A, 1991.
- [46] "Chapitre 2 Sélection de variables",Thèse Université lyon.
- [47] P. Dangauthier, "Sélection automatique de variables Pertinentes ", nbr de page=49, 2007.
- [48] K .Derdour, H .Mouss ,"Combinaison de classifieurs KPPV pour la classification des chiffres imprimé",Article , Université de Batna, Algérie,nombre de page=5, 2014.
- [49] A.MARTIN, "Fusion de classifieurs pour la classification d'images sonar", nbr de page =13,2012.
- [50] A. Elisseeff,I.Guyon, " An introduction to variable and feature selection",*Journal of Machine learning Research* 3, 2003.

- [51] D.Heckerman, D.Geiger, and D.M Chickering, " Learning bayesiannetworks : The combination of knowledge and statistical data ". In KDD Workshop, pages (85,96), 1994.
- [52] D.Koller and M. Sahami, "Toward optimal feature selection ",In International Conference on Machine Learning, Stanford University,pages (284,292), 1996.
- [53] Edwin E. Ghiselli. Theory of Psychological Measurement. McGraw-Hill Book Company, 1964. Pas trouvé
- [54] M. Hall. Correlation-based feature selection for machine learning, for the degree of Doctor, University of Waikato.1999.
- [60] A.Cornuéjols, L.Miclet, Y.Kodratoff, "*Apprentissage Artificiel: Concepts et algorithmes*", Pages : 638,2002.
- [61] O.Chapelle,S.Bernhard, A.Zien, "Semi-Supervised Learning", 2006, The MIT Press,Cambridge, Massachusetts,London, England.
- [62] Zhu,Ghahramani, "Learning fromlabeled and unlabeled data with label propagation", 2002, Carnegie Mellon University.[63] L.Claire, "Aspects fondamentaux et évolutifs de la reconnaissance de formes". 1991.
- [64] Lam, Suen, C.Y, "Application of Majority Voting to Pattern Recognition : An Analysis of Its Behavior and Performance", IEEE Transactions on Systems, Man Cybernetics 1997, 27(5), pp 553-568.)
- [68] N.Abbas, "Développement de modèles de fusion et de classification contextuelle d'images satellitaires par la théorie de l'évidence et la théorie du raisonnement plausible et paradoxal", Université des sciences et de la technologie Houari Boumediene, 2009.
- [69] DEMPSTER, A.P, "A generalisation of Bayesian Inference". Journal of the Royal Statistical Society, 30: pp. 205-247, 1968.
- [70] DENOEU, T, "A k-nearest neighbour classification rule based on Dempster- Shafer Theory". IEEE Transactions on Systems, Man and Cybernetics, Vol. 25, N° 5, pp. 805-813, may 1995.
- [71] SMETS, PH. AND KENNES, R., "The transferable belief model", Artificial Intelligence 66, pp.191-234, 1994.
- [72] A. APPRIOU, "Probabilités et incertitude en fusion de données multisenseurs", Revue Scientifique et Technique de la Défense. Vol. 11, pp. 27-40, 1991.
- [73] A.DROMIGNY-BADIN, "Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles", Ecole Doctorale des Sciences de l'Ingénieur de Lyon ,1998.
- [74] F. Faux, F. Luthon, "Modélisation de visage par fusion d'information couleur dans le cadre de la théorie de l'évidence et suivi par filtrage particulière",Laboratoire LIUPPA.
- [75] Sophie-Gariépy Major D.O, "Troubles thyroïdiens", 15 pages, 2007.

Résumé

Les systèmes d'aide au diagnostic ont su des expansions très frappant ainsi ils ont mobilisé durant ces dernières années une large communauté de chercheurs, leurs but principal est de concevoir un système intelligent informatique qui permet la précaution ou l'identification des maladies d'une manière automatique.

Le système conçu dans ce travail vise à aider les médecins dans leurs routines cliniques plus exactement dans le diagnostic de l'hypothyroïdie, nous avons choisi cette pathologie parce que c'est une maladie très fréquente touche surtout les femmes et elle propage rapidement dans l'environnement, ses effets sont durable comme elle peut causer des complications assez grave.

L'approche proposée dans ce mémoire est basée sur le test de quatre différentes techniques de classification indépendamment sur la base de données d'hypothyroïdie a savoir les réseaux de neurones artificiel (RNA), les machines à vecteurs de supports (SVM), et les K-plus proches voisins (KNN) et les arbres de décision (ADD) par la suite effectuant une combinaison homogène et hétérogène de ces classifieurs par la méthode de Dempster-Shafer (DST)cette dernière elle base sur un fondement théorique solide destiner pour combiner toutes source d'information afin d'obtenir une décision plus précise.

Les résultats obtenus sont satisfaisants et prometteurs, ils montrent que l'approche de combinaison donne des réponses plus fiable ainsi ils montrent la cohérence et la robustesse de notre système proposé.

Mots clés : Hypothyroïdie, classification, RNA, SVM, KNN, ADD, Sélection, Relief, Rank, Vote majoritaire, DST

Summary:

The diagnostic support system expansions knew very striking and they have mobilized in recent years a large community of researchers, their main goal is to develop a computer system that allows smart precaution or identification of a disease automatically.

The system designed in this work is to help physicians in their clinical routines, more precisely in the diagnosis of hypothyroidism, we chose this disease because it is a very common disease mainly affects women and is spreading rapidly in the environment; its effects are sustainable as it can cause quite serious complications.

The approach proposed in this paper is based on the test four different classification techniques independently on hypothyroidism database namely artificial neural networks (RNA), support vectors machines (SVM), and K-nearest neighbor (KNN) and decision trees (ADD) subsequently performing a homogeneous and heterogeneous combination of these classifiers by the method of Dempster-Shafer (DST) it is based on a solid theoretical basis for combining destine All information source to obtain a decision more.

The results are satisfactory and promising, they show that the combination approach provides more reliable answers and they show the consistency and robustness of our proposed system.

Keywords: Hypothyroidism, classification, RNA, SVM, KNN, ADD, Selection, Relief, Rank, Majority Voting, DST

المخلص

انظمة دعم التشخيص عرفت تطورا ملفتا للغاية لقد حشدت في السنوات الأخيرة مجموعة كبيرة من الباحثين حيث هدفهم الرئيسي هو تطوير نظام حاسوبي ذكي يمكنه التعرف على المرض تلقائيا.

النظام المصمم في هذا العمل يسمح بمساعدة الأطباء في الإجراءات السريرية، وبالتحديد في تشخيص قصور الغدة الدرقية، اخترنا هذا المرض لأنه مرض شائع جدا يصيب بشكل رئيسي النساء وينتشر بسرعة في البيئة، وأثاره مستدامة لأنها يمكن أن تسبب مضاعفات خطيرة للغاية.

ويستند النهج المقترح في هذه الورقة على اختبار اربع تقنيات تصنيف مختلفة بشكل مستقل على قاعدة بيانات الغدة الدرقية: الشبكات العصبية الاصطناعية، خوارزمية الأقرب جار، شجرة القرارات، آلات دعم الناقلات ثم بعد ذلك تنفيذ مزيج متجانس وغير متجانس من هذه المصنفات عن طريق تقنية ديمبستر شيفر التي تقوم على أساس نظرية متينة للجمع بين خصص كل مصدر معلومات للحصول على قرار أكثر دقة النتائج كانت مرضية واعدة، تدل على أن طريقة الجمع توفر أجوبة دقيقة موثوق فيها وأنها تظهر اتساق ومتانة نظامنا المقترح

الكلمات المفتاحية: الغدة الدرقية، التصنيف، الشبكات العصبية الاصطناعية، خوارزمية الأقرب جار، شجرة القرارات، آلات دعم الناقلات، اختيار، التصويت بالأغلبية، ديمبستر شيفر