



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABOU-BEKR BELKAID – TLEMCCEN

THÈSE LMD

Présentée à :

FACULTE DES SCIENCES – DEPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

DOCTORAT

Spécialité: *Informatique*

Par :

Mme Ismahan Baghli

Sur le thème

Reconnaissance automatique des plasmocytes pour l'aide au diagnostic du Myélome

Soutenue publiquement le 4 Février 2018 à Tlemcen devant le jury composé de :

Mr BENAMMAR Abdelkrim	Maître de Conférences A	Université de Tlemcen	Président
Mr CHIKH Mohamed Amine	Professeur	Université de Tlemcen	Directeur de thèse
Mr BENZAOUZ Mourtada	Maître de Conférences A	Université de Tlemcen	Co-Directeur de thèse
Mr MESSAIDI Mohamed	Maître de Conférences A	Université de Tlemcen	Examineur
Mr BOUAMRANE Karim	Professeur	Université d'Oran	Examineur

*Laboratoire Génie Biomédicale
BP 119, 13000 Tlemcen - Algérie*

Remerciements

Le chercheur a cette chance de pouvoir s'exprimer dans une page loin des formules mathématiques, des courbes et des observations. Après des années de travail, vient le moment où cette thèse voit le jour et 'Al hamdoulilah', et par conséquent l'expression de la gratitude à tous ceux qui ont contribué de près ou de loin à sa réalisation.

Tout d'abord, mes sincères remerciements vont à mon directeur de thèse, le Professeur Mohamed Amine CHIKH, pour la confiance qu'il m'a attribué tout en m'encourageant à aller de l'avant, en me laissant autonome, avec des échanges toujours aussi riches et diverses.

Je tiens à exprimer ma gratitude à mon co-directeur de thèse Monsieur Mourtada BENZAOUZ, tout d'abord, de m'avoir introduit à ce domaine si riche dans mon mémoire de master et d'avoir continué à me guider et m'orienter tout au long de mes travaux de thèse, je vous exprime mes sincères remerciements.

J'adresse mes sincères remerciements à Monsieur Abdelkrim BENAMMAR, Maître de conférence et chef du département d'informatique de l'université de Tlemcen, qui m'a fait l'honneur de présider le jury de thèse.

Je remercie très sincèrement Monsieur Mohamed MESSAIDI, Maître de conférence au département d'informatique de l'université de Tlemcen, Monsieur Karim BOUAMRANE, Professeur à l'université d'Oran 1, pour l'intérêt qu'ils ont bien voulu porter à ce travail de thèse en acceptant de l'examiner et d'en être rapporteurs.

Je tiens à remercier, Monsieur Amir NAKIB, Professeur associé à l'université Paris-est Créteil, de m'avoir accueilli à plusieurs reprises au sein du laboratoire LISSI (France), et de m'avoir bien orienter dans mes travaux de thèse, tout en m'encourageant, je vous exprime ma gratitude.

J'adresse mes remerciements les plus sincères à tous les enseignants du département d'informatique et spécialement ceux qui ont contribué à mon cursus, aux membres du laboratoire Génie Biomédicale et spécialement les membres de l'équipe CREDOM.

Cela va de soi, **la family**, je vous remercie tous, du plus grand au plus petit. Tout d'abord, à mon mari Selmane et ma fille Afnene, votre soutien et votre présence m'a encouragé à continuer, je vous remercie du fond de mon coeur. À mes parents, Abderrahim et Naoual, depuis ma naissance à ce jour, vous êtes toujours à mes côtés, vos précieuses valeurs et votre amour propre

m'ont toujours guidés et me guideraient pour toujours, moi et mes frères, notre expression de notre gratitude ne sera jamais à la hauteur de nos sentiments. Et à l'occasion un grand merci pour ma petite soeur Ikram (la benjamine) et mon frère Mehdi (le lion de la famille) et sa femme Wafaa , vous êtes toujours présents pour me soutenir. Je remercie très sincèrement mes beaux-parents, Saad et Farida, pour m'avoir toujours encouragé à continuer de progresser. Je souhaite du succès pour mon amie Fadia, qui a toujours su m'encourager. Je remercie mes beaux-frères et mes belles soeurs : Anas, Mouad, Rouzia, Imane, Amina, et leurs petits : Yousra, Taha, Assia, Yacine et Nawfel. Un remerciement spécial à mes grands parents : Aicha, Nourredine et Rachida, et à la mémoire de mon grand père Abdelghani. Un clin d'oeil à mes tantes, mes oncles et toute la grande famille.

Résumé

La maladie du Myélome multiple appelée aussi maladie de Kahler, est parmi les maladies du cancer du sang. Ce type de cancer est caractérisé par la prolifération des globules blancs de type plasmocyte. Le diagnostic précoce de cette maladie, améliore le taux de survie du patient atteint. Le diagnostic manuel repose sur l'examen visuel, par le pathologiste, des images microscopiques de la moelle osseuse pour chercher des signes de prolifération de cellules. Cette analyse nécessite des efforts considérables pour une détection fiable et précise, ce qui la rend à la fois lente et fastidieuse.

Les systèmes automatiques basés sur l'identification et le comptage des globules blancs, peuvent donner des résultats plus précis que les méthodes manuelles. Ces systèmes sont basés principalement sur deux étapes essentielles : la segmentation des cellules, et la caractérisation et la classification des globules blancs. Dans le modèle automatique que nous proposons, les images microscopiques de la moelle osseuse sont tout d'abord segmentées par la transformation watershed et ensuite les régions sont fusionnées par l'intégration de l'incertitude sur la couleur via la théorie de l'évidence. Ces globules blancs segmentés sont caractérisés par les attributs de formes du noyau et de la cellule, et les attributs de textures couleurs du noyau et du cytoplasme, et au final ces attributs sont introduits à trois classifieurs à apprentissage supervisé SVM (Support Vector Machine), Kppv (K plus proche voisins) et arbre de décision. Les résultats expérimentaux montrent que la reconnaissance du plasmocyte par le Kppv est prometteuse, atteignant un taux de reconnaissance de 97% avec une spécificité de 100%, ce qui la rend insensible aux faux positifs.

Mots clés : {Myélome, Plasmocyte, Images de la moelle osseuse, Images microscopiques, Segmentation, Théorie de l'évidence, Transformation watershed, Caractérisation, Forme, Couleur, Texture, Classification.}

Table des matières

Introduction générale	1
1 Aide au diagnostic du Myélome multiple	6
1.1 Introduction	6
1.2 Diagnostic de la maladie de Kahler	6
1.2.1 Critères de diagnostic	6
1.2.2 Ponction médullaire	7
1.2.3 frottis médullaire	8
1.2.4 Lecture du frottis au microscope	9
1.2.4.1 Zone de lecture	9
1.2.4.2 Images microscopiques	9
1.3 Le système d'aide au diagnostic	9
1.3.1 L'acquisition des images microscopiques	11
1.3.2 La segmentation des images microscopiques	11
1.3.3 La classification automatique des globules blancs	12
1.4 Conclusion	12
2 Segmentation des cellules microscopiques	14
2.1 Enjeux de la segmentation	14
2.2 Méthodes de segmentation des images microscopiques des cellules	15
2.2.1 Méthodes basées régions	16
2.2.1.1 Seuillage	16
2.2.1.2 Classification pixellaire	16
2.2.1.3 Croissance des régions	17
2.2.1.4 Clusters des régions	17
2.2.1.5 Segmentation morphologiques - Transformation Watershed	17
2.2.2 Méthodes basées contours	17
2.2.2.1 Détection des bordures	17
2.2.2.2 Modèles déformables ou les contours actifs	18
2.3 Quelques méthodes appliquées à la segmentation des images cy- tologiques	18
2.4 Méthode proposée	20
2.4.1 Notions préliminaires	20
2.4.1.1 Projection sur différents espaces couleurs	20

2.4.2	Principe de la théorie de l'évidence	23
2.4.3	Segmentation évidentielle des cellules	25
2.4.3.1	Segmentation initiale par transformation watershed	25
2.4.3.2	Agrégation évidentielle des régions	26
2.5	Conclusion	28
3	Caractérisation et classification des cellules	29
3.1	Introduction	29
3.2	Étude des cellules sanguines	29
3.2.1	Lymphopoïèse	30
3.2.1.1	Lymphoblaste	30
3.2.1.2	Prolymphocyte	30
3.2.1.3	Petit lymphocyte	31
3.2.1.4	Grand lymphocyte	31
3.2.1.5	Plasmocyte	32
3.2.2	Mégacaryopoïèse	32
3.2.2.1	Mégacaryoblaste	32
3.2.2.2	Mégacaryocyte	33
3.2.2.3	Thrombocyte	33
3.2.3	Monopoïèse	34
3.2.3.1	Promonocyte	34
3.2.3.2	Monocyte	35
3.2.4	Myélopoïèse	35
3.2.4.1	Myeloblaste	35
3.2.4.2	Promyélocyte	35
3.2.4.3	Myélocyte	36
3.2.4.4	Metamyélocyte	37
3.2.4.5	Granulocyte à bande neutrophile	38
3.2.4.6	Granulocyte à neutrophile segmenté	38
3.2.4.7	Granulocyte éosinophile	38
3.2.4.8	Granulocyte basophile	39
3.2.5	Érythropoïèse	39
3.2.5.1	Proerythroblast	39
3.2.5.2	Érythroblaste basophile	40
3.2.5.3	Erythroblaste polychromatique	41
3.2.5.4	Erythroblaste orthochromatique	41
3.2.5.5	Erythrocyte polychromatique	42
3.2.5.6	Erythrocyte	42
3.3	Caractérisation des cellules	43
3.3.1	Attributs géométriques	43
3.3.2	Attributs de textures	45
3.4	Classification des cellules	46
3.4.1	Apprentissage supervisé	46

3.4.2	Apprentissage par renforcement	47
3.4.3	Apprentissage non supervisé	47
3.4.4	Algorithmes utilisés	47
3.5	Méthodes de reconnaissance automatique des cellules dans la littérature	48
3.6	Modèle proposé pour la détection des plasmocytes	49
3.6.1	Caractérisation des globules blancs	50
3.6.2	Reconnaissance des plasmocytes	51
3.7	Conclusion	51
4	Approche proposée pour l'aide au diagnostic du Myélome	52
4.1	Introduction	52
4.2	Modèle automatique proposé	52
4.2.1	Segmentation évidentielle des cellules	53
4.2.2	Caractérisation des globules blancs	56
4.2.3	Reconnaissance automatique des plasmocytes	58
4.3	Évaluation du modèle de reconnaissance	59
4.3.1	Résultats de la segmentation	59
4.3.2	Résultats de la caractérisation et la classification	67
4.4	Conclusion	71
	Conclusion générale et perspectives	75

Table des figures

1	Système automatique pour la reconnaissance des plasmocytes . . .	3
1.1	Préparation des lames pour l'inspection cytologique [12]	8
1.2	Image microscopique de cytologie	10
1.3	Automatisation de l'analyse des images microscopiques	10
1.4	Système d'aide au diagnostic	11
1.5	Module d'acquisition des images microscopiques	11
1.6	Segmentation des images microscopiques ((a) Image originale (b) Image segmentée ; vert :noyau, jaune :cytoplasme, rouge :glo- bule rouge et noir :fond)	12
1.7	Reconnaissance des plasmocytes	12
2.1	Évolution de l'application de techniques de segmentation sur les images de cellules [43]	16
2.2	(a) Image originale (b) Projection sur le canal Magenta	22
2.3	(a) Image originale (b) Projection sur le canal Magenta (c) Pro- jection sur l'espace couleur HSV	23
3.1	Le lymphoblaste [3]	30
3.2	Le prolymphocyte [3]	31
3.3	Le petit lymphocyte [3]	31
3.4	Le grand lymphocyte [3]	32
3.5	Le plasmocyte [3]	32
3.6	Le mégacaryoblaste [3]	33
3.7	Le mégacaryocyte [3]	33
3.8	Le thrombocyte [3]	34
3.9	Le promonocyte [3]	34
3.10	Le monocyte [3]	35
3.11	Le myeloblaste [3]	36
3.12	Le promyélocyte [3]	36
3.13	Le myélocyte [3]	37
3.14	Le metamyélocyte [3]	37
3.15	Le granulocyte à bande neutrophile [3]	38
3.16	Le granulocyte à neutrophile segmenté [3]	39
3.17	Le granulocyte éosinophile [3]	39
3.18	Le granulocyte basophile [3]	40

3.19	Le proerythroblast [3]	40
3.20	L'érythroblaste basophile [3]	41
3.21	L'érythroblaste polychromatique [3]	41
3.22	L'érythroblaste orthochromatique [3]	42
3.23	L'érythrocyte polychromatique [3]	42
3.24	L'érythrocyte [3]	43
3.25	Apprentissage supervisé [53]	46
3.26	Apprentissage par renforcement [53]	47
3.27	Caractérisation du globule blanc	49
3.28	Reconnaissance du plasmocyte	49
4.1	Système proposé pour l'aide au diagnostic du Myélome	53
4.2	Transformation watershed : (a) Image originale (b) Image sur-segmentée par la transformation watershed	54
4.3	Blanc Vs rouge : (a) Image sur-segmentée par la transformation watershed (b) Segmentation globule blanc (blanc) Vs globule rouge (rouge)	55
4.4	Segmentation finale : (a) Segmentation globule blanc Vs globule rouge (b) Image segmentée : noyau(vert), cytoplasme(jaune), globule rouge (rouge) et fond (noir)	56
4.5	Processus de reconnaissance des plasmocytes	58
4.7	Illustration de la sensibilité par rapport aux paramètres γ_i des RBC et γ_i des WBC : (a) image originale, (b) image vérité terrain, (c) à (k) résultats de segmentation pour différentes valeurs de γ_i de RBC, et γ_i de WBC.	62
4.8	Taux et précisions : (a) Données RBC (b) Données WBC	63
4.9	L'impact des distances sur les degrés de croyance	64
4.10	Comparaison : (a) Image originale (b) Controlled watershed (vert : marqueurs) (c) Stochastic watershed (d) Segmentation ESA (e) Vérité terrain	65
4.6	Résultats ESA : (a) image originale (b) segmentation ESA (c) Vérité terrain	73
4.11	Courbes ROC : (a) pour k=5 (b) pour k=7 (c) pour k=10 (d) pour k=15	74

Liste des tableaux

4.1	Métriques de performances	61
4.2	Comparaison des segmentations SDR et ESA	67
4.3	Performances de SVM, Kppv et arbre de décision pour $k=5$. . .	69
4.4	Performances de SVM, Kppv et arbre de décision pour $k=7$. . .	69
4.5	Performances de SVM, Kppv et arbre de décision pour $k=10$. .	69
4.6	Performances de SVM, Kppv et arbre de décision pour $k=15$. .	70
4.7	AUCs pour SVM, Kppv et arbre de décision	70

Introduction générale

Le développement de la technologie informatique s'accompagne de nouveaux dispositifs, et de nouvelles solutions. En effet, le monde de l'informatique et les matériels technologiques évoluent rapidement, au point que nous avons du mal à connaître tous les produits existants. Nous sommes aujourd'hui loin, des ordinateurs qui occupaient des salles et mettant des jours et des jours pour les calculs complexes. De nos jours, les composants électroniques sont basés sur une technologie de miniaturisation accompagnée d'une capacité de traitement toujours plus puissante, que ça a évolué vers du matériel de pointe dans tous les domaines : militaire, santé, éducation, commerce,...etc

Le domaine médical a pu bénéficier de nouvelles technologies, liées aux matériels et logiciels de pointe pour accompagner et alléger le travail du médecin au quotidien. Nous citons les imageries médicales (radiographies, microscopies, IRM,...), les logiciels de gestion des dossiers de patients, et bien d'autres. L'idée n'est pas de remplacer le médecin, mais de lui apporter l'aide et la précision dans le diagnostic. Néanmoins, ces technologies nécessitent des systèmes d'aide à l'interprétation précis et fiables, parce que la vie humaine en dépend. De plus ces systèmes doivent être ergonomiques, car ils sont destinés à un utilisateur non technique, qui devrait pouvoir faire fonctionner le système intuitivement.

Problématique

La prolifération des plasmocytes, détectée par l'étude anatomo-pathologique, oriente le médecin vers le diagnostic du Myélome Multiple. Les plasmocytes sont des cellules de type globule blanc. La prolifération des plasmocytes est détectée à travers l'analyse visuelle des lames médullaires au microscope. L'analyse visuelle se fait en général manuellement, engendrant un processus lent avec un risque d'erreur élevé. En effet, cette analyse nécessite des efforts considérables pour une détection fiable et précise, ce qui rend cette tâche lente et

fastidieuse.

L'analyse des différentes cellules sanguines commence d'abord par l'étude quantitative à l'aide de l'hémogramme, complétée par l'étude qualitative via le frottis de sang. La découverte de certaines anomalies qualitatives ou quantitatives de ces cellules implique la réalisation d'un frottis médullaire (myélogramme). Les systèmes s'appuyant sur le traitement automatique des images peuvent aider le médecin pathologiste. Cependant, les pathologistes observent un manque de consensus et de standardisation des protocoles d'imagerie, à la fois, pour le diagnostic des patients nouvellement atteints du myélome, ainsi que pour le suivi des patients en cours de traitement [42].

Les méthodes de traitement d'image et de reconnaissance de forme offrent la possibilité d'assister le pathologiste dans l'étude des différentes formes/cellules dans les images microscopiques. Il existe différents types de cellules dans le sang périphérique, la moelle osseuse et les sections tissulaires. Cela incluent les globules rouges, les plaquettes et les globules blancs. Ces derniers, connus comme les cellules inflammatoires, font partie du système immunitaire, protégeant le corps contre les maladies infectieuses.

Pour détecter la présence anormale des plasmocytes, il est nécessaire de différencier les plasmocytes des autres cellules. Et donc, l'identification des plasmocytes reste la première étape du diagnostic du myélome. Dans la majorité des cas, l'identification est faite manuellement, cependant elle reste imprécise, subjective et fastidieuse. Les techniques de traitement d'images offrent l'avantage de la rapidité et l'accessibilité pour la caractérisation et le comptage des globules blancs. L'automatisation va se reposer sur la segmentation des cellules et la classification des globules blancs.

Contexte et motivation

Cette thèse rentre dans le cadre d'un projet nationale de recherche : projet N18-santé intitulé SARCY (Développement d'un Système Adaptable pour la Reconnaissance des cellules Cytologiques). Ce travail constitue la continuité du mémoire de master sous l'encadrement du Docteur Mourtada BENAZZOUZ. Le sujet du master faisait partie de ses travaux de thèse [12].

Nous avons acquis les images microscopiques du service d'Hémobiologie au Centre Hospitalo Universitaire de Tlemcen. Les images concernaient quatre différents patients, tous atteints de la maladie du Myélome. Nous nous sommes

intéressés à l'élaboration d'un modèle automatique, où le but est d'aider le médecin à diagnostiquer ce type précis de cancer, qui est le Myélome.

Dans les systèmes d'analyse automatique des images médicales, la réussite de la classification et le comptage automatique des cellules reposent sur l'étape de segmentation. En effet, beaucoup de travaux ont été présentés dans ce domaine [43] et [22]. La segmentation sert à identifier automatiquement les différentes cellules présentes dans l'image. La segmentation en elle-même n'est pas le but du système d'analyse automatique, mais la réussite de la chaîne du diagnostic repose sur l'efficacité de cette étape, puisque la bonne identification des cellules permettra la bonne classification de celles-ci.

Le modèle automatique que nous proposons (voir figure 1) doit au préalable segmenter les images afin de différencier les globules blancs (parties noyau et cytoplasme séparés), les globules rouges et le fond. Après la segmentation, les globules blancs sont caractérisés dans le but de les introduire à un classifieur, pour pouvoir reconnaître les plasmocytes des autres cellules.

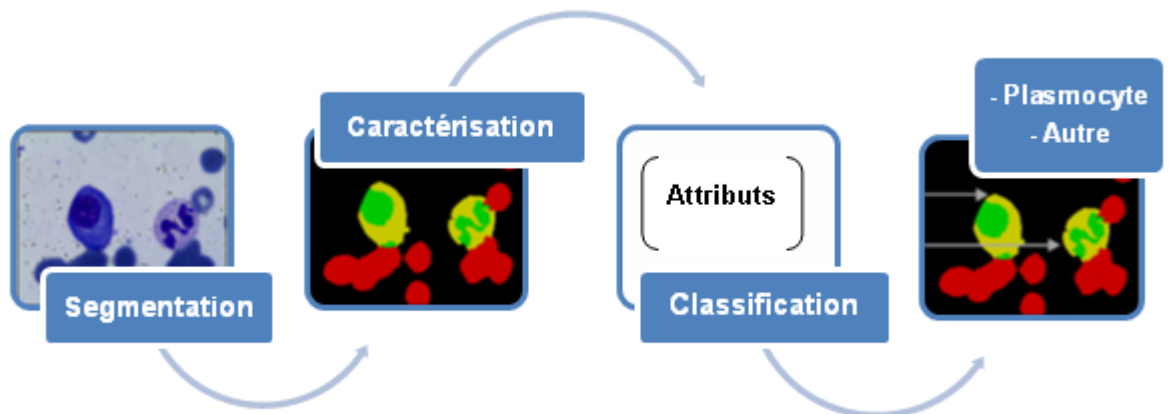


FIGURE 1 – Système automatique pour la reconnaissance des plasmocytes

Contributions

Nous proposons dans cette thèse un modèle automatique pour la reconnaissance des plasmocytes. Ce modèle constitue l'objectif majeure du projet national de recherche SARCY (Développement d'un Système Adaptable pour la Reconnaissance des cellules Cytologiques). Lors de notre travail de thèse, nous avons proposé deux contributions principales, qui concernent respective-

ment la segmentation des cellules, et la caractérisation et la classification des globules blancs.

Dans la première contribution, nous avons proposé une approche pour segmenter les images microscopiques [11]. En effet, nous nous sommes intéressés à la résolution du problème de la sur-segmentation générée par la transformation watershed. Notre solution s’est portée sur l’intégration de l’incertitude via la théorie de l’évidence liée aux régions watershed sur-segmentées. Pour cette solution, nous nous sommes intéressés aux propriétés colorimétriques des régions sur-segmentées. Effectivement, outre l’espace couleur RGB, nous sommes passés à l’espace CMYK et l’espace HSV, et en exploitant la puissance de la théorie de l’évidence à intégrer l’incertitude, nous sommes arrivés à résoudre le problème de la sur-segmentation, et par conséquent, nous atteignons un niveau de segmentation satisfaisant.

La deuxième contribution, est celle de la caractérisation choisit pour reconnaître les plasmocytes [10]. Cette caractérisation nous l’avons voulu être proche de ce que le médecin décrit, puisque le médecin reste à nos jours réticent par rapport aux systèmes automatiques, et d’une grande partie à cause de l’interprétabilité. En plus de la description médicale, nous avons rajouté les attributs majoritairement cités sur les travaux dans ce domaine [57], et de là, nous avons caractérisé le globule blanc selon l’aspect de forme du noyau et de la cellule, et selon l’aspect de texture du noyau et du cytoplasme, nous avons voulu profiter des propriétés couleurs, et nous avons donc calculé la texture sur le canal Magenta et sur le canal Vert. Les attributs obtenus, sont introduits aux classifieurs automatiques (SVM, Kppv et arbre de décision), après une phase d’apprentissage et une phase de test, nous avons aboutit à des résultats prometteurs et encourageants.

Organisation

Cette thèse est organisé en quatre chapitres, en plus de l’introduction générale et la conclusion générale. Dans le **chapitre 1** nous définissons la maladie du myélome multiple et son diagnostic ; nous mettons le point sur l’importance de l’analyse visuelle dans le diagnostic et le risque d’erreur lié au facteur de fatigue et d’expertise, nous introduisons alors le système automatique d’aide au diagnostic ; nous définissons brièvement les différentes étapes de l’automatisation de l’analyse visuelle. Dans le **chapitre 2**, nous passons à l’étape clé

du système automatique qui est la segmentation des images cytologiques, nous définissons ce qui est la segmentation des cellules, les principales méthodes utilisées dans la segmentation des images microscopiques, en insistant sur la non existence d'une méthode spécifique qui fonctionne sur toutes les images, nous passons ensuite à l'explication technique de notre méthode de segmentation, nous définissons les notions préliminaires dont nous avons fait appel, et les différentes étapes de notre technique. Nous continuons d'expliquer l'étape visible du système qui est la reconnaissance automatique des plasmocytes, dans le **chapitre 3**. Dans ce dernier, nous définissons les étapes de la reconnaissance automatique du plasmocyte, qui commence tout d'abord par la caractérisation du globule blanc à travers plusieurs attributs : géométriques, textures et statistiques, ensuite ces attributs sont introduits à un classifieur afin de lui faire apprendre à différencier entre les cellules. Nous présentons ensuite notre caractérisation basé sur un rapprochement entre la description médicale et les aspects techniques de caractérisation, et après nous expliquons comment ces attributs obtenus vont être introduits aux différents classifieurs dont nous avons fait appel. Les résultats obtenus des deux phases, les métriques, les comparaisons et les critiques sont exposés dans le **chapitre 4**. Nous terminons notre thèse par la **conclusion générale** et nous donnerons des perspectives aux contributions réalisées dans cette thèse.

Chapitre 1

Aide au diagnostic du Myélome multiple

1.1 Introduction

Le Myélome Multiple (MM) des os ou maladie de Kahler est une hémopathie maligne caractérisée par le développement d'un clone de plasmocytes tumoraux envahissant la moelle hématopoïtique. Le MM représente 1% de l'ensemble des cancers et 10 % des hémopathies malignes. L'incidence s'accroît avec l'âge et l'âge moyen du diagnostic est d'environ 65 ans, l'incidence du MM a augmenté de 45% en 30 ans). Le myélome n'existe pas chez l'enfant. Les causes du MM sont inconnues [1]. Cependant, les facteurs favorisant cette maladie sont d'ordre :

- Génétiques : l'incidence est double dans la race noire américaine et elle est au contraire très faible chez les asiatiques ;
- Environnementaux : benzène, pesticides, exposition aux radiations, stimulation antigénique chronique.

1.2 Diagnostic de la maladie de Kahler

1.2.1 Critères de diagnostic

Les principales manifestations du myélome résultent de l'accumulation de plasmocytes malins au niveau de la moelle osseuse, pouvant entraîner [37] :

- La production et la sécrétion d'une protéine monoclonale dans le sang

et/ou les urines ;

- Des lésions de l'os environnant la cavité médullaire ;
- Une insuffisance médullaire avec anémie et/ou leucopénie et thrombopénie ;
- Une immunodépression avec inhibition de la production des immunoglobulines normales et sensibilité accrue aux infections

Les critères de diagnostic à rechercher ont été établis par le Southwest Oncology Group [63], regroupé dans [14] en critères majeures et mineurs :

1. Critères majeurs

- Plasmocytome sur biopsie tissulaire
- Infiltration plasmocytaire monoclonale $> 30 \%$
- Présence d'une Ig monoclonale sérique (IgG > 35 g/L ; IgA > 20 g/L et/ou urinaire (chaîne légères libres et urinaires responsables d'une protéinurie de Bence Jones > 1 g/24 h en l'absence d'amylose)

2. Critères mineurs

- Plasmocytose médullaire comprise entre 10 et 30 %
- Pic monoclonal (IgG < 35 g/L ; IgA < 20 g/L) et/ou protéinurie de Bence Jones < 1 g/24 h)
- Lésions osseuses lytiques
- Diminution des Ig polyclonales inférieure à 50 % des intervalles de référence

Le diagnostic est confirmé si sont associées au moins un critère majeur et un critère mineur ou trois critères mineurs dont au moins a + b [14].

Pour diagnostiquer le myélome, il est donc nécessaire de :

- Mettre en évidence un composant monoclonal sérique et/ou urinaire ;
- Mettre en évidence une prolifération plasmocytaire médullaire ;
- Analyser le retentissement de ces deux paramètres.

Dans notre travail, nous nous sommes intéressés à la la prolifération plasmocytaire médullaire, mise en évidence par le myélogramme. Cette dernière est effectuée par ponction sternale.

1.2.2 Ponction médullaire

Il s'agit d'un prélèvement des cellules de la moelle "hématopoïétique", considérée comme l'usine de production de toutes les cellules du sang. Le principe est de piquer l'os pour aspirer le "suc médullaire", niché à l'intérieur des

loquettes osseuses et de le déposer sur une lame (frottis médullaire) pour analyser les cellules. Au laboratoire d'hématologie, l'hématologue-biologiste analysera le prélèvement et déterminera les proportions des différentes variétés de cellules de la moelle osseuse : celles qui deviendront des globules blancs, des globules rouges ou des plaquettes et leur degré de maturation [13].

1.2.3 frottis médullaire

Le frottis médullaire consiste à étaler de la moelle osseuse sur une lame, dans le but d'observer les cellules médullaires, et aussi de les dénombrer.

Le principe est d'étaler une goutte de moelle uniformément sur une lame de verre, de manière à obtenir une seule couche de cellules. La technique consiste à marquer la lame de verre, déposer une petite goutte de moelle à 1 cm du bord de la lame, placer le bord de l'étaleur en contact avec la lame, puis entrer en contact avec la goutte de moelle, pousser l'étaleur rapidement, régulièrement, et sans trop appuyer, tout en gardant la même inclinaison, soulever progressivement l'étaleur, en fin d'étalement, sécher immédiatement le frottis à l'air libre [12], voir figure 1.1.

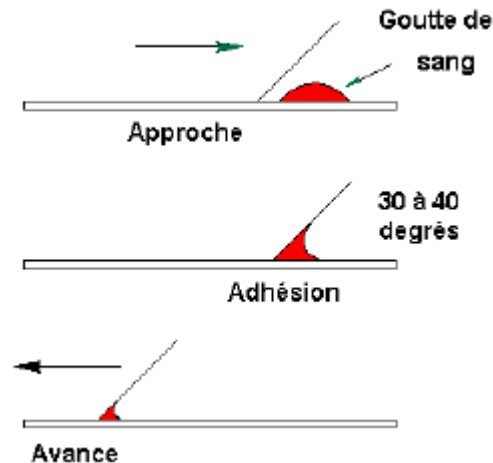


FIGURE 1.1 – Préparation des lames pour l'inspection cytologique [12]

Après coloration avec le protocole MGG (May-Grünwald Giemsa) et fixation, l'hématologue-biologiste pourra, via le microscope optique, effectuer l'étude morphologique des cellules, et déterminer s'il y a anomalies de présence, d'aspect ou de nombre de cellules.

1.2.4 Lecture du frottis au microscope

Pour pouvoir observer et reconnaître les différents types cellulaires présentes, le frottis médullaire doit être coloré avec le protocole MGG et introduit au microscope optique pour pouvoir le lire. Dans ce présent travail, nous avons collaboré avec le service d'hémobiologie du Centre Hospitalo-Universitaire de Tlemcen. Dans ce service, la bonne zone de lecture est élaboré comme décrit dans la section 1.2.4.1, ce qui nous a permis d'obtenir notre base d'images microscopiques de cytologie.

1.2.4.1 Zone de lecture

Une bonne lecture d'un frottis nécessite deux qualités : une zone de lecture suffisante (cellule bien étalée) et une bonne technique de coloration. Et dans les deux cas, les hématies jouent le rôle de témoin. La caméra est fixée sur un microscope Leica, dans un premier temps, un objectif de grossissement à x10 est utilisé afin de repérer une zone de travail en évitant les zones cellulaires entassées pour une reconnaissance aisée des différents types cellulaires. Une fois que la zone optimale de lecture a été obtenue, le zoom du microscope est changé en augmentant le grossissement à x100 ; afin de localiser les images intéressantes pour les acquérir et les stocker et les analyser ultérieurement [12].

1.2.4.2 Images microscopiques

Les images microscopiques ont été obtenues via le système LAS EZ (Leica Application Suite 1.7.0). Ce dernier est un environnement qui permet d'acquérir des images de haute qualité à l'aide du microscope et caméra Leica, reliés à un ordinateur fonctionnant sous Windows. Les images ainsi obtenues (voir figure 1.2), sont des images couleurs RGB au format Bitmap de dimension 1024x768 pixels.

1.3 Le système d'aide au diagnostic

Pour diagnostiquer le myélome, l'analyse des images microscopiques de la moelle osseuse tient une place importante pour la détection des plasmocytes au nombre et/ou forme anormales, afin d'établir un diagnostic fiable et précis. Cette analyse visuelle des images microscopiques constitue généralement une

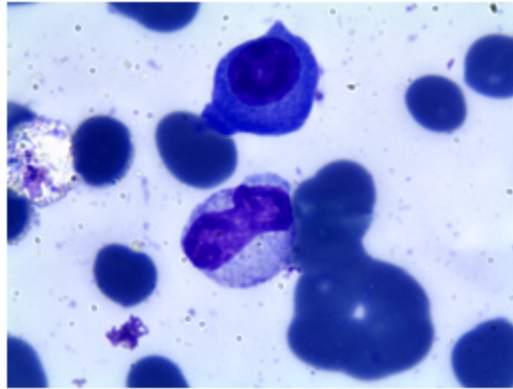


FIGURE 1.2 – Image microscopique de cytologie

difficulté majeur, et nécessite des efforts considérables de la part de l'expert humain, afin d'éviter les mauvaises interprétations.

Pour pallier à cette difficulté, nous proposons dans ce qui suit un système automatique pour l'aide au diagnostic du Myélome multiple (voir figure 1.3).

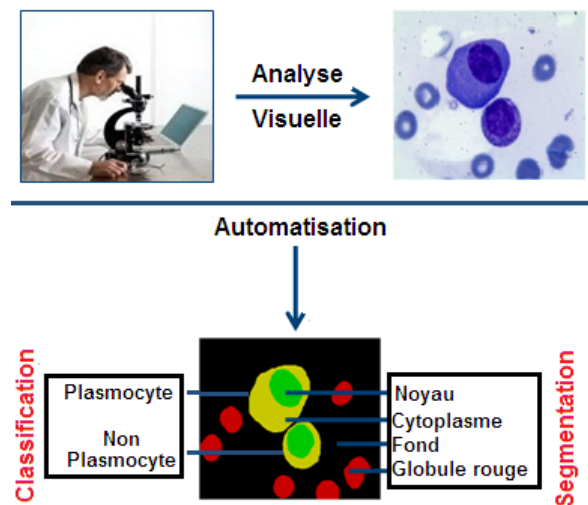


FIGURE 1.3 – Automatisation de l'analyse des images microscopiques

En générale, l'automatisation de l'analyse des images microscopiques cytologiques passe par les étapes d'acquisition, de segmentation et de classification (voir figure 1.4).

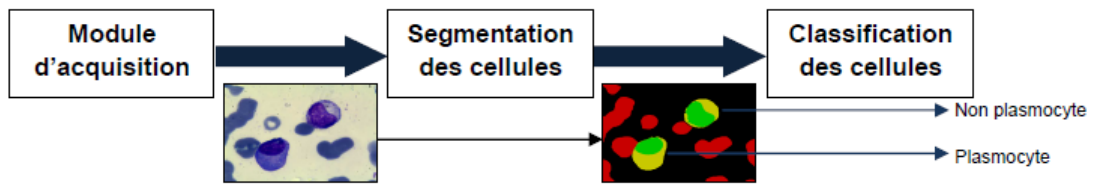


FIGURE 1.4 – Système d'aide au diagnostic

1.3.1 L'acquisition des images microscopiques

Une fois la lame de moelle préparée, et placé sous le microscope, qui est relié à un ordinateur, l'acquisition des images microscopiques se déroule de la manière suivante (voir figure1.5) :

1. Repérage de la zone de lecture ;
2. Grossissement de la zone ;
3. Acquisition et enregistrement des images.

Nous notons, que plusieurs chercheurs optent pour le pré-traitement des images obtenus, dans le but de les simplifier, éliminer le bruit,...etc. En générale le pré-traitement nécessite des heuristiques. Pour notre part, nous avons choisit de n'appliquer aucune technique de pré-traitement, et ainsi rendre notre méthodologie automatique.

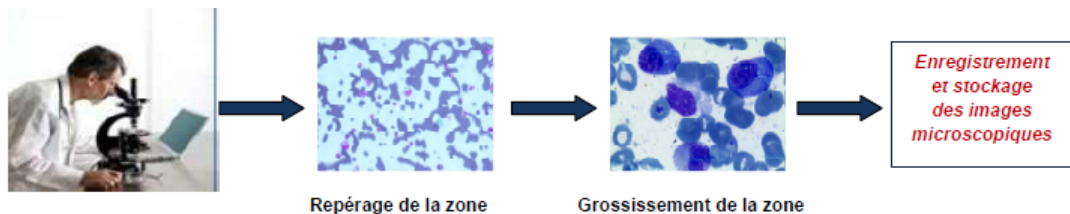


FIGURE 1.5 – Module d'acquisition des images microscopiques

1.3.2 La segmentation des images microscopiques

À l'aide de technique(s) dédiée(s), la segmentation de l'image sert à simplifier et/ou changer la représentation de l'image, de façon à la rendre plus facile à analyser. Dans le cas de la segmentation des images cytologiques, le but étant de retrouver les cellules composantes, i.e., les globules rouges et les globules blancs (avec délimitation de la partie noyau, de la partie cytoplasme), voir

figure 1.6. La réussite de la segmentation agit directement sur la suite du processus d'analyse des images, par conséquent cette étape est la plus cruciale et la plus délicate de tout le système.

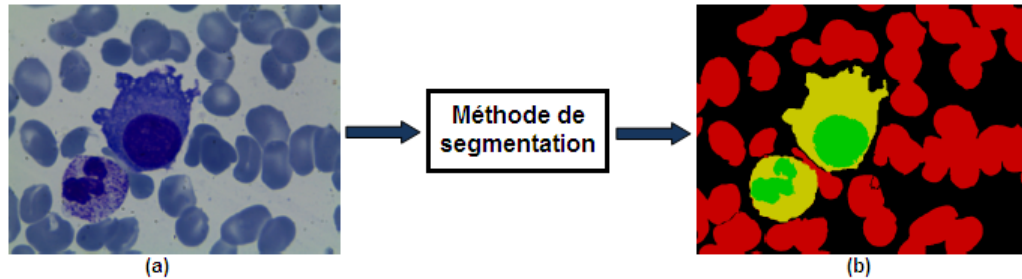


FIGURE 1.6 – Segmentation des images microscopiques ((a) Image originale (b) Image segmentée; vert :noyau, jaune :cytoplasme, rouge :globule rouge et noir :fond)

1.3.3 La classification automatique des globules blancs

À partir des cellules segmentées, des attributs sont extraits et introduits à un modèle prédictif afin de reconnaître la classe du globule blanc. Dans notre cas, nous distinguons les plasmocytes des autres cellules (voir figure 3.28).

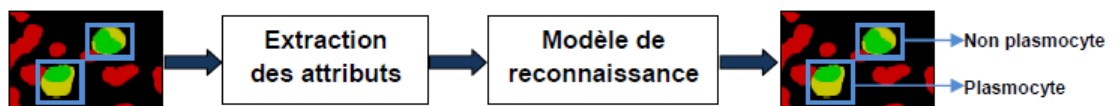


FIGURE 1.7 – Reconnaissance des plasmocytes

1.4 Conclusion

La prolifération des plasmocytes, détectée par l'étude anatomo-pathologique, oriente le médecin vers le diagnostic du Myélome Multiple. La prolifération des plasmocytes, est détectée à travers l'analyse visuelle des lames médullaires au microscope. Cette analyse nécessite des efforts considérables pour une détection fiable et précise, ce qui rend cette tâche lente et fastidieuse. Une solution consiste à automatiser l'analyse, ce qui peut aider le médecin à établir un

diagnostic fiable et rapide. L'automatisation va reposer sur la segmentation des cellules et la classification des globules blancs.

Le chapitre 2 présente les différentes techniques utilisées pour la segmentation des images de cytologie.

Chapitre 2

Segmentation des cellules microscopiques

2.1 Enjeux de la segmentation

La segmentation est le processus qui subdivise l'image en objets ou en régions qui la compose. Le niveau de subdivision est dépendant du problème à résoudre, c'ad que le processus de segmentation doit s'arrêter dès lors que l'objet d'intérêt est isolé [68].

L'imagerie médicale est le processus de création d'une représentation visuelle des composants du corps, afin de procéder à une analyse clinique ou une intervention médicale. L'augmentation du nombre des images médicales a nécessité le recours aux algorithmes de traitement et d'analyse d'images, et plus particulièrement les algorithmes de segmentation qui permettent d'isoler la région d'intérêt. La difficulté de la segmentation des images médicales réside dans l'immense variabilité concernant l'anatomie du corps humain et les différentes modalités d'imagerie médicale (rayon-X, IRM, Microscopie, Endoscopie et beaucoup d'autres). Par conséquent, l'algorithme de segmentation doit intégrer l'information du domaine d'intervention.

L'analyse du sang est importante pour connaître l'état de santé d'un patient (histologie), cependant pour des investigations approfondies le recours à la biopsie est nécessaire (cytologie). La segmentation des cellules est fondamentale pour l'analyse quantitative et qualitative des images cytologiques et histologiques. Dans ce domaine, deux niveaux de segmentation sont à atteindre : la séparation des cellules du fond, et la séparation des composants du

globule blanc (noyau et cytoplasme). Parmi les variantes des images des cellules, celles obtenues de la biopsie sont plus difficile à segmenter, à cause de la diversité des structures des images, les variations du fond dues aux problèmes liés à la fixation et au chevauchement des cellules. La plupart des caractéristiques malins de la cellule sont contenues dans le noyau du globule blanc, et par conséquent l'isolation du noyau de la cellule prend une place importante dans la segmentation des images des cellules [31].

Dans ce qui suit, nous citons quelques travaux de l'état de l'art sur les techniques les plus utilisées dans le domaine de la segmentation des cellules microscopiques.

2.2 Méthodes de segmentation des images microscopiques des cellules

Durant les 50 dernières années, différents algorithmes et techniques ont été développés pour résoudre le problème de la segmentation des images microscopiques des cellules [43]. L'auteur, qui s'est intéressé à tous les types d'image de cellules, fait remarquer qu'aucune méthode n'a été spécialement développée pour la segmentation des cellules, et que le développement de nouvelles techniques d'analyse d'image, conduisait généralement, à l'application sur les images des cellules. Dans la figure 2.1, l'auteur trace l'évolution de l'application de méthodes sur les images de cellules et le nombre de papier traitant le sujet ¹ ; la période du milieu des années 2000 a été fructueuse pour le domaine, et où différentes techniques ont été appliquées, allant des plus simples (seuillage) aux méthodes plus pointues (machine learning, méthode hybride,...etc).

Concernant les images d'hématologie, les auteurs dans [22] regroupent les techniques de segmentation des cellules en deux grandes familles : les méthodes basées régions et les méthodes basées contour. Nous mettons l'accent sur le fait que beaucoup plus d'auteurs traitent les images du sang périphériques que les images de la moelle osseuse.

1. lustrum= période de cinq ans

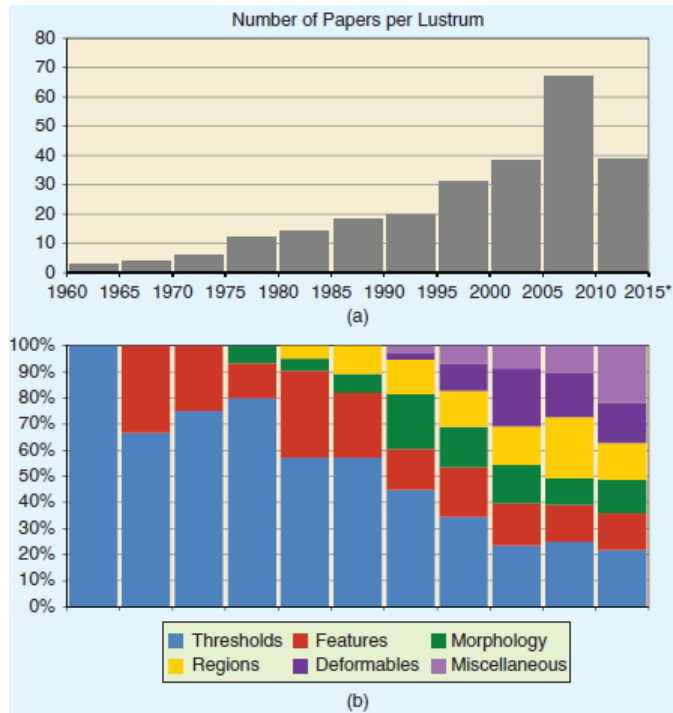


FIGURE 2.1 – Évolution de l’application de techniques de segmentation sur les images de cellules [43]

2.2.1 Méthodes basées régions

2.2.1.1 Seuillage

Le seuillage est la méthode la plus simple utilisée pour la segmentation d’une image. Elle permet de séparer l’objet du fond en utilisant une valeur d’un attribut du pixel qui est comparé à un seuil afin de déterminer la classe du pixel [22].

2.2.1.2 Classification pixellaire

Étant donné une image I , la classification d’un pixel $I(x, y)$ est faite par rapport au vecteur descripteur de ce pixel $l = (u_1..u_k)$. La stratégie dans ces méthodes se décompose en deux étapes : l’extraction des descripteurs du pixel et l’identification de la classe de ce pixel[22].

2.2.1.3 Croissance des régions

Ce sont des algorithmes classiques dans lesquelles un ensemble de régions sont définies, chacune est constitué initialement par un seul pixel, et après une croissance itérative des régions est effectuée selon un certain critère[22].

2.2.1.4 Clusters des régions

Les approches des clusters des régions sont similaires aux méthodes de croissance des régions, cependant le cluster des régions focalise sur la région directement sans aucun marquage. Et contrairement aux approches de classification pixellaire, le cluster des régions identifie les régions cohérentes au lieu de classer chaque pixel indépendamment[22].

2.2.1.5 Segmentation morphologiques - Transformation Watershed

La transformation dite watershed [64] est un outil fondamentale de la morphologie mathématique pour la segmentation des images. Le concept de la transformation repose sur le fait de considérer l'image comme une surface topographique, où les niveaux de gris correspondent aux altitudes des terrains. La segmentation est effectuée par la simulation d'inondation, et les bassins versant résultants² constituent la segmentation finale. La transformation dite watershed classique est appliquée sur l'image gradient, qui contient plusieurs minimas, ce qui conduit à la sur-segmentation [22]. Beaucoup de propositions ont été faites pour parvenir au problème de la sur-segmentation, dont nous faisons partie.

2.2.2 Méthodes basées contours

2.2.2.1 Détection des bordures

Les approches classiques de détection des bordures sont basées sur la détection du changement dans un voisinage des valeurs des pixels. Dans de nombreux cas, les limites entre les cellules et leurs composants ne sont pas clairement définies, et les bordures sont mal détectées sur ces images, et par conséquent la combinaison avec d'autres techniques est proposée [22].

2. (correspondant initialement aux régions minimas de l'image)

2.2.2.2 Modèles déformables ou les contours actifs

Un modèle déformable est une courbe qui évolue vers la limite des contours en utilisant les informations locales de l'image. Deux modèles principaux de contours actifs ont été proposés : les modèles explicites (snake) et les modèles implicites (level set) [22].

2.3 Quelques méthodes appliquées à la segmentation des images cytologiques

Dans le domaine de segmentation des images cytologiques, une multitude de méthodes ont été proposées. Les travaux de ces dix dernières années vont des méthodes les plus simples comme le seuillage globale [50], aux combinaisons de méthodes.

La théorie de l'évidence est combiné avec le Fuzzy C-Means afin de segmenter les cellules dans [17], l'utilisation de Mean Shift et le contour déformable GVF (Gradient Vector Flow) pour segmenter le noyau et le cytoplasme dans [36], la segmentation du noyau par contour actif et la segmentation du cytoplasme par le seuillage de Zack dans [25], un seuillage pour éliminer les globules rouges et le fond et ensuite un autre seuillage pour séparer le noyau du cytoplasme dans [47], un contour actif avec l'orthogonalisation Gram-Schmidt pour segmenter le noyau et le cytoplasme [54] et un réhaussement du noyau suivi d'un seuillage d'Otsu et une estimation de la taille des globule rouge par seuillage pour segmenter les globules blancs dans [32].

L'application de la transformation watershed sur les images cytologiques est omniprésente. Néanmoins, son application directe crée de la sur-segmentation par rapport aux minimas locaux dues aux bruits présents sur les images microscopiques. Par conséquent, nous trouvons dans la littérature plusieurs solutions afin d'éviter la sur-segmentation. Ces solutions reposent sur la combinaison avec d'autres techniques avant et/ou après l'application de la transformation watershed. Dans [59] une transformation watershed contrôlée par marqueur a été appliquée pour segmenter les globules rouges, cette même technique a été appliquée dans [27] pour segmenter les globules blancs. Les régions watershed sont fusionnées par un critère basé sur la granularité pour segmenter les globules blancs dans [46], croissance des régions watershed par regroupement des plans couleurs pour la segmentation des cellules dans [38], un gradient basé

sur la texture et l'intensité comme marqueur de la transformation watershed pour la segmentation des globules blancs dans [26]. Dans [45], les globules blancs sont segmentés par l'extraction du gradient basé sur la saturation, la reconstruction morphologique et la transformation watershed. Dans les travaux de [24] et [23], la segmentation du noyau est faite par la transformation watershed et la méthode level set, tandis que la segmentation du cytoplasme est réalisée par l'analyse granulométrique et les opérations morphologiques. La transformation watershed pour séparer les cellules et le seuillage d'Otsu pour segmenter les globules rouges dans [60]. Une transformation watershed contrôlée par marqueur interne et externe pour segmenter les globules rouges dans les images acquises par caméra DHM (Digital Holographic Microscopy) dans [67]. Les cellules sont segmentées par la fusion basé sur les critères de dissimilarités des régions obtenues par transformation watershed, cette transformation a été appliquée sur deux marqueurs ; le marqueur interne est obtenu par les opérations morphologiques d'ouverture, d'érosion et de dilatation sur les images binaires, alors que le marqueur externe est extrait par la distance et la transformation watershed dans [66]. Les globules blancs, les globules rouges et les plaquettes sont extraits de trois différentes transformation watershed appliquées sur trois différents marqueurs dans [61]. L'application de la transformation watershed basée sur les distances map (fermeture et érosion sur les images binaires) pour la division de l'image en bassins versants, chaque bassin correspond à une cellule dans [40] et [41]. Les images couleurs des images cytologiques sont segmentées par k-means et le seuillage globale, et ensuite les cellules sanguines sont séparées par l'algorithme Sobel et la transformation watershed dans [58]. Les globules blancs sont identifiés par le mean shift basé couleur, et la transformation watershed est appliquée afin de faire face au problème d'adhésion des cellules dans [39]. Pour segmenter le noyau du globe blanc présent dans des conditions normales et leucémiques, la méthode proposée dans [6] est décomposée en plusieurs étapes : conversion des espaces couleurs, le seuillage colorimétriques, le filtrage, une transformation watershed contrôlée par marqueur et différentes opérations morphologiques.

Afin d'obtenir de meilleurs résultats, certains travaux cités précédemment requièrent des améliorations au niveau de : l'estimation du contour du cytoplasme car il dépasse le contour réel dans [24], le marqueur du cytoplasme et le pouvoir de séparer le grand chevauchement dans [59], l'ajustement des images binaires comme c'est l'entrée de la transformation watershed dans [46],

la régularisation de la magnitude du gradient pour localiser les régions d'intérêt incluant le noyau et le cytoplasme comme une cellule entière dans [45] et l'ajustement des marqueurs afin de traiter le grand chevauchement puisque le but du travail dans [61] est de compter les cellules.

Des limites ont été observés dans [67], [38], [66] et [26]. En effet, les paramètres doivent être régularisés pour les différents temps de stockage des globules rouges dans [67], l'ajout d'un autre critère sur les régions non-labéllisées afin de bien segmenter les variations du cytoplasme dans [38], l'ajustement du seuillage pour éviter la sous-segmentation dans [66] et la base d'images doit être étendue, pour avoir des résultats concluants, puisque elle ne contient que 13 images et aucune scène complexe dedans [26].

2.4 Méthode proposée

2.4.1 Notions préliminaires

La transformation dite watershed considère les minimas de l'image comme les objets d'intérêt et les maximas comme les frontières séparatrices entre les objets. Cependant, l'existence de plusieurs minimas locaux provoque une sur-segmentation. Les attributs couleurs des régions watershed offrent la possibilité de rassembler ceux qui sont similaires. Les différentes représentations colorimétriques mènent à faire face à l'incertitude, cette dernière est issue de l'imprécision, l'incomplétude et la complémentarité. Le processus de fusion des données permet de tirer profit de l'incertitude liée à la couleur de la région sur-segmentée, et à partir de là, obtenir la meilleure décision parmi les hypothèses considérées.

2.4.1.1 Projection sur différents espaces couleurs

Les descripteurs couleurs sont parmi les attributs les plus importants utilisés dans le domaine de traitement d'image. Ils fournissent une information puissante pour simplifier la détection et l'identification des objets. Chaque pixel d'une image couleur est spécifique dans un espace couleur, ce qui sert de système de coordonnées couleurs, et le système couleur commun est le système RGB (Red Green Blue).

Il est toujours difficile de prédire le comportement d'une méthode sur un espace couleur donné, puisque chaque espace couleur s'intéresse à différents

stimulus [8].

Dans les images cytologiques, les globules rouges ont toujours la même apparence, cependant, les globules blancs ont diverses formes dues aux différentes classes. Ces différences se portent essentiellement sur la forme du noyau et la couleur du cytoplasme. Nous tirons profit des différentes représentations colorimétriques, qui sont le RGB, HSV(Hue Saturation Value) et le CMYK (Cyan Magenta Yellow black), afin de distinguer les globules blancs (noyau et cytoplasme) des globules rouges. Ces trois espaces ont les propriétés suivantes :

* **Espace RGB** : C'est le modèle couleur le plus utilisé. Les images couleurs dans les moniteurs et les caméras vidéos sont représentées dans l'espace RGB, et sont communément nommés des images RGB. Les couleurs dans le modèle RGB sont obtenues par la combinaison linéaire des couleurs primaires rouge, vert et bleu [16].

* **Espace CMYK** : Le modèle CMY prend son nom de Cyan (bleu-vert), Yellow (jaune) et Magenta (rose). Ce sont des couleurs secondaires, elles sont les couleurs primaires des pigments. Elles sont aussi appelées des primaires soustraites car elles sont obtenues par la soustraction de la lumière du blanc[16]. Ce modèle est utilisé en imprimerie. La transformation pour obtenir ces couleurs est :

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 - C \\ 1 - M \\ 1 - Y \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1)$$

Où $R, G, B \in [0, 1]$

L'espace CMYK est une variation de l'espace CMY. Il ajoute du noir (black). L'espace couleur CMYK remplit l'écart entre la théorie et la pratique. En théorie, la quantité du noir n'est pas nécessaire. Cependant, l'expérience avec plusieurs encres et papiers a démontré que lorsque les composants Cyan, Magenta et Yellow sont égales, les encres sont mixés ce qui donne un marron foncé, et non un noir. La solution est de rajouter l'encre noir pour remédier à ça. d'où la naissance de ce modèle.

* **Espace HSV** : Le modèle HSV, où HSV est pour Hue (Nuance), Saturation (saturation) et Value (valeur). Les coordonnées HSV peuvent être obtenues du modèle RGB par les transformations qui vont suivre [16].

Premièrement les valeurs RGB sont normalisées par :

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \quad (2.2)$$

Et ensuite les valeurs H,S et V sont calculées ainsi :

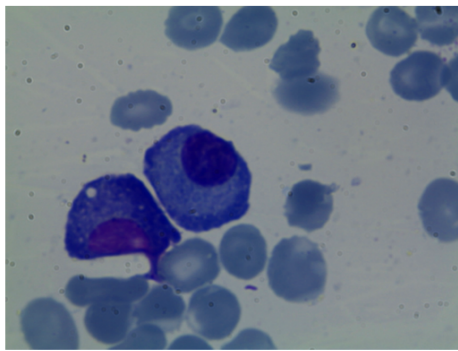
$$V = \max(r, g, b) \quad (2.3)$$

$$S = \begin{cases} 0 & \text{si } V = 0 \\ V - \frac{\min(r,g,b)}{V} & \text{si } V > 0 \end{cases} \quad (2.4)$$

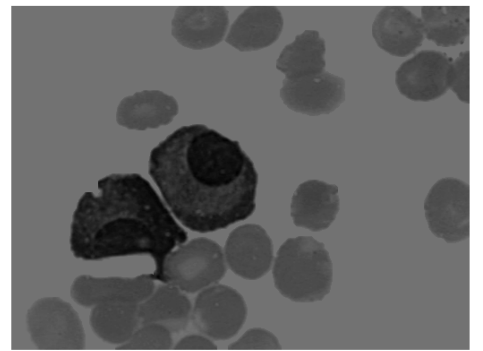
$$H = \begin{cases} 0 & \text{si } S = 0 \\ \frac{60*(g-b)}{S*V} & \text{si } V = r \\ 60 * [2 + \frac{b-r}{S*V}] & \text{si } V = g \\ 60 * [4 + \frac{r-g}{S*V}] & \text{si } V = b \end{cases} \quad (2.5)$$

$$H = H + 360 \text{ si } H < 0 \quad (2.6)$$

Le choix de ces espaces couleurs n'est pas anodin mais lié au fait que ces représentations colorimétriques différencient au mieux les cellules entre elles ; voir la figure 2.2 : les globules rouges sont plus claires que les globules blancs sur le canal Magenta (espace CMYK). Et voir figure 2.3 : tous les cytoplasmes sont similaires aux globules rouges sur le canal Magenta, cependant la différence est clairement observée sur l'espace HSV.



(a)



(b)

FIGURE 2.2 – (a) Image originale (b) Projection sur le canal Magenta

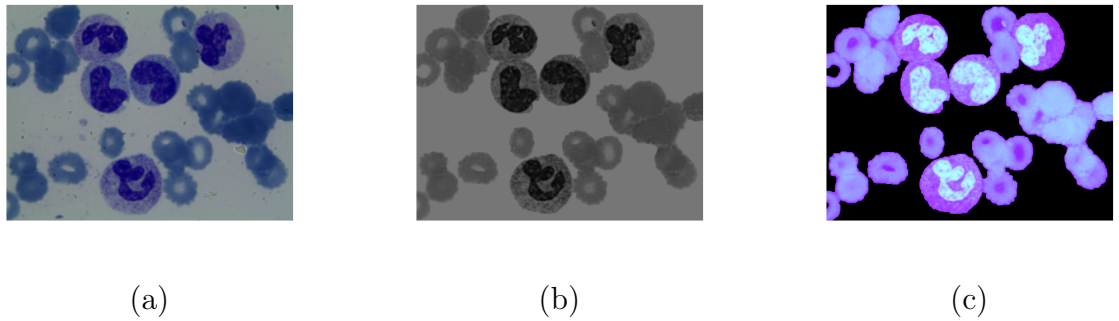


FIGURE 2.3 – (a) Image originale (b) Projection sur le canal Magenta (c) Projection sur l'espace couleur HSV

2.4.2 Principe de la théorie de l'évidence

L'objectif de la fusion d'informations est d'améliorer la qualité des décisions en augmentant la quantité globale d'informations tout en diminuant leur incertitude, en utilisant la redondance et la complémentarité des informations entre les sources. Et donc, c'est une technique pour intégrer les données hétérogènes provenant de différentes sources, en vue d'optimiser l'estimation. La fusion est basée sur le raisonnement incertain, développé dans la théorie des probabilités bayésiennes, la théorie de Dempster-Shafer (théorie de l'évidence), la théorie des possibilité et bien d'autres. L'approche bayésienne a une théorie de prise de décision, mais elle nécessite une connaissance complexe de la combinaison des probabilités conditionnelle, et la spécification des connaissances à priori de la distribution des probabilités, démontrant que la preuve est présente. De plus, la limite majeure de l'approche bayésienne est qu'elle ne peut pas modéliser l'imprécision. En d'autres termes, la probabilité bayésienne ne peut pas mesurer les évidences avec une imprécision dans les mesures des probabilités.

La théorie de l'évidence est une extension générale de la théorie bayésienne, qui peut traiter les données incomplètes avec robustesse [5]. Il existe trois bonnes raisons de considérer la théorie de l'évidence pour la fusion d'informations :

1. Elle modélise l'imprécision et l'incertitude, elle est donc considérée comme une approche plus flexible et générale que la théorie des probabilités.
2. Elle offre la possibilité de traiter les probabilités de plusieurs hypothèses, tandis que la la théorie des probabilités classique traite seulement une

hypothèse.

3. Et finalement, la force majeure de la théorie de l'évidence, est sa capacité de considérer les informations manquantes et incomplètes.

La théorie de l'évidence de Dempster-Shafer a été développée par Dempster, qui s'est intéressé aux probabilités inférieures et supérieures, et ensuite Shafer introduit sa contribution en offrant les fonctions de croyances pour modéliser la connaissance incertaine, en se basant sur des fondements mathématiques [33].

Cette théorie est basée sur deux idées principales : obtenir un degrés de croyance sur une hypothèse (appelée la "fonction de masse") à partir de probabilités subjectives, et la règle de Dempster (somme orthogonale) pour combiner ces degrés de croyances quand ils sont basés sur des éléments d'évidences indépendants [29]. D'une manière plus formelle :

Soit Ω un ensemble fini d'hypothèses exhaustives et mutuellement exclusives, appelé le cadre de discernement :

$$\Omega = \{H_1, \dots, H_n\}$$

L'ensemble de probabilités $P(\Omega)$ composé avec les 2^n propositions de Ω définit par :

$$P(\Omega) = \{\phi, \{H_1\}, \{H_2\}, \dots, \{H_n\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Omega\}$$

Quand le cadre de discernement est déterminé, la fonction de masse (the Basic Probability Assignment (BPA)), est définie par la table de correspondance de l'ensemble des probabilités $P(\Omega)$ à un nombre entre 0 et 1, i.e. : $m : P(\Omega) \rightarrow [0, 1]$, qui satisfie les conditions ci-dessous :

$$\begin{cases} m(\phi) = 0 \\ \sum_{A \subseteq \Omega} m(A) = 1 \end{cases} \quad (2.7)$$

Si A est considéré comme un sous ensemble de $\Omega (A \subseteq \Omega)$, $m(A)$ indique le degrés de croyance attribué à exactement l'ensemble A et non à aucun sous ensemble de A .

Partant du fait que la théorie de l'évidence a la capacité de calculer le degrés d'incertitude, elle est utilisée aussi, pour combiner différentes évidences, afin d'augmenter la valeur de la certitude, ce qui conduit à des décisions plus précises.

La combinaison par la règle de Dempster, des différentes évidences, dans la-

quelle deux sources d'informations indépendantes (m_1 et m_2) sont fusionnées pour créer une nouvelle fonction de croyance, est la suivante :

$$m(C) = \frac{\sum_{A \cap B = C} m_1(A) \times m_2(B)}{1 - \sum_{A \cap B = \phi} m_1(A) \times m_2(B)} \quad (2.8)$$

Où le dénominateur peut être interprété comme un critère de conflit entre les évidences indépendantes à être combinées [29]. L'augmentation de $m(A)$ correspond à renforcer la croyance en A .

Il est important de bien modéliser la connaissance des différentes sources d'informations S afin d'initialiser les fonctions de masses associées. De plus, modéliser la croyance sur un ensemble de probabilités peut aboutir à une complexité exponentielle des opérations réalisées sur les fonctions de masses relatives au nombre de classes. Cependant, la fonction de masse est souvent interprétée comme une confiance associée à une certaine hypothèse, vu que la définition de la fonction de masse dans la théorie de l'évidence reste un problème non résolu et il n'y a pas de solution générale.

2.4.3 Segmentation évidentielle des cellules

Dans cette section, nous décrivons la méthode de segmentation proposée (cette méthode et ses résultats préliminaires ont été présentés dans [11]). Notre idée, est d'utiliser la puissance de la théorie de l'évidence pour classer les régions issues de la transformation watershed.

La transformation watershed donne une sur-segmentation due à l'existence des minimas locaux. Cependant, la considération simultanée des régions sur-segmentées et des différentes projections couleurs (vu dans la section 2.4.1.1), permettra l'amélioration de la segmentation des images. C'est à dire, la combinaison des informations issues de différentes sources (l'énergie watershed et les informations couleurs), réduira la sur-segmentation.

2.4.3.1 Segmentation initiale par transformation watershed

Notre but, à cette étape, est d'avoir une segmentation initiale des contours connectés. À cette fin, la transformation watershed a été appliquée sur le gradient de l'image originale. En effet, cette méthode permet d'avoir une partition initiale de l'image en régions.

L'image sur-segmentée est considérée comme une partition initiale de l'image,

sur laquelle est appliquée une procédure d'agrégation évidentielle de régions (voir la section suivante). Il est évident, que la probabilité d'une agrégation erronée, est liée à la qualité de la segmentation initiale. De plus, la complexité computationnelle de la procédure d'agrégation, dépend du nombre des partitions initiales. Il est possible de limiter ce nombre de partitions, en empêchant la sur-segmentation dans les régions homogènes, où la magnitude du gradient est faible.

2.4.3.2 Agrégation évidentielle des régions

Dans ce travail, quelques modèles de la littérature ont été testés : Shafer [29], Appriou [4], et Denoeux [21]. Ce dernier a donné de meilleurs résultats, étant donné que la connaissance sur chaque classes est modélisée par la distance entre les exemples et le centre de la classe. Par conséquent, la fonction de masse est définie par :

$$m(H_i) = \alpha_i \times \exp(-\gamma_i \times \text{dist}(x, x_i)) \quad (2.9)$$

avec $\alpha_i \in [0, 1]$ est le paramètre d'affaiblissement, $\gamma_i > 0$ relié à l'hypothèse H_i , $\gamma_i = \frac{1}{d_i}$ avec, d_i la distance moyenne entre les exemples représentatifs de (H_i) , et $\text{dist}(x, x_i)$ la distance euclidienne entre la valeur moyenne de la zone traitée x , et x_i la valeur moyenne des exemples représentatifs de H_i .

L'algorithme proposée (Algorithme 1) est présenté comme suit :

Algorithm 1 AGRÉGATION ÉVIDENTIELLE DES RÉGIONS

ENTREE: r_k {Régions watershed}**SORTIE:** Image segmentée : {NOYAU}, {CYTOPLASME}, {RBC}**Etape1 :****for all** r_k **do**

-Calculer les fonctions de masses avec eq 2.9 sur le canal magenta.

 $m(H_i) = \alpha_i \times \exp(-\gamma_i \times dist(x, x_i))$ avec $i = \{WBC, RBC\}$

-Combinee les degrés de croyances avec la somme orthogonale eq 2.8

-Attribuer la classe j (BPA plus élevée) à r_k i.e :**if** $i = WBC$ **then**WBC = {WBC} + $\{r_k\}$ {WBC est l'ensemble des WBC identifiés}**else if** $i = RBC$ **then**RBC = {RBC} + $\{r_k\}$ {RBC est l'ensemble des RBC identifiés}**end if****end for****Etape2 :****for all** RBC **do**

-Calculer BPA sur l'espace HSV

if $m(RBC) < m(WBC)$ **then**

WBC = {WBC} + {RBC}

RBC = {RBC} - {RBC}

end if**end for****Etape3 :****for all** WBC **do**

-Calculer BPA sur le canal vert {Hypothèses sur Noyau et Cytoplasme}

if $m(Noyau) > m(Cytoplasms)$ **then**

NOYAU = {NOYAU} + {Noyau}

else

CYTOPLASME = {CYTOPLASME} + {Cytoplasme}

end if**end for**

En d'autres termes l'agrégation de régions est basée sur un ordre hiérarchique, permettant la segmentation des différents composants de l'image microscopique de la moelle osseuse. Les trois étapes de cette procédure s'exécutent comme suit :

1. Nous associons à chaque région labellisée un degré de croyances pour les hypothèses WBC et RBC. Dans la majorité des cas, nous avons remarqué que le canal Magenta, issu de l'espace couleur CMYK, permet de différencier les RBCs des autres cellules (voir figure 2.2 : RBCs sont plus clairs que le WBC). Cependant, dans quelques cas le cytoplasme est

confondu avec le RBC, ce qui nous ramène à l'étape suivante.

2. Nous sommes passé à l'espace HSV et nous avons appliqué une classification évidentielle. Pour ce faire, nous avons considéré une nouvelle fonction de masse. Cette dernière consiste à représenter chaque région RBC par sa valeur moyenne sur l'espace HSV :

$$m(H_i) = \frac{1}{[R_i]} \sum_{j \in R_i}^{[R_i]} I(j) \quad (2.10)$$

où H_i est l'hypothèse sur la région R_i , $[R_i]$ est le cardinal de la région R_i , et $I(j)$ est l'intensité du pixel j appartenant à la région considérée. (voir figure 2.3 : Cytoplasme du WBC d'en haut est similaire au RBC, mais la différence est clairement faite dans l'espace HSV). Ensuite, l'heuristique suivante est appliquée : **si** $m(H_i) > 0.5$, alors la région est reclassée WBC, **sinon**, elle reste une région RBC.

3. Et finalement, pour tous les WBCs identifiés, la valeur moyenne sur le canal vert est calculé, afin de séparer le globule blanc en cytoplasme et noyau.

2.5 Conclusion

La segmentation est une étape importante du traitement dans les recherches médicales et les applications cliniques, où la prise de décision reste un point critique. L'analyse automatique des cellules, peut réduire le risque d'erreur et le temps de traitement pour établir un diagnostic.

La transformation watershed est un outil simple et robuste, néanmoins il présente l'inconvénient de mener à la sur-segmentation due aux bruits présents dans les images microscopiques. Plusieurs solutions sont proposées dans la littérature pour remédier au problème de la sur-segmentation, dont nous faisons partie. Nous avons proposé d'intégrer l'incertitude liée à la couleur des régions watershed, pour segmenter les globules blancs (noyau et cytoplasme), les globules rouges et le fond.

Chapitre 3

Caractérisation et classification des cellules

3.1 Introduction

La reconnaissance et le comptage automatique des leukocytes dans les images microscopiques, sont d'une importance primordiale dans le diagnostic des maladies du sang. La classification des cellules, s'avère essentielle pour connaître le type et le développement de la maladie, ce qui peut être d'une grande utilité dans le processus thérapeutique clinique.

Les valeurs numériques des globules blancs, provenant des images microscopiques du sang et des sections tissulaires, représentent un challenge différent à relever. En effet, le bon diagnostic repose sur la détection précise et fiable, par conséquent, l'erreur peut être soit minimale, soit inexistante.

La reconnaissance automatique du type du globule blanc passe par deux étapes : la caractérisation et la classification. Dans ce qui suit nous présentons, tout d'abord une généralité sur l'étude des cellules sanguines, ensuite nous passons à la méthode de caractérisation du globule blanc, et la méthode de reconnaissance du plasmocyte.

3.2 Étude des cellules sanguines

Le corps humain a la capacité de produire quotidiennement 440 milliards de cellules sanguines, grâce à l'hématopèse. C'est un phénomène physiologique, dynamique et continu qui assure une production quotidienne de 440 milliards

de cellules sanguines. Ceci correspond, à la production de plus de 200 milliards d'hématies (globules rouges), d'environ 200 milliards de plaquettes et de 40 milliards de polynucléaires (globules blancs). Cette activité considérable de production est assurée par une petite population de cellules de la moelle osseuse appelées cellules souches hématopoïétiques [2]. Les cellules souches se divisent en 5 catégories suivantes :

3.2.1 Lymphopoïèse

Dans cette catégorie nous retrouvons :

3.2.1.1 Lymphoblaste

Le lymphoblaste (voir figure 3.1) est caractérisé par [3] :

- 10-20 microns
- Noyau rond à ovale avec une structure de chromatine fine et uniformément colorée, 1-2 nucléoles
- Cytoplasme bleu avec une zone claire proche du noyau
- Pas de granules

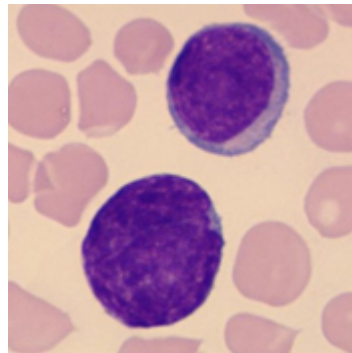


FIGURE 3.1 – Le lymphoblaste [3]

3.2.1.2 Prolymphocyte

Le prolymphocyte (voir figure 3.2) est caractérisé par [3] :

- 10-20 microns
- Noyau rond, structure un peu plus petite que le lymphoblaste
- Cytoplasme bleu
- Pas de granules

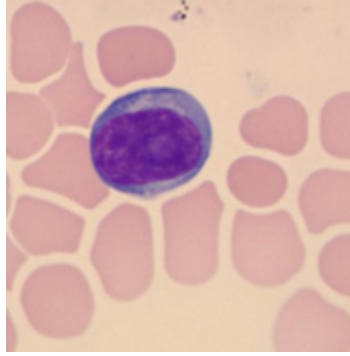


FIGURE 3.2 – Le prolymphocyte [3]

3.2.1.3 Petit lymphocyte

Le petit lymphocyte (voir figure 3.3) est caractérisé par [3] :

- 7-8 microns
- Noyau rond à ovale, avec une structure de la chromatine condensée. Pas de nucléoles
- Bande de cytoplasme basophile (un bleu clair ou foncé)
- Quelques granules violets brillants

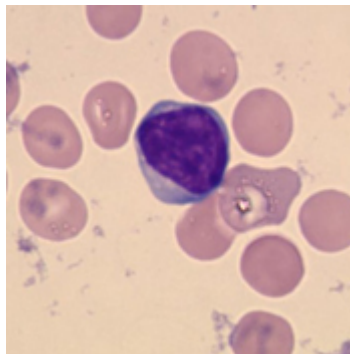


FIGURE 3.3 – Le petit lymphocyte [3]

3.2.1.4 Grand lymphocyte

Le grand lymphocyte (voir figure 3.4) est caractérisé par [3] :

- 9-15 microns
- Noyau ovale/bombé, avec une structure de la chromatine plus petite que celle du petit lymphocyte. Pas de nucléoles
- Cytoplasme est incolore à bleu clair

- Quelques granules violets brillants

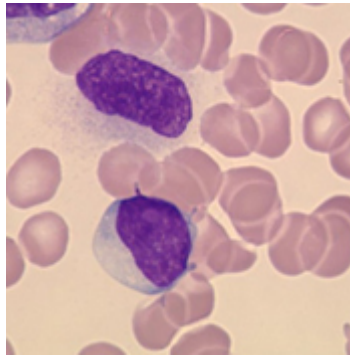


FIGURE 3.4 – Le grand lymphocyte [3]

3.2.1.5 Plasmocyte

Le plasmocyte (voir figure 3.5) est caractérisé par [3] :

- 7-15 microns
- Noyau rond et excentré. Structure de chromatine compacte (radiale)
- Cytoplasme bleu foncé avec zone périnucléaire

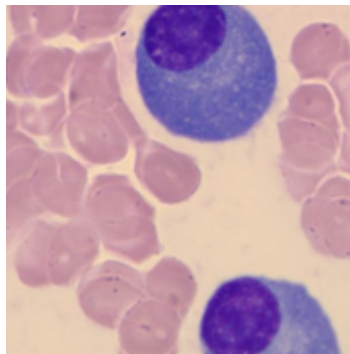


FIGURE 3.5 – Le plasmocyte [3]

3.2.2 Mégacaryopoïèse

Dans cette catégorie nous retrouvons :

3.2.2.1 Mégacaryoblaste

Le mégacaryoblaste (voir figure 3.6) est la cellule qui engendre le mégacaryocyte, difficile à identifier morphologiquement [3] :

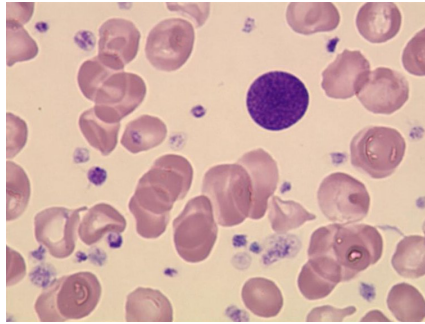


FIGURE 3.6 – Le mégacaryoblaste [3]

3.2.2.2 Mégacaryocyte

Le mégacaryocyte (voir figure ??) est caractérisé par [3] :

- 25-90 microns
- Cellule ronde avec paroi cellulaire irrégulière
- Noyau avec 2-16 lobes. La taille de la cellule varie en fonction du nombre de lobes
- Chromatine finement dispersée dans le noyau. Des petits nucléoles peuvent être présents.
- Cytoplasme bleu à rose

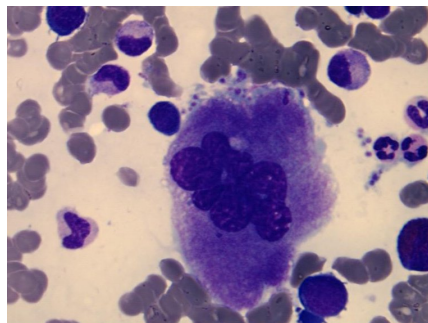


FIGURE 3.7 – Le mégacaryocyte [3]

3.2.2.3 Thrombocyte

Le thrombocyte (voir figure 3.8) est caractérisé par [3] :

- 1-15 microns
- Fragment de cytoplasme de la mégacaryocyte
- Cytoplasme incolore à bleu clair

- Granules rouges éparses

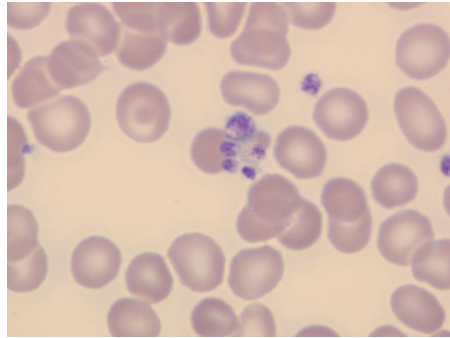


FIGURE 3.8 – Le thrombocyte [3]

3.2.3 Monopoièse

Dans cette catégorie nous retrouvons :

3.2.3.1 Promonocyte

Le promonocyte (voir figure 3.9) est caractérisé par [3] :

- 15-20 microns
- Forme du noyau irrégulière, sans lobes, et une structure de la chromatine fine avec nucléole
- Cytoplasme bleu-gris
- Quelques granules peuvent être présentes

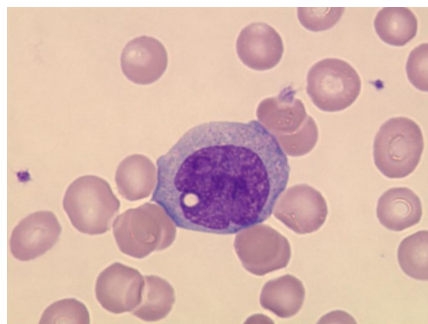


FIGURE 3.9 – Le promonocyte [3]

3.2.3.2 Monocyte

Le monocyte (voir figure 3.10) est caractérisé par [3] :

- 12-24 microns
- Noyau de forme rénale, bombé ou segmenté, avec une structure de la chromatine très petite
- les granules violets peuvent se produire
- Les vacuoles apparaissent souvent dans le cytoplasme

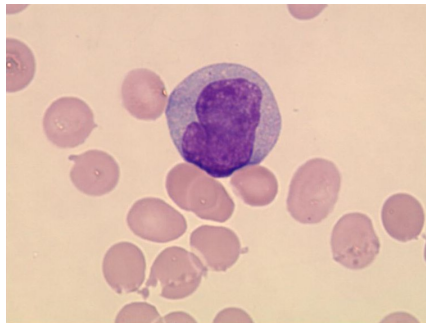


FIGURE 3.10 – Le monocyte [3]

3.2.4 Myélopoïèse

Dans cette catégorie nous retrouvons :

3.2.4.1 Myeloblaste

Le myeloblaste (voir figure 3.11) est caractérisé par [3] :

- 15-20 microns
- Grand noyau rond au milieu de la cellule, structure fine de la chromatine avec 2-5 nucléoles
- Petite quantité de cytoplasme bleu, parfois avec une zone claire
- Pas de granules

3.2.4.2 Promyélocyte

Le promyélocyte (voir figure 3.12) est caractérisé par [3] :

- 12-24 microns

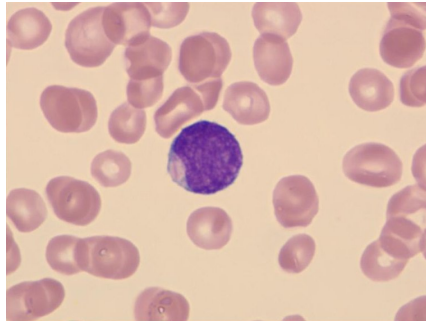


FIGURE 3.11 – Le myeloblaste [3]

- Noyau ovale légèrement bombé, souvent avec une position excentrique. Structure de la chromatine légèrement plus grossière que le myeloblaste, avec 3-4 nucléoles visibles
- Cytoplasme bleu pâle
- Granules violettes de taille inégale

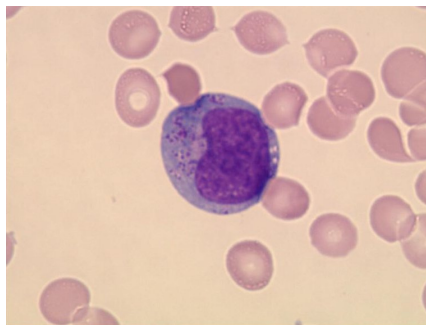


FIGURE 3.12 – Le promyélocyte [3]

3.2.4.3 Myélocyte

Le myélocyte (voir figure 3.13) est caractérisé par [3] :

- 10-18 microns
- Noyau rond à ovale, souvent excentré
- Structure de la chromatine grossière et plus condensée, et qui ne contient pas de nucléoles
- Cytoplasme rose à incolore (plus immature, bleuâtre)
- Granulation modérée très fine à dense (granules secondaires)
- Myélocyte neutrophile : granules violettes
- Myélocyte éosinophile : granules roses vifs

- Myélocyte basophile : granules bleu-violet (noir)

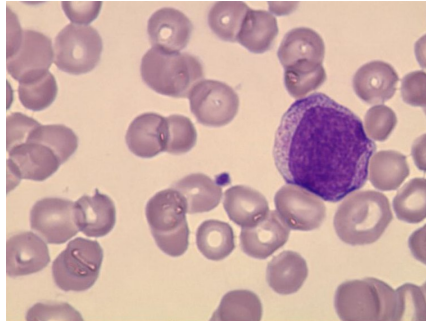


FIGURE 3.13 – Le myélocyte [3]

3.2.4.4 Metamyélocyte

Le metamyélocyte (voir figure 3.14) est caractérisé par [3] :

- 10-16 microns
- Noyau indenté avec une structure grossière de la chromatine agglomérée, qui ne contient pas de nucléoles.
- Cytoplasme rose claire à incolore
- Beaucoup de granules secondaires
- Metamyélocyte neutrophile : granules violettes
- Metamyélocyte éosinophile : granules roses
- Myélocyte basophile : granules bleu-violet (noir)

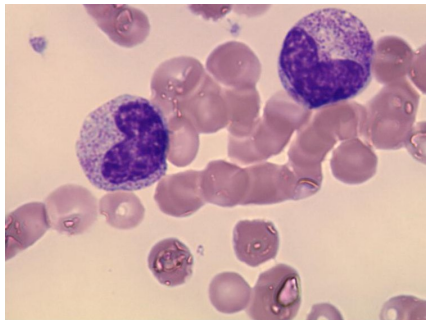


FIGURE 3.14 – Le metamyélocyte [3]

3.2.4.5 Granulocyte à bande neutrophile

Le granulocyte à bande neutrophile (voir figure 3.15) est caractérisé par [3] :

- 10-15 microns
- Noyau en forme de bande avec structure de chromatine grossière et agglomérée. La partie la plus étroite est au moins 1/3 de la plus large. Pas de nucléoles
- Cytoplasme rose claire à incolore
- Granules violette très fines

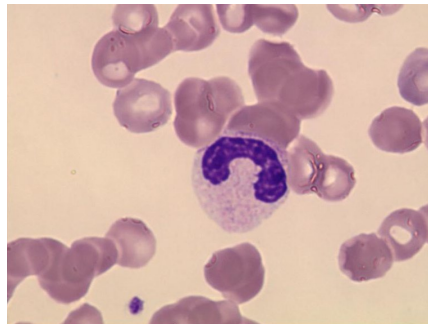


FIGURE 3.15 – Le granulocyte à bande neutrophile [3]

3.2.4.6 Granulocyte à neutrophile segmenté

Le granulocyte à neutrophile segmenté (voir figure 3.16) est caractérisé par [3] :

- 10-15 microns
- Structure de la chromatine grossière et agglomérée, avec 3-5 lobes. Pas de nucléoles.
- Cytoplasme rose claire à incolore
- Granules violette secondaires abondantes

3.2.4.7 Granulocyte éosinophile

Le granulocyte éosinophile (voir figure 3.17) est caractérisé par [3] :

- 10-15 microns
- Structure de la chromatine grossière et agglomérée, avec 2-3 lobes. Pas de nucléoles.
- Cytoplasme rose claire à incolore

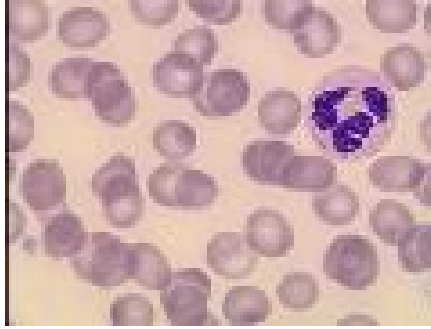


FIGURE 3.16 – Le granulocyte à neutrophile segmenté [3]

- Granules rouge-brique secondaires abondantes

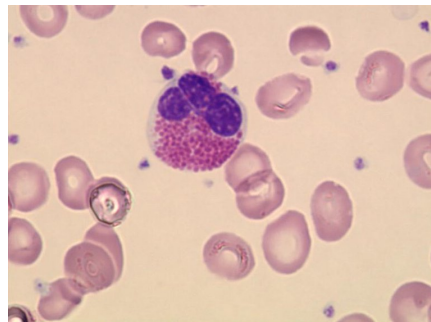


FIGURE 3.17 – Le granulocyte éosinophile [3]

3.2.4.8 Granulocyte basophile

Le granulocyte basophile (voir figure 3.18) est caractérisé par [3] :

- 10-15 microns
- Structure de la chromatine grossière et agglomérée, avec 3-4 lobes. Pas de nucléoles.
- Cytoplasme rose claire à incolore
- Granules bleu-violet (noir) secondaires en distribution inégale

3.2.5 Érythropoïèse

Dans cette catégorie nous retrouvons :

3.2.5.1 Proerythroblast

Le proerythroblast (voir figure 3.19) est caractérisé par [3] :

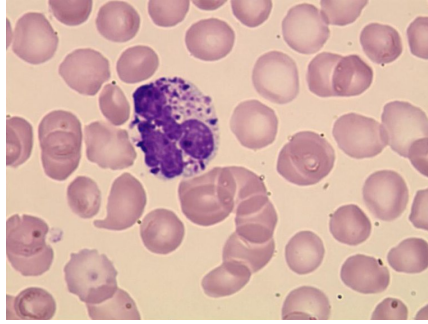


FIGURE 3.18 – Le granulocyte basophile [3]

- 14-24 microns
- Grand noyau rond avec structure fine de la chromatine avec 2-5 nucléoles
- Cytoplasme étroit bleu foncé, avec une structure irrégulière avec zone claire périnucléaire bien définie
- Pas de granules

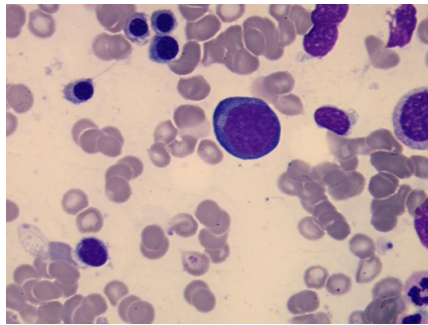


FIGURE 3.19 – Le proerythroblast [3]

3.2.5.2 Érythroblaste basophile

L'érythroblaste basophile (voir figure 3.20) est caractérisé par [3] :

- 10-17 microns
- Noyau rond plutôt avec une structure de la chromatine légèrement condensée. Aucun nucléole
- Bordure du cytoplasme foncée
- Pas de granules

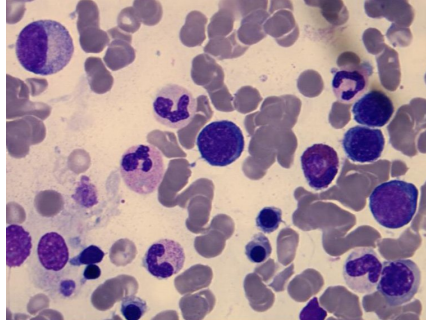


FIGURE 3.20 – L'érythroblaste basophile [3]

3.2.5.3 Erythroblaste polychromatique

L'érythroblaste polychromatique (voir figure 3.21) est caractérisé par [3] :

- 10-15 microns
- Noyau rond petit, avec une structure de la chromatine assez condensée et sans nucléoles
- Cytoplasme bleu-gris
- Pas de granules

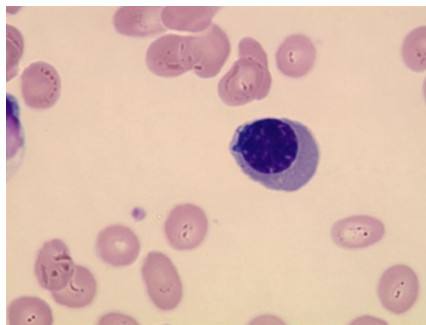


FIGURE 3.21 – L'érythroblaste polychromatique [3]

3.2.5.4 Erythroblaste orthochromatique

L'érythroblaste orthochromatique (voir figure 3.22) est caractérisé par [3] :

- 8-12 microns
- Noyau rond et petit, souvent au centre de la cellule, avec une structure de la chromatine condensée
- Cytoplasme rougeâtre abondant avec des composants légèrement bleus
- Pas de granules

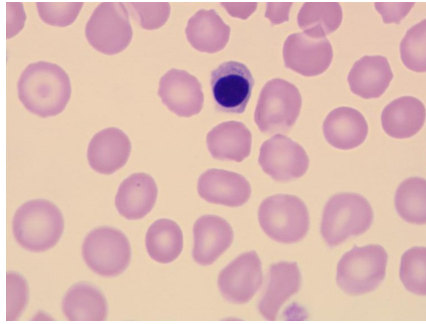


FIGURE 3.22 – L'erythroblaste orthochromatique [3]

3.2.5.5 Erythrocyte polychromatique

L'erythrocyte polychromatique (voir figure 3.23) est caractérisé par [3] :

- 7-10 microns
- Pas de noyau
- Cytoplasme rougeâtre avec une composante bleue plus faible
- Pas de granules

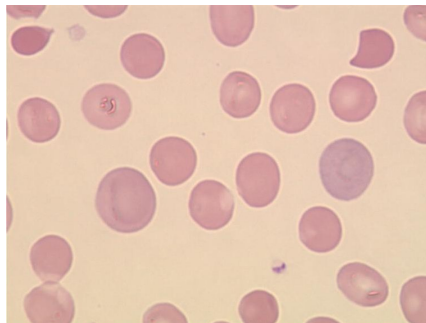


FIGURE 3.23 – L'erythrocyte polychromatique [3]

3.2.5.6 Erythrocyte

L'erythrocyte (voir figure 3.24) est caractérisé par [3] :

- 6-8 microns
- Pas de noyau
- Cytoplasme légèrement rougeâtre, clair vers le centre
- Pas de granules

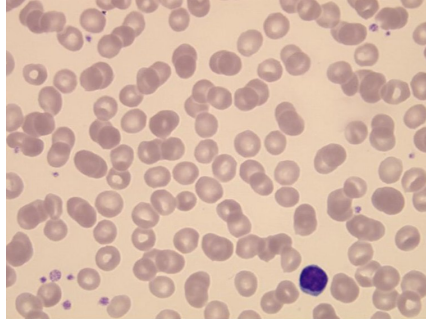


FIGURE 3.24 – L'erythrocyte [3]

3.3 Caractérisation des cellules

La littérature récente de la classification des cellules dans les images microscopiques, se focalise essentiellement sur les attributs de texture, de taille et de formes. Généralement, les attributs utilisés dans la caractérisation des globules blancs sont divisés en deux types ; attributs géométriques et attributs de textures [57].

3.3.1 Attributs géométriques

Pour la différenciation entre les cellules, les attributs géométriques jouent un rôle vital, ils décrivent la structure et la taille du globule blanc. La surface et le périmètre du globule blanc sont les attributs utilisés pour représenter la taille de la cellule, quant aux attributs de formes, ils peuvent être divisés en attributs régions et en attributs contours. Pour extraire les attributs, l'image microscopique cytologique est convertit en image binaire, où les pixels de la cellule sont représentés par les valeurs non nulles [57], les attributs les plus présents dans la littérature [57] sont décrit dans ce qui suit :

- *Surface* : La surface de la cellule est représentée par le nombre totale des pixels non nuls avec le contour de la cellule.
- *Rapport de surface* : Il est défini comme le rapport de nombre de pixels dans le cytoplasme au nombre de pixels dans le noyau.
- *Surface convexe* : Dans certains cas, la coque convexe est calculée et sa surface est appelée comme le nombre de pixels se trouvant à l'intérieur de son contour.

- *Solidité* : Le rapport de la surface à la surface convexe

$$Solidite = \frac{Surface}{Surface\ convexe} \quad (3.1)$$

- *Périmètre* : Il est calculé à travers la mesure de la somme des distances entre les pixels du contours successifs.
- *Éccentricité* : Le rapport de la longueur des axes majeur et mineur.

$$Eccentricite = \frac{Longueur\ de\ l'axe\ majeur}{Longueur\ de\ l'axe\ mineur} \quad (3.2)$$

- *Compacité* : Le rapport de la surface au périmètre au carré

$$Compacite = \frac{Surface}{Perimetre^2} \quad (3.3)$$

- *Nombre de lobes* : Certaines cellules (poly-nuclée) ont différents nombre de noyau (lobes) dans leur cytoplasme, ce qui peut être utilisé comme attributs important dans la classification des globules blancs.
- *Orientation* : C'est l'angle entre l'axe des 'x' et l'axe majeur de la cellule.
- *Circularité* : Elle est définie par le rapport du plus petit périmètre du cercle de délimitation au périmètre.
- *Rectangularité* : C'est le rapport du plus petit périmètre du rectangle de délimitation au périmètre.
- *Facteur de forme* : C'est un paramètre sans dimension.

$$Facteur\ de\ forme = \frac{4 * \pi * Surface}{Perimetre^2} \quad (3.4)$$

- *Symétrie* : Elle représente la différence entre les lignes qui sont perpendiculaire à l'axe majeur du contour de la cellule dans les deux directions.
- *Concavité* : Elle est définie comme l'extension à laquelle le contour actuel de la cellule se trouve à l'intérieur de chaque corde entre les points du contour non adjacents
- *Élongation* : C'est la mesure du rapport entre les distances maximales et les distances minimales, du centre de gravité au contour du noyau.

3.3.2 Attributs de textures

D'une façon générale, il n'y a pas une définition exacte de la texture. C'est une perception visuelle des objets, qui ont des caractéristiques similaires. La texture est définie dans [18] comme une fonction de la variation spatiale dans l'intensité des pixels. La texture de l'image est définie à travers des propriétés d'uniformité, de variance, d'entropie, ...etc. Les attributs de textures statistiques (attributs de Haralick) définis dans [30] sont aussi connus dans le domaine de l'analyse de texture [57]. Dans ce qui suit, nous présentons les attributs de textures les plus utilisés pour caractériser les globules blancs :

- *Moyenne* : La valeur moyenne de l'intensité du pixel (μ) à l'intérieur du globule blanc, peut être calculée par la somme des valeurs de l'intensité des pixels de la cellule divisée par le nombre des pixels de la cellule.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (3.5)$$

où x_i est la valeur du i^{me} pixel, et n est le nombre total des pixels.

- *Variance* : C'est la différence au carré par rapport à la moyenne

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} \quad (3.6)$$

- *Déviation standard* : C'est la racine carrée de la variance

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (3.7)$$

- *Entropie* : C'est la mesure du désordre.

$$Entropie = - \sum_i \sum_j p_d(i, j) \log p_d(i, j) \quad (3.8)$$

avec $p_d(i, j)$ représente le nombre d'occurrences de la paire (i, j) distant de d .

- *Énergie* : c'est la mesure de l'uniformité à l'intérieur de la cellule.

$$Energie = - \sum_i \sum_j p_d^2(i, j) \quad (3.9)$$

- *Homogénéité* : c'est le degré de la variation.

$$Homogeniete = - \sum_i \sum_j \frac{p_d(i, j)}{1 + |i - j|} \quad (3.10)$$

- *Corrélation* : c'est la corrélation entre les valeurs de pixels et son voisinage.

$$Correlation = - \sum_i \sum_j \frac{(i - \mu)(j - \mu)p_d(i, j)}{\sigma_x \sigma_y} \quad (3.11)$$

3.4 Classification des cellules

Un classifieur automatique est une fonction X , s'appuyant sur un modèle d'apprentissage, cette fonction associe une classe cl à un vecteur d'attributs $v = \{v_1, v_2, \dots, v_n\}$; la classe cl réunit des caractéristiques communes.

En générale, l'apprentissage peut se faire selon trois types : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

3.4.1 Apprentissage supervisé

L'objectif principale dans l'apprentissage supervisé est d'apprendre un modèle à partir des données annotées, qui permettent de faire des prédictions sur des données nouvelles. Le terme supervisé fait référence à un ensemble d'exemples où la sortie désirée (la classe) est déjà connue [53], voire figure 3.25.

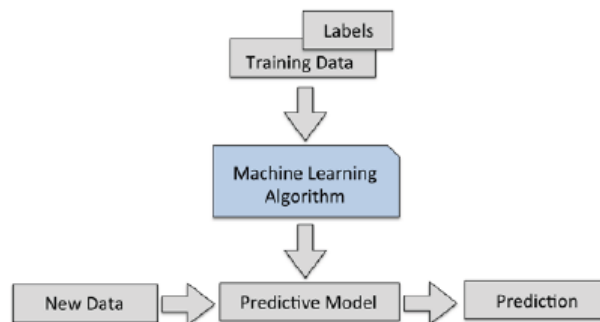


FIGURE 3.25 – Apprentissage supervisé [53]

3.4.2 Apprentissage par renforcement

Dans l'apprentissage par renforcement, le but est de développer un système (agent) qui améliore ses performances en se basant sur les interactions avec l'environnement. Puisque l'information sur l'état courant de l'environnement inclut un signal de récompense, l'un peut penser à une catégorie de l'apprentissage supervisé. Cependant, dans l'apprentissage par renforcement, le retour arrière ne correspond pas à la vérité terrain, mais à la mesure, par la fonction de récompense, de combien l'action a été bien accompli. À travers l'interaction avec l'environnement, un agent peut alors utiliser l'apprentissage par renforcement afin d'apprendre des séries d'actions qui maximise la fonction de récompense, via une approche d'exploration essai-et-erreur ou une planification délibérative [53], voir figure 3.26.

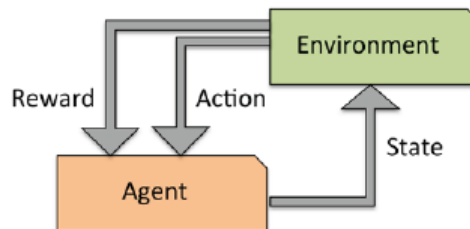


FIGURE 3.26 – Apprentissage par renforcement [53]

3.4.3 Apprentissage non supervisé

Dans l'apprentissage supervisé, la réponse correcte est connue lors de l'entraînement du modèle, et dans l'apprentissage par renforcement, une mesure de récompense est définie pour les actions particulières par un agent. Cependant, dans l'apprentissage non supervisé les données sont non labellisées ou ont des structures inconnues. En utilisant les techniques d'apprentissage non supervisés, il est possible d'explorer les structures des données afin d'en extraire des informations significatives, sans avoir recours de la sortie de variables ou la fonction de récompense [53].

3.4.4 Algorithmes utilisés

Dans nos travaux, un médecin au CHU de Tlemcen nous a annoté les globules blancs de type plasmocyte. Par conséquent, nous nous sommes intéressés

à l'apprentissage supervisé. En particulier trois algorithmes SVM(Support Vector Machines), K-plus proches voisins, et l'arbre de décisions, ont été utilisés. Nous allons les décrire brièvement dans ce qui suit :

- **SVM** : C'est un algorithme d'apprentissage supervisé. Les SVM reposent sur la théorie de l'apprentissage statistique de Vapnik et de Chervnenkis [19]. Ils ont la capacité de traiter des données de grandes dimensions et ne dépendent pas de l'espace de représentation de ces données. Ils reposent sur la marge maximale et la fonction noyau (polynomiale, rbf,...).
- **Kppv** : C'est un algorithme d'apprentissage supervisé, qui est considéré, jusqu'à aujourd'hui comme un algorithme de classification puissant. Il s'appuie sur la mémorisation des données d'apprentissage labélisées, et attribuer à un nouveau exemple la classe la plus proche parmi les k données d'apprentissage. Il est recommandé en pratique de choisir un k impaire et pas trop petit.
- **Arbre de décision supervisé** : c'est une structure de données hiérarchique implémentant la stratégie "diviser pour régner". C'est une méthode non paramétrique efficace, et peut être utilisée tant bien dans la classification que dans la regression [7].

3.5 Méthodes de reconnaissance automatique des cellules dans la littérature

Après des résultats satisfaisants de segmentation, les papiers suivants traitent la deuxième partie i.e la caractérisation et la classification des cellules. D'une part, il y a ceux qui se sont intéressés aux globules rouges, en les comptant dans [59], et dans [60], les globules rouges sont caractérisés par des attributs géométriques et statistiques afin de détecter les cellules infectées à l'aide d'un classifieur SVM, ensuite un autre classifieur SVM est entraîné sur des attributs couleurs et géométriques des cellules infectées, dans le but d'identifier l'état d'évolution du parasite Malariya. Et d'une autre part, beaucoup de chercheurs se sont intéressés aux globules blancs, dont nous faisons partis, dans [47] le classifieur LDA (Linear Discriminant Analysis) a été entraîné afin de reconnaître cinq types de globules blancs. Le SFS (Sequential Forward Selection) pour sélectionner les attributs significatifs des globules blancs segmentés, et après une comparaison des performances de classification du SVM et du per-

ceptron multi-couches a été réalisé dans [32]. La reconnaissance de cinq types de globules blancs dans [27] et dans [24] et de trois types dans [50], basés sur Naive Bayes, Kppv et classification par règles, respectivement. Les attributs géométriques, statistique et de texture sont utilisés dans [41] pour identifier dix classes de globules blanc à l'aide du classifieur SVM, et dans [40] pour reconnaître dix sept classes des cellules leucémiques. Les attributs des noyaux et cytoplasmes segmentés sont réduits par le SFS et ensuite introduit au classifieur Bayes pour reconnaître cinq type de globules blancs dans [49]. La classification des cellules est basée sur l'apprentissage neuronal fuzzy Min Max dans [34]. Les régions des globules blancs sont classifiés séparément, sur la base des attributs de la forme, la taille, la couleur et la texture, en faisant appel à différentes techniques floues et non floues, dans le but d'identifier cinq types de globules blancs dans [28].

3.6 Modèle proposé pour la détection des plasmocytes

Dans cette section, nous décrivons notre méthodologie, introduite dans [10], pour la reconnaissance automatique des plasmocytes. Ce processus intervient après la phase de segmentation. L'identification des plasmocytes se fait en deux étapes : la caractérisation (figure 3.27) et la classification (figure 3.28).

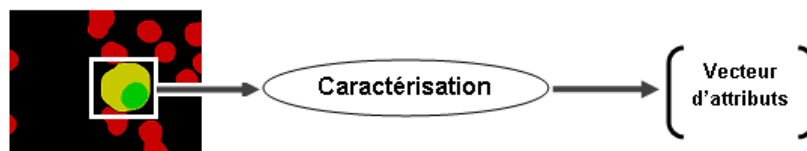


FIGURE 3.27 – Caractérisation du globule blanc



FIGURE 3.28 – Reconnaissance du plasmocyte

3.6.1 Caractérisation des globules blancs

Notre but dans l'étape de caractérisation est de se rapprocher le plus de la description du médecin; le plasmocyte a une forme ovale ou ronde. Il a une grande taille, avec un important rapport noyau-cytoplasme. Son noyau se caractérise par son éccentricité et sa forme ovale. Le cytoplasme est basophile, est contient une zone pale (d'apparence blanche et distincte). Combiné à la description médicale, nous avons ajouté les contraintes techniques, ce qui nous a mené à la caractérisation avec trente et un attributs suivants :

- *Éccentricité de la cellule*
- *Solidité de la cellule*
- *Compacité de la cellule*
- *Circularité de la cellule*
- *Rectangularité de la cellule*
- *Facteur de forme de la cellule*
- *Solidité du noyau*
- *Compacité du noyau*
- *Circularité du noyau*
- *Rectangularité du noyau*
- *Facteur de forme du noyau*
- *Moyenne du noyau sur le canal vert et sur le canal Magenta*
- *Variance du noyau sur le canal vert et sur le canal Magenta*
- *Déviation standard du noyau sur le canal vert et sur le canal Magenta*
- *Entropie du noyau sur le canal vert et sur le canal Magenta*
- *Moyenne du cytoplasme sur le canal vert et sur le canal Magenta*
- *Variance du cytoplasme sur le canal vert et sur le canal Magenta*
- *Déviation standard du cytoplasme sur le canal vert et sur le canal Magenta*
- *Entropie du cytoplasme sur le canal vert et sur le canal Magenta*
- *Rapport du noyau à la cellule*
- *Différence de la valeur maximale et la valeur minimale sur la canal magenta du noyau*
- *Différence de la valeur maximale et la valeur minimale sur la canal vert du noyau*
- *Différence de la valeur maximale et la valeur minimale sur la canal magenta du cytoplasme*

- *Différence de la valeur maximale et la valeur minimale sur la canal vert du cytoplasme*

3.6.2 Reconnaissance des plasmocytes

Pour que le plasmocyte soit identifié, il faut tout d’abord entraîner un algorithme, afin que ce dernier aura le pouvoir de distinguer les plasmocytes, des non plasmocytes. Dans nos travaux, comme mentionné précédemment, nous nous sommes intéressé aux algorithmes d’apprentissage supervisé. Par conséquent, cette phase de reconnaissance va se dérouler en deux phases : la première concerne l’apprentissage (il est parfois appelé phase hors ligne), nous entraînons les algorithmes sur des données d’apprentissage (un ensemble des attributs des globules blancs labellisés $\{Plasmocyte \text{ ou } non \text{ plasmocyte}\}$, ce qui va conduire à la construction d’un modèle prédictif. Ce dernier est utilisé dans la deuxième phase, c’est la phase de reconnaissance, ici nous introduisons les attributs d’un nouvel globule blanc, et le modèle doit pouvoir prédire sa classe.

3.7 Conclusion

Dans le modèle automatique de reconnaissance de globules blancs, c’est la phase de classification qui reste visible du système. En amont, la segmentation reste la phase clé (pour la différenciation des composants cellulaires), ensuite la caractérisation, pour donner une description au globule blanc, cette dernière sera utilisée à l’entrée du classifieur.

Nous proposons un modèle de reconnaissance de plasmocytes (cellule importante pour le diagnostic du myélome), nous nous sommes rapprochés de la description médicale du plasmocyte et nous avons rajoutés des aspects techniques liés à la texture couleur. Les attributs obtenus de cette caractérisation, ont introduit à trois différents classifieurs, en se basant sur un apprentissage supervisé, dans le but de reconnaître les plasmocytes.

Chapitre 4

Approche proposée pour l'aide au diagnostic du Myélome

4.1 Introduction

L'analyse manuelle des images microscopiques nécessitent des efforts considérables de la part de l'expert, et par conséquent le diagnostic établi reste subjectif, étant donné qu'il est étroitement lié à l'expérience de l'expert et son niveau de concentration. L'interprétation des images microscopiques médicales peut bénéficier des avancées des outils et techniques d'analyse automatique des images.

Les systèmes d'aide au diagnostic permettent de faire un diagnostic établi uniquement par l'expert humain. Ce dernier utilise la sortie de l'analyse automatique des images médicales, comme opinion secondaire dans l'interprétation des scènes et la détection des éventuelles anomalies. La décision finale revient toujours à l'expert humain, d'où l'appellation 'aide au diagnostic'.

4.2 Modèle automatique proposé

Notre modèle proposé est présenté dans la figure 4.1. Les trois étapes principales de notre approche sont : la segmentation évidentielle des cellules, la caractérisation des globules blancs et la classification automatique des plasmocytes.

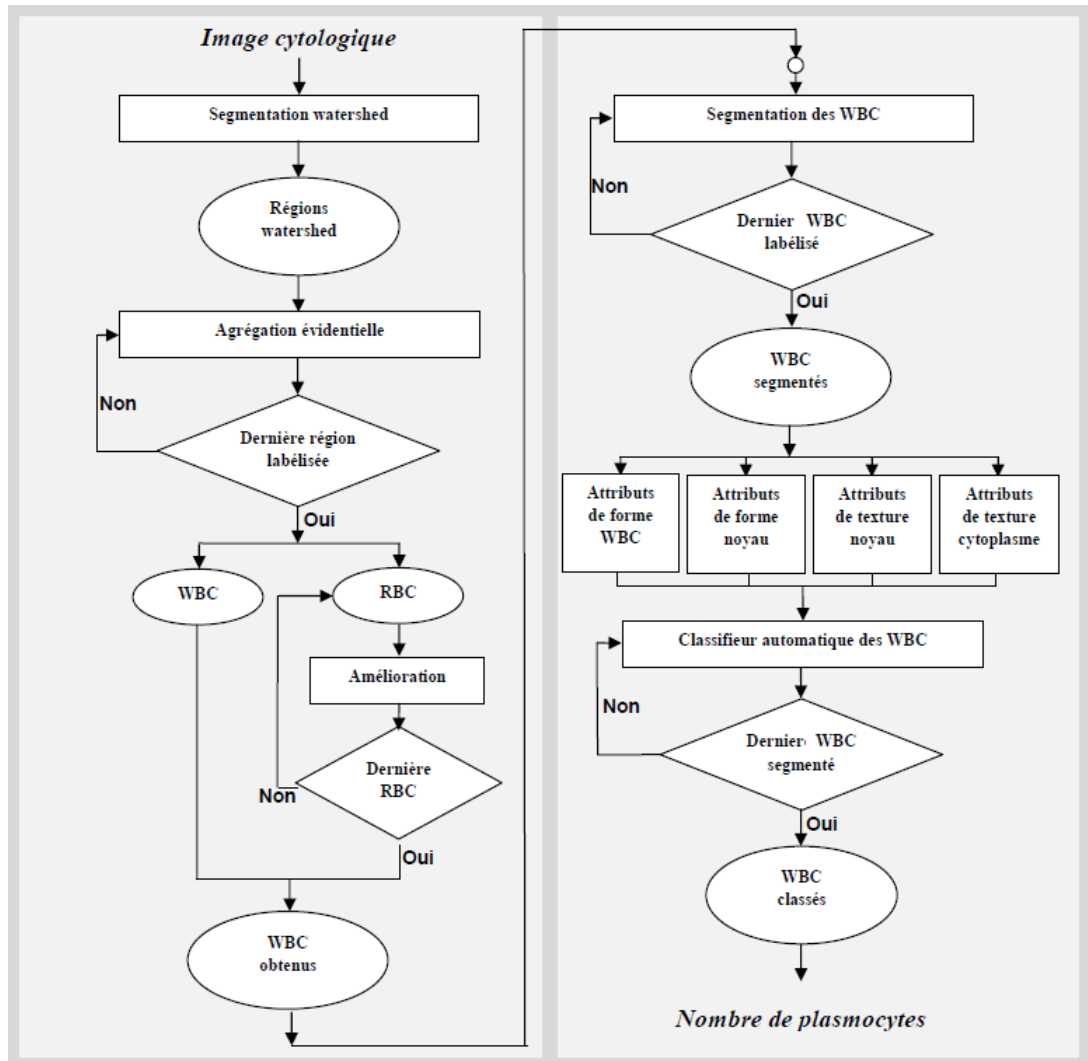
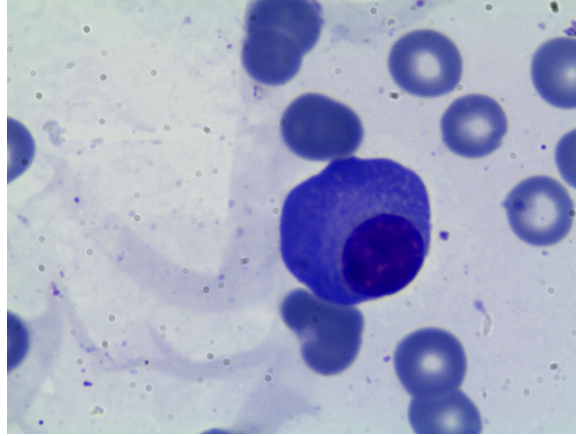


FIGURE 4.1 – Système proposé pour l’aide au diagnostic du Myélome

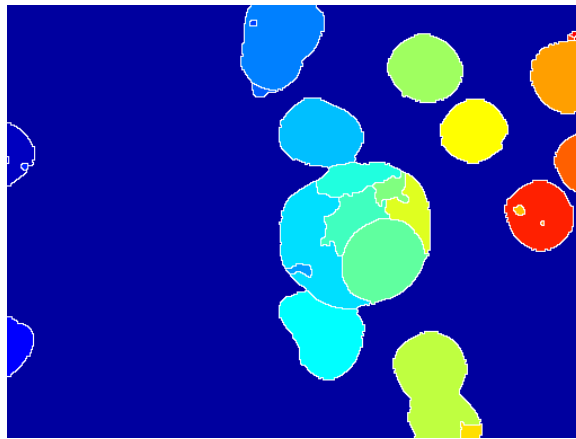
4.2.1 Segmentation évidentielle des cellules

Notre technique de segmentation évidentielle est basée sur un ordre hiérarchique, permettant la segmentation des différents composants de l’image microscopique de la moelle osseuse. Les quatre étapes du processus de segmentation se déroulent de la façon suivante :

1. Segmentation initiale par transformation watershed (voir figure 4.2).



(a)



(b)

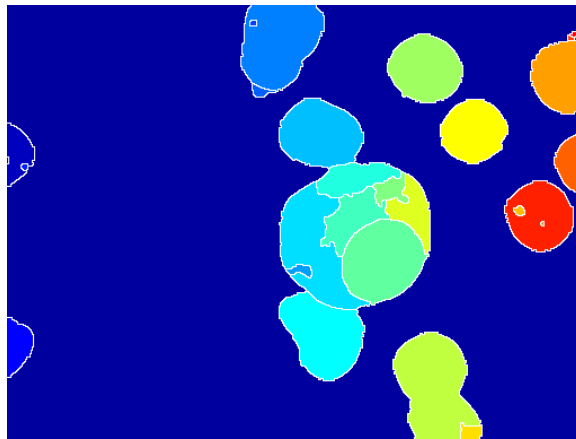
FIGURE 4.2 – Transformation watershed : (a) Image originale (b) Image sur-segmentée par la transformation watershed

2. Nous associons à chaque région labellisée un degrés de croyance pour les hypothèses WBC et RBC. Dans la majorité des cas, nous avons remarqué que le canal Magenta, issu de l'espace couleur CMYK, permet de différencier les RBCs des autres cellules. Cependant, dans quelques cas le cytoplasme est confondu avec le RBC, ce qui nous ramène à l'étape suivante.
3. Nous sommes passés à l'espace HSV et nous avons appliqué une classification évidentielle. Pour ce faire, nous avons considéré une nouvelle fonction de masse. Cette dernière consiste à représenter chaque région

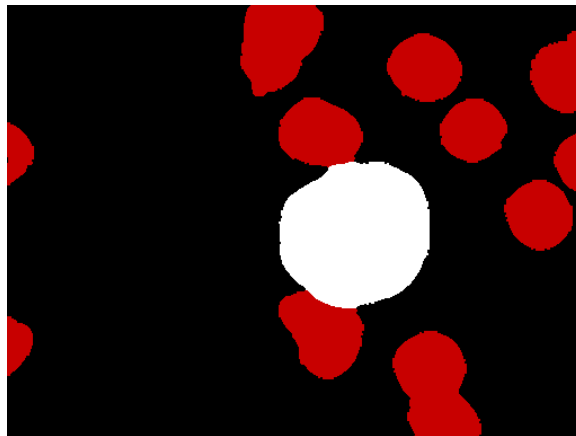
RBC par sa valeur moyenne sur l'espace HSV :

$$m(H_i) = \frac{1}{|R_i|} \sum_{j \in R_i} I(j) \quad (4.1)$$

où H_i est l'hypothèse sur la région R_i , $|R_i|$ est le cardinal de la région R_i , et $I(j)$ est l'intensité du pixel j appartenant à la région considérée. Ensuite, l'heuristique suivante est appliquée : **si** $m(H_i) > 0.5$, alors la région est reclassée WBC, **sinon**, elle reste une région RBC. La figure 4.3 expose la segmentation évidentielle en globule blanc Vs globule rouge.



(a)

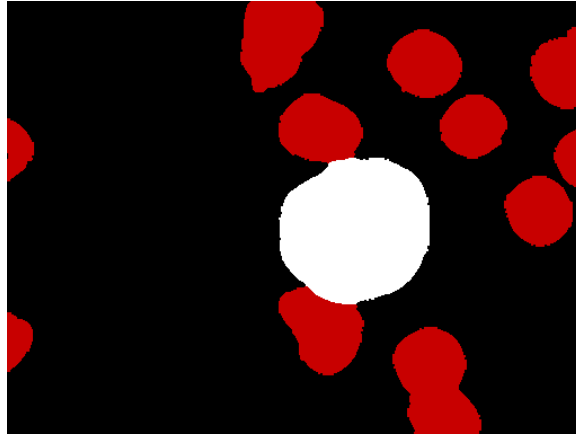


(b)

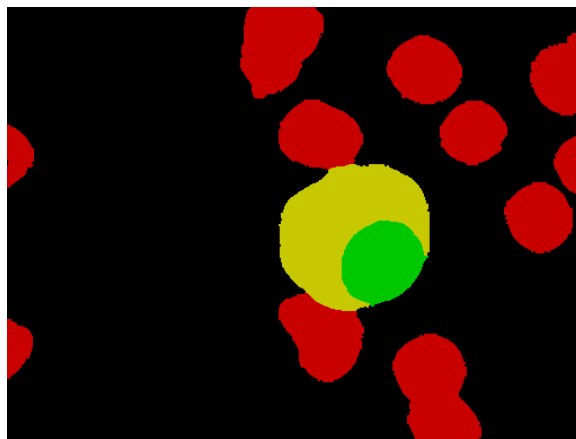
FIGURE 4.3 – Blanc Vs rouge : (a) Image sur-segmentée par la transformation watershed (b) Segmentation globule blanc (blanc) Vs globule rouge (rouge)

4. Et finalement, pour tous les WBCs identifiés, la valeur moyenne sur le

canal vert est calculé, afin de séparer le globule blanc en cytoplasme et noyau. La figure 4.4 montre la segmentation finale en noyau, cytoplasme, globule rouge et fond.



(a)



(b)

FIGURE 4.4 – Segmentation finale : (a) Segmentation globule blanc Vs globule rouge (b) Image segmentée : noyau(vert), cytoplasme(jaune), globule rouge (rouge) et fond (noir)

4.2.2 Caractérisation des globules blancs

Le plasmocyte a une forme ovale ou ronde. Il a une grande taille, avec un important rapport noyau-cytoplasme. Son noyau se caractérise par son éccentricité et sa forme ovale. Le cytoplasme est basophile, est contient une zone pale (d'apparence blanche et distincte).

Dans les systèmes automatiques de classification de leukocytes, la majorité

des chercheurs font appel aux attributs de formes et de textures, afin de caractériser le globule blanc, et par la même occasion, augmenter la précision du classifieur [57]. Pour aider les cliniciens, nous avons considéré les informations d'aspects médicales, ainsi que les informations d'aspects techniques, afin de bien caractériser les globules blancs. Pour ce faire, nous avons considéré les attributs suivants :

1. Attributs de formes :

- *Éccentricité*
- *Solidité*
- *Compacité*
- *Circularité*
- *Rectangularity*
- *Facteur de forme*

2. Attributs de texture :

- *Moyenne*
- *Variance*
- *Déviation standard*
- *Entropie*

En effet, nous avons pris les attributs de formes de la cellule entière, ensuite, nous avons considéré les attributs de formes et de texture pour le noyau. Aussi, nous avons pris les attributs de textures du cytoplasme. Nous avons tenu compte des projections couleurs pour l'extraction des attributs de textures, par conséquent, les attributs ont été calculés sur le canal magenta et sur le canal vert. Enfin, nous avons ajouté l'aspect du plasmocyte : le rapport du noyau à la cellule et les différences sur la canal magenta et sur le canal vert pour le noyau et pour le cytoplasme. Cela nous a conduit à obtenir les 31 attributs suivants :

- Attributs de forme de la cellule.
- Attributs de forme du noyau.
- Attributs de texture du noyau sur le canal magenta et le canal vert.
- Attributs de texture du cytoplasme sur le canal magenta et le canal vert.
- Rapport du noyau à la cellule.
- les différences de la valeur maximale et la valeur minimale sur la canal magenta et sur le canal vert, repectivement.

4.2.3 Reconnaissance automatique des plasmocytes

Le but principale de ce travail est de concevoir, un modèle automatique pour la reconnaissance des plasmocytes, dans les images microscopiques de la moelle osseuse.

À la différence de l'observation humaine, la classification des globules blancs est un processus complexe pour les systèmes informatiques, ce qui a motivé les chercheurs à étudier ce problème dans le contexte de l'intelligence artificielle [57]. Dans ce travail de thèse, nous nous sommes intéressé spécialement aux algorithmes d'apprentissage supervisé, étant donné que notre base est labellisée. Le processus se déroule en deux phases (voir la figure 4.5) :

- Phase d'apprentissage (Hors ligne) : Introduction des attributs des leucocytes extraits et leurs labels à l'algorithme d'apprentissage, ce qui donne un modèle prédictif.
- Phase de reconnaissance (En ligne) : Les attributs du leucocyte segmenté sont introduits au modèle prédictif, généré de la phase d'apprentissage, pour être reconnu plasmocyte ou non plasmocyte.

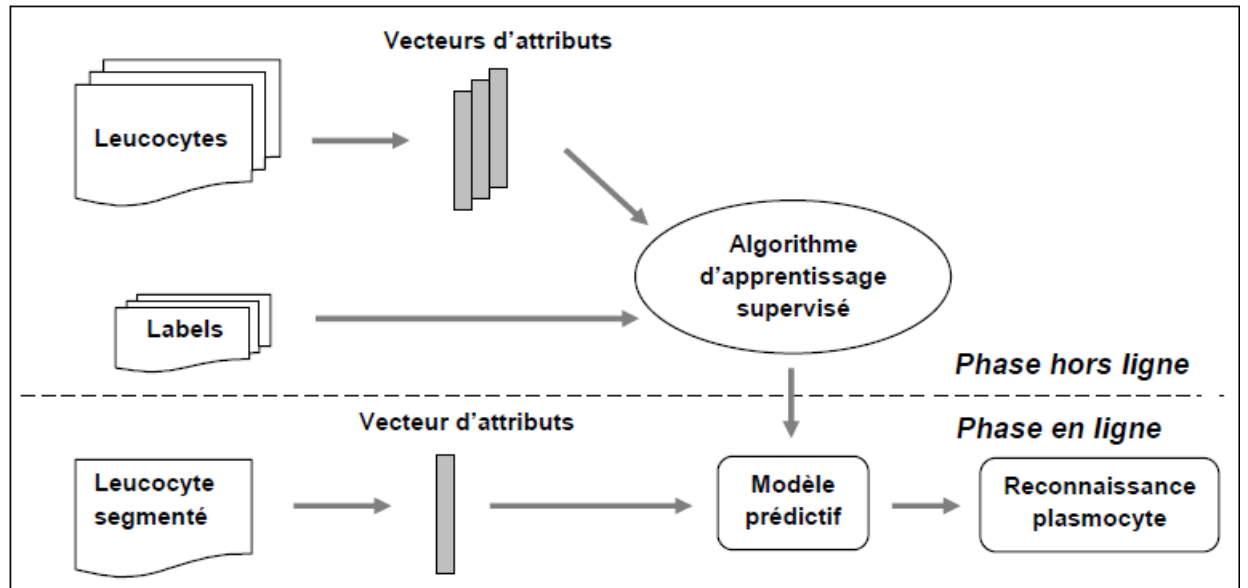


FIGURE 4.5 – Processus de reconnaissance des plasmocytes

Notons, que nous nous sommes intéressés à trois algorithmes d'apprentissage : Support Vector Machine (SVM) [62], K-Plus-Proche-Voisins (KPPV), et arbre de décision. Nous justifions ce choix par l'efficacité reconnu des SVMs

pour la classification binaire, et la simplicité combiné à la robustesse du Kppv et l'arbre de décision de type CART.

4.3 Évaluation du modèle de reconnaissance

Nous avons validé notre modèle proposé sur une base de 87 images de la moelle osseuse. Nous avons collecté nous même les données de cette base au service d'hémobiologie du Centre Hospitalo-Universitaire de Tlemcen. Notons que les images proviennent de quatre patients souffrant du Myélome. Les lames de la moelle osseuse ont été fixées par la coloration MGG (May-Grunwald Giemsa), et ont été introduites au système LEICA (microscope avec caméra) sous un zoom de 1000x. Cette acquisition nous a permis d'obtenir des images RGB de 24 bits avec une résolution de 1024×768 pixels. Nous allons, dans les sections suivantes, présenter et discuter les résultats obtenus de la segmentation (section 4.3.1) et de la classification (section 4.3.2).

4.3.1 Résultats de la segmentation

Nous illustrons les résultats de segmentation obtenus sur 8 images différentes de la moelle osseuse (voir figure 4.6 (a)-1 à (a)-8).

Dans cette étape, le but étant de différencier les globules rouges des globules blancs. Pour ensuite traiter les globules blancs, étant donné qu'ils sont des cellules clés pour l'identification du cancer du sang, plus spécifiquement, nous nous sommes intéressé au Myélome. Et par conséquent, nous ne traiterons pas le cas des globules rouges chevauchés.

Les globules blancs sont présents sous plusieurs aspects, la différence réside dans la forme du noyau et la couleur (et/ou texture) du cytoplasme. Dans la figure 4.6, la forme du noyau est circulaire ou ovale dans les images (a)-1, en bas du (a)-2, (a)-4, en haut et milieu du (a)-5, (a)-6, (a)-7, (a)-8, et de forme d'haricot en haut du (a)-2 et dans (a)-3. Concernant la couleur du cytoplasme, nous la retrouvons foncé et/ou clair sur l'ensemble des images, par exemple, nous retrouvons les deux aspects dans l'image (a)-2.

Les résultats de segmentation obtenus sont satisfaisants, étant donné, qu'ils se rapprochent des vérités terrain. En effet, les cellules ont été toutes bien segmentées dans les images (b)-1, (b)-3, (b)-4 et (b)-6. L'algorithme proposé est bien adapté pour le cas du chevauchement entre globules rouges et globules

blancs, cependant, dans l'image (b)-5 le globule blanc du milieu est mal reconnu. Par ailleurs, nous avons observé quelques limites : une petite partie du noyau a été identifiée comme cytoplasme dans les images (b)-2 (le globule blanc d'en haut) et (b)-7 (celui sur la droite), et vice-versa sur la gauche de l'image (b)-8, une partie du cytoplasme a été reconnue comme noyau. Malgré ces limites, cela n'affecte pas l'identification du globule blanc comme cellule complète.

Notons que le plasmocyte, sous sa forme normale et dystrophique, est présent dans les images : en bas du (a)-2, (a)-4, milieu de (a)-5, en bas de (a)-6, (a)-7 et (a)-8. Fort heureusement, les deux formes ont été bien reconnues, ce qui est encourageant pour la réussite du système.

Pour mesurer qualitativement les performances de notre algorithme de segmentation proposé, il est impératif de faire recours aux métriques de performances afin de déterminer la similarité entre la vérité terrain et le résultat obtenu. Par conséquent, nous avons choisit le score de Dice [51] et la similarité de Jaccard [48]. Ces deux métriques, ont été développées pour mesurer les similarités dans les études écologiques, cependant grâce à leurs natures d'opérations sur les ensembles, elles sont devenues applicables pour mesurer la similarité des segmentations.

Le score de Dice est un indice de chevauchement spatial et une validation de reproductibilité, ce score est également connu sous le nom de chevauchement moyen. Pour ce qui est de la similarité de Jaccard, c'est le rapport de l'intersection de la zone segmentée A et la zone de vérité terrain B, à l'union de la zone segmentée A et la zone de vérité terrain B, et donc, il prend en compte l'aspect qualitatif du chevauchement. Ces deux métriques se calculent comme suit :

$$Score\ de\ Dice = 2 * \frac{TP}{((TP + FP) + (TP + FN))} \quad (4.2)$$

$$Similarit\ de\ Jaccard = \frac{TP}{(FP + TP + FN)} \quad (4.3)$$

Avec, TP (True Positives) : intersection entre la segmentation obtenue et la vérité terrain, FP (false positives) : les parties segmentées qui ne se chevauchent pas avec la vérité terrain, FN (false negatives) : les parties manquantes de la vérité terrain, et TN (true negatives) : partie de l'image au-delà de l'union entre la segmentation et la vérité terrain.

Les valeurs des deux métriques, varient de 0, indiquant un non cheva-

chement entre les deux ensembles de segmentations, à 1, indiquant un chevauchement total. Nous avons calculé ces deux métriques sur l'ensemble des résultats, et nous avons enregistré les valeurs présentées dans le tableau 4.1, le chevauchement spatial est à 93% avec une qualité du chevauchement égale à 87%.

TABLE 4.1 – Métriques de performances

	Score de Dice	Similarité de Jaccard
Valeurs	0.93	0.87

La figure 4.7 attire l'attention sur l'analyse de la qualité de la segmentation par rapport à la variation du paramètre $\gamma_i = d_i^{-1}$ défini dans l'équation 2.9, où d_i est la distance moyenne entre les exemples de la classe i . Nous avons observé que l'ajustement de ce paramètre, est plus important dans le cas du canal magenta, que dans le cas de l'espace HSV. En effet, la projection HSV a été réalisée juste pour la vérification de la classification des globules rouges. L'illustration de l'analyse est présentée dans la figure 4.7 ((c) à (k)); tous les cas sont rapportés. Un résumé des cas est donné ci-dessous :

- Cas (c) à cas (e) le γ_i du RBC sur le canal Magenta est égale à 10×10^{-3} et le γ_i du WBC est supérieur à 4.8×10^{-3} , égale à 4.8×10^{-3} et inférieur à 48×10^{-4} , respectivement.
- Cas (f) à cas (h), le γ_i du WBC sur le canal Magenta est fixé sur la valeur retenue 4.8×10^{-3} et le γ_i du RBC est supérieur à 16×10^{-3} , entre 11×10^{-3} et 15×10^{-3} et inférieur à 9×10^{-3} , respectivement.
- Cas (i) à cas (k) le γ_i du RBC sur l'espace HSV est supérieur à 10^{-1} , égale à 9×10^{-2} et inférieur à 4×10^{-2} , respectivement.

Nous observons sur la figure 4.7, les mauvaises segmentations sur les cas (c), (e), (f), (g), (h), (i) et (k) (plus importante sur le canal magenta que sur l'espace HSV). Pour encore mieux mettre en évidence l'impact de l'ajustement, nous avons calculé le taux de TP et la précision $((TP+TN)/(TP+TN+FP+FN))$ des données des WBC et des données des RBC pour tous les cas rapportés (figure 4.8). Comme nous avons pu constater, sur les cas (c), (e), (f), (g), (h), (i) et (k) une bonne précision est au détriment d'un mauvais taux des TP, i.e une bonne reconnaissance mais avec une précision basse et vice versa.

Cependant le cas (g) représente un bon compromis.

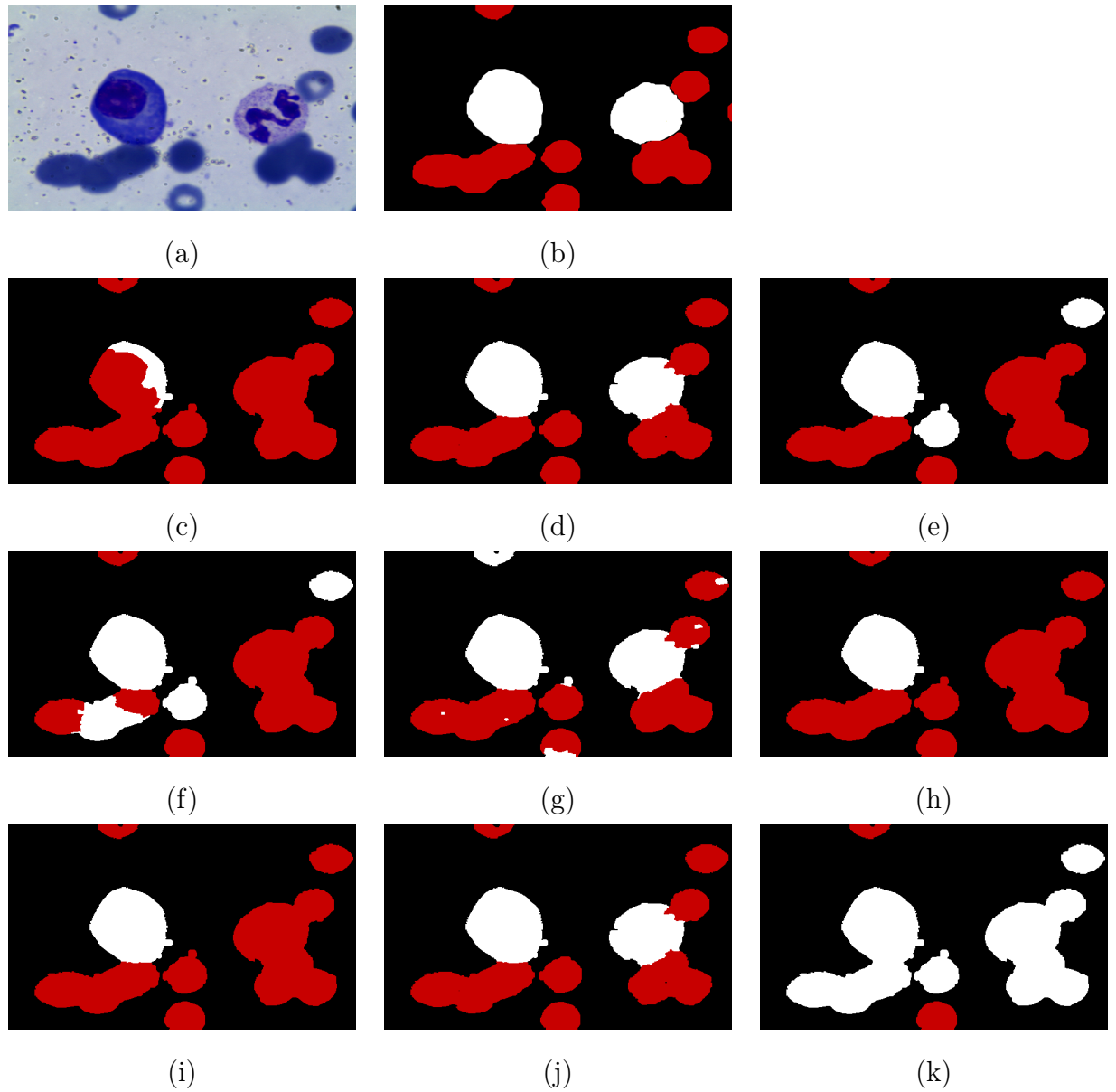
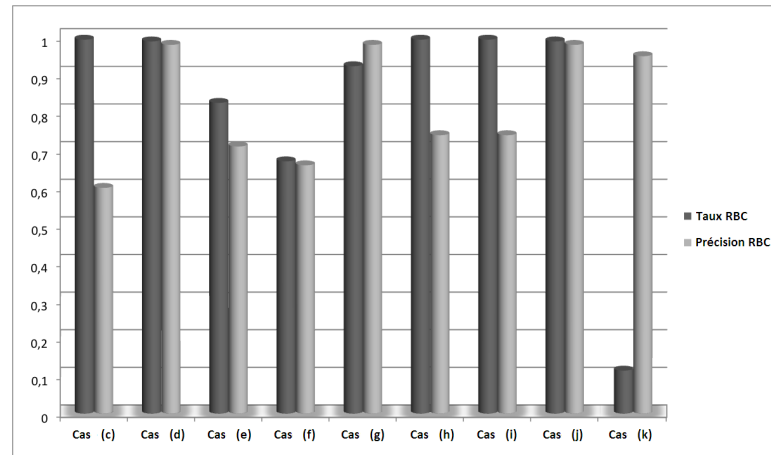
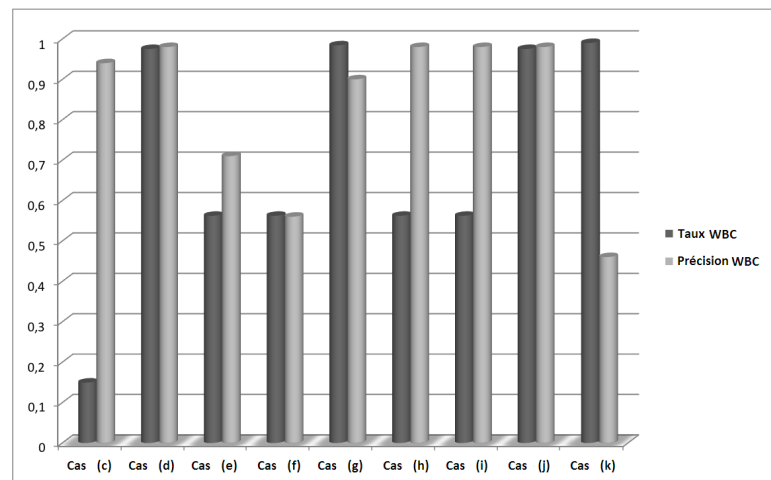


FIGURE 4.7 – Illustration de la sensibilité par rapport aux paramètres γ_i des RBC et γ_i des WBC : (a) image originale, (b) image vérité terrain, (c) à (k) résultats de segmentation pour différentes valeurs de γ_i de RBC, et γ_i de WBC.



(a)



(b)

FIGURE 4.8 – Taux et précisions : (a) Données RBC (b) Données WBC

En tenant en compte la crucialité des données du canal magenta, sur les résultats de segmentation, nous avons analysé toutes les valeurs possible des degrés de croyance. Nous avons alors calculé les degrés de croyances pour les deux ensembles de données (WBC et RBC) en respectant les valeurs possible de la région traitée. Les résultats obtenus sont rapportés sur la figure 4.9, nous avons constaté que les degrés de croyances des RBC décroient plus rapidement que ceux des WBC, cela est expliqué par le fait que l'apparence du noyau se ressemble sur l'ensemble des images.

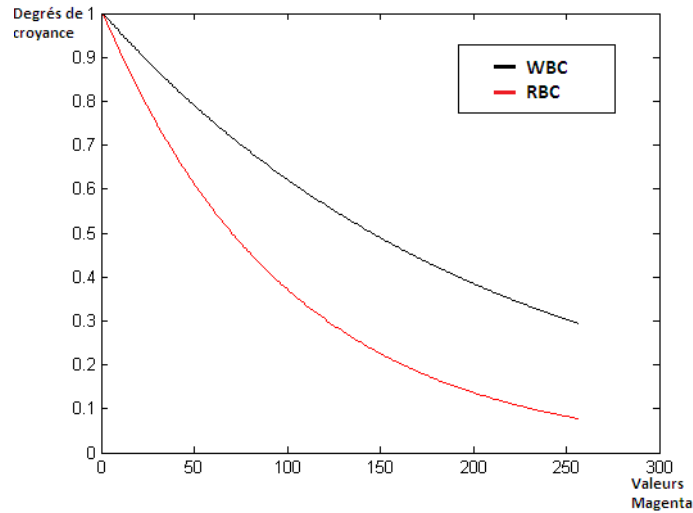


FIGURE 4.9 – L’impact des distances sur les degrés de croyance

Nous avons mentionné auparavant, qu’ils existaient plusieurs méthodes pour traiter le problème de la sur-segmentation du watershed. Et dans l’optique de démontrer l’efficacité de notre proposition, nous avons alors comparé notre méthode avec deux autres techniques : stochastic watershed ([9]) et controlled watershed ([35]). Nous présentons dans la figure 4.2, la comparaison en termes de détection de contours des trois méthodes, illustrée sur quatre images de la moelle osseuse ((a)-1 à (a)-4). Les contours obtenus du controlled watershed, du stochastic watershed, de la segmentation ESA et la vérité terrain sont superposés sur les images originales de la figure 4.2 sur (b), (c), (d) et (e), respectivement.

Nous insistons sur le fait que les images de la figure 4.2(a)-2 et figure 4.2 (a)-4, sont difficiles à segmenter. Celà est dû au grand chevauchement des globules rouges, ce qui mène à une confusion avec le noyau des globules blancs.

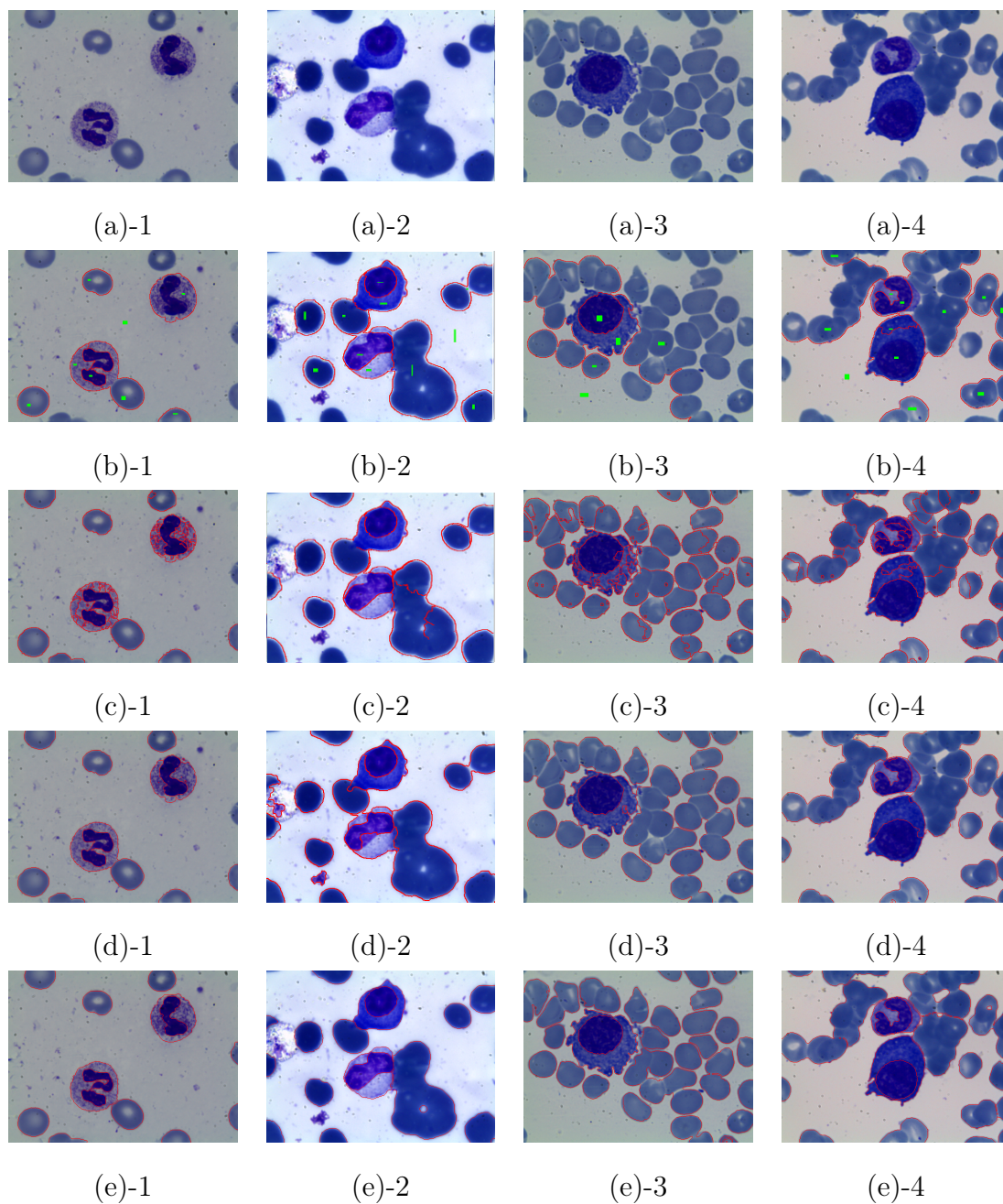


FIGURE 4.10 – Comparaison : (a) Image originale (b) Controlled watershed (vert : marqueurs) (c) Stochastic watershed (d) Segmentation ESA (e) Vérité terrain

Afin d'obtenir une bonne segmentation avec le controlled watershed, nous devons marquer chaque région manuellement, et parfois à plusieurs reprises : dans la figure 4.2 (b)-1, les deux WBCs sont segmentés différemment ; celui d'en haut (un marqueur) est segmenté comme une région, i.e que le cytoplasme et

le noyau ne sont pas séparés, alors que le noyau du WBC d'en bas (deux marqueurs) est bien segmenté, contrairement au cytoplasme, où sa partie droite a été agrégée avec le globule rouge d'en dessous. Par conséquent, plusieurs cellules sont ignorées parcequ'il n'y avait pas de marqueurs au dessus (montré clairement sur la figure 4.2 (b)-1 à (b)-4), Ainsi, la qualité de la segmentation du controlled watershed est étroitement liée à la position et le nombre des marqueurs. Concernant le stochastic watershed (figure 4.2 (c)-1 à (c)-4), et malgré que ce ne fut pas notre objectif, nous avons observé qu'il se comporte bien pour la séparation des globules rouges chevauchés, spécialement sur l'image (c)-3. Malheureusement, il sur-segmente le cytoplasme (voir les deux sur l'image (c)-1)). Pour ce qui est de la segmentation ESA (figure 4.2 (d)-1 à (d)-4), nous notons que les contours ont été bien détectés, Malgré un dépassement entre le noyau et le cytoplasme sur le WBC en bas de l'image (d)-2, cependant, il est important de mentionner que le grand chevauchement des RBC de l'image (a)-(2) et l'image (a)-4, n'a pas aboutit à la confusion avec le noyau, de plus, il a bien été détecté comme étant RBC, ce qui est encourageant pour la suite des traitement et les travaux futures.

La performance en terme de détection de contours a été évaluée par une comparaison avec le controlled watershed et le stochastic watershed. De plus, d'un côté, pour obtenir une bonne segmentation avec le controlled watershed, il s'avère nécessaire de marquer chaque région manuellement, et parfois à plusieurs reprises, et donc, cette technique ne peut pas être sans intervention, et dépend aussi de l'expertise du praticien. De l'autre côté, pour le stochastic watershed, nous avons opté pour 20 marqueurs sur 20 itérations, et après, nous retenons les contours redondants pour terminer la segmentation. Les vrais contours sont alors renforcés et les faux contours sont affaiblis. Cependant, le nombre d'itérations et le nombre de marqueurs ont des conséquences direct sur les résultats, et malgré que le temps d'exécution des deux méthodes est similaire avec notre paramétrage, augmenter le nombre de marqueurs et/ou le nombre d'itérations va augmenter considérablement le temps d'exécution du stochastic watershed.

Nous avons comparé notre approche avec une technique que nous avons proposé précédemment (SDR : Segmentation based on Dimensionality Reduction) [44], où nous avons juste 27 images. Maintenant, l'approche ESA a été testée sur 87 images, et alors nous l'avons comparé à la SDR, en mesurant le score de Dice et la similarité de Jaccard. Dans le tableau 4.2, les valeurs

démontrent l'efficacité de la segmentation ESA. Nous insistons sur le fait que les deux méthodes exploitent que les informations couleurs. De plus, la segmentation ESA est assez rapide, le traitement en entier dure 5 secondes, alors que la segmentation SDR prend 138 secondes.

TABLE 4.2 – Comparaison des segmentations SDR et ESA

	ESA	SDR
Score de Dice	0.93	0.74
Similarité de Jaccard	0.87	0.59

Les résultats obtenus nous ont démontré que la théorie de l'évidence est bien adapté pour la segmentation des cellules sanguines. en effet, le tableau 4.2 met en avant sa force, i.e, sa capacité d'intégrer l'incertitude liée à l'imprécision, l'incomplétude, et la complémentarité des informations. De plus, les résultats de segmentation sont prometteurs et se rapprochent de la segmentation de l'expert (vérité terrain). Cependant, quelques limites ont été observées : la classification de quelques cytoplasmes comme étant RBCs, principalement dû aux différentes classes de WBCs, contrairement aux RBCs qui ont la même apparence.

Dans notre approche, nous nous sommes focalisés sur la distinction basée sur l'incertitude colorimétrique. À cette fin, la théorie de l'évidence a été utilisée sur le canal magenta et l'espace HSV, afin de distinguer les globules rouges des globules blancs. Par la suite, les globules blancs sont segmentés en noyau et cytoplasme, en prenant l'incertitude sur le canal vert. Les globules blancs segmentés de cette étape, cruciale, sont directement introduits aux étapes suivantes (section 4.3.2) qui sont la caractérisation et la reconnaissance des plasmocytes.

4.3.2 Résultats de la caractérisation et la classification

Dans la phase de caractérisation du globule blanc de notre approche, nous avons procédé à l'extraction des attributs de forme de la cellule entière et du noyau, ainsi que les attributs de texture du noyau et du cytoplasme. Cette extraction nous a permis d'obtenir un vecteur de 31 attributs. Une partie de

l'ensemble des données a été utilisé pour l'apprentissage de l'identification des plasmocytes à l'aide du SVM, Kppv et l'arbre de décision. Les classifieurs, ont été ensuite évalué par la technique de la k-validation croisée, avec différentes valeurs de k ; $k = \{5, 7, 10, 15\}$. Après la validation, les trois classifieurs ont été testés sur tout l'ensemble des données.

Nous avons opté pour les métriques d'évaluations suivantes :

- Taux de reconnaissance= $TP/\text{nombre totale}$
- Sensibilité= $TP/(TP+FN)$
- Specificité= $TN/(TN+FP)$
- Précision= $TP/(TP+FP)$
- F-score= $2*((\text{précision}*\text{sensibilité})/(\text{précision}+\text{sensibilité}))$

Où TP (true positive) : le nombre de cellules qui sont plasmocytes et ont été détectées plasmocytes, FN (false negative) : nombre des cellules qui sont plasmocytes mais ont été détectées comme non plasmocytes, TN (true negative) : nombre de cellules qui sont non plasmocytes et ont été détectées comme non plasmocytes, et FP (false positive) : le nombre de cellules qui sont non plasmocytes et ont été détectées comme plasmocytes. Les mesures choisies pour l'évaluation des classifieurs se justifient par le fait que :

- La sensibilité est la probabilité que la cellule va être détectée comme plasmocyte dans le cas où elle l'est réellement.
- La Specificité est la probabilité que la cellule va être détectée comme non plasmocyte dans le cas où elle est non plasmocyte.
- La précision est la probabilité que la détection actuelle du plasmocyte est associé à un vrai plasmocyte.
- Le F-score est la balance entre la précision et la sensibilité (précision Vs robustesse).

Nous avons observé, (voir tableau 4.3, tableau 4.4, tableau 4.5 et tableau 4.6), que les performances sont satisfaisantes pour tous les classifieurs. En effet, les taux de reconnaissances sont proche de 1 et par conséquent un taux d'erreur proche de 0. D'autre part, les 3 classifieurs ont une précision et une spécificité élevées, quand nous rajoutons à cela un F-score élevé, cela rend les classifieurs à la fois précis et robustes. Et par conséquent nos classifieurs sont considéré comme étant assez bon car ils ont un taux d'erreur faible, et d'autre part, une précision et une spécificité élevée, avec un bon compromis entre la précision et la robustesse. Cependant, le classifieur Kppv présente les résultats les plus intéressants où la spécificité et la précision sont égales à 1. Et donc il a la

capacité d'éviter les fausses alarmes, et une bonne détection des cas positifs.

TABLE 4.3 – Performances de SVM, Kppv et arbre de décision pour k=5

	SVM	Kppv	Arbre de décision
Taux de reconnaissance	0.92	0.95	0.95
Sensibilité	0.92	0.92	0.94
Spécificité	0.93	0.97	0.95
Précision	0.92	0.97	0.94
F-score	0.92	0.97	0.94

TABLE 4.4 – Performances de SVM, Kppv et arbre de décision pour k=7

	SVM	Kppv	Arbre de décision
Taux de reconnaissance	0.94	0.96	0.92
Sensibilité	0.92	0.92	0.89
Spécificité	0.95	1	0.95
Précision	0.94	1	0.94
F-score	0.93	0.96	0.92

TABLE 4.5 – Performances de SVM, Kppv et arbre de décision pour k=10

	SVM	Kppv	Arbre de décision
Taux de reconnaissance	0.94	0.97	0.94
Sensibilité	0.92	0.94	0.89
Spécificité	0.95	1	0.97
Précision	0.94	1	0.97
F-score	0.93	0.97	0.93

TABLE 4.6 – Performances de SVM, Kppv et arbre de décision pour k=15

	SVM	Kppv	Arbre de décision
Taux de reconnaissance	0.94	0.97	0.95
Sensibilité	0.92	0.94	0.94
Spécificité	0.95	1	0.95
Précision	0.94	1	0.94
F-score	0.93	0.97	0.94

Pour évaluer au mieux notre technique de reconnaissance de plasmocyte, nous avons fait appel à la courbe de ROC (Receiver Operating Characteristics), l'évaluation est illustrée sur la figure 4.11). Notons que le choix de la courbe ROC n'est pas anodin mais lié au fait qu'il est connu pour sa non sensibilité à la distribution des données, et par conséquent, il délivre une analyse objective du classifieur. De plus, la AUC (Aire Under the Curve; la surface sous la courbe) permet de répondre sur son caractère discriminant entre les classes, quand c'est égale à 1 le classifieur est parfait et quand c'est égale à 0.5 il est non discriminant. Pour notre cas, les AUCs sont reportés sur le tableau 4.7 Étant donné que ces valeurs se rapprochent de 1, les résultats obtenus sont prometteurs. La courbe de ROC est aussi connu pour son pouvoir de comparaison entre les méthodes, dans notre cas, nous remarquons une légère supériorité du classifieur Kppv par rapport au SVM et l'arbre de décision.

TABLE 4.7 – AUCs pour SVM, Kppv et arbre de décision

	SVM	Kppv	Arbre de décision
AUC	0.94	0.97	0.93

Dans l'état de l'art, les travaux sur les frottis de sang périphérique sont beaucoup plus importants que sur la moelle osseuse. En comparaison avec notre méthode, malgré la différence sur la base de données traitée et/ou le globule blanc ciblé, nous notons que sur le même type d'image (moelle osseuse), les auteurs dans [24] ont atteint 78.51% de taux de reconnaissance avec le

classifieur Kppv, sur des image au niveau de gris. Et pour la reconnaissance des plasmocytes, les auteurs dans [55] une sensibilité de 96.52%, une spécificité de 93.04% et une précision de 95.28% avec le classifieur SVM.

En considérant toutes les performances atteintes, et sachant que la classification prend 0.45 secondes, nous estimons que le Kppv, et malgré sa simplicité, a le potentiel de bien identifier les plasmocytes ($AUC=0.97$) et surtout éviter les fausses alarmes (spécificité=1). Cette capacité est d'un grand intérêt pour le processus de diagnostic, vu que le pourcentage des plasmocytes par rapport aux autres globules blancs informe sur la présence ou non du cancer du myélome.

4.4 Conclusion

Nous avons proposé un modèle automatique complet pour la reconnaissance des plasmocytes, considérées comme des cellules clés pour le diagnostic du cancer du Myélome. Notre système se déroule en trois phases essentielles : la segmentation des cellules sanguines, la caractérisation des globules blancs et la classification des globules blancs en vue de reconnaître les plasmocytes.

La phase de segmentation est l'étape la plus cruciale, car sa réussite (détection de toutes les cellules présentes) engendre le succès de tout le système. Pour notre part, la technique de segmentation que nous avons proposé nous a donné des résultats convaincants et prometteurs. Cette technique basée sur l'hybridation de la transformation watershed et la théorie de l'évidence, se positionne dans l'état de l'art de la segmentation d'images dans le traitement du problème de la sur-segmentation issue de la transformation watershed, qui est un domaine très actif. Néanmoins, la technique a présenté quelques limites, comme les confusions entre cytoplasme et globule rouge, et qui nécessitent des améliorations au niveau des caractéristiques traitées, i.e. ajouter d'autres attributs en plus de la couleur. Pour ce qui est de la suite du traitement, nous avons proposé une caractérisation qui additionne les techniques de caractérisation avec les formulations de l'expert, cela nous a conduit à l'obtention d'un vecteur d'attributs de 31 valeurs. Ces attributs ont été présentés directement aux différents classifieurs afin de procéder à l'apprentissage pour la reconnaissance automatique des plasmocytes. L'évaluation des classifieurs nous a permis de reconnaître l'efficacité du classifieur Kppv pour la reconnaissance des plasmocytes.

Pour finir, notre approche présente dans l'ensemble l'avantage d'être efficace et rapide, en effet le modèle proposé prend 5.45 secondes pour traiter une image ; 5 pour la segmentation et 0.45 secondes pour la classification. Compte tenu de cette performance, nous considérons que c'est un atout convaincant pour le cyto-technicien.

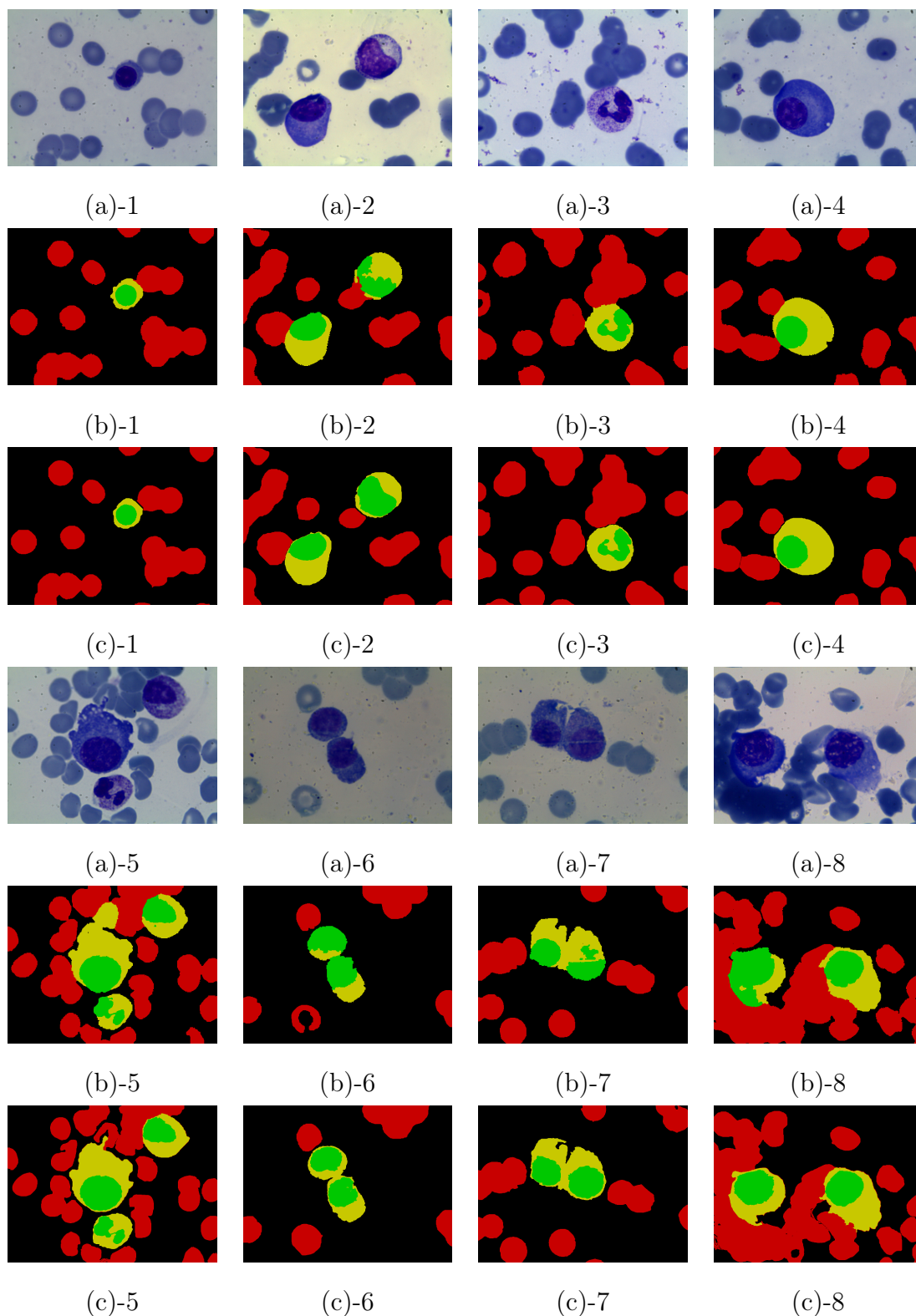
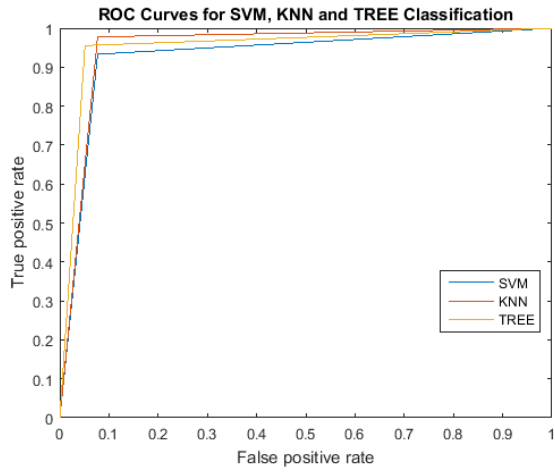
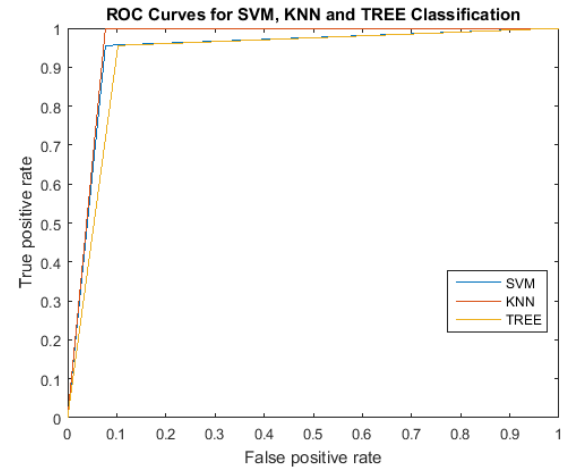


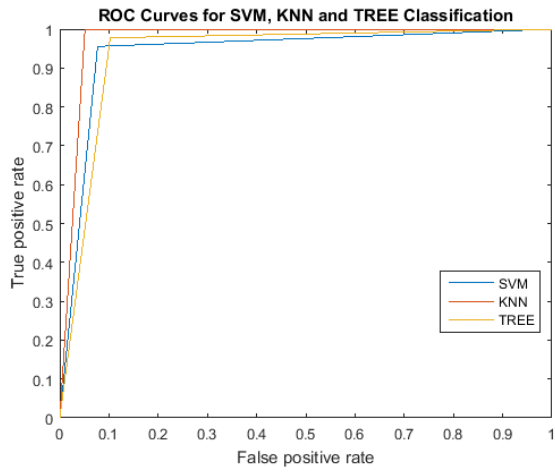
FIGURE 4.6 – Résultats ESA : (a) image originale (b) segmentation ESA (c) Vérité terrain



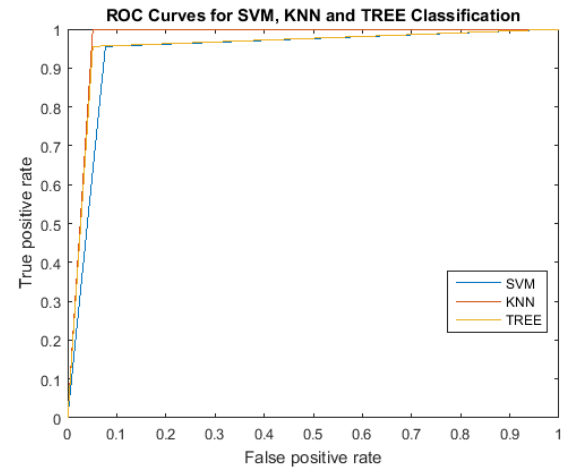
(a)



(b)



(c)



(d)

FIGURE 4.11 – Courbes ROC : (a) pour $k=5$ (b) pour $k=7$ (c) pour $k=10$ (d) pour $k=15$

Conclusion générale et perspectives

L'analyse du sang est importante pour connaître l'état de santé d'un patient (sang périphérique), cependant pour des investigations approfondies le recours à la biopsie est nécessaire. Les patients atteints de la maladie de Myélome, sont diagnostiqués après la biopsie médullaire. Les images obtenues de cette technique sont visualisées manuellement, d'où le risque d'erreur, i.e, la génération des faux positifs avec un temps important pour établir un diagnostic. Nous nous sommes positionnées dans la problématique d'automatiser cette analyse visuelle des images de la moelle osseuse, afin d'aider le médecin dans son diagnostic. Notre approche comporte trois phases principales : l'acquisition des images, la segmentation des cellules et la classification des globules blancs.

L'acquisition des images microscopiques est réalisée à partir de l'observation par un microscope, lié à un ordinateur, des lames préparées à partir d'une goutte de moelle osseuse retirée par ponction médullaire.

Les images microscopiques sont automatiquement segmentées par notre technique ESA (segmentation évidentielle), nous proposons d'utiliser des informations issues de différentes sources : l'énergie watershed, et les informations couleurs provenant de différents espaces couleurs, afin de segmenter les images microscopiques de la moelle osseuse. Dans les images cytologiques, le nombre des globules rouges est beaucoup plus important que celui des globules blancs, par exemple, une image peut contenir cent globules rouges pour seulement un à trois globules blancs, ce qui mène à un chevauchement important des globules rouges. De plus, la qualité de l'image dépend de la fixation et la qualité du microscope. Notre approche de segmentation a permis de fusionner les différentes informations hiérarchiquement, c'est pour cette raison qu'à chaque étape la segmentation est optimisée jusqu'à l'obtention de la segmentation finale. Dans notre approche hiérarchique, nous avons fait appel à la transformation watre-

shed, connue par sa simplicité et sa rapidité, cependant cette technique est aussi connue par sa limite majeure qui est la sur-segmentation, nous avons alors proposé de profiter de la segmentation initiale par la transformation watershed, et après nous procédons à une agrégation des régions en intégrant l'incertitude liée à la couleur des régions watersheds. Cette approche nous a permis d'obtenir de bons résultats, ce qui nous a conduit à la troisième phase de notre modèle, qui est la reconnaissance des plasmocytes.

La reconnaissance des plasmocytes est la partie visible par le médecin dans tout le modèle, sa qualité dépend bien sûr de la qualité de segmentation, cependant la qualité discriminatoire du classifieur automatique est aussi importante. Dans notre approche, trois différents classifieurs sont entraînés sur un ensemble d'attributs labellisés. Ces attributs sont obtenus par la caractérisation des globules blancs segmentés auparavant. Nous avons basé la caractérisation selon des aspects techniques mais aussi médicales, pour pouvoir donner une description interprétable du plasmocyte. Nous nous sommes alors intéressés aux attributs de formes de la cellule et du noyau, mais aussi aux attributs de texture couleur du noyau et du cytoplasme calculés sur le canal vert et le canal magenta. Les trois classifieurs SVM, Kppv et l'arbre de décision nous ont conduit à un bon taux de reconnaissance, avec une meilleure performance pour le Kppv qui s'est bien comporté vis à vis des faux positifs, ce qui est très important dans la fiabilité du système d'aide au diagnostic.

Notre technique de segmentation, peut être améliorée au niveau de la transformation watershed, appliquer des marqueurs fiables pour détecter toutes les cellules sans avoir recours à l'agrégation, cela peut devenir compliqué dû à la grande variabilité d'aspect et de couleur des différents globules blancs. Cependant, l'introduction de l'incertitude sur les régions watersheds sur-segmentées est possible, il suffit d'adopter d'autres hypothèses avant la fusion, ou alors faire appel à une autre méthode d'intégration d'incertitude qui est la théorie des possibilités.

Les images sont acquises à travers un ajustement fixe, pour que les classifieurs donnent aussi sur d'autres zooms, les attributs de caractérisation doivent être remplacés par des attributs obtenus par les techniques dites invariantes au zoom, à la rotation et bien d'autres, par exemple la méthode SURF.

L'automatisation de l'analyse des lames, reste une actualité, comme c'est le cas de l'étude dans [56] d'une éventualité d'analyse automatique du sang, par mobile dans les villes intelligentes (smart cities).

Bibliographie

- [1] Université médicale virtuelle francophone, 2010. Support de cours consulté en Février 2017.
- [2] Infocancer, 2016. <http://www.arcagy.org/infocancer/localisations/hemopathies-malignes-cancers-du-sang/leucemies-aigues/maladie/les-quatre-compartiments.html>.
- [3] Cell atlas, 2017. <http://www.cellavision.com/en/cellavision-cellatlas/hematopoiesis-cellatlas>.
- [4] Appriou A. Multisensor signal processing in the framework of the theory of evidence. *NATO/RTO Ő Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, page 5Ő31, 1999.
- [5] Rakar A., Juricic D., and Ballé P. Transferable belief model in fault diagnosis. *Engineering Applications of Artificial Intelligence*, 12 :555Ő567, 1999.
- [6] Rakibull Ahasan, Ahasan Ulla Ratul, and ASM Bakibillah. White blood cells nucleus segmentation from microscopic images of strained peripheral blood film during leukemia and normal condition. In *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on*, pages 361–366. IEEE, 2016.
- [7] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2014.
- [8] Koschan Andreas and Abidi Mongi. Digital color image processing. *John Wiley and Sons, Inc.*, 2008.
- [9] J. Angulo and D. Jeulin. Stochastic watershed segmentation. *Proc. of the 8th International Symposium on Mathematical Morphology (ISMM'2007)*, pages 265–276, october 2007.
- [10] Ismahan BAGHLI, Mourtada BENAZZOUZ, and Mohamed Amine CHIKH. Plasma cell identification based on evidential segmentation and

- supervised learning. *In press, to appear in International Journal of Bio-medical Engineering and Technology*, 2017.
- [11] Ismahan Baghli, Amir Nakib, Elie Sellam, Mourtada Benazzouz, Amine Chikh, and Eric Petit. Hybrid framework based on evidence theory for blood cell image segmentation. *In proc. SPIE 9038, Medical Imaging 2014, San Diego, (USA)*, Feb. 15-19, 2014, <http://dx.doi.org/10.1117/12.2042142>.
- [12] Mourtada Benazzouz. *analyse intelligente des images médicales : application aux images microscopiques de cytologie*. PhD thesis, 2014.
- [13] Dominique Bordessoule. CHU de limoges, 2013. Livret d’information du patient.
- [14] A Bouatay, S Hizem, Y Ben Youssef, F Sayari, N Braham, A Khélif, and M Kortas. Myélome multiple : aspect clinique, diagnostic biologique et pronostic. *Immuno-analyse et Biologie Spécialisée*, 28(1) :30–35, 2013.
- [15] Mohamed BOUMHRAS. Evaluation de la toxicité des moules (*mytilus galloprovincialis*) issues de jorf lasfar (jl) et oualidia (ol) sur la moelle osseuse chez le rat. 2008. Université Hassan 1er, Faculté des Sciences et Techniques de Settat, DESA 2008.
- [16] Francesco Camastra, Alessandro Vinciarelli, and Jie Yu. Machine learning for audio, image and video analysis. *J. Electronic Imaging*, 18(2) :029901, 2009.
- [17] S. Ben Chaabane, M. Sayadi, F. Fnaiech, and E. Brassart. Dempster-shafer evidence theory for image segmentation : Application in cells images. *International Journal of Information and Communication Engineering*, 5(2) :126–132, 2009.
- [18] Chi-hau Chen. *Handbook of pattern recognition and computer vision*. World Scientific, 2015.
- [19] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, 20(3) :273–297, 1995.
- [20] Camille Couprie, Leo Grady, Laurent Najman, and Hugues Talbot. Power watershed : A unifying graph-based optimization framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7) :1384–1399, 2011.

- [21] Thierry Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.
- [22] Gloria Díaz and Antoine Manzanera. Automatic analysis of microscopic images in hematological cytology applications. *Biomedical Image Analysis and Machine Learning Technologies : Applications and Techniques : Applications and Techniques*, 167, 2009.
- [23] Leyza Baldo Dorini, Rodrigo Minetto, and Neucimar Jerônimo Leite. White blood cell segmentation using morphological operators and scale-space analysis. *in Proc. 20th Brazilian Symposium Computer Graphics Image Processing*, pages 100–107, 2007.
- [24] Leyza Baldo Dorini, Rodrigo Minetto, and Neucimar Jerônimo Leite. Semiautomatic white blood cell segmentation based on multiscale analysis. *IEEE journal of biomedical and health informatics*, 17(1) :250–256, 2013.
- [25] Sadeghian F, Seman Z, Ramli AR, Abdul Kahar BH, and Saripan MI. A framework for white blood cell segmentation in microscopic blood images using digital image processing. *Biological Procedures Online*, 11(1) :196–206, 2009.
- [26] Wei Gao, Yinggan Tang, and Xiaoli Li. Segmentation of microscopic images for counting leukocytes. *Proceeding of The 2nd International Conference on Bio-informatics and Biomedical Engineering (ICBBE'08), Shanghai, China*, pages 2609 – 2612, 2008.
- [27] Madhumala Ghosh, Devkumar Das, Subhodip Mandal, Chandan Chakraborty, Mallika Pal, Ashok K Maity, Surjya K Pal, and Ajoy KRay. Statistical pattern analysis of white blood cell nuclei morphometry. *Proceedings of the 2010 IEEE Students Technology Symposium, IIT Kharagput*, pages 59–66, 3-4 April 2010.
- [28] Pramit Ghosh, Debotosh Bhattacharjee, and Mita Nasipuri. Blood smear analyzer for white blood cell counting : A hybrid microscopic image analyzing technique. *Applied Soft Computing*, 2016.
- [29] Shafer Glenn. A mathematical theory of evidence. *Princeton University Press*, 1976.
- [30] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621, 1973.

- [31] Min Hu, Xijian Ping, and Yihong Ding. Automated cell nucleus segmentation using improved snake. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 4, pages 2737–2740. IEEE, 2004.
- [32] Der-Chen Huang, Kun-Ding Hung, and Yung-Kuan Chan. A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images. *The Journal of Systems and Software*, 85 :2104–2118, 2012.
- [33] Guan J. W. and Bell D. A. Evidence theory and its applications. *North-Holland, New York*, 1991.
- [34] C. Kalaiselvi and R. Asokan. A classification of chronic leukaemia using new extension of k-means clustering and efmm based on digital microscopic blood images. *Int. J. of Biomedical Engineering and Technology*, 23(2/3/4) :232–241, 2017.
- [35] Mandeep Kaur and Gagandeep Jindal. Medical image segmentation using marker controlled watershed transformation. *IJCST*, 2(4) :548–551, october-decembre 2011.
- [36] Byoung Chul Ko, Ja-Won Gim, and Jae-Yeal Nam. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron*, 42 :695–705, 2011.
- [37] S. Manier ; X. Leleu. Myélome multipleã : diagnostic clinique et perspective de traitement. recommandations de l'international myeloma working group (imwg). *Immuno-analyse et Biologie Spécialisée*, 26, 2011.
- [38] Olivier Lezoray. Supervised automatic histogram clustering and watershed segmentation. application to microscopic medical color images. *Image Anal Stereol*, 22 :113–120, 2003.
- [39] Zhi Liu, Jing Liu, Xiaoyan Xiao, Hui Yuan, Xiaomei Li, Jun Chang, and Chengyun Zheng. Segmentation of white blood cells through nucleus mark watershed operations and mean shift clustering. *sensors*, 15(9) :22561–22586, 2015.
- [40] Tomasz Markiewicz, Stanislaw Osowski, Bonenza Marianska, and Leszek Moszczynski. Automatic recognition of the blood cells of myelogenous leukemia using svm. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 4, pages 2496–2501. IEEE, 2005.

- [41] Tomasz Markiewicz, Stanisław Osowski, and Bożena Mariańska. White blood cell automatic counting system based on support vector machine. *Adaptive and Natural Computing Algorithms, Springer*, pages 318–326, 2007.
- [42] Mulligan ME and Badros AZ. Pet/ct and mr imaging in myeloma. *Skeletal radiol*, 36 :5–16, 2007.
- [43] E. Meijering. Cell segmentation : 50 years down the road. *IEEE Signal Processing Magazine*, 29(5) :140–145, 2012.
- [44] Benazzouz Mourtada, Baghli Ismahan, and Med. Amine Chikh. Microscopic image segmentation based on pixel classification and dimensionality reduction. *International Journal of Imaging Systems and Technology*, 23 :22–28, March 2013.
- [45] MLim Huey Nee, Mohd Yusoff Mashor, and Rosline Hassan. White blood cell segmentation for acute leukemia bone marrow images. *International Conference on Biomedical Engineering (ICoBE'12), Penang, Malaysia*, pages 357–361, 27-28 February 2012.
- [46] J.B. Nemane and V. A. Chakkarwar. A novel method of white blood cell segmentation and counting. *International Journal on Advanced Computer Engineering and Communication Technology*, 1(1) :44–49, 2012.
- [47] Ramesh Nisha, Dangott Bryan, E. Salama Mohammed, and Tolga Tasdizen. Isolation and two-step classification of normal white blood cells in peripheral blood smears. *Journal of pathology informatics*, 3(1), march 2012.
- [48] Jaccard P. Etude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la société Vaudoise des Sciences Naturelles*, 37 :547–579, 1901.
- [49] Jaroonrut Prinyakupt and Charnchai Pluempitiwiriwawej. Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers. *Biomedical engineering online*, 14(1) :63, 2015.
- [50] P.S.Hiremath, Parashuram Bannigidad, and Sai Geeta. Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. *IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition"*, pages 59–63, 2010.
- [51] Dice L. R. Measures of the amount of ecologic association between species. *Ecology*, 26(3) :297–302, 1945.

- [52] Rajkumar and S Vincent. Multiple myeloma : 2013 update on diagnosis, risk-stratification, and management. *American journal of hematology*, 88(3) :226–235, March 2013.
- [53] Sebastian Raschka. *Python machine learning*. Packt Publishing Ltd, 2015.
- [54] Seyed Hamid Rezafofighi and Hamid Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, 35 :333–343, 2011.
- [55] Z Saeedizadeh, A Mehri Dehnavi, A Talebi, H Rabbani, O Sarrafzadeh, and A Vard. Automatic recognition of myeloma cells in microscopic images using bottleneck algorithm, modified watershed and svm classifier. *Journal of microscopy*, 261(1) :46–56, 2015.
- [56] Muhammad Sajjad, Siraj Khan, Zahoor Jan, Khan Muhammad, Hyeon-joon Moon, Jin Tae Kwak, Seungmin Rho, Sung Wook Baik, and Irfan Mehmood. Leukocytes classification and segmentation in microscopic blood smear : a resource-aware healthcare service in smart cities. *IEEE Access*, 5 :3475–3489, 2017.
- [57] Mukesh Saraswat and KV Arya. Automated microscopic image analysis for leukocytes identification : A survey. *Micron*, 65 :20–33, 2014.
- [58] SS Savkare and SP Narote. Blood cell segmentation from microscopic blood images. In *Information Processing (ICIP), 2015 International Conference on*, pages 502–505. IEEE, 2015.
- [59] J. M. Sharif, M. F. Miswan, M. A. Ngadi, Md Sah Hj Salam, and Muhammad Mahadi bin Abdul Jamil. Red blood cell segmentation using masking and watershed algorithm : A preliminary study. *Proceeding of ICoBE, Penang, Malaysia*, pages 258–262, 27-28 February 2012.
- [60] S.S.Savkare and S.P.Narote. Automatic system for classification of erythrocytes infected with malaria and identification of parasite’s life stage. *Procedia Technology : 2nd International Conference on Communication, Computing and Security (ICCCS’12)*, pages 405 – 410, 2012.
- [61] Hemant Tulsani, Saransh Saxena, and Naveen Yadav. Segmentation using morphological watershed transformation for counting blood cells. *International Journal of Computer Applications and Information Technology*, 2(3) :28–36, April-May 2013.
- [62] Vapnik V. Statistical learning theory. *John Wiley and Sons, Chichester*, 1998.

- [63] AMW Van Marion, HM Lokhorst, and JG Van Den Tweel. Pathology of multiple myeloma. *Current Diagnostic Pathology*, 9(5) :322–327, 2003.
- [64] L. Vincent and P. Soille. Watersheds in digital spaces : an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 13 :583–598, 1991.
- [65] L. Vincent and P. Soille. Watersheds in digital spaces : an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 13 :583–598, 1991.
- [66] Weixing Wang, Hao Song, and Qi Zhao. A modified watersheds image segmentation algorithm for blood cell. *International Conference on Communications, Circuits and Systems Proceedings*, 1 :450 – 454, June 2006.
- [67] Faliu Yi, Inkyu Moon, Bahram Javidi, Daniel Boss, and Pierre Marquet. Automated segmentation of multiple red blood cells with digital holographic microscopy. *Journal of Biomedical Optics*, 18 :DOI : 10.1117/1.JBO.18.2.026006, February 2013.
- [68] Musbah J. Zaitoun, Nida M. ; Aqel. Survey on image segmentation techniques. *Procedia Computer Science*, 65, 2015.

Plasma cell identification based on evidential segmentation and supervised learning

Ismahan Baghli, Mourtada Benazzouz and Med Amine Chikh

Université de Tlemcen, Laboratoire GBM, Tlemcen, Algeria

{i_baghli,m_benazzouz,mea_chikh}@mail.univ-tlemcen.dz

July 10, 2017

Abstract

Myeloma disease is among the most common type of cancer, it is characterized by proliferation of plasma cells, kind of white blood cell (WBC). Early diagnosis of the disease can improve the patient's survival rate. The manual diagnosis involves clinicians to visually examine microscopic bone marrow images for any signs of cells proliferation. This step is often laborious and can be highly subjective due to clinician's expertise. Automatic system based on WBC identification and counting provides more accurate result than manual method. This system is mainly based on three major steps: cell's segmentation, cell's characterization and cell's classification. In the proposed system, microscopic images of bone marrow blood are segmented by the combination of watershed transform and the evidence theory, the segmented cells are characterized with shape and Colour texture features, and then classified into plasma cells or not plasma cells with three supervised classifiers; Support Vector Machines, K Nearest Neighbour and Decision Tree. Experimental results show that the recognition of plasma cells with the K nearest neighbour achieved 97% of correct rate with 100% of specificity.

keywords: Myeloma, Plasma cell, Bone marrow images, Segmentation, Evidence theory, Watershed, Characterization, Shape, Colour texture, Classification.

1 Introduction

The main components of blood include white blood cells (WBC), Red Blood Cells (RBC) and platelets. Each component provides vital functionality: WBC protect the body from infective organisms and foreign substances, RBC transport oxygen and Platelets stop bleeding. Besides, there are different kind of WBC, characterized by different nucleus's shape and cytoplasm's texture; nucleus and cytoplasm are both components of white blood cell.

The blood cancer is a type of disease in which a kind of WBC exhibits irregular cell growth cycle. In the case of myeloma cancer, the plasma cells (kind of WBC) grow uncontrollably. The diagnosis requires 10% or more clonal plasma cells on bone marrow examination or a biopsy proven plasmacytoma plus evidence of associated end-organ damage [1]. Early diagnosis of the disease can improve the patient's survival rate. Figure 1 shows a bone marrow image, where the cell in the top is a plasma cell and the one in the bottom is not a plasma cell.

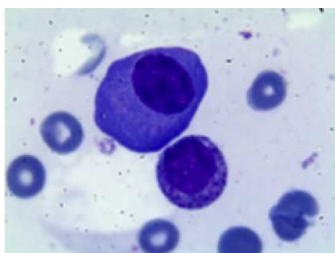


Figure 1: Bone marrow image

Actually, clinicians visually examine microscopic bone marrow images for any signs of plasma cells proliferation. This step is used to be laborious and can be highly subjective due to clinician's expertise. The automatic counting system of plasma cells indicates the total number and the percentage regarding to other WBCs. It is important to develop an image description language for pathologists to aid in reducing the variation in reporting [2]. Therefore, the use of automatic system may render the diagnosis more efficient and reliable. The automatic system is mainly based on three major steps: cell's segmentation, WBC characterization and WBC classification.

The literature review reveals that more works have been done in segmentation process, for blood cell [3] and for WBC [4]. Since, the segmentation is the crucial step for the efficiency of the whole system. Also, a general review of segmentation algorithms used in blood cancer detection [5]. Nevertheless, interesting works have been done in characterization and WBC classification [6].

Here in, we focus on papers dealing with the whole process. Cell's segmentation with global threshold-

ing, cell's characterization by shape features, and cell's classification into three WBC kinds by rules based classification in [7]. A segmentation based on two thresholding, the first to remove RBC and background and the second to separate the detected WBC into nucleus and cytoplasm, and then, shape and texture features are extracted to automatically identify five WBC kinds by LDA (Linear Discriminant Analysis) in [8]. the color blood cell images are segmented by k-means and global threshold, and than blood cells are separated by sobel edge detector and watershed transform [9]. The Gram-Schmidt orthogonalization with a snake algorithm to segment nucleus and cytoplasm, Colour, morphological and texture features are extracted from nucleus and cytoplasm, afterward a SFS (Sequential Forward Selection) is applied to select relevant features, and then compare classification performances of a neural network and SVM (Support Vector Machines) in [10]. The microscopic blood smear are segmented by thresholding, morphology operation and ellipse curve, features from segmented nucleus and cytoplasm are extracted by SFS and introduced to Bayes classifier to recognize five kind of WBC[11]. The colour scale of digital microscope blood image is pre-processed, then the image is segmented by an extension of k-means clustering, and the classification is based on Fuzzy Min Max neural network in [12]. WBC are segmented by color space decomposition and k-means clustering[13]. Nucleus enhancer followed by Otsu threshold and RBC size estimation to segment WBC, PCA (Principal Component Analysis) reduction on texture and shape features and K-means to classify segmented leukocytes into five types, in [14]. Watershed transform and level set method to segment nucleus, granulometric analysis and morphological operations to segment cytoplasm, shape features to characterize WBC, and KNN (K-Nearest Neighbors) classifier to identify five kind of WBC in [15]. Microscopic blood cell are segmented in parallel and than fused, the obtained cells are separated by distance and watershed transform[16]. The WBCs are identified by color mean shift clustering, image enhancement techniques are applied to obtain a complete set of WBCs and the watershed transform is also applied to adress cel adhesion problem[17]. The central areas of RBC are extracted by a mask based on line operator, and than central areas are used as seed points to improve performances of the watershed transform[18]. Application of the watershed algorithm based on map distances (closing and erosion operations on binary image) for the final division of the image into catchment's basins, each corresponding to one cell in [19] and [20]. Textural, geometrical and statistical features were used to identify ten WBC classes by the SVM [20] and recognition of seventeen classes of blood cells of myelogenous leukemia [19]. Cells are segmented by region based segmentation. Afterward, gradient based region growing with neighbourhood influence is applied to restore region boundary. WBCs regions are separately classified on the basis of shape, size, Colour and texture features independently using different fuzzy and non-fuzzy

techniques, in order to recognize five kinds of WBC [21].

To tackle the problem of plasma cells identification, we developed an automated plasma cell's recognition based on blood cell segmentation, WBC characterization and finally, plasma cell's recognition. In the first step, blood cells are segmented by the hybridization of watershed regions and the evidence theory, based on Colour and lighting uncertainties. In the second step, the white blood cells are characterized by colour texture features and shape features. And finally, WBCs are automatically classified into plasma cells or not plasma cells with three supervised classifiers: Support Vector Machines, K Nearest Neighbour and Decision Tree.

The next section describes the details of the proposed system. We show results and discuss the obtained performances in section 3. Finally, we conclude the paper in section 4.

2 Materials and method

The whole process of our proposed method for the automatic recognition of plasma cells is presented in figure 2. The proposed system is explained as follows. Section 2.1 describes the cell's segmentation process based on watershed and evidence theory. WBC characterization is presented in section 2.2. And, the section 2.3 details the classification algorithm.

2.1 Cell's segmentation

A microscopic bone marrow image contains RBC, WBC, platelets and background, often with noise, which makes the segmentation process more difficult. To tackle this problem, we used ESA (Evidential Segmentation Algorithm) segmentation described in [22]. This method is based on a segmentation by the watershed transform, and then, the obtained regions are merged by the evidence theory based on uncertainty of different colour informations.

2.1.1 Watershed segmentation

The watershed transform is a simple and efficient method [23], known also for its speedness. The latter is of great interest for clinician, since it is facing hundreds to thousands of images. Indeed, it is very used in biomedical image analysis [24].

The watershed transform considers minima of the image as the objects of interest and the maxima as the separation boundaries between objects. We obtain an initial segmentation, with the application of the

watershed transform on the gradient image. However, the existence of noises causes an over-segmentation (see figure 3). One can consider the colour descriptors, and hence, it is possible to merge similar regions to deal with the over-segmentation, moreover, hierarchical and fusion process is important to obtain robust and efficiency cell segmentation results[25]. Also, region merging based on color differences were successfully applied on image segmentation[26]. Therefore, we pass to the next step.

2.1.2 Evidence merging

Colour descriptors are among the most important features used in image processing. They provide powerful information that often simplifies object detection and identification. In blood cell images, the RBCs have always the same appearance, however, WBCs have diverse classes such as basophil, eosinophil etc. Those differences mainly concern the shape of nucleus and the colour and/or texture of cytoplasm. Being based on colour differences, we projected our blood cell images on several colour spaces. We noticed that one colour space is insufficient to make all the difference, therefore, three colour spaces kept our interest, which are RGB (Red Green Blue), HSV (Hue Saturation Value) and CMYK (Cyan Magenta Yellow black), to distinguish between RBCs and WBCs (nucleus and cytoplasm).

In figure 4, RBCs are lighter than the WBC on the Magenta channel. While in figure 5, all WBC's cytoplasm are similar to RBC in the Magenta channel, but the difference can be seen clearly when we used the HSV projection. O. Lezoray [27], has shown the ability of the green channel in the separation of the white blood cell into nucleus and cytoplasm.

Yet, different colour representations lead to handle with uncertainty resulting from imprecise, incomplete, and complementary colour informations. Our idea is to deal with the colour uncertainty related to the oversegmentation, with using a data fusion principle based on evidence theory [28]. Evidence theory, is till now, used in image segmentation, and had proven its efficiency within complex scenes [29].

The Dempster-Shafer evidence theory was originally developed by Dempster, who concerned about the lower and upper probabilities, and later Shafer made his contribution by offering belief functions to model uncertain knowledge on the basis of mathematical foundations [30]. It is based on two main ideas: obtaining a degree of belief on a hypothesis (called the "mass function") from subjective probabilities, and Dempster's rule (orthogonal sum) for combining such degrees of beliefs when they are based on independent items of evidence [28]. In a more formal way:

Let Ω be a finite set of mutually exclusive and exhaustive hypotheses, called the frame of discernment:

$$\Omega = \{H_1, \dots, H_n\}$$

The power set $P(\Omega)$ composed with the 2^n propositions of Ω defined by:

$$P(\Omega) = \{\phi, \{H_1\}, \{H_2\}, \dots, \{H_n\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Omega\}$$

When the frame of discernment is determined, the Basic Probability Assignment (BPA) or mass function, is defined as a mapping of the power set $P(\Omega)$ to a number between 0 and 1, i.e.: $m : P(\Omega) \rightarrow [0, 1]$, that satisfies conditions below:

$$\begin{cases} m(\phi) = 0 \\ \sum_{A \subseteq \Omega} m(A) = 1 \end{cases} \quad (1)$$

The combination of the different evidences using Dempster's rule in which two independent information sources (m_1 and m_2) are fused to create a new belief function, is presented below:

$$m(C) = \frac{\sum_{A \cap B = C} m_1(A) \times m_2(B)}{1 - \sum_{A \cap B = \phi} m_1(A) \times m_2(B)} \quad (2)$$

where the denominator can be interpreted as a conflict criterion between independent evidences to be combined [28]. The increase of $m(A)$ corresponds to stronger the belief on the proposition A , we choose the mass as the decision function.

An important key point within the evidence theory framework is to properly model the knowledge given by different sources of information S in order to initialize the associated BPA. Since the definition of the BPA function in evidence theory remains a largely unsolved problem and there is not one general solution, we tested some models proposed by Denoeux [31], Shafer [28], and Appriou [32]. Then, the Denoeux model seems to be suitable for our blood cell images, because the knowledge about each pattern is modeled by distances between samples and class center. The Appriou and Shafer models integrate also the modelisation of scenes with a large variances in the same class (natural scenes), which is not our concern. Thus, the considered is:

$$m(H_i) = \alpha_i \times \exp(-\gamma_i \times \text{dist}(x, x_i)) \quad (3)$$

with $\alpha_i \in [0, 1]$ is the weakening parameter, $\gamma_i > 0$ related to the hypothesis H_i , $\gamma_i = \frac{1}{d_i}$ (with, d_i the average distance between representatives samples of H_i), $\text{dist}(x, x_i)$ the euclidian distance between average values of the processed zone x , and x_i (the average value of a representatives samples of H_i).

WBC (nucleus vs cytoplasm) and RBC cells are segmented after applying Algorithm 1. The microscopic blood image is given as input and it returns a segmented image, as shown in figure 6. (in the segmented

image, the background, RBC, nucleus and cytoplasm are represented by black, red, green and yellow, respectively.)

Algorithm 1 EVIDENCE MERGING

Input: r_k {Watershed regions}

Output: Segmented image: {NUCLEUS}, {CYTOPLASM}, {RBC}

Step1:

for all r_k **do**

-Calculate masses functions with eq 3 on magenta channel

$m(H_i) = \alpha_i \times \exp(-\gamma_i \times dist(x, x_i))$ with $i = \{WBC, RBC\}$

-Combine degrees of beliefs with orthogonal sum eq 2

-Affect the class j (the higher BPA) to r_k i.e:

if $i = WBC$ **then**

WBC = {WBC} + $\{r_k\}$ {WBC is the set of identified WBC}

else if $i = RBC$ **then**

RBC = {RBC} + $\{r_k\}$ {RBC is the set of identified RBC}

end if

end for

Step2:

for all RBC **do**

-Calculate BPA on HSV space

if $m(RBC) < m(WBC)$ **then**

WBC = {WBC} + {RBC}

RBC = {RBC} - {RBC}

end if

end for

Step3:

for all WBC **do**

-Calculate BPA on green channel {Nucleus and Cytoplasm hypothesis}

if $m(Nucleus) > m(Cytoplasm)$ **then**

NUCLEUS = {NUCLEUS} + {*Nucleus*}

else

CYTOPLASM = {CYTOPLASM} + {*Cytoplasm*}

end if

end for

The proposed algorithm was based on a hierarchical order that allowed to segment the different components. The main ideas of our three steps are as below:

1. We associated to each labeled region a degree of belief for the hypothesis WBC and RBC. In most cases, we noticed that the magenta channel from the CMYK colour space allows to differentiate RBCs from others cells (see figure 4: RBCs are lighter than the WBC). However, in some cases the cytoplasm is confused with RBC.
2. New projection was performed on the HSV colour space to check the certainty of the identified RBC. (see figure 5: Cytoplasm of cell on top is similar to RBC, but difference is clearly made in the HSV

projection). Then, the following heuristic was applied: **if** $m(\text{WBC}) > m(\text{RBC})$, then the region was reclassified as WBC, **Else**, it was left as RBC region.

3. For each identified WBC, the BPAs on the green channel was computed to separate nucleus and cytoplasm.

2.2 wbc's characterization

The plasma cells are round or oval shaped. they are large with a considerable nucleus-to-cytoplasm ratio. The nucleus too is oval in shape and some what eccentric. The cytoplasm is basophilic, and contains a pale zone (appears white and distinct).

In automatic leukocytes classification system, most of the researchers used both texture features and geometric features to characterize the white blood cell, and by the way increase the accuracy of the classifier [6]. In our work, we used the following shape and texture features:

1. Shape features:

- *Eccentricity*: The ratio of the major and minor axes length

$$Eccentricity = \frac{Length\ of\ Major\ Axis}{Length\ of\ Minor\ Axis} \quad (4)$$

- *Solidity*: The ratio of area to convex area

$$Solidity = \frac{Area}{Convex\ Area} \quad (5)$$

- *Compactness*: The ratio of area and square of the perimeter

$$Compactness = \frac{Area}{Perimeter^2} \quad (6)$$

- *Circularity*: The circularity is defined by the ratio of the tightest bounding circle perimeter to the perimeter.
- *Rectangularity*: It is the ratio between the perimeter of the tightest bounding rectangle and the perimeter.

- *Form Factor*: It is a dimensionless parameter.

$$Form\ factor = \frac{4 * \pi * Area}{Perimeter^2} \quad (7)$$

2. Texture features:

- *Mean*:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (8)$$

where x_i is the value of the i^{th} pixel, and n is the total number of pixels.

- *Variance*: It is the average of the squared differences from the mean

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} \quad (9)$$

- *Standard deviation*: The square root of the variance

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (10)$$

- *Entropy*: It is used to measure the randomness

$$Entropy = - \sum_i \sum_j p_d(i, j) \log p_d(i, j) \quad (11)$$

with $p_d(i, j)$ represents the number of occurrences of the pair (i, j) having apart by a distance d .

To help the clinicians, we have tried to consider both medical informations and technical aspect of white blood cell characterization. We have taken the shape feature of the whole cell, afterward, we have considered both shape and texture features of the nucleus. Also, we have extracted the texture features of cytoplasm. Taking into account the colour projection, we have calculated the texture on the magenta channel and on the green channel. Finally, we have added some plasma cell aspect: the ratio of nucleus to cell and the differences on both magenta and green channel of cytoplasm. And then, we have extracted 31 features, as follows:

- Cell shape features.

- Nucleus shape features.
- Nucleus texture features on both magenta and green channels.
- Cytoplasm texture features on both magenta and green channels.
- Ratio of nucleus to cell.
- Difference of the maximum value and the minimum value in cytoplasm, of both magenta and green channels.

2.3 wbc classification

The main goal of this work is to design an automatic system to recognize the plasma cells in a microscopic bone marrow images. Leukocytes classification is the complex process for computer systems as compared to human observer, which has motivated the researchers to study this field in the perspective of artificial intelligence [6]. In our work, we have been interested by three supervised machine learning algorithms: Support Vector Machine (SVM) [33], K-Nearest-Neighbor (KNN), and Decision Tree. This choice is justified by the efficiency of the SVM for binary classification, and the simplicity and the robustness of the Knn and the decision tree.

3 Results and discussion

The processed data base is constructed from 87 bone marrow images. We have acquired those images from four different patients suffering from myeloma cancer. The bone marrow blades have been fixed by the MGG (May-Grunwald Giemsa) colouration. We have obtained 24-bits RGB images of 1024×768 pixels by the use of LEICA environment (camera and microscope with a 1000x magnification). The segmentation results and the classification results of our method, are presented in section 3.1 and section 3.2, respectively.

3.1 Segmentation results

We present four bone marrow images to illustrate the obtained results of the proposed algorithm (see figure 7 (a)-1 to (a)-4).

In these experiments, the goal was to identify the red blood cell regions and the white blood cell regions, especially, the white blood cells since, it remains important, for the diagnosis of myeloma cancer.

We handle with several kinds of white blood cells, the difference concern the shape of nucleus and the colour of cytoplasm. In figure 7, the shape of nucleus is circular or oval in images (a)-1, bottom (a)-2, top and middle (a)-3,(a)-4, and has kidney shape in top (a)-2. Regarding the colour of cytoplasm, one can notice the dark and the light cytoplasm, in some cases within the same image (for example, image (a)-2).

The obtained segmentation results compared to the ground truth are good enough, especially in the case of images (b)-1 and (b)-4: cells were totally recognized. The proposed algorithm provides good results in the case of overlapping cells in all images (RBC vs WBC and WBC vs WBC), except in image (b)-3 where the WBC in the middle is missclassified. Furthermore, some limitations can be observed: a small part nucleus has been identified as cytoplasm in images (b)-2 (the top one). However, it does not affect the identification of the WBC.

The plasma cell is present in images: bottom of (a)-2, (a)-4, and middle of (a)-3. We notice, the plasma cell is present in normal and distorted shape, fortunatly, the two shapes were well recognized, which is a promising results.

Segmentation is an important step of plasma cell recognition and highly affects the performances of the classifiers. We evaluate the segmentation quality to determine the similarity between the ground truth data and the segmentation results. We calculate two metrics: Dice score [34] and Jaccard similarity [35]; those metrics, were first developed to measure similarities in ecological studies, but their nature of set operations made them applicable for segmentation similarity.

The Dice score is a spatial overlap index and a reproducibility validation metric it also known as the mean overlap. The Jaccard similarity is the ratio of the intersection of segmented area A and ground truth area B, to the union of segmented area A and ground truth area B. It takes the qualitative aspect overlap into account. The corresponding formulas are given below:

$$Dice\ score = 2 * \frac{TP}{((TP + FP) + (TP + FN))} \quad (12)$$

$$Jaccard\ similarity = \frac{TP}{(FP + TP + FN)} \quad (13)$$

where, TP (True Positives): intersection between segmentation and ground truth), FP (false positives): segmented parts not overlapping the ground truth, FN (false negatives): missed parts of the ground truth, and TN (true negatives): part of the image beyond the union between segmentation and ground truth).

These two metrics values range from 0, indicating no spatial overlap between two sets of segmentation results, to 1, indicating complete overlap. When applied to all our data, the Dice score and the Jaccard

similarity have achieved a promising results: 0.93 and 0.87, respectively.

We have compared our work with (SDR: Segmentation based on Dimensionality Reduction) [36], since it concerns the same data base; they worked on only 27 images. In the present work, we process 87 images, we have compared both methods with dice score and the jaccard similarity. In table 1, it is clear that ESA segmentation is better than SDR segmentation. We emphasize on the fact that both methods work with colour informations. The table 1 shows the power of evidence theory. Since it is able to handle with uncertainty resulting from imprecise, incomplete, and complementary informations. In addition, the segmentation is fast enough, indeed, the whole process take 5 seconds.

Table 1: Dice score and Jacard similarity comparison

	ESA segmentation	SDR segmentation
Dice score	0.97	0.74
Jaccard similarity	0.94	0.59

3.2 Characterization and classification results

A vector of 31 features was extracted for every segmented white blood cell. The resulting dataset has been used to identify plasma cells using SVM, Knn and Tree decision. The classifiers were evaluated using a k-fold cross-validation technique, with differents value of k ; $k = \{5, 7, 10, 15\}$. After validation, the three classifiers were tested on the whole data set.

We have evaluated the classifiers in terms of:

- correct rate= $TP/Total\ number$
- sensitivity= $TP/(TP+FN)$
- specificity= $TN/(TN+FP)$
- precision= $TP/(TP+FP)$
- F-score= $2*((precision*sensitivity)/(precision+sensitivity))$

Where TP (true positive) are the number of cells which are plasma cells and detected as plasma cells, FN (false negative) number of cells which are plasma cells but are detected as other cells, TN (truc negative) are the number of cells which are other cells and detected as other cells, and FP (false positive) are the number of cells which are other cells but are detected plasma cells. Therefore, the measures were

calculated in order to well evaluate the performances of the automatic classification of plasma cells. Since, sensitivity is the probability that a cell will be detected plasma cell, in case it is plasma cell. Specificity is the probability that a cell will be detected another cell, in case it is another WBC. Precision is the probability that the detection of plasma cell is actually associated with a really plasma cell. The F-score is the balance between precision and sensitivity.

We observed, (see tab. 2, tab. 3 and tab. 4), that a good performances were obtained for all the classifiers, with an interesting results for the knn classifier where both specificity and precision are equal to 1. Therefore, It is able to avoid false alarms, and a good enough of detecting positifs cases. In the state of art, works on the blood smear was more important than those on the bone marrow. Compraed to our method, although the difference of the target white blood cell and the processed data base, we have observed on the same kind of images (bone marrow), the authors in [15] had achieved 78.51% of correct rate on gray-scale image with the Knn classifier.

Table 2: SVM, Knn and Tree Performances for k=5

Classifier	Correct Rate	Sensitivity	Specificity	Precision	F-score
SVM	0.92	0.92	0.93	0.92	0.92
Knn	0.95	0.92	0.97	0.97	0.97
Tree	0.95	0.94	0.95	0.94	0.94

Table 3: SVM, Knn and Tree Performances for k=7

Classifier	Correct Rate	Sensitivity	Specificity	Precision	F-score
SVM	0.94	0.92	0.95	0.94	0.93
Knn	0.96	0.92	1	1	0.96
Tree	0.92	0.89	0.95	0.94	0.92

Table 4: SVM, Knn and Tree Performances for k=10

Classifier	Correct Rate	Sensitivity	Specificity	Precision	F-score
SVM	0.94	0.92	0.95	0.94	0.93
Knn	0.97	0.94	1	1	0.97
Tree	0.94	0.89	0.97	0.97	0.93

To further justify our Plasma cell recognition, we have chosen the ROC (Receiver Operating Characteristics) curve to evaluate our classifiers (see figure 8). The ROC curve is famous for its insensitivity to

the data distribution, so, it provides an objective analysis of the classifier. Furthermore, the AUC (Area Under the Curve) permit to respond of its discrimination between classes, when it equal to 1 the classifier is perfect and when it equal to 0.5 it is a non-discriminant classifier. In our case the AUC is equal to 0.94, 0.97 and 0.93 for SVM, Knn and Tree decision, respectively. Those values are considered as promising results.

We have choosen the WBC characterization near to the clinician's description, which is important for results interpretability. Moreover, since the specificity (for the Knn) is equal to 1, i.e. the system avoid false alarms, the clinician is more confident in the automatic system.

Considering all the achievements, we estimate that, the Knn, despite his simplicity, has potential to well identify number of plasma cells (AUC=0.97) and it is able to avoid false alarms (specificity=1). This ability is useful in treatment process, since the percentage of number of plasma cells to all cells proves myeloma cancer. Moreover, cell classification need 0.45 second. Finally, our method take approximately 6 second to process an image, which is an asset to convince clinician.

4 Conclusion

In this paper, we propose an automatic system for plasma cells recognition in microscopic bone marrow images. The microscopic bone marrow image is taken as input, to be initially segmented with the watershed transform, and then the application of evidence theory based on colour uncertainty integration to segment RBC and WBC cells. Afterward, the segmented WBC is characterized with shape and colour texture features closely to clinician description. Those features are, finally, the input of three different classifiers.

We have obtained a good results for segmentation and, specially, classification with a specificity and precision equal to 1 for the Knn classifier. The robustness and generalization of the system can be improved by using an unified data base. Until now, the segmentation step keep more interest, indeed the integration of shape factor is under work to improve the quality of the segmentation.

References

- [1] Rajkumar and S Vincent. Multiple myeloma: 2013 update on diagnosis, risk-stratification, and management. *American journal of hematology*, 88(3):226–235, March 2013.
- [2] Jon Plumley, Richard Gatward, Alison Todman, Raouf NG Naguib, and Mark K Bennett. An experiment in the reporting of observed features of cancer images to aid in the development of an

- image description language for pathologists. *International Journal of Biomedical Engineering and Technology*, 3(3-4):319–328, 2010.
- [3] Robiyanti Adollah, MY Mashor, NF Mohd Nasir, H Rosline, H Mahsin, and H Adilah. Blood cell image segmentation: a review. In *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, pages 141–144. Springer, 2008.
- [4] Luis E Hasbon Reyes, Lola X Bautista Rozo, and Fernando A Rojas Morales. Automatic leukocyte image segmentation: A review. In *Signal Processing, Images and Computer Vision (STSIWA), 2015 20th Symposium on*, pages 1–9. IEEE, 2015.
- [5] Mashiat Fatma and Jaya Sharma. A survey on image segmentation techniques used in blood cancer detection. *Int Journal of Engineering Research and Applications www. ijera. com ISSN*, pages 2248–9622, 2014.
- [6] Mukesh Saraswat and KV Arya. Automated microscopic image analysis for leukocytes identification: A survey. *Micron*, 65:20–33, 2014.
- [7] P.S.Hiremath, Parashuram Bannigidad, and Sai Geeta. Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. *IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition"*, pages 59–63, 2010.
- [8] Ramesh Nisha, Dangott Bryan, E. Salama Mohammed, and Tolga Tasdizen. Isolation and two-step classification of normal white blood cells in peripheral blood smears. *Journal of pathology informatics*, 3(1), march 2012.
- [9] SS Savkare and SP Narote. Blood cell segmentation from microscopic blood images. In *Information Processing (ICIP), 2015 International Conference on*, pages 502–505. IEEE, 2015.
- [10] Seyed Hamid Rezatofghi and Hamid Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, 35:333–343, 2011.
- [11] Jaroonrut Prinyakupt and Charnchai Pluempitiwiriwawej. Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers. *Biomedical engineering online*, 14(1):63, 2015.

- [12] C. Kalaiselvi and R. Asokan. A classification of chronic leukaemia using new extension of k-means clustering and efm based on digital microscopic blood images. *Int. J. of Biomedical Engineering and Technology*, 23(2/3/4):232–241, 2017.
- [13] Congcong Zhang, Xiaoyan Xiao, Xiaomei Li, Ying-Jie Chen, Wu Zhen, Jun Chang, Chengyun Zheng, and Zhi Liu. White blood cell segmentation by color-space-based k-means clustering. *Sensors*, 14(9):16128–16147, 2014.
- [14] Der-Chen Huang, Kun-Ding Hung, and Yung-Kuan Chan. A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images. *The Journal of Systems and Software*, 85:2104–2118, 2012.
- [15] Leyza Baldo Dorini, Rodrigo Minetto, and Neucimar Jerônimo Leite. Semiautomatic white blood cell segmentation based on multiscale analysis. *IEEE journal of biomedical and health informatics*, 17(1):250–256, 2013.
- [16] Eric Dahai Cheng, Subhash Challa, and Rajib Chakravorty. Parallel microscopic cell image segmentation and multiple fusions. *International Journal of Signal and Imaging Systems Engineering*, 4(2):96–114, 2011.
- [17] Zhi Liu, Jing Liu, Xiaoyan Xiao, Hui Yuan, Xiaomei Li, Jun Chang, and Chengyun Zheng. Segmentation of white blood cells through nucleus mark watershed operations and mean shift clustering. *sensors*, 15(9):22561–22586, 2015.
- [18] Hassan Khajepour, Alireza Mehri Dehnavi, Hossein Taghizad, Esmat Khajepour, and Mohammadreza Naeemabadi. Detection and segmentation of erythrocytes in blood smear images using a line operator and watershed algorithm. *Journal of medical signals and sensors*, 3(3):164, 2013.
- [19] Tomasz Markiewicz, Stanislaw Osowski, Bonenza Marianska, and Leszek Moszczynski. Automatic recognition of the blood cells of myelogenous leukemia using svm. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 4, pages 2496–2501. IEEE, 2005.
- [20] Tomasz Markiewicz, Stanisław Osowski, and Bożena Mariańska. White blood cell automatic counting system based on support vector machine. *Adaptive and Natural Computing Algorithms, Springer*, pages 318–326, 2007.

- [21] Pramit Ghosh, Debotosh Bhattacharjee, and Mita Nasipuri. Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique. *Applied Soft Computing*, 2016.
- [22] Ismahan Baghli, Amir Nakib, Elic Sellam, Mourtada Benazzouz, Amine Chikh, and Eric Petit. Hybrid framework based on evidence theory for blood cell image segmentation. *In proc. SPIE 9038, Medical Imaging 2014, San Diego, (USA)*, Feb. 15-19, 2014, <http://dx.doi.org/10.1117/12.2042142>.
- [23] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Machine Intell.*, 13:583–598, 1991.
- [24] Yu Jinhua and Tan Jinglu. Object density-based image segmentation and its applications in biomedical image analysis. *computer methods and programs in biomedicine*, pages 193–204, 2009.
- [25] Eric Dahai Cheng, Subhash Challa, and Rajib Chakravorty. Multi-channel and multi-layer cell segmentations and decision fusion based on an hierarchical structure. *International Journal of Biomedical Engineering and Technology*, 6(1):14–45, 2011.
- [26] Almond DSouza and V Seenivasagam. Colour image segmentation using region-growing and region-merging methods. *International Journal of Biomedical Engineering and Technology*, 7(2):165–173, 2011.
- [27] Olivier Lezoray, Abderrahim Elmoataz, Hubert Cardot, Gilles Gougeon, Michel Lecluse, Hubert Elie, and Marinette Revenu. Segmentation of cytological image using color and mathematical morphology. In *European conference on Stereology*, pages 10–pp, 1998.
- [28] Shafer Glenn. A mathematical theory of evidence. *Princeton University Press*, 1976.
- [29] Chunlei Yang, Jiexin Pu, Yongsheng Dong, Zhonghua Liu, Lingfei Liang, and Xiaohong Wang. Salient object detection in complex scenes via d-s evidence theory based region classification. *The Visual Computer*, pages 1–14, 2016.
- [30] Guan J. W. and Bell D. A. Evidence theory and its applications. *North-Holland, New York*, 1991.
- [31] Thierry Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5):804–813, 1995.

- [32] Appriou A. Multisensor signal processing in the framework of the theory of evidence. *NATO/RTO Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, page 531, 1999.
- [33] Vapnik V. Statistical learning theory. *John Wiley and Sons, Chichester*, 1998.
- [34] Dice L. R. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [35] Jaccard P. Etude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la socit Vaudoise des Sciences Naturelles*, 37:547–579, 1901.
- [36] Benazzouz Mourtada, Baghli Ismahan, and Med. Amine Chikh. Microscopic image segmentation based on pixel classification and dimensionality reduction. *International Journal of Imaging Systems and Technology*, 23:22–28, March 2013.

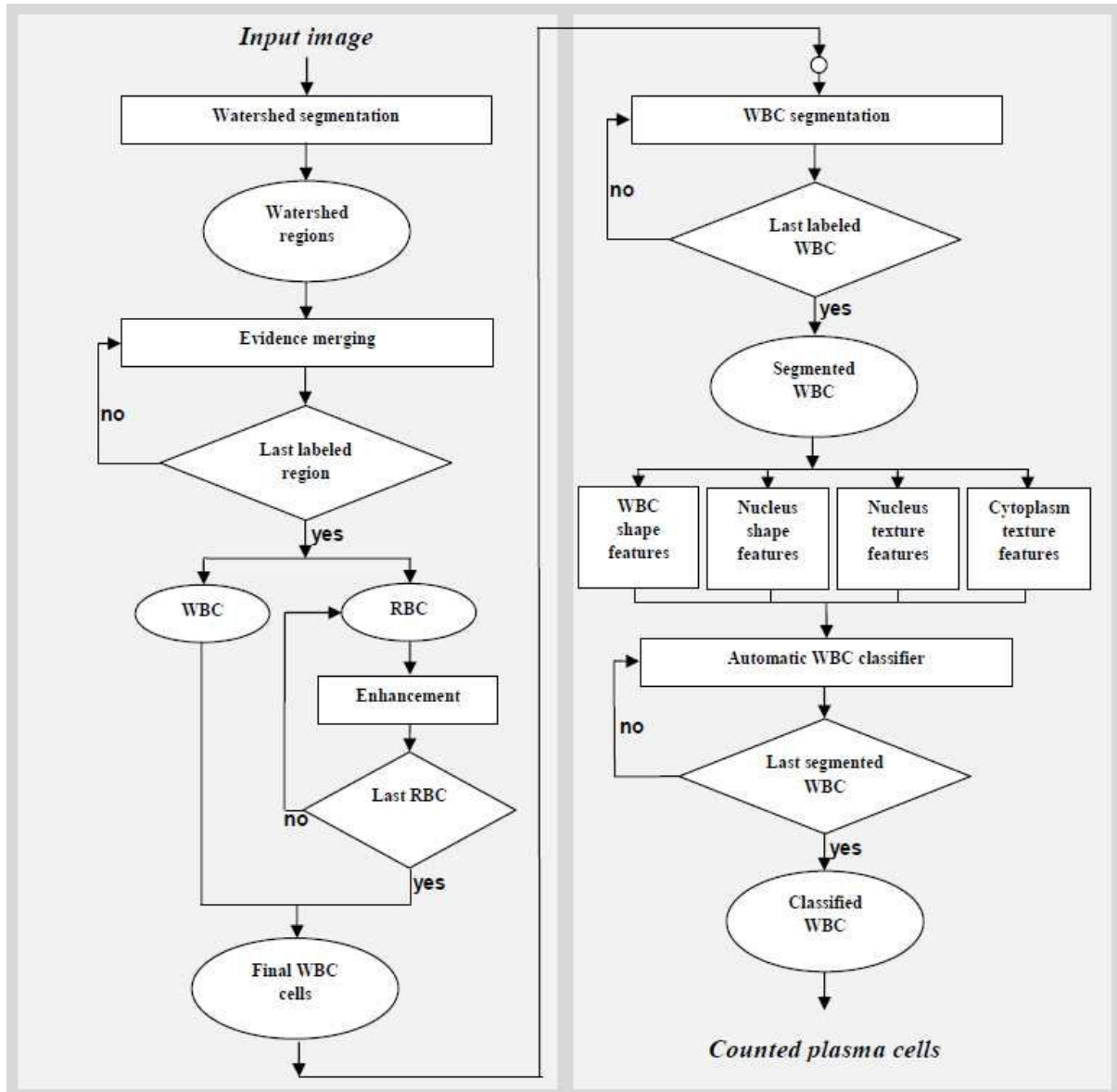


Figure 2: The proposed method scheme

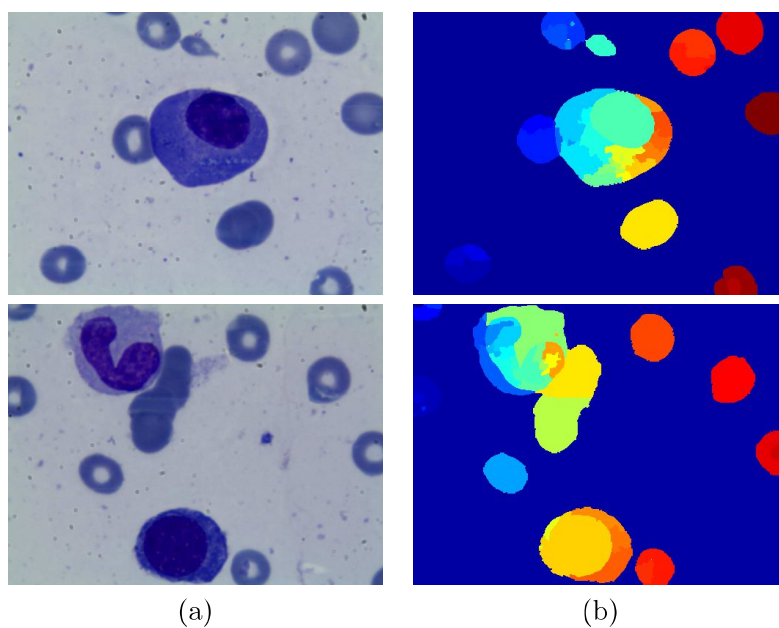


Figure 3: (a) Original images (b) Watershed regions

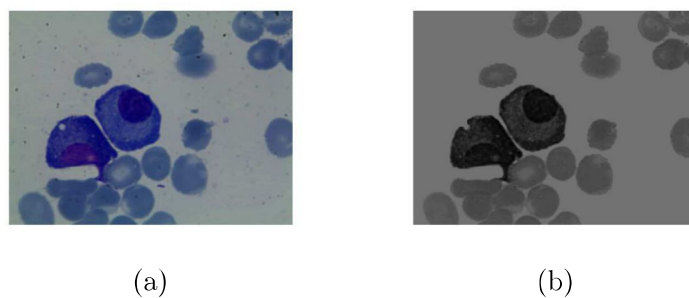


Figure 4: (a) original image (b) image projection on Magenta channel (CMYK SPACE)

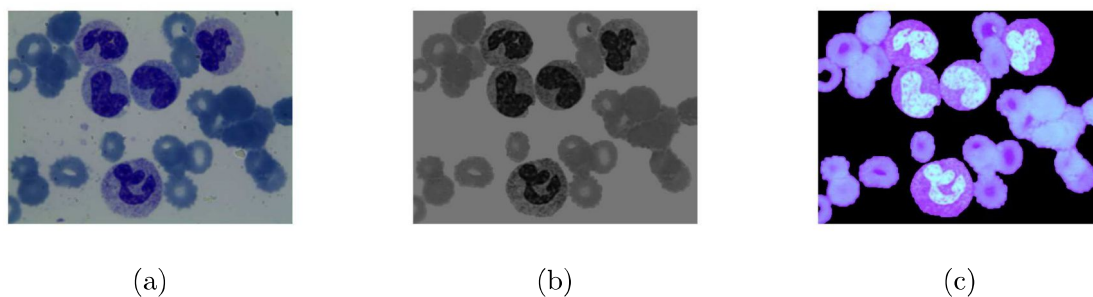


Figure 5: Illustration of the projection of the Microscopic dataset in the different colour spaces. (a) original image (b) image projection on Magenta channel (c) image projection on HSV colour space

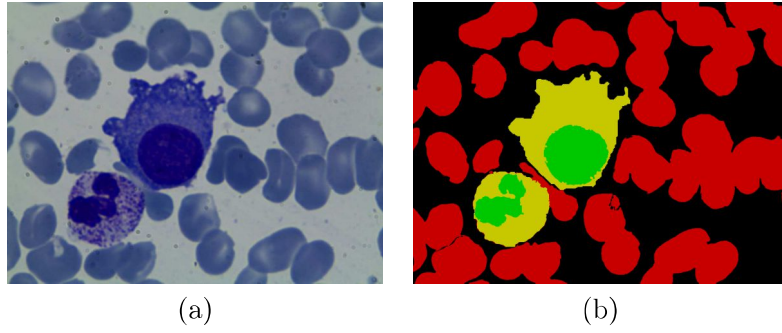


Figure 6: (a) Microscopic blood image (b) segmented image

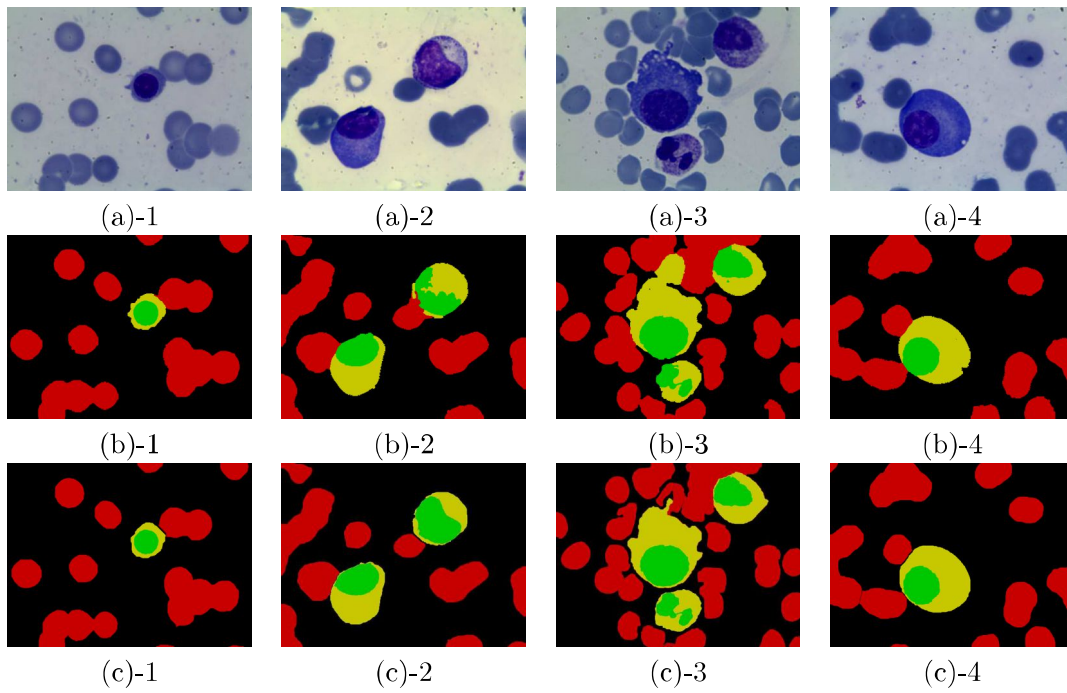


Figure 7: ESA results: (a) original image (b) ESA segmentation (c) Ground truth

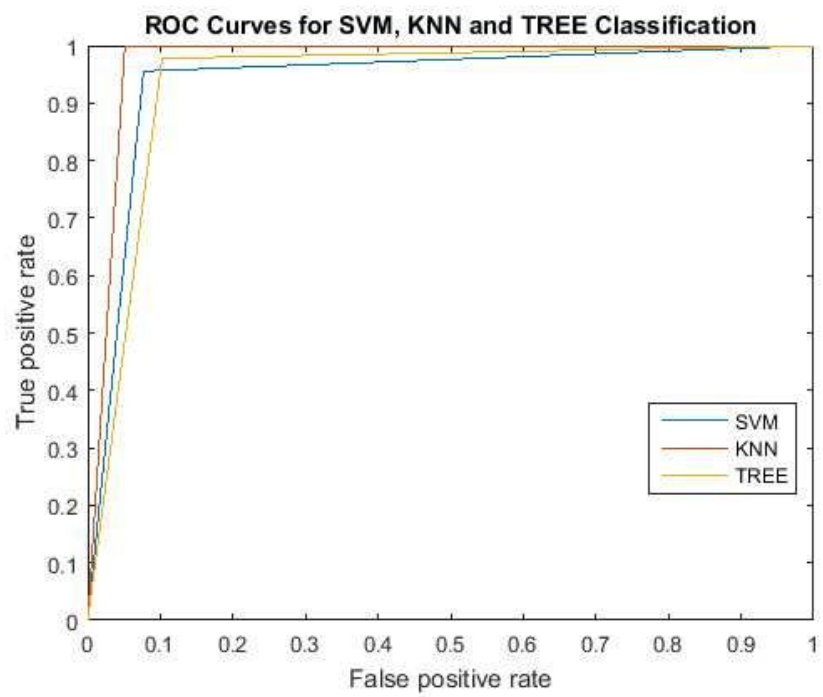


Figure 8: ROC curves of the Plasma cell recognition