

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



UNIVERSITÉ D'EL-OUED

FACULTÉ DES SCIENCES ET DE TECHNOLOGIE

Mémoire de fin d'étude

LICENCE ACADEMIQUE

Domaine: Mathématique et Informatique

Filière: Informatique

Spécialité: Informatique fondamentale

Présenté par: *M^{elle} Settou Tarablesse*

Thème

Reconnaissance du groupe sanguin par les
Supports Vecteurs Machines « SVM »

Soutenu le ...juin 2013

Devant le jury composé de:

Mr. Adamou Abdelrohman	MC (B) Univ. El Oued	Président
Mr. Berdjouh Chafik	MA (B) Univ. ElOued	examinateur
Mr. ZAIZ Faouzi	MA (B) Univ. ElOued	Rapporteur

Année universitaire 2012 – 2013



Remerciements

*Je remercie Allah le tout puissant, qui nous a donné la force et la patience pour
l'accomplissement de ce travail.*

*Mes remerciements, les plus vifs, ma profonde gratitude et mes respects s'adressent à
mon directeur de recherche*

M. ZAIZ Faouzi

*Pour avoir accepté de m'encadrer, pour les conseils et orientations tant précieux
qu'il m'a prodigué durant ce Mémoire.*

Sans son aide, notre travail n'aurait pas vu la lumière.

Je remercie vivement les membres du jury qui m'ont fait l'honneur

D'accepter de juger notre travail.

Je remercie vivement l'enseignante Melle. Leila AMRANE

Je remercie vivement l'enseignant Khelifa abdennacer

Notre reconnaissance va aussi à tous ceux qui ont collaboré à notre

Formation en particulier les enseignants du département

D'Informatique, de l'université d'El-Oued.

Aussi à nos collègues de la promotion 2012-2013

Licence Informatique,

Je remercie également tous ceux qui ont participé de près

Ou de loin à élaborer ce travail.

Résumé

La reconnaissance d'image joue un rôle très important dans le monde actuel. Elle est capable de résoudre des problèmes complexes et rendre les activités de l'homme plus simple. Notamment, dans les analyses médicales et biologiques.

Dans ce travail on s'intéresse à faire une étude concernant le domaine de la reconnaissance d'image. Nous avons présenté une étude générale sur les SRI, puis elle été affinée par un intérêt particulier sur deux phases important dans le processus de reconnaissance : phase de segmentation et phase de classification.

Nous avons présenté un état de l'art des différents systèmes de groupage (ABO, Rhésus, KELL,...etc.). Ensuite, la méthode de classification utilisée SVM est présentée en détail.

Finalement les différentes phases du système développé ont été exposées avec nos choix de méthodes de prétraitement, segmentation, classification et post-traitement.

Mots clés : Reconnaissance de formes, segmentation, groupe sanguin, pré et post-traitement, SVM.

ملخص

التعرف على الصور " محتوى الصور " يلعب دورا جُداً هام في وقتنا الحالي. فهو يسمح لنا بحل المشاكل المعقدة كما يسهل العديد من الأنشطة التي يقوم بها الانسان, خاصة في مجال التحاليل الطبية والبيولوجية.

هذا العمل يمثل دراسة حول مجال التعرف على الصور. قمنا بعرض دراسة عامة لأنظمة التعرف على الصور, ثم التطرق لمرحلتين مهمتين في هذه العملية : مرحلة التجزئة ومرحلة التصنيف.

قمنا بتقديم لمحة على مختلف أنظمة تحديد فصيلة الدم (ABO, Rhésus, KELL,...etc.). ثم عرضنا طريقة التصنيف المختارة SVM بالتفصيل.

وفي الأخير, تم شرح مختلف مراحل النظام المطور وكذا الطرق المختارة في التنقية, التجزئة, التصنيف والمعالجة المتممة.

كلمات البحث: التعرف على الصور, التجزئة, فصائل الدم, قرب وبعد التجهيز, SVM.

Abstract

Image recognition plays a very important role in today's world, it's capable to resolve complex problems and simplify human activities, especially in medical and biologic analysis.

This work, presents a survey concerning image recognition domain. A general study concerning image recognition systems is presented, than a special view to the two important phases of the process: segmentation phase and classification phase.

We had presented a panorama of different blood type test systems (ABO, Rhesus, KELL ...etc.). Then, the chosen method of classification SVM is detailed.

Finally, the different developed system phase had been exposed with our choices of segmentation, classification...etc.

Keywords: Pattern recognition, segmentation, blood group, pre-and post-treatment, SVM.

Glossaire

GS: Groupe Sanguin.

KNN: K Nearest Neighbor.

SRI: Système Reconnaissance des Images.

SVM: Support Vector Machines.

RI: Reconnaissance des Images.

RH: RHESUS

TI: Traitement d'Image.

Liste des figures

I. Reconnaissance des images	2
Figure I. 1: Diffèrent aspects de SRI.....	4
Figure I. 2: Exemple de contour(A) et de région(B).....	8
Figure I. 3: Schéma générique d'un SRI.....	9
Figure I. 4: Schéma générale d'un SRI.....	10
Figure I. 5: Exemple de prétraitement.....	11
Figure I. 6: Schéma présentant les différentes méthodes de segmentation [12, 16, 17].....	13
II. Groupe sanguin «GS »	21
Figure II. 1: Différents types des systèmes à partir des différentes formes des membranes.....	23
Figure II. 2: Source génétique et la façon de former l'antigène H.....	24
Figure II. 3: la source génétique et la façon de former l'antigène A.....	25
Figure II. 4: Source génétique et la façon de former l'antigène B.....	25
Figure II. 5: loi de compatibilité.....	27
Figure II. 6: Principe réactionnel des groupes sanguins [36].....	28
Figure II. 7: Réalisation du groupage sanguin ABO-RH1.....	29
III. Support Vector Machines	32
Figure III. 1: Schéma des différents types d'apprentissage.....	34
Figure III. 2: A: Cas Sur-apprentissage, B : Cas sous-apprentissage.....	35
Figure III. 3: Exemple d'un hyperplan séparateur [15].....	36
Figure III. 4: Exemple de vecteurs de support [15].....	37
Figure III. 5: Exemple d'hyperplan séparateur.....	37
Figure III. 6: a) Hyperplan avec faible marge, b) Meilleur hyperplan séparateur [15].....	38
Figure III. 7: Exemple de classification d'un nouvel élément.....	39
Figure III. 8: Schéma spectacles les différents modèle des SVM.....	39
Figure III. 9: a) Cas linéairement séparable, b) Cas non linéairement séparable [21,15].....	40
Figure III. 10: Exemple de marge maximal (SVM binaire à marge dure).....	41
Figure III. 11: SVM binaire à marge souple.....	43

Figure III. 12: Exemple de changement de l'espace de données	44
Figure III. 13: le principe général de technique SVM dans ce cas.....	44
Figure III. 14: Illustration de passage à R3 [41]	45
Figure III. 15: Architecture d'une machine à vecteur support	47
IV. Conception & Mise en œuvre	49
Figure IV. 1: Illustration des modules du système.....	50
Figure IV. 2: Exemple de prétraitement.....	52
Figure IV. 3: Phase de segmentation.....	53
Figure IV. 4: Phase d'extraction des caractéristiques	57
Figure IV. 5: Nombre de segments de goutte agglutinée et non agglutinée	58
Figure IV. 6 : Taille de goutte agglutinée et non agglutinée.....	58
Figure IV. 7: La forme des gouttes agglutinée et non agglutinée	59
Figure IV. 8: Illustration du processus de segmentation.....	60
Figure IV. 9: Phase de classification.....	61
Figure IV. 10: Illustration de la relation entre l'application utilisateur et le package SVM.....	63
Figure IV. 11: Relation en détail entre l'application utilisateur et le package libsvm-2.83	64
Figure IV. 12: Illustration de la fenêtre principale de l'application	66
Figure IV. 13: Illustration du chargement d'une image.....	67
Figure IV. 14: Extraction des caractéristiques des exemples d'apprentissage	67
Figure IV. 15: Illustration du mode d'apprentissage	68
Figure IV. 16: Illustration du mode de test	68
Figure IV. 17: Illustration du résultat du traitement.....	69

Liste des Tableaux

I. Reconnaissance des images

Tableau I. 1: Comparaison brève entre les approches en-ligne et hors-ligne [1].....5

II. Groupe sanguin «GS »

Tableau II. 1: Système ABO26

Tableau II. 2 : Principe réactionnel des groupes sanguin.....28

Tableau II. 3: lois de compatibilité des groupes sanguins.....30

IV. Conception & Mise en œuvre

Tableau IV. 1: Chaine de définition associée à chaque groupe sanguin65

Tableau IV. 2: Résultats de test.....70

Sommaire

Glossaire	I
Liste des figures.....	II
Liste des tableaux.....	IV
Sommaire.....	V
Introduction générale	1
I. Reconnaissance des images	2
Introduction	3
I.1 Différents aspects d' un SRI	3
I.1.1 Type d'acquisition	4
1.1 La reconnaissance en-ligne (on-line)	4
1.2 La reconnaissance hors-ligne (off-line)	4
I.1.2 Approches de reconnaissance.....	5
2.1 Approche globale	5
2.2 Approche analytique	5
I.1.3 Nature des traits caractéristiques.....	6
3.1 Caractéristiques topologiques ou métrique	6
3.2 Caractéristiques structurelles	6
3.3 Caractéristiques statistiques	6
3.4 Caractéristiques globales ou locales	7
3.5 Superposition des modèles et corrélation	7
I.2 Problèmes liés à la Reconnaissance d'images « RI »	7
I.3 Organisation générale d'une SRI	7
I.3.1 quelques concepts de base d'imagerie	7
I.3.2 Qu'est-ce qu'un SRI ?	9

I.3.3	Les phases d'un SRI.....	10
3.1	Acquisition.....	11
3.2	Prétraitement	11
3.3	Segmentation.....	12
3.5	Classification.....	17
3.6	Post –traitement	19
	Conclusion.....	20
II.	Groupe sanguin «GS »	21
	Introduction	22
II.1	Systèmes de groupage	22
II.2	Le système ABO et Hh.....	23
II.2.1	Les antigènes du système ABO-Hh.....	24
II.2.2	Les Anticorps ANTI-A et ANTI-B.....	26
II.2.3	les lois de compatibilité ABO.....	27
II.2.4	Principe réactionnel des groupes sanguin	27
II.3	Le système Rhésus « RH ».....	28
II.3.1	Aspects génétique et biochimique	28
II.3.2	Anticorps ANTI-Rhésus	29
II.3.3	Anticorps ANTI-Rhésus	29
II.3.4	Réalisation du groupage sanguin ABO-RH1.....	29
II.4	Le système KELL.....	30
II.5	Autres systèmes d'intérêt clinique en transfusion sanguine.....	30
	Conclusion.....	31
III.	Support Vector Machines	32
	Introduction	33
III.1	Pourquoi les Machine à Vecteurs de Support « SVM » ?.....	33

III.2 Apprentissage statistique et SVM	33
III.2.1 Objectif de l'apprentissage statistique	34
III.2.2 Sur-apprentissage et sous-apprentissage	35
A. Sur-apprentissage :	35
B. Sous-apprentissage :	35
III.3 Notions de base	35
III.3.1 Hyperplan	36
III.3.2 Marge.....	36
III.3.3 Support vecteur.....	36
III.3.4 Séparateur Optimale « vaste marge »	37
III.4 Pourquoi maximiser la marge ?.....	38
III.5 Linéarité et non-linéarité	39
III.5.1 Cas linéairement séparable	40
1.1 SVM à marge dure (Sans bruit) :	40
1.2 Cas non séparable :- SVM à marge molle «souple » (Avec bruit)	42
III.5.2 Cas non linéairement séparable	43
2.1 Fonction noyau (kernel).....	44
2.2 Condition de Mercer	46
III.6 Complexité	46
III.8 Les domaines d'applications	48
Conclusion.....	48
IV. Conception & Mise en œuvre	49
Introduction	50
IV.1 Mise en œuvre du système	50
IV.1.1 Acquisition	51
IV.1.2 Pré-traitement	51

Binairisation.....	51
IV.1.3 Segmentation	52
3.1 Segmentation niveau 1	52
3.2 Segmentation niveau 2.....	52
IV.1.4 Extraction des caractéristiques	56
4.1 Caractéristiques statistiques	56
4.2 Caractéristiques structurelles	57
IV.1.5 Classification.....	59
5.1 Classes à reconnaître.....	59
5.2 Classifieur SVM.....	60
IV.1.6 Post-traitement.....	64
IV.2 Résultats et bilan	65
IV.2.1 Choix du langage de programmation	65
IV.2.2 Interface et Fenêtres	66
2.1 Mode Apprentissage	66
2.2 Mode Test	68
IV.2.3 Test et Résultats.....	69
Conclusion.....	70
Conclusion générale	71

Introduction générale

La reconnaissance d'image est un domaine très vaste et pluridisciplinaire qui fait l'objet de nombreux travaux de recherches. Elle est très utilisée dans beaucoup d'applications et domaines tels que la reconnaissance des visages, reconnaissance de la parole, reconnaissance de caractères imprimés ou manuscrit...etc. En plus, les SRI sont utilisés pour résoudre des problèmes d'analyse médicale ou biologique.

Jusqu'aujourd'hui, la plupart des hôpitaux utilisent un processus de groupage manuel, ce qui pose de nombreux problèmes comme : l'absence d'un biologiste avec une présence d'un grand nombre de patients.

Notre travail vise à développer une application de groupage sanguin semi-automatique par la création d'un système qui apprend d'un ensemble d'images des différents groupes sanguins et par la suite il permet de reconnaître de nouveaux exemples (nouvelles images de groupes) de groupes sanguins et les classifie à l'aide d'un classificateur SVM.

De telle application va être très efficace pour :

- les centres d'analyse,
- les hôpitaux,
- les pharmacies,
- la police dans l'identification des suspects.

Pour réaliser notre application, nous avons organisé le mémoire en quatre chapitres comme suit :

Le premier chapitre présente une vue générale des SRI, ainsi les différents problèmes d'un SRI ont été exposés.

Le second chapitre vient d'introduire les systèmes de groupage sanguin plus connus.

Le troisième chapitre, va mettre l'accent sur la méthode de classification choisie, par l'étude de la méthode de machines à vecteurs de support « SVM ».

Le quatrième chapitre constitue la liste de nos choix pour chaque phase du processus de reconnaissance et les tests et résultats obtenus.

Chapitre I:

Reconnaissance des images « SRI »

- *Introduction*
- *Différents aspects d'un SRI*
- *Problèmes liés à la Reconnaissance d'images « RI »*
- *Organisation générale d'une SRI*
- *Conclusion*

Introduction

On désigne par reconnaissance de formes «ou parfois reconnaissance de motifs » un ensemble de techniques et méthodes visant à identifier des motifs informatiques à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à ce motif. On considère que c'est une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques.

Le mot forme est au sens très général, il ne s'agit pas que de forme géométrique. Les formes ou motifs à reconnaître peuvent être de natures très variées. Il peut s'agir de contenu visuel (code barre, visage, empreinte digitale...) ou sonore (reconnaissance de parole), d'images médicales (rayon X, EEG, IRM...) ou multi- spectrales (images satellitaires) et bien d'autres.

Dans ce chapitre, on s'intéresse d'une part à introduire et présenter un état de l'art du domaine de la reconnaissance des images, et d'autre part à exposer les différentes approches, méthodes et technique réalisées depuis plusieurs années.

I.1 Différents aspects d' un SRI

Il est très difficile de créer un système SRI capable de reconnaître n'importe quelle image ou forme. Tout dépend de l'application visée ou voulu [1] et des données à traiter [2, 3]. Généralement, les SRI sont classés en se basant sur trois critères:

- Outil d'acquisition: Les systèmes qualifiés de « en-ligne » ou « hors-ligne »,
- Approches de reconnaissance: approches globales ou analytiques selon que l'analyse s'opère sur la totalité d'image, ou par segmentation en des segments.
- Nature des traits caractéristiques: approches statistiques, structurelles ou stochastiques.

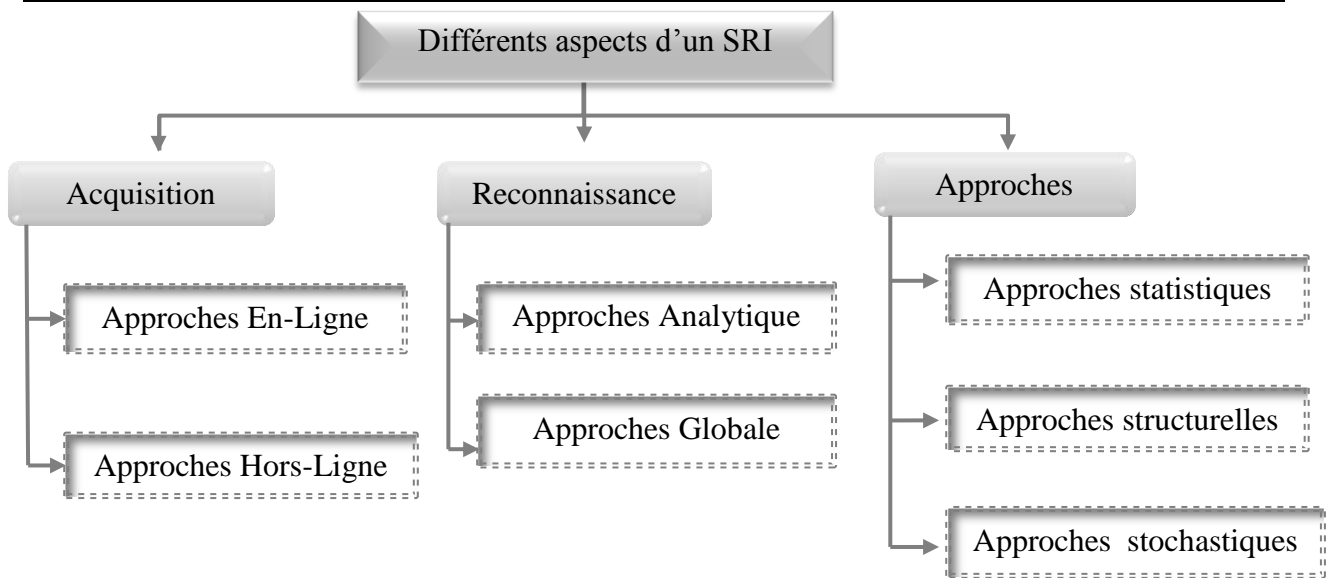


Figure I. 1: Différent aspects de SRI

I.1.1 Type d'acquisition

Les méthodes de reconnaissance se distinguent tout d'abord par le type d'acquisition des données. La reconnaissance est dite en-ligne lorsque les données dynamiques sont acquises pendant l'écriture. On pense ici à une tablette graphique ou un papier électronique où l'utilisateur écrit avec un stylo. La reconnaissance est dite hors-ligne lorsqu'il s'agit de reconnaître l'image d'objet obtenue avec un scanner. On s'accorde généralement à dire que la reconnaissance hors-ligne est plus complexe que la reconnaissance en-ligne pour plusieurs raisons.

1.1 La reconnaissance en-ligne (on-line)

Dans le cas de la reconnaissance en-ligne (dynamique [4]) le système reçoit les images d'entrée de données en temps réel, ce qui permet d'intégrer les mouvements du stylo électronique (Tracé Dynamique [5]) et pression information [6], et calcule la relation entre les points pour extraire les caractéristiques afin de reconnaître les objets au fur et à mesure qu'ils sont écrits à la main [1]. La reconnaissance en-ligne présente un avantage majeur c'est la possibilité de correction et de modification de manière interactive vu la réponse en continu du système [3].

1.2 La reconnaissance hors-ligne (off-line)

La reconnaissance hors-ligne (ou en différé, ou encore statique) est obtenue par la saisie d'une image déjà existante, obtenue par un scanner ou une caméra. Ayant perdu toute information temporelle sur l'ordre des points. De plus, ce mode introduit une difficulté

supplémentaire relative à la variabilité du tracé en épaisseur et en connectivité, nécessitant l'application de techniques de prétraitement [4].

Le tableau suivant présente une brève comparaison des deux approches :

Tableau I. 1: Comparaison brève entre les approches en-ligne et hors-ligne [1]

<i>Critère de comparaison</i>	<i>En-ligne</i>	<i>Hors-ligne</i>
Outils d'acquisition	▣ Stylo électronique plus tablette graphique.	▣ Scanner ou caméra.
Bruit d'image	▣ presque absent.	▣ existence d'un bruit
Informations disponibles	▣ la position, ▣ la direction du mouvement, ▣ les points de fin, ▣ les points de début, ▣ ordre des traits	▣ absence d'informations contextuelles

I.1.2 Approches de reconnaissance

Durant les dernières décennies, beaucoup de méthodes de segmentation ont été développées dans le but d'avoir un système de reconnaissance d'image plus robuste. Malgré tous les efforts, la situation reste loin d'atteindre les ambitions. En se basant sur le processus de segmentation, deux approches ont été appliquées [7]:

- ▣ Approche globale (segmentation libre).
- ▣ Approche analytique (basée segmentation).

2.1 Approche globale

L'approche globale essaye de reconnaître la représentation intégrale des objets de l'image d'entrée [7] et de le décrire indépendamment des sous objets qui le constituent.

2.2 Approche analytique

L'approche analytique au contraire de celle présentée précédemment, isole les différents sous objets des objets [7]. L'idée de base de l'approche analytique est de segmenter l'image d'objet en entrée en sous objets ou en fragments morphologiques significatifs inférieurs au objet appelés graphèmes[2,6].

I.1.3 Nature des traits caractéristiques

La nature des caractéristiques varie d'une approche à une autre. Généralement, les caractéristiques peuvent être classés en cinq groupes principaux [2, 8, 9] :

- caractéristiques topologiques,
- caractéristiques structurelles,
- caractéristiques statistiques,
- globales ou locales, et
- superposition des modèles et corrélation.

3.1 Caractéristiques topologiques ou métrique

Ce type de primitives est basé sur des densités de pixels. Dans ce type de primitives, on compte également les profils et histogrammes. Pour maintenir un vecteur de taille fixe, on divise l'image en un nombre fixe de bandes horizontales et verticales. Les caractéristiques sont les moyennes des valeurs sur ces bandes [9].

Elles consistent à compter dans une forme le nombre de trous, évaluer les concavités, mesurer des pentes et autres paramètres de courbures et évaluer des orientations, mesurer la longueur et l'épaisseur des traits, détecter les croisements et les jonctions des traits, mesurer les surfaces et les périmètres, ... [8]

3.2 Caractéristiques structurelles

Elles ressemblent beaucoup aux primitives topologiques. La différence est qu'elles sont généralement extraites non pas de l'image brute mais à partir du squelette ou du contour de la forme en donnant ses propriétés globales et locales. Ainsi, on ne parle plus de trous, mais de boucles ou de cycles dans une représentation filiforme de l'objet. Parmi ces caractéristiques on peut citer [8]:

- Les traits et les anses dans les différentes directions ainsi que leurs tailles.
- Les points terminaux.
- Les points d'intersections, Les boucles.....etc.

3.3 Caractéristiques statistiques

Les caractéristiques statistiques décrivent une forme en termes d'un ensemble de mesures extraites à partir de cette forme.

3.4 Caractéristiques globales ou locales

Les primitives globales cherchent à représenter au mieux la forme générale d'un objet et sont donc calculées sur des images relativement grandes (ex : transformée de Fourier et transformée de Hough). Les primitives locales sont calculées lors d'un parcours des pixels de l'image avec un pas d'analyse qui dépend de la modélisation, du type de primitive et de la taille de l'image [2,8].

3.5 Superposition des modèles et corrélation

La méthode de « template matching » appliquée à une image binaire (en niveaux de gris ou squelettes), consiste à utiliser l'image de la forme comme vecteur de caractéristiques pour être comparé à un modèle (template) pixel par pixel dans la phase de reconnaissance, et une mesure de similarité est calculée [2].

I.2 Problèmes liés à la Reconnaissance d'images « RI »

Divers problèmes compliquent le processus de reconnaissance, ce qui rend la tâche de l'SRI très délicates, parmi ces problèmes nous pouvons citer [4]:

- Disposition spatiale des objets,
- Nombre des objets,
- Taille des objets.

I.3 Organisation générale d'une SRI

Avant d'entamer la définition de la reconnaissance des images, il est indispensable de savoir quelques notions de base:

I.3.1 quelques concepts de base d'imagerie

Les notions plus utilisé sont :

- Image numérique,
- Une région,
- Un contour,
- Traitement des images,
- Analyse d'une image.

A. Image numérique

Une image numérique est une image (dessin, icône, photographie, ...) créée, traitée, stockée sous forme binaire. Lorsqu'on agrandit une image numérique, on voit que celle-ci est composée d'un ensemble de points, appelés « pixels », est désigné par ses coordonnées entières.

Un **Pixel** est l'unité de base d'une image numérique. À chaque pixel est associée une couleur, généralement décomposée en 3 couleurs primaires. Il y a deux types d'image numérique :

- Une image matricielle (image en mode point).
- Une image vectorielle (image en mode trait).

B. Une région

Une région est un ensemble de pixels connexes ayant des propriétés communes qui les différencient des pixels des régions voisines [10].

C. Un contour

Un contour est la frontière qui sépare deux objets dans une image numérique. On peut dire aussi une discontinuité de l'image ou est une variation brusque d'intensité. » [11].

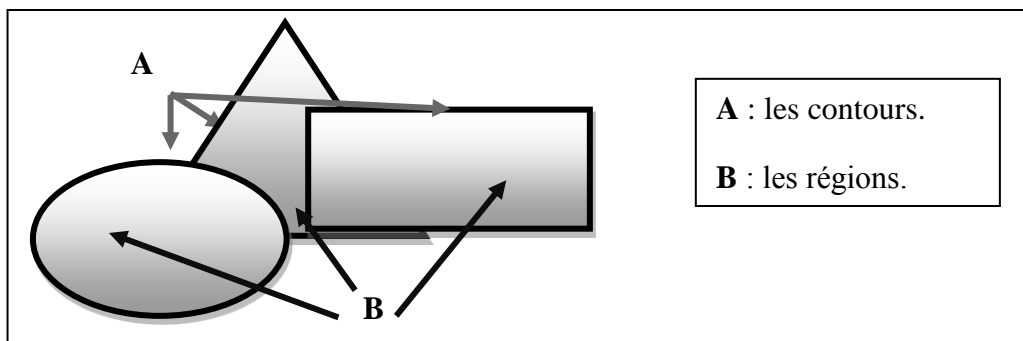


Figure I. 2: Exemple de contour(A) et de région(B)

D. Traitement des images

Le traitement d'images « TI » est souvent définie comme ensemble automatisé qui permet, à partir d'images numérisées de produire d'autres images numériques ou d'en extraire de l'information [12].

E. Analyse d'image

L'analyse d'image à travers ses attributs consiste souvent à extraire un certain nombre de propriétés caractéristiques et à les exprimer sous forme paramétriques. Ces derniers, permettent de décrire, de caractériser, d'analyser et de segmenter les images.

I.3.2 Qu'est-ce qu'un SRI ?

Un SRI est un ensemble des techniques informatiques de représentation et de décision permettant aux machines de simuler un comportement « sensible » (de l'être humain tel que la vision ... etc.).



Figure I. 3: Schéma générique d'un SRI

Il y a plusieurs domaines d'application de la reconnaissance d'images « RI » par exemples :

- Imagerie médicale et vidéo numérique.
- Télédétection (éléments topographiques, détection de changements, gestion des catastrophes.....etc.).
- Reconnaissance des formes, Communication visuelle (vidéo conférence, vidéophone, ...).
- Reconnaissance des caractères (imprimés, manuscrits)... etc.

I.3.3 Les phases d'un SRI

Le schéma général d'un système de reconnaissance d'images est présenté dans « Figure I.4 »

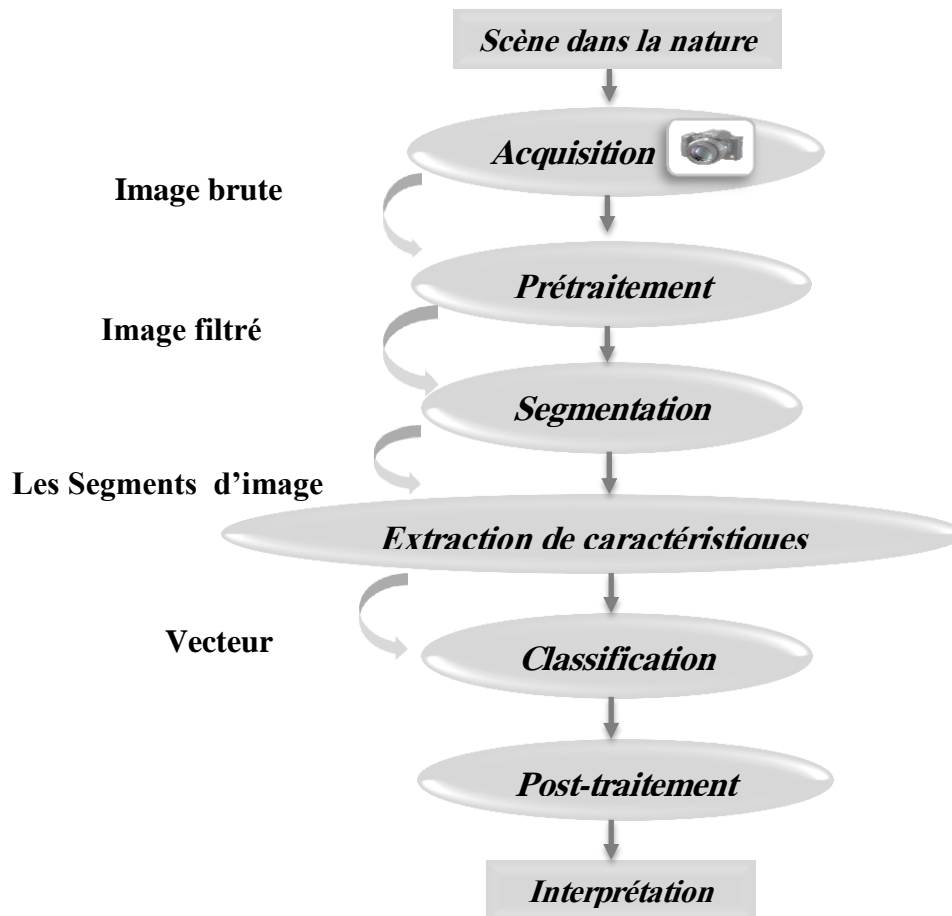


Figure I. 4: Schéma générale d'un SRI

1. **Scène dans la nature:** c'est une scène dans la nature (image continue).
2. **Image brute :** c'est une image obtenu après la phase d'acquisition.
3. **Image filtré :** c'est une image obtenu après la phase de prétraitement.
4. **Les segments d'image :** Ils sont les parties d'images qu'on obtient après la segmentation.
5. **Vecteurs :** c'est un tableau contenant les caractéristiques d'un segment d'image.
6. **Interprétation:** c'est le résultat de la reconnaissance.

3.1 Acquisition

La phase d'acquisition consiste à capter l'image au moyen des capteurs physiques (scanner, caméra, appareil photos numérique...) et de la convertir en grandeurs numériques adaptés aux systèmes de traitement, avec un minimum de dégradation possible [13].

La numérisation des images c'est la conversion analogique numérique dont c'est succession de trois effets sur l'image analogique «naturel » :

- L'échantillonnage.
- La quantification.
- Le codage.

3.2 Prétraitement

La phase de prétraitement consiste essentiellement à réduire le bruit superposé aux données et essayer de ne garder que l'information significative de la forme représentée.

Le bruit peut être dû aux conditions d'acquisition (éclairage, mise incorrecte de l'image,...) ou encore à la qualité de l'image d'origine [14,15]. Parmi les opérations de prétraitement généralement utilisées on cite:

- L'extraction des composantes connexes,
- Le redressement d'image,
- Le lissage,
- La normalisation,
- la squelettisation, et
- Le contraste.

L'image obtenue dans la phase d'acquisition n'est qu'une image brute (bruitée). Durant la phase de prétraitement nous allons appliquer des filtres afin de rendre cette image prête pour des traitements futurs [13]. (La figure I.5 illustre un exemple d'application).

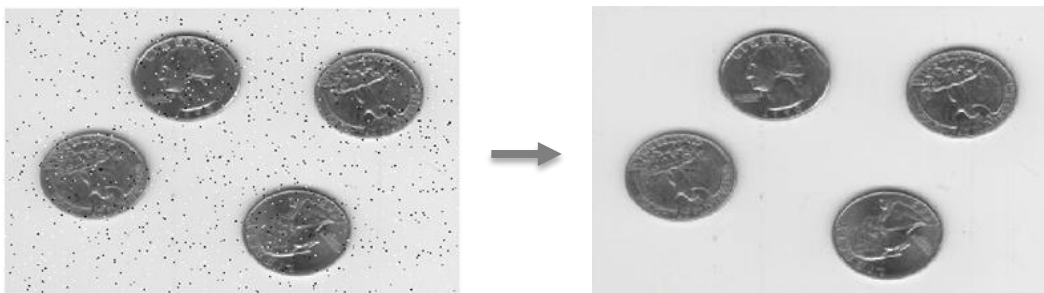


Figure I. 5: Exemple de prétraitement

Pour réaliser cette phase on passe par deux étapes:

- Binairisation,
- Filtrage.

- A. Binairisation:** L'étape de binairisation consiste à transformer la matrice de l'image brute obtenue par un procédé d'acquisition en une image représentée en image binaire (noir et blanc).
- B. Filtrage :** le filtrage consiste à suivre certains processus afin d'éliminer ces taches et de représenter en mieux les points d'intérêt.

3.3 Segmentation

La segmentation d'images peut être définie [10] comme « un traitement de bas niveau qui consiste à créer de l'image a en un ensemble de sous-ensembles R_i (sous-images ou segments), appelés régions tels qu'aucune région ne soit vide, l'intersection entre deux région soit vide et l'ensemble des régions recouvre toute l'image.

Formellement : la segmentation d'une image numérique I utilisant un prédicat d'homogénéité P est communément définie comme une partition [12, 16, 17]: $S=R_1, R_2, R_3, \dots, R_N$, tels que :

1. $I = \cup R_i, i \in [1 \dots N]$
2. R_i est connexe, $\forall i \in [1 \dots N]$
3. $P(R_i) = \text{vrai}, \forall i \in [1 \dots N]$
4. $P(R_i \cup R_j) = \text{faux} \forall i \neq j, (i, j) \in [1 \dots N]^2$

Où N est le nombre de régions résultant de la segmentation, R_i est une région.

3.3.1 Les objectifs de la segmentation

Il y a plusieurs objectifs de la segmentation, on cite quelques une :

- Extraire les entités d'une image pour y appliquer un traitement spécifique et interpréter le contenu de l'image [18].
- Elle aide à localiser et à délimiter les entités présentes dans l'image [19].

3.3.2 Les méthodes de segmentation

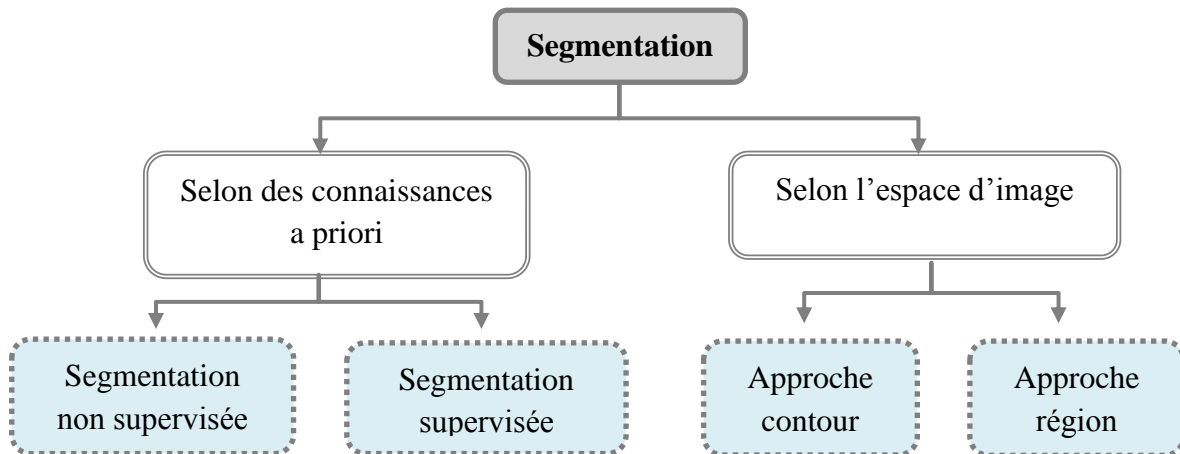


Figure I. 6: Schéma présentant les différentes méthodes de segmentation [12, 16, 17]

2.1 Segmentation selon des connaissances à priori

La connaissance a priori, comme définie par [20], fait référence à toute l'information disponible sur le problème en plus des données d'apprentissage. Il existe deux types de méthodes de segmentation d'image selon la connaissance à priori.

- Segmentation supervisée,
- Segmentation non supervisée.

A. Segmentation supervisée

Ces sont des méthodes dans lesquelles les classes des segments sont connues a priori avant d'effectuer l'opération d'identification des éléments de l'image.

B. Segmentation non supervisée

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune base d'apprentissage et par là même aucune tâche préalable d'étiquetage manuel n'est requise [21].

2.2 Segmentation selon l'espace d'image

Il existe plusieurs méthodes de segmentation selon l'espace d'image, on peut les diviser en deux approches :

- Approche basée sur les régions et
- Approche basée sur les contours.

A. La segmentation basée sur les régions

Les méthodes fondées sur les régions se focalisent sur l'extraction des régions en considérant leur homogénéité vis à vis des caractéristiques pertinentes (intensité, texture, ...etc.) au niveau des pixels, alors elles manipulent directement des régions [17].

B. La segmentation basée sur les contours

Les méthodes fondées sur les contours cherchent à délimiter les objets selon leurs contours présents, en se basant généralement sur les variations d'intensité dans l'image. Ces variations sont significatives aux frontières entre les régions [17].

3.4 Extraction des caractéristiques

L'extraction des caractéristiques consiste à extraire les mesures des entrées pour distinguer entre les classes, Devijver et Kittler dans, suggèrent que le problème d'extraction des caractéristiques des données d'entrée soit fait par la sélection des informations les plus significatives et appropriées pour les buts de classification et capable de discriminer entre classes, dans le sens de [13]:

- ▣ minimiser la variabilité entre les exemples de la même classe.
- ▣ maximiser la variabilité entre les exemples des classes différentes.

3.4.1 Les différents types de descripteurs

La littérature dans ce domaine étant très riches, mais nous allons dans ce qui suit présenter les méthodes les plus importantes et les plus pertinentes qui sont de trois types :

- ▣ La description globale de l'image
- ▣ La description locale de l'image
- ▣ Les descripteurs spécifiques.

A. La description globale de l'image

La recherche est faite approximativement sur toute l'image. Pour se faire, on indexe les images par les caractéristiques symboliques comme la couleur, la texture, la forme...etc.

A.1 Descripteurs de la couleur

La couleur est une caractéristique riche d'information et très utilisée pour la représentation des images. Elle forme une partie significative de la vision humaine. L'utilisation de la couleur pour l'extraction des primitives passe par trois principales étapes qui sont: le choix de l'espace couleur, la sélection d'un quantificateur et finalement le calcul des descripteurs couleurs.

A.2 Descripteurs de texture

La texture est définie comme la répétition d'un motif créant une image visuellement homogène. Plus précisément, la texture peut être vue comme un ensemble de pixels spatialement agencés selon un certain nombre de relations spatiales, ainsi créant une région homogène. De ces définitions, les recherches sur la modélisation des textures se sont portées sur la caractérisation de ces relations spatiales [22,23].

De nombreuses approches et modèles [24] ont été proposées pour la caractérisation de la texture. Parmi les plus connues, on peut citer : les méthodes statistiques, les méthodes fréquentielles et les méthodes géométriques.

Dans ce travail, l'accent sera mis sur les méthodes statistiques, en particulier sur les statistiques du premier ordre à calculer la moyenne, la variance, l'entropie,...etc. à l'aide des équations suivantes :

La première équation, donne la valeur moyenne des niveaux de gris appartenant à tous les pixels de la région considérée dont:

$$\text{Moy} = \frac{1}{N} \sum_{i,j} g(i,j). \text{ Tel que :}$$

Le $g(i,j)$ représente la valeur du niveau de gris du pixel (i,j) , et N est un facteur de normalisation qui correspond au nombre total de pixels.

La deuxième équation, calcul la variance. Elle correspond au moment d'ordre 2. Elle mesure la répartition des niveaux de gris autour de la valeur moyenne dont :

$$\text{Var} = \frac{1}{N} \sum_{i,j} (g(i,j) - \text{Moy})^2 .$$

La troisième équation, correspond au moment d'ordre 3 centré autour de la moyenne (le skewness). Ce paramètre mesure la déviation de la distribution des niveaux de gris par rapport à une distribution symétrique. Pour une déviation par les valeurs élevées, le skewness est positif, alors que pour une déviation vers les basses valeurs, il est négatif, dont :

$$\text{SKEW} = \frac{1}{N} \sum_{i,j} (g(i,j) - \text{Moy})^3 .$$

La dernière équation, corresponde au moment d'ordre 4 centré autour de la moyenne (le Kurtosis). Il caractérise la forme du sommet de l'histogramme : plus le Kurtosis est faible et plus le sommet de l'histogramme est arrondi :

$$KURT = \frac{1}{N} \sum_{i,j} (g(i,j) - Moy)^4.$$

A.3 Descripteurs forme

La forme est l'un des attributs bas niveau également le plus utilisé pour décrire le contenu visuel des images. L'importance de la forme pour la recherche d'images peut être constatée par le simple fait que plusieurs systèmes d'indexation et recherche d'image par le contenu incorporent d'une façon ou d'une autre des descripteurs de formes. Ces derniers sont utilisés pour décrire la structure géométrique générique du contenu visuel [22,25].

B. La description locale de l'image

Puisque les utilisateurs ont besoin de chercher des objets dans une image, le système doit utiliser les caractéristiques locales puisque les caractéristiques globales ne peuvent pas donner les résultats désirés toujours [26].

On fait la segmentation de l'image pour la diviser en des régions locales. On souhaite que chaque région locale soit plus proche d'un objet dans l'image. Et puis on calcule les caractéristiques pour chacune des régions extraites. C'est difficile à choisir une "bonne" méthode de segmentation valide pour toutes les images.

C. Les descripteurs spécifiques (essentiellement biométrie)

Pour des applications où les images présentent moins de variance de l'une à l'autre, ces descripteurs ne sont pas suffisants, notamment dans des applications comme la reconnaissance de visages et la reconnaissance d'empreintes digitales. Des descripteurs spécifiques ont été proposés pour ces domaines particuliers en utilisant les connaissances expertes. Donc des descripteurs spécifiques ont été proposés pour ce domaine particulier en utilisant les connaissances expertes.

Les caractéristiques de visages comme les caractéristiques de la couleur de peau, les positions des yeux, de la bouche,....sont extraits. Ensuite, le système fait la comparaison avec les échantillons [22].

3.5 Classification

La phase de classification consiste à déterminer à quelle classe appartient une entrée de données (voir figure I.4). La classification dans un SRI regroupe deux sous-phases : l'apprentissage et la décision. Les caractéristiques de la phase précédente sont utilisées pour identifier un segment d'image et l'attribuer à un modèle de référence [13].

La classification passe par deux sous-phases :

- Apprentissage,
- Décision (Test).

3.5.1 Apprentissage

Avant de pouvoir prendre une décision, il faut acquérir des connaissances les organiser en modèles de référence ou classes. Ce travail est réalisé durant la phase d'apprentissage. Lorsque le concepteur (ou professeur) indique le nom de la forme d'entrée, l'apprentissage est dit supervisé. Si la construction des classes est automatique, on parle d'apprentissage non supervisé [27].

3.5.2 Décision

Le module de décision cherche parmi les modèles de référence une description des paramètres plus proche de celle du caractère. La décision peut conduire à un succès si la réponse est unique (un seul modèle répond à la description de la forme du caractère). Elle peut conduire à une confusion si la réponse est multiple. Enfin elle peut conduire à un rejet de la forme si aucun modèle ne correspond à sa description.

Dans les deux premiers cas, la décision peut être accompagnée d'une mesure de vraisemblance, appelée aussi score ou taux de reconnaissance. Les approches de décision peuvent être regroupées en quatre groupes principaux [27,2]:

- Approches statistiques
- Approches structurelles.
- Approches stochastiques.
- Et approches hybrides.

2.1 Approches statistiques

Dans l'approche statistique, la forme est représentée par un vecteur de n composant correspondant à la mesure de n caractéristiques observé sur elles. La décision consiste à trouver

la classe à laquelle le symbole possède la plus grande probabilité d'appartenir et à évaluer le risque lié à la prise de la décision [27].

A. Méthode de K plus proche voisin (KNN)

L'algorithme KNN (K Nearest Neighbors) affecte une forme inconnue à la classe de son plus proche voisin en le comparant aux formes stockées dans une classe de références nommée prototypes [28].

Il renvoie les K formes les plus proches de la forme à reconnaître suivant un critère des similarités. Une stratégie de décision permet d'affecter des valeurs de confiance à chacune des classes en compétition et d'attribuer la classe la plus vraisemblable (au sens de la métrique choisie) à la forme inconnue [29].

Cette méthode présente l'avantage d'être facile à mettre en œuvre et fournit de bons résultats. Son principal inconvénient est lié à la faible vitesse de classification due au nombre important de distances à calculer [29, 2].

B. Réseaux de neurones

Un réseau de neurones est un graphe orienté pondéré. Les nœuds de ce graphe sont des automates simples appelés neurones formels. Les neurones sont dotés d'un état interne, l'activation, par lequel ils influencent les autres neurones du réseau. Cette activité se propage dans le graphe le long d'arcs pondérés appelés liens synaptiques [28, 30].

C. Machines à Vecteurs de Support (SVM)

L'algorithme des machines à vecteurs de support a été développé dans les années 90 par Vapnik. Initialement il est développé comme un algorithme de classification binaire supervisée. Il s'avère particulièrement efficace de par le fait qu'il peut traiter des problèmes mettant en jeu de grands nombres de descripteurs, qu'il assure une solution unique (pas de problèmes de minimum local comme pour les réseaux de neurones) et il a fourni de bons résultats sur des problèmes réels [31].

L'algorithme sous sa forme initiale revient à chercher une frontière de décision linéaire entre deux classes, mais ce modèle peut considérablement être enrichi en se projetant dans un autre espace permettant d'augmenter la séparabilité des données. On peut alors appliquer le même

algorithme dans ce nouvel espace, ce qui se traduit par une frontière de décision non linéaire dans l'espace initial [13]. Nous reviendrons aux détails à cette méthode dans le troisième chapitre.

2.2 Approche Syntaxique « Structurelle »

Les approches structurelles reposent sur des représentations de formes quelconques sous forme d'agencement, d'assemblage de formes plus simples, les primitives, considérées, comme motifs élémentaire. Les structures les plus simples sont les représentations par chaînes, et nous verrons dans un premier temps comment les comparer pour effectuer la décision. Les chaînes sont par exemple utilisées pour comparer et reconnaître des formes à partir de leur squelette ou de leurs contours. Les grammaires et les automates permettent une analyse structurée, imitant celle du langage naturel.

Des modèles plus généraux reposent sur des graphes (avec le cas particulier des arbres), et sont souvent utilisés pour la mise en correspondance et l'analyse de scènes complexes, dans lesquelles l'agencement spatial des structures est de grande importance [32].

2.3 Approche stochastique

L'approche stochastique considère la forme comme un signal continu dans le temps, observable à différents endroits constituant les "états d'observation" (elle n'est donc utilisable que sur des objets comportant une information temporelle). Le modèle décrit ces états à l'aide de probabilités, de transition d'état à état, et d'observation d'états [27, 33].

2.4 Approche hybride

Pour améliorer les performances de reconnaissance, la tendance aujourd'hui est de construire des systèmes hybrides qui utilisent différents types de caractéristiques, et qui combinent plusieurs classificateurs en couches [27,34].

3.6 Post –traitement

Cette phase consiste à faire une sélection de la solution en utilisant des niveaux d'information plus élevés (syntaxique, lexicale, sémantiques...). Le post-traitement se charge également de vérifier si la réponse est correcte (même si elle est unique) en se basant sur d'autres informations non disponibles au classificateur [34].

Conclusion

Nous avons présenté, au long de ce chapitre, les différentes notions de base de la reconnaissance des images. En plus, les principaux problèmes et défis rencontrés dans ce domaine. Par la suite, nous avons abordé les différentes phases et étapes intervenant dans la conception d'un SRI. Où nous avons vu les différentes méthodes et techniques connus dans la boîte à outils de la littérature pour chaque phase du processus de la reconnaissance.

Dans le prochain chapitre, nous allons présenter les différents systèmes de groupage sanguin afin de connaître les étapes de ce processus.

Chapitre II:

Groupe sanguin «GS »

- *Introduction*
- *Systèmes de groupage*
- *Le système ABO et Hh*
- *Le système Rhésus*
- *Le système KELL*
- *Autres systèmes d'intérêt clinique en transfusion sanguine*
- *Conclusion*

Introduction

Le sang a toujours fasciné les humains. La perte de sang accompagnant souvent la perte de vie, on a, de tous temps, tenté de restituer sinon la vie du moins la vigueur avec du sang. Dès lors, de multiples essais de transfusions, aux succès inégaux, ont été tentés avec du sang d'animaux et du sang humain. Au cours de la guerre franco-prussienne de 1870, la transfusion fut largement utilisée et sauva de nombreux blessés.

Dans ce chapitre, nous allons présenter les différents systèmes de groupage du sang connus à nos jours.

II.1 Systèmes de groupage

Les groupes sanguins, ou phénotypes érythrocytaires, correspondent à des antigènes membranaires de l'érythrocyte, dont l'expression est déterminée par une série de systèmes génétiques polymorphes.

Ces antigènes, introduits dans un organisme qui les reconnaît comme étrangers, peuvent être la cible d'anticorps sériques naturels ou immuns, responsables d'une lyse cellulaire parfois grave, voire mortelle. Cette notion s'exprime dans 2 domaines de la pathologie : les accidents immunologiques transfusionnels et l'incompatibilité fœto-maternelle.

Plus de 23 systèmes de groupes sanguins ont été identifiés depuis la découverte du système ABO. Certains de nature glucidique, comme les systèmes " ABO, P, Hh ou Lewis", dont les extrémités terminales glycoprotéiques ou glycolipidiques membranaires portent les antigènes. D'autres, de nature peptidique comme les systèmes "Kidd, Duffy, RH", représentent l'expression directe des gènes et sont ancrés dans la membrane des hématies.

Les implications cliniques des conflits immunologiques mettent en jeu, de façon considérable, les antigènes de groupes sanguins. Il faut distinguer deux situations très différentes:

A. La présence d'anticorps naturels dans le système ABO représente un obstacle infranchissable à toute transfusion " incompatible " dans ce système,

B. L'immunisation et l'apparition d'anticorps irréguliers vis à vis du système Rhésus ou d'un autre système " majeur " imposent de sélectionner des hématies (donneurs) compatibles pour les transfusions ultérieures [35]

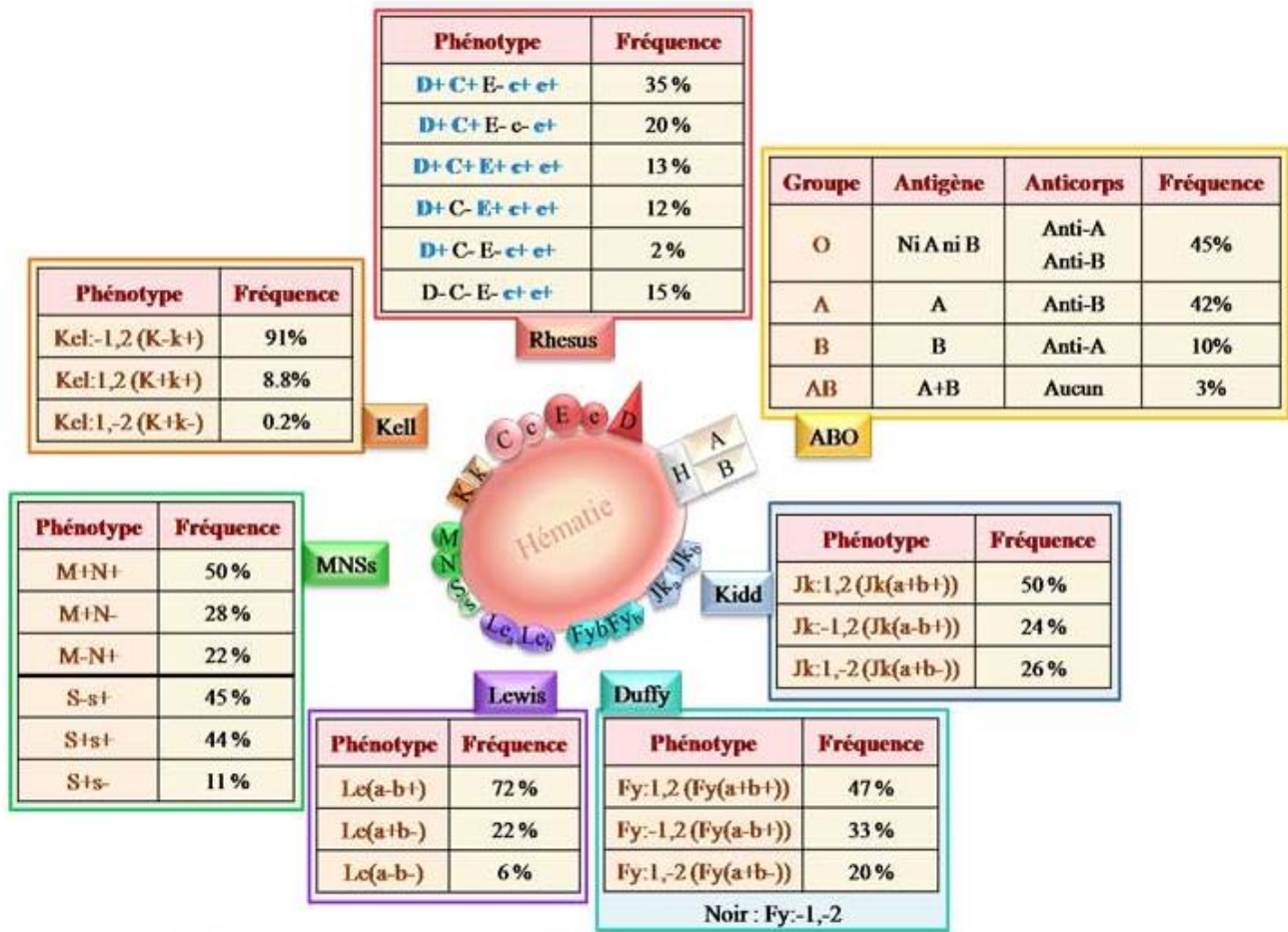


Figure II. 1: Différents types des systèmes à partir des différentes formes des membranes

II.2 Le système ABO et Hh

En 1900, Landsteiner observe que le plasma de différents sujets agglutine les hématies de nombreux autres sujets et, poursuivant ses études, il en déduit l'existence des groupes A, B et O. Un an plus tard, De Castillo décrit un quatrième groupe: AB [35,36].

En 1924, Bernstein démontre la transmission héréditaire selon les lois de Mendel des facteurs de groupes sanguins [37].

II.2.1 Les antigènes du système ABO-Hh

Les antigènes A, B et H sont des oligosaccharides portés par des glycolipides membranaires des hématies, des cellules épithéliales et endothéliales [38].

Le locus Hh sur le chromosome 19 présente deux variantes alléliques : H et h. L'allèle H code pour une fucose-transférase qui ajoute un fucose à l'extrémité terminale de la chaîne oligosaccharidique de base, formant l'antigène H. La synthèse ultérieure éventuelle des antigènes A et B nécessite la présence de l'antigène H. Il convient de noter l'extrême rareté de l'allèle h, gène amorphe, non fonctionnel. De plus, sa présence à l'état homozygote détermine le phénotype Bombay [38].

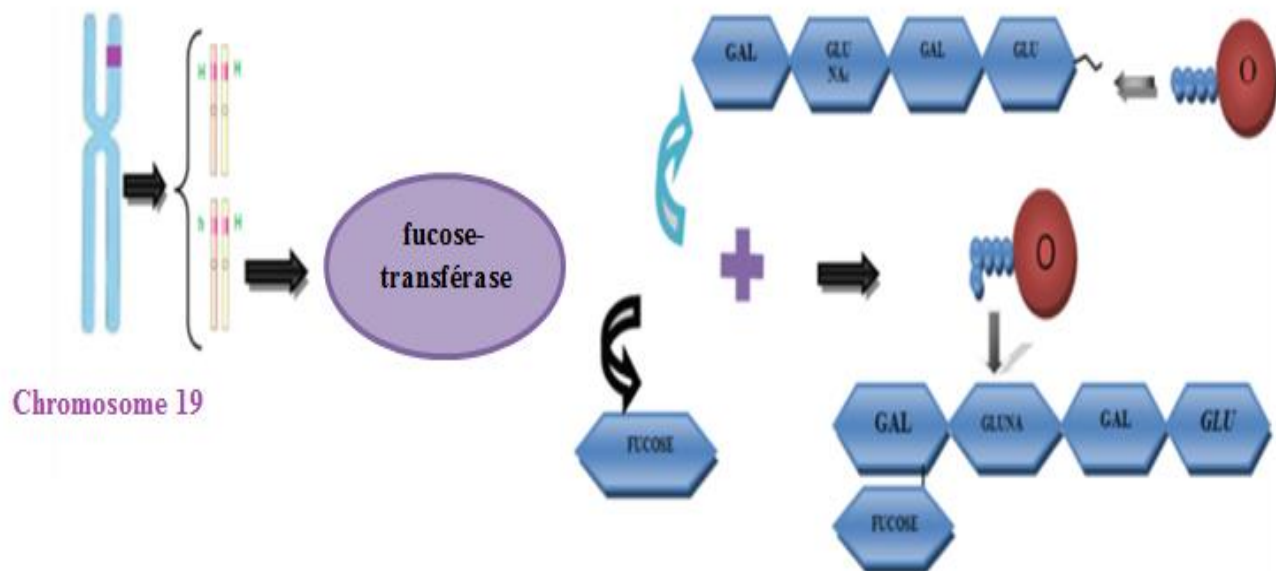


Figure II. 2: Source génétique et la façon de former l'antigène H

Les allèles A1 et A2 codent pour une N-acétyl-galactosamine-transférase. Chez les sujets de phénotype A2, l'antigène H persiste à la surface cellulaire. Les sujets de phénotype A1 possèdent, au contraire, une enzyme très active et l'antigène H, totalement masqué, ne peut plus être détecté. La distinction A1/A2 ne présente pas d'intérêt clinique majeur.

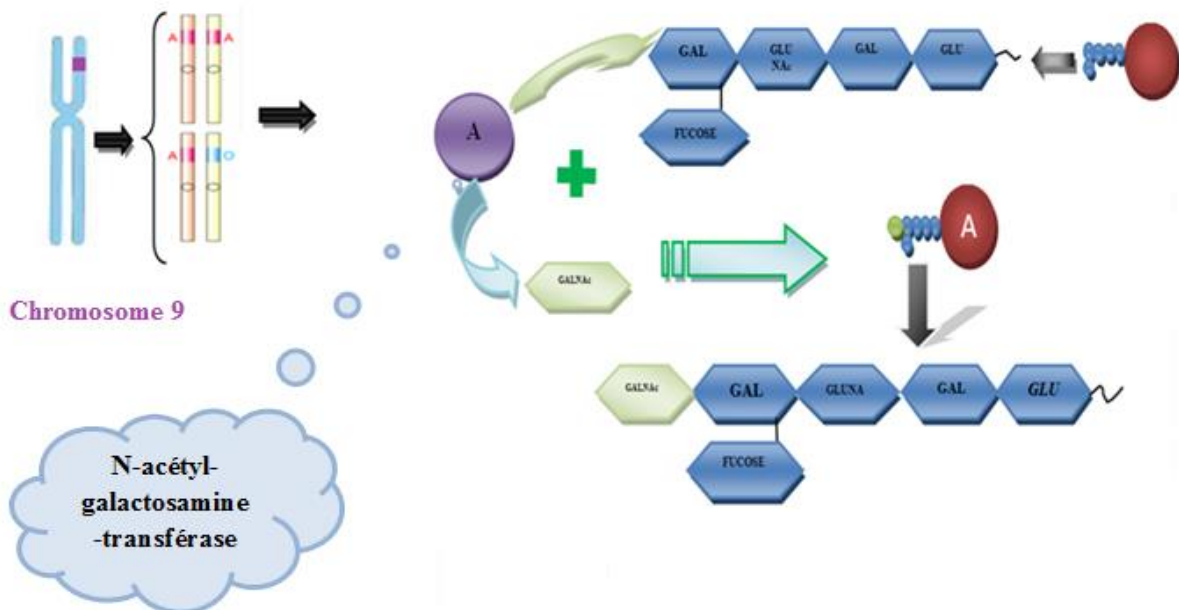


Figure II. 3: la source génétique et la façon de former l'antigène A

L'allèle B produit une galactose-transférase qui ajoute un résidu galactose et forme l'antigène B, toujours sous la condition que H soit présent.

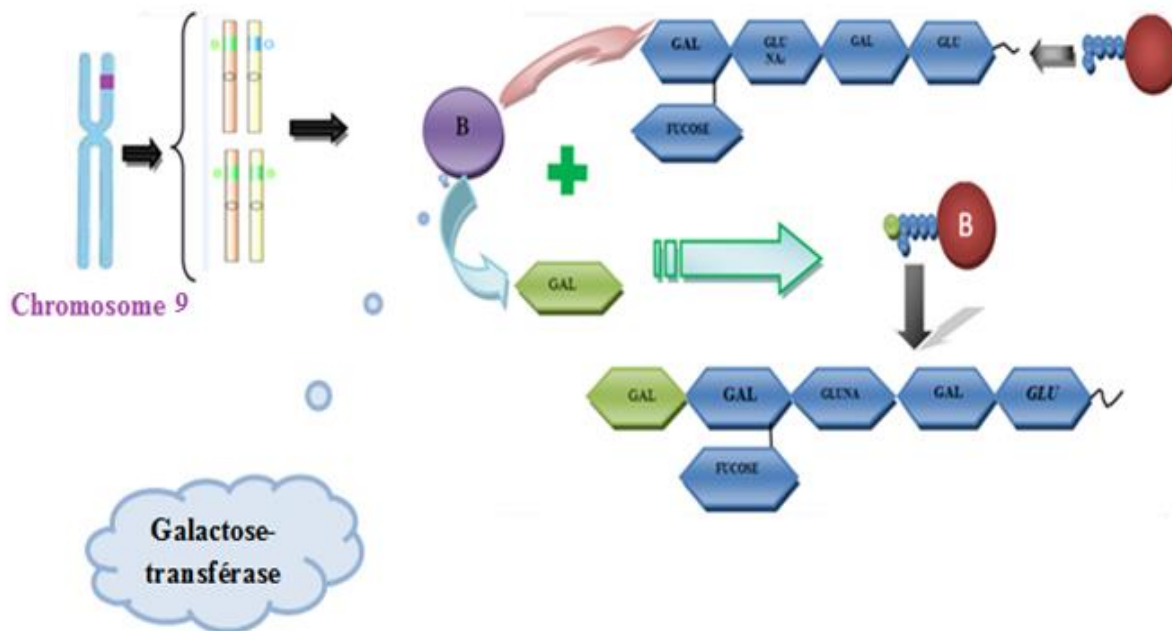


Figure II. 4: Source génétique et la façon de former l'antigène B








Remarque : Les locus des antigènes ABO sur les surfaces des globules rouges peuvent se situer sur 800 sites [38].

II.2.2 Les Anticorps ANTI-A et ANTI-B

Les anticorps anti-A et anti-B, dirigés contre les antigènes du système ABO, sont des anticorps naturels réguliers, c'est à dire qu'ils existent de façon constante chez tout individu adulte qui ne possède pas le(s) antigène(s) A et/ou B, en dehors de toute stimulation antigénique. En fait, les antigènes A et B se trouvent largement répandus dans l'environnement, en particulier chez les bactéries. Ces anticorps dits "naturels" correspondent en réalité à une immunisation acquise vis-à-vis d'antigènes étrangers ubiquitaires.

Ainsi, les individus de groupe A produisent des anti-B, les individus de groupe B produisent des anti-A et les individus de groupe O produisent à la fois des anti -A et des anti-B. Les personnes de groupe AB ne possèdent pas d'anticorps naturel dans le système ABO.

Tableau II. 1: Système ABO

type de sang	Type A	Type B	Type AB	Type O
Rouge surface de la cellule sanguine (phénotype)	 A agglutinogènes seulement	 B agglutinogènes seulement	 A et B agglutinogènes	 Pas agglutinogènes
Anticorps plasma (phénotype)	 b agglutinine seulement	 a agglutinine seulement	<i>NONE.</i> Pas agglutinine	 a et b agglutinine seulement

Il faut noter l'intérêt clinique de ces anticorps naturels anti-A et anti-B : en se fixant à la surface d'hématies étrangères non compatibles dans le système ABO, ils sont capables d'induire une réaction d'hémolyse massive souvent mortelle.

II.2.3 les lois de compatibilité ABO

- Un sujet de groupe O possède des anti-A et anti-B et ne peut être transfusé qu'avec des globules O.
- Un sujet de groupe A possède des anti-B et ne peut être transfusé qu'avec des globules A ou O.
- Un sujet de groupe B possède des anti-A et ne peut être transfusé qu'avec des globules B ou O.
- Un sujet de groupe AB ne possède pas d'anticorps naturels et peut être transfusé avec des globules A, B, AB ou O [50, 44,39].

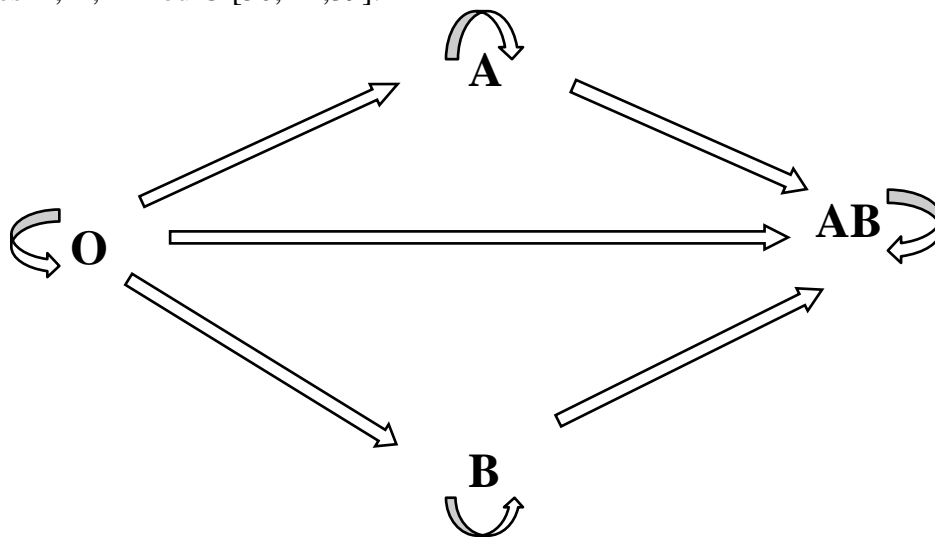


Figure II. 5: loi de compatibilité

II.2.4 Principe réactionnel des groupes sanguin

Ce système utilise trois types d'antisérums, à savoir : Anti-A, Anti-B et Anti-AB (voir la figure ci-dessous). Ces antisérums sont mis par la suite sur trois gouttes du sang dans le but de connaître le groupe de ce sang. Si l'antisérum possède une réaction avec le sang on trouve une agglutination dans la goutte du sang (la figure et le tableau ci-dessous montrent le principe réactionnel pour chaque groupe du sang) [36].

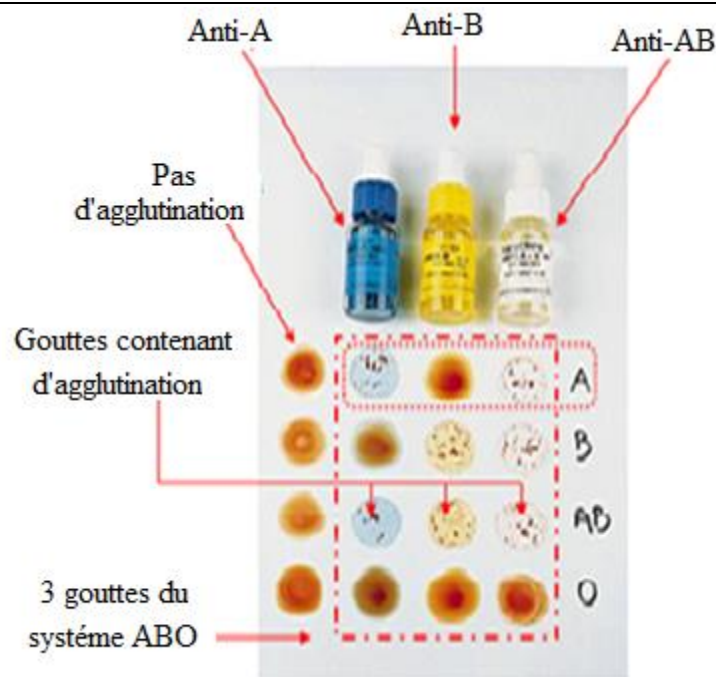


Figure II. 6: Principe réactionnel des groupes sanguins [36]

Tableau II. 2 : Principe réactionnel des groupes sanguin

	Anti-A	Anti-B	Anti-AB
Groupe A	+	-	+
Groupe B	-	+	+
Groupe AB	+	+	+
Groupe O	-	-	-

II.3 Le système Rhésus « RH »

En 1940 Landsteiner et son élève Wiener, sont à l'origine de la découverte du système rhésus. En injectant au lapin des hématies du singe *Macacus rhésus*, ils obtiennent des anticorps qu'ils dénomment anti-rhésus.

II.3.1 Aspects génétique et biochimique

Système RH comprend une cinquantaine d'antigènes de nature polypeptidique. Seuls 5 d'entre eux présentent un intérêt clinique en médecine transfusionnelle. Il s'agit des antigènes D (RH1), C (RH2), E (RH3), c(RH4) et e (RH5). Deux gènes (RHD et RHCE), adjacents et de structure très voisine, localisés sur le chromosome 1, contrôlent l'expression de ces antigènes.

Le gène RHD détermine l'expression d'une protéine exprimant l'anti gène D. On note sa présence chez 85% des individus en France dits : Rhésus positifs (Rh +). Chez les autres, dits

Rhésus négatifs « Rh - », il existe une délétion complète du locus RHD, à l'état homozygote qui conduit à l'absence de protéine RHD sur la membrane érythrocytaire et donc à l'absence d'antigène D. Le phénotype de ces individus s'écrit D- « RH :-1 » (l'appellation " d " est incorrecte car il n'existe pas d'antigène d).

II.3.2 Anticorps ANTI-Rhésus

Contrairement aux anticorps anti-A ou anti-B dits naturels, la grande majorité des anticorps dans le système Rhésus résulte d'une réponse immunitaire induite par une grossesse ou une transfusion sanguine incompatible. Cependant, pour une raison inconnue, il n'est pas rare de détecter des anticorps "naturels" anti-E par exemple, chez des sujets E négatifs qui n'ont jamais été en contact avec l'antigène E.

II.3.3 Anticorps ANTI-Rhésus

Contrairement aux anticorps anti-A ou anti-B dits naturels, la grande majorité des anticorps dans le système Rhésus résulte d'une réponse immunitaire induite par une grossesse ou une transfusion sanguine incompatible.

II.3.4 Réalisation du groupage sanguin ABO-RH1

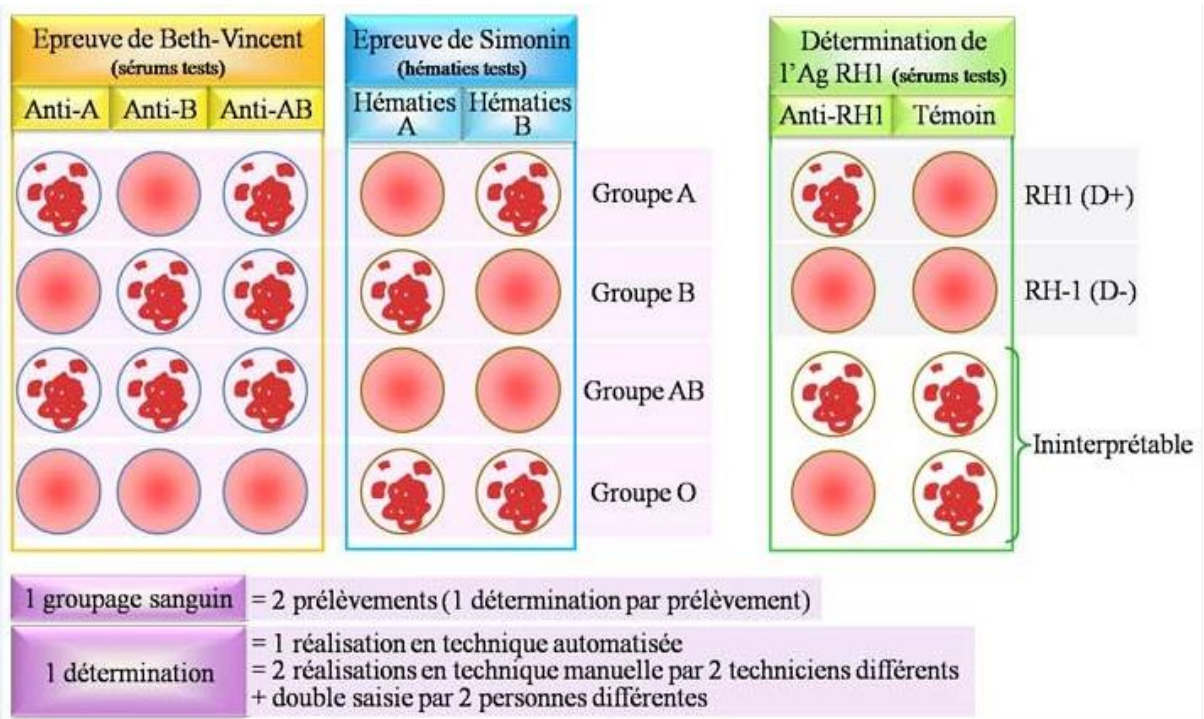


Figure II. 7: Réalisation du groupage sanguin ABO-RH1

De ce fait, nous pouvons dire qu'il existe huit « 08 » groupes sanguins avec les compatibilités suivantes :

Tableau II. 3: lois de compatibilité des groupes sanguins

		DONNEUR							
		O-	O+	B-	B+	A-	A+	AB-	AB+
RECEVEUR	AB+	●	●	●	●	●	●	●	●
	AB-	●		●		●		●	
	A+	●	●			●	●		
	A-	●				●			
	B+	●	●	●	●				
	B-	●		●					
	O+	●	●						
	O-	●							

II.4 Le système KELL

Il s'agit du système le plus immunogène après le système Rhésus. Le système Kell possède 2 antigènes principaux : K « KEL1 » et k « KEL2 », portés par une glycoprotéine membranaire dont l'expression se trouve restreinte à la lignée érythrocytaire.

Les anticorps anti-K « KEL1 » fréquents et dangereux, occasionnent des accidents hémolytiques post transfusionnels, des anémies fœtales sévères (avec pan cytopénie) et des maladies hémolytiques du nouveau-né. Ceci justifie le respect du phénotype Kell, comme le phénotype Rhésus, en particulier chez les femmes avant la ménopause et chez les sujets polytransfusés.

II.5 Autres systèmes d'intérêt clinique en transfusion sanguine

Trois autres systèmes d'antigènes " secondaires " doivent être connus et pris en considération dans les conflits immunologiques potentiels provoqués par une transfusion ou une grossesse incompatible : les systèmes Duffy « FY », Kidd « JK » et MNS [37].

- Le système DUFFY,
- Le système KIDD,
- LE SYSTEME MNS.

Conclusion

Cette étude nous a permis d'avoir une vue globale sur quelques systèmes de groupage qui permettent la détection de groupe sanguin. Nous avons donné pour chaque groupe les antigènes et les anticorps de chaque système et aussi le principe réactionnel des groupes sanguins avec les lois de comptabilité et la relation entre le donneur et le récepteur.

Dans ce travail, on s'intéresse beaucoup plus aux systèmes de groupage ABO/Hh et Rhésus « RH ». Les autres systèmes ne sont pas trop utilisés dans le monde réel, mais on peut les exploiter dans des cas d'une personne à besoin de beaucoup de sang « dans une fois », ou une personne a besoin d'opérations de transfusion de manière fréquente.

Chapitre III

Support Vector Machines

- *Introduction.*
- *Pourquoi les Machine à Vecteurs de Support (SVM) ?*
- *Apprentissage statistique et SVM.*
- *Notions de base.*
- *Pourquoi maximiser la marge ?*
- *Linéarité et non-linéarité*
- *Temps de calcul et convergence.*
- *Les domaines d'applications.*
- *Conclusion.*

Introduction

Dans le premier chapitre nous avons abordé brièvement quelques méthodes de classification alors que dans ce chapitre, nous allons prendre plus en détail une méthode de classification SVM qui est utilisée pour résoudre le problème de cette travaille.

Les machines à vecteur support se situent sur l'axe de développement de la recherche humaine des techniques d'apprentissage. Les SVMs sont dans leur origine utilisées pour la classification binaire et. Aujourd'hui, elles sont utilisées dans différents domaines de recherche et d'ingénierie tel que le diagnostic médical, le marketing, la biologie, la reconnaissance de caractères manuscrits et de visages humains. Le présent chapitre introduit les machines à vecteur de support, et leurs origines théoriques.

III.1 Pourquoi les Machine à Vecteurs de Support « SVM » ?

L'algorithme des machines à vecteurs de support a été développé dans les années 90 par le russe Vladimir Vapnik. Initialement, les SVM ont été développé comme un algorithme de classification binaire supervisée. Il s'avère particulièrement efficace de par le fait qu'il peut traiter des problèmes mettant en jeu de grands nombres de descripteurs, qu'il assure une solution unique (pas de problèmes de minimum local comme pour les réseaux de neurones) et il a fourni de bons résultats sur des problèmes réels [38].

L'algorithme sous sa forme initiale revient à chercher une frontière de décision linéaire entre deux classes, mais ce modèle peut considérablement être enrichi en se projetant dans un autre espace permettant d'augmenter la séparabilité des données. On peut alors appliquer le même algorithme dans ce nouvel espace, ce qui se traduit par une frontière de décision non linéaire dans l'espace initial [38].

III.2 Apprentissage statistique et SVM

La notion d'apprentissage étant importante, nous allons commencer par effectuer un rappel. L'apprentissage par induction permet d'arriver à des conclusions par l'examen d'exemples particuliers. Il se divise en apprentissage supervisé et non supervisé. Le cas qui concerne les SVM est l'apprentissage supervisé. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie. Le but est d'apprendre une fonction qui correspond aux exemples vus

et qui prédit les sorties pour les entrées qui n'ont pas encore été vues. Les entrées peuvent être des descriptions d'objets et les sorties la classe des objets donnés en entrée [15]. Donc on peut schématiser par :

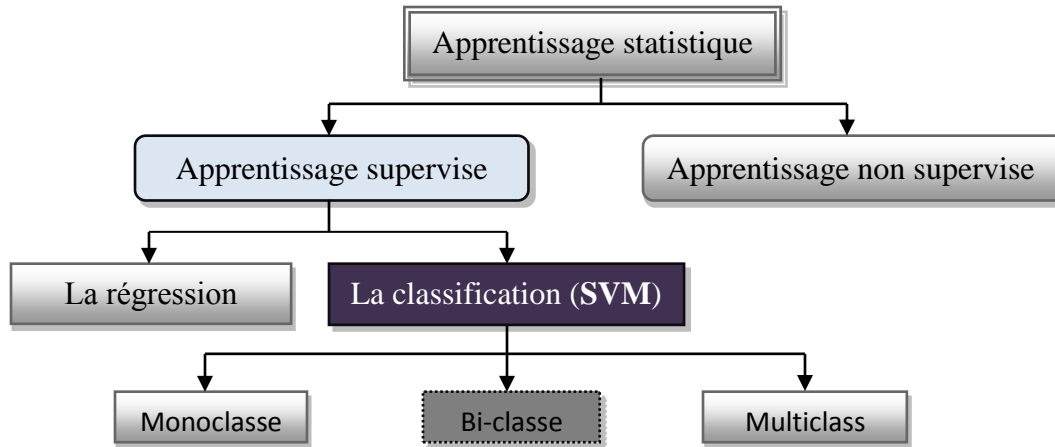


Figure III. 1: Schéma des différents types d'apprentissage

III.2.1 Objectif de l'apprentissage statistique

Effectuer une classification consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi plusieurs. Le cas le plus simple consiste à discriminer deux classes. D'une manière plus formelle, la classification bi-classe revient à estimer une fonction $f : x \rightarrow \{+1, -1\}$ à partir d'un ensemble d'apprentissage constitué de couples (x_i, y_i) , suivant une distribution de probabilité $P(x, y)$ inconnue, tels que : $(x_i, y_i) \in X \times Y$ où $i = 1, \dots, N_x$ et $y = \{+1, -1\}$,

De sorte à ce que f classe correctement des exemples inconnus (x_t, y_t) . Par exemple, on peut assigner x_t à la classe (+1) si $f(x_t) \geq 0$, et à la classe (-1) sinon. Les exemples inconnus sont supposés suivre la même distribution de probabilité $P(x, y)$ que ceux de l'ensemble d'apprentissage. La meilleure fonction f est celle obtenue en minimisant le risque :

$$(2.1). \quad R[f] = \int L[f(x), y] dP(x, y).$$

Où L désigne une fonction de coût, comme par exemple: $L[f(x), y] = (f(x) - y)^2$.

III.2.2 Sur-apprentissage et sous-apprentissage

Quand on parle de la phase d'apprentissage deux notions de base très importante doivent être discutées :

- Le sur-apprentissage, et
- Le sous-apprentissage.

A. Sur-apprentissage :

Le sur -apprentissage (appelé overfitting en anglais) survient lorsqu' on cherche à trop « coller » aux données d'entraînement (comme dans le cas d'un apprentissage par cœur) donc c'est le cas où la classificateur est classifieur les exemples d'autre classe à la classe, vue (figure suivant) Le modèle de haut degré est en situation de sur-apprentissage

B. Sous-apprentissage :

C'est la situation au contraire si la classe de fonctions considérée par l'algorithme d'apprentissage n'est pas assez « riche » pour pouvoir décrire la diversité présente dans les données alors que c'est le cas où la classificateur est classifieur les exemples d'un classe à l'autre classe, vue (figure suivant) Le modèle linéaire est en situation de sous-apprentissage

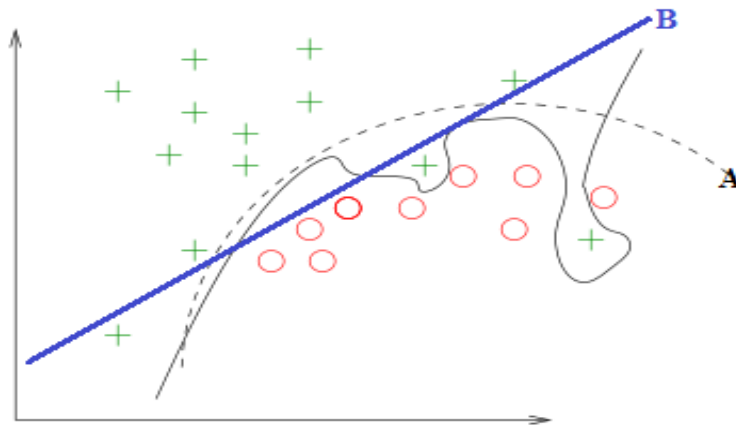


Figure III. 2: A: Cas Sur-apprentissage, B : Cas sous-apprentissage

III.3 Notions de base

D'abord nous allons voir quelques notions de base de SVM comme hyperplan, marge, support vecteur et séparateur Optimale :

III.3.1 Hyperplan

Un hyperplan qui sépare les deux ensembles de points [15] (les deux classes) qu'est montré dans figure III.3.

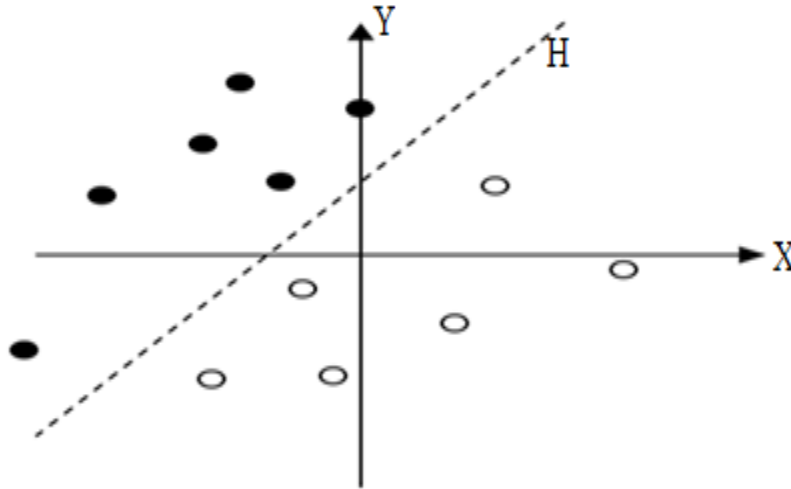


Figure III. 3: Exemple d'un hyperplan séparateur [15]

Supposons pour l'instant que les échantillons de l'ensemble d'apprentissage sont séparables par un hyperplan (Figure III.4), i.e. on choisit des fonctions de décision de la forme :

$$(2.4). \quad f(x) = \langle w, x \rangle + b.$$

III.3.2 Marge

La marge est la distance minimale entre les échantillons de l'ensemble d'apprentissage et la frontière de décision. La marge peut à son tour être mesurée grâce au vecteur poids w : puisque nous supposons que les échantillons sont séparables, on peut redéfinir w et b de sorte à ce que les échantillons x les plus proches de l'hyperplan satisfassent $|\langle w, x \rangle + b| = 1$.

III.3.3 Support vecteur

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

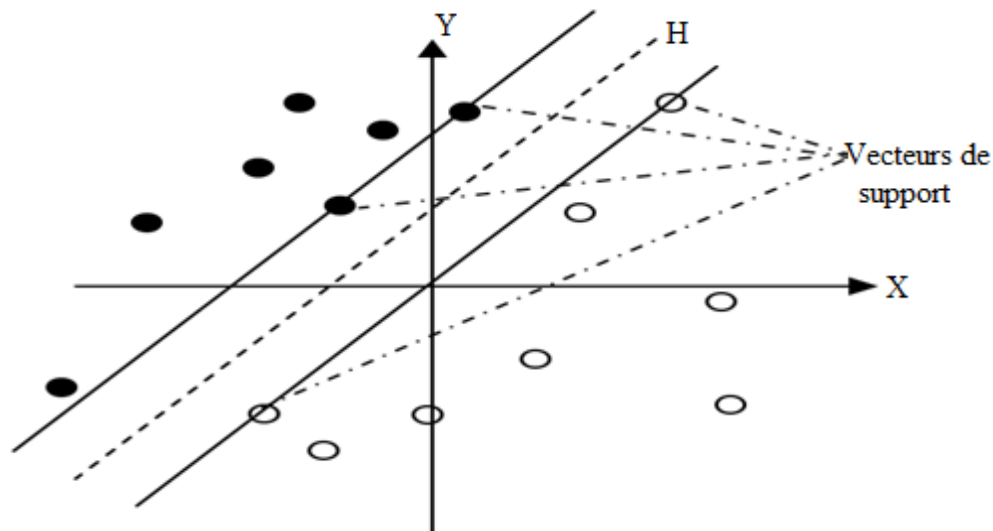


Figure III. 4: Exemple de vecteurs de support [15]

III.3.4 Séparateur Optimale « vaste marge »

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe « au milieu » des points des deux classes d'exemples « voire figure III.5 ». Intuitivement, cela revient à chercher l'hyperplan le « plus sûr » [39].

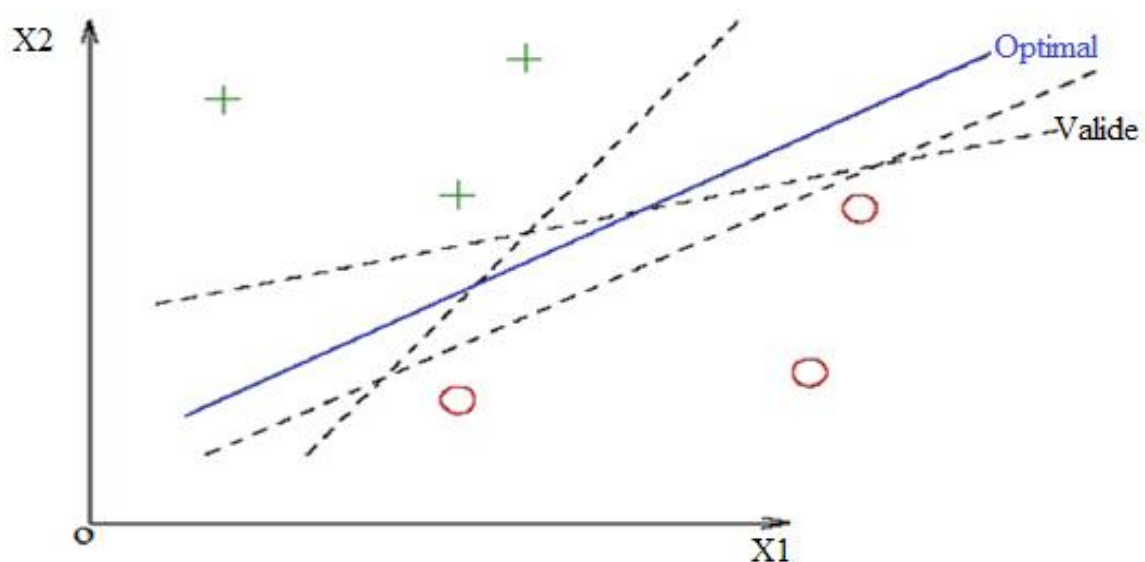


Figure III. 5: Exemple d'hyperplan séparateur

En effet, supposons qu'un exemple n'ait pas été décrit parfaitement, une petite variation ne modifiera pas sa classification si sa distance à l'hyperplan est grande. Formellement, cela

revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale [39].

On appelle cette distance « marge » entre l'hyperplan et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge [39].

III.4 Pourquoi maximiser la marge ?

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé [15].

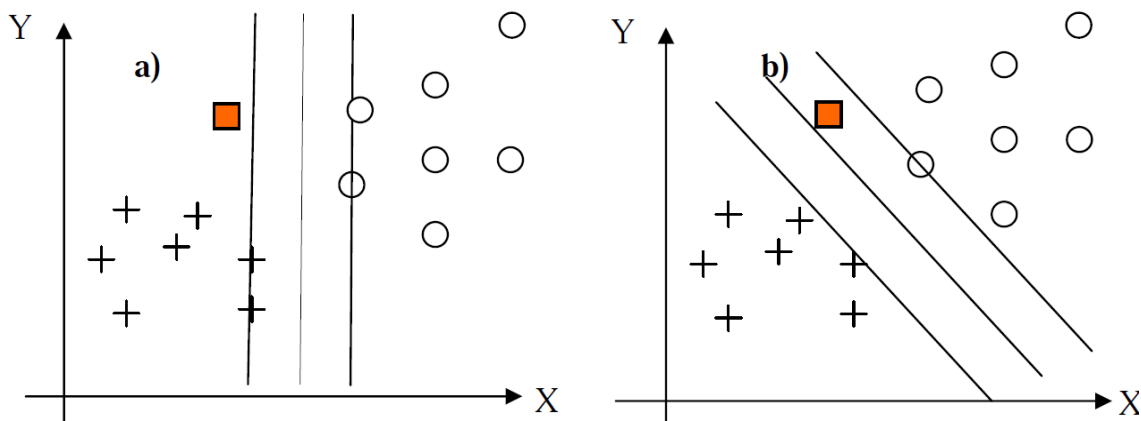


Figure III. 6: a) Hyperplan avec faible marge, b) Meilleur hyperplan séparateur [15]

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma suivant, le nouvel élément sera classé dans la catégorie des «+».

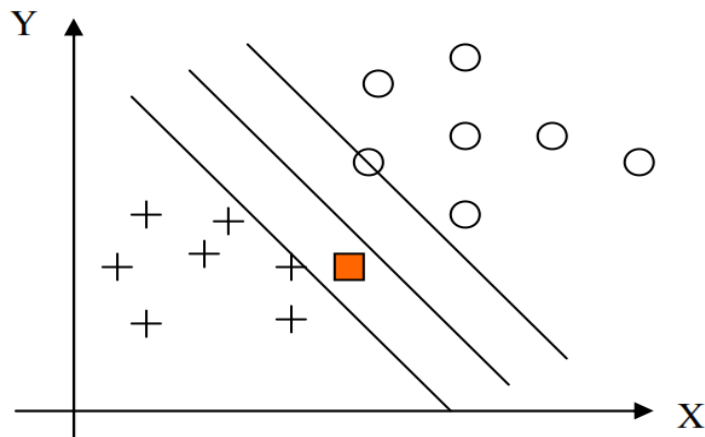


Figure III. 7: Exemple de classification d'un nouvel élément

III.5 Linéarité et non-linéarité

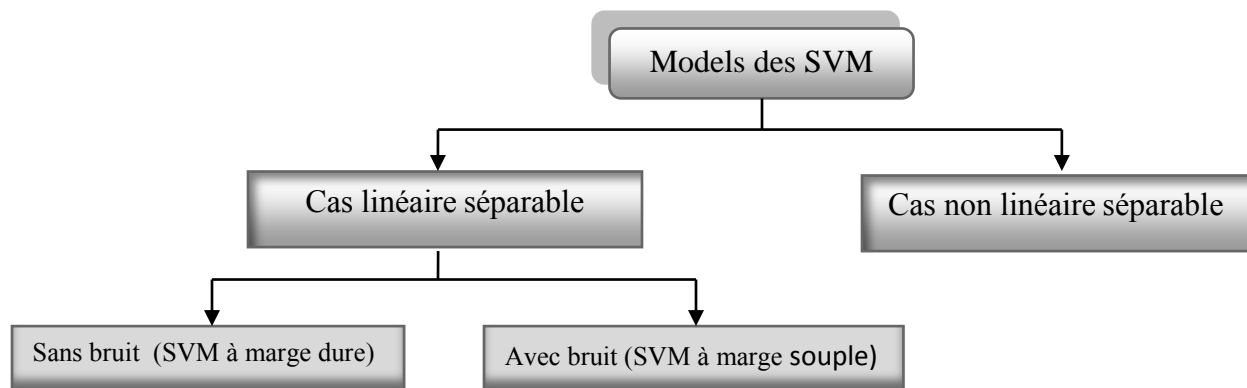


Figure III. 8: Schéma hiérarchique des différents modèles de SVM

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. Les premiers sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [15].

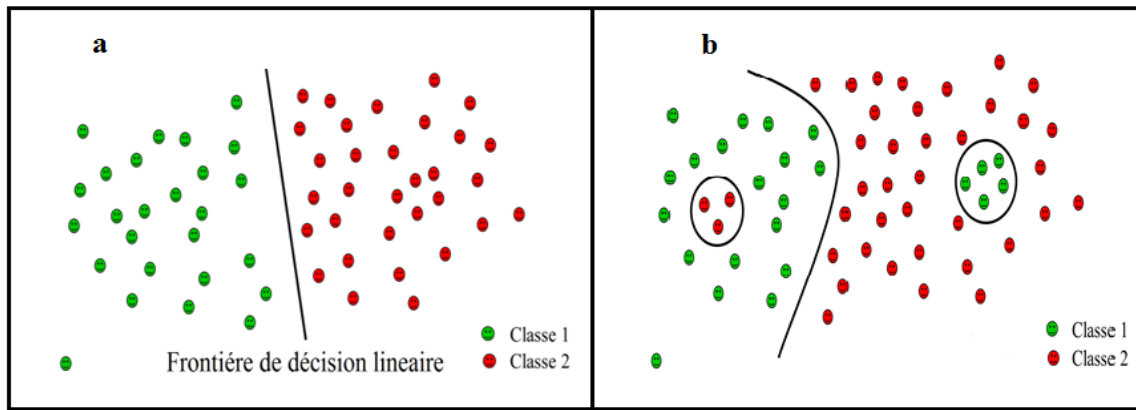


Figure III. 9: a) Cas linéairement séparable, b) Cas non linéairement séparable [21,15]

III.5.1 Cas linéairement séparable

Ces cas sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire et dans un modèle linéaire, on a $f(x) = w \cdot x + b$.

- L'hyperplan séparateur (frontière de décision) a donc pour équation $w \cdot x + b = 0$.
- La distance d'un point au plan est donnée par $d(x) = |w \cdot x + b| / \|w\|$
- L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (marge) est maximale.

1.1 SVM à marge dure (Sans bruit) :

C'est le cas où on ne trouve pas des exemples dans l'espace de généralisation.

A. Maximisation de la marge

La marge est la distance du point le plus proche à l'hyperplan. Soient x_1 et x_2 eux points de classes différentes : ($f(x_1) = +1$ et $f(x_2) = -1$) , $w \cdot x_1 + b = +1$ et $w \cdot x_2 + b = -1$

donc $(w \cdot (x_1 - x_2)) = 2$, d'où : $(w / \|w\| \cdot (x_1 - x_2)) = 2 / \|w\|$ [41].

Considérons maintenant deux échantillons x_1 et x_2 de classes différentes telles qu'on ait $w \cdot x_1 + b = +1$ et $w \cdot x_2 + b = -1$.

- La distance entre x_i et l'hyperplan est donnée comme ça :

$$(2.5). \quad d_i(w, b, x_i) = \frac{|w \cdot x_i + b|}{\|w\|}$$

- La marge γ correspond alors à la distance entre x_1 et x_2 mesurée perpendiculairement à l'hyperplan :

$$\begin{aligned}
\delta(w, b) &= \min_{x_i, y_i = -1} d(w, b, x_i) + \min_{x_i, y_i = 1} d(w, b, x_i) \\
&= \min_{x_i, y_i = -1} \frac{|w \cdot x_i + b|}{\|w\|} + \min_{x_i, y_i = 1} \frac{|w \cdot x_i + b|}{\|w\|} \\
&= \frac{1}{\|w\|} [\min_{x_i, y_i = -1} |w \cdot x_i + b| + \min_{x_i, y_i = 1} |w \cdot x_i + b|] \\
&= \frac{1}{\|w\|} [1 + 1] \\
&= \frac{2}{\|w\|} \quad [22].
\end{aligned}$$

On peut donc en déduire que maximiser la marge (maximiser $\frac{2}{\|w\|}$) revient à minimiser $\|w\|$ alors que minimiser $(\frac{1}{2} \|w\|^2)$ sous certaines contraintes que nous verrons dans les paragraphes suivants.

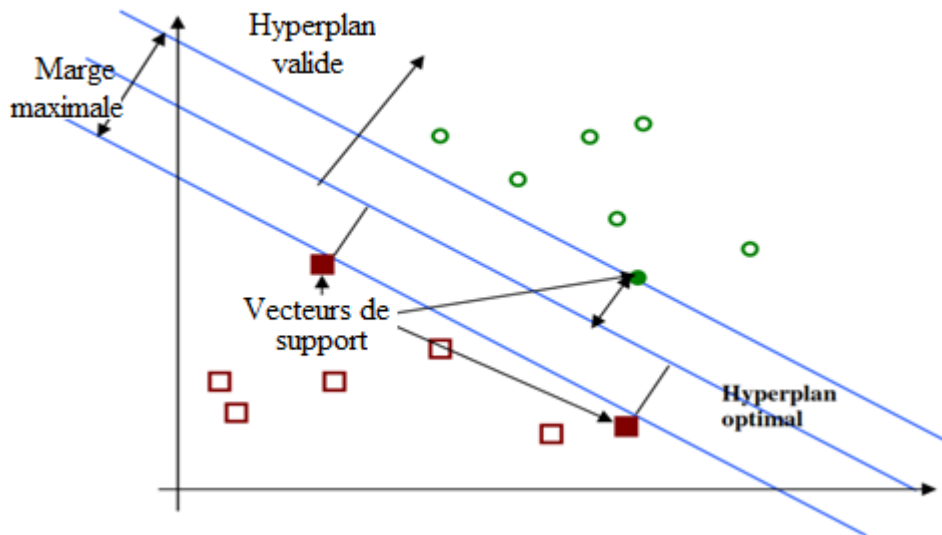


Figure III. 10: Exemple de marge maximal (SVM binaire à marge dure)

B. Problème primal

Un point (x, y) est bien classé si et seulement si $yf(x) > 0$. Comme le couple (w, b) est défini à un coefficient multiplicatif près, on s'impose $yf(x) \geq 1$. On en déduit (en s'appuyant également sur le paragraphe précédent), le problème de minimisation sous contraintes suivantes :

$$(2.6) \quad \begin{cases} \min \frac{1}{2} \|w\|^2 \\ \forall i, y_i (w \cdot x_i + b) \geq 1 \end{cases}$$

Il peut être en effet plus aisé de minimiser $\|w\|^2$ plutôt que directement $\|w\|$ [39,40].

C. Problème dual

On passe du problème primal au problème dual en introduisant des *multiplieurs de Lagrange* pour chaque contrainte. Ici on a une contrainte par exemple d'apprentissage :

$$(2.7) \quad \begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \forall i, 0 \leq \alpha_i \leq c \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

C'est un problème de programmation quadratique de dimension n (nombre d'exemples). et qu'introduit la fonction d'objective :

$$(2.8) \quad Q(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i \{y_i (w^T x_i + b) - 1\}$$

Où les α_i sont les multiplieurs non négatifs de Lagrange. L'optimum de la fonction objective Q peut être obtenu en la minimisant par rapport à w et b et en la maximisant par rapport aux α_i . On définit ainsi la matrice suivante appelée « matrice hessienne » : $(x_i \cdot x_j)_{i, j}$ qui représente la matrice des produits des entrées X (La notation matricielle permettant de résoudre plus facilement le problème en informatique) [15].

On montre que si les α_i^* sont solutions de ce problème alors on a :

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad [39].$$

Seuls les α_i à correspondant aux points les plus proches sont non nuls. On parle de vecteurs de support. La fonction de décision associée est donc :

$$f(x) = \sum_{i=1}^n \alpha_i^* y_i x_i \cdot x + b$$

1.2 Cas non séparable :- SVM à marge molle «souple» (Avec bruit)

En réalité, un hyperplan séparateur n'existe pas toujours, et même s'il existe, il ne représente pas généralement la meilleure solution pour la classification. En plus une erreur d'étiquetage dans les données d'entraînement (un exemple étiqueté +1 au lieu de -1 par exemple) affectera crucialement l'hyperplan. Dans le cas où les données ne sont pas linéairement séparables, ou contiennent du bruit (outliers : données mal étiquetées) les contraintes de l'équation (2.5) ne peuvent être vérifiées, et il y a nécessité de les relaxer un peu. Ceci peut être fait en admettant une certaine erreur de classification des données (figure III.11) ce qui est appelé "SVM à marge souple (Soft Margin).

On part du problème primal linéaire et on introduit des variables « ressort » pour assouplir les

contraintes [39] :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \varepsilon_i \\ \forall i, y_i (w \cdot x_i + b) \geq 1 - \varepsilon_i \end{cases}$$

On pénalise par le dépassement de la contrainte. La seule différence est la borne supérieure C sur les α . Où C est un paramètre positif libre (mais fixe) qui représente une balance entre les deux termes de la fonction objective (la marge et les erreurs permises) c.-à-d. entre la maximisation de la marge et la minimisation de l'erreur de classification

On en déduit le problème dual qui a la même forme que dans le cas séparable [39]:

$$\begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \forall i, 0 \leq \alpha_i \leq c \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

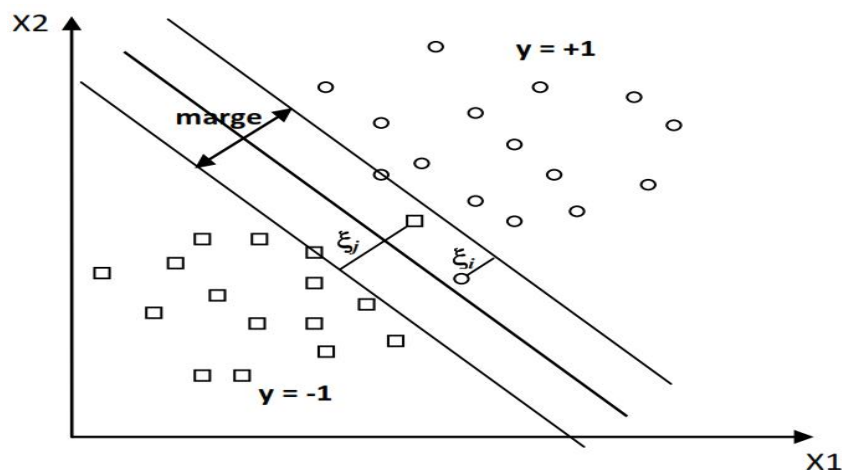


Figure III. 11: SVM binaire à marge souple

III.5.2 Cas non linéairement séparable

Il existe néanmoins des cas où on ne peut pas classer les entrées de façon linéaire [39]. Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de re-description ».

En effet, intuitivement, plus la dimension de l'espace de re-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant [15].

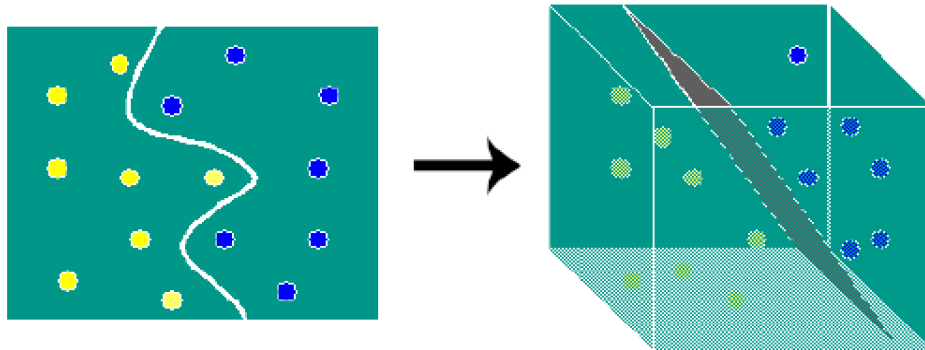


Figure III. 12: Exemple de changement de l'espace de données

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau [15].

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien [15].

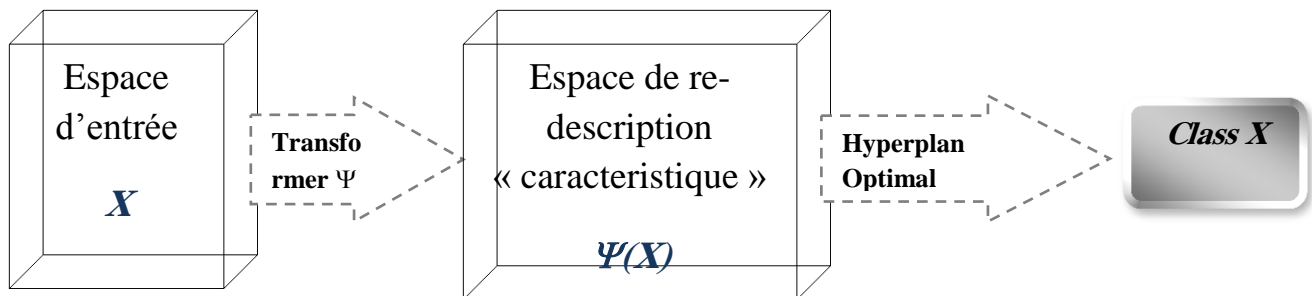


Figure III. 13: le principe général de technique SVM dans ce cas

2.1 Fonction noyau (kernel)

Dans les cas linéaire, on pouvait transformer les données dans un espace où la classification serait plus aisée. Dans ce cas, l'espace de re-description utilisé le plus souvent est \mathbb{R}

(ensemble des nombres réels). Il se trouve que pour des cas non linéaires, cet espace ne suffit pas pour classer les entrées. On passe donc dans un espace de grande dimension [41].

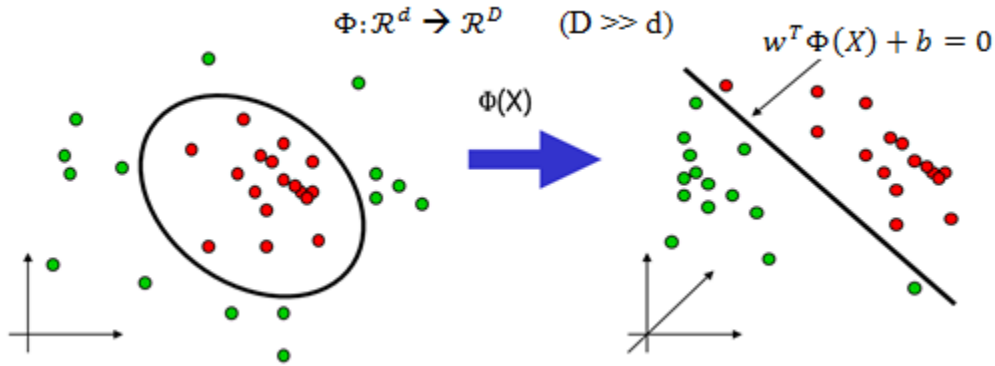


Figure III. 14: Illustration de passage à R3 [41]

Le passage dans $F = R^3$ (i.e. $D=3$). Rend possible la séparation linéaire des données. On doit donc résoudre :

$$\left\{ \begin{array}{l} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \varphi(x_i) \varphi(x_j) \\ \forall i, 0 \leq \alpha_i \leq c \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right.$$

Et la solution à la forme: $f(x) = \sum_{i=1}^n \alpha_i^* y_i \varphi(x_i) \cdot \varphi(x) + b$

Le problème et sa solution ne dépendent que du produit scalaire $\varphi(x) \times \varphi(x')$. Plutôt que de choisir la transformation non-linéaire $\varphi : X \rightarrow F$, on choisit une fonction $K : X \times X \rightarrow R$ (nombres réels) appelés fonction noyau.

Elle représente un produit scalaire dans l'espace de représentation intermédiaire. Du Coup k est linéaire (ce qui nous permet de faire le rapprochement avec le cas linéaire des paragraphes précédents). Cette fonction traduit donc la répartition des exemples dans cet espace $k(x, x') = \varphi(x) \cdot \varphi(x')$. Lorsque k est bien choisie, on n'a pas besoin de calculer la représentation des exemples dans cet espace pour calculer φ .

Exemple:

soit $x = (x_1, x_2)$ et $\varphi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$. Dans l'espace intermédiaire, le produit scalaire donne

$$\begin{aligned}\varphi(x)\varphi(x') &= x_1^2 x_1'^2 + 2x_1 x_2 x_1' x_2' + x_2^2 x_2'^2 \\ &= (x_1 x_1' + x_2 x_2')^2 = (x \cdot x')^2\end{aligned}$$

On peut donc calculer $\varphi(x) \times \varphi(x')$ sans calculer φ : $k(x, x') = (x \cdot x')^2$. K représentera donc le noyau pour les entrées correspondantes mais devra néanmoins remplir certaines conditions [42].

Exemples de noyaux:

- Linéaire $K(x, x') = x \cdot x'$.
- Polynomial $K(x, x') = (x \cdot x')^d$ ou $(c + x \cdot x')^d$.
- Gaussien $K(x, x') = e^{-\|x-x'\| \cdot \frac{2}{\sigma}}$.
- Laplacien $K(x, x') = e^{-\|x-x'\| \cdot \frac{1}{\sigma}}$.

On remarque en général que le noyau gaussien donne de meilleurs résultats et groupe les données dans des paquets nets. En pratique, on combine des noyaux simples pour en obtenir de plus complexe [41].

2.2 Condition de Mercer

Une fonction k symétrique est un noyau si $(k(x_i, x_j))_{i,j}$ est une matrice définie positive. Dans ce cas, il existe un espace F et une fonction φ tels que $k(x, x') = \varphi(x) \times \varphi(x')$ [41]

Problèmes :

- Cette condition est très difficile à vérifier.
- Elle ne donne pas d'indication pour la construction de noyaux.
- Elle ne permet pas de savoir comment est φ .

III.6 Complexité

Nous allons évaluer la complexité (temps de calcul) de l'algorithme SVM. Elle ne dépend que du nombre des entrées à classer (d) et du nombre de données d'apprentissage (n).

On montre que cette complexité est polynomiale en n taille de la matrice hessienne = n^2

$$dn^2 \leq \text{Complexité} \leq dn^3$$

En effet, on doit au moins parcourir tous les éléments de la matrice ainsi que toutes les entrées. Pour un très grand nombre de données d'apprentissage, le temps de calcul explose. C'est Pourquoi les SVMs sont pratiques pour des « petits » problèmes de classification [15].

III.7 Architecture générale d'une machine à vecteur support

Une machine à vecteur support, recherche à l'aide d'une méthode d'optimisation, dans un ensemble d'exemples d'entraînement, des exemples, appelés vecteurs support, qui caractérisent la fonction de séparation. La machine calcule également des multiplicateurs associés à ces vecteurs. Les vecteurs supports et leurs multiplicateurs sont utilisés pour calculer la fonction de décision pour un nouvel exemple. Le schéma de la relation (2.6) résume l'architecture générale d'une SVM dans le cas de la reconnaissance des chiffres manuscrits [12].

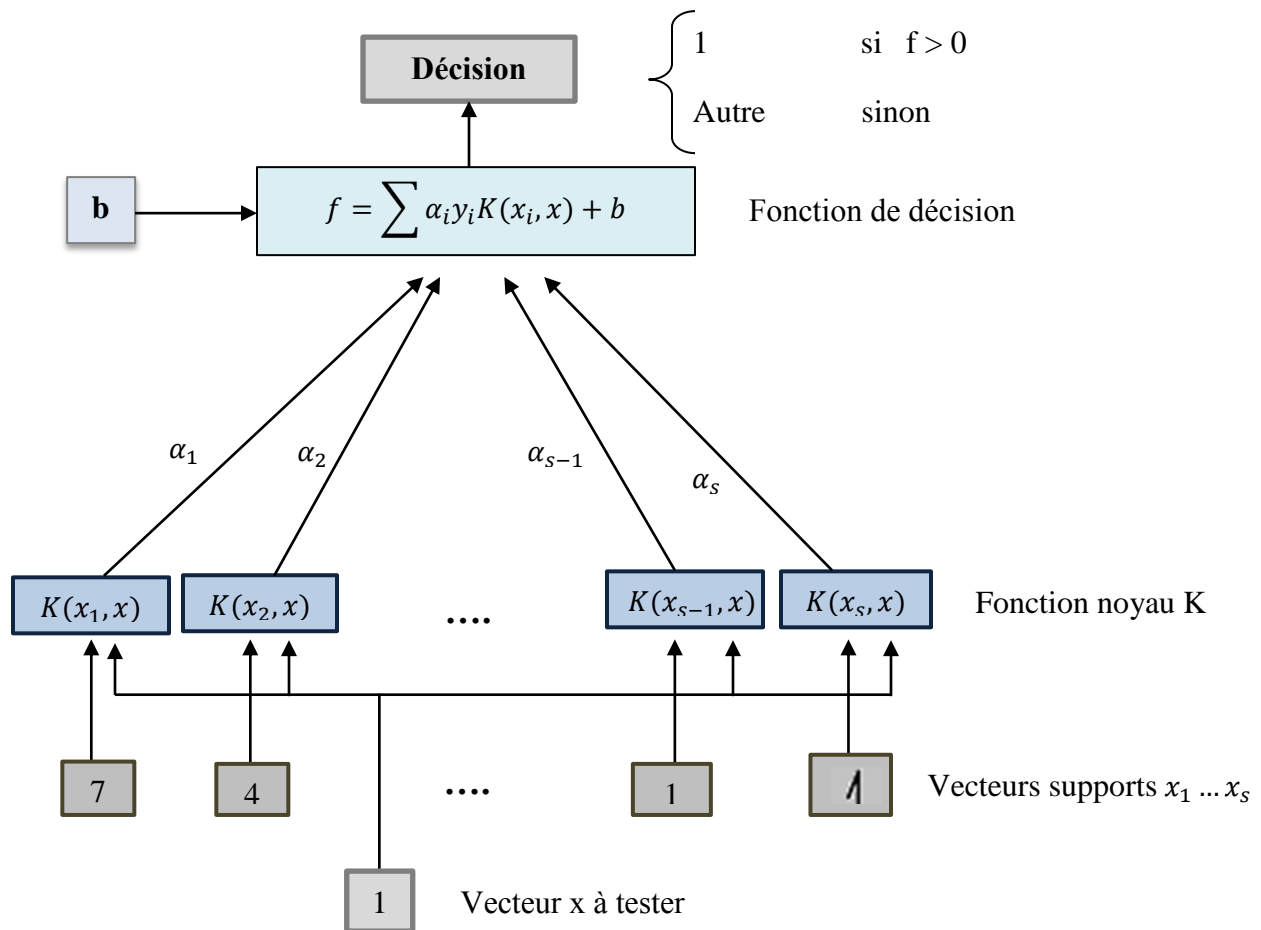


Figure III. 15: Architecture d'une machine à vecteur support

III.8 Les domaines d'applications

SVM est une méthode de classification qui montre de bonnes performances dans la résolution de problèmes variés. Cette méthode a montré son efficacité dans de nombreux domaines d'applications tels que le traitement d'image, la catégorisation de textes ou le diagnostics médicales et ce même sur des ensembles de données de très grandes dimensions.

Conclusion

Dans ce chapitre, nous avons tenté de présenter de manière simple et complète le concept de système d'apprentissage introduit par Vladimir Vapnik, les Machines à Vecteurs de Support. Nous avons donné une vision générale et une vision purement mathématiques des SVM.

Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données. Nous avons exposé les cas linéairement séparable et les cas non linéairement séparables qui nécessitent l'utilisation de fonction noyau (kernel) pour changer d'espace. Cette méthode est applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multi classe. Nous nous sommes ensuite intéressés aux différents domaines d'application. Il existe des extensions que nous n'avons pas présentées, parmi lesquelles l'utilisation des SVM pour des tâches de régression, c'est-à-dire de prédiction d'une variable continue en fonction d'autre variable

Chapitre IV:

Conception & Mise en œuvre

- *Introduction*
- *Mise en œuvre du système*
- *Bilan et résultats*
- *Conclusion*

Introduction

Dans les chapitres précédents nous avons montré comment déterminer le groupe sanguin manuellement, après nous avons présenté la méthode de classification binaire SVM, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik introduite en 1995. Dans ce chapitre nous allons présenter une conception par affinement successif du système en donnant son architecture générale, puis nous détaillons en étudiant séparément chacun de ses composants, ensuite nous allons voir les résultats obtenus.

IV.1 Mise en œuvre du système

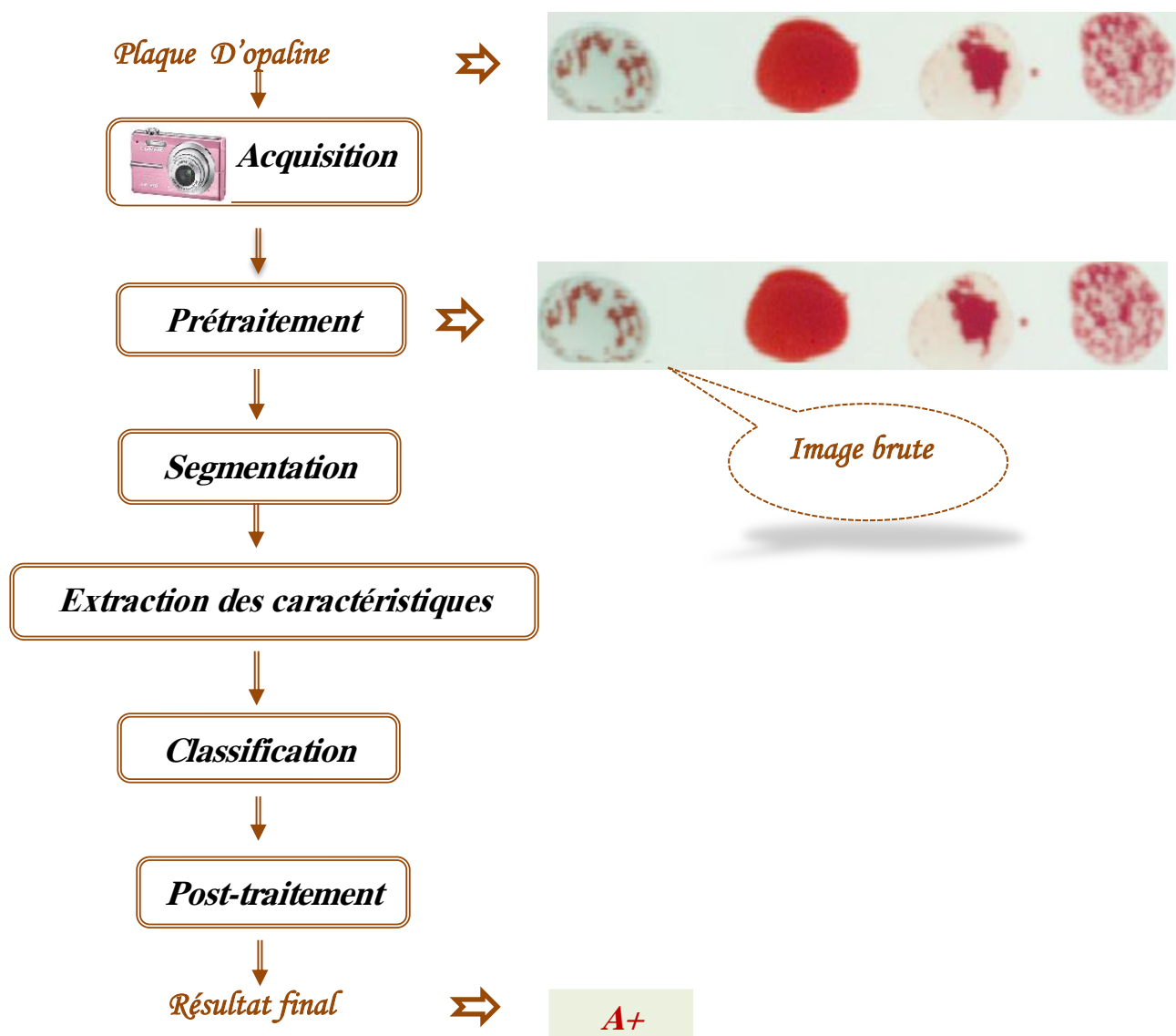


Figure IV. 1: Illustration des modules du système

IV.1.1 Acquisition

Ce module consiste tout simplement à acquérir l'image du groupe sanguin et la sauvegarder en un format d'image connu (par exemple : jpeg, gif, bmp...etc.) et /ou la transformer en un matrice de pixels dans le but de l'utiliser directement.

IV.1.2 Pré-traitement

Le prétraitement est le processus qui couvre de la forme originale de l'image d'entrée à une forme squelettique ce qui permet de simplifier les phases suivants [27]. Il s'agit essentiellement de réduire le bruit superposé aux données et essayer de ne garder que l'information significative de la forme représentée. Le bruit peut être dû aux conditions d'acquisition (éclairage, pose incorrecte de l'objet, ...) [14, 43]. En générale pour réaliser cette phase on passe pratiquement par deux étapes :

- Binairisation,
- Filtrage,

Binairisation

L'étape de binairisation consiste à transformer la matrice de l'image brute obtenue dans la phase d'acquisition représentée en couleurs (RGB) en une image représentée en deux couleurs, « noir et blanc ». Pour le faire on utilise les formules de conversion RBG/YUV, à savoir :

$$Y = 299 * R + 0.587 * G + 0.114 * B$$

$$U = 0.492 (B - Y)$$

$$V = 0.877 (R - Y)$$

Dans la réalité, et en pratique on utilise seulement la première équation, ensuite en définissant un seuil S on peut décider si un pixel est noir ou blanc.

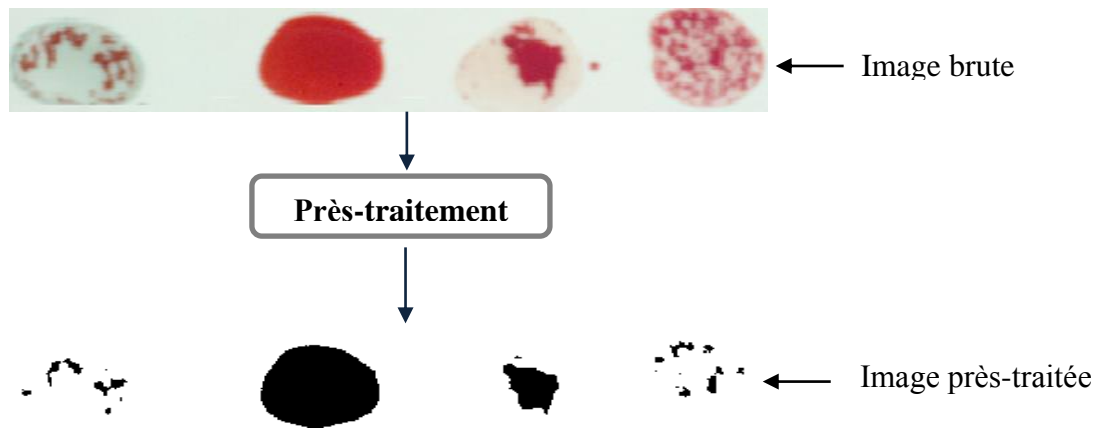


Figure IV. 2: Exemple de prétraitement

IV.1.3 Segmentation

Dans cette phase, on s'intéresse à déterminer les coordonnées des 04 gouttes du sang pour pouvoir faire cette tâche on passe par 02 étape :

- Segmentation niveau 1,
- Segmentation niveau 2.

3.1 Segmentation niveau 1

On appliquant une méthode de seuillage sur l'image binaire obtenue dans la phase précédente, nous obtenons les coordonnées des petits segments représentant les gouttes du sang.

3.2 Segmentation niveau 2

A partir des coordonnées des petits segments obtenus précédente « dans l'étape de segmentation niveau 1 », nous pouvons déterminer les coordonnées globales des segments de chacune des gouttes du sang.

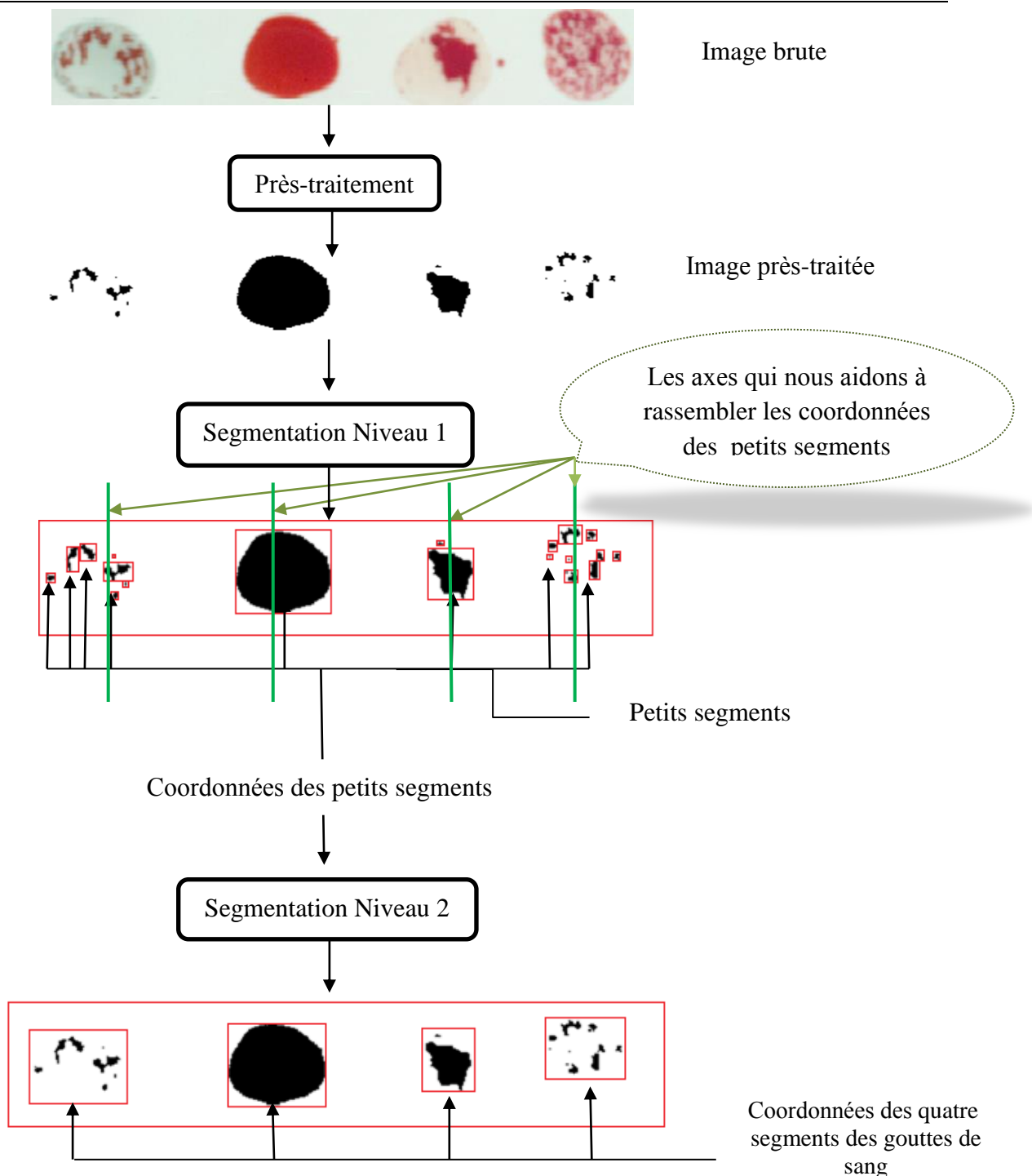


Figure IV. 3: Phase de segmentation

Ces coordonnées vont être utilisées par la suite dans la phase d'extraction des caractéristiques.

Cette tâche est réalisée à l'aide des deux modules :

- ConstructSegment
- Segmentry

Procédure ConstructSegment (Point p)

FULL_MARKER: booléen;

m: matrice d'entier [3,3];

x,y,h,k: entier;

Début

// Initialiser l'indicateur du marquage

FULL_MARKER=false;

Répétez**Pour** h de 0 à 3 pas 1 **Faire****Pour** k de 0 à 3 pas 1 **Faire**

x= (p.x-1) +h;

y= (p.y-1) +k;

Si ((x >= 0) et (x <= img.getWidth())) **Alors****Si** ((y >= 0) et (y <= img.getHeight())) **Alors****Si** (IN_BF ((img.getRGB(x,y))) **Alors****Si** ((x<>p.x)ou(y<>p.y)) **Alors**

m[h,k]=1;

Sinon

m[h,k]=0;

FinSi

FULL_MARKER=vrai;

Si (x<Xmin) **Alors** Xmin=x; **FinSi****Si** (y<Ymin) **Alors** Ymin=y; **FinSi****Si** (x>Xmax) **Alors** Xmax=x; **FinSi****Si** (y>Ymax) **Alors** Ymax=y; **FinSi**

// Changer le pixel correspondante de l'image temporaire « en noir=0 »

tmpimg.setRGB(x,y,0);

// Changer le pixel correspondante de l'image source « en blanc = 16777215 »

img.setRGB(x, y, 16777215);

// Add point to the list if it's not already in the list

exist =false: booléen;

Pour s de 0 à tsize pas 1 **Faire**

```

    Si (pv[s]≠ null) Alors
        Si (( pv[s].x = x) et ( pv[s].y =y )) Alors
            exist=true;
        Si ( (!exist) et ((x ≠ p.x) ou (y ≠ p.y)) ) Alors
            Si (in<pv.length) Alors
                pv[in] = new Point(x,y);
                in=in+1;
            Sinon
                System.out.println("\nPoint index exceeds array dimension!...\n");
                System.exit(0);
            FinSi
        FinSi
    Sinon m[h][k]=0;
    FinSi // IN_BF
FinPour
// get the next point p
in=in-1;
Si (in >= 0)
P = new Point( pv[in].x,pv[in].y);
Si (in > max_in) max_in=in;
.jusqu'un (in >= 0)

```

Fin

Où:

Img: image source,*Tmpimg*: image résultante de la segmentation,*getWidth*(): fonction permettant d'obtenir la largeur d'une image (ex: img.),*getHeight*(): fonction permettant d'obtenir la hauteur d'une image (ex: img.),*getRGB(x,y)*: fonction permettant d'obtenir les composantes du couleur du pixel en position (x,y),*setRGB(x,y,v)*:fonction permettant de changer les composantes du couleur du pixel en position (x,y) à la valeur v (v appartient à [0, 16777215]),*IN_BF*: fonction permet de connaître si la couleur d'un pixel est dans la plage du couleur noir ou non.

Le résultat de cette étape est un ensemble d'images nommée « Image_NumSeq.gif » et des fichiers nommés « Segment_NumSeq.pos » contenant la position de chaque image produite dans l'image initiale «Coordonnées des petits segments», tout ça, à l'aide d'une classe Java appelée **Segmenter** (Segmenter (nom_Image)) et ces résultat est donner à une autre classe Java appelée **Segmentry** qui fait l'association des petits segments à un segment globale représentant le goutte de sang à l'aide des axes de localisation.

IV.1.4 Extraction des caractéristiques

L'extraction des caractéristiques consiste à utiliser une techniques d'analyse (statistiques, hybride, structurelle,...etc.) pour obtenir les caractéristiques qui donnent une bonne description de l'image de la goutte pour enfin décider si il y'a d'agglutination ou non.

Dans notre système nous allons utiliser deux catégories de caractéristiques (caractéristiques hybrides):

- Caractéristiques statistiques (descripteurs de formes), et
- Caractéristiques structurelles.

4.1 Caractéristiques statistiques

Le sang semble beaucoup à une texture aléatoire, ce qui nous permettre d'utiliser des techniques d'analyse de texture afin d'extraire des caractéristiques définissant l'apparence d'une goutte du sang.

Dans ce travail, nous avons choisi d'utiliser une méthode de premier ordre dont laquelle on s'intéresse à calculer la moyenne, la variance, l'entropie,...etc. à l'aide des équations suivantes :

$$Moy = \frac{1}{N} \sum_{i,j} g(i,j),$$

$$Var = \frac{1}{N} \sum_{i,j} (g(i,j) - Moy)^2 ,$$

$$SKEW = \frac{1}{N} \sum_{i,j} (g(i,j) - Moy)^3,$$

$$KURT = \frac{1}{N} \sum_{i,j} (g(i,j) - Moy)^4$$

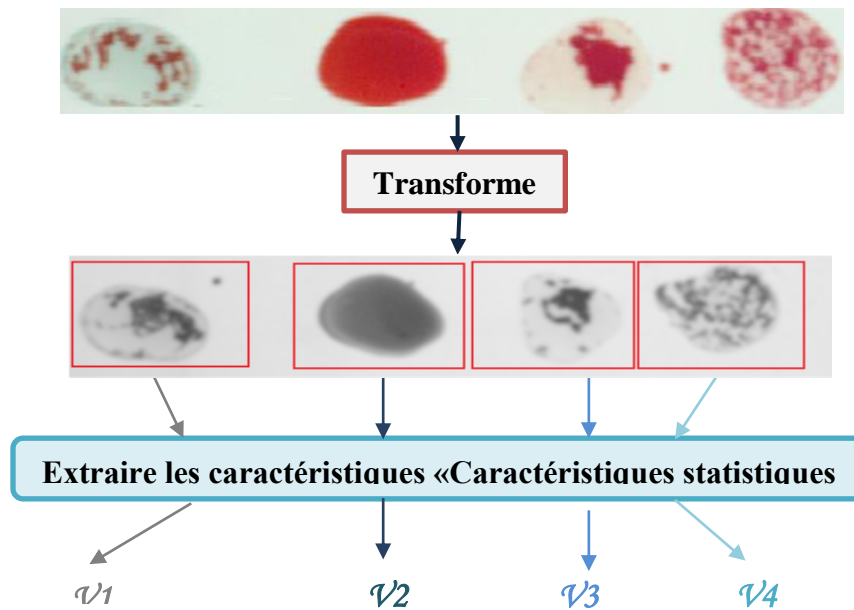


Figure IV. 4: Phase d'extraction des caractéristiques

4.2 Caractéristiques structurales

Pour enrichir en mieux la définition de la goutte du sang nous avons ajouté au vecteur caractéristique des descripteurs de structure :

- Le nombre des segments composant la goutte,
- La taille des segments composant la goutte, et
- La forme des segments composant la goutte.

4.2.1 Nombre de segments

Il représente tout simplement le nombre de segments généré dans l'étape de segmentation de niveau 1. En examinant la figure IV. 6 nous pouvons remarquer que le nombre de segments généré pour une goutte agglutinée est plus grand que celui généré pour une goutte non agglutinée.

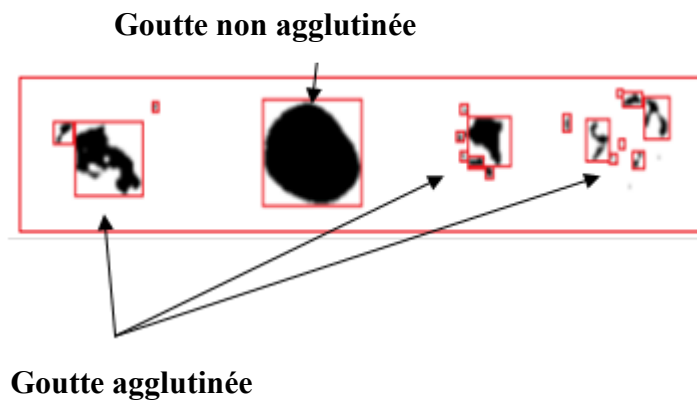


Figure IV. 5: Nombre de segments de goutte agglutinée et non agglutinée

4.2.2 Taille des segments

Dans la réalité, le nombre de segments peut être parfois confus, parce qu'il y'a des cas où le nombre de segments d'une goutte non agglutinée est égale (ou même plus grand) que celui d'une goutte agglutinée. Pour cela la taille de chaque segment représente une information clé à ce problème (voir la figure ci-dessous).

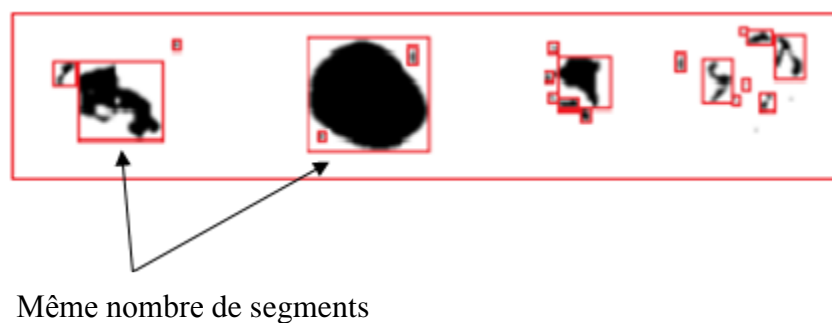


Figure IV. 6 : Taille de goutte agglutinée et non agglutinée

4.2.3 Forme des segments

Jusqu'ici on remarque que le nombre des segments et la taille de chaque segment présente une définition pas malle, mais nous avons ajouté des descripteurs de forme pour chaque segment afin d'avoir des résultats plus mieux.

Les gouttes non agglutinées sont de forme presque circulaire (plus uniforme) par rapport à la forme des segments d'une goutte agglutinée.

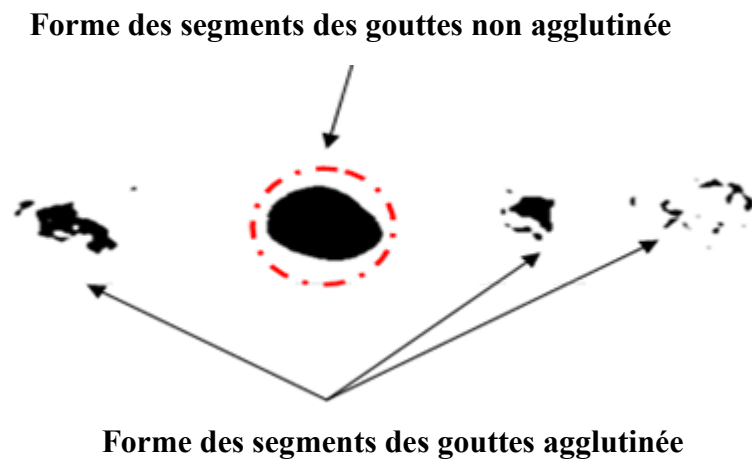


Figure IV. 7: La forme des gouttes agglutinée et non agglutinée

Dans cette phase le système va donner comme résultat un ensemble $V = \{V_1, V_2, V_3, V_4\}$ des vecteurs caractéristiques pour l'image de niveau de gris, et l'ensemble $V' = \{V'_1, V'_2, V'_3\}$ des vecteurs caractéristiques associés à l'image près-traitée.

Ensuite, le système va utiliser ces vecteurs en tant que couples : $C = \{C_1 = (V_1, V'_1), C_2 = (V_2, V'_2), C_3 = (V_3, V'_3), C_4 = (V_4, V'_4)\}$ à fin d'enrichir la définition de chaque segments d'images. Les couples décrits représentent le vecteur caractéristique de l'image en cours de traitement.

IV.1.5 Classification

Elle consiste à utiliser une méthode de classification (dans notre cas SVM) pour prendre la décision c.-à-d. à partir de vecteur caractéristique obtenu dans l'étape précédente nous allons maintenant «base des caractéristiques » classer l'image, mais avant de discuter ce point, nous allons voir premièrement :

- quelles sont les classes utilisées ou plus précisément quelle hiérarchie de classes nous allons utiliser,
- la méthode de classification SVM en détail.

5.1 Classes à reconnaître

Selon la figure IV.9 nous pouvons dire que le nombre de classes effectives à reconnaître est 08 (nombre des groupes sanguins), mais on analysant l'exemple de l'image brute utilisé dans

la figure IV. 1 (ci-dessus) nous allons constater qu'il est possible à minimiser le nombre des classes en faisons une segmentation de l'image.

Chaque exemple doit contenir quatre gouttes de test et par conséquent 04 petites images. Ensuite, chacune des petites images va être avec agglutination ou non, alors 02 classes réelles.

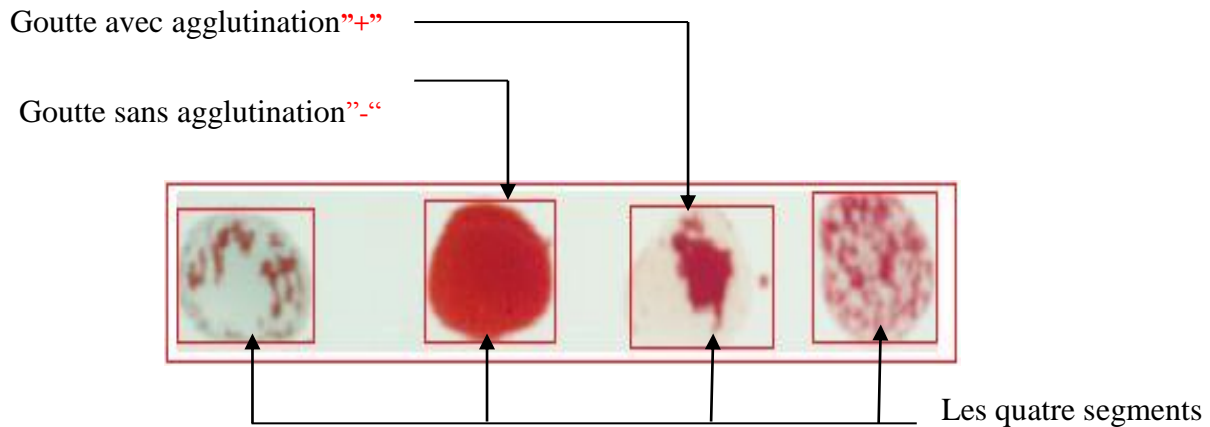


Figure IV. 8: Illustration du processus de segmentation

5.2 Classifieur SVM

Un classifieur SVM est chargé de prendre le vecteur de caractéristiques généré par la phase précédente comme une donnée d'entrée, rechercher un hyperplan séparateur qui sépare les exemples dans la phase d'apprentissage et faire une décision de classification dans la phase d'identification « décision ».

Dans le module SVM, il y a deux phases : une pour l'apprentissage et l'autre pour la classification. La figure suivante représente la relation entre ces deux phases

Généralement, nous allons répondre à la question : *Y'a-t 'il d'agglutination ?*

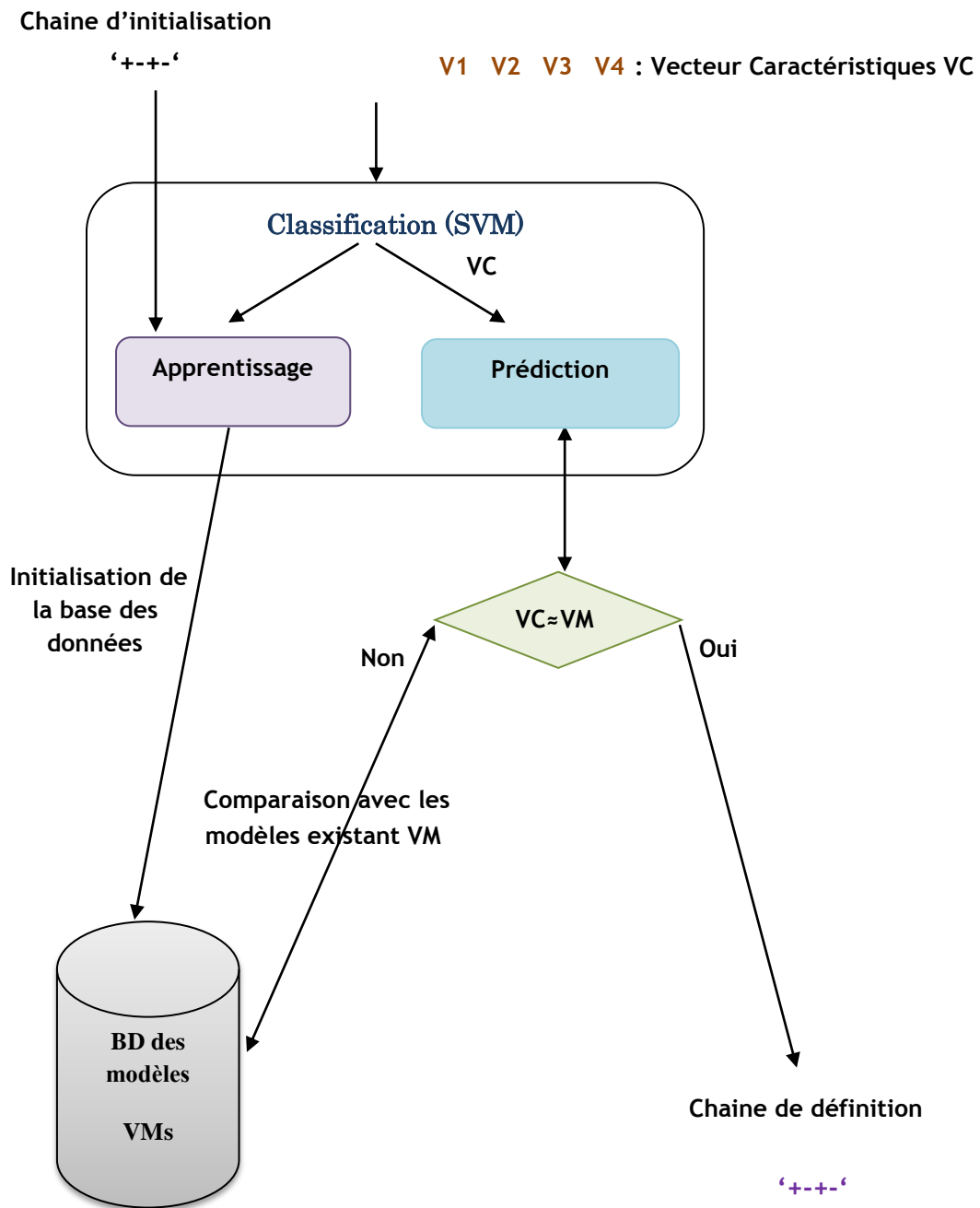


Figure IV. 9: Phase de classification

5.2.1 Algorithme générale de SVM

1. Entrée:

Q,b,y,Cp,Cn, et an des points initial faisable alpha;

l: la taille des vecteurs et matrices;

eps: critère d'arrêt.

2. Sortie:

alpha: vecteur contenant la solution;

obj: la valeur objective;

3. Variables locales:

G,G_bar: vecteurs utiliser pour les calculs du gradient;

alpha_status: vecteur contenant l'état de alpha;

active_size: variable contenant la taille du vecteur active_set;

active_set: vecteur contenant l'ensemble de travail courant;

Début

1. Initialisation des paramètres : taille de l'échantillon (l), paramètre c , largeur de bande du noyau (gaussien), précision numérique ϵ (zéro_tolérance).
2. Initialiser le vecteur `alpha_status` (vecteur contenant l'état d'alpha).
3. Initialiser la variable `active_size` et le vecteur `active_set` (utilisés durant l'opération de shrinking).
4. Initialiser le gradient.
5. Commencer l'optimisation:
 - a. Faire l'opération de shrinking.
 - b. Reconstruire tout le gradient.
 - c. Réinitialiser `active_size` et choisir un ensemble de travail.
 - d. Mis à jour d' α_i , α_j , G , G_bar et `alpha_status`.
6. Calculer la valeur objective.
7. Affichage des résultats.

Fin

L'algorithme présenté ci-dessus ne montre que les grandes étapes de résolution des SVM; pour plus d'informations il faut consulter l'ad-doc jointe avec le package `libsvm-2.83` [44].

5.2.2 Bibliothèque LIBSVM utilisée

La bibliothèque LIBSVM est développée dans le but de simplifier l'utilisation des SVM comme un outil, dans ce travail nous utilisant la version 2.83 qui est présentée par le package `java libsvm-2.83`.

La relation entre une application utilisateur et le package libsvm-2.83 peut être vu selon un modèle de trois couches:

- Application: qui représente l'application utilisateur,
- Interface: qui permet la communication entre l'application et les modules de calcul, et
- Calcule: qui permet de réaliser les calculs nécessaires.

La figure suivante illustre cette relation.

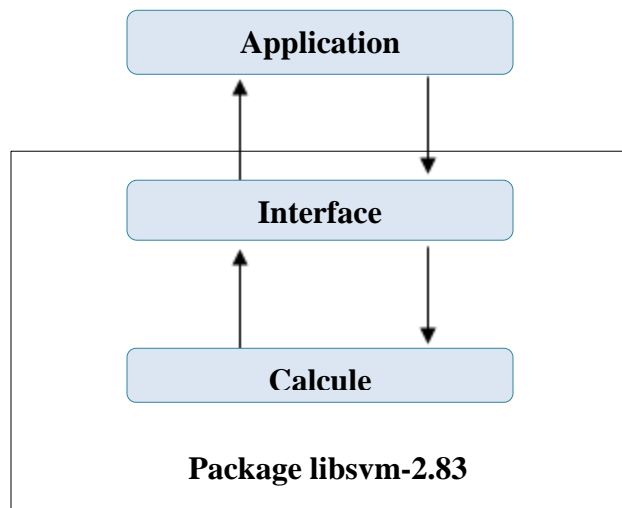


Figure IV. 10: Illustration de la relation entre l'application utilisateur et le package SVM

La couche est présentée par les deux modules:

- `svm_train(CarVec interface.txt)` : pour la phase d'apprentissage, et
- `svm_predict (CarVec.txt, CarVector.model, classLabels.txt)` : pour la phase de test (prédiction).

Où:

`CarVec.txt`: fichier contenant les caractéristiques extraites.

`CarVector.model`: fichier contenant les paramètres (w, b) après la phase d'apprentissage.

`classLabels.txt`: fichier contenant les étiquettes des classes après la phase de test ou décision.

Autant que la couche calcule est présentée par les modules:

- `svm`: comme module principale,
- `svm_node`, `svm_parameter`, et `svm_problem`: comme des modules complémentaire.

Les modules de la couche calcule sont invisible à l'utilisateur du package. Un utilisateur ne peut se communiquer avec le package qu'à travers les deux modules de la couche interface présentée ci-dessus. La relation entre l'application et le package libsvm-2.83 ainsi que les différents modules du package est présentée en détail par la figure suivante:

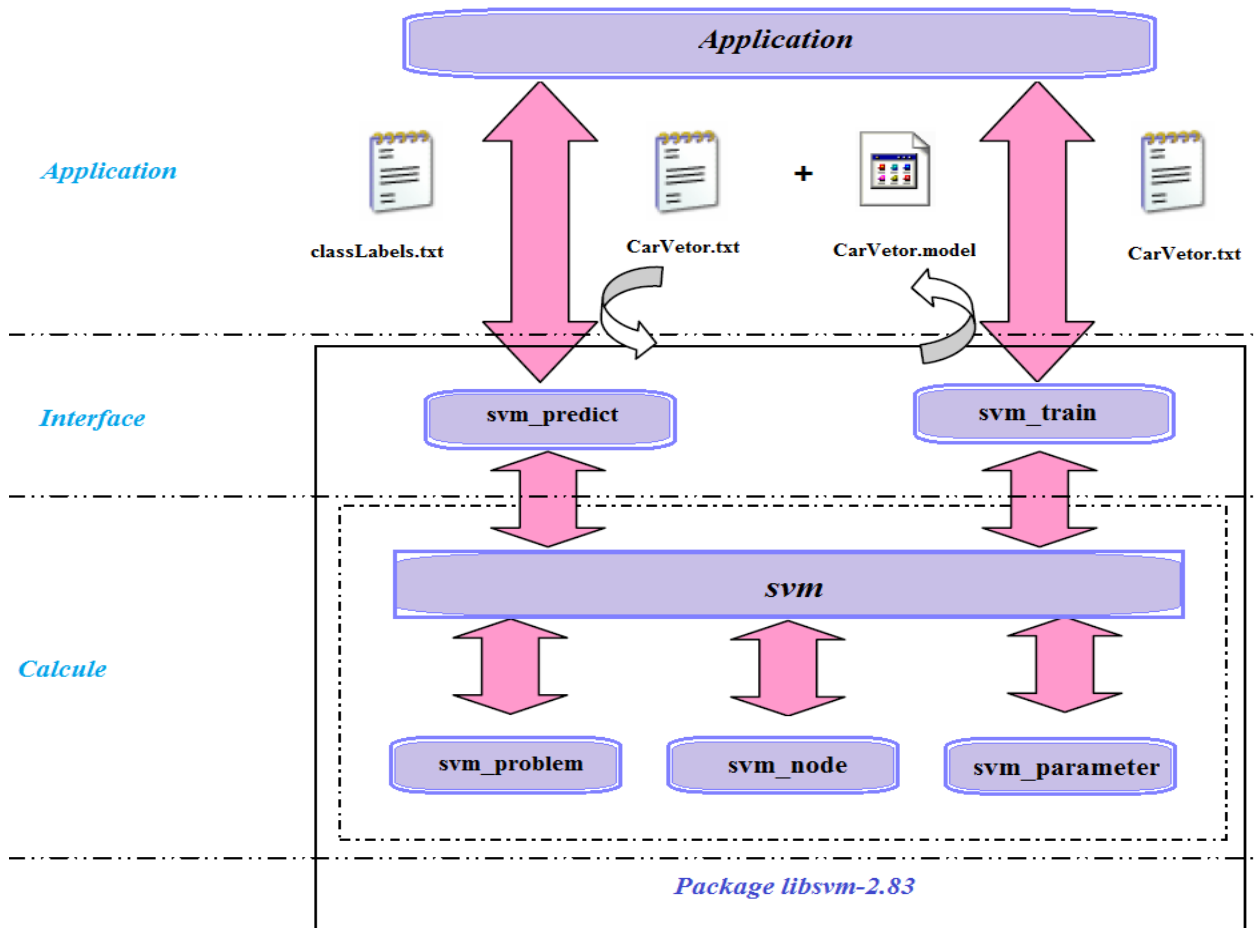


Figure IV. 11: Relation en détail entre l'application utilisateur et le package libsvm-2.83

Premièrement, dans la phase d'apprentissage la couche application génère un fichier texte (`CarVetor.txt`) contenant les vecteurs caractéristiques (X_i) et les étiquettes des classes (Y_i), en utilisant ce fichier comme entrée au module `svm_train` nous obtenant un fichier de modèle (`CarVetor.model`) qui contient les paramètres (w, b) nécessaire à la phase de test ou prédiction.

En revanche, dans la phase de test on utilise le fichier de modèle généré précédemment plus un fichier texte (`CarVetor.txt`) généré par l'application nous obtenant comme résultat un autre fichier texte (`classLabels.txt`) qui contient les étiquettes des classes.

IV.1.6 Post-traitement

Dans le cas du groupe sanguin, l'étape de post-traitement peut faire l'association des différents résultats de la phase de classification « chaîne de définition » à chaque groupe sanguin donné comme le montre le tableau ci-dessous et l'algorithme `SetGroupeSanguin`.

Tableau IV. 1: Chaîne de définition associée à chaque groupe sanguin

	Anti-A	Anti-B	Anti-AB	Anti-D	Chaîne de définition
Groupe A+	+	-	+	+	'+++'
Groupe A-	+	-	+	-	'+-'
Groupe B+	-	+	+	+	'-+++'
Groupe B-	-	+	+	-	'-++'
Groupe AB+	+	+	+	+	'++++'
Groupe AB-	+	+	+	-	'+++'
Groupe O+	-	-	-	+	'---+'
Groupe O-	-	-	-	-	'----'

Suivant le tableau ci-dessus expliquant le principe réactionnel des groupes sanguin nous pouvons déduire le code suivant :

Cas chaîne_def **De**

'+++' : Ecrire ('Le groupe sanguin est A+') ;
 '+-' : Ecrire ('Le groupe sanguin est A-') ;
 '-+++': Ecrire ('Le groupe sanguin est B+') ;
 '-++': Ecrire ('Le groupe sanguin est B-') ;
 '++++': Ecrire ('Le groupe sanguin est AB+') ;
 '+++': Ecrire ('Le groupe sanguin est AB-') ;
 '---+' : Ecrire ('Le groupe sanguin est O+') ;
 '----' : Ecrire ('Le groupe sanguin est O-') ;

Fin**IV.2 Résultats et bilan**

Cette section, présentera le choix du langage de programmation, les différentes interfaces et fenêtres principales du système, les tests et résultats obtenus.

IV.2.1 Choix du langage de programmation

Dans ce travail, nous avons choisi comme environnement de programmation le langage JAVA qui possède une richesse et offre une grande simplicité de manipulation d'images, soit en acquisition ou en génération des fichiers images.

Ce langage possède des avantages très intéressants tel que :

- La portabilité des logiciels,
- La réutilisation de certaines classes déjà développées,

- La possibilité d'ajouter à l'environnement de base des composants fournis par l'environnement lui-même,
- La quasi-totalité de contrôle de Windows (boutons, boîtes de saisies, listes déroulantes, menus ...etc.) qui sont représentés par classes.

IV.2.2 Interface et Fenêtres

On lançant le logiciel nous allons voir premièrement une image d'entrée (splash window) suivie de la fenêtre principale comme nous montre la figure suivant.

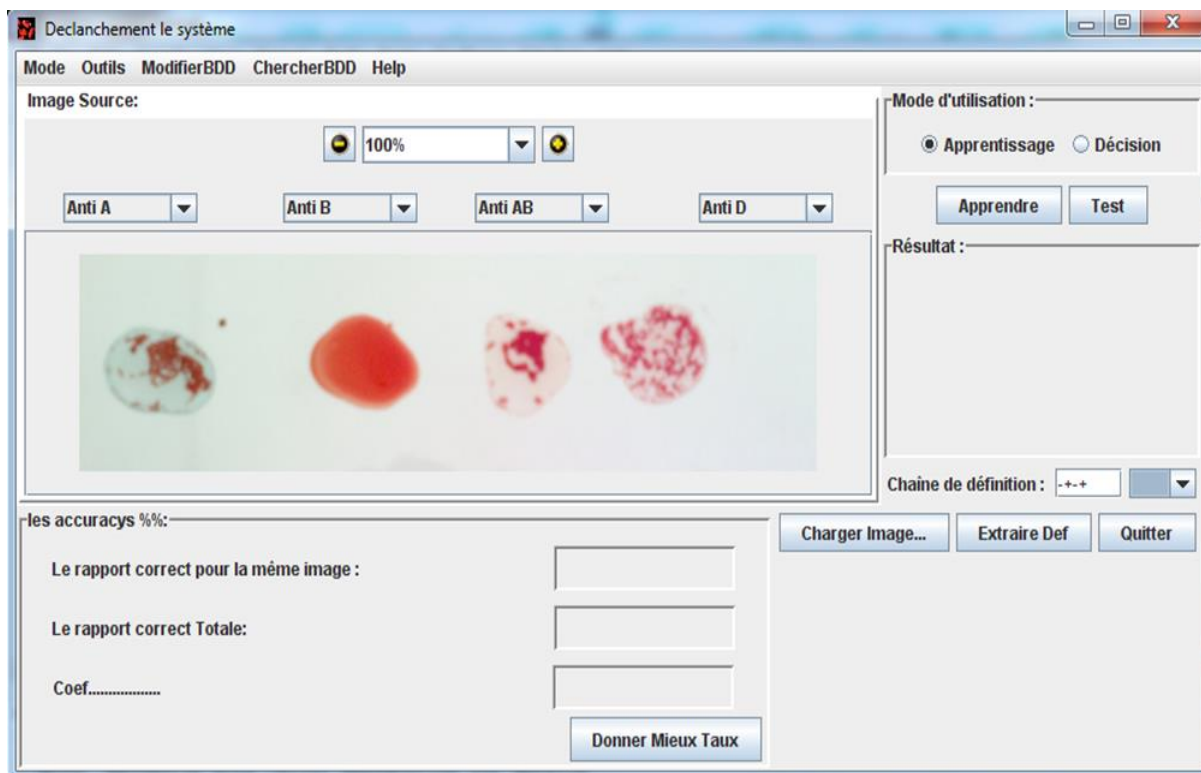


Figure IV. 12: Illustration de la fenêtre principale de l'application

L'application peut être utilisée suivant deux modes:

- Apprentissage: comme phase initiale pour aider le système à apprendre les différentes classes (Formes primitives),
- Test: pour tester et calculer le taux de reconnaissance, et éventuellement utiliser OSF.

2.1 Mode Apprentissage

Ce mode peut être vu comme phase initiale ou d'initialisation de la base de connaissance du système, pour le faire on procède comme suit:

1. Charger une image contenant les exemples à l'aide du bouton *Charger Image*. Déterminer le type de ces exemples à l'aide du combo box (voire figure IV. 14).

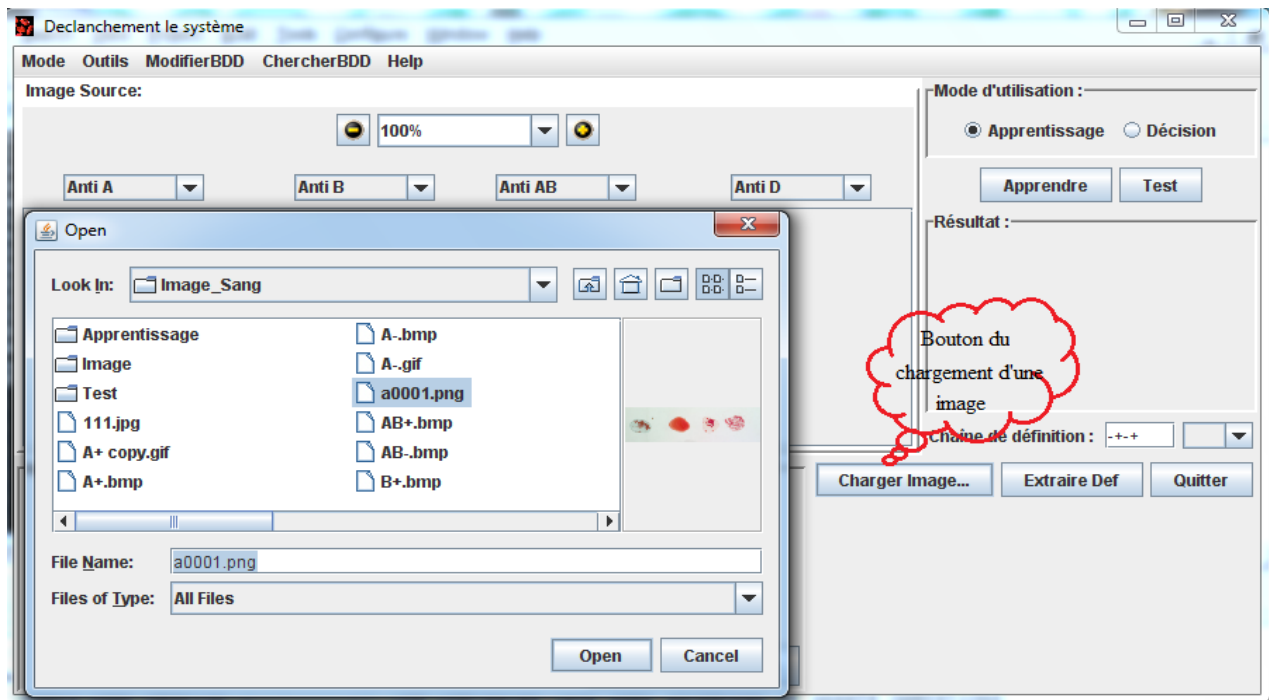


Figure IV. 13: Illustration du chargement d'une image

2. cliquer sur le bouton *Extraire Def* Pour extraire les caractéristiques de ces exemples. On répète cette étape jusqu'en compléter tous les exemples destiner à faire l'apprentissage.

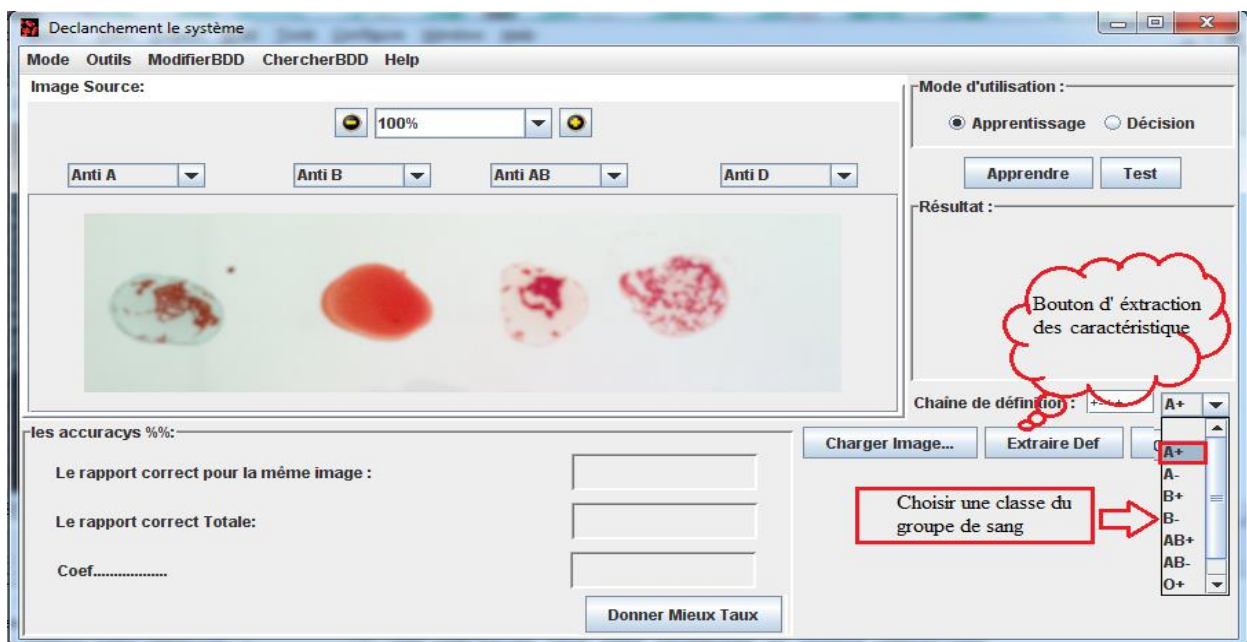


Figure IV. 14: Extraction des caractéristiques des exemples d'apprentissage

Ces étapes sont répétées pour toutes les classes.

3. Cliquer sur le bouton *Apprendre* qui permet lancer l'opération d'apprentissage des exemples données

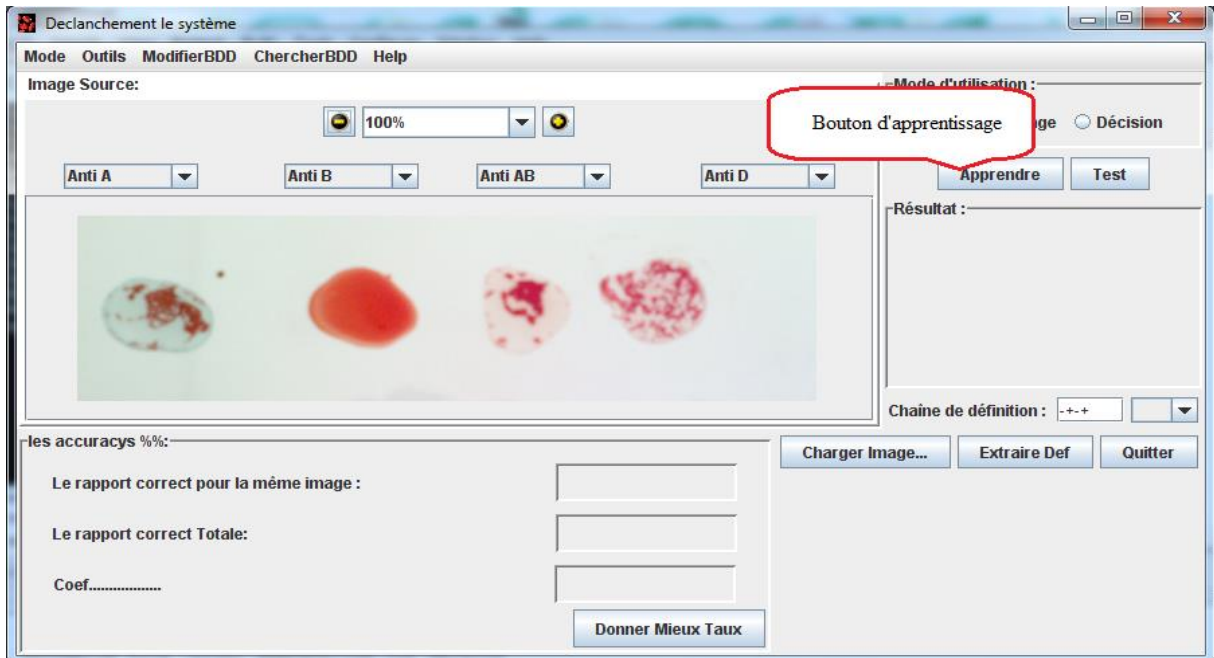


Figure IV. 15: Illustration du mode d'apprentissage

2.2 Mode Test

Ce mode ne peut être exploitable qu'après avoir terminé la phase d'apprentissage et il suit presque les mêmes étapes. L'étape 1 est la même, pour l'étape 2, on clique sur le bouton de test *Test* (voire figure suivante).

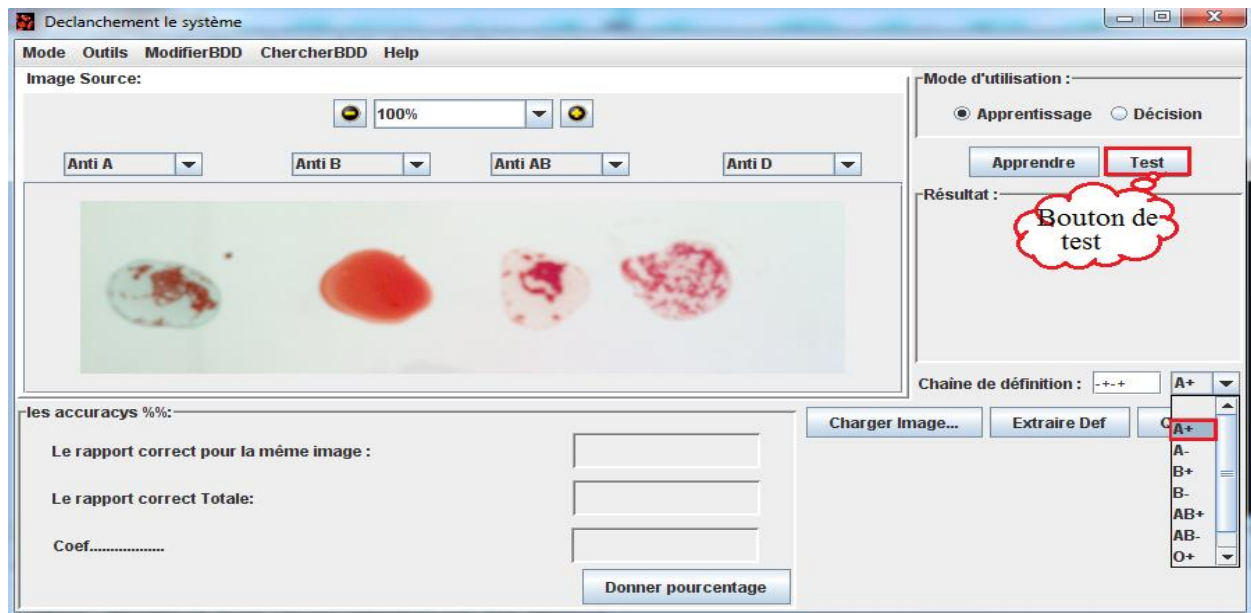


Figure IV. 16: Illustration du mode de test

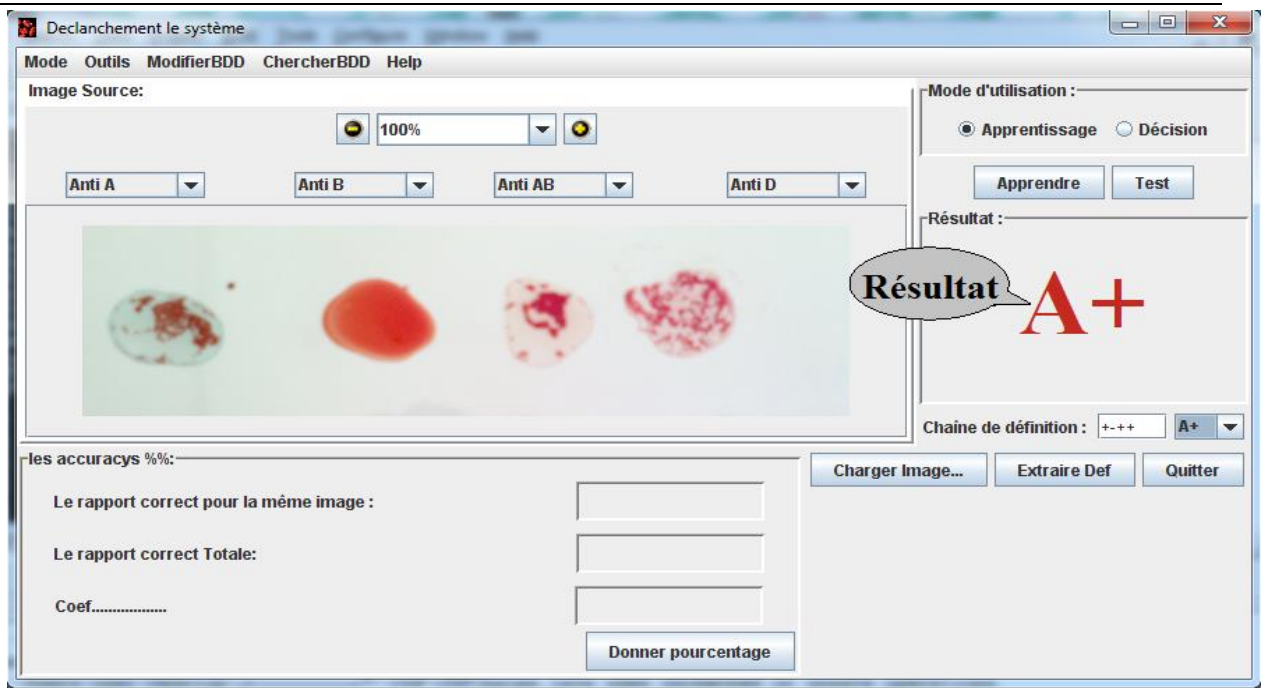


Figure IV. 17: Illustration du résultat du traitement

Dans la réalité, ce mode peut être étendu en mode d'utilisation normal, tous simplement on charge l'image et clique sur la buttons *Test*, ce qui veut dire qu'on ne s'intéresse pas à calculer le taux de reconnaissance mais seulement à reconnaître ce qui à quel groupe représenté.

Le résultat du traitement va être présenté dans Panel *résultat*.

Dans ce qui est passé, nous avons présentés seulement les fenêtres principales, mais d'autres détails comme :

- Les Modification de BDD « Ajouter Personne, Supprimer Personne.....etc. ».
- Chercher dans BDD « Chercher les donneur,..... »
- Filtre, les paramètres du classifieur, Analyse,...etc, il faut consulter l'aide de l'application.

IV.2.3 Test et Résultats

Pour pouvoir jugé la performance du système nous avons utilisé un ensemble d'exemples pour chaque groupe sanguin (A+, A-, B+, B-,...etc.).Ces ensemble sont divisé en deux parties une partie contenant les images d'apprentissage et l'autre les images de test. Ces exemples sont pris à l'aide d'un appareil photo numérique de type samsung es70

Tableau IV. 2: Résultats de test

Groupe Sanguin	Nombre d'exemples		Taux de reconnaissance
	Apprentissage	Test	
A+	10	10	75%
AB+	10	10	100%
AB-	10	10	50%
O+	10	10	75%
O-	10	10	100%

Conclusion

Nous avons présenté dans ce chapitre les différentes étapes qui peuvent conduire à une conception convenable d'un système de reconnaissance de groupe sanguin, nous avons choisi pour l'extraction des caractéristiques une méthode hybrides dont on combine les vecteurs de descripteur statistique et descripteur de structurelle et la méthode SVM pour la reconnaissance des différentes classes. Les résultats de la classification par SVM obtenus peuvent être améliorés en optimisant les paramètres suivants :

- Les paramètres C et gamma ;
- Le nombre d'exemples ;
- ...etc.

Comme solution au problème des classes qui ne sont pas facilement séparables, nous pouvons augmenter le nombre d'exemples.

Conclusion générale

Malgré les efforts et les travaux intensifs réalisés dans le domaine de la reconnaissance d'images, aucun SRI n'est jugé fiable à 100%, Mais on a essayé et à mesure les autres ont essayé d'améliorer les scores pour de meilleurs résultats.

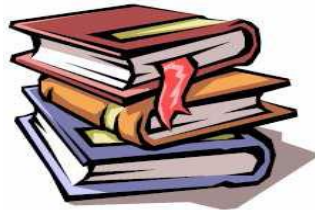
Dans ce travail nous avons présenté les différentes étapes nécessaires au développement d'un système de reconnaissance automatique du groupe sanguin (système de groupage automatique).

De tels systèmes sont normalement destinés à être intégrés ou utilisés avec des machines qui possèdent un environnement d'acquisition conditionné (éclairage fixe, résolution bien définie, ...etc.), c'est la raison pour laquelle nous avons choisi d'utiliser une méthode de seuillage fixe.

De plus, une technique de segmentation à deux niveaux est utilisée pour pouvoir avoir la possibilité d'extraire deux types de caractéristiques :

- Caractéristiques structurelles (Segmentation niveau 1),
- Caractéristiques statistiques (Segmentation niveau 2).

Ensuite un classificateur SVM est utilisé pour classer les différentes images des groupes sanguins. Les résultats obtenus ont été acceptables, mais pour un taux de reconnaissance bien défini pour de tels systèmes des tests réels sont exigés.



Bibliographies

- [1] Al-Rashaideh H: « Preprocessing phase for Arabic Word Handwritten Recognition », Institut d'informatique et automatique, Tom 6, N0 1,2006, cmp.11-19, Russie, February 26, 2006.
- [2] S. Haitamar : « segmentation de texte en caractère pour la reconnaissance optique de l'écriture arabe», Université EL-HADJ LAKHDHAR Batna, Juillet 2007.
- [3] M. Côté : « Utilisation d'un modèle d'accès lexical et de concepts perceptifs pour la reconnaissance d'images de mots cursifs ». Thèse de Docteur de l'école Nationale Supérieure de Télécommunications, France, Juin 1997.
- [4] A. Belaïd: « Reconnaissance automatique de l'écriture et du document », LORIA-CNRS, Campus scientifique B.P. 239, 54506 Vandoeuvre-Lès-nancy, France.
- [5] T. Paquet, L. Heutte, Y. Lecourtier : «Problématique de la Reconnaissance de l'Écriture», ASTI'2001 des Sons, des Images et des Documents à leur Interprétation, France, 2001.
- [6] P. Smrž et al: « Off-line Recognition of Cursive Handwritten Czech text», Université de Masaryk, Février 1998.
- [7] Shubair A et al: « Off-line Arabic handwritten word segmentation using rotational invariant segments features », The international Arab journal of information technology, Vol. 5, No. 2, April 2008.
- [8] S. Chevalier et al : « Étude de primitives spectrales pour la reconnaissance de caractères manuscrits dans le cadre d'une approche markovienne 2D », DGA/Centre d'Expertise Parisien, France, Novembre 2005.
- [9] F. Menasri : «Segmentation d'image Application aux documents anciens », Thèse Docteur de l'Université Paris Descartes en Informatique, France, Juin 2008.
- [10] Djerou.L, Batouche.M: « Résolution collective du problème de segmentation », Université de Biskra, 2003.

- [11] Rousselle.A : « La contour actif une méthode de segmentation application à l'imagerie médicale », Thèse de Doctorat, Université François Robelais de Tours, Juillet 2003.
- [12] BEY. R : « Une segmentation hybride d'images IRM cérébrales par combinaison de l'algorithme K-Means et Quad-Tree Génétique », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, 2012.
- [13] ZAIZ Faouzi : « Les Supports Vecteurs Machines(SVM) pour la reconnaissance des caractères manuscrits arabes », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, 2010.
- [14] A. Madaan, R. Mehta : «Comparative Evaluation of Learning Algorithms on Hand written Character Data », Cours CS 698, Mars 2004.
- [15] Mohamadally .H, Fomani.B : « SVM machine à vecteurs de support ou séparateur à vaste marge », BD Web, ISTY3, Versailles St Quentin, France, janvier 2006.
- [16] GUEDIRI.K : « Une méthode floue adaptative (AF²PSO) pour la segmentation d'images IRM cérébrales », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, 2012.
- [17] DJOUADIM : « La segmentation Markovienne des IRM cérébrales », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, 2012.
- [18] Jean-Christophe Baillie : « Cours de segmentation. ENSTA ParisTech (France) », 2009.
- [19] Sebari.I et Dong-Chenhe : « Les approches de segmentation d'image par coopération régions-contours », Université de Sherbrooke, 10 avril 2007.
- [20] Mahdjane.K: « Détection déanomalies sur des données biologiques par SVM », 2010.
- [21] Michéle Jardino: « Fouille de données dans les corpus de textes (Classification supervisée SVM) ».
- [22] TOURQUI.W : « Recherche d'images par le contenu symbolique (approche statistique) », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, 2012.
- [23] Gueguen. L: « Extraction d'information et compression conjointes des séries temporelles d'images satellitaires », Ecole Nationale Supérieure des Télécommunications Paris. Thèse de Doctorat, Octobre 2007.

- [24] ZHANG. J: « Brief review of invariant texture analysis methods. Pattern Recognition », Vol. 35, iss. 3 March 2001.
- [25] BELHASSEN.M : « Une Approche d'Indexation et de Recherche d'Images Pulmonaires TDM par le Contenu Visuel », Mémoire pour l'obtention du MASTER EN INFORMATIQUE, Université de MONASTIR, 2008.
- [26] VU NGOV Son: « Indexation et recherche d'images par le contenu », Travail d'Intérêt Personnel Encadré (TIPE), Rapport final, Superviseur : Dr. NGUYEN Thi Hoang Lan, 2011.
- [27] J. Y. Ramel : « Lecture automatique des partitions musicales», Mémoire de DEA ingénierie informatique, LISPI - Equipe de Reconnaissance des Formes et Diagnostics, Université Lyon 1, France, 1993.
- [28] M. Pom H. Wu: « Handwritten Character Recognition, Queensland University, October 2003.
- [29] M. Eldawy: « A Survey of Classification techniques», Presentation, May 2006.
- [30] A. Boutarfa : «Reconnaissance de formes 3D par approche neuronale associant la transformée de Hough en robotique mobile: application à la productique». Thèse de doctorat Es-Sciences en électronique industrielle, Université de Batna, 2006.
- [31] C. Bahlmann, B. Haasdonk, H. Burkhardt : « On-line Handwriting Recognition with Support Vector Machines—A Kernel Approach», Publ. in Proc. of the 8th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 49–54, Germany, 2002.
- [32] Bloch.I : « Classification et reconnaissance des formes », 2003.
- [33] M. Volker, E. Haikal, M. Pechwitz: « Offline Handwritten Arabic Word Recognition Using HMM – a Character Based Approach without Explicit Segmentation», Technical University of Braunschweig Institut for Communications Technology (IfN), Germany, Mars 2008.
- [34] F. Menasri : « Segmentation d'image Application aux documents anciens », Thèse Docteur de l'Université Paris Descartes en Informatique, France, Juin 2008.
- [35] Y.Halam, Y.Djebali: «Ggroupes sanguin», Pharmacie Djebali, Décembre 2012.

- [36] K. Landsteiner: « Les groupes sanguins ».L'Atelier scientifique de Pablo Picasso, France, Février 2011.
- [37] D. M. Tazerout, Y. Galinier: « Les groupes sanguins», Direction Régionale des Affaires Sanitaires et Sociales Midi-Pyrénées, France, Mars 2009.
- [38] P. Mahé : «Noyaux pour graphes et Support Vector Machines pour le criblage virtuel de molécules », Rapport de stage, DEA MVA 2002/2003, Septembre 2003.
- [39] P. Mahé, L. Ait-Ali : « Projet d'apprentissage statistique SVM pour l'apprentissage non supervisé», DEA MVA, Février 2003.
- [40] P. Mahé, L. Ait-Ali : « Projet d'apprentissage statistique SVM pour l'apprentissage non supervisé», DEA MVA, Février 2003.
- [41] O. Bousquet : « Introduction aux ' Support Vector Machines '(SVM) », Centre de Mathématiques Appliquées Ecole Polytechnique, Palaiseau, Orsay, Novembre 2001.
- [42] P. Vincent : «Modèles à noyaux à structure locale », Thèse présentée à la Faculté des études supérieures en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.) en informatique, Université de Montréal, Octobre 2003.
- [43] J. Park: «Hierarchical character recognition and it's use in handwritten word/phrase recognition », Thèse de phd, Université de New York, Novembre 1999.
- [44] Chih-Chung Chang et Chih-Jen Lin : « LIBSVM: a Library for Support Vector Machines », Technical_Report, Septembre 2006.