

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOHAMED BOUDIAF - M'SILA

FACULTÉ : TECHNOLOGIE
DEPARTEMENT : ELECTRONIQUE
N° : 2019/ ESEM /



DOMAINE : Sciences et Technologies
FILIERE : Électronique
OPTION : Électronique des Systèmes
Embarqués

Mémoire présenté pour l'obtention du diplôme de Master Académique

Par: ABASSI Brahim

Intitulé

**APPROCHE DE SÉLECTION DES DONNÉES
BIOMÉDICALES POUR L'IDENTIFICATION
DE PATHOLOGIES EN UTILISANT LES
SUPPORTS VECTORS MACHINES**

Soutenu publiquement le 26 juin 2019 devant un jury composé de:

Mr. DJERIOUI Mohamed	Université de M'sila	Président
Mr. LADJAL Mohamed	Université de M'sila	Encadreur
Mr. BRIK Youcef	Université de M'sila	Co-Encadreur
Mr. MEZACHE Hatem	Université de M'sila	Examineur
Mr. DJAIDJA Youcef	Médecin général	Invité

Année universitaire : 2018 /2019

Remerciements

*Tout d'abord, Nous remercions **ALLAH** miséricorde de nous avoir donné la volonté, le courage et la patience qui nous ont permis de réaliser ce travail.*

*Nous tenons à exprimer nos vifs remerciements a nos encadreurs Mr. **LADJAL Mohamed** et Mr. **BRIK Youcef** d'avoir proposé ce sujet et pour avoir dirigé notre travail durant la période d'élaboration de ce mémoire.*

Nous exprimons notre profonde gratitude pour nous avoir fait profiter de leurs connaissances, mais aussi de leurs méthodes de travail, et surtout de leur rigueur scientifique.

Nos remerciements s'adressent également à tous les membres de Jury, qui ont accepté de nous honorer, de leur présence et de juger notre travail.

Nos remerciements vont aussi à tous les enseignants du département d'électronique sans oublier toutes les personnes de ce même département.

Enfin, nous remercions A tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

Dédicace

Je Dédie ce travail

A

*Les deux êtres les plus chères au monde
les bougies qui m'ont toujours guidé vers le bon chemin :
Ma mère, la lumière de ma vie, qui a tout fait pour
ma réussite et mon bonheur.*

*Mon père, à lui, je dois offrir tout le respect et
l'amour pour son soutien et sa tendresse.*

A mes chers frères et sœurs.

A Toute la famille.

A tous mes enseignants.

A Tous mes amis.

ABASSI Brahim

Table des matières

LISTE DES FIGURES	iv
LISTE DES TABLEAUX	v
ABREVIATION ET NOTATION	vi
INTRODUCTION GENERALE	1

CHAPITRE I La biomédicale et l'apprentissage des machines

I.1. INTRODUCTION	03
I.2. BIOMEDICAL ET L'ELECTRONIQUE	03
I.2.1 Electromyogramme	03
I.2.2 Electrocardiogramme	04
I.2.3 Maladies cardiaques (Heart Diseases)	05
I.2.4 Dépistage du cancer	05
I.3 INTELLIGENCE ARTIFICIELLE	06
I.3.1 Historique	06
I.3.2 Multiples applications de l'intelligence artificielle	07
I.3.3 Futur de l'intelligence artificielle	10
I.4. APPRENTISSAGE DES MACHINES ET SES APPLICATIONS A LA BIOMEDICALE	10
I.5. CATEGORIES DES MACHINES D'APPRENTISSAGE	10
I.5.1. Machines par séparation	10
I.5.1.1. Machines à Vecteurs de Support (SVMs)	10
I.5.1.2. Réseaux de neurones artificiels (RNAs)	11
I.5.2. Machines par modélisation	12
I.5.2.1. Modèle de mélange gaussiens	12
I.5.2.2. Chaines de Markov cachées	13
I.6. TYPES D'APPRENTISSAGE	14
I.7. CONCLUSION	15

CHAPITRE II Système proposé

II.1. INTRODUCTION	16
II.2. ARCHITECTURE GLOBALE DU SYSTEME DE RECONNAISSANCE PROPOSE.....	16
II.3. SELECTION DES CARACTERISTIQUES s	17
II.3.1. Définition	18
II.3.2. Schéma général de la sélection des caractéristiques	18
II.3.3. Neighborhood Component Analysis (NCA).....	19
II.4. CLASSIFICATION DE DONNEES	21
II.4.1. Machines à Vecteurs de Supports	23
II.4.1.1. Principe des SVMs	23
II.4.2. Variantes de SVM	29
II.4.2.1. Twin Support Vector Machine	31
II .5. CONCLUSION	34

CHAPITRE III Résultats expérimentaux

III.1. INTRODUCTION	35
III.2. BASE DE DONNÉES UTILISÉE	35
III.2.1. Heart Disease	35
III.2.2. Brest cancer	36
III.2.3. Ovarien cancer	37
III.3. PROTOCOLE EXPÉRIMENTAL	38
III.3.1. Taux de classification	38
III.3.2. Taux d'erreur	38
III.4. RÉSULTATS DE SIMULATION	40
III.4.1. Sans sélection des caractéristiques	40
III.4.1.1. Avec la base Heart Disease	40
III.4.1.2. Avec la base Breast Cancer	42
III.4.1.3. Avec la base Ovarian Cancer	43
III.4.2. Avec sélection des caractéristiques	44

III.4.2.1. Avec la base Heart Disease	44
III.4.2.2. Avec la base Breast cancer	47
III.4.2.3. Avec la base Ovarian Cancer	51
III.5. CONCLUSION	55
CONCLUSION GENERALE.....	56
REFERENCES.....	57

LISTE DES FIGURES

Fig. I.1. la forme d'un signal EMG	04
Fig. I.2. Signal ECG	04
Fig. I.3. L'architecture d'un réseau de neurones de type MLP	12
Fig. I.4. Le principe de GMM	13
Fig. I.5. Le principe de MMC	13
Fig. I.6. Schéma d'un modèle supervisé	14
Fig. I.7. Schéma d'un modèle non-supervisé	14
Fig. I.8. Schéma d'un modèle semi-supervisé	15
Fig. II.1. Le système de reconnaissance de formes.	16
Fig. II.2. Principe de la sélection de caractéristiques	17
Fig. II.3. Principe de l'extraction de caractéristiques	18
Fig. II.4. Processus de sélection d'attributs	19
Fig. II.5. Taxonomie des méthodes utilisées pour la reconnaissance	22
Fig. II.6. Apprentissage supervisé d'une machine.	23
Fig. II.7. Hyperplan de séparation linéaire pour des données linéairement séparables (Les vecteurs de support sont encerclés).....	24
Fig. II.8. Exemple d'une application rendant les données linéairement séparables.....	26
Fig. II.9. Schéma synoptique de l'implémentation un contre tous.	28
Fig. II.10. Extension of Twin Support Vector Machine.	30
Fig. II.11. Classification binaire à l'aide de TWSVM.	31
Fig. III.1. Taux de classification illustré dans la matrice de confusion.	38
Fig. III.2. Taux d'erreur illustré dans la matrice de confusion.	39
Fig. III.3. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-Laplacien sur la base Heart disease.	41
Fig. III.4. Sorties réelle et calculée pour l'apprentissage et le test avec TWIN-SVM sur la base Heart disease.....	41
Fig. III.5. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-Linear sur la base Breast Cancer.....	42
Fig. III.6. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-RBF sur la base Ovarian Cancer.....	44
Fig. III.7. Application de NCA sur les données: a) Ordre avant NCA et b) Ordre après NCA	44

Fig.III.8. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Heart Disease utilisant les différentes variantes de SVM.....	47
Fig. III.9. Application de NCA sur les données de la base Breast Cancer: a) Ordre avant NCA et b) Ordre après NCA	48
Fig.III.10. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Breast Cancer utilisant les différentes variantes de SVM.....	51
Fig. III.11. Application de NCA sur les données de la base Ovarian Cancer: a) Ordre avant NCA et b) Ordre après NCA	51
Fig.III.12. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Ovarian Cancer en utilisant les différentes variantes de SVM.	54

LISTE DES TABLEAUX

Tableau III.1. Informations d'attributs de la base Heart Disease.	36
Tableau III.2. Informations d'attributs de la base Brest Cancer.....	36
Tableau III.3. Résultats obtenu de la base Heart disease avec les différentes variantes de SVM.	40
Tableau III.4. Résultats obtenu de la base Breast Cancer avec les différentes variantes de SVM.....	42
Tableau III.5. Résultats obtenu de la base Ovarian Cancer avec les différentes variantes de SVM.....	43
Tableau III.6. Ordre des caractéristiques avant et après l'application de NCA sur la base Heart Disease	45
Tableau III.7. Meilleure performance obtenu avec la sélection des caractéristiques de la base Heart Disease.....	45
Tableau III.8. Ordre des caractéristiques avant et après l'application de NCA sur la base Breast Cancer	48
Tableau III.9. Meilleure performance obtenu avec la sélection des caractéristiques de la base Breast Cancer.....	49
Tableau III.10. Meilleure performance obtenu avec la sélection des caractéristiques de la base Ovarian Cancer	52

LISTE DES ABREVIATIONS

EMG : Electromyogramme
ECG : Electrocardiogramme
NHLBI : National Heart, Lung, and Blood Institute
GPS : genaral problem solver
IA : Intelligence Artificielle
FAIR : Facebook Artificial Intelligence Research
SVM : Support Vector Machines
RNA : Réseaux de Neurones Artificiels
MLP : Multi Layer Perceptron
MMG : Modèle de Mélange Gaussien
MMC : Modèles de Markov Cachée
NCA : Neighborhood Component Analysis
ML : Machine Learning
VC : Vapnik Chervonenkis
MRS : Minimisation du Risque Structurel
KKT : Karush Kuhn Tucker
SMO : Sequential Minimal Optimisation
PQ : problème d'optimisation
OAo : one against one
OAA : one against all
RBF : Radial Basis Function
SRM : Minimisation des Risques Structurels
QPP : problème de programmation quadratique
T : Ensemble d'échantillons d'entraînement
 \mathbf{x}_i : Vecteur caractéristique de dimension d
 y_i : Étiquette de classe
N : Nombre d'échantillons
W : vecteur de pondération des caractéristiques

Introduction générale

INTRODUCTION GENERALE

En médecine, l'un des principaux enjeux consiste à prendre un prélèvement ou un signal vitale de patient à travers un capteur spécifique ou via une analyse physico-chimique pour les étudier, les traiter et en dégager les informations pertinentes pour identifier la pathologie en lien avec la conduite à accomplir. À cause de coût très élevé des ces traitements et aussi le nombre de patients qui augmente jour après jour, la nécessité d'automatiser ce traitement ou cette conduite est devenu indispensable [1]. Afin de remédier à ce besoin, l'intelligence artificielle présente une voix technologique capable de remplacer le professionnel de la santé par une machine. En effet, elle peut servir comme un outil de base pour l'aide à la décision.

Nous pouvons supposer que le problème de la reconnaissance de pathologies est vu comme un problème de reconnaissance de formes, où les formes représentent l'ensemble des observations ou du prélèvement lié aux patients. Actuellement, la reconnaissance de formes a été largement étudiée au cours des dernières années dans la littérature scientifiques. Cependant, malgré toutes les recherches intensives, les techniques de reconnaissance disponibles dans ce domaine restent au-dessous des capacités visuelles humaines.

Afin de mettre en service la reconnaissance de formes, le machine learning (ML) peut s'intervenir pour séparer les forme des une des autres. Plusieurs techniques ont été développées pour optimiser les performances du système biomédical comme la reconnaissance des tumeurs, la segmentation des images médicales, la classification des cellules et l'identification des pathologies. Malheureusement, dans les systèmes qui existes dans la littérature, nous trouvons des techniques performantes mais un peut long dans leur exécution ou rapides mais qui ne présentent pas de résultats assez satisfaisants.

Dance ce mémoire, nous présentons un système de reconnaissance de pathologie à partir de données biomédicale. Ce système s'articule sur deux modules principaux. Le premier est un classifieur de type Support Vector Machines (SVMs). Ces machines se

démarquent des autres outils par leur capacité d'apprentissage et de généralisation. Leur rôle réside au niveau de décision en séparant les données en deux classes.

Le deuxième module est une technique de sélection des caractéristiques qui s'appelle Neighborhood Components Analysis (NCA). Cette technique permet de réduire la dimension de données pour minimiser le coût du traitement en terme de temps, espace mémoire et nombre de prélèvement médical tout en augmentant la performance de la reconnaissance.

Afin de mener ce travail à terme, notre mémoire est structuré en trois chapitres comme suit:

Le premier chapitre présente une généralité sur La biomédicale et sa relation avec le machine learning.

Dans le deuxième chapitre, nous allons aborder les différents module de notre système de reconnaissance de pathologie notamment la partie de classification avec les SVMs et la partie de sélection de caractéristique en utilisant le NCA.

Le troisième chapitre présente et discute les résultats obtenus sur trois différentes base de données réelles..

Finalement, nous rappelons les objectifs du travail, les résultats obtenus aussi les perspectives pour améliorer les performances du système de reconnaissance de pathologies.

Chapitre I
**Biomédicale et l'apprentissage
des machines**

I.1. INTRODUCTION

Les termes « biomédecine » et « biomédical » ont des origines et des trajectoires différentes en anglais, en allemand «Biomedizin» et en français « biomédecine », mais leur sens est très similaire aujourd'hui. «Biomédical» est apparu pour la première fois dans les écrits d'auteurs américains et britanniques dans les années 1920, suivi dix ans plus tard par la «biomédecine». Le dictionnaire médical américain (1923) le définit comme «une médecine clinique fondée sur les principes de la physiologie et de la biochimie». Ainsi, dès le début, la biomédecine et la recherche biomédicale ont été comprises comme une sorte de médicament étroitement associé à l'expérimentation et au laboratoire plutôt qu'au savoir du médecin et de la clinique.

I.2. BIOMEDICAL ET L'ELECTRONIQUE

Dans nos jours, l'étude des signaux biomédicaux et l'extraction de l'information utile d'une façon automatique sont très importantes et aident les médecins à bien comprendre le comportement de différents tissus cellulaires afin de faire un bon diagnostic et à construire les bases de données de patients. Dans la suite, nous citons les signaux biomédicaux les plus utilisés dans le traitement automatiques des patients à l'aide de l'apprentissage des machines:

I.2.1 Electromyogramme

L'électromyogramme (EMG) est enregistré en utilisant un appareil d'électromyographie, qui mesure le potentiel électrique du muscle. Le système nerveux central (le cerveau, la moelle épinière et les nerfs périphériques) contrôle l'action des fibres musculaires qui se traduit en général aux mouvements. Le muscle se compose de cellules spéciales qui sont capables de faire des dilatations et des relaxations et qui sont contrôlées par des neurones [1].

Un EMG peut être enregistré par deux méthodes, EMG de surface (en mettant des électrodes sur la surface de la peau) qui est plus populaire que la deuxième appelée EMG intramusculaire (où une électrode à aiguille est insérée dans la peau) parce qu'elle est non invasive.

Un EMG de surface mesure un seul ou plusieurs unités motrices, il est appelé potentiel d'action de l'unité motrice (PAUM). Le potentiel réel voisine les 100 mV mais à cause des couches de tissus conjonctives et de la peau, l'EMG est un signal complexe avec une amplitude beaucoup plus petite (environ 5 mV) (Figure I.1).

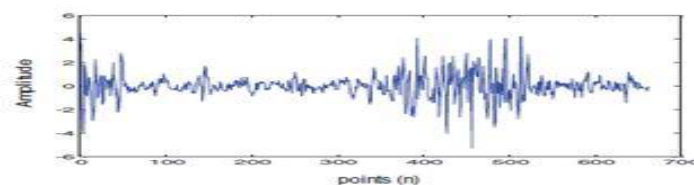


Fig. I.1. la forme d'un signal EMG.

I.2.2 Electrocardiogramme

L'électrocardiogramme (ECG) est la représentation de l'activité électrique du cœur, ce rythme cardiaque est contrôlé par la cellule stimulatrice connu sous le nom du nœud sinusal (SA).

Les ondes PQRST et parfois U constituent un cycle ECG (Figure I.2) : l'onde P est aperçue lorsque le nœud SA déclenche l'impulsion qui se propage dans les oreillettes et crée la contraction auriculaire ; l'intervalle PQ (où segment isométrique) est la durée de propagation de l'impulsion de l'oreillette vers les ventricules, permettant au flux sanguin de s'écouler dans la même direction; le complexe QRS se produit quand l'impulsion arrive aux ventricules et traduit la contraction ventriculaire.

Le segment ST est la période de la dépolarisation des ventricules, dans le cas normal, il doit être isoélectrique, sinon, le niveau d'amplitude et la pente de ce segment sont des indicateurs de l'état ischémique du myocarde et enfin l'onde T qui exprime la repolarisation ventriculaire, c'est l'état de repos des ventricules (relaxation) [1, 2].

Les impulsions électriques sont enregistrées par le biais des électrodes placées sur des surfaces particulières du corps, en utilisant un ECG 12-dérivations ou 12-canal (dans les hôpitaux) ou ECG 3-canal.

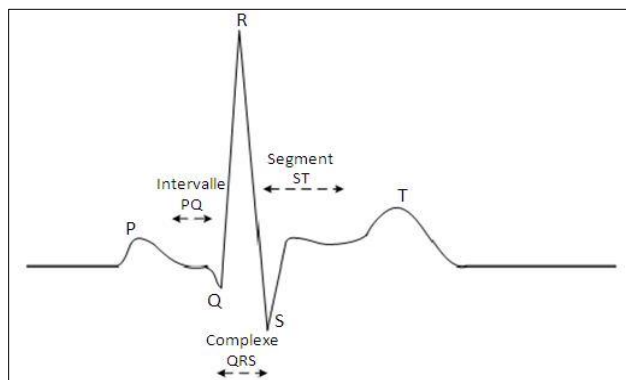


Fig. I.2. Signal ECG.

I.2.3 Maladies cardiaques (Heart Diseases)

Le nombre de personnes souffrant de maladies cardiaques est en augmentation. Un diagnostic précis à un stade précoce suivi d'un traitement ultérieur approprié peut permettre de sauver un nombre des vies considérable. De nouvelles données publiées par l'Institut national du cœur, des poumons et du sang (NHLBI) montrent que les femmes plus âgées, en particulier, sont plus à risque de contracter une maladie cardiaque.

Une étude récente sur le terrain d'une maladie cardiaque peut être contrôlée efficacement si elle est diagnostiquée à un stade précoce [3]. Mais ce n'est pas facile de faire un diagnostic précis en raison de nombreux facteurs complexes de maladies cardiaques. Par exemple, de nombreux symptômes cliniques sont associés à de nombreux organes humains autres que le cœur et très souvent, les maladies cardiaques peuvent présenter divers syndromes.

En raison de cette complexité, il est nécessaire d'automatiser le processus de diagnostic médical qui peut aider les médecins dans le processus de diagnostic.

I.2.4 Dépistage du cancer

Le dépistage consiste en une série d'exams médicaux, répétés à intervalles réguliers, qui permettent de découvrir la présence éventuelle d'une maladie bien avant qu'elle ne commence à causer des anomalies perceptibles (symptômes). Seuls certains cancers sont actuellement concernés par le dépistage. Le dépistage permet de détecter certains cancers (ou certaines anomalies précancéreuses) longtemps avant l'apparition des premiers symptômes. Son objectif est clair :

- éviter un cancer grâce au traitement de lésions précancéreuses.
- augmenter les chances de guérison (ou permettre un traitement nettement moins agressif) grâce à une prise en charge précoce de la maladie.

Le dépistage n'est actuellement possible que pour une liste limitée de cancers. D'autres dépistages ne font pas l'objet d'un consensus scientifique, même si certains patients y ont recours après discussion avec leur médecin traitant.

I.3 INTELLIGENCE ARTIFICIELLE

C'est une discipline scientifique relative au traitement des connaissances et au raisonnement dans le but de permettre à une machine d'exécuter des fonctions normalement associées à l'être humain. L'intelligence artificielle tente de reproduire les processus cognitifs humains dans le but de réaliser des actions « intelligente ». Elle est comme « la construction des programmes informatique 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 perceptuel.
- L'organisation de la mémoire et le raisonnement critique.

La norme ISO 2382-28 définit l'intelligence artificielle comme la « capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage ». Qualifiée de prochaine révolution informatique, l'intelligence artificielle est au cœur de tous les sujets d'actualités, il semble indispensable de définir cette technologie de rupture et de clarifier son régime juridique, mais aussi d'identifier les applications en cours ou en développement dans les entreprises et les bénéfices qu'elles en tirent.

I.3.1 Historique

Il est intéressant de revenir sur les origines et l'historique de l'intelligence artificielle afin de bien comprendre ses premières orientations et ses perspectives pour l'avenir [4] :

- 1950 : Alan M. Turing, mathématicien et théoricien précurseur de l'informatique, lance le concept d'intelligence artificielle.
- 1955-1956: Lancement du premier programme d'intelligence artificielle par Allen Newell, John C. Shaw et Herbert A. Simon, Logic Theorist.
- 1957: Modélisation des jeux d'échec.
- 1958: John McCarthy invente le Lisp (list processing), langage de programmation interactif (développement au MIT).
- 1958: Construction du premier réseau neuronal, le Perceptron, de Frank Rosenblatt, machine dite connexionniste.

- 1959: Élaboration du premier GPS (general problem solver) -fin de la première période de l'intelligence artificielle.
- 1970: Néoconnectionnisme.
- 1989: DeepThought, supercalculateur d'IBM, deux millions de coups par seconde.
- 1990 - 1997: Développement de Deep Blue rebaptisé DeeperBlue: conception d'un système de 256 processeurs fonctionnant en parallèle, chaque processeur peut calculer environ trois millions de coups par seconde.
- 2009: Le MIT a lancé un projet visant à repenser la recherche en intelligence artificielle.
- 2011: Watson, le superordinateur d'IBM remporte deux des trois manches du jeu télévisé Jeopardy! La performance a consisté pour cette intelligence artificielle à répondre à des questions de culture générale.
- 2013: Humain Brain Project. Google ouvre un laboratoire de recherches dans les locaux de la NASA.
- 2014: DeepKnowledge Ventures : nomme à son conseil d'administration VITAL, un algorithme capable d'élaborer ses décisions en analysant les bilans comptables des entreprises potentiellement intéressantes, les tests cliniques, la propriété intellectuelle et les précédents investissements.
- 2015: Facebook Artificial Intelligence Research (FAIR). Google rend sa technologie d'intelligence artificielle TensorFlow accessible à tous. Développement d'une crainte que l'intelligence artificielle dépasse à terme les performances de l'intelligence humaine.
- 2016: Amelia d'IPSoft un agent virtuel. Aussi, AlphaGo bat trois fois consécutives le champion du monde de jeu de go, Lee Se-Dol en cinq manches.

I.3.2 Multiples applications de l'intelligence artificielle

L'intelligence artificielle a déjà trouvé de multiples usages au sein de la société :

- **Les moteurs de recherche** : Tous les moteurs de recherche (y compris des agences de voyage) sont basés sur des systèmes intelligents d'extraction, d'analyse, et de classification de données pour produire le plus rapidement possible un résultat pertinent à la requête de l'utilisateur. C'est ainsi que Google a implémenté un système utilisant les techniques de machine learning pour son moteur de recherche en octobre 2015, intitulé RankBrain9. Ce système convertit de grandes quantités de texte en vecteurs

mathématiques pour aider le système à deviner le sens des mots ou des phrases qu'il ne connaît pas et ainsi traiter les 15 % de requêtes jamais effectuées auparavant qu'il reçoit chaque jour.

- **Les moteurs de recommandation** : En s'appuyant sur les données issues de la navigation et des achats d'un utilisateur, des sites comme Amazon ou Netflix sont capables de lui proposer d'autres produits similaires qui pourraient l'intéresser. Ces technologies prédictives sont aussi utilisées pour les plateformes de publicité en ligne (Google, Criteo) pour proposer aux visiteurs des contenus d'annonceurs en rapport avec les pages qu'ils ont visité.
- **La traduction automatique** : Elle s'appuie sur des algorithmes de modélisation statistique du langage naturel. Ils intègrent les règles de construction de chaque langue.
- **Les assistants personnels (Siri, Cortana, Google Now...)** : Ils sont déployés sur les Smartphones qui s'appuient sur plusieurs briques technologiques: la reconnaissance vocale pour convertir le son en texte, le langage naturel pour comprendre le sens des mots, un moteur de recherche pour trouver réponse à la question et la synthèse vocale pour communiquer la réponse à l'utilisateur, la planification pour la gestion d'événements, etc.
- **Les agents conversationnels** : Ils sont utilisés dans les domaines du support client et du télémarketing et consistent en des fenêtres de chat qui s'ouvrent toutes seules sur un site web, ou de serveur vocal qui répond aux questions 24h sur 24. Ils utilisent le langage naturel et leur accès à de vastes bases de données leur permet de répondre aux questions les plus simples.
- **Les véhicules autonomes** : Si certains prototypes roulent déjà sur les routes au contact des autres véhicules, les voitures qui se garent toutes seules ou qui freinent par anticipation sont déjà une réalité. Le pilotage automatique des avions ou la gestion de trajectoire des véhicules spatiaux, ou encore les drones se basent aussi sur l'intelligence artificielle.
- **Les systèmes de navigation GPS** : Développé en 1968 par l'Institut de recherche de Stanford, cet algorithme permet d'optimiser le cheminement entre plusieurs points dans un réseau en se basant sur le coût du trajet ou la distance parcourue.

- **Les finances** : Elles sont gérées par des systèmes intelligents pour organiser leurs opérations, investir en bourse et gérer leurs biens, mais aussi pour repérer des transactions qui sortent de l'ordinaire. Les banques possèdent aussi des systèmes experts d'évaluation de risques liés à l'octroi de crédits (credit-scoring).
- **Le cyber sécurité** : Les acteurs du cyber sécurité ont adopté les techniques du machine learning afin de détecter des comportements anormaux dans les systèmes d'information, et de déceler les menaces persistantes pour éviter des opérations d'espionnage ou d'extraction de données¹⁰. Près de 300 paramètres (heures et IP des connexions et des machines, téléchargements, etc.) sont pris en compte pour établir le modèle d'analyse comportementale dont la première phase d'apprentissage dure environ une semaine. Citons la jeune pousse lyonnaise Sentryo qui intègre des algorithmes de machine learning pour sécuriser les sites industriels critiques.
- **Les jeux vidéo** : Ils emploient des techniques d'intelligence artificielle pour donner vie aux personnages non joueurs ou encore pour créer des univers entiers à partir d'algorithmes. En 1997, DeepBlue, le superordinateur d'IBM avait battu Garry Kasparov, champion du monde d'échecs en titre. En 2016, c'est DeepMind, le programme d'intelligence artificielle de Google qui a annoncé la victoire de son programme AlphaGo contre le champion d'Europe de go en titre, Fan Hui¹³. Ce résultat prend appui sur la technologie des réseaux neuronaux, que nous décrivons précédemment. Deepmind avait déjà développé un système d'intelligence artificielle capable de déterminer l'action la plus judicieuse pour battre l'homme dans une vingtaine de jeux d'arcade [5].
- **La médecine** : L'ordinateur Watson d'IBM vise à compiler la plus large base de données dédiée à la santé dans le monde, portant sur 300 millions de patients. La branche santé d'IBM a acquis quatre sociétés médicales depuis sa création : Phytel (santé & population), Explorys (fichiers santé de cliniques), Merge Healthcare (imagerie médicale) et Truven (analytics médicales). Ce supercalculateur a notamment fait ses preuves dans un service d'oncologie pour des services personnalisés : analysant ADN, dossier du patient et autres publications et essais cliniques. L'ordinateur propose un protocole adapté au patient. L'entreprise américaine Enlitic propose des technologies d'intelligence artificielle pour analyser l'imagerie médicale, qui permettraient de déceler des fractures de façon plus performante que les radiologues d'après les dires de la société.

I.3.3 Futur de l'intelligence artificielle

Nous venons de voir que l'intelligence artificielle est déjà une réalité pour de multiples applications. Mais au vu des recherches qui sont menées dans le monde entier, elle a encore de beaux jours devant elle. Les applications de l'intelligence artificielle pour le futur peuvent aller encore plus loin notamment dans le domaine de : voitures autonomes, robotique, bâtiments connectés et la médecine.

I.4. APPRENTISSAGE DES MACHINES ET SES APPLICATIONS A LA BIOMEDICALE

Les chercheurs en intelligence artificielle visent toujours à programmer des machines capables d'effectuer des tâches qui requièrent de l'intelligence. Cependant, programmer des machines capables de s'adapter à toutes les situations et éventuellement d'évoluer en fonction de nouvelles contraintes est difficile [6]. L'enjeu est de contourner cette difficulté en dotant la machine de capacités d'apprentissage lui permettant de tirer profit de son expérience. C'est pourquoi parallèlement aux recherches sur le raisonnement automatique se sont développées des recherches sur l'apprentissage par les machines en anglais « machine Learning ». Le principal objectif de ses recherches est la résolution automatique des problèmes complexes par la prise de décision sur la base d'observations de ces problèmes [6].

L'utilisation de l'apprentissage automatique pour les applications biomédicales connaît une augmentation considérable [7]. Ce regain d'intérêt a plusieurs causes. D'une part, l'application réussie des techniques d'apprentissage automatique dans différents domaines D'autre part, le développement le plus récent est l'avènement des dossiers médicaux électroniques.

I.5. CATEGORIES DES MACHINES D'APPRENTISSAGE

I.5.1. Machines par séparation

I.5.1.1. Machines à Vecteurs de Support (SVMs)

Elles ont été inventées par Boser, Guyon et Vapnik [8, 9] et présentées pour la première fois dans la conférence Computational Learning Theory (COLT) de 1992. Les SVM utilisent une approche géométrique pour classer les données en deux catégories. En modélant les données

comme des points (vecteurs) dans l'espace, elles construisent un plan qui sépare les données dans chacune des catégories. Une fois la frontière de décision construite, la SVM sera capable de classer de nouvelles données en observant de quel côté de la frontière elles tombent, et en leur assignant la catégorie correspondante. Finalement, l'utilisation des SVM peut facilement être étendue à la classification de plus de deux catégories ainsi qu'à la prévision de valeurs continues.

Les machines à support de vecteurs ont été appliquées avec succès dans divers domaines comme la vérification et la reconnaissance, telle que l'identification de visages, la reconnaissance de caractères manuscrits et des chiffres, la vérification et reconnaissance du discours et du parlant et la prédiction. Les SVM constituent une classe spécifique d'algorithmes qui est caractérisée par, l'utilisation de kernels (fonctions noyaux), la capacité d'arriver à un résultat optimal global, la faible densité de la solution et sa capacité de contrôle en agissant sur la marge ou sur une autre quantité indépendante de la dimension comme le nombre de vecteurs de support.

Ces caractéristiques font en sorte qu'elle soit une des techniques dont l'application peut être très adéquate pour la classification des données biomédicales, comme nous verrons dans le chapitre 2 où on va expliquer les SVM en détails.

I.5.1.2. Réseaux de neurones artificiels (RNAs)

Les premiers modèles de réseaux de neurones ont été introduits en 1943 par les neurologues McCulloch et Pitts. Toute fois, la technologie de l'époque ne leur a pas permis d'obtenir beaucoup des progrès. D'autres chercheurs comme Donald Hebb, qui a présenté en 1949 une série d'idées sur la structure et le fonctionnement des systèmes biologiques de neurones [10] et Frank Rosenblatt, qui a développé entre 1957 et 1959 [11, 12] le perceptron, un algorithme neuronal simple, ont contribué au développement de ce type d'algorithmes. Toutefois on a dû attendre le milieu des années 1980 pour que cette approche acquiert une nouvelle force, grâce à l'algorithme d'apprentissage de rétro-propagation (Back Propagation) introduit par Rumelhart et McClelland en 1986, à partir du quel ils ont montré que les réseaux de neurones de multiples couches ont une capacité exceptionnelle de discrimination en étant capables d'apprendre des patrons complexes [13].

Pour comprendre le fonctionnement des réseaux de neurones artificiels, il est utile de savoir comment fonctionnent les neurones naturels. Les neurones naturels reçoivent des signaux par des synapses situées sur les dendrites ou membrane du neurone. Quand les signaux reçus sont assez

forts (surpassant un certain seuil), le neurone est activé et émet un signal à travers l'axone. Ce signal pourrait être envoyé à une autre synapse, et pourrait activer d'autres neurones [14]. Inspirés de ce mécanisme, les réseaux de neurones artificiels sont représentés par des Nœuds qui constituent les neurones artificiels.

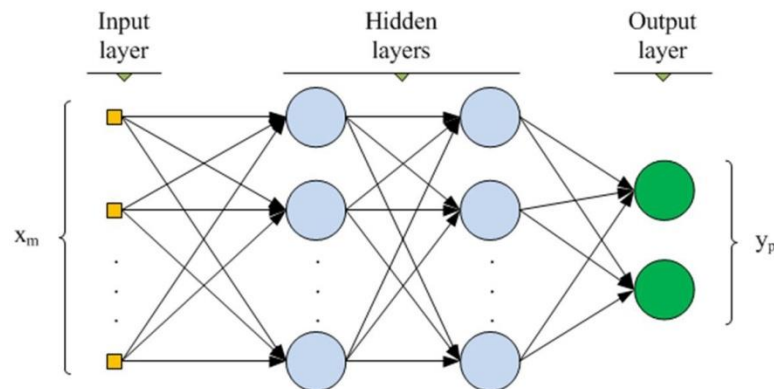


Fig. I.3. L'architecture d'un réseau de neurones de type MLP.

I.5.2. Machines par modélisation

I.5.2.1. Modèle de mélange gaussiens

Un modèle de mélange gaussien (MMG) est une catégorie de modèle probabiliste qui stipule que tous les points de données générés sont dérivés d'un mélange de distributions gaussiennes finies sans paramètres connus. Les paramètres des modèles de mélange gaussien sont dérivés d'une estimation maximum a posteriori ou d'un algorithme itératif d'anticipation-maximisation à partir d'un modèle antérieur bien entraîné. Les modèles de mélange gaussien sont très utiles pour modéliser des données, en particulier des données provenant de plusieurs groupes.

Les modèles de mélange gaussien sont utilisés dans les systèmes biométriques, où le modèle paramétrique aide à comprendre les caractéristiques ou les mesures associées à des caractéristiques telles que les caractéristiques spectrales du tractus vocal.

Les modèles de mélange gaussien sont également utilisés pour l'estimation de la densité et sont considérés comme les techniques les plus matures sur le plan statistique pour la classification.

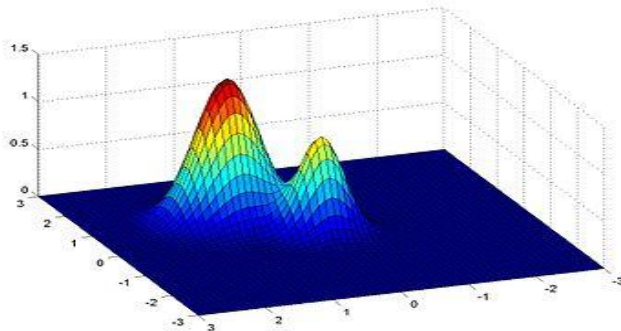


Fig. I.4. Le principe de MMG.

I.5.2.2. Chaines de Markov cachées

Les modèles de Markov cachés MMCs sont des modèles statistiques, riches et largement utilisés en traitement du signal. Ils sont développés par Andrew Markov (étudiant de Tchebychev), et ils sont premièrement, orientés vers des objectifs linguistiques dans des travaux de littérature Russe. Ceux sont des outils efficaces en modélisation des données séquentielles ou 'time-series Data'. Utilisés par la suite dans des problèmes de reconnaissance de la parole par Baker, leur théorie de base est introduite par Baum et ses collègues à la fin des années soixante.

Actuellement, ces modèles sont, de plus en plus adoptés en reconnaissance automatique de la parole, pour l'analyse des séquences ADN, et dans des problèmes liés à l'écriture et le traitement de texte. Aussi, leur utilisation en visionique est vaste. Ils sont implémentés en segmentation d'images, reconnaissance des visages, interprétation des gestes, ainsi que la modélisation de l'arrière-plan et récemment en traitement vidéo.

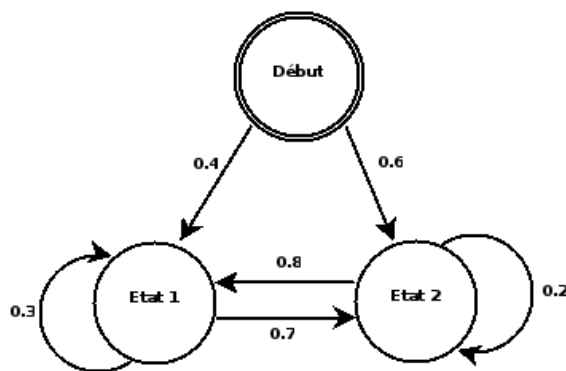


Fig. I.5. Le principe de MMC.

I.6. TYPES D'APPRENTISSAGE

On peut avoir à mettre en place différents types d'apprentissage :

Apprentissage Supervisé : Cette approche a pour objectif la conception d'un modèle reliant des données d'apprentissage à un **ensemble de valeurs de sortie** (un comportement).

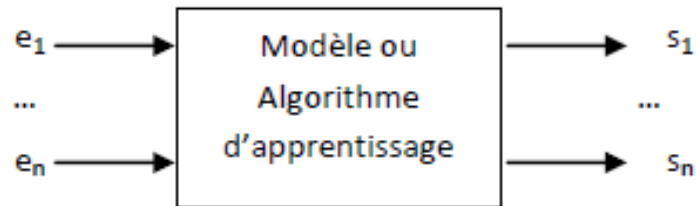


Fig. I.6. Schéma d'un modèle supervisé.

Apprentissage par Renforcement : Les données en entrée sont les mêmes que pour l'apprentissage supervisé, cependant l'apprentissage est guidé par l'environnement sous la forme de récompenses ou de pénalités données en fonction de l'erreur commise lors de l'apprentissage.

Apprentissage Non-Supervisé : Il vise à concevoir un modèle structurant l'information. La différence ici est que les comportements (ou catégories ou encore les classes) des données d'apprentissage ne sont pas connus, c'est ce que l'on cherche à trouver.

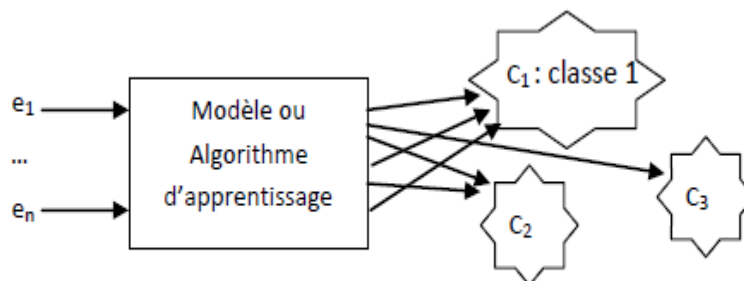


Fig. I.7. Schéma d'un modèle non-supervisé.

Apprentissage Semi-Supervisé : Les données d'entrée sont constituées d'exemples étiquetés et non étiquetés. Ce qui peut être très utile quand on a deux types de données, car cela permet de ne pas en laisser de côté et d'utiliser toute l'information.

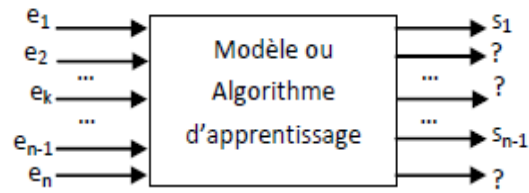


Fig. I.8. Schéma d'un modèle semi-supervisé.

I.7. CONCLUSION

Dans ce chapitre, nous avons présenté des notions sur la biomédicale et sa relation avec les machines d'apprentissage. Nous avons montré que les machines d'apprentissage permettent d'améliorer les traitements et les décisions médicales par les professionnels de la santé.

Nous avons ainsi présenté les différentes catégories des machines d'apprentissage et également les différents types d'apprentissage. Dans le prochain chapitre, nous allons utiliser ces notions pour une étude expérimentale et comparative entre différents filtres.

Chapitre II

Système propose

II.1. INTRODUCTION

Dans le cadre de notre travail, la solution adoptée est une reconnaissance de pathologie par les techniques de reconnaissance de formes. Dans ce chapitre, nous allons donc pouvoir passer en revue des méthodes d'apprentissage statistique appliquées à la sélection des caractéristiques et à la classification. Après une brève introduction, où nous allons au premier temps donnons les étapes et les techniques essentielles pour chaque module de reconnaissance en commençant par la sélection des caractéristiques comme une technique de prétraitement des données utilisée en vue d'une réduction de la dimension des variables d'entrée. Nous allons rappeler les aspects théoriques et fondements de la méthode des Machines à vecteurs de supports (SVM) appliquée à la classification. Un grand intérêt sera focalisé sur les SVM où nous leurs nouvelle variante comme les TWIN-SVM.

II.2. ARCHITECTURE GLOBALE DU SYSTEME DE RECONNAISSANCE PROPOSE

La conception d'un système de reconnaissance de pathologie est réalisée en deux phases :

- Phase d'apprentissage (training) ;
- Phase de test (testing).

La figure II.1 représente les différents modules constituant le système de reconnaissance de pathologie

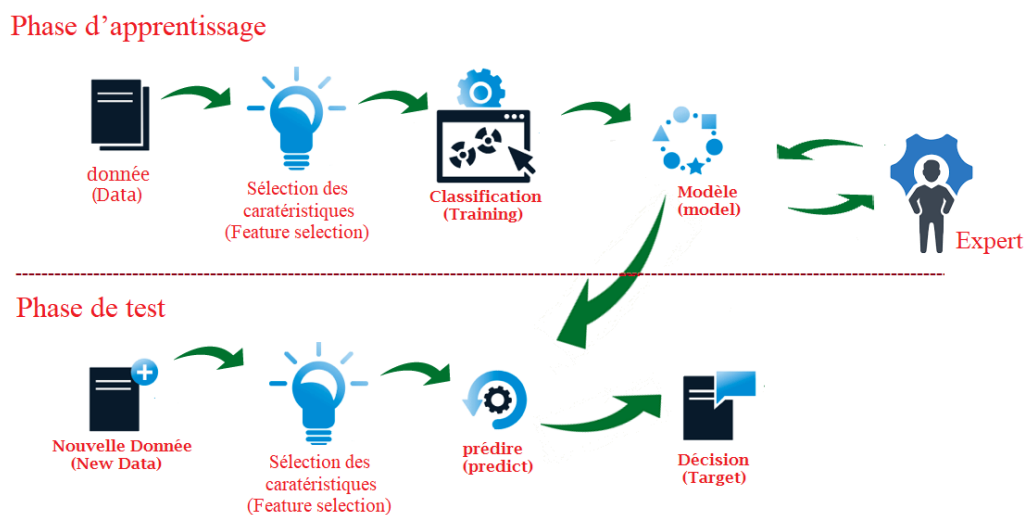


Fig. II.1. Système de reconnaissance de formes.

Phase d'apprentissage : consiste à apprendre au classifier les différentes formes. Celle-ci se déroule en deux étapes : extraction de caractéristiques et génération du modèle de classification.

Phase de test : consiste à valider grâce à une fonction de décision le modèle de classification

II.3. SELECTION DES CARACTERISTIQUES

Plusieurs paramètres peuvent influencer sur les performances d'un système de classification. Les caractéristiques extraites à partir des entités afin de représenter ces dernières peuvent être considérées parmi les paramètres les plus importants. La réduction de la dimensionnalité, via l'extraction et la sélection d'attributs, est l'une des étapes les plus importantes dans le traitement de données. Son but est de sélectionner ou d'extraire un sous-ensemble optimal de caractéristiques pertinentes pour un critère fixé auparavant. Les principaux objectifs de la réduction de dimension sont :

- Faciliter la visualisation et la compréhension des données.
- Réduire l'espace de stockage nécessaire.
- Réduire le temps d'apprentissage et d'utilisation.
- Identifier les facteurs pertinents.
- Améliorer la précision du module de classification.
- Améliorer la fiabilité de l'estimation de la performance.

La sélection de caractéristiques permet de sélectionner un sous-ensemble de caractéristiques parmi un ensemble de départ, en utilisant divers critères et différentes méthodes (Figure II.2).

L'extraction de caractéristiques permet de créer de nouveaux ensembles de caractéristiques, en utilisant une transformation ou une combinaison d'un espace de départ et en effectuant une réduction du nombre de dimensions (Figure II.3). Dans le cadre de notre travail, nous nous intéressons aux méthodes de sélection de caractéristique [15].

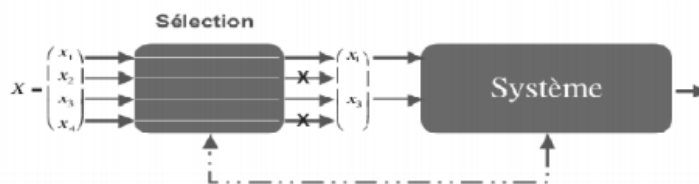


Fig. II.2. Principe de la sélection de caractéristiques [16].

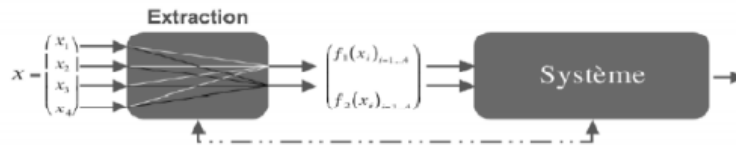


Fig. II.3. Principe de l'extraction de caractéristiques [17].

II.3.1. Définition

La sélection de caractéristiques est généralement définie comme un processus de recherche permettant de trouver un sous-ensemble "pertinent" de caractéristiques parmi celles de l'ensemble de départ. La notion de pertinence d'un sous-ensemble de caractéristiques dépend toujours des objectifs et des critères du système. En général, le problème de sélection de Caractéristiques peut être décrit par :

Soit $F = (f_1, f_2, \dots, f_N)$ un ensemble de caractéristiques de taille N ou N représente le nombre total de caractéristiques étudiées. Soit Ev une fonction qui permet d'évaluer un sous-ensemble de caractéristiques. Nous supposons que la plus grande valeur de Ev soit obtenue pour le meilleur sous-ensemble de caractéristiques. L'objectif de la sélection est de trouver un sous-ensemble F' ($F' \subset F$) de taille N' ($N' \leq N$) tel que :

$$Ev(F') = \max_{Z \subset F} Ev(Z) \quad (\text{II.1})$$

Où $|Z| = N'$ et N' est soit un nombre prédéfini par l'utilisateur ou soit contrôlé par une des méthodes de génération de sous-ensembles

II.3.2. Schéma général de la sélection des caractéristiques

Les différentes méthodes proposées dans la littérature pour la sélection d'attributs peuvent être décrites par un schéma général [18] dans lequel on trouve les éléments clés suivants (Figure II.4):

1. Une procédure de génération de sous ensembles candidats qui détermine l'exploration de l'espace de recherche.
2. Une fonction d'évaluation donnant la qualité de sous ensembles candidats.
3. Une condition d'arrêt.
4. Un processus de validation pour vérifier si l'objectif souhaité est atteint.

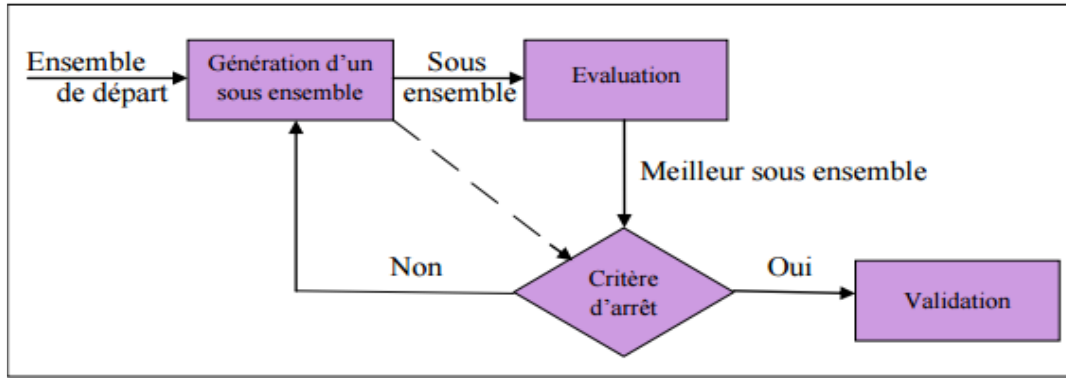


Fig. II.4. Processus de sélection d'attributs.

Dans la section suivante, nous expliquons une technique de sélection des caractéristiques que nous allons les utilisées dans notre étude.

II.3.3. Neighborhood Component Analysis (NCA)

Soit $T = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)\}$ l'ensemble d'échantillons d'entraînement, où x_i est un vecteur caractéristique de dimension d , $y_i \in \{1, \dots, C\}$ est son étiquette de classe correspondante et N est un nombre d'échantillons. Le but est de trouver un vecteur de pondération w qui se prête à sélectionner le sous-ensemble d'entités en optimisant la classification des plus proches voisins. En termes de vecteur de pondération w , nous désignons la distance pondérée entre deux échantillons x_i et x_j par :

$$D_w(x_i, x_j) = \sum_{l=1}^d w_l^2 |x_{il} - x_{jl}| \quad (\text{II.2})$$

Où w_l est un poids associé à la caractéristique l . Pour que la classification des plus proches voisins réussisse, une intuitive et une stratégie efficace est de maximiser la précision *leave-one-out* classification sur l'ensemble d'apprentissage T . Cependant, en raison de la véritable précision de *leave-one-out* qui sélectionne le plus proche voisin comme un point de référence de classification, est un non différentiable fonction. Une approximation efficace est que le point de référence est déterminé par une distribution de probabilité. Ici, la probabilité que x_i sélectionne x_j comme point de référence est définie comme :

$$p_{ij} = \begin{cases} \frac{\kappa(D_w(x_i, x_j))}{\sum_{k \neq i} \kappa(D_w(x_i, x_k))} & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \quad (\text{II.3})$$

Où $\kappa(z) = \exp(-z / \sigma)$ est une fonction du noyau et le σ la largeur du noyau est un paramètre d'entrée qui influence la probabilité de chaque point étant sélectionné comme point de référence. En particulier, si $\sigma \rightarrow 0$, seul le voisin le plus proche de l'échantillon de requête

peut être sélectionné comme point de référence. D'autre part, si $\sigma \rightarrow +\infty$, tous les points ont la même chance d'être sélectionné en dehors du point de requête. Sur la base de la définition ci-dessus, la probabilité de la requête x_i pour qu'il soit correctement classifié est donné par:

$$p_i = \sum_j y_{ij} p_{ij} \quad (\text{II.4})$$

Où $y_{ij} = 1$ si et seulement si $y_i = y_j$ et $y_{ij} = 0$ sinon. Par conséquent, l'approximation de la classification peut s'écrire comme suit:

$$\xi(w) = \frac{1}{N} \sum_i p_i = \frac{1}{N} \sum_i \sum_j y_{ij} p_{ij} \quad (\text{II.5})$$

Notez que quand $\sigma \rightarrow 0$, $\xi(w)$ est la vraie précision de *leave-one-out* classification. De plus, pour effectuer la sélection de caractéristiques et atténuer le sur-ajustement, nous introduisons un terme de régularisation et donc obtenir la fonction objective suivante:

$$\xi(w) = \sum_i \sum_j y_{ij} p_{ij} - \lambda \sum_{l=1}^d w_l^2 \quad (\text{II.6})$$

où $\lambda > 0$ est un paramètre de régularisation qui peut être réglé par validation croisée. Notez que le coefficient $1/N$ dans (4) est négligé car cela signifie seulement que le paramètre λ apporte un changement correspondant et le vecteur final est le même. Puisque la fonction objective $\xi(w)$ est différentiable, sa dérivée par rapport à w peut être calculée comme suit:

$$\frac{\partial \xi(w)}{\partial w_l} = \sum_i \sum_j y_{ij} \left[\frac{\partial}{\partial \sigma} p_{ij} (\sum_{k \neq i} p_{ik} |x_{il} - x_{kl}| - |x_{il} - x_{jl}| w_l) - 2\lambda w_l \right] \quad (\text{II.7})$$

$$= \frac{\partial}{\partial \sigma} \sum_i (p_i \sum_{k \neq i} p_{ik} |x_{il} - x_{kl}| - \sum_j y_{ij} p_{ij} |x_{il} - x_{jl}|) w_l - 2\lambda w_l \quad (\text{II.8})$$

$$= 2 \left(\frac{\partial}{\partial \sigma} \sum_i (p_i \sum_{j \neq i} p_{ij} |x_{il} - x_{jl}| - \sum_j y_{ij} p_{ij} |x_{il} - x_{jl}|) - \lambda \right) w_l \quad (\text{II.9})$$

L'utilisation de la dérivée ci-dessus conduit à l'équation de gradient. Le pseudo code de l'algorithme de NCA est donné ci-dessous [19].

Algorithme : calcul du NCA

1 : **procédure** NCFS ($T, \alpha, \sigma, \lambda, \eta$) où T : ensemble d'apprentissage, α : longueur de l'étape initiale, σ : largeur du noyau, λ : paramètre de régularisation, η : petite constante positive;

2 : Initialisation: $\mathbf{w}^{(0)} = (\mathbf{1}, \mathbf{1}, \dots, \mathbf{1}), \epsilon^{(0)} = -\infty, \mathbf{t} = \mathbf{0}$

3 : **répéter**

4 : **for** $i = 1, \dots, N$ **do**

5 : Calculer \mathbf{p}_{ij} et \mathbf{p}_i en utilisant $\mathbf{w}^{(t)}$ selon les équations (2) et (3)

6 : **for** $l = 1, \dots, d$ **do**

7 : $\Delta_t = 2 \left(\frac{1}{\sigma} \sum_i (\mathbf{p}_i \sum_{j \neq i} \mathbf{p}_{ij} |x_{il} - x_{jl}| - \sum_j y_{ij} \mathbf{p}_{ij} |x_{il} - x_{jl}|) - \lambda \right) \mathbf{w}_l^{(t)}$

8 : $\mathbf{t} = \mathbf{t} + \mathbf{1}$

9 : $\mathbf{w}^{(t)} = \mathbf{w}^{(t-1)} + \alpha \Delta$

10 : $\epsilon^{(t)} = \xi(\mathbf{w}^{(t-1)})$

11 : **end for**

12 : **end for**

13 : **if** $\epsilon^{(t)} > \epsilon^{(t-1)}$ **then**

14 : $\alpha = 1.01\alpha$

15 : **else**

16 : $\alpha = 0.4\alpha$

17 : **fin** jusqu'à $|\epsilon^{(t)} - \epsilon^{(t-1)}| < \eta$

18 : $\mathbf{w} = \mathbf{w}^{(t)}$

19 : **revenir** \mathbf{w}

II.4. CLASSIFICATION DE DONNEES

Après la sélection de descripteurs, l'étape de reconnaissance consiste à évaluer la fonction discriminante pour affecter chaque vecteur à la classe ayant la plus grande vraisemblance. Le terme de classification regroupe trois approches de traitement de données. Dans la première, les observations de la base de données sont étiquetées et il s'agit d'une classification supervisée. Dans la seconde approche, aucune des observations n'est étiquetée et c'est le cas de la classification automatique ou non supervisée. Enfin, dans la troisième approche seulement une partie des observations est étiquetée et on parle d'une classification semi-

supervisée [20]. Théoriquement, toutes les techniques de classification supervisée peuvent être exploitées pour la Reconnaissance mais en pratique, seuls les classifieurs pouvant confronté la complexité présente dans les données sont utilisés. Ces classifieurs sont généralement, répartis selon deux taxonomies. La première partage les techniques de classification en deux familles, une approche paramétrique qui concerne les classifieurs qui présupposent la forme des lois et une approche non paramétrique qui les estiment sans connaissance a priori. Dans la première famille, on trouve tous les classifieurs statistiques et probabilistes tels que les modèles de mélange et les classifieurs Bayesiens [21]. Par ailleurs, dans l'approche non paramétrique on distingue les classifieurs Plus Proches Voisins, les réseaux de neurones et les SVMs. La deuxième taxonomie quant à elle, distingue les approches procédant par modélisation ou par séparation. L'approche par modélisation est basée sur la construction d'un modèle pour chaque classe. Ainsi, les données sont affectées selon les mesures de vraisemblance [22]. Dans cette catégorie, on distingue d'une part, les classifieurs KPPV qui ne nécessitent pas de compilation et utilisent tous les exemples disponibles comme des références. En effet, la décision se fait à l'aide d'une mesure de distance qui peut être la distance Euclidienne [22], la distance élastique [23] ou encore la distance tangente [24]. D'autre part, les classifieurs procédant par compilation de données sont les classifieurs Bayesiens et les HMM. Par ailleurs, l'approche par séparation consiste plutôt à partager l'espace de décision en plusieurs zones dont les frontières sont définies par des hyperplans. Les classifieurs typiques de cette catégorie sont les réseaux de neurones et les SVMs. La figure (2.6) illustre la position des SVMs qui font l'objet de notre travail [20, 25, 26].

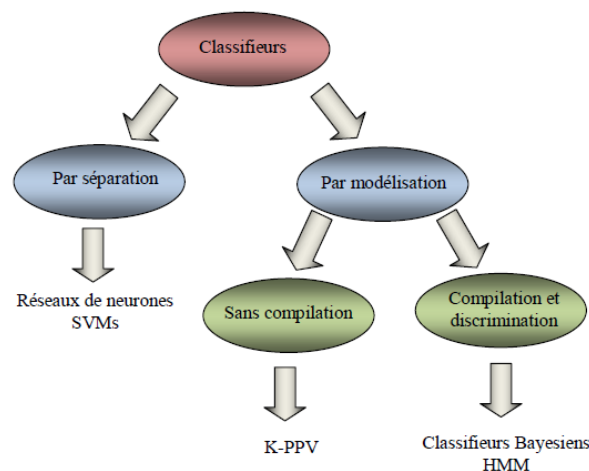


Fig. II.5. Taxonomie des méthodes utilisées pour la reconnaissance [22].

II.4.1. Machines à Vecteurs de Supports

Les machines à vecteurs de support SVMs ou séparateurs à vaste marge constituent une technique d'apprentissage supervisé destinée à résoudre des problèmes de classification [27]. La classification de données se fait dans un contexte supervisé où on cherche à estimer une fonction $f(x)$ à partir des exemples x , comme l'indique la figure (II.6).

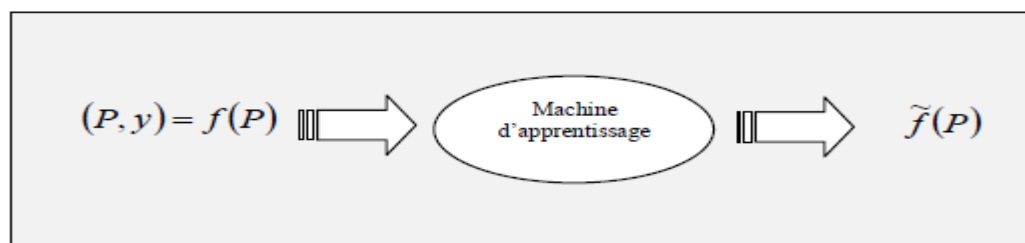


Fig. II.6. Apprentissage supervisé d'une machine.

L'origine des SVMs remonte en 1975 lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la dimension de VC (VC : Vapnik Chervonenkis) pour caractériser la capacité d'une machine d'apprentissage. A cette époque ce principe n'a pas trouvé place car il n'existait pas encore un modèle de classification solidement appréhendé pour être utilisé avec cette théorie. Il a fait attendre jusqu'en 1982 pour que Vapnik propose un premier classifieur basé sur la Minimisation du Risque Structurel (MRS) baptisé SVM. Ce modèle était toutefois linéaire et l'on ne connaissait pas encore le moyen d'introduire des frontières de décision non-linéaires. En 1992 Boser et al [28] proposèrent d'introduire des noyaux non-linéaires pour étendre les SVMs aux cas non-linéaires. Par la suite, Cortes et Vapnik [29] ont introduit des paramètres de relâchement qui tolèrent quelques erreurs d'apprentissage pour mettre fin à l'algorithme d'optimisation du SVM qui peut autrement, tourner indéfiniment. Dans ce qui suit nous présentons les classifieurs SVMs mais nous passons d'abord en revue quelques notions sur la théorie de l'apprentissage statistique.

II.4.1.1. Principe des SVMs

Les SVMs sont conçus pour mettre en œuvre le principe du risque structurel. Grâce à leur base théorique solide, ces classifieurs ont requis un très grand succès dans les différents domaines de traitement de données. En outre, ils répondent à deux problèmes majeurs de la théorie de l'apprentissage statistique qui sont:

- Le contrôle de la capacité du classifieur.
- Le sur-apprentissage de données.

Aussi, deux autres avantages sont particulièrement offerts par les SVMs. D'une part, contrairement aux machines d'apprentissage classiques dont le nombre de paramètres à définir est important, les SVMs possèdent une formulation élégante qui laisse très peu de place aux paramètres utilisateurs. D'autre part, puisque l'erreur ne dépend pas de la dimension de l'espace d'entrée, les SVMs sont bien adaptés pour le traitement des données de très haute dimension telles que les textes et les images. L'apprentissage d'un SVM consiste à chercher l'hyperplan permettant une séparation optimale entre deux classes de données. L'erreur est minimisée tout en maximisant la marge de séparation et en contrôlant la dimension VC du classifieur.

a. SVMs linéaires

Soit un ensemble de données d'apprentissage : $\{x_i, y_i\}, i = 1, \dots, l$ avec : $x_i \in R^d$ et $y_i \in \{-1, +1\}$. Supposons qu'on dispose d'un hyperplan séparant les données positives des données négatives. Les x_i qui appartiennent à l'hyperplan vérifient la relation: $x_i \cdot w + b = 0$. Où w est la normale de l'hyperplan tandis que $|b| / \|w\|$ est la distance perpendiculaire entre l'hyperplan et l'origine (Figure II.7).

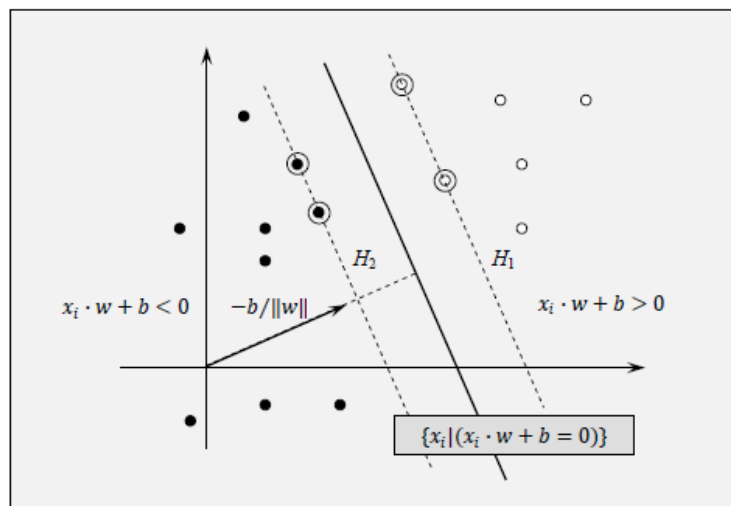


Fig. II.7. Hyperplan de séparation linéaire pour des données linéairement séparables (Les vecteurs de support sont encerclés)

L'algorithme des vecteurs de support cherche simplement l'hyperplan séparateur permettant la plus large marge entre les exemples des deux classes. Ceci est formulé par les inégalités suivantes:

$$\begin{aligned} x_i \cdot w + b &\geq +1, y_i = +1 \\ x_i \cdot w + b &\leq -1; y_i = -1 \end{aligned} \quad (\text{II.10})$$

Et l'on écrira aussi :

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \quad (\text{II.11})$$

Le Lagrangien primal, correspond à la différence entre la fonction objective et la multiplication de la contrainte dans (II.11) par des multiplieurs α_i .

$$L_P = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^l \alpha_i \quad (\text{II.12})$$

L'objectif est la minimisation du Lagrangien par rapport à w et b sous la contrainte que les dérivées premières par rapport à tous les multiplieurs α_i , disparaissent et que $\forall i, \alpha_i \geq 0$. Ce problème étant convexe et puisque la fonction objective est elle-même convexe, on travaille plutôt sur sa forme duale. Alors, l'optimisation du SVM revient donc à maximiser le Lagrangien dual (L_D) par rapport aux α_i , sous les contraintes que ses dérivées premières par rapport à w et b disparaissent. En imposant que les gradients de L_P par rapport à w et b disparaissent, nous obtenons :

$$\begin{aligned} w &= \sum_i \alpha_i y_i x_i \\ \sum_i \alpha_i y_i &= 0 \end{aligned} \quad (\text{II.13})$$

En substituant les équations (II.13) dans (II.12), nous obtenons le Lagrangien Dual L_D :

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (\text{II.14})$$

Après l'optimisation, les points dont les α_i sont strictement supérieurs à 0, appelés vecteurs de support, appartiennent à l'un des hyperplans H_1 ou H_2 . Ces points sont les plus proches de la frontière de décision et forment le plan séparateur. Les autres points n'interviennent pas dans le calcul de la fonction de décision. Par ailleurs, les conditions de **Karush-Kuhn-Tucker** (KKT) résument les contraintes que le processus d'optimisation doit vérifier.

Conditions de Karush-Kuhn-Tucker

Les conditions de *KKT* pour le problème d'optimisation du SVM sont les suivantes :

$$\begin{aligned} \frac{\partial}{\partial w_p} Lp &= w_p - \sum_i \alpha_i y_i x_{iv} = 0, v = 1, \dots, d \\ \frac{\partial}{\partial w_p} Lp &= - \sum_i \alpha_i y_i = 0 \\ y_i(x_i \cdot w + b) - 1 &\geq 0, i = 1, \dots, l \\ \alpha_i &\geq 0, \forall i \\ \alpha_i(y_i(x_i \cdot w + b) - 1) &= 0, \forall i \end{aligned} \quad (\text{II.15})$$

d : Dimension des données. Ces conditions sont nécessaires et suffisantes pour que les w et b et α_i ainsi obtenus soient une solution du problème. En résumé, l'optimisation d'un SVM revient à trouver une solution aux conditions *KKT*. Enfin, la fonction de décision est donnée par :

$$g(x) = \text{sign}\left(\sum_{i=1}^{vs} y_i x_{iv} x_i^T x + b\right) \quad (\text{II.16})$$

b. SVMs non linéaires

Naturellement, en reconnaissance de formes les données nécessitent une fonction de décision non linéaire pour être séparées. Pour généraliser les SVMs aux cas non linéaires, on introduit des noyaux de *Mercer* permettant de trouver une hyper-courbe qui marquerait la frontière de décision entre les exemples positifs et négatifs. La figure (II.8) illustre le passage de l'espace d'entrée vers l'espace augmenté.

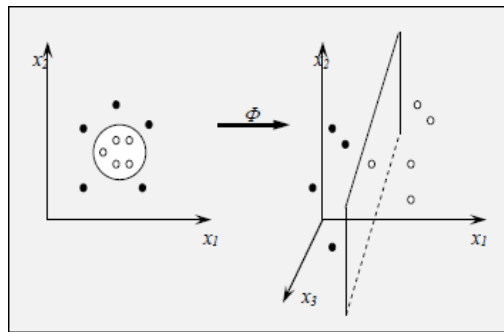


Fig. II.8. Exemple d'une application rendant les données linéairement séparables.

Le problème d'optimisation reste typiquement le même. Le Lagrangien dual de la fonction objective à maximiser devient alors :

$$\text{Maximiser} \quad \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, y_j) \quad (\text{II.17})$$

Nous obtenons ainsi un SVM fonctionnant dans un espace de dimension infinie où toutes les considérations des sections précédentes sont vérifiées puisque nous maintenons toujours la séparation linéaire, mais dans un espace différent. La fonction de décision devient :

$$g(x) = \sum_{i=1}^{vs} \alpha_i y_{iv} K(x_i, x) + b \quad (\text{II.18})$$

c. Algorithmes d'optimisation du SVM

A cause de sa taille considérable, le problème de *PQ* d'un *SVM* ne peut pas être facilement résolu avec la technique standard. La matrice de gram contient un nombre d'éléments égal au carré du nombre total de données. Depuis plusieurs années, deux algorithmes ont été proposés pour ce fait. Il s'agit de l'algorithme procédant par tronçon ou *Chunking* et l'algorithme de décomposition. Plus récemment, Platt a proposé un algorithme beaucoup plus rapide baptisé optimisation minimale et séquentielle (ou *Sequential Minimal Optimisation SMO*).

Optimisation minimale et séquentielle : *SMO*

Le *SMO* est un algorithme simple et rapide proposé pour résoudre le problème de *PQ* du *SVM* sans la nécessité de stocker une grande matrice en mémoire. A l'instar des autres algorithmes, il décompose le problème d'optimisation en plusieurs sous-problèmes mais à chaque étape il optimise le plus petit problème possible. En effet, à chaque étape les valeurs de deux multiplicateurs du Lagrangien sont optimisées. Ceci dit, l'avantage principal du *SMO* réside dans le fait que l'optimisation de deux multiplicateurs peut être effectuée analytiquement. En plus, le *SMO* ne nécessite pas un espace mémoire important. Cet algorithme se compose principalement de trois éléments :

1. Une méthode analytique pour résoudre le problème de *PQ*.
2. Des heuristiques pour choisir les multiplicateurs à optimiser.
3. Une méthode pour calculer le seuil b .

d. SVM multi-classes

A l'origine, les *SVMs* ont été conçus pour la séparation de deux classes. Cependant, plusieurs approches permettent d'étendre cet algorithme aux problèmes multi-classes. En effet, il existe deux approches classiques permettant l'implémentation multi-classes des *SVMs*.

Un contre un (OAO : one against one).

Dans cette approche chaque SVM sépare deux classes de données. Pour un problème de M classes, on doit construire $M(M-1)/2$ SVMs. On définit f_{ij} la fonction de décision séparant la classes i de la classes j par : $f_{ij} = -f_{ji}(p)$. durant la phase de classification, on calcule le score de chaque classe tel que :

$$S_i(p) = \sum_{j \neq i, j=1}^M \text{sign}(f_{ij}(p)) \quad (\text{II.19})$$

En fin, p est affecté à la classe ayant le plus grand score :

$$\text{arg max}_{i=1, \dots, M} S_i(p) \quad (\text{II.20})$$

Un contre tous (OAA : one against all)

C'est la première approche proposée pour l'implémentation multi-classes des SVMs. Le principe consiste à déterminer pour chaque classe un hyperplan séparant celle-ci de toutes les autres classes. Ainsi, pour M classes on doit déterminer M fonction de décision. Tous les exemples appartenant à la classe considérée sont étiquetés positivement (+1) et tous les exemples n'appartenant pas à la classe sont étiquetés négativement (-1). La figure () décrit le schéma synoptique de cette séparation multi-classes. Par ailleurs, pour chaque exemple de test, on calcule M fonctions de décision. Ensuite, l'exemple est affecté à la classe qui correspond à la classe ayant la plus grande valeur de sortie.

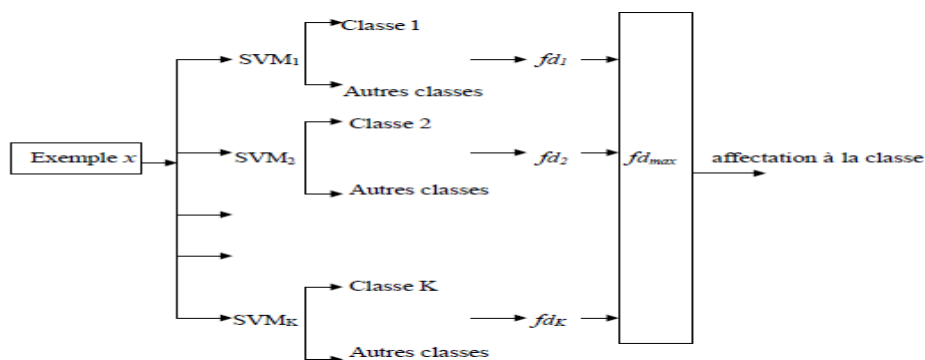


Fig. II.9. Schéma synoptique de l'implémentation un contre tous.

SVM et noyaux

Les techniques SVM (non linéaires) font appel à une fonction implicite Φ transformant l'espace d'entrée $X \subset R^d$ en un espace hilbertien $(H, (\cdot, \cdot))$ de plus grande dimension. L'apprentissage s'effectue alors à partir du modèle $(\phi(X), Y)$ dans l'espace H , de dimension plus grande certes, mais dans lequel on espère que les données soient "davantage linéairement séparables". Du point de vue pratique, il convient de noter que le calcul des projections $\phi(X)$ n'est pas utilisé dans la méthode, seuls les produits scalaires $\{\Phi(x), \Phi(x')\}, (x, x') \in X^2$, sont requis. Or, ceux-ci sont donnés par un noyau K , via la relation ("kernel trick")

$$K(x, x') = \{\Phi(x), \Phi(x')\} \quad (\text{II.21})$$

La méthode requiert donc une sélection d'un noyau avec ses paramètres éventuels. Parmi les choix possibles, on cite en particulier [30] :

- Le noyau linéaire : $K(x, x') = (x, x')$ (correspondant aux SVM linéaires)
- Le noyau Gaussien radial (Gaussian RBF) $K(x, x') = \exp(-\sigma \|x - x'\|^2)$
- Les noyaux polynomiaux $K(x, x') = (\alpha + \beta(x, x'))^\delta$
- Le noyau radial de Laplace (Laplace RBF) $K(x, x') = \exp(-\sigma \|x - x'\|)$
- Le noyau tangente hyperbolique (sigmoïde) $K(x, x') = \tanh(\alpha + \beta(x, x'))$
- Normalized poly: $((x' * x + \alpha) / \sqrt{(x' * x + \alpha) * (x * x' + \alpha)})^\delta$

II.4.2. Variantes de SVM

SVM présente de nombreux avantages, notamment le fait de fournir une solution globale pour la classification des données. Il génère un hyper-plan global unique pour séparer les échantillons de données de différentes classes plutôt que de limites locales par rapport à d'autres approches de classification de données existantes. Dans la mesure où la SVM respecte le principe de minimisation des risques structurels (SRM), elle réduit l'occurrence des risques pendant la phase de formation et améliore ses capacités de généralisation. En raison de sa meilleure performance, SVM est l'un des plus utilisés des techniques de classification des mines de données qui a des applications dans de nombreux domaines allant de la détection de la maladie, la catégorisation de textes, défaut du logiciel prédiction, reconnaissance de la parole, identification du visage, prévision de la faillite, détection des intrusions, prévision des séries chronologiques, détection des musicales, etc [13~38] Mais l'un des principaux problèmes du SVM conventionnel est d'obtenir la solution d'un problème complexe de programmation quadratique (QPP).

Récemment, Mangasarian et al. introduit un SVM proximal à valeur propre généralisée (GEP-SVM) qui génère deux hyper-plans non parallèles pour deux problèmes de classe [31]. Dans cette approche, les modèles ou les échantillons de données de chaque classe se trouvent à proximité immédiate d'un hyper-plan et maintiennent une séparation nette avec les autres. Sur la base de SVM et de GEP-SVM, Jayadeva et al. a proposé un nouveau classificateur binaire, TWSVM (Twin Support Vector Machine), qui classe les modèles de deux classes en utilisant deux hyper-plans non parallèles [32]. TWSVM résout une paire de QPP au lieu d'un seul QPP complexe comme dans le SVM traditionnel. En SVM, tous les échantillons de données constituent une contrainte pour la QPP, à savoir la formulation double pour le SVM dépend du nombre de tous les échantillons de données dans le jeu d'entraînement. Dans TWSVM, les modèles d'une classe créent des contraintes pour les autres QPP et inversement.

Pour les échantillons de données d'apprentissage 'n' size, la complexité de calcul de SVM est $O(n^3)$. Si le nombre d'échantillons de données dans chaque classe est approximativement égal à $n/2$, la complexité de TWSVM est alors égale à $O(2 \times (n/2)^3)$, ce qui est quatre fois plus rapide que celle du SVM traditionnel. Le ratio de complexité de calcul de SVM et TWSVM est :

$$\frac{n^2}{2 \times (\frac{n}{2})^3} = 4 \tag{II.22}$$

Il existe deux manières d'améliorer le TWSVM conventionnel, comme illustré à la Figure II.10. Dans la première méthode, des améliorations sont apportées à la formulation de la méthode originale afin d'améliorer ses performances. Dans la deuxième méthode, TWSVM classique est étendu des classes binaires aux multi-classes.

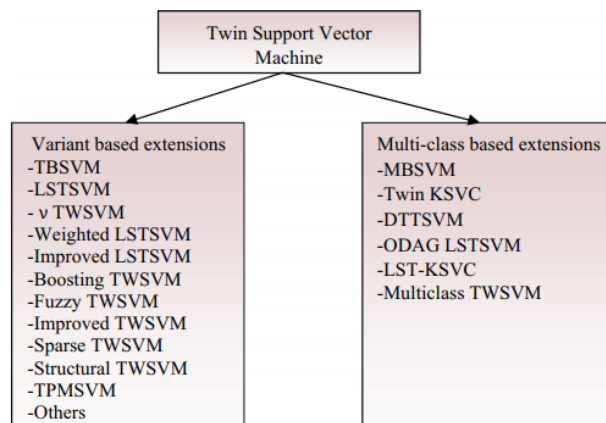


Fig. II.10. Extension of Twin Support Vector Machine.

II.4.2.1. Twin Support Vector Machine

TWSVM est l'une des nouvelles approches d'apprentissage automatique qui convient aux problèmes de classification et de régression. Le but de TWSVM est de construire deux plans non parallèles pour chaque classe en optimisant une paire de QPP de telle sorte que chaque hyper-plan soit plus proche des échantillons de données d'une classe et également soit éloigné des échantillons de données de l'autre classe [32]. Ainsi, TWSVM résout une paire de QPP de plus petite taille au lieu d'un QPP complexe comme dans le SVM conventionnel. La Figure II.11 représente la catégorisation de deux classes à l'aide de TWSVM. Comme le montre la figure, il existe deux classes - les classes 1 et 2 - qui sont divisées en utilisant deux plans non parallèles de manière à ce que chaque plan soit plus proche des échantillons de données d'une classe et plus éloignés de l'autre.

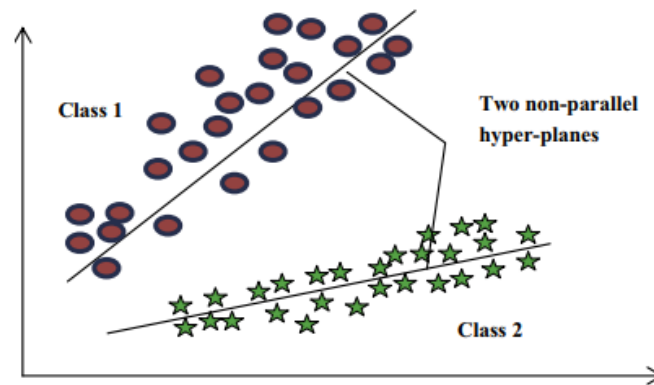


Fig. II.11. Classification binaire à l'aide de TWSVM.

Laissons les échantillons des classes positives et négatives sont désignés par ' m ' et ' n ' respectivement et les échantillons de données de classes positives et négatives sont représentés par des matrices $X_1 \in R^{m \times k}$ et $X_2 \in R^{n \times k}$ conséquence, où R indique un espace réel de dimension k . Les équations de deux plans d'hyper-non parallèles à k de dimension espace réel R^k sont donnés ci - dessous [33] :

$$x^T w_1 + b_1 = 0 \text{ et } x^T w_2 + b_2 = 0 \quad (\text{II.23})$$

Ici, symboles w_1 et w_2 indiquent des vecteurs normaux vers l'hyper-plan, b_1 et b_2 sont des termes de partialité. Les formulations de TWSVM pour les cas linéaires et non linéaires sont obtenues.

TWIN-SVM Linear

Le problème fondamental de TWSVM linéaire est formulé comme suit:

$$\begin{aligned}
 \min \quad & (w_1, b_1, \varepsilon) \frac{1}{2} \|X_1 w_1 + e_1 b_1\|^2 + c_1 e_2^T \varepsilon \\
 \text{s. t} \quad & -(X_2 w_1 + e_2 b_1) + \varepsilon \geq e_2, \varepsilon \geq 0 \\
 \min \quad & (w_{21}, b_2, \varepsilon) \frac{1}{2} \|X_2 w_2 + e_2 b_2\|^2 + c_2 e_1^T \eta \\
 \text{s. t} \quad & (X_1 w_2 + e_1 b_2) + \eta \geq e_1, \eta \geq 0
 \end{aligned} \tag{II.24}$$

Où les variables de jeu et les paramètres de pénalité sont représentés par ξ, η et c_1, c_2 respectivement. e_1 et e_2 représentent deux vecteurs de dimension appropriée et ayant toutes les valeurs égales à 1. Eqs. (II.24) ont ensuite été utilisés pour développer un classifieur TWSVM. Lagrangien de l'équation (II.24) est obtenu sous la forme :

$$L(w_1, b_1, \varepsilon, \alpha, \beta) = \frac{1}{2} \|X_1 w_1 + e_1 b_1\|^2 + c_1 e_2^T \varepsilon + \alpha^T ((X_2 w_1 + e_2 b_1) - \varepsilon + e_2) - \beta^T \varepsilon \tag{II.25}$$

Les multiplicateurs lagrangiens sont désignés par deux vecteurs α et β . Karush – Kuhn – Tucker (KKT) conditions d'équation (4) sont formulés comme :

$$\begin{aligned}
 \frac{\partial L}{\partial w_1} &= X_1^T (X_1 w_1 + e_1 b_1) + X_2^T \alpha = 0 \\
 \frac{\partial L}{\partial b_1} &= e_1^T (X_1 w_1 + e_1 b_1) + e_2^T \alpha = 0 \\
 \frac{\partial L}{\partial \varepsilon} &= c_1 e_2^T - \beta^T - \alpha^T = 0 \\
 &-(X_2 w_1 + e_2 b_1) + \varepsilon \geq e_2, \varepsilon \geq 0 \\
 &\alpha^T ((X_2 w_1 + e_2 b_1) - \varepsilon + e_2) = 0, \beta^T \varepsilon = 0 \\
 &\alpha \geq 0, \beta \geq 0
 \end{aligned} \tag{II.26}$$

Après Eq. (II.24) est obtenu après la fusion des équations, comme suit:

$$\begin{bmatrix} X_1^T \\ e_1^T \end{bmatrix} [X_2 \ e_1] \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} X_2^T \\ e_2^T \end{bmatrix} \alpha = 0$$

Laisser $A = [X_1 \ e_1]$ et $B = [X_2 \ e_2] u_1 = \begin{bmatrix} w_1 \\ b_1 \end{bmatrix}$.
reformulé comme :

$$A^T A u_1 + B^T \alpha = 0 \tag{II.27}$$

$$u_1 = -(A^T A)^{-1} B^T \alpha$$

Parfois, il est difficile d'obtenir l'inverse de $A^T A$. Cette condition est gérée en ajoutant un terme de régularisation δI edans l'équation ci-dessus, où ' I ' représente une matrice d'identité de dimension appropriée. Éq. (2.27) est reformulé en [32] :

$$u_1 = -(A^T A + \delta I)^{-1} B^T \alpha \quad (\text{II.28})$$

De la même manière, le vecteur normal et le biais pour la deuxième classe sont obtenus en résolvant l'équation ci-dessous:

$$u_2 = (B^T B + \delta I)^{-1} A^T \gamma \quad (\text{II.29})$$

Les vecteurs normaux et les biais sont en outre utilisés pour générer des plans non parallèles. De cette manière, TWSVM détermine l'hyper-plan pour chaque classe et un nouvel échantillon de données est affecté à une classe ' i ' à l'aide de la fonction de décision suivante :

$$\text{Class } i = \min |x^T w_i + b_i| \text{ for } i = 1, 2 \quad (\text{II.30})$$

La distance perpendiculaire de l'échantillon de données de test est calculée à partir de chaque hyper-plan et le motif est attribué à la classe à partir de laquelle sa distance est inférieure.

TWIN-SVM non Linear

TWSVM classe également les échantillons de données difficiles à classer en utilisant des limites de classe linéaires. Pour faciliter la séparation de ce type d'échantillons de données, les fonctions du noyau sont utilisées pour mapper les échantillons de données dans un espace de fonctions de dimension supérieure [32]. Le problème primordial de la TWSVM non linéaire est formulé de la manière suivante:

$$\begin{aligned} \min(u_1, b_1, \varepsilon) \quad & \frac{1}{2} \|K_1(X_1, D^T)u_1 + e_1 b_1\|^2 + c_1 e_2^T \varepsilon \\ \text{s. t.} \quad & -K(X_2, D^T)u_1 + e_2 b_1 + \varepsilon \geq e_2, \varepsilon \geq 0 \\ \text{Et} \end{aligned} \quad (\text{II.31})$$

$$\begin{aligned} \min(u_2, b_2, \eta) \quad & \frac{1}{2} \|K(X_2, D^T)u_2 + e_2 b_2\|^2 + c_2 e_1^T \eta \\ \text{s. t.} \quad & -K(X_1, D^T)u_2 + e_1 b_2 + \eta \geq e_1, \eta \geq 0 \end{aligned}$$

où $D = [X_1 \ X_2]^T$ et la fonction du noyau est indiquée par ' K '. Les surfaces des noyaux dans un espace de grande dimension sont formulées comme suit:

$$K(x^T, D^T)u_1 + b_1 = 0 \text{ et } \ll. \gg + b_2 = 0 \quad (\text{II.32})$$

$$L(u_1, b_1, \varepsilon, \beta) = \frac{1}{2} \|K(X_1, D^T)u_1 - e_1 b_1\|^2 + c_1 e_2^T \varepsilon + \alpha^T ((K(X_2, D^T)u_1 - \varepsilon + e_2) - \beta^T \varepsilon$$

Le vecteur normal et le biais pour deux hyper-plans non parallèles sont obtenus comme :

$$\begin{aligned} Z_1 = \begin{bmatrix} u_1 \\ b_1 \end{bmatrix} &= -(P^T P)^{-1} Q^T \alpha \\ Z_2 = \begin{bmatrix} u_2 \\ b_2 \end{bmatrix} &= -(Q^T Q)^{-1} P^T \alpha \end{aligned} \quad (\text{II.33})$$

où $P = [k(X_1, D^T)e_1]$ et $Q = [k(X_2, D^T)e_2]$.

Le nouvel échantillon de données est classé comme :

$$\text{Class } i = \min |K(x^T, D^T)u_i + b_i| \text{ for } i = 1, 2 \quad (\text{II.34})$$

Pour un nouvel échantillon de données, sa distance est mesurée à partir des deux surfaces du noyau et est attribuée à la classe à partir de laquelle sa distance est plus petite

II .5. CONCLUSION

Ce chapitre a fait l'objet de rappels des fondements des méthodes et techniques d'apprentissage statistique appliqués à la classification constituant notre système de reconnaissance de pathologie. En premier, nous avons rappelé les concepts fondamentaux de la sélection des caractéristiques. Un exposé assez exhaustif de quelques méthodes appliquées à la sélection des caractéristiques. La première étape consiste à extraire des caractéristiques plus pertinentes et plus discriminantes tout en gagnant l'espace et le temps de calcul. Cependant, la deuxième étape a pour objectif d'affecter chaque ensemble de caractéristiques à une classe appropriée. Dans le chapitre suivant, nous testons tout ce que nous avons vu dans ce chapitre afin d'évaluer les techniques optées et plusieurs expériences seront conduites pour maximiser les performances du système.

Le chapitre suivant est consacré totalement à la simulation, est dédié à l'évaluation des performances des techniques étudiées. L'objectif, vise l'étude et la contribution de ces méthodes en vue de leur application dans le domaine de reconnaissance de pathologie.

Chapitre III
Résultats expérimentaux

III.1. INTRODUCTION

Ce dernier chapitre est dédié spécifiquement à la mise en œuvre et l'évaluation de notre système proposé qui se base sur l'application de Machines à Vecteurs de Supports (SVMs) avec ses variantes sur trois différentes bases de données biomédicales réelles (Heart disease, Breast cancer et Ovarian cancer). Les expériences effectuées dans cette partie subdivisent sur deux axes principaux. Le premier consiste à évaluer la classification des pathologies en utilisant plusieurs variantes des SVMs. Par contre, le deuxième axe consiste à appliquer une étape de réduction de dimension afin de réduire un peu le nombre des caractéristiques requis pour le traitement. La technique utilisée dans cette phase est l'Analyse en Composantes Voisines (NCA: Neighborhood Component Analysis). Également, NCA permet de sélectionner les caractéristiques les plus pertinentes tout en éliminant celle qui peuvent augmenter la redondance entre les variables des données. Les résultats de simulation est aussi bien décrit pour bien évaluer les performances de la méthode présentée. Afin d'optimiser les paramètres de chaque technique utilisée avant et après la sélection des caractéristiques, nous ferons plusieurs testes. Finalement, une étude récapitulative clôture le présent chapitre.

III.2. BASE DE DONNÉES UTILISÉE

Afin de bien évaluer notre système proposé, trois différentes bases de données ont été exploitées pour sortir les performances globales. Ces données viennent de sources biomédicales réelles. Dans la suite, nous décrivons les caractéristiques de chaque base de données.

III.2.1. Heart Disease

Cette base de données contient 76 attributs mais toutes les expériences publiées font référence à l'utilisation d'un sous-ensemble de 14 d'entre eux. En particulier, cette base de données est la seule qui ait été utilisée par les chercheurs du Machine Learning (ML) à ce jour. Le champ "objectif" représente l'étiquette de classe, il fait référence à la présence d'une maladie cardiaque chez le patient. C'est un entier valant de 0 (pas de présence) à 4 [34, 35].

N° de paramètre	Abréviation	signification
1	Age	âge en années
2	sexe	1 = male; 0 = femelle
3	cp	type de douleur thoracique
4	trestbps	tension artérielle au repos
5	chol	cholestérol sérique en mg/dl
6	fbs	glycémie à jeun > 120 mg / dl) (1 = vrai; 0 = faux
7	restecg	Résultats électro cardiographiques au repos
8	thalach	fréquence cardiaque maximale atteinte
9	exang	angine de poitrine induite par l'exercice (1 = oui; 0 = non)
10	oldpeak	Dépression ST induite par l'exercice relatif au repos
11	Slope	la pente du segment ST d'exercice de pointe
12	ca	nombre de vaisseaux majeurs (0-3)
13	thal	3 = normal; 6 = défaut fixe; 7 = défaut réversible
14	num	Caractéristique (attribut) prévu

Tableau III.1. Informations d'attributs de la base Heart Disease.

III.2.2. Brest cancer

Cet ensemble de données peut être utilisé pour concevoir l'état du patient, en fonction des caractéristiques des échantillons de biopsies [36].

✓ Cancer Inputs - une matrice 9x699 définissant neuf attributs de 699 biopsies :

N° de paramètre	Signification
1	Épaisseur des mottes
2	Uniformité de la taille de la cellule
3	Uniformité de la forme des cellules
4	Adhérence marginale
5	Taille unique des cellules épithéliales
6	Noyaux nus
7	Bland Chomatin
8	Nucléoles normales
9	Mitoses

Tableau III.2. Informations d'attributs de la base Brest Cancer.

- ✓ Cancer Target - Matrice 2x699 où chaque colonne indique une catégorie correcte avec l'élément 1 ou l'élément 2:
 - 1 = Sain (Tumeur bénigne).
 - 2 = Cancéreux (Tumeur maline).

III.2. 3. Ovarien cancer

Cette base de données a été construite pour l'objectif de détecter le cancer à partir de données de spectrométrie de masse sur des profils de protéines. Ses caractéristiques sont les niveaux d'intensité des ions à des valeurs spécifiques de masse / charge. Le défi de cette base est de concevoir un classifieur capable de distinguer les patients cancéreux et les patients sains par le contrôle des données de spectrométrie de masse sur des profils de protéines [37, 38].

Les données de cet exemple proviennent de la banque de données du programme de clinique protéomique FDA-NCI. Les données contiennent les variables Y, MZ et grp, chaque colonne de Y représentant 216 patientes, parmi lesquelles 121 sont des patientes atteintes de cancer de l'ovaire et 95 des patientes normales, chaque ligne de Y représente le niveau d'intensité des ions à une valeur de charge en masse spécifique indiquée en MZ. Il existe 15 000 valeurs de charge en masse en MZ et chaque ligne en Y représente les niveaux d'intensité ionique des patients à cette valeur de charge en masse particulière. La variable grp contient les informations d'index indiquant lesquels de ces échantillons représentent des patients cancéreux et lesquels représentent des patients normaux [37, 38].

Cet ensemble de données peut être utilisé pour concevoir un réseau de neurones permettant de classer les patientes selon qu'elles sont atteintes d'un cancer de l'ovaire ou non .

- ✓ Cancer Inputs - une matrice 100x216 définissant 100 attributs pour 216 patients différents.
- ✓ Cancer Target - une matrice 2x216 où chaque élément indique un patient atteint de cancer de l'ovaire avec un [1; 0], ou un patient normal avec un [0; 1]. Il y a 121 patients cancéreux et 95 patients normaux.

III.3. PROTOCOLE EXPÉRIMENTAL

Après avoir décrit les caractéristiques de chaque base de données. Maintenant, l'opération suivante consiste à implémenter un modèle et obtenir des résultats sous forme de probabilité ou de classe. L'étape suivante consiste à déterminer l'efficacité du modèle basé sur une métrique utilisant des jeux de données de test. Différentes mesures de performance sont utilisées pour évaluer différents algorithmes d'apprentissage automatique, métrique pour l'évaluation d'algorithmes d'apprentissage automatique est la précision (en Anglais : *Accuracy*), erreurs, intervenons pour voir ce que sont ces métriques.

III.3.1. Taux de classification

Le taux de classification est le nombre de prédictions correctes faites par le modèle sur toutes les sortes de prédictions effectuées.

		Actual	
		Positives(1)	Negatives(0)
Predicted	Positives(1)	TP	FP
	Negatives(0)	FN	TN

Fig.III.1. Taux de classification illustré dans la matrice de confusion.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} * 100\% \quad (\text{III.1})$$

Dans le numérateur, nos prédictions correctes (vrais positifs et vrais négatifs) (marquées en rouge dans la figure ci-dessus) et dans le dénominateur, sont le genre de toutes les prédictions faites par l'algorithme (correctes ou non).

III.3.2. Taux d'erreur

L'erreur dans les problèmes de classification est le nombre de prédictions incorrectes faites par le modèle sur tous les types de prédictions effectuées.

		Actual	
		Positives(1)	Negatives(0)
Predicted	Positives(1)	TP	FP
	Negatives(0)	FN	TN

Fig. III.2. Taux d'erreur illustré dans la matrice de confusion.

$$\text{Erreur} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} * 100\% \quad (\text{III.2})$$

Dans le numérateur, nos prédictions incorrectes (faux positifs et faux négatifs) (marquées en rouge dans la figure ci-dessus) et dans le dénominateur, sont le genre de toutes les prédictions faites par l'algorithme (correctes ou non).

Termes associés à la matrice de confusion :

Pour bien comprendre les quatre terminologies principales d'une matrice de confusion :

- **Vrai positif VP (True Positives TP)** : les cas où la prédiction est positive, et où la valeur réelle est effectivement positive.
- **Vrai négatif VN (True Négatives TN)** : les cas où la prédiction est négative, et où la valeur réelle est effectivement négative.
- **Faux positif FP (False Positive FP)** : les cas où la prédiction est positive, mais où la valeur réelle est négative.
- **Faux négatif FN (False Négative FN)** : les cas où la prédiction est négative, mais où la valeur réelle est positive.

III.4. RÉSULTATS DE SIMULATION

III.4.1. Sans sélection des caractéristiques

Dans cette section, notre objectif est de décortiquer les Machines à Vecteurs de Supports (SVMs) avec leurs variantes pour classifier efficacement les pathologies. Les tests sont appliqués sur trois différentes bases de données biomédicales réelles pour évaluer les performances de chaque machine. Les Tableaux III.3, III.4 et III.5 présentent les résultats de classification en utilisant différentes variantes de SVMs.

III.4.1.1. Avec la base Heart Disease

Dans cette partie, nous montrons les résultats obtenus de la base Heart Disease avec les variantes SVM. Tableau III.3 résume les performances obtenues pour chaque variante de SVM ainsi que ses paramètres optimaux (les paramètres des noyaux) avec la base de données Heart Disease.

Variante	Paramètres	Accuracy Train %	Time train(ms)	Accuracy Test %	Time test(ms)
SVM-Linear	$C = 1$	88,079	15	84.105	0
SVM-Polynomial	$C = 1$ $\beta = 0$ $\delta = 1$ $\alpha = 5$	86.754	27	83.443	13
SVM-Rbf	$C = 420$ $\beta = -1$	97,3505	3	75,496	3
SVM-Laplacian	$C = 1000$ $\beta = -1$	100	14	64,238	12
SVM-Sigmoid	$C = 1000$ $\beta = 0$ $\alpha = 5$	53,642	15	55,629	10
SVM-Lagrangian	$C = 2200$	99.337	15	72.847	0
SVM-Normalized Poly	$C = 1000$ $\beta = 0$ $\delta = 2$ $\alpha = 5$	53,642	16	55,629	12
Twin SVM (Linear)	-	89.403	55	86.754	47

Tableau III.3. Résultats obtenu de la base Heart disease avec les différentes variantes de SVM.

Ensuite, nous avons comparé les sorties réelle avec les sorties de classificateurs SVM dans la base d'apprentissage et de test. La figure III.3 rapporte la correspondance entre les deux sorties.

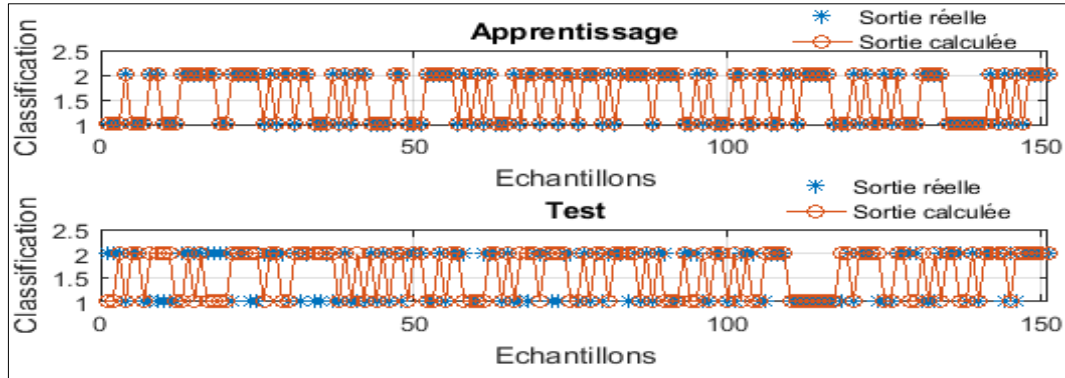


Fig. III.3. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-Laplacien sur la base Heart disease.

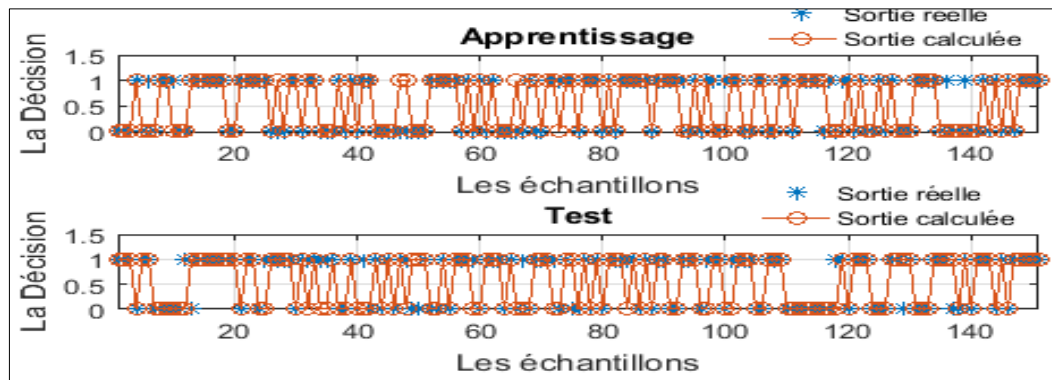


Fig. III.4. Sorties réelle et calculée pour l'apprentissage et le test avec TWIN-SVM sur la base Heart disease.

Après avoir ajusté les paramètres de chaque variante (noyau) de SVM pour améliorer les résultats dans le Tableau (III.3), nous remarquons que le meilleur taux d'apprentissage est obtenu avec SVM-Laplacien où 100% a été achevé (Fig. III.3). Par contre, Twin SVM (Linear) donne le meilleur taux de test avec 89.40%. Cette variante présente une précision importante dans le test comme il est montré dans la figure (III.4).

III.4.1.2. Avec la base Breast Cancer

Dans cette partie, nous montrons les résultats obtenus de la base Breast Cancer avec les variantes SVM. Tableau III.4 résume les performances obtenues pour chaque variante de SVM ainsi que ses paramètres optimaux (les paramètres des noyaux) avec la base de données Breast Cancer.

Variante	Paramètre	Accuracy Train %	Time train (ms)	Accuracy Test %	Time test (ms)
SVM-Linéar	$C=4300$	96,275	16	98,567	0
SVM-Polynomial	$C = 1500$ $\beta = 0$ $\delta = 2$ $\alpha = 5$	97,13	5	97,994	5
SVM-Rbf	$C = 600$ $\beta = 0$	97,13	31	98,280	16
SVM-Laplacian	$C = 1000$ $\beta = 0$	100	31	98,567	15
SVM-Sigmoid	$C = 1000$ $\beta = 2$ $\alpha = 5$	95,98	31	98,567	31
SVM-Lagrangian	$C = 1000$	100	13	95.415	12
SVM-Normalized Poly	$C = 1000$ $\beta = 1$ $\delta = 2$ $\alpha = 5$	65.61	22	65.329	15
SVM-Twin SVM (Linear)	-	95.41	147	97.421	139

Tableau III.4. Résultats obtenu de la base Breast Cancer avec les différentes variantes de SVM.

Nous avons aussi comparé les sorties réelle avec les sorties de classificateurs SVM dans la base d'apprentissage et de test. La figure III.5 rapporte la correspondance entre les deux sorties.

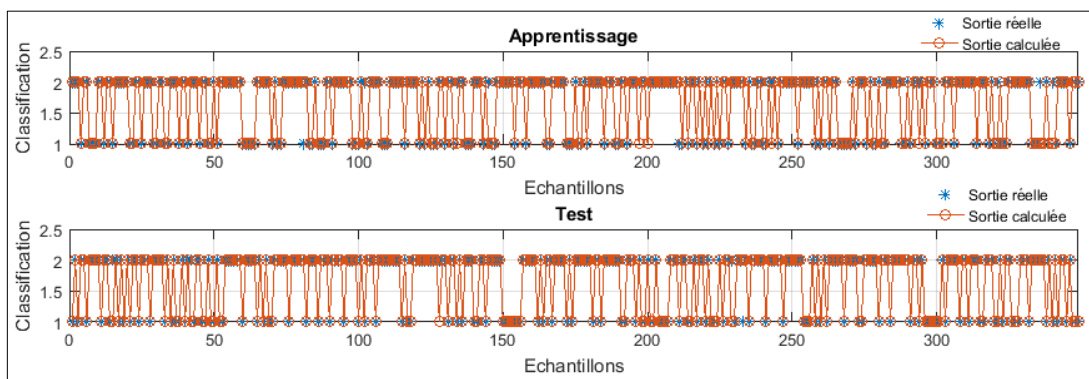


Fig. III.5. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-Linéar sur la base Breast Cancer.

Avec cette base de données, le meilleur taux d'apprentissage est obtenu avec SVM-Laplacien et Lagrangian. Mais le bon taux de test est obtenu avec SVM (Linear, Laplacien et Sigmoid). Puisque le taux de test est similaire pour ces trois variantes nous choisissons celle qui a le temps d'exécution le plus faible. À partir de tableau III.4 et en termes de temps d'exécution, nous prenons comme un meilleur classificateur le SVM-linear.

III.4.1.3. Avec la base Ovarian Cancer

Dans cette partie, nous montrons les résultats obtenus de la base Ovarian Cancer avec les variantes de SVMs.

Variante	Paramètre	Accuracy train	Time train (ms)	Accuracy test	Time test (ms)
SVM-Linear	$C = 1000$	100	15	97,19	15
SVM-Polynomial	$C = 1000$ $\beta = 0$ Degré=3 $\alpha = 0$	100	63	98.13	31
SVM-Rbf	$C = 1200$ $\beta = 0$	100	16	99,06	15
SVM-Laplacian	$C = 1000$ $\beta = 0$	100	16	96,26	15
SVM-Sigmoid	$C = 1340$ $\beta = 0$ $\alpha = 5$	82,40	16	87,85	16
SVM-Lagrangian	$C = 1000$	100	15	92,52	15
SVM-Normalized Poly	$C = 1$ $\beta = -1$ $\delta = 3$ $\alpha = 0$	55.55	118	57.00	58
Twin SVM (Linear)	-	99,07	26	99.06	15

Tableau III.5. Résultats obtenu de la base Ovarian Cancer avec les différentes variantes de SVM.

Selon le tableau III.5, la performance de la base d'apprentissage a atteint le maximum dans la majorité des variantes de SVMs. Néanmoins, le meilleur taux de test est obtenu avec la variante SVM-RBF qui donne 99.06% . La figure III.6 montre la superposition

des deux sorties (réelles et calculées), ce qui s'interprète par la précision de ce type de SVM.

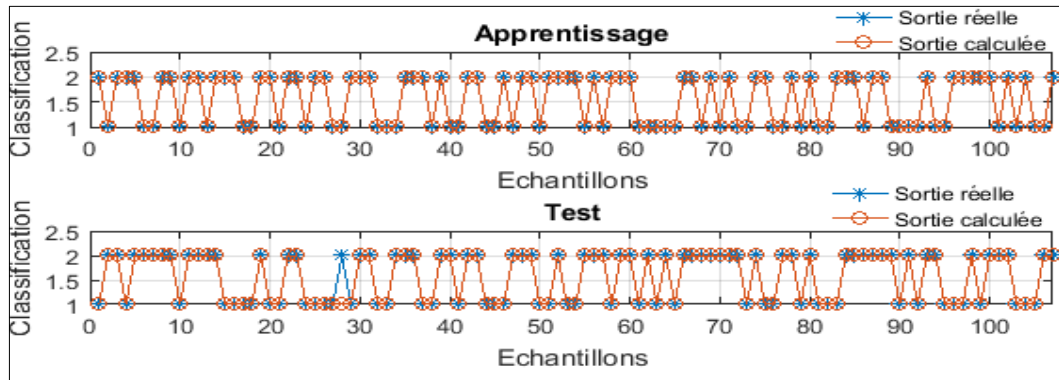


Fig. III.6. Sorties réelle et calculée pour l'apprentissage et le test avec SVM-RBF sur la base Ovarian Cancer.

III.4.2. Avec sélection des caractéristiques

La méthodologie suivie dans cette partie consiste à sélectionner un ensemble réduit de caractéristiques qui peuvent être utilisées pour bien classifier les pathologies.

III.4.2.1. Avec la base Heart Disease

Dans cette expérience, nous appliquons l'algorithme de sélection des caractéristiques NCA sur notre base de données. Pour l'évaluation, nous rajoutons à chaque test une nouvelle caractéristique aux vecteurs descripteurs tout en respectant l'ordre d'importance. La Figure III.7 et le Tableau III.6 montrent le nouveau ordre des caractéristiques obtenu lorsque nous appliquons le NCA sur la base de données Heart Disease.

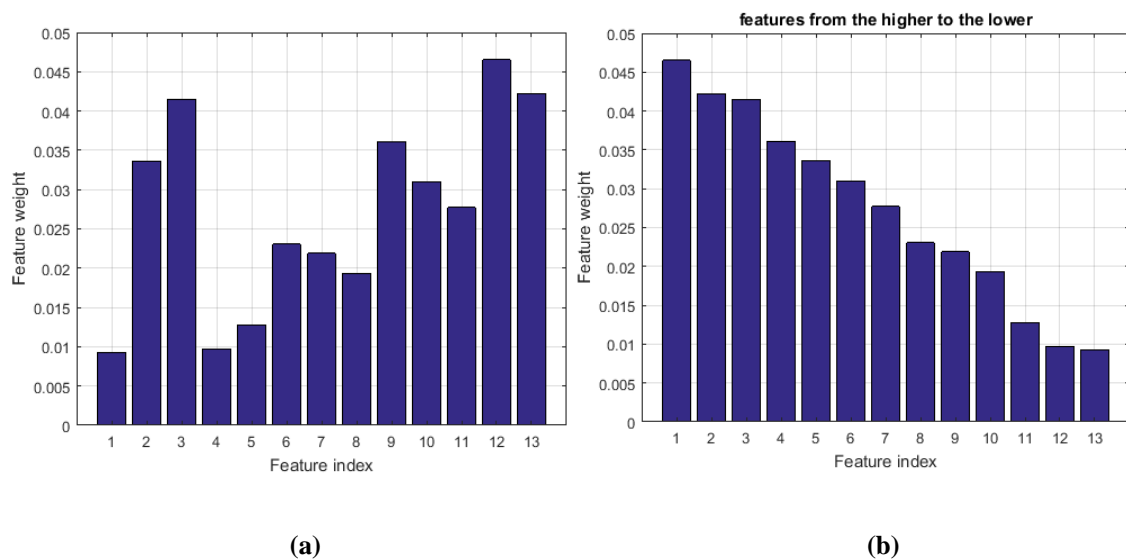


Fig. III.7. Application de NCA sur les données: a) Ordre avant NCA et b) Ordre après NCA.

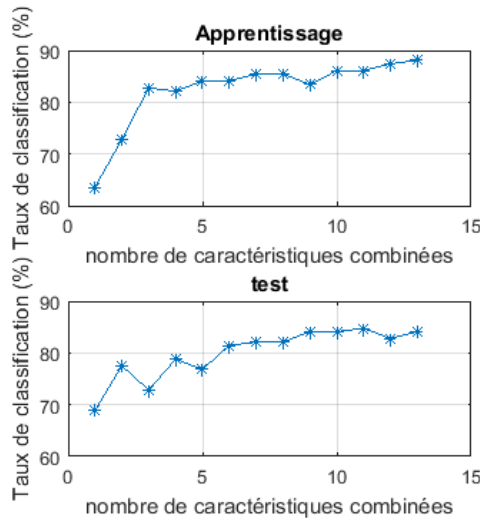
Ordre des caractéristiques avant NCA	Ordre des caractéristiques après NCA
1 : age	12 : ca
2 : sexe	3 :cp
3 :cp	13 :thal
4 : trestbps	9 :exang
5 : chol	2 :sexe
6 :fbs	10 :oldpeak
7 : restecg	11 :Slope
8 : thalach	6 :fbs
9 : exang	7 :restecg
10 : oldpeak	8 :thalach
11 : Slope	1 :age
12 : ca	4 :trestbps
13 : thal	5 :chol

Tableau III.6. Ordre des caractéristiques avant et après l'application de NCA sur la base Heart Disease.

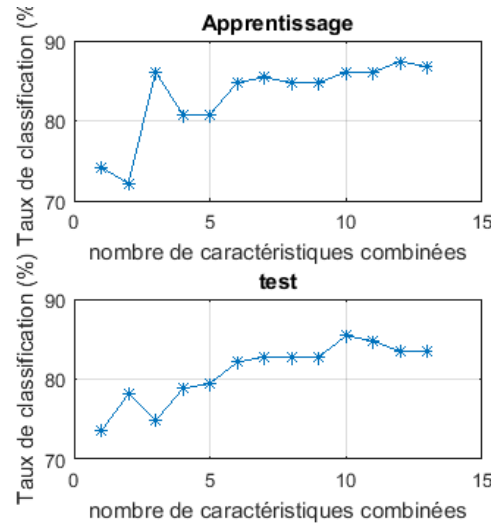
À partir de la Figure III.7. et le tableau III.6, nous remarquons que la caractéristique (nombre de vaisseaux majeurs) à une valeur de poids plus élevée par rapport à les autres. Ça vaut dire que celui-ci est le plus représentatif parmi les autres. Afin de vérifier cet ordonnancement, nous avons fait plusieurs tests en introduisant différentes combinaisons de caractéristiques selon l'ordre d'importance pour toutes les variantes SVM.

Variante	Configuration	N° Indice	Accuracy train	Time train (ms)	Accuracy test	Time test (ms)
SVM-Linear	C =1	11	86.092	4	84.768	4
		13	88,079	15	84.105	0
SVM-Polynomial	C=1 $\beta =0$ $\delta =1$ $\alpha =5$	10	86.092	9	85.430	6
		13	86.754	27	83.443	13
SVM-Rbf	C = 420 $\beta =-1$	5	92.052	3	79.470	3
		13	97,350	15	75,496	15
SVM-Laplacian	C =1000 $\beta =-1$	8	98.013	15	83.443	0
		13	100	14	64,238	12
SVM-Sigmoid	C= 1000 $\beta =0$ $\alpha =5$	2	72.847	15	77.483	0
		13	53,642	15	55,629	10
SVM-Lagrangian	C = 2200	3	89.403	0	79.470	0
		13	99.337	15	72.847	0
SVM-Normalized POLY	C =1000 $\beta =0$ $\delta =2$ $\alpha =5$	1	53,642	31	55,629	15
		13	53,642	16	55,629	12
Twin SVM (Linear)	-	10	86.754	15	84.768	11
		13	89.403	55	86.754	47

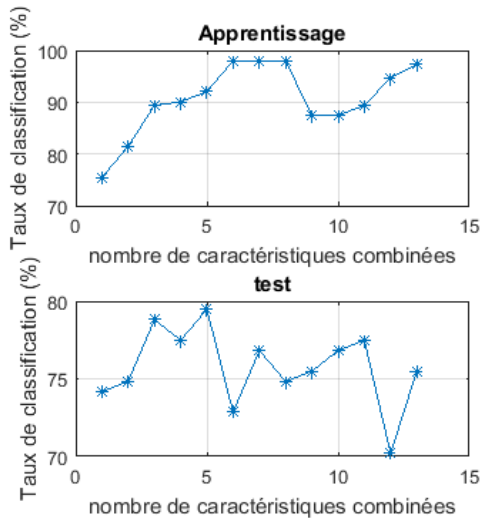
Tableau III.7. Meilleure performance obtenu avec la sélection des caractéristiques de la base Heart Disease.



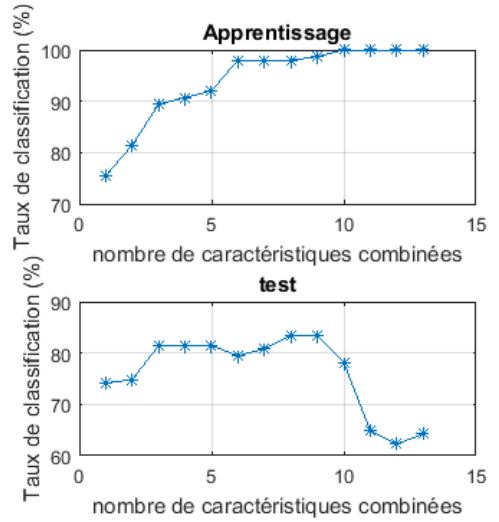
(a) Linear



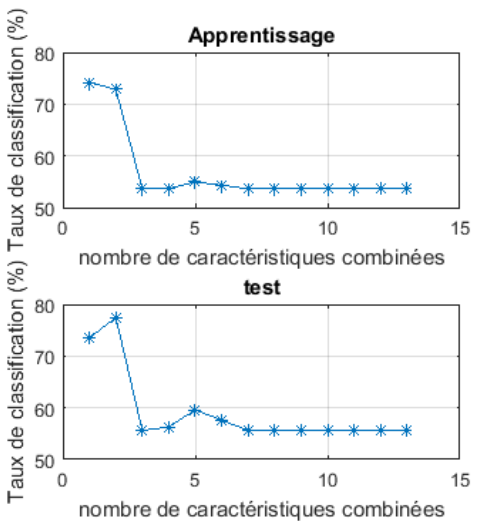
(b) Polynomial



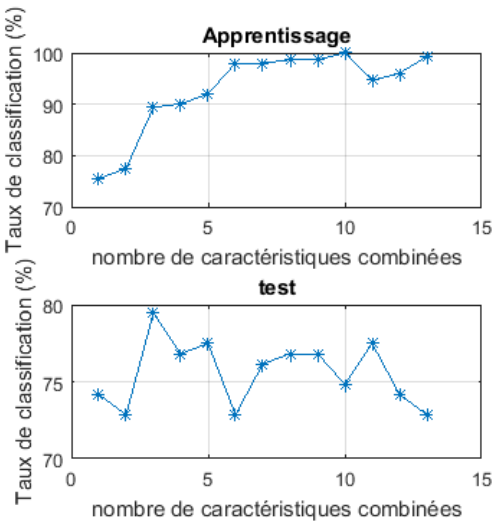
(c) RBF



(d) Laplacian



(e) Sigmoid



(f) Lagrangian

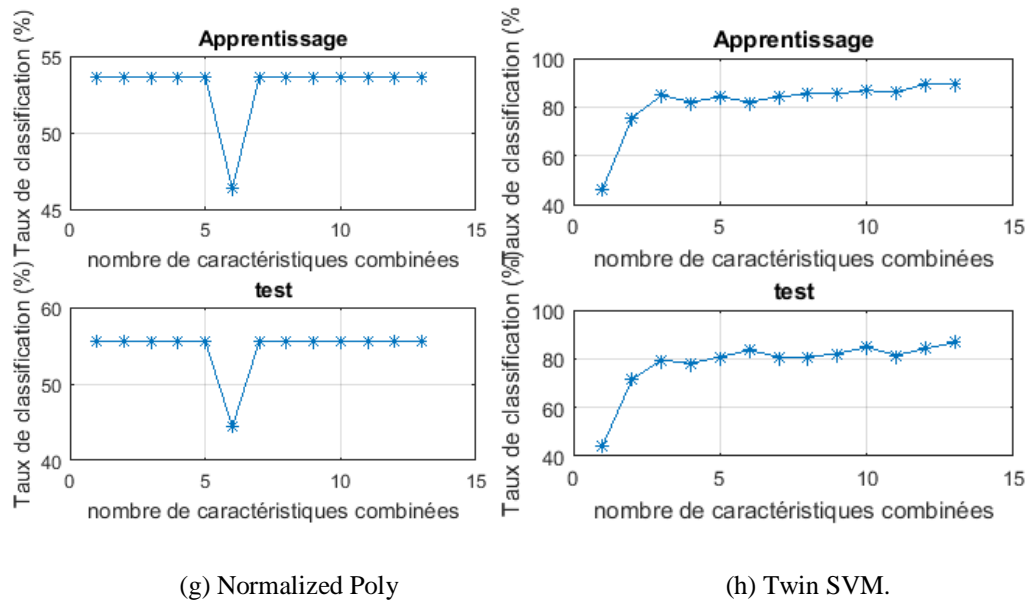


Fig.III.8. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Heart Disease utilisant les différentes variantes de SVM.

D'après le tableau III.7, nous trouvons que les performances de classification ont connu des améliorations importantes lorsque en utilisant la sélection des caractéristiques avec toutes les variantes de SVM sauf la dernière (Twin-SVM). Nous avons arrivé à réduire la taille des caractéristiques tout en gardant un taux de test considérable. Par exemple, avec SVM-RBF nous avons atteint un taux de test égale à 79.74% en utilisant les 5 premières caractéristiques. Par contre, toute les caractéristiques donnent un taux de test égale à 75,496% avec le même type de SVM. Aussi, le temps d'exécution avec 5 caractéristiques est évidemment plus rapide que tout l'ensemble.

III.4.2.2. Avec la base Breast cancer

Dans cette expérience, nous appliquons l'algorithme de sélection des caractéristiques NCA sur la base Breast Cancer. Pour l'évaluation, nous rajoutons à chaque test une nouvelle caractéristique aux vecteurs descripteurs tout en respectant l'ordre d'importance. La Figure III.9 et le Tableau III.8 montrent le nouveau ordre des caractéristiques obtenu lorsque nous appliquons le NCA sur la base de données Breast Cancer.

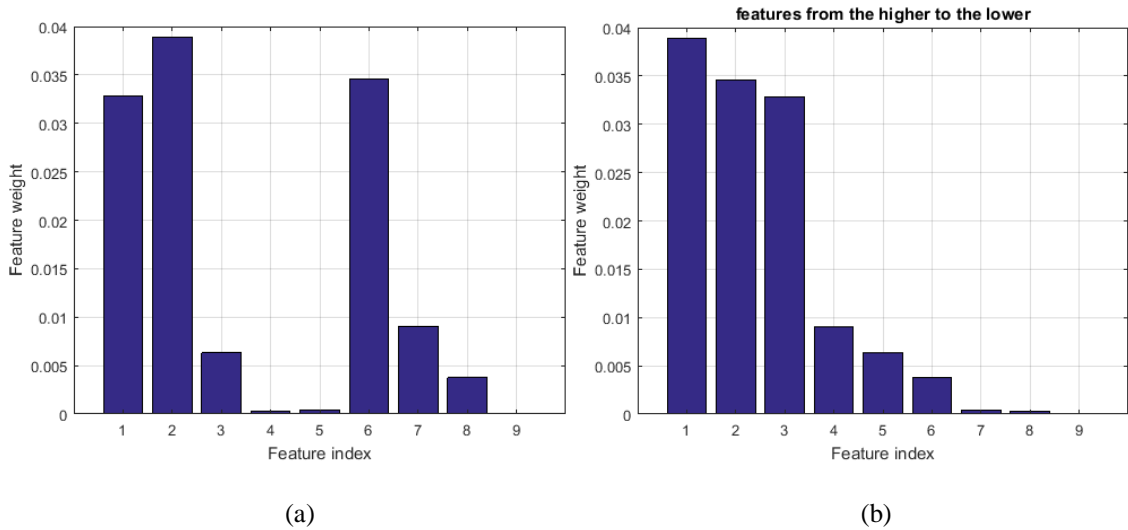


Fig. III.9. Application de NCA sur les données de la base Breast Cancer: a) Ordre avant NCA et b) Ordre après NCA.

Ordre des caractéristiques avant la sélection	Ordre des caractéristiques après la sélection
1 : Épaisseur des mottes	2 : Uniformité de la taille de la cellule
2 : Uniformité de la taille de la cellule	6 : noyaux nus
3 : Uniformité de la forme des cellules	1 : Épaisseur des mottes
4 : Adhérence marginale	7 : Bland Chomatin
5 : Taille unique des cellules épithéliales	3 : Uniformité de la forme des cellules
6 : noyaux nus	8 : nucléoles normales
7 : Bland Chomatin	4 : Adhérence marginale
8 : nucléoles normales	5 : Taille unique des cellules épithéliales
9 : Mitoses	9 : Mitoses

Tableau III.8. Ordre des caractéristiques avant et après l'application de NCA sur la base Breast Cancer.

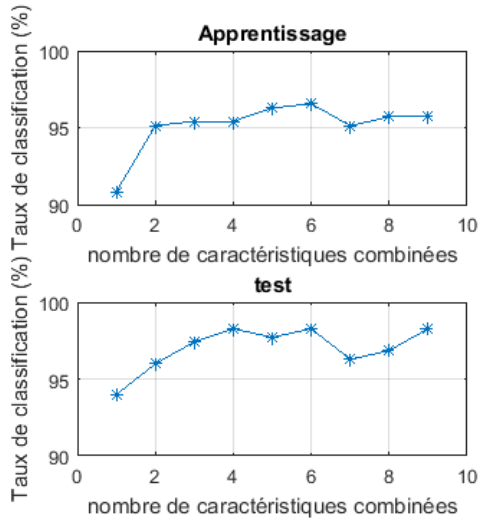
À partir de la Figure III.9. et le tableau III.8, nous remarquons que l'ordonnement des caractéristiques a été changé selon le poids de voisinage pour chaque une. Grâce à NCA, nous pouvons connaître quel est le sous-ensemble le plus pertinents parmi l'ensemble total. La caractéristique (uniformité de la taille de la cellule) possède la valeur de poids la plus élevée par rapport à les autres. Ça vaut dire que cette caractéristique est le plus représentatif parmi les autres. Afin de fixer la taille de sous-ensemble le plus représentatif, nous avons fait plusieurs tests en variant le nombre des caractéristiques selon l'ordre d'importance pour toutes les variantes SVM.

Variante	Paramètre	N° Indice	Accuracy train	Time train	Accuracy Test (ms)	Time Test (ms)
SVM-Linear	C=4300	4	95.415	15	98.280	0
		9	96,275	16	98,567	0
SVM-Polynomial	C=1500 $\beta =0$ $\delta =2$ $\alpha =5$	5	96.275	15	97.707	0
		9	97,134	5	97,994	5
SVM-RBF	C = 600 $\beta =0$	4	95.988	15	97.134	15
		9	97,134	31	98,280	16
SVM-Laplacian	C = 1000 $\beta =0$	4	100	31	97.421	16
		9	100	31	98,567	15
SVM-Sigmoid	C = 1000 $\beta =2$ $\alpha =5$	8	95,988	31	98,567	16
		9	95,988	31	98,567	31
SVM-Lagrangian	C =1000	2	95.988	0	96.561	01
		9	100	13	95.415	12
SVM-Normalized Poly	C =1000 $\beta =1$ $\delta =2$ $\alpha =5$	1	65.616	47	65.329	31
		9	65.6160	22	65.329	15
SVM-Twin SVM (Linear)	-	5	95.415	33	96.561	32
		9	95.415	147	97.421	139

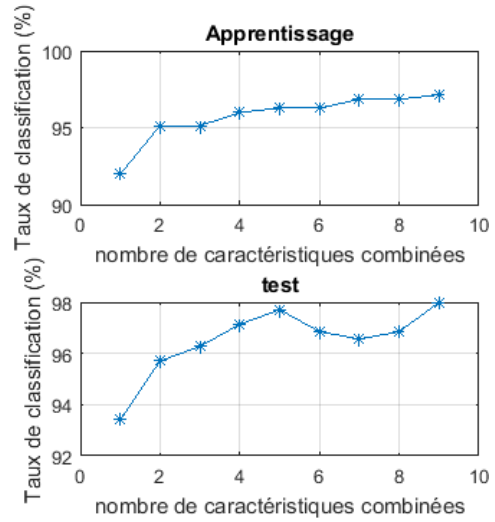
Tableau III.9. Meilleure performance obtenu avec la sélection des caractéristiques de la base Breast Cancer.

D'après le tableau III.9, les résultats de la classification avec sélection n'a pas vraiment apporté une amélioration importante dans cette base de données. Cela peut justifier au nombre de caractéristiques originale qui est très petit. Donc , il semble que toutes ces caractéristiques sont discriminantes. Excepte, avec la variante SVM-Lagrangien où nous avons eu une amélioration significative d'une marge de 1.15% avec une réduction de 7 caractéristiques.

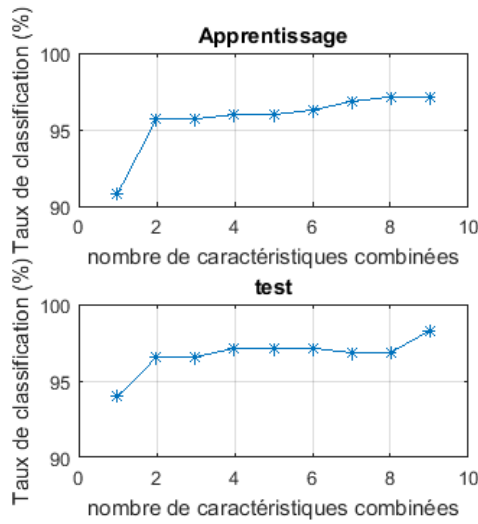
Également, les variations des performance du système (apprentissage et test) de classification de pathologie en fonction de nombre des caractéristiques combinées sur la base Breast Cancer en utilisant toutes les variantes de SVMs sont reportées dans la figure III.10.



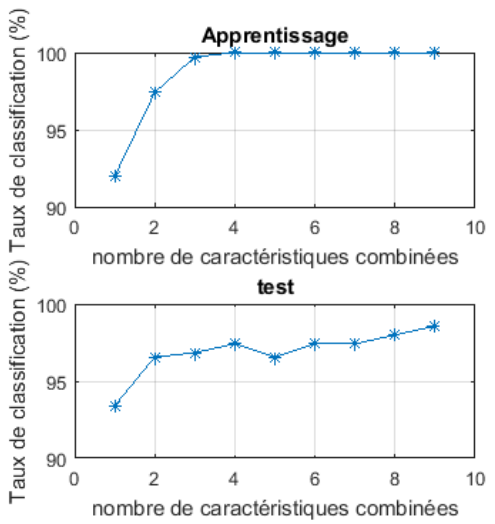
(a) Linear



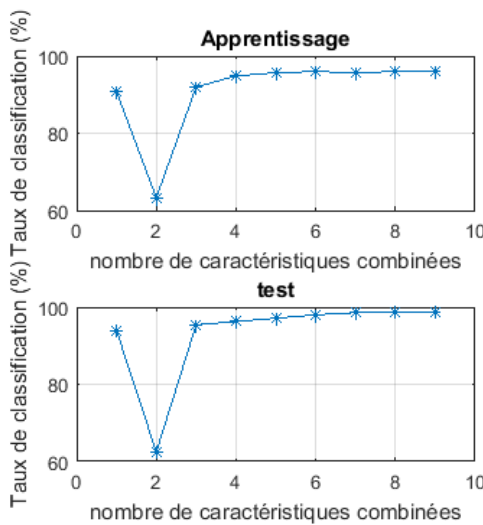
(b) Polynomial



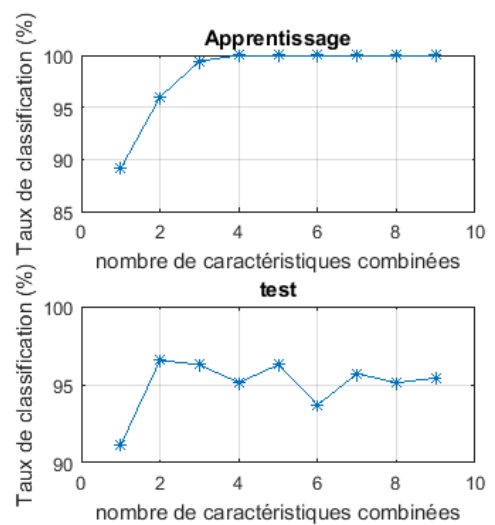
(c) RBF



(d) Laplacian



(e) Sigmoid



(f) Lagrangian

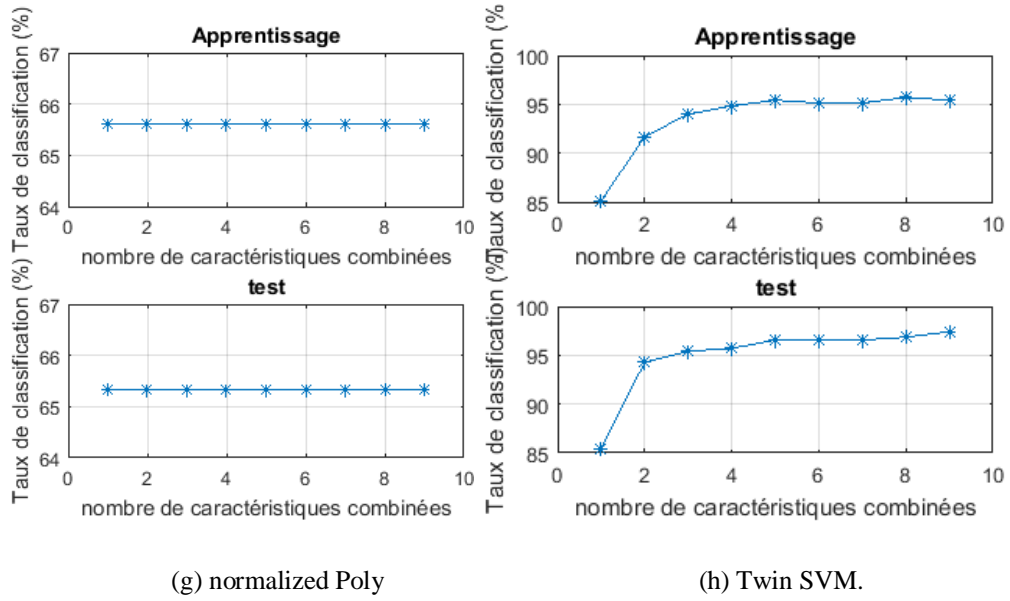


Fig.III.10. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Breast Cancer utilisant les différentes variantes de SVM.

III.4.2.3. Avec la base Ovarian Cancer

Comme dernière expérience, nous avons évalué la base de données Ovarian Cancer pour voir quel sous-ensemble dominant sur le reste des données. La Figure III.11 montre les résultats de l'ordonnancement selon le poids d'importance en utilisant l'algorithme de NCA.

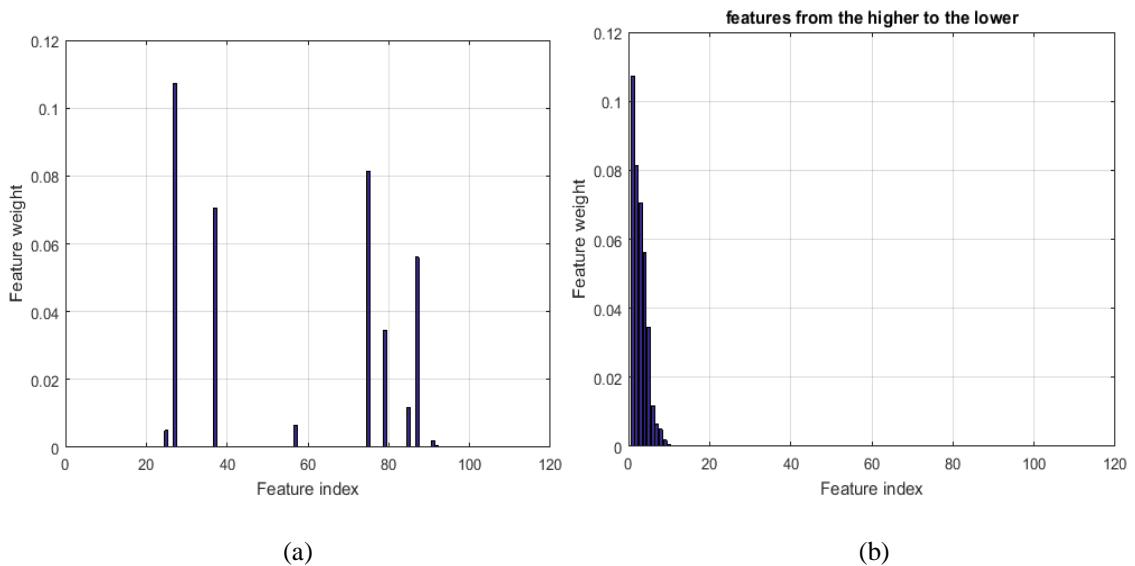


Fig. III.11. Application de NCA sur les données de la base Ovarian Cancer: a) Ordre avant NCA et b) Ordre après NCA.

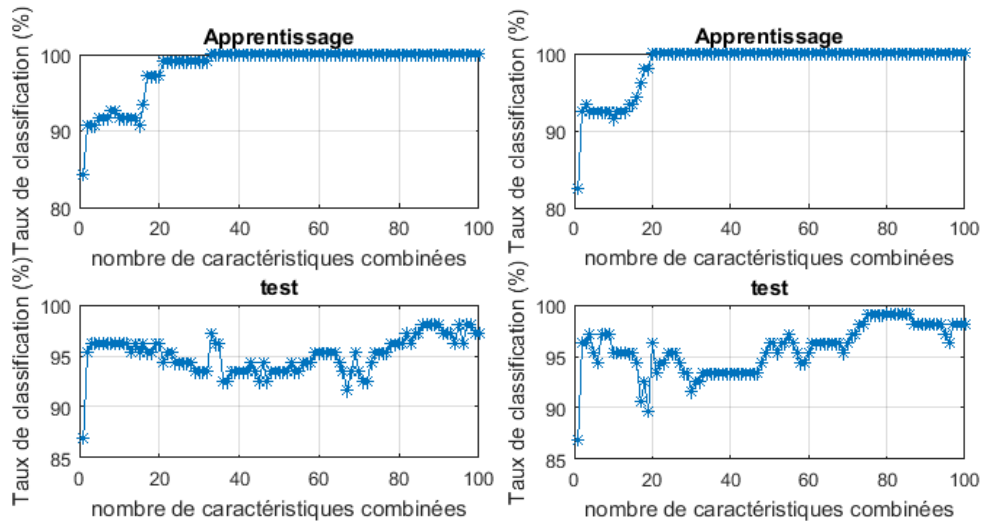
Variante	Paramètre	N° Indice	Accuracy train	Time train	Accuracy Test (ms)	Time Test (ms)
SVM-Linear	C=1000	86	100	31	98.130	47
		100	100	15	97,196	15
SVM-Polynomial	C =1000 $\beta =0$ $\delta =3$ $\alpha =0$	73	100	47	99.065	31
		100	100	63	98.130	31
SVM-RBF	C = 1200 $\beta = 0$	71	100	31	99,065	63
		100	100	15	99,065	16
SVM-Laplacian	C = 1000 $\beta =0$	64	100	31	97.196	78
		100	100	15	96,261	16
SVM-Sigmoid	C = 1340 $\beta =0$ $\alpha =5$	11	88.888	16	96.261	0
		100	82,407	16	87,850	16
SVM-Lagrangian	C =1000	57	100	31	97.196	16
		100	100	15	92,523	15
SVM-Normalized Poly	C =1 $\beta =-1$ $\delta =3$ $\alpha =0$	1	55.555	16	57.009	16
		100	55.555	58	57.009	118
Twin SVM (Linear)	-	61	96.296	7	100	6
		100	99.074	26	99.065	15

Tableau III.10. Meilleure performance obtenu avec la sélection des caractéristiques de la base Ovarian Cancer.

À partir de la Figure III.11, nous remarquons que les premières caractéristiques ont des poids importants contrairement au les dernières. Ça vaut dire que ces premières caractéristiques sont plus représentatives. Avec cette base de données, le NCA peut éliminer plusieurs.

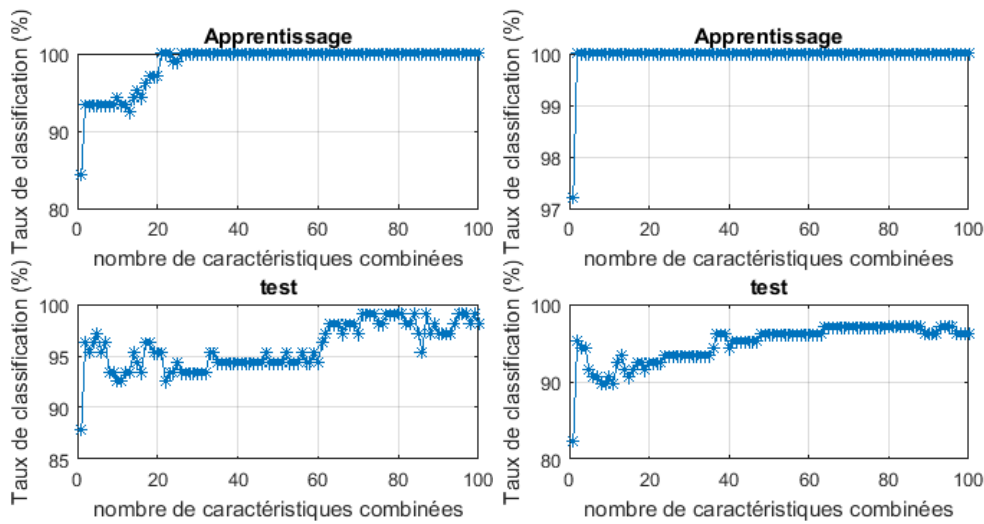
Le tableau III.10 reporte les résultats de la classification avec sélection des caractéristiques sur les données de la base Ovarian Cancer. Le taux de bonne classification a achevé le maximum 100% avec la variante TWIN-SVM en utilisant 61 caractéristiques seulement (presque 2/3 de toutes les données). Ce résultat montre l'efficacité de la sélection des caractéristiques dans une telle application biomédicale où les professionnels de la santé peuvent minimiser les moyens de collection des données en basant seulement sur les paramètres les plus représentatifs. Également, y a un avantage qui réside dans la réduction de la main d'œuvre et du temps de calcul pour sortir les résultats.

Outre, les variations des performances du système (apprentissage et test) de classification de pathologie en fonction de nombre des caractéristiques combinées sur la base Ovarian Cancer en utilisant toutes les variantes de SVMs sont reportées dans la figure III.12.



(a) Linear

(b) Polynomial



(c) RBF

(d) Laplacian

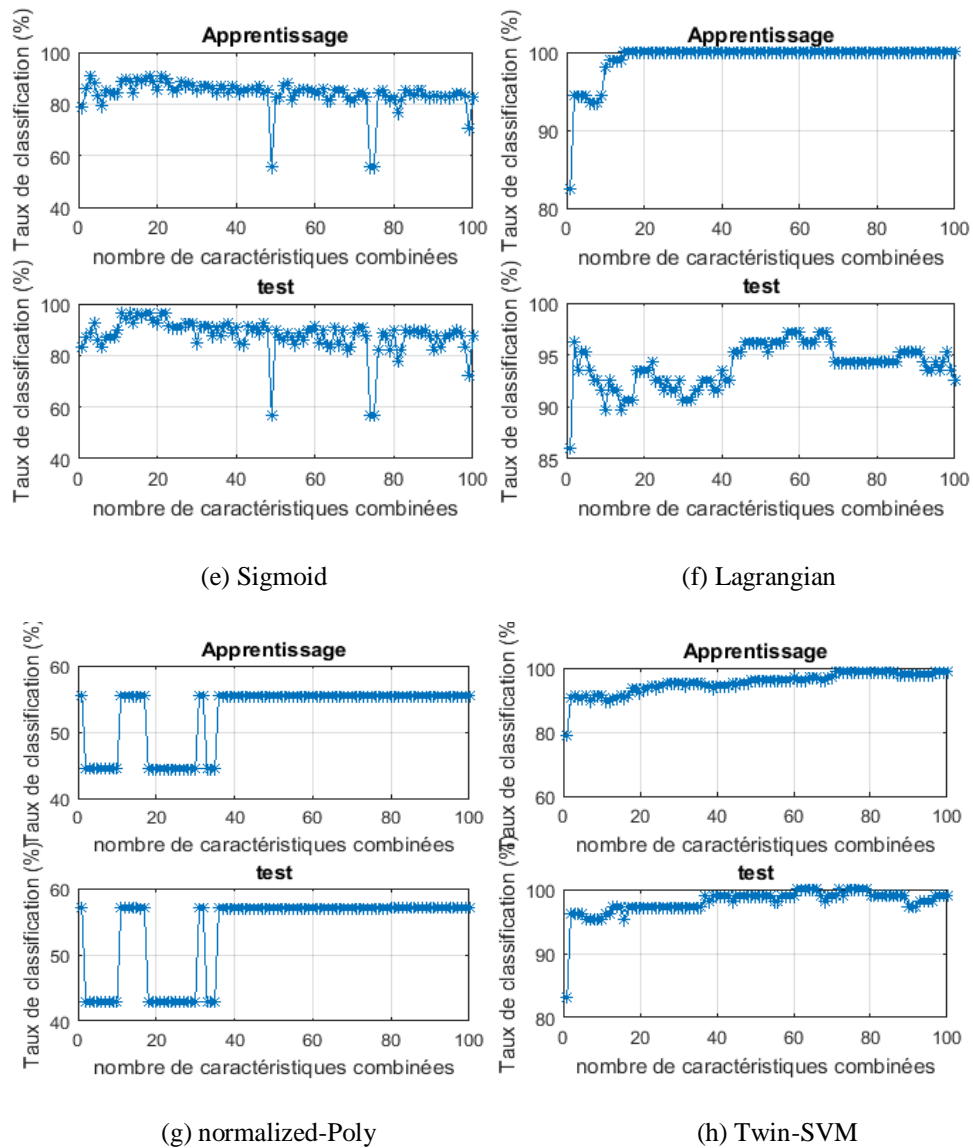


Fig.III.12. Variation de performance du système (apprentissage et test) en fonction de nombre des caractéristiques combinées sur la base Ovarian Cancer en utilisant les différentes variantes de SVM.

À partir des résultats illustrés dans la figure III.12, la sélection des caractéristiques est une étape très importante surtout avec une base de données de taille énorme. En effet, la sélection des caractéristiques retient les variables les plus représentatives des données. Elle peut être utile comme étape préliminaire à tout processus d'analyse et de classification de données afin de se concentrer sur les variables les plus importantes et réduire le coût d'exécution d'un tel processus

III.5. CONCLUSION

Ce dernier chapitre a fait l'objet d'une étude en simulation concernant la mise en œuvre de la technique d'apprentissage SVM appliquée dans la reconnaissance de pathologies, ainsi que la mise en œuvre de la technique NCA permettant de sélectionner les entrées discriminantes et parallèlement éliminer les informations redondantes. Il permet aussi de voir l'intérêt de la réduction de dimension. Cette étude a permis la validation et l'évaluation des performances de chacune des méthodes présentées.

Une étude détaillée a été faite dans le but d'un choix décisif de l'approche proposée. Les paramètres liés au taux de reconnaissance ont été les facteurs pertinents qui ont permis d'évaluer les méthodes étudiées. Les résultats obtenus ont montré que l'approche SVM-NCA donne des performances considérables en matière de classification de reconnaissance de pathologies.

Conclusion générale

CONCLUSION GENERALE

L'objectif de notre travail a porté sur l'implémentation d'un système de reconnaissance de pathologie en utilisant une technique de sélection des données (NCA) et un classifieur puissant (SVMs).

Dans la phase de sélection de données, nous avons présenté une technique qui se base sur l'analyse statistique (NCA) pour réduire au-maximum le nombre de capteurs utilisés. NCA a permis de déterminer les corrélations existantes entre les variables caractéristiques des bases puis de ne conserver que les caractéristiques apportant réellement une information pertinente.

Dans le module de la classification, nous avons évalué plusieurs variantes de SVMs afin de dégager les avantages et les inconvénients des uns des autres. Aussi, notre expérimentale a été testée plusieurs fois pour trouver les paramètres optimaux de chaque variante. Trois bases de données biomédicales réelles ont été exploitées pour valider le système proposé. Le taux de bonne classification a été pris comme un critère d'évaluation, et les résultats ont montré l'efficacité de la combinaison NCA/SVM.

Notre principale contribution dans ce travail est, d'une part, l'utilisation de la sélection des caractéristiques pour minimiser le nombre de traitements conduits ce qui fait la minimisation des équipements médicaux et le personnel correspondant. De l'autre part, SVM a présenté son efficacité en terme de temps de calcul et de performance pour prendre une décision importante concernant la pathologie.

Nous pensons que ce travail peut être amélioré en intégrant des algorithmes de l'apprentissage profond dans tel système pour augmenter les performances globales.

Bibliographie

Bibliographié

- [1] Begg, R., Lai, D. T., & Palaniswami, M. (2007). Computational intelligence in biomedical engineering. CRC Press.
- [2] R. Palaniappan, Biological Signal Analysis. Ventus Publishing, 2010.
- [3] Mediline Plus: Information related to Heart Diseases. Mayo Foundation for Medical Education and Research). Mai 2019.
- [4] Auteur inconnu, "Gouvernance de l'intelligence artificielle dans les entreprises: Stratégies, gouvernances et challenges de la data intelligence". CIGREF publication. Paris. Octobre 2018.
- [5] Coraline Cherbit, L'intelligence artificielle en plein boum. Dossier thématique, ARDI Rhone-Alpes, Février 2016.
- [6] A. Cornuéjols and L. Miclet, Apprentissage artificiel : Concepts et algorithmes, vol. 50, no. 5. 2003.
- [7] C. Sammut and G. I. Webb, Encyclopedia of Machine Learning, Springer. 2011.
- [8] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In Proceedings of the 5th annual workshop on Computational learning theory, Taylor & Francis, pp. 144-152, 1992.
- [9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine learning, 20(3) :273-297, 1995.
- [10] Klein, R. M. The Hebb legacy. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 53(1), 1. 1999.
- [11] Dimitri Siganos and Christos Stergiou. Neural networks. Web site, Consulted November 2014.
- [12] PACE University. "Rosenblatt's Contributions". Web site, 2011. Consulted Feb 2014.
- [13] PALIT, Ajoy K. et POPOVIC, Dobrivoje. Computational intelligence in time series forecasting: theory and engineering applications. Springer Science & Business Media, 2006.

- [14] GERSHENSON, Carlos. Artificial neural networks for beginners. arXiv preprint cs/0308031, 2003.
- [15] MOLINA, Luis Carlos, BELANCHE, Lluís, et NEBOT, Àngela. Feature selection algorithms: A survey and experimental evaluation. In : 2002 IEEE International Conference on Data Mining, 2002. Proceedings. IEEE, 2002. pp. 306-313.
- [16] Chandrashekar, Girish, and Ferat Sahin. "A survey on feature selection methods." Computers & Electrical Engineering 40.1 (2014): 16-28.
- [17] Philippe Biondi, Architecture expérimentale pour la détection d'intrusions dans un système informatique, Article de recherche, Sptembre, 2001.
- [18] Dash, M. et Liu, H. (1997). Feature selection for classification. Intelligent Data Analysis,1:131-156.
- [19] Yang, Wei, Kuanquan Wang, and Wangmeng Zuo. "Neighborhood Component Feature Selection for High-Dimensional Data." JCP 7.1 (2012): 161-168.
- [20] Webb, A.R. Statistical pattern recognition, Edition John Wiley & Sons, Londres,491 pages. 2002.
- [21] Arrivault, D. Apport des graphes dans la reconnaissance Non-contrainte de caractères manuscrits Anciens, Thèse de doctorat : Université de Poitiers, 200 pages. 2002.
- [22] Schwenk, S., Amélioration des classifieurs neuronaux par incorporation de connaissances explicites: Application à la reconnaissance de caractères manuscrits, Thèse de doctorat : Université de Pierre et Marie Curie, 148 pages. 1996.
- [23] Uchida, S., et Sakoe, H., Handwritten character recognition using elastic matching based on a class-dependent deformation model, ICDAR'03, 3-6, Edinburgh, Vol.
- [24] Simard, P., Le Cun, Y., Denker, J., et Victorri, B., 1993. Efficient pattern recognition using a new transformation distance, Advances in Neural Information Processing Systems, Vol. 5, 50-58.
- [25] Duda, R.O., Hart, P.E., et Stork, D.G., Pattern classification, Edition John Wiley and Sons, Londres, 738 pages. 1997.

- [26] Cheriet. M., Nawwaf, K., Liu, C-L., et Suen, C-Y., Character recognition systems : A guide for students and practioners, Edition John Wiley and Sons, Londres, 321 pages. 2007.
- [27] Aboulaiche, N. and Meskine, N., Implémentation d'un système de reconnaissance de mots arabes manuscrits basé sur la transformée en Ridgelets et les SVMs. Mémoire de Master, USTHB. 2010.
- [28] Boser, B., Guyon, I., et Vapnik, V., A training algorithm for optimal margin classifiers, In fifth annual workshop on computational learning theory, Proceedings of the 5th annual workshop on Computational learning theory, Pittsburg , pp.144-152. 1992.
- [29] Cortes, C., et Vapnik, V., Support vector network, Machine Learning Journal, Vol. 20, 273-197. 1992.
- [30] Vapnik, V. Statistical Learning Theory. Wiley, NY. 1998.
- [31] Mangasarian OL, Wild EW. Multisurface proximal support vector lassification via generalized eigenvalues. IEEE Trans PAMI. 2006; 28(1): pp. 69–74.
- [32] Jayadeva, Khemchandani R, Chandra S. Twin support vector machine for pattern classification. IEEE Trans Pattern Anal Mach Intell 2007;29(5):905–10.
- [33] Kim KJ. Financial time series forecasting using support vector machines. Neurocomputing 2003;55(1):307–19.
- [34] <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>, Consulté Mai 2019.
- [35] <https://www.kaggle.com/ronitf/heart-disease-uci>, Consulté Mai 2019.
- [36] Murphy,P.M., Aha, D.W. UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mllearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science. 1994.
- [37] T.P. Conrads, et al., "High-resolution serum proteomic features for ovarian detection", Endocrine-Related Cancer, 11, 2004, pp. 163-178.
- [38] E.F. Petricoin, et al., "Use of proteomic patterns in serum to identify ovarian cancer", Lancet, 359(9306), 2002, pp. 572-577.

_____:

هذا العمل هو جزء من مجال الذكاء الاصطناعي لحل المشاكل المتعلقة بالصحة. قمنا بتطبيق نظام التعرف على علم الأمراض باستخدام تقنية اختيار البيانات (NCA) ومصنف قوي (SVMs).
تتمثل مساهمتنا الرئيسية في هذا العمل، من ناحية، في استخدام مجموعة من الميزات لتقليل عدد العلاجات التي تقلل من المعدات الطبية والموظفين اللازمين. من ناحية أخرى، قدمت SVM كفاءتها من حيث حساب الوقت والأداء لاتخاذ قرار مهم بشأن علم الأمراض.
كلمات مفتاحية : أمراض القلب، سرطان الثدي، سرطان المبيض، التصنيف، تعلم الآلات، استخراج البيانات،
. NCA, SVM

Résumé :

Ce travail s'inscrit dans le domaine de l'intelligence artificielle pour résoudre des problèmes liés à la santé. Nous avons implémenté un système de reconnaissance de pathologie en utilisant une technique de sélection des données (NCA) et un classifieur puissant (SVMs). Notre principale contribution dans ce travail est, d'une part, l'utilisation de la sélection des caractéristiques pour minimiser le nombre de traitement conduite ce qui fait la minimisation des équipements médicaux et le personnel correspondant. De l'autre part, SVM a présenté son efficacité en terme de temps de calcul et de performance pour prendre une décision importante concernant la pathologie.

Mots clés : heart disease, breast cancer, ovarian cancer, classification, Apprentissage des machines, Sélection des caractéristiques, SVM, NCA.